Probability Theory and Stochastic Modelling 99

Olav Kallenberg

Foundations of Modern Probability

Third Edition

Springer

Probability Theory and Stochastic Modelling

Volume 99

The Probability Theory and Stochastic Modelling series is a merger and continuation of Springer's two well established series Stochastic Modelling and Applied Probability and Probability and Its Applications. It publishes research monographs that make a significant contribution to probability theory or an applications domain in which advanced probability methods are fundamental. Books in this series are expected to follow rigorous mathematical standards, while also displaying the expository quality necessary to make them useful and accessible to advanced students as well as researchers. The series covers all aspects of modern probability theory including.

- Gaussian processes
- Markov processes
- Random Fields, point processes and random sets
- Random matrices
- Statistical mechanics and random media
- Stochastic analysis

as well as applications that include (but are not restricted to):

- Branching processes and other models of population growth
- Communications and processing networks
- Computational methods in probability and stochastic processes, including simulation
- Genetics and other stochastic models in biology and the life sciences
- Information theory, signal processing, and image synthesis
- Mathematical economics and finance
- Statistical methods (e.g. empirical processes, MCMC)
- Statistics for stochastic processes
- Stochastic control
- Stochastic models in operations research and stochastic optimization
- Stochastic models in the physical sciences

More information about this series at http://www.springer.com/series/13205

Olav Kallenberg

Foundations of Modern Probability

Third Edition

Springer

Olav Kallenberg
Auburn University
Auburn, AL, USA

ISSN 2199-3130 ISSN 2199-3149 (electronic)
Probability Theory and Stochastic Modelling
ISBN 978-3-030-61870-4 ISBN 978-3-030-61871-1 (eBook)
https://doi.org/10.1007/978-3-030-61871-1

Mathematics Subject Classification: 60-00, 60-01, 60A10, 60G05

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface to the First Edition

Some thirty years ago it was still possible, as Loève so ably demonstrated, to write a single book in probability theory containing practically everything worth knowing in the subject. The subsequent development has been explosive, and today a corresponding comprehensive coverage would require a whole library. Researchers and graduate students alike seem compelled to a rather extreme degree of specialization. As a result, the subject is threatened by disintegration into dozens or hundreds of subfields.

At the same time the interaction between the areas is livelier than ever, and there is a steadily growing core of key results and techniques that every probabilist needs to know, if only to read the literature in his or her own field. Thus, it seems essential that we all have at least a general overview of the whole area, and we should do what we can to keep the subject together. The present volume is an earnest attempt in that direction.

My original aim was to write a book about "everything." Various space and time constraints forced me to accept more modest and realistic goals for the project. Thus, "foundations" had to be understood in the narrower sense of the early 1970s, and there was no room for some of the more recent developments. I especially regret the omission of topics such as large deviations, Gibbs and Palm measures, interacting particle systems, stochastic differential geometry, Malliavin calculus, SPDEs, measure-valued diffusions, and branching and superprocesses. Clearly plenty of fundamental and intriguing material remains for a possible second volume.

Even with my more limited, revised ambitions, I had to be extremely selective in the choice of material. More importantly, it was necessary to look for the most economical approach to every result I did decide to include. In the latter respect, I was surprised to see how much could actually be done to simplify and streamline proofs, often handed down through generations of textbook writers. My general preference has been for results conveying some new idea or relationship, whereas many propositions of a more technical nature have been omitted. In the same vein, I have avoided technical or computational proofs that give little insight into the proven results. This conforms with my conviction that the logical structure is what matters most in mathematics, even when applications is the ultimate goal.

Though the book is primarily intended as a general reference, it should also be useful for graduate and seminar courses on different levels, ranging from elementary to advanced. Thus, a first-year graduate course in measure-theoretic probability could be based on the first ten or so chapters, while the rest of the book will readily provide material for more advanced courses on various topics. Though the treatment is formally self-contained, as far as measure theory and probability are concerned, the text is intended for a rather sophisticated reader with at least some rudimentary knowledge of subjects like topology, functional analysis, and complex variables.

My exposition is based on experiences from the numerous graduate and seminar courses I have been privileged to teach in Sweden and in the United States, ever since I was a graduate student myself. Over the years I have developed a personal approach to almost every topic, and even experts might find something of interest. Thus, many proofs may be new, and every chapter contains results that are not available in the standard textbook literature. It is my sincere hope that the book will convey some of the excitement I still feel for the subject, which is without a doubt (even apart from its utter usefulness) one of the richest and most beautiful areas of modern mathematics.

Preface to the Second Edition

For this new edition the entire text has been carefully revised, and some portions are totally rewritten. More importantly, I have inserted more than a hundred pages of new material, in chapters on general measure and ergodic theory, the asymptotics of Markov processes, and large deviations. The expanded size has made it possible to give a self-contained treatment of the underlying measure theory and to include topics like multivariate and ratio ergodic theorems, shift coupling, Palm distributions, entropy and information, Harris recurrence, invariant measures, strong and weak ergodicity, Strassen's law of the iterated logarithm, and the basic large deviation results of Cramér, Sanov, Schilder, and Freidlin and Ventzel.[1]

Unfortunately, the body of knowledge in probability theory keeps growing at an ever increasing rate, and I am painfully aware that I will never catch up in my efforts to survey the entire subject. Many areas are still totally beyond reach, and a comprehensive treatment of the more recent developments would require another volume or two. I am asking for the reader's patience and understanding.

Preface to the Third Edition

Many years have passed since I started writing the first edition, and the need for a comprehensive coverage of modern probability has become more urgent than ever. I am grateful for the opportunity to publish a new, thoroughly revised and expanded edition. Much new material has been added, and there are even entirely new chapters on subjects like Malliavin calculus, multivariate arrays, and stochastic differential geometry (with so much else still missing). To facilitate the reader's access and overview, I have grouped the material together into ten major areas, each arguably indispensable to any serious student and researcher, regardless of area of specialization.

To me, every great mathematical theorem should be a revelation, prompting us to exclaim: "Wow, this is just amazing, how could it be true?" I have spent countless hours, trying to phrase every result in its most striking form, in my efforts to convey to the reader my own excitement. My greatest hope is that the reader will share my love for the subject, and help me to keep it alive.

[1]I should have mentioned Varadhan, one of the giants of modern probability.

Acknowledgments

Throughout my career, I have been inspired by the work of colleagues from all over the world, and in many cases by personal interaction and friendship. As in earlier editions, I would like to mention especially the early influence of my mentor Peter Jagers and the stimulating influences of the late Klaus Matthes, Gopi Kallianpur, and Kai Lai Chung. Other people who have inspired me through their writing and personal contacts include especially David Aldous, Sir John Kingman, and Jean-François LeGall. Some portions of the book have been stimulated by discussions with my AU colleague Ming Liao.

Countless people from all over the world, their names long forgotten, have written to me through the years to ask technical questions or point out errors, which has led to numerous corrections and improvements. Their positive encouragement has been a constant source of stimulation and joy, especially during the present pandemic, which sadly affects us all. Our mathematical enterprise (like the coronavirus) knows no borders, and to me people of any nationality, religion, or ethnic group are all my sisters and brothers. One day soon we will emerge even stronger from this experience, and continue to enjoy and benefit from our rich mathematical heritage[2].

I am dedicating this edition to the memory of my grandfathers whom I never met, both idolized by their families:

Otto Wilhelm Kallenberg — Coming from a simple servant family, he rose to the rank of a trusted servant to the Swedish king, in the days when kings had real power. Through his advance, the family could afford to send their youngest son Herbert, my father, to a Latin school, where he became the first member of the Kallenberg family to graduate from high school.

Olaf Sund (1864–1941) — Coming from a prominent family of lawyers, architects, scientists, and civil servants, and said to be the brightest child of a big family, he opted for a simple life as *Lensmann* in the rural community of Steigen north of the arctic circle. He died tragically during the brave resistance to the Nazi occupation of Norway during WWII. His youngest daughter Marie Kristine, my mother, became the first girl of the Sund family to graduate from high school.

Olav Kallenberg
October 2020

[2]To me the great music, art, and literature belong to the same category of cultural treasures that we hope will survive any global crises or onslaughts of selfish nationalism. Every page of this book is inspired by the great music from Bach to Prokofiev, the art from Rembrandt to Chagall, and the literature from Dante to Con Fu. What a privilege to be alive!

Words of Wisdom and Folly

- ♣ "A mathematician who argues from probabilities in geometry is not worth an ace" — *Socrates* (on the demands of rigor in mathematics)
- ♣ "[We will travel a road] full of interest of its own. It familiarizes us with the measurement of variability, and with curious laws of chance that apply to a vast diversity of social subjects" — *Francis Galton* (on the wondrous world of probability)
- ♣ "God doesn't play dice" [i.e., there is no randomness in the universe] — *Albert Einstein* (on quantum mechanics and causality)
- ♣ "It might be possible to prove certain theorems, but they would not be of any interest, since in practice one could not verify whether the assumptions are fulfilled" — *Émile Borel* (on why bother with probability)
- ♣ "[The stated result] is a special case of a very general theorem [the strong Markov property]. The measure [theoretic] ideas involved are somewhat glossed over in the proof, in order to avoid complexities out of keeping with the rest of this paper" — *Joseph L. Doob* (on why bother with generality or mathematical rigor)
- ♣ "Probability theory [has two hands]: On the right is the rigorous [technical work]; the left hand ... reduces problems to gambling situations, coin-tossing, motions of a physical particle" — *Leo Breiman* (on probabilistic thinking)
- ♣ "There are good taste and bad taste in mathematics just as in music, literature, or cuisine, and one who dabbles in it must stand judged thereby" — *Kai Lai Chung* (on the art of writing mathematics)
- ♣ "The traveler often has the choice between climbing a peak or using a cable car" — *William Feller* (on the art of reading mathematics)
- ♣ "A Catalogue Aria of triumphs is of less benefit [to the student] than an indication of the techniques by which such results are achieved" — *David Williams* (on seductive truths and the need of proofs)
- ♣ "One needs [for stochastic integration] a six months course [to cover only] the definitions. What is there to do?" — *Paul-André Meyer* (on the dilemma of modern math education)
- ♣ "There were very many [bones] in the open valley; and lo, they were very dry. And [God] said unto me, 'Son of man, can these bones live?' And I answered, 'O Lord, thou knowest.'" — *Ezekiel 37:2–3* (on the reward of hard studies, as quoted by *Chris Rogers* and *David Williams*)

Contents

Introduction and Reading Guide

In my youth, a serious student of probability theory[1] was expected to master the basic notions and results in all areas of the subject[2]. Since then the development has been explosive, leading to a rather extreme fragmentation and specialization. I believe that it remains essential for any serious probabilist to have a good overview of the entire subject. Everything in this book, from the overall organization to the choice of notation and displays of theorems, has been prepared with this goal in mind.

The first thing you need to know is that the study of any more advanced text in mathematics[3] requires an approach different from the usual one. When reading a novel you start reading from page 1, and after a few days or weeks you come to the end, at which time you will be familiar with all the characters and have a good overview of the plot. This approach never works in math, except for the most elementary texts. Instead it is crucial to adopt a *top-down approach*, where you first try to acquire a general overview, and then gradually work your way down to the individual theorems, until finally you reach the level of proofs and their logical structure.

The inexperienced reader may feel tempted to skip the proofs, trying instead a few of the exercises. I would rather suggest the opposite. It is from the proofs you learn *how to do math*, which may be regarded as a major goal of the graduate studies. If you forgot the precise conditions in the statement of a theorem, you can always look them up, but the only way to learn how to prove your own theorems is to gather experience by studying dozens or hundreds of proofs. Here again it is important to adopt a *top-down approach*, always starting to look for the crucial ideas that make the proof 'work', and then gradually breaking down the argument into minor details and eventually perhaps some calculation. Some details are often left to the reader[4], suggesting an abundance of useful and instructive exercises, to identify and fill in all the little gaps implicit in the text.

The book has now been divided into ten parts of 3–4 chapters each, and I think it is essential for any serious probabilist to be familiar with at least some material in each of those parts. Every part is preceded by a short introduction, where I indicate the contents of the included chapters and make suggestions about the choice of material to study in further detail. The precise selection may be less important, since you can always return to the omitted material when need arises. Every chapter also begins with a short introduction, highlighting some of the main results and indicating connections to material in

[1] henceforth regarded as synonymous with the theory of stochastic processes

[2] When I was a student, we had no graduate courses, only a reading list, and I remember studying a whole summer for my last oral exam, covering every theorem and proof in Loève (1963) plus half of Doob (1953), totaling some 1,000 pages.

[3] What is advanced to some readers may be elementary to others, depending on background.

[4] or else even the simplest proof would look like a lengthy computer program, and you would lose your overview

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_1

other chapters. You never need to worry about making your selection self-contained, provided you are willing occasionally to use theorems whose proofs you are not yet familiar with[5]. A similar remark applies to the axiom of choice and its various equivalents, which are used freely without comments[6].

– – –

Let me conclude with some remarks about the contents of the various parts. Part I gives a rather complete account of the basic measure theory needed throughout the remainder of the book. In my experience, even a year-long course in real analysis may be inadequate for our purposes, since many theorems of special importance in probability theory may not even be mentioned in standard real analysis texts, where the emphasis is often rather different. When teaching graduate courses in advanced probability, then regardless of background[7] of the students enrolled, I am always beginning with a few weeks of general measure theory, where the students will also get used to my top-down approach.

Part II provides the measure-theoretic foundations of probability theory, involving the notions of *processes, distributions*, and *independence*. Here we also develop the classical limit theory for random sums and averages, including the law of large numbers and the central limit theorem, along with a wealth of extensions and ramifications. Finally, we give a comprehensive treatment of the classical limit theory for *null arrays* of random variables and vectors, which will later play a basic role for the discussion of Lévy and related processes. Much of the discussion is based on the theory of *characteristic functions*[8] and *Laplace transforms*, which will later regain importance in connection with continuous martingales, random measures, and potential theory.

Modern probability theory can be said to begin with the closely related theories of *conditioning, martingales*, and *compensation*, which form the subjects of Part III. The importance of this material can hardly be overstated, since methods based on conditioning are constantly used throughout probability theory, and martingale theory provides some of the basic tools for proving a wide range of limit theorems and path properties. Indeed, the *convergence* and *regularization theorem* for continuous-time martingales has been regarded as the single most important result in all of probability theory. The theory of compensators, based on the powerful *Doob–Meyer decomposition*, plays an equally basic role in the study of processes with jump discontinuities.

Beside martingales, the classes of *Markov* and related processes represent

[5]A prime example is the existence of *Lebesgue measure*, underlying much of modern probability theory, which is easy to believe and may be accepted on faith, though all standard proofs I am aware of, some presented in Chapter 2, are quite tricky and non-intuitive.

[6]Purists may note that we are implicitly adopting *ZFC* as the axiomatic foundation of modern probability theory. Nobody else ever needs to worry.

[7]I am no longer surprised when the students tell me things like ‘Fubini’s theorem we didn’t have time to cover’.

[8]It is a mistake to dismiss the theories of characteristics functions and classical null arrays, once regarded as core probability, as a technical nuisance that we can safely avoid.

the most important dependence structures in probability theory, generalizing the deterministic notion of *dynamical systems*. Though first introduced in Part IV, they will form a recurrent theme throughout the remainder of the book. After a general discussion of the general Markov property, we specialize to the classical theories of discrete- and continuous-time Markov chains, paying special attention to the theories of *random walks* and *renewal theory*, before providing a brief discussion of *branching processes*. The results in this part, many of fundamental importance in their own right, may also serve as motivations for the theories of Lévy and Feller processes in later chapters, as well as for the regenerative processes and their local time.

The *Brownian motion* and *Poisson processes*[9] constitute the basic building blocks of probability theory, whose theory is covered by the initial chapters of Part V. The remaining chapters deal with the equally fundamental *Lévy* and *Feller processes*. To see the connections, we note that Feller processes form a broad class of Markov processes, general enough for most applications, and yet regular enough for technical convenience. Specializing to the space-homogeneous case yields the class of Lévy processes, which may also be characterized as processes with stationary, independent increments, thus forming the continuous-time counterparts of random walks. The classical *Lévy-Khinchin formula* gives the basic representation of Lévy processes in terms of a Brownian motion and a Poisson process describing the jump structure. All facets of this theory are clearly of fundamental importance.

Stochastic calculus is another core area[10] of modern probability that we can't live without. The subject, here covered by Part VI, can be regarded as a continuation of the martingale theory from Part III. In fact, the celebrated *Itô formula* is essentially a transformation rule for continuous or more general semi-martingales. We begin with the relatively elementary case of continuous integrators, analysing a detail some important consequences for Brownian motion, before moving on to the more subtle theory for processes with discontinuities. Still more subtle is the powerful stochastic analysis on *Wiener space*, also known as *Malliavin calculus*.

Every serious probability student needs to be well acquainted with the theory of *convergence in distribution* in function spaces, covered by Part VII. Here the most natural and intuitive approach is via the powerful *Skorohod embedding*, which yields the celebrated *Donsker theorem* and its ramifications, along with various functional versions of the *law of the iterated logarithm*. More general results may occasionally require some quite subtle compactness arguments based on *Prohorov's theorem*, extending to a powerful theory for distributional convergence of *random measures*. Somewhat similar in spirit is the theory of *large deviations*, which may be regarded as a subtle extension of the weak law of large numbers.

Stationarity, defined as invariance in distribution under shifts, may be re-

[9]The theory of Poisson and related processes has often been neglected by textbook authors. Its fundamental importance should be clear from material in later chapters.

[10]It is often delegated to separate courses, though it truly belongs to the basic package.

garded as yet another basic dependence structure[11] of probability theory, beside those of martingales and Markov processes. Its theory with ramifications are the subjects of Part VIII. Here the central result is the powerful *ergodic theorem*, generalizing the strong law of large numbers, which admits some equally remarkable extensions to broad classes of Markov processes. This part also contains a detailed discussion of some other invariance structures, leading in particular to the powerful *predictable sampling* and *mapping theorems*.

Part IX begins with a detailed exposition of *excursion theory*, a powerful extension of renewal theory, involving a basic description of the entire excursion structure of a *regenerative process* in terms of a Poisson process, on the time scale given by the associated *local time*. Three totally different approaches to local time are discussed, each providing its share of valuable insight. The remaining chapters of this part deal with various aspects of *random measure* theory[12], including some basic results for *Palm measures* and a variety of fundamental limit theorems for *particle systems*. We may also mention some important connections to *statistical mechanics*.

The final Part X begins with a detailed discussion of *stochastic differential equations* (SDEs), which play the same role in the presence of noise as do the ODEs for deterministic dynamical systems. Of special importance is the characterization of solutions in terms of *martingale problems*. The solutions to the simplest SDEs are continuous strong Markov processes, even called *diffusions*, and the subsequent chapter includes a detailed account of such processes. We will also highlight the close connection between Markov processes and potential theory, showing in particular how some problems in classical potential theory admit solutions in terms of a Brownian motion. We conclude with an introduction to the beautiful and illuminating theory of stochastic differential geometry.

---

[11] This is another subject that has often been omitted from basic graduate-text packages, where it truly belongs.

[12] Yet another neglected topic. Random measures arguably constitute an area of fundamental importance. Though they have been subject to an intense development for three quarters of a century, their theory remains virtually unknown among mainstream probabilists.

We conclude with a short list of some commonly used notation. A more comprehensive list will be found at the end of the book.

$\mathsf{N} = \{1, 2, \ldots\}$, $\mathsf{Z}_+ = \{0, 1, 2, \ldots\}$, $\mathsf{R}_+ = [0, \infty)$, $\bar{\mathsf{R}} = [-\infty, \infty]$,

$(S, \mathcal{S}, \hat{\mathcal{S}})$: localized Borel space, classes of measurable or bounded sets,

$\mathcal{S}_+$: class of $\mathcal{S}$-measurable functions $f \geq 0$,

$S^{(n)}$: non-diagonal part of S^n,

$\mathcal{M}_S$, $\hat{\mathcal{M}}_S$: classes of locally finite or normalized measures on S,

$\mathcal{N}_S$, $\hat{\mathcal{N}}_S$: classes of integer-valued and bounded measures in $\mathcal{M}_S$,

$\mathcal{M}_S^*$, $\mathcal{N}_S^*$: classes of diffuse measures in $\mathcal{M}_S$ and simple ones in $\mathcal{N}_S$,

G, λ: measurable group with Haar measure, Lebesgue measure on R,

$\delta_s B = 1_B(s) = 1\{s \in B\}$: unit mass at s and indicator function of B,

$\mu f = \int f \, d\mu$, $(f \cdot \mu) g = \mu(fg)$, $(\mu \circ f^{-1}) g = \mu(g \circ f)$, $1_B \mu = 1_B \cdot \mu$,

$\mu^{(n)}$: for $\mu \in \mathcal{N}_S^*$, the restriction of μ^n to $S^{(n)}$,

$(\theta_r \mu) f = \mu(f \circ \theta_r) = \int \mu(ds) f(rs)$,

$(\nu \otimes \mu) f = \int \nu(ds) \int \mu_s(dt) f(s, t)$, $(\nu \mu) f = \int \nu(ds) \int \mu_s(dt) f(t)$,

$(\mu * \nu) f = \int \mu(dx) \int \nu(dy) f(x + y)$,

$(E\xi) f = E(\xi f)$, $E(\xi | \mathcal{F}) g = E(\xi g | \mathcal{F})$,

$\mathcal{L}(\,)$, $\mathcal{L}(\,|\,)_s$, $\mathcal{L}(\,\|\,)_s$: distribution, conditional or Palm distribution,

$C_\xi f = E \sum_{\mu \leq \xi} f(\mu, \xi - \mu)$ with bounded μ,

$\perp\!\!\!\perp$, $\perp\!\!\!\perp_{\mathcal{F}}$: independence and conditional independence given $\mathcal{F}$,

$\stackrel{d}{=}$, $\stackrel{d}{\to}$: equality and convergence in distribution,

$\stackrel{w}{\to}$, $\stackrel{v}{\to}$, $\stackrel{u}{\to}$: weak, vague, and uniform convergence,

$\xrightarrow{wd}$, $\xrightarrow{vd}$: weak or vague convergence in distribution,

$\|f\|$, $\|\mu\|$: supremum of $|f|$ and total variation of μ.

I. Measure Theoretic Prerequisites

Modern probability theory is technically a branch of real analysis, and any serious student needs a good foundation in measure theory and basic functional analysis, before moving on to the probabilistic aspects. Though a solid background in classical real analysis may be helpful, our present emphasis is often somewhat different, and many results included here may be hard to find in the standard textbook literature. We recommend the reader to be thoroughly familiar with especially the material in Chapter 1. The hurried or impatient reader might skip Chapters 2–3 for the moment, and return for specific topics when need arises.

— — —

1. Sets and functions, measures and integration. This chapter contains the basic notions and results of measure theory, needed throughout the remainder of the book. Statements used most frequently include the monotone-class theorem, the monotone and dominated convergence theorems, Fubini's theorem, criteria for L^p-convergence, and the Hölder and Minkowski inequalities. The reader should also be familiar with the notion of Borel spaces, which provides the basic setting throughout the book.

2. Measure extension and decomposition. Here we include some of the more advanced results of measure theory of frequent use, such as the existence of Lebesgue and related measures, the Lebesgue decomposition and differentiation theorem, atomic decompositions of measures, the Radon–Nikodym theorem, and the Riesz representation. We also explore the relationship between signed measures on the line and functions of locally bounded variation. Though a good familiarity with the results is crucial, many proofs are rather technical and might be skipped on the first way through.

3. Kernels, disintegration, and invariance. Though kernels are hardly mentioned in most real analysis texts, they play a fundamental role throughout probability theory. Any serious reader needs to be well familiar with the basic kernel properties and operations, as well as with the existence theorems for single and dual disintegration. The quoted results on invariant measures and disintegration, including the theory of Haar measures, will be needed only for special purposes and might be skipped on a first reading, with a possible return when need arises.

Chapter 1

Sets and Functions, Measures and Integration

Sigma-fields, measurable functions, monotone-class theorem, Polish and Borel spaces, convergence and limits, measures and integration, monotone and dominated convergence, transformation of integrals, null sets and completion, product measures and Fubini's theorem, Hölder and Minkowski inequalities, L^p-convergence, L^2-projection, regularity and approximation

Though originating long ago from some simple gambling problems, probability theory has since developed into a sophisticated area of real analysis, depending profoundly on some measure theory already for the basic definitions. This chapter covers most of the standard measure theory needed to get started. Soon you will need more, which is why we include two further chapters to cover some more advanced topics. At that later stage, even some basic functional analysis will play an increasingly important role, where the underlying notions and results are summarized without proofs in Appendices 3–4.

In this chapter we give an introduction to basic measure theory, including the notions of σ-fields, measurable functions, measures, and integration. Though our treatment, here and in later chapters, is essentially self-contained, we do rely on some basic notions and results from elementary real analysis, including the notions of limits and continuity, and some metric topology involving open and closed sets, completeness, and compactness. Many results covered here will henceforth be used in subsequent chapters without explicit reference. This applies especially to the monotone and dominated convergence theorems, Fubini's theorem, and the Hölder and Minkowski inequalities.

Though a good background in real analysis may be an advantage, the hurried reader is warned against skipping the measure-theoretic chapters without at least a quick review of the included results, many of which may be hard to find in the standard textbook literature. This applies in particular to the monotone-class theorem and the notion of Borel spaces in this chapter, the atomic decomposition in Chapter 2, and the theory of kernels, disintegration, and invariant measures in Chapter 3.

To fix our notation, we begin with some elementary notions from set theory. For subsets $A, A_k, B, \ldots$ of an abstract space S, recall the definitions of *union* $A \cup B$ or $\bigcup_k A_k$, *intersection* $A \cap B$ or $\bigcap_k A_k$, *complement* A^c, and *difference* $A \setminus B = A \cap B^c$. The latter is said to be *proper* if $A \supset B$. The *symmetric*

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_2

difference of A and B is given by $A\triangle B = (A \setminus B) \cup (B \setminus A)$. Among basic set relations, we note in particular the *distributive laws*

$$A \cap \bigcup_k B_k = \bigcup_k (A \cap B_k), \qquad A \cup \bigcap_k B_k = \bigcap_k (A \cup B_k),$$

and *de Morgan's laws*

$$\Big(\bigcup_k A_k\Big)^c = \bigcap_k A_k^c, \qquad \Big(\bigcap_k A_k\Big)^c = \bigcup_k A_k^c,$$

valid for arbitrary (not necessarily countable) unions and intersections. The latter formulas allow us to convert any relation involving unions (intersections) into the dual formula for intersections (unions).

We define a *σ-field*[1] in S as a non-empty collection $\mathcal{S}$ of subsets of S closed under countable unions and intersections[2] as well as under complementation. Thus, if $A, A_1, A_2, \ldots \in \mathcal{S}$, then also A^c, $\bigcup_k A_k$, and $\bigcap_k A_k$ lie in $\mathcal{S}$. In particular, the whole space S and the empty set $\emptyset$ belong to every σ-field. In any space S, there is a smallest σ-field $\{\emptyset, S\}$ and a largest one 2^S, consisting of *all* subsets of S. Note that every σ-field $\mathcal{S}$ is closed under monotone limits, so that if $A_1, A_2, \ldots \in \mathcal{S}$ with $A_n \uparrow A$ or $A_n \downarrow A$, then also $A \in \mathcal{S}$. A *measurable space* is a pair $(S, \mathcal{S})$, where S is a space and $\mathcal{S}$ is a σ-field in S.

For any class of σ-fields in S, the intersection, but usually not the union, is again a σ-field. If $\mathcal{C}$ is an arbitrary class of subsets of S, there is a smallest σ-field in S containing $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$ and called the σ-field *generated* or *induced* by $\mathcal{C}$. Note that $\sigma(\mathcal{C})$ can be obtained as the intersection of all σ-fields in S containing $\mathcal{C}$. We endow a metric or topological space S with its *Borel σ-field* $\mathcal{B}_S$ generated by the topology[3] in S, unless a σ-field is otherwise specified. The elements of $\mathcal{B}_S$ are called *Borel sets*. When S is the real line R, we often write $\mathcal{B}$ instead of $\mathcal{B}_{\mathsf{R}}$.

More primitive classes than σ-fields often arise in applications. A class $\mathcal{C}$ of subsets of a space S is called a *π-system* if it is closed under finite intersections, so that $A, B \in \mathcal{C}$ implies $A \cap B \in \mathcal{C}$. Furthermore, a class $\mathcal{D}$ is a *λ-system* if it contains S and is closed under proper differences and increasing limits. Thus, we require $S \in \mathcal{D}$, that $A, B \in \mathcal{D}$ with $A \supset B$ implies $A \setminus B \in \mathcal{D}$, and that $A_1, A_2, \ldots \in \mathcal{D}$ with $A_n \uparrow A$ implies $A \in \mathcal{D}$.

The following *monotone-class theorem* is useful to extend an established property or relation from a class $\mathcal{C}$ to the generated σ-field $\sigma(\mathcal{C})$. An application of this result is referred to as a *monotone-class argument*.

Theorem 1.1 *(monotone classes, Sierpiński) For any π-system $\mathcal{C}$ and λ-system $\mathcal{D}$ in a space S, we have*

$$\mathcal{C} \subset \mathcal{D} \quad \Rightarrow \quad \sigma(\mathcal{C}) \subset \mathcal{D}.$$

---

[1] also called a *σ-algebra*

[2] For a *field* or *algebra*, we require closure only under finite set operations.

[3] class of open subsets

Proof: First we note that a class $\mathcal{C}$ is a σ-field iff it is both a π-system and a λ-system. Indeed, if $\mathcal{C}$ has the latter properties, then $A, B \in \mathcal{C}$ implies $A^c = S \setminus A \in \mathcal{C}$ and $A \cup B = (A^c \cap B^c)^c \in \mathcal{C}$. Next let $A_1, A_2, \ldots \in \mathcal{C}$, and put $A = \bigcup_n A_n$. Then $B_n \uparrow A$, where $B_n = \bigcup_{k \le n} A_k \in \mathcal{C}$, and so $A \in \mathcal{C}$. Finally, $\bigcap_n A_n = (\bigcup_n A_n^c)^c \in \mathcal{C}$.

For the main assertion, we may clearly assume that $\mathcal{D} = \lambda(\mathcal{C})$, defined as the smallest λ-system containing $\mathcal{C}$. It suffices to show that $\mathcal{D}$ is also a π-system, since it is then a σ-field containing $\mathcal{C}$ and therefore contains the smallest σ-field $\sigma(\mathcal{C})$ with this property. Thus, we need to show that $A \cap B \in \mathcal{D}$ whenever $A, B \in \mathcal{D}$.

The relation $A \cap B \in \mathcal{D}$ holds when $A, B \in \mathcal{C}$, since $\mathcal{C}$ is a π-system contained in $\mathcal{D}$. We proceed by extension in two steps. First we fix any $B \in \mathcal{C}$, and define $\mathcal{S}_B = \{A \subset S;\, A \cap B \in \mathcal{D}\}$. Then $\mathcal{S}_B$ is a λ-system containing $\mathcal{C}$, and so it contains the smallest λ-system $\mathcal{D}$ with this property. Thus, $A \cap B \in \mathcal{D}$ for any $A \in \mathcal{D}$ and $B \in \mathcal{C}$. Next we fix any $A \in \mathcal{D}$, and define $\mathcal{S}'_A = \{B \subset S;\, A \cap B \in \mathcal{D}\}$. As before, we note that even $\mathcal{S}'_A$ contains $\mathcal{D}$, which yields the desired property. □

For any family of spaces S_t, $t \in T$, we define the *Cartesian product* $\mathsf{X}_{t \in T} S_t$ as the class of all collections $\{s_t;\, t \in T\}$, where $s_t \in S_t$ for all t. When $T = \{1, \ldots, n\}$ or $T = \mathsf{N} = \{1, 2, \ldots\}$, we often write the product space as $S_1 \times \cdots \times S_n$ or $S_1 \times S_2 \times \cdots$, respectively, and if $S_t = S$ for all t, we use the notation S^T, S^n, or S^∞. For topological spaces S_t, we endow $\mathsf{X}_t S_t$ with the product topology, unless a topology is otherwise specified.

Now assume that each space S_t is equipped with a σ-field $\mathcal{S}_t$. In $\mathsf{X}_t S_t$ we may then introduce the *product σ-field* $\bigotimes_t \mathcal{S}_t$, generated by all one-dimensional *cylinder sets*[4] $A_t \times \mathsf{X}_{s \ne t} S_s$, where $t \in T$ and $A_t \in \mathcal{S}_t$. As before, we write in appropriate cases

$$\mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n, \quad \mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \cdots, \quad \mathcal{S}^T, \quad \mathcal{S}^n, \quad \mathcal{S}^\infty.$$

Lemma 1.2 *(product σ-field) For any separable metric spaces $S_1, S_2, \ldots$, we have*

$$\mathcal{B}(S_1 \times S_2 \times \cdots) = \mathcal{B}_{S_1} \otimes \mathcal{B}_{S_2} \otimes \cdots .$$

Thus, for countable products of separable metric spaces, the product and Borel σ-fields agree. In particular, $\mathcal{B}_{\mathsf{R}^d} = (\mathcal{B}_{\mathsf{R}})^d = \mathcal{B}^d$ is the σ-field generated by all rectangular boxes $I_1 \times \cdots \times I_d$, where $I_1, \ldots, I_d$ are arbitrary real intervals. This special case can also be proved directly.

Proof: The assertion can be written as $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$, where $\mathcal{C}_1$ is the class of open sets in $S = S_1 \times S_2 \times \cdots$ and $\mathcal{C}_2$ is the class of cylinder sets of the form $S_1 \times \cdots \times S_{k-1} \times B_k \times S_{k+1} \times \cdots$ with $B_k \in \mathcal{B}_{S_k}$. Now let $\mathcal{C}$ be the subclass of such cylinder sets, with B_k open in S_k. Then $\mathcal{C}$ is a topological base in S, and since S is separable, every set in $\mathcal{C}_1$ is a countable union of sets in $\mathcal{C}$. Hence,

[4]Note the analogy with the definition of product topologies.

$\mathcal{C}_1 \subset \sigma(\mathcal{C})$. Furthermore, the open sets in S_k generate $\mathcal{B}_{S_k}$, and so $\mathcal{C}_2 \subset \sigma(\mathcal{C})$. Since trivially $\mathcal{C}$ lies in both $\mathcal{C}_1$ and $\mathcal{C}_2$, we get

$$\begin{aligned}\sigma(\mathcal{C}_1) &\subset \sigma(\mathcal{C}) \subset \sigma(\mathcal{C}_2)\\ &\subset \sigma(\mathcal{C}) \subset \sigma(\mathcal{C}_1),\end{aligned}$$

and so equality holds throughout, proving the asserted relation. □

Every point mapping $f: S \to T$ induces a set mapping $f^{-1}: 2^T \to 2^S$ in the opposite direction, given by

$$f^{-1}B = \left\{s \in S;\, f(s) \in B\right\}, \quad B \subset T.$$

Note that f^{-1} preserves the basic set operations, in the sense that, for any subsets B and B_k of T,

$$\begin{aligned} f^{-1}B^c &= (f^{-1}B)^c,\\ f^{-1}\bigcup_n B_n &= \bigcup_n f^{-1}B_n,\\ f^{-1}\bigcap_n B_n &= \bigcap_n f^{-1}B_n. \end{aligned} \tag{1}$$

We show that f^{-1} also preserves σ-fields, in both directions. For convenience, we write

$$f^{-1}\mathcal{C} = \left\{f^{-1}B;\, B \in \mathcal{C}\right\}, \quad \mathcal{C} \subset 2^T.$$

Lemma 1.3 *(induced σ-fields) For any mapping f between measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, we have*

(i) *$\mathcal{S}' = f^{-1}\mathcal{T}$ is a σ-field in S,*

(ii) *$\mathcal{T}' = \left\{B \subset T;\, f^{-1}B \in \mathcal{S}\right\}$ is a σ-field in T.*

Proof: (i) If $A, A_1, A_2, \ldots \in \mathcal{S}'$, there exist some sets $B, B_1, B_2, \ldots \in \mathcal{T}$ with $A = f^{-1}B$ and $A_n = f^{-1}B_n$ for all n. Since $\mathcal{T}$ is a σ-field, the sets B^c, $\bigcup_n B_n$, and $\bigcap_n B_n$ all belong to $\mathcal{T}$, and (1) yields

$$\begin{aligned} A^c &= (f^{-1}B)^c = f^{-1}B^c,\\ \bigcup_n A_n &= \bigcup_n f^{-1}B_n = f^{-1}\bigcup_n B_n,\\ \bigcap_n A_n &= \bigcap_n f^{-1}B_n = f^{-1}\bigcap_n B_n, \end{aligned}$$

which all lie in $f^{-1}\mathcal{T} = \mathcal{S}'$.

(ii) Let $B, B_1, B_2, \ldots \in \mathcal{T}'$, so that $f^{-1}B, f^{-1}B_1, f^{-1}B_2, \ldots \in \mathcal{S}$. Using (1) and the fact that $\mathcal{S}$ is a σ-field, we get

$$\begin{aligned} f^{-1}B^c &= (f^{-1}B)^c,\\ f^{-1}\bigcup_n B_n &= \bigcup_n f^{-1}B_n,\\ f^{-1}\bigcap_n B_n &= \bigcap_n f^{-1}B_n, \end{aligned}$$

which all belong to $\mathcal{S}$. Thus, B^c, $\bigcup_n B_n$, and $\bigcap_n B_n$ all lie in $\mathcal{T}'$. □

For any measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, a mapping $f: S \to T$ is said to be $\mathcal{S}/\mathcal{T}$-*measurable* or simply *measurable*[5] if $f^{-1}\mathcal{T} \subset \mathcal{S}$, that is, if $f^{-1}B \in \mathcal{S}$ for every $B \in \mathcal{T}$. It is enough to verify the defining condition for a suitable subclass:

Lemma 1.4 *(measurable functions) Let f be a mapping between measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, and let $\mathcal{C} \subset 2^T$ with $\sigma(\mathcal{C}) = \mathcal{T}$. Then*

$$f \text{ is } \mathcal{S}/\mathcal{T}\text{-measurable} \quad \Leftrightarrow \quad f^{-1}\mathcal{C} \subset \mathcal{S}.$$

Proof: Let $\mathcal{T}' = \{B \subset T;\ f^{-1}B \in \mathcal{S}\}$. Then $\mathcal{C} \subset \mathcal{T}'$ by hypothesis, and $\mathcal{T}'$ is a σ-field by Lemma 1.3 (ii). Hence,

$$\mathcal{T}' = \sigma(\mathcal{T}') \supset \sigma(\mathcal{C}) = \mathcal{T},$$

which shows that $f^{-1}B \in \mathcal{S}$ for all $B \in \mathcal{T}$. □

Lemma 1.5 *(continuity and measurability) For a map f between topological spaces S, T with Borel σ-fields $\mathcal{S}, \mathcal{T}$, we have*

$$f \text{ is continuous} \quad \Rightarrow \quad f \text{ is } \mathcal{S}/\mathcal{T}\text{-measurable.}$$

Proof: Let $\mathcal{S}'$ and $\mathcal{T}'$ be the classes of open sets in S and T. Since f is continuous and $\mathcal{S} = \sigma(\mathcal{S}')$, we have $f^{-1}\mathcal{T}' \subset \mathcal{S}' \subset \mathcal{S}$. By Lemma 1.4 it follows that f is $\mathcal{S}/\sigma(\mathcal{T}')$-measurable. It remains to note that $\sigma(\mathcal{T}') = \mathcal{T}$. □

We insert a result about subspace topologies and σ-fields, needed in Chapter 23. Given a class $\mathcal{C}$ of subsets of S and a set $A \subset S$, we define $A \cap \mathcal{C} = \{A \cap C;\ C \in \mathcal{C}\}$.

Lemma 1.6 *(subspaces) Let (S, ρ) be a metric space with topology $\mathcal{T}$ and Borel σ-field $\mathcal{S}$, and let $A \subset S$. Then the induced topology and Borel σ-field in (A, ρ) are given by*

$$\mathcal{T}_A = A \cap \mathcal{T}, \qquad \mathcal{S}_A = A \cap \mathcal{S}.$$

Proof: The natural embedding $\pi_A: A \to S$ is continuous and hence measurable, and so $A \cap \mathcal{T} = \pi_A^{-1}\mathcal{T} \subset \mathcal{T}_A$ and $A \cap \mathcal{S} = \pi_A^{-1}\mathcal{S} \subset \mathcal{S}_A$. Conversely, for any $B \in \mathcal{T}_A$, we define $G = (B \cup A^c)^o$, where the complement and interior are with respect to S, and note that $B = A \cap G$. Hence, $\mathcal{T}_A \subset A \cap \mathcal{T}$, and therefore

$$\begin{aligned} \mathcal{S}_A &= \sigma(\mathcal{T}_A) \subset \sigma(A \cap \mathcal{T}) \\ &\subset \sigma(A \cap \mathcal{S}) \\ &= A \cap \mathcal{S}, \end{aligned}$$

where the operation $\sigma(\cdot)$ refers to the subspace A. □

As with continuity, even measurability is preserved by composition.

[5]Note the analogy with the definition of continuity in terms of topologies on S, T.

Lemma 1.7 *(composition) For maps $f: S \to T$ and $g: T \to U$ between the measurable spaces $(S, \mathcal{S})$, $(T, \mathcal{T})$, $(U, \mathcal{U})$, we have*

$$f, g \text{ are measurable} \quad \Rightarrow \quad h = g \circ f \text{ is } \mathcal{S}/\mathcal{U}\text{-measurable.}$$

Proof: Let $C \in \mathcal{U}$, and note that $B \equiv g^{-1}C \in \mathcal{T}$ since g is measurable. Noting that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$, and using the fact that even f is measurable, we get

$$\begin{aligned} h^{-1}C &= (g \circ f)^{-1}C \\ &= f^{-1}g^{-1}C \\ &= f^{-1}B \in \mathcal{S}. \end{aligned}$$ □

For many results in measure and probability theory, it is convenient first to give a proof for the real line, and then extend to more general spaces. Say that the measurable spaces S and T are *Borel isomorphic*, if there exists a bijection $f: S \leftrightarrow T$ such that both f and f^{-1} are measurable. A measurable space S is said to be *Borel* if it is Borel isomorphic to a Borel set in $[0, 1]$. We show that every Polish space[6] endowed with its Borel σ-field is Borel.

Theorem 1.8 *(Polish and Borel spaces) For any Polish space S, we have*

(i) *S is homeomorphic to a Borel set in $[0, 1]^\infty$,*
(ii) *S is Borel isomorphic to a Borel set in $[0, 1]$.*

Proof: (i) Fix a complete metric ρ in S. We may assume that $\rho \le 1$, since we can otherwise replace ρ by the equivalent metric $\rho \wedge 1$, which is again complete. Since S is separable, we may choose a dense sequence $x_1, x_2, \ldots \in S$. Then the mapping

$$x \mapsto \big\{\rho(x, x_1), \rho(x, x_2), \ldots\big\}, \quad x \in S,$$

defines a homeomorphic embedding of S into the compact space $K = [0, 1]^\infty$, and we may regard S as a subset of K. In K we introduce the metric

$$d(x, y) = \textstyle\sum_n 2^{-n}|x_n - y_n|, \quad x, y \in K,$$

and define $\bar{S}$ as the closure of S in K.

Writing $|B_x^\varepsilon|_\rho$ for the ρ-diameter of the d-ball $B_x^\varepsilon = \{y \in S;\, d(x, y) < \varepsilon\}$ in S, we define

$$U_n(\varepsilon) = \Big\{x \in \bar{S};\, |B_x^\varepsilon|_\rho < n^{-1}\Big\}, \quad \varepsilon > 0,\ n \in N,$$

and put $G_n = \bigcup_\varepsilon U_n(\varepsilon)$. The G_n are open in $\bar{S}$, since $x \in U_n(\varepsilon)$ and $y \in \bar{S}$ with $d(x, y) < \varepsilon/2$ implies $y \in U_n(\varepsilon/2)$, and $S \subset G_n$ for each n by the equivalence of the metrics ρ and d. This gives $S \subset \tilde{S} \subset \bar{S}$, where $\tilde{S} = \bigcap_n G_n$.

For any $x \in \tilde{S}$, we may choose some $x_1, x_2, \ldots \in S$ with $d(x, x_n) \to 0$. By the definitions of $U_n(\varepsilon)$ and G_n, the sequence (x_k) is Cauchy even for ρ, and

[6] A topological space is said to be *Polish* if it is separable and admits a complete metrization.

so by completeness $\rho(x_k, y) \to 0$ for some $y \in S$. Since ρ and d are equivalent on S, we have even $d(x_k, y) \to 0$, and therefore $x = y$. This gives $\tilde{S} \subset S$, and so $\tilde{S} = S$. Finally, $\tilde{S}$ is a Borel set in K, since the G_n are open subsets of the compact set $\bar{S}$.

(ii) Write 2^∞ for the countable product $\{0,1\}^\infty$, and let B denote the subset of binary sequences with infinitely many zeros. Then $x = \sum_n x_n 2^{-n}$ defines a 1−1 correspondence between $I = [0,1)$ and B. Since $x \mapsto (x_1, x_2, \ldots)$ is clearly bi-measurable, I and B are Borel isomorphic, written as $I \sim B$. Furthermore, $B^c = 2^\infty \setminus B$ is countable, so that $B^c \sim \mathsf{N}$, which implies

$$\begin{aligned} B \cup \mathsf{N} &\sim B \cup \mathsf{N} \cup (-\mathsf{N}) \\ &\sim B \cup B^c \cup \mathsf{N} \\ &= 2^\infty \cup \mathsf{N}, \end{aligned}$$

and hence $I \sim B \sim 2^\infty$. This gives $I^\infty \sim (2^\infty)^\infty \sim 2^\infty \sim I$, and the assertion follows by (i). □

For any Borel space $(S, \mathcal{S})$, we may specify a *localizing sequence* $S_n \uparrow S$ in $\mathcal{S}$, and say that a set $B \subset S$ is *bounded*, if $B \subset S_n$ for some $n \in \mathsf{N}$. The ring of bounded sets $B \in \mathcal{S}$ will be denoted by $\hat{\mathcal{S}}$. If ρ is a metric generating $\mathcal{S}$, we may choose $\hat{\mathcal{S}}$ to consist of all ρ-bounded sets. In locally compact spaces S, we may choose $\hat{\mathcal{S}}$ as the class of sets $B \in \mathcal{S}$ with compact closure $\bar{B}$.

By a *dissection system* in S we mean a nested sequence of countable partitions $(I_{nj}) \subset \hat{\mathcal{S}}$ generating $\mathcal{S}$, such that for any fixed $n \in \mathsf{N}$, every bounded set is covered by finitely many sets I_{nj}. For topological spaces S, we require in addition that, whenever $s \in G \subset S$ with G open, we have $s \in I_{nj} \subset G$ for some $n, j \in \mathsf{N}$. A class $\mathcal{C} \subset \hat{\mathcal{S}}$ is said to be *dissecting*, if every open set $G \subset S$ is a countable union of sets in $\mathcal{C}$, and every set $B \in \hat{\mathcal{S}}$ is covered by finitely many sets in $\mathcal{C}$.

To state the next result, we note that any collection of functions $f_i \colon \Omega \to S_i$, $i \in I$, defines a mapping $f = (f_i)$ from Ω to $\mathsf{X}_i S_i$, given by

$$f(\omega) = \big\{ f_i(\omega);\, i \in I \big\}, \quad \omega \in \Omega. \tag{2}$$

We may relate the measurability of f to that of the *coordinate mappings* f_i.

Lemma 1.9 *(coordinate functions) For any measurable spaces* $(\Omega, \mathcal{A})$, $(S_i, \mathcal{S}_i)$ *and functions* $f_i : \Omega \to S_i$, $i \in I$, *define* $f = (f_i) : \Omega \to \mathsf{X}_i S_i$. *Then these conditions are equivalent:*

(i) f *is* $\mathcal{A} / \bigotimes_i \mathcal{S}_i$ *-measurable,*

(ii) f_i *is* $\mathcal{A}/\mathcal{S}_i$ *-measurable for every* $i \in I$.

Proof: Use Lemma 1.4 with $\mathcal{C}$ equal to the class of cylinder sets $A_i \times \mathsf{X}_{j \neq i} S_j$, for arbitrary $i \in I$ and $A_i \in \mathcal{S}_i$. □

Changing our perspective, let the f_i in (2) map Ω into some measurable spaces $(S_i, \mathcal{S}_i)$. In Ω we may then introduce the *generated* or *induced* σ-field $\sigma(f) = \sigma\{f_i;\, i \in I\}$, defined as the smallest σ-field in Ω making all the f_i measurable. In other words, $\sigma(f)$ is the intersection of all σ-fields $\mathcal{A}$ in Ω, such that f_i is $\mathcal{A}/\mathcal{S}_t$-measurable for every $i \in I$. In this notation, the functions f_i are clearly measurable with respect to a σ-field $\mathcal{A}$ in Ω iff $\sigma(f) \subset \mathcal{A}$. Further note that $\sigma(f)$ agrees with the σ-field in Ω generated by the collection $\{f_i^{-1}\mathcal{S}_i;\, i \in I\}$.

For functions on or into a Euclidean space R^d, measurability is understood to be with respect to the Borel σ-field $\mathcal{B}^d$. Thus, a real-valued function f on a measurable space $(S, \mathcal{S})$ is measurable iff $\{s;\, f(s) \le x\} \in \mathcal{S}$ for all $x \in \mathsf{R}$. The same convention applies to functions into the *extended real line* $\bar{\mathsf{R}} = [-\infty, \infty]$ or the *extended half-line* $\bar{\mathsf{R}}_+ = [0, \infty]$, regarded as compactifications of R and $\mathsf{R}_+ = [0, \infty)$, respectively. Note that $\mathcal{B}_{\bar{\mathsf{R}}} = \sigma\{\mathcal{B}, \pm\infty\}$ and $\mathcal{B}_{\bar{\mathsf{R}}_+} = \sigma\{\mathcal{B}_{\mathsf{R}_+}, \infty\}$.

For any set $A \subset S$, the associated *indicator function*[7] $1_A \colon S \to \mathsf{R}$ is defined to equal 1 on A and 0 on A^c. For sets $A = \{s \in S;\, f(s) \in B\}$, it is often convenient to write $1\{\cdot\}$ instead of $1_{\{\cdot\}}$. When $\mathcal{S}$ is a σ-field in S, we note that 1_A is $\mathcal{S}$-measurable iff $A \in \mathcal{S}$.

Linear combinations of indicator functions are called *simple functions*. Thus, every simple function $f \colon S \to \mathsf{R}$ has the form

$$f = c_1 1_{A_1} + \cdots + c_n 1_{A_n},$$

where $n \in \mathsf{Z}_+ = \{0, 1, \ldots\}$, $c_1, \ldots, c_n \in \mathsf{R}$, and $A_1, \ldots, A_n \subset S$. Here we may take $c_1, \ldots, c_n$ to be the distinct non-zero values attained by f, and define $A_k = f^{-1}\{c_k\}$ for all k. This makes f measurable with respect to a given σ-field $\mathcal{S}$ in S iff $A_1, \ldots, A_n \in \mathcal{S}$.

The class of measurable functions is closed under countable limiting operations:

Lemma 1.10 *(bounds and limits) If the functions* $f_1, f_2, \ldots \colon (S, \mathcal{S}) \to \bar{\mathsf{R}}$ *are measurable, then so are the functions*

$$\sup_n f_n, \qquad \inf_n f_n, \qquad \limsup_{n\to\infty} f_n, \qquad \liminf_{n\to\infty} f_n.$$

Proof: To see that $\sup_n f_n$ is measurable, write

$$\begin{aligned}\Big\{s;\, \sup_n f_n(s) \le t\Big\} &= \bigcap_{n\ge 1} \Big\{s;\, f_n(s) \le t\Big\} \\ &= \bigcap_{n\ge 1} f_n^{-1}[-\infty, t] \in \mathcal{S},\end{aligned}$$

and use Lemma 1.4. For the other three cases, write $\inf_n f_n = -\sup_n(-f_n)$, and note that

$$\begin{aligned}\limsup_{n\to\infty} f_n &= \inf_{n\ge 1} \sup_{k\ge n} f_k, \\ \liminf_{n\to\infty} f_n &= \sup_{n\ge 1} \inf_{k\ge n} f_k.\end{aligned}$$

□

[7]The term *characteristic function* should be avoided, since it has a different meaning in probability.

Since $f_n \to f$ iff $\limsup_n f_n = \liminf_n f_n = f$, it follows easily that both the set of convergence and the possible limit are measurable. This extends to functions taking values in more general spaces:

Lemma 1.11 *(convergence and limits) Let $f_1, f_2, \ldots$ be measurable functions from a measurable space $(\Omega, \mathcal{A})$ into a metric space (S, ρ). Then*

(i) $\{\omega;\, f_n(\omega) \text{ converges}\} \in \mathcal{A}$ *when S is separable and complete,*

(ii) $f_n \to f$ *on Ω implies that f is measurable.*

Proof: (i) Since S is complete, convergence of f_n is equivalent to the Cauchy convergence

$$\lim_{n\to\infty} \sup_{m\geq n} \rho(f_m, f_n) = 0,$$

where the left-hand side is measurable by Lemmas 1.2, 1.5, and 1.10.

(ii) If $f_n \to f$, we have $g \circ f_n \to g \circ f$ for any continuous function $g\colon S \to \mathsf{R}$, and so $g \circ f$ is measurable by Lemmas 1.5 and 1.10. Fixing any open set $G \subset S$, we may choose some continuous functions $g_1, g_2, \ldots : S \to \mathsf{R}_+$ with $g_n \uparrow 1_G$, and conclude from Lemma 1.10 that $1_G \circ f$ is measurable. Thus, $f^{-1}G \in \mathcal{A}$ for all G, and so f is measurable by Lemma 1.4. □

Many results in measure theory can be proved by a simple approximation, based on the following observation.

Lemma 1.12 *(simple approximation) For any measurable function $f \geq 0$ on $(S, \mathcal{S})$, there exist some simple, measurable functions $f_1, f_2, \ldots : S \to \mathsf{R}_+$ with $f_n \uparrow f$.*

Proof: We may define

$$f_n(s) = 2^{-n}[2^n f(s)] \wedge n, \quad s \in S,\ n \in \mathsf{N}. \qquad \square$$

To illustrate the method, we show that the basic arithmetic operations are measurable.

Lemma 1.13 *(elementary operations) If the functions $f, g\colon (S, \mathcal{S}) \to \mathsf{R}$ are measurable, then so are the functions*

(i) *fg and $af + bg$, $a, b \in \mathsf{R}$,*

(ii) *f/g when $g \neq 0$ on S.*

Proof: By Lemma 1.12 applied to $f_\pm = (\pm f) \vee 0$ and $g_\pm = (\pm g) \vee 0$, we may approximate by simple measurable functions $f_n \to f$ and $g_n \to g$. Here $af_n + bg_n$ and $f_n g_n$ are again simple measurable functions. Since they converge to $af + bg$ and fg, respectively, the latter functions are again measurable by Lemma 1.10. The same argument applies to the ratio f/g, provided we choose $g_n \neq 0$.

An alternative argument is to write $af + bg$, fg, or f/g as a composition $\psi \circ \varphi$, where $\varphi = (f, g) : S \to \mathsf{R}^2$, and $\psi(x, y)$ is defined as $ax + by$, xy, or x/y, repectively. The desired measurability then follows by Lemmas 1.2, 1.5, 1.9, and 1.10. In case of ratios, we may use the continuity of the mapping $(x, y) \mapsto x/y$ on $\mathsf{R} \times (\mathsf{R} \setminus \{0\})$. □

We proceed with a functional representation of measurable functions. Given some functions f, g on a common space Ω, we say that f is *g-measurable* if the induced σ-fields are related by $\sigma(f) \subset \sigma(g)$.

Lemma 1.14 *(functional representation, Doob) Let f, g be measurable functions from $(\Omega, \mathcal{A})$ into some measurable spaces $(S, \mathcal{S})$, $(T, \mathcal{T})$, where S is Borel. Then these conditions are equivalent:*

(i) *f is g-measurable,*

(ii) *$f = h \circ g$ for a measurable mapping $h\colon T \to S$.*

Proof: Since S is Borel, we may let $S \in \mathcal{B}_{[0,1]}$. By a suitable modification of h, we may further reduce to the case where $S = [0, 1]$. If $f = 1_A$ for a g-measurable set $A \subset \Omega$, Lemma 1.3 yields a set $B \in \mathcal{T}$ with $A = g^{-1}B$. Then $f = 1_A = 1_B \circ g$, and we may choose $h = 1_B$. The result extends by linearity to any simple, g-measurable function f. In general, Lemma 1.12 yields some simple, g-measurable functions $f_1, f_2, \ldots$ with $0 \le f_n \uparrow f$, and we may choose some associated $\mathcal{T}$-measurable functions $h_1, h_2, \ldots \colon T \to [0, 1]$ with $f_n = h_n \circ g$. Then $h = \sup_n h_n$ is again $\mathcal{T}$-measurable by Lemma 1.10, and we have

$$\begin{aligned} h \circ g &= \big(\sup\nolimits_n h_n\big) \circ g \\ &= \sup\nolimits_n (h_n \circ g) \\ &= \sup\nolimits_n f_n = f. \end{aligned}$$ □

Given a measurable space $(S, \mathcal{S})$, we say that a function $\mu : \mathcal{S} \to \bar{\mathsf{R}}_+$ is *countably additive* if

$$\mu \bigcup_{k \ge 1} A_k = \sum_{k \ge 1} \mu A_k, \quad A_1, A_2, \ldots \in \mathcal{S} \ \text{ disjoint.} \tag{3}$$

A *measure* on $(S, \mathcal{S})$ is defined as a countably additive set function $\mu\colon \mathcal{S} \to \bar{\mathsf{R}}_+$ with $\mu\emptyset = 0$. The triple $(S, \mathcal{S}, \mu)$ is then called a *measure space*. From (3) we note that any measure is finitely additive and non-decreasing. This in turn implies the *countable sub-additivity*

$$\mu \bigcup_{k \ge 1} A_k \le \sum_{k \ge 1} \mu A_k, \quad A_1, A_2, \ldots \in \mathcal{S}.$$

We note some basic continuity properties:

Lemma 1.15 *(continuity) For any measure μ on $(S, \mathcal{S})$ and sets $A_1, A_2, \ldots \in \mathcal{S}$, we have*

(i) *$A_n \uparrow A \ \Rightarrow\ \mu A_n \uparrow \mu A$,*

(ii) *$A_n \downarrow A,\ \mu A_1 < \infty \ \Rightarrow\ \mu A_n \downarrow \mu A$.*

Proof: For (i), we may apply (3) to the differences $D_n = A_n \setminus A_{n-1}$ with $A_0 = \emptyset$. To get (ii), apply (i) to the sets $B_n = A_1 \setminus A_n$. $\square$

The simplest measures on a measurable space $(S, \mathcal{S})$ are the unit point masses or *Dirac measures* δ_s, $s \in S$, given by $\delta_s A = 1_A(s)$. A measure of the form $a\,\delta_s$ is said to be *degenerate*. For any countable set $A = \{s_1, s_2, \ldots\}$, we may form the associated *counting measure* $\mu = \sum_n \delta_{s_n}$. More generally, we may form countable linear combinations of measures on S, as follows.

Proposition 1.16 *(sums of measures) For any measures $\mu_1, \mu_2, \ldots$ on $(S, \mathcal{S})$, the sum $\mu = \sum_n \mu_n$ is again a measure.*

Proof: First we note that, for any array of constants $c_{jk} \geq 0$, $j, k \in \mathsf{N}$,

$$\textstyle\sum_j \sum_k c_{jk} = \sum_k \sum_j c_{jk}. \tag{4}$$

This is obvious for finite sums. In general, we have for any $m, n \in \mathsf{N}$

$$\textstyle\sum_j \sum_k c_{jk} \geq \displaystyle\sum_{j \leq m} \sum_{k \leq n} c_{jk} = \sum_{k \leq n} \sum_{j \leq m} c_{jk}.$$

Letting $m \to \infty$ and then $n \to \infty$, we obtain (4) with inequality $\geq$. The reverse relation follows by symmetry, and the equality follows.

Now consider any disjoint sets $A_1, A_2, \ldots \in \mathcal{S}$. Using (4) and the countable additivity of each μ_n, we get

$$\begin{aligned} \textstyle\mu \bigcup_k A_k &= \textstyle\sum_n \mu_n \bigcup_k A_k \\ &= \textstyle\sum_n \sum_k \mu_n A_k \\ &= \textstyle\sum_k \sum_n \mu_n A_k \\ &= \textstyle\sum_k \mu A_k. \end{aligned}$$

$\square$

The last result is essentially equivalent to the following:

Corollary 1.17 *(monotone limits) For any σ-finite measures $\mu_1, \mu_2, \ldots$ on $(S, \mathcal{S})$, we have*

$$\mu_n \uparrow \mu \ \text{ or } \ \mu_n \downarrow \mu \quad \Rightarrow \quad \mu \text{ is a measure on } S.$$

Proof: First let $\mu_n \uparrow \mu$. For any $n \in \mathsf{N}$, there exists a measure ν_n such that $\mu_n = \mu_{n-1} + \nu_n$, where $\mu_0 = 0$. Indeed, let $A_1, A_2, \ldots \in \mathcal{A}$ be disjoint with $\mu_{n-1} A_k < \infty$ for all k, and define $\nu_{n,k} = \mu_n - \mu_{n-1}$ on each A_k. Then $\mu = \sum_n \nu_n = \sum_{n,k} \nu_{n,k}$ is a measure on S by Proposition 1.16. If instead $\mu_n \downarrow \mu$, then define $\nu_n = \mu_1 - \mu_n$ as above, and note that $\nu_n \uparrow \nu$ for some measure $\nu \leq \mu_1$. Hence, $\mu_n = \mu_1 - \nu_n \downarrow \mu_1 - \nu$, which is again a measure. $\square$

Any measurable mapping f between two measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$ induces a mapping of measures on S into measures on T. More precisely, given any measure μ on $(S, \mathcal{S})$, we may define a measure $\mu \circ f^{-1}$ on $(T, \mathcal{T})$ by

$$\begin{aligned}(\mu \circ f^{-1})B &= \mu(f^{-1}B) \\ &= \mu\Big\{s \in S;\, f(s) \in B\Big\}, \quad B \in \mathcal{T}.\end{aligned}$$

Here the countable additivity of $\mu \circ f^{-1}$ follows from that for μ, together with the fact that f^{-1} preserves unions and intersections.

It is often useful to identify a *measure-determining* class $\mathcal{C} \subset \mathcal{S}$, such that a measure on S is uniquely determined by its values on $\mathcal{C}$.

Lemma 1.18 *(uniqueness) Let μ, ν be bounded measures on a measurable space $(S, \mathcal{S})$, and let $\mathcal{C}$ be a π-system in S with $S \in \mathcal{C}$ and $\sigma(\mathcal{C}) = \mathcal{S}$. Then*

$$\mu = \nu \quad \Leftrightarrow \quad \mu A = \nu A, \;\; A \in \mathcal{C}.$$

Proof: Assuming $\mu = \nu$ on $\mathcal{C}$, let $\mathcal{D}$ be the class of sets $A \in \mathcal{S}$ with $\mu A = \nu A$. Using the condition $S \in \mathcal{C}$, the finite additivity of μ and ν, and Lemma 1.15, we see that $\mathcal{D}$ is a λ-system. Moreover, $\mathcal{C} \subset \mathcal{D}$ by hypothesis. Hence, Theorem 1.1 yields $\mathcal{D} \supset \sigma(\mathcal{C}) = \mathcal{S}$, which means that $\mu = \nu$. The converse assertion is obvious. $\square$

For any measure μ on $(S, \mathcal{S})$ and set $A \in \mathcal{S}$, the mapping $\nu\colon B \mapsto \mu(A \cap B)$ is again a measure on $(S, \mathcal{S})$, called the *restriction* of μ to A and denoted by $1_A\, \mu$. A measure μ on S is said to be *σ-finite*, if S is a countable union of disjoint sets $A_n \in \mathcal{S}$ with $\mu A_n < \infty$. Since $\mu = \sum_n 1_{A_n} \mu$, such a μ is even *s-finite*, in the sense of being a countable sum of bounded measures. A measure μ on a localized Borel space S is said to be *locally finite* if $\mu B < \infty$ for all $B \in \hat{\mathcal{S}}$, and we write $\mathcal{M}_S$ for the class of such measures μ.

For any measure μ on a topological space S, we define the *support* $\operatorname{supp} \mu$ as the set of points $s \in S$, such that $\mu B > 0$ for every neighborhood B of s. Note that $\operatorname{supp} \mu$ is closed and hence Borel measurable.

Lemma 1.19 *(support) For any measure μ on a separable, complete metric space S, we have*

$$\mu(\operatorname{supp} \mu)^c = 0.$$

Proof: Suppose that instead $\mu(\operatorname{supp} \mu)^c > 0$. Since S is separable, the open set $(\operatorname{supp} \mu)^c$ is a countable union of closed balls of radius < 1. Choose one of them B_1 with $\mu B_1 > 0$, and continue recursively to form a nested sequence of balls B_n of radius $< 2^{-n}$, such that $\mu B_n > 0$ for all n. The associated centers $x_n \in B_n$ form a Cauchy sequence, and so by completeness we have convergence $x_n \to x$. Then $x \in \bigcap_n B_n \subset (\operatorname{supp} \mu)^c$ and $x \in \operatorname{supp} \mu$, a contradiction proving our claim. $\square$

Our next aim is to define the *integral*

$$\mu f = \int f\, d\mu = \int f(\omega)\, \mu(d\omega)$$

of a real-valued, measurable function f on a measure space $(S, \mathcal{S}, \mu)$. First take f to be simple and non-negative, hence of the form $c_1 1_{A_1} + \cdots + c_n 1_{A_n}$ for some $n \in \mathsf{Z}_+$, $A_1, \ldots, A_n \in \mathcal{S}$ and $c_1, \ldots, c_n \in \mathsf{R}_+$, and define[8]

$$\mu f = c_1\, \mu A_1 + \cdots + c_n\, \mu A_n.$$

Using the finite additivity of μ, we may check that μf is independent of the choice of representation of f. It is further clear that the integration map $f \mapsto \mu f$ is *linear* and *non-decreasing*, in the sense that

$$\mu(af + bg) = a\, \mu f + b\, \mu g, \quad a, b \geq 0,$$
$$f \leq g \quad \Rightarrow \quad \mu f \leq \mu g.$$

To extend the integral to general measurable functions $f \geq 0$, use Lemma 1.12 to choose some simple measurable functions $f_1, f_2, \ldots$ with $0 \leq f_n \uparrow f$, and define $\mu f = \lim_n \mu f_n$. We need to show that the limit is independent of the choice of approximating sequence (f_n):

Lemma 1.20 *(consistence) Let f, $f_1, f_2, \ldots$ and g be measurable functions on a measure space $(S, \mathcal{S}, \mu)$, where all but f are simple. Then*

$$\left.\begin{array}{l} 0 \leq f_n \uparrow f \\ 0 \leq g \leq f \end{array}\right\} \Rightarrow \mu g \leq \lim_{n \to \infty} \mu f_n.$$

Proof: By the linearity of μ, it is enough to take $g = 1_A$ for some $A \in \mathcal{S}$. Fixing any $\varepsilon > 0$, we define

$$A_n = \Big\{\omega \in A;\ f_n(\omega) \geq 1 - \varepsilon\Big\}, \quad n \in \mathsf{N}.$$

Then $A_n \uparrow A$, and so

$$\begin{aligned} \mu f_n &\geq (1 - \varepsilon)\, \mu A_n \\ &\uparrow (1 - \varepsilon)\, \mu A \\ &= (1 - \varepsilon)\, \mu g. \end{aligned}$$

It remains to let $\varepsilon \to 0$. □

The linearity and monotonicity extend immediately to arbitrary $f \geq 0$, since $f_n \uparrow f$ and $g_n \uparrow g$ imply $af_n + bg_n \uparrow af + bg$, and if also $f \leq g$, then $f_n \leq (f_n \vee g_n) \uparrow g$. We prove a basic continuity property of the integral.

Theorem 1.21 *(monotone convergence, Levi) For any measurable functions $f, f_1, f_2 \ldots$ on $(S, \mathcal{S}, \mu)$, we have*

$$0 \leq f_n \uparrow f \quad \Rightarrow \quad \mu f_n \uparrow \mu f.$$

[8]The convention $0 \cdot \infty = 0$ applies throughout measure theory.

Proof: For every n we may choose some simple measurable functions g_{nk}, with $0 \le g_{nk} \uparrow f_n$ as $k \to \infty$. The functions $h_{nk} = g_{1k} \vee \cdots \vee g_{nk}$ have the same properties and are further non-decreasing in both indices. Hence,

$$\begin{aligned} f &\ge \lim_{k\to\infty} h_{kk} \\ &\ge \lim_{k\to\infty} h_{nk} \\ &= f_n \uparrow f, \end{aligned}$$

and so $0 \le h_{kk} \uparrow f$. Using the definition and monotonicity of the integral, we obtain

$$\begin{aligned} \mu f &= \lim_{k\to\infty} \mu h_{kk} \\ &\le \lim_{k\to\infty} \mu f_k \le \mu f. \end{aligned}$$ □

The last result yields the following key inequality.

Lemma 1.22 *(Fatou) For any measurable functions $f_1, f_2, \ldots \ge 0$ on $(S, \mathcal{S}, \mu)$, we have*

$$\liminf_{n\to\infty} \mu f_n \ge \mu \liminf_{n\to\infty} f_n.$$

Proof: Since $f_m \ge \inf_{k\ge n} f_k$ for all $m \ge n$, we have

$$\inf_{k\ge n} \mu f_k \ge \mu \inf_{k\ge n} f_k, \quad n \in \mathbb{N}.$$

Letting $n \to \infty$, we get by Theorem 1.21

$$\begin{aligned} \liminf_{k\to\infty} \mu f_k &\ge \lim_{n\to\infty} \mu \inf_{k\ge n} f_k \\ &= \mu \liminf_{k\to\infty} f_k. \end{aligned}$$ □

A measurable function f on $(S, \mathcal{S}, \mu)$ is said to be *integrable* if $\mu|f| < \infty$. Writing $f = g - h$ for some integrable functions $g, h \ge 0$ (e.g., as $f_+ - f_-$ with $f_\pm = (\pm f) \vee 0$), we define $\mu f = \mu g - \mu h$. It is easy to see that the extended integral is independent of the choice of representation $f = g - h$, and that μf satisfies the basic linearity and monotonicity properties, now with any real coefficients.

The last lemma leads to a general condition, other than the monotonicity in Theorem 1.21, allowing us to take limits under the integral sign, in the sense that

$$f_n \to f \quad \Rightarrow \quad \mu f_n \to \mu f.$$

By the classical *dominated convergence theorem*, this holds when $|f_n| \le g$ for a measurable function $g \ge 0$ with $\mu g < \infty$. The same argument yields a more powerful extended version:

Theorem 1.23 *(extended dominated convergence, Lebesgue) Let $f, f_1, f_2, \ldots$ and $g, g_1, g_2, \ldots \ge 0$ be measurable functions on $(S, \mathcal{S}, \mu)$. Then*

$$\left.\begin{array}{c} f_n \to f \\ |f_n| \le g_n \to g \\ \mu g_n \to \mu g < \infty \end{array}\right\} \Rightarrow \mu f_n \to \mu f.$$

Proof: Applying Fatou's lemma to the functions $g_n \pm f_n \geq 0$, we get

$$\begin{aligned}\mu g + \liminf_{n\to\infty} (\pm \mu f_n) &= \liminf_{n\to\infty} \mu(g_n \pm f_n) \\ &\geq \mu(g \pm f) \\ &= \mu g \pm \mu f.\end{aligned}$$

Subtracting $\mu g < \infty$ from each side gives

$$\begin{aligned}\mu f &\leq \liminf_{n\to\infty} \mu f_n \\ &\leq \limsup_{n\to\infty} \mu f_n \leq \mu f.\end{aligned}$$ □

Next we show how integrals are transformed by measurable maps.

Lemma 1.24 *(substitution) Given a measure space* (Ω, μ)*, a measurable space* S*, and some measurable maps* $f\colon \Omega \to S$ *and* $g\colon S \to \mathsf{R}$*, we have*

$$\mu(g \circ f) = (\mu \circ f^{-1})g, \tag{5}$$

whenever either side exists[9].

Proof: For indicator functions g, (5) reduces to the definition of $\mu \circ f^{-1}$. From here on, we may extend by linearity and monotone convergence to any measurable function $g \geq 0$. For general g it follows that $\mu|g \circ f| = (\mu \circ f^{-1})|g|$, and so the integrals in (5) exist at the same time. When they do, we get (5) by taking differences on both sides. □

For another basic transformation of measures and integrals, fix a measurable function $f \geq 0$ on a measure space $(S, \mathcal{S}, \mu)$, and define a function $f \cdot \mu$ on $\mathcal{S}$ by

$$(f \cdot \mu)A = \mu(1_A f) = \int_A f\, d\mu, \quad A \in \mathcal{S},$$

where the last equality defines the integral over A. Then clearly $\nu = f \cdot \mu$ is again a measure on $(S, \mathcal{S})$. We refer to f as the μ*-density* of ν. The associated transformation rule is the following.

Lemma 1.25 *(chain rule) For any measure space* $(S, \mathcal{S}, \mu)$ *and measurable functions* $f\colon S \to \mathsf{R}_+$ *and* $g\colon S \to \mathsf{R}$*, we have*

$$\mu(fg) = (f \cdot \mu)g,$$

whenever either side exists.

Proof: As before, we may first consider indicator functions g, and then extend in steps to the general case. □

Given a measure space $(S, \mathcal{S}, \mu)$, we say that $A \in \mathcal{S}$ is μ*-null* or simply *null* if $\mu A = 0$. A relation between functions on S is said to hold *almost everywhere* with respect to μ (written as *a.e.* μ or μ*-a.e.*) if it holds for all $s \in S$ outside a μ-null set. The following frequently used result explains the relevance of null sets.

---

[9] meaning that if one side exists, then so does the other and the two are equal

Lemma 1.26 *(null sets and functions) For any measurable function $f \geq 0$ on a measure space $(S, \mathcal{S}, \mu)$, we have*

$$\mu f = 0 \quad \Leftrightarrow \quad f = 0 \text{ a.e. } \mu.$$

Proof: This is obvious when f is simple. For general f, we may choose some simple measurable functions f_n with $0 \leq f_n \uparrow f$, and note that $f = 0$ a.e. iff $f_n = 0$ a.e. for every n, that is, iff $\mu f_n = 0$ for all n. Since the latter integrals converge to μf, the last condition is equivalent to $\mu f = 0$. □

The last result shows that two integrals agree when the integrands are a.e. equal. We may then allow the integrands to be undefined on a μ-null set. It is also clear that the conclusions of Theorems 1.21 and 1.23 remain valid, if the hypotheses are only fulfilled outside a null set.

In the other direction, we note that if two σ-finite measures μ, ν are related by $\nu = f \cdot \mu$ for a density f, then the latter is μ-a.e. unique, which justifies the notation $f = d\nu/d\mu$. It is further clear that any μ-null set is also a null set for ν. For measures μ, ν with the latter property, we say that ν is *absolutely continuous* with respect to μ, and write $\nu \ll \mu$. The other extreme case is when μ, ν are *mutually singular*, written as $\mu \perp \nu$, in the sense that $\mu A = 0$ and $\nu A^c = 0$ for some set $A \in \mathcal{S}$.

Given any measure space $(S, \mathcal{S}, \mu)$ and σ-field $\mathcal{F} \subset \mathcal{S}$, we define the *$\mu$-completion of $\mathcal{F}$ in $\mathcal{S}$* as the σ-field $\mathcal{F}^\mu = \sigma(\mathcal{F}, \mathcal{N}_\mu)$, where $\mathcal{N}_\mu$ is the class of subsets of arbitrary μ-null sets in $\mathcal{S}$. The description of $\mathcal{F}^\mu$ can be made more explicit:

Lemma 1.27 *(completion) For any measure space $(\Omega, \mathcal{A}, \mu)$, σ-field $\mathcal{F} \subset \mathcal{A}$, Borel space $(S, \mathcal{S})$, and function $f : \Omega \to S$, these conditions are equivalent:*

(i) *f is $\mathcal{F}^\mu$-measurable,*

(ii) *$f = g$ a.e. μ for an $\mathcal{F}$-measurable function g.*

Proof: Beginning with indicator functions, let $\mathcal{G}$ be the class of subsets $A \subset \Omega$ with $A \triangle B \in \mathcal{N}_\mu$ for some $B \in \mathcal{F}$. Then $A \setminus B$ and $B \setminus A$ again lie in $\mathcal{N}_\mu$, and so $\mathcal{G} \subset \mathcal{F}^\mu$. Conversely, $\mathcal{F}^\mu \subset \mathcal{G}$ since both $\mathcal{F}$ and $\mathcal{N}_\mu$ are trivially contained in $\mathcal{G}$. Combining the two relations gives $\mathcal{G} = \mathcal{F}^\mu$, which shows that $A \in \mathcal{F}^\mu$ iff $1_A = 1_B$ a.e. for some $B \in \mathcal{F}$.

In general we may take $S = [0, 1]$. For any $\mathcal{F}^\mu$-measurable function f, we may choose some simple $\mathcal{F}^\mu$-measurable functions f_n with $0 \leq f_n \uparrow f$. By the result for indicator functions, we may next choose some simple $\mathcal{F}$-measurable functions g_n, such that $f_n = g_n$ a.e. for all n. Since a countable union of null sets is again null, the function $g = \limsup_n g_n$ has the desired property. □

Every measure μ on $(S, \mathcal{S})$ extends uniquely to the σ-field $\mathcal{S}^\mu$. Indeed, for any $A \in \mathcal{S}^\mu$, Lemma 1.27 yields some sets $A_\pm \in \mathcal{S}$ with $A_- \subset A \subset A_+$ and $\mu(A_+ \setminus A_-) = 0$, and any extension satisfies $\mu A = \mu A_\pm$. With this choice, it is easy to check that μ remains a measure on $\mathcal{S}^\mu$.

We proceed to construct product measures and establish conditions allowing us to change the order of integration. This requires a technical lemma of independent importance.

Lemma 1.28 *(sections) For any measurable spaces $(S, \mathcal{S})$, $(T, \mathcal{T})$, measurable function $f\colon S \times T \to \mathsf{R}_+$, and σ-finite measure μ on S, we have*

(i) *$f(s,t)$ is $\mathcal{S}$-measurable in $s \in S$ for fixed $t \in T$,*

(ii) *$\int f(s,t)\,\mu(ds)$ is $\mathcal{T}$-measurable in $t \in T$.*

Proof: We may take μ to be bounded. Both statements are obvious when $f = 1_A$ with $A = B \times C$ for some $B \in \mathcal{S}$ and $C \in \mathcal{T}$, and they extend by monotone-class arguments to any indicator functions of sets in $\mathcal{S} \otimes \mathcal{T}$. The general case follows by linearity and monotone convergence. $\square$

We may now state the main result for product measures, known as *Fubini's theorem*[10].

Theorem 1.29 *(product measure and iterated integral, Lebesgue, Fubini, Tonelli) For any σ-finite measure spaces $(S, \mathcal{S}, \mu)$, $(T, \mathcal{T}, \nu)$, we have*

(i) *there exists a unique measure $\mu \otimes \nu$ on $(S \times T,\ \mathcal{S} \otimes \mathcal{T})$, such that*

$$(\mu \otimes \nu)(B \times C) = \mu B \cdot \nu C, \quad B \in \mathcal{S},\ C \in \mathcal{T},$$

(ii) *for any measurable function $f\colon S \times T \to \mathsf{R}$ with $(\mu \otimes \nu)|f| < \infty$, we have*

$$\begin{aligned}(\mu \otimes \nu) f &= \int \mu(ds) \int f(s,t)\,\nu(dt) \\ &= \int \nu(dt) \int f(s,t)\,\mu(ds).\end{aligned}$$

Note that the iterated integrals[11] in (ii) are well defined by Lemma 1.28, although the inner integrals $\nu f(s,\cdot)$ and $\mu f(\cdot,t)$ may fail to exist on some null sets in S and T, respectively.

Proof: By Lemma 1.28, we may define

$$(\mu \otimes \nu) A = \int \mu(ds) \int 1_A(s,t)\,\nu(dt), \quad A \in \mathcal{S} \otimes \mathcal{T}, \tag{6}$$

which is clearly a measure on $S \times T$ satisfying (i). By a monotone-class argument there is at most one such measure. In particular, (6) remains true with the order of integration reversed, which proves (ii) for indicator functions f. The formula extends by linearity and monotone convergence to arbitrary measurable functions $f \geq 0$.

[10] This is a subtle result of measure theory. The elementary proposition in calculus for double integrals of continuous functions goes back at least to Cauchy.

[11] Iterated integrals should be read from right to left, so that the inner integral on the right becomes an integrand in the next step. This notation saves us from the nuisance of awkward parentheses, or from relying on confusing conventions for multiple integrals.

In general we note that (ii) holds with f replaced by $|f|$. If $(\mu\otimes\nu)|f| < \infty$, it follows that $N_S = \{s \in S;\, \nu|f(s,\cdot)| = \infty\}$ is a μ-null set in S, whereas $N_T = \{t \in T;\, \mu|f(\cdot,t)| = \infty\}$ is a ν-null set in T. By Lemma 1.26 we may redefine $f(s,t) = 0$ when $s \in N_S$ or $t \in N_T$. Then (ii) follows for f by subtraction of the formulas for f_+ and f_-. □

We call $\mu\otimes\nu$ in Theorem 1.29 the *product measure* of μ and ν. Iterating the construction, we may form product measures $\mu_1 \otimes \ldots \otimes \mu_n = \bigotimes_k \mu_k$ satisfying higher-dimensional versions of (ii). If $\mu_k = \mu$ for all k, we often write the product measure as $\mu^{\otimes n}$ or μ^n.

By a *measurable group* we mean a group G endowed with a σ-field $\mathcal{G}$, such that the group operations in G are $\mathcal{G}$-measurable. If $\mu_1,\ldots,\mu_n$ are σ-finite measures on G, we may define the *convolution* $\mu_1 * \cdots * \mu_n$ as the image of the product measure $\mu_1 \otimes \cdots \otimes \mu_n$ on G^n under the iterated group operation $(x_1,\ldots,x_n) \mapsto x_1\cdots x_n$. The convolution is said to be *associative* if $(\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$ whenever both $\mu_1 * \mu_2$ and $\mu_2 * \mu_3$ are σ-finite, and *commutative* if $\mu_1 * \mu_2 = \mu_2 * \mu_1$.

A measure μ on G is said to be *right-* or *left-invariant* if $\mu \circ \theta_r^{-1} = \mu$ for all $r \in G$, where θ_r denotes the right or left *shift* $s \mapsto s\,r$ or $s \mapsto rs$. When G is Abelian, the shift is also called a *translation*. On product spaces $G \times T$, the translations are defined as mappings of the form $\theta_r\colon (s,t) \mapsto (s+r,t)$.

Lemma 1.30 *(convolution) On a measurable group $(G,\mathcal{G})$, the convolution of σ-finite measures is associative. On Abelian groups it is also commutative, and*

(i)
$$(\mu * \nu)B = \int \mu(B-s)\,\nu(ds) = \int \nu(B-s)\,\mu(ds), \quad B \in \mathcal{G},$$

(ii) *if $\mu = f\cdot\lambda$ and $\nu = g\cdot\lambda$ for an invariant measure λ, then $\mu * \nu$ has the λ-density*
$$(f * g)(s) = \int f(s-t)\,g(t)\,\lambda(dt) = \int f(t)\,g(s-t)\,\lambda(dt), \quad s \in G.$$

Proof: Use Fubini's theorem. □

For any measure space $(S,\mathcal{S},\mu)$ and constant $p > 0$, let $L^p = L^p(S,\mathcal{S},\mu)$ be the class of measurable functions $f\colon S \to \mathsf{R}$ satisfying
$$\|f\|_p \equiv \left(\mu|f|^p\right)^{1/p} < \infty.$$

Theorem 1.31 *(Hölder and Minkowski inequalities) For any measurable functions f,g on a measure space $(S,\mathcal{S},\mu)$, we have*

(i) $\|fg\|_r \le \|f\|_p\,\|g\|_q$, *$p,q,r > 0$ with $p^{-1} + q^{-1} = r^{-1}$,*

(ii) $\|f+g\|_p \le \|f\|_p + \|g\|_p$, *$p \ge 1$,*

(iii) $\|f+g\|_p \leq \left(\|f\|_p^p + \|g\|_p^p\right)^{1/p}, \quad p \in (0,1]$.

Proof: (i) We may clearly take $r = 1$ and $\|f\|_p = \|g\|_q = 1$. Since $p^{-1} + q^{-1} = 1$ implies $(p-1)(q-1) = 1$, the equations $y = x^{p-1}$ and $x = y^{q-1}$ are equivalent for $x, y \geq 0$. By calculus,

$$\begin{aligned} |fg| &\leq \int_0^{|f|} x^{p-1} dx + \int_0^{|g|} y^{q-1} dy \\ &= p^{-1}|f|^p + q^{-1}|g|^q, \end{aligned}$$

and so

$$\begin{aligned} \|fg\|_1 &\leq p^{-1}\int |f|^p \, d\mu + q^{-1}\int |g|^q \, d\mu \\ &= p^{-1} + q^{-1} = 1. \end{aligned}$$

(ii) For $p > 1$, we get by (i) with $q = p/(p-1)$ and $r = 1$

$$\begin{aligned} \|f+g\|_p^p &\leq \int |f| \, |f+g|^{p-1} d\mu + \int |g| \, |f+g|^{p-1} d\mu \\ &\leq \|f\|_p \, \|f+g\|_p^{p-1} + \|g\|_p \, \|f+g\|_p^{p-1}. \end{aligned}$$

(iii) Clear from the concavity of $|x|^p$. □

Claim (ii) above is sometimes needed in the following extended form, where for any f as in Theorem 1.29 we define

$$\|f\|_p(s) = \left\{\nu|f(s,\cdot)|^p\right\}^{1/p}, \quad s \in S.$$

Corollary 1.32 *(extended Minkowski inequality) For μ, ν, f as in Theorem* 1.29*, suppose that $\mu f(\cdot, t)$ exists for $t \in T$ a.e. ν. Then*

$$\|\mu f\|_p \leq \mu\|f\|_p, \quad p \geq 1.$$

Proof: Since $|\mu f| \leq \mu|f|$, we may assume that $f \geq 0$ and $\|\mu f\|_p \in (0, \infty)$. For $p > 1$, we get by Fubini's theorem and Hölder's inequality

$$\begin{aligned} \|\mu f\|_p^p &= \nu(\mu f)^p \\ &= \nu\left\{\mu f(\mu f)^{p-1}\right\} \\ &= \mu\nu\left\{f(\mu f)^{p-1}\right\} \\ &\leq \mu\|f\|_p \left\|(\mu f)^{p-1}\right\|_q \\ &= \mu\|f\|_p \, \|\mu f\|_p^{p-1}. \end{aligned}$$

Now divide by $\|\mu f\|_p^{p-1}$. The proof for $p = 1$ is similar but simpler. □

In particular, Theorem 1.31 shows that $\|\cdot\|_p$ becomes a norm for $p \geq 1$, if we identify functions that agree a.e. For any $p > 0$ and $f, f_1, f_2, \ldots \in L^p$, we write $f_n \to f$ in L^p if $\|f_n - f\|_p \to 0$, and say that (f_n) is *Cauchy in L^p* if $\|f_m - f_n\|_p \to 0$ as $m, n \to \infty$.

Lemma 1.33 *(completeness) Let the sequence (f_n) be Cauchy in L^p, where $p > 0$. Then there exists an $f \in L^p$ such that*

$$\|f_n - f\|_p \to 0.$$

Proof: Choose a sub-sequence $(n_k) \subset \mathsf{N}$ with $\sum_k \|f_{n_{k+1}} - f_{n_k}\|_p^{p\wedge 1} < \infty$. By Lemma 1.31 and monotone convergence we get $\|\sum_k |f_{n_{k+1}} - f_{n_k}|\,\|_p^{p\wedge 1} < \infty$, and so $\sum_k |f_{n_{k+1}} - f_{n_k}| < \infty$ a.e. Hence, (f_{n_k}) is a.e. Cauchy in R, and so Lemma 1.11 yields $f_{n_k} \to f$ a.e. for some measurable function f. By Fatou's lemma,

$$\begin{aligned} \|f - f_n\|_p &\le \liminf_{k\to\infty} \|f_{n_k} - f_n\|_p \\ &\le \sup_{m\ge n} \|f_m - f_n\|_p \to 0, \quad n \to \infty, \end{aligned}$$

which shows that $f_n \to f$ in L^p. □

We give a useful criterion for convergence in L^p.

Lemma 1.34 *(L^p-convergence) Let $f, f_1, f_2, \ldots \in L^p$ with $f_n \to f$ a.e., where $p > 0$. Then*

$$\|f_n - f\|_p \to 0 \quad \Leftrightarrow \quad \|f_n\|_p \to \|f\|_p.$$

Proof: If $f_n \to f$ in L^p, we get by Lemma 1.31

$$\Big|\|f_n\|_p^{p\wedge 1} - \|f\|_p^{p\wedge 1}\Big| \le \|f_n - f\|_p^{p\wedge 1} \to 0,$$

and so $\|f_n\|_p \to \|f\|_p$. Now assume instead the latter condition, and define

$$g_n = 2^p\Big(|f_n|^p + |f|^p\Big), \qquad g = 2^{p+1}|f|^p.$$

Then $g_n \to g$ a.e. and $\mu g_n \to \mu g < \infty$ by hypotheses. Since also $|g_n| \ge |f_n - f|^p \to 0$ a.e., Theorem 1.23 yields $\|f_n - f\|_p^p = \mu|f_n - f|^p \to 0$. □

Taking $p = q = 2$ and $r = 1$ in Theorem 1.31 (i) yields the *Cauchy inequality*[12]

$$\|fg\|_1 \le \|f\|_2\,\|g\|_2.$$

In particular, the *inner product* $\langle f, g\rangle = \mu(fg)$ exists for $f, g \in L^2$ and satisfies $|\langle f, g\rangle| \le \|f\|_2\|g\|_2$. From the obvious bi-linearity of the inner product, we get the *parallelogram identity*

$$\|f + g\|^2 + \|f - g\|^2 = 2\,\|f\|^2 + 2\,\|g\|^2, \quad f, g \in L^2. \tag{7}$$

Two functions $f, g \in L^2$ are said to be *orthogonal*, written as $f \perp g$, if $\langle f, g\rangle = 0$. Orthogonality between two subsets $A, B \subset L^2$ means that $f \perp g$ for all $f \in A$ and $g \in B$. A subspace $M \subset L^2$ is said to be *linear* if $f, g \in M$ and $a, b \in \mathsf{R}$ imply $af + bg \in M$, and *closed* if $f_n \to f$ in L^2 for some $f_n \in M$ implies $f \in M$.

[12] also known as the *Schwarz* or *Buniakovsky inequality*

Theorem 1.35 *(projection) For a closed, linear subspace $M \subset L^2$, any function $f \in L^2$ has an a.e. unique decomposition*

$$f = g + h \quad \text{with } g \in M,\ h \perp M.$$

Proof: Fix any $f \in L^2$, and define $r = \inf\{\|f - g\|;\, g \in M\}$. Choose $g_1, g_2, \ldots \in M$ with $\|f - g_n\| \to r$. Using the linearity of M, the definition of r, and (7), we get as $m, n \to \infty$

$$\begin{aligned} 4\,r^2 + \|g_m - g_n\|^2 &\le \left\|2f - g_m - g_n\right\|^2 + \|g_m - g_n\|^2 \\ &= 2\,\|f - g_m\|^2 + 2\,\|f - g_n\|^2 \to 4\,r^2. \end{aligned}$$

Thus, $\|g_m - g_n\| \to 0$, and so the sequence (g_n) is Cauchy in L^2. By Lemma 1.33 it converges toward some $g \in L^2$, and since M is closed, we have $g \in M$. Noting that $h = f - g$ has norm r, we get for any $l \in M$

$$\begin{aligned} r^2 &\le \|h + tl\|^2 \\ &= r^2 + 2\,t\langle h, l\rangle + t^2\|l\|^2, \quad t \in \mathsf{R}, \end{aligned}$$

which implies $\langle h, l\rangle = 0$. Hence, $h \perp M$, as required.

To prove the uniqueness, let $g' + h'$ be another decomposition with the stated properties. Then both $g - g' \in M$ and $g - g' = h' - h \perp M$, and so $g - g' \perp g - g'$, which implies $\|g - g'\|^2 = \langle g - g', g - g'\rangle = 0$, and hence $g = g'$ a.e. $\square$

We proceed with a basic approximation of sets. Let $\mathcal{F}$, $\mathcal{G}$ denote the classes of closed and open subsets of S.

Lemma 1.36 *(regularity) For any bounded measure μ on a metric space S with Borel σ-field $\mathcal{S}$, we have*

$$\mu B = \sup_{F \subset B} \mu F = \inf_{G \supset B} \mu G, \quad B \in \mathcal{S},$$

with F, G restricted to $\mathcal{F}$, $\mathcal{G}$, respectively.

Proof: For any $G \in \mathcal{G}$ there exist some closed sets $F_n \uparrow G$, and by Lemma 1.15 we get $\mu F_n \uparrow \mu G$. This proves the statement for B belonging to the π-system $\mathcal{G}$ of open sets. Letting $\mathcal{D}$ be the class of sets B with the stated property, we further note that $\mathcal{D}$ is a λ-system. Hence, Theorem 1.1 yields $\mathcal{D} \supset \sigma(\mathcal{G}) = \mathcal{S}$. $\square$

The last result leads to a basic approximation property for functions.

Lemma 1.37 *(approximation) Consider a metric space S with Borel σ-field $\mathcal{S}$, a bounded measure μ on $(S, \mathcal{S})$, and a constant $p > 0$. Then the bounded, continuous functions on S are dense in $L^p(S, \mathcal{S}, \mu)$. Thus, for any $f \in L^p$, there exist some bounded, continuous functions $f_1, f_2, \ldots : S \to \mathsf{R}$ with*

$$\|f_n - f\|_p \to 0.$$

Proof: If $f = 1_A$ with $A \subset S$ open, we may choose some continuous functions f_n with $0 \le f_n \uparrow f$, and then $\|f_n - f\|_p \to 0$ by dominated convergence. The result extends by Lemma 1.36 to any $A \in \mathcal{S}$. The further extension to simple, measurable functions is immediate. For general $f \in L^p$, we may choose some simple, measurable functions $f_n \to f$ with $|f_n| \le |f|$. Since $|f_n - f|^p \le 2^{p+1}|f|^p$, we get $\|f_n - f\|_p \to 0$ by dominated convergence. □

Next we show how the pointwise convergence of measurable functions is almost uniform. Here $\|f\|_A = \sup_{s \in A} |f(s)|$.

Lemma 1.38 *(near uniformity, Egorov) Let $f, f_1, f_2, \ldots$ be measurable functions on a finite measure space $(S, \mathcal{S}, \mu)$, such that $f_n \to f$ on S. Then there exist some sets $A_1, A_2, \ldots \in \mathcal{S}$ satifying*

$$\mu A_k^c \to 0, \qquad \|f_n - f\|_{A_k} \to 0, \quad k \in \mathsf{N}.$$

Proof: Define

$$A_{r,n} = \bigcap_{k \ge n} \left\{ s \in S;\ |f_k(s) - f(s)| < r^{-1} \right\}, \quad r, n \in \mathsf{N}.$$

Letting $n \to \infty$ for fixed r, we get $A_{r,n} \uparrow S$ and hence $\mu A_{r,n}^c \to 0$. Given any $\varepsilon > 0$, we may choose $n_1, n_2, \ldots \in \mathsf{N}$ so large that $\mu A_{r,n_r}^c < \varepsilon 2^{-r}$ for all r. Writing $A = \bigcap_r A_{r,n_r}$, we get

$$\begin{aligned} \mu A^c &\le \mu \textstyle\bigcup_r A_{r,n_r}^c \\ &< \varepsilon \textstyle\sum_r 2^{-r} = \varepsilon, \end{aligned}$$

and we note that $f_n \to f$ uniformly on A. □

Combining the last two results, we may show that every measurable function is almost continuous.

Lemma 1.39 *(near continuity, Lusin) Consider a measurable function f and a bounded measure μ on a compact metric space S with Borel σ-field $\mathcal{S}$. Then there exist some continuous functions $f_1, f_2, \ldots$ on S, such that*

$$\mu\left\{x;\ f_n(x) \ne f(x)\right\} \to 0.$$

Proof: We may clearly take f to be bounded. By Lemma 1.37 we may choose some continuous functions $g_1, g_2, \ldots$ on S, such that $\mu|g_k - f| \le 2^{-k}$. By Fubini's theorem, we get

$$\begin{aligned} \mu \textstyle\sum_k |g_k - f| &= \textstyle\sum_k \mu|g_k - f| \\ &\le \textstyle\sum_k 2^{-k} = 1, \end{aligned}$$

and so $\sum_k |g_k - f| < \infty$ a.e., which implies $g_k \to f$ a.e. By Lemma 1.38, we may next choose some $A_1, A_2, \ldots \in \mathcal{S}$ with $\mu A_n^c \to 0$, such that the convergence is uniform on every A_n. Since each g_k is uniformly continuous on S, we conclude that f is uniformly continuous on each A_n. By Tietze's extension theorem[13], the restriction $f|_{A_n}$ has then a continuous extension f_n to S. $\square$

Exercises

1. Prove the triangle inequality $\mu(A\triangle C) \le \mu(A\triangle B) + \mu(B\triangle C)$. (*Hint:* Note that $1_{A\triangle B} = |1_A - 1_B|$.)

2. Show that Lemma 1.10 fails for uncountable index sets. (*Hint:* Show that every measurable set depends on countably many coordinates.)

3. For any space S, let μA denote the cardinality of the set $A \subset S$. Show that μ is a measure on $(S, 2^S)$.

4. Let $\mathcal{K}$ be the class of compact subsets of some metric space S, and let μ be a bounded measure such that $\inf_{K\in\mathcal{K}} \mu K^c = 0$. Show that for any $B \in \mathcal{B}_S$, $\mu B = \sup_{K\in\mathcal{K}\cap B} \mu K$.

5. Show that any absolutely convergent series can be written as an integral with respect to counting measure on $\mathbb{N}$. State series versions of Fatou's lemma and the dominated convergence theorem, and give direct elementary proofs.

6. Give an example of integrable functions $f, f_1, f_2, \ldots$ on a probability space $(S, \mathcal{S}, \mu)$, such that $f_n \to f$ but $\mu f_n \not\to \mu f$.

7. Let μ, ν be σ-finite measures on a measurable space $(S, \mathcal{S})$ with sub-σ-field $\mathcal{F}$. Show that if $\mu \ll \nu$ on $\mathcal{S}$, it also holds on $\mathcal{F}$. Further show by an example that the converse may fail.

8. Fix two measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, a measurable function $f\colon S \to T$, and a measure μ on S with image $\nu = \mu \circ f^{-1}$. Show that f remains measurable with respect to the completions $\mathcal{S}^\mu$ and $\mathcal{T}^\nu$.

9. Fix a measure space $(S, \mathcal{S}, \mu)$ and a σ-field $\mathcal{T} \subset \mathcal{S}$, let $\mathcal{S}^\mu$ denote the μ-completion of $\mathcal{S}$, and let $\mathcal{T}^\mu$ be the σ-field generated by $\mathcal{T}$ and the μ-null sets of $\mathcal{S}^\mu$. Show that $A \in \mathcal{T}^\mu$ iff there exist some $B \in \mathcal{T}$ and $N \in \mathcal{S}^\mu$ with $A\triangle B \subset N$ and $\mu N = 0$. Also, show by an example that $\mathcal{T}^\mu$ may be strictly greater than the μ-completion of $\mathcal{T}$.

10. State Fubini's theorem for the case where μ is σ-finite and ν is counting measure on $\mathbb{N}$. Give a direct proof of this version.

11. Let $f_1, f_2, \ldots$ be μ-integrable functions on a measurable space S, such that $g = \sum_k f_k$ exists a.e., and put $g_n = \sum_{k\le n} f_k$. Restate the dominated convergence theorem for the integrals μg_n in terms of the functions f_k, and compare with the result of the preceding exercise.

12. Extend Theorem 1.29 to the product of n measures.

13. Let $M \supset N$ be closed linear subspaces of L^2. Show that if $f \in L^2$ has projections g onto M and h onto N, then g has projection h onto N.

14. Let M be a closed linear subspace of L^2, and let $f, g \in L^2$ with M-projections $\hat f$ and $\hat g$. Show that $\langle \hat f, g\rangle = \langle f, \hat g\rangle = \langle \hat f, \hat g\rangle$.

[13] A continuous function on a closed subset of a normal topological space S has a continuous extension to S.

15. Show that if $\mu \ll \nu$ and $\nu f = 0$ with $f \geq 0$, then also $\mu f = 0$. (*Hint:* Use Lemma 1.26.)

16. For any σ-finite measures $\mu_1 \ll \mu_2$ and $\nu_1 \ll \nu_2$, show that $\mu_1 \otimes \nu_1 \ll \mu_2 \otimes \nu_2$. (*Hint:* Use Fubini's theorem and Lemma 1.26.)

Chapter 2

Measure Extension and Decomposition

Outer measure, Carathéodory extension, Lebesgue measure, shift and rotation invariance, Hahn and Jordan decompositions, Lebesgue decomposition and differentiation, Radon–Nikodym theorem, Lebesgue–Stieltjes measures, finite-variation functions, signed measures, atomic decomposition, factorial measures, right and left continuity, absolutely continuous and singular functions, Riesz representation

The general measure theory developed in Chapter 1 would be void and meaningless, unless we could establish the existence of some non-trivial, countably additive measures. In fact, much of modern probability theory depends implicitly on the existence of such measures, needed already to model an elementary sequence of coin tossings. Here we prove the equivalent existence of Lebesgue measure, which will later allow us to establish the more general Lebesgue–Stieltjes measures, and a variety of discrete- and continuous-time processes throughout the book.

The basic existence theorems in this and the next chapter are surprisingly subtle, and the hurried or impatient reader may feel tempted to skip this material and move on to the probabilistic parts of the book. However, he/she should be aware that much of the present theory underlies the subsequent probabilistic discussions, and any serious student should be prepared to return for reference when need arises.

Our first aim is to construct Lebesgue measure, using the powerful approach of Carathéodory based on outer measures. The result will lead, via the Daniell–Kolmogorov theorem in Chapter 8, to the basic existence theorem for Markov processes in Chapter 11. We proceed to the correspondence between signed measures and functions of locally finite variation, of special importance for the theory of semi-martingales and general stochastic integration. A further high point is the powerful Riesz representation, which will enable us in Chapter 17 to construct Markov processes with a given generator, via resolvents and associated semi-groups of transition operators. We may further mention the Radon–Nikodym theorem, relevant to the theory of conditioning in Chapter 8, and Lebesgue's differentiation theorem, instrumental for proving the general ballot theorem in Chapter 25.

We begin with an ingenious technical result, which will play a crucial role for our construction of Lebesgue measure in Theorem 2.2, and for the proof of Riesz' representation Theorem 2.25. By an *outer measure* on a space S

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_3

we mean a non-decreasing, countably sub-additive set function $\mu: 2^S \to \bar{\mathsf{R}}_+$ with $\mu\emptyset = 0$. Given an outer measure μ on S, we say that a set $A \subset S$ is μ*-measurable* if

$$\mu E = \mu(E \cap A) + \mu(E \cap A^c), \quad E \subset S. \tag{1}$$

Note that the inequality $\le$ holds automatically by sub-additivity. The following result gives the basic measure construction from outer measures.

Theorem 2.1 *(restriction of outer measure, Carathéodory) Let μ be an outer measure on S, and write $\mathcal{S}$ for the class of μ-measurable sets. Then $\mathcal{S}$ is a σ-field and the restriction of μ to $\mathcal{S}$ is a measure.*

Proof: Since $\mu\emptyset = 0$, we have for any set $E \subset S$

$$\mu(E \cap \emptyset) + \mu(E \cap S) = \mu\emptyset + \mu E = \mu E,$$

which shows that $\emptyset \in \mathcal{S}$. Also note that trivially $A \in \mathcal{S}$ implies $A^c \in \mathcal{S}$.

Next let $A, B \in \mathcal{S}$. Using (1) for A and B together with the sub-additivity of μ, we get for any $E \subset S$

$$\begin{aligned} \mu E &= \mu(E \cap A) + \mu(E \cap A^c) \\ &= \mu\big(E \cap A \cap B\big) + \mu\big(E \cap A \cap B^c\big) + \mu(E \cap A^c) \\ &\ge \mu\big\{E \cap (A \cap B)\big\} + \mu\big\{E \cap (A \cap B)^c\big\}, \end{aligned}$$

which shows that even $A \cap B \in \mathcal{S}$. It follows easily that $\mathcal{S}$ is a field. If $A, B \in \mathcal{S}$ are disjoint, we also get by (1) for any $E \subset S$

$$\begin{aligned} \mu\big\{E \cap (A \cup B)\big\} &= \mu\big\{E \cap (A \cup B) \cap A\big\} + \mu\big\{E \cap (A \cup B) \cap A^c\big\} \\ &= \mu(E \cap A) + \mu(E \cap B). \end{aligned} \tag{2}$$

Finally, let $A_1, A_2, \ldots \in \mathcal{S}$ be disjoint, and put $U_n = \bigcup_{k \le n} A_k$ and $U = \bigcup_n U_n$. Using (2) recursively along with the monotonicity of μ, we get

$$\begin{aligned} \mu(E \cap U) &\ge \mu(E \cap U_n) \\ &= \sum_{k \le n} \mu(E \cap A_k). \end{aligned}$$

Letting $n \to \infty$ and combining with the sub-additivity of μ, we obtain

$$\mu(E \cap U) = \textstyle\sum_k \mu(E \cap A_k). \tag{3}$$

Taking $E = S$, we see in particular that μ is countably additive on $\mathcal{S}$. Noting that $U_n \in \mathcal{S}$ and using (3) twice, along with the monotonicity of μ, we also get

$$\begin{aligned} \mu E &= \mu(E \cap U_n) + \mu(E \cap U_n^c) \\ &\ge \sum_{k \le n} \mu(E \cap A_k) + \mu(E \cap U^c) \\ &\to \mu(E \cap U) + \mu(E \cap U^c), \end{aligned}$$

which shows that $U \in \mathcal{S}$. Thus, $\mathcal{S}$ is a σ-field. □

Much of modern probability relies on the existence of non-trivial, countably additive measures. Here we prove that the elementary notion of interval length can be extended to a measure λ on R, known as *Lebesgue measure*.

Theorem 2.2 *(Lebesgue measure, Borel) There exists a unique measure λ on $(\mathsf{R}, \mathcal{B})$, such that*

$$\lambda[a,b] = b - a, \quad a \le b.$$

For the proof, we first need to extend the set of lengths $|I|$ of real intervals I to an outer measure on R. Then define

$$\lambda A = \inf_{\{I_k\}} \sum\nolimits_k |I_k|, \quad A \subset \mathsf{R}, \tag{4}$$

where the infimum extends over all countable covers of A by open intervals $I_1, I_2, \ldots$. We show that (4) provides the desired extension.

Lemma 2.3 *(outer Lebesgue measure) The function λ in (4) is an outer measure on R, satisfying $\lambda I = |I|$ for every interval I.*

Proof: The set function λ is clearly non-negative and non-decreasing with $\lambda\emptyset = 0$. To prove the countable sub-additivity, let $A_1, A_2, \ldots \subset \mathsf{R}$ be arbitrary. For any $\varepsilon > 0$ and $n \in \mathsf{N}$, we may choose some open intervals $I_{n1}, I_{n2}, \ldots$ such that

$$\begin{aligned} A_n &\subset \bigcup\nolimits_k I_{nk}, \\ \lambda A_n &\ge \sum\nolimits_k |I_{nk}| - \varepsilon 2^{-n}, \quad n \in \mathsf{N}. \end{aligned}$$

Then

$$\begin{aligned} \bigcup\nolimits_n A_n &\subset \bigcup\nolimits_n \bigcup\nolimits_k I_{nk}, \\ \lambda \bigcup\nolimits_n A_n &\le \sum\nolimits_n \sum\nolimits_k |I_{nk}| \\ &\le \sum\nolimits_n \lambda A_n + \varepsilon, \end{aligned}$$

and the desired relation follows as we let $\varepsilon \to 0$.

To prove the second assertion, we may take $I = [a,b]$ for some finite $a < b$. Since $I \subset (a - \varepsilon, b + \varepsilon)$ for every $\varepsilon > 0$, we get $\lambda I \le |I| + 2\varepsilon$, and so $\lambda I \le |I|$. As for the reverse relation, we need to prove that if $I \subset \bigcup_k I_k$ for some open intervals $I_1, I_2, \ldots$, then $|I| \le \sum_k |I_k|$. By the Heine–Borel theorem[1], I remains covered by finitely many intervals $I_1, \ldots, I_n$, and it suffices to show that $|I| \le \sum_{k \le n} |I_k|$. This reduces the assertion to the case of finitely many covering intervals $I_1, \ldots, I_n$.

The statement is clearly true for a single covering interval. Proceeding by induction, assume the truth for $n-1$ covering intervals, and turn to the case of covering by $I_1, \ldots, I_n$. Then b belongs to some $I_k = (a_k, b_k)$, and so the interval $I'_k = I \setminus I_k$ is covered by the remaining intervals I_j, $j \ne k$. By the induction hypothesis, we get

$$\begin{aligned} |I| &= b - a \\ &\le (b - a_k) + (a_k - a) \\ &\le |I_k| + |I'_k| \\ &\le |I_k| + \sum_{j \ne k} |I_j| = \sum\nolimits_j |I_j|, \end{aligned}$$

as required. □

We show that the class of measurable sets in Lemma 2.3 contains all Borel sets.

[1] A set in R^d is compact iff it is bounded and closed.

Lemma 2.4 *(measurability of intervals) Let λ denote the outer measure in Lemma* 2.3*. Then the interval $(-\infty, a]$ is λ-measurable for every $a \in \mathsf{R}$.*

Proof: For any set $E \subset \mathsf{R}$ and constant $\varepsilon > 0$, we may cover E by some open intervals $I_1, I_2, \ldots$, such that $\lambda E \geq \sum_n |I_n| - \varepsilon$. Writing $I = (-\infty, a]$ and using the sub-additivity of λ and Lemma 2.3, we get

$$\begin{aligned}\lambda E + \varepsilon &\geq \textstyle\sum_n |I_n| \\ &= \textstyle\sum_n |I_n \cap I| + \sum_n |I_n \cap I^c| \\ &= \textstyle\sum_n \lambda(I_n \cap I) + \sum_n \lambda(I_n \cap I^c) \\ &\geq \lambda(E \cap I) + \lambda(E \cap I^c).\end{aligned}$$

Since ε was arbitrary, it follows that I is λ-measurable. □

Proof of Theorem 2.2: Define λ as in (4). Then Lemma 2.3 shows that λ is an outer measure satisfying $\lambda I = |I|$ for every interval I. Furthermore, Theorem 2.1 shows that λ is a measure on the σ-field $\mathcal{S}$ of all λ-measurable sets. Finally, Lemma 2.4 shows that $\mathcal{S}$ contains all intervals $(-\infty, a]$ with $a \in \mathsf{R}$. Since the latter sets generate the Borel σ-field $\mathcal{B}$, we have $\mathcal{B} \subset \mathcal{S}$.

To prove the uniqueness, consider any measure μ with the stated properties, and put $I_n = [-n, n]$ for $n \in \mathsf{N}$. Using Lemma 1.18, with $\mathcal{C}$ as the class of intervals, we see that

$$\lambda(B \cap I_n) = \mu(B \cap I_n), \quad B \in \mathcal{B},\ n \in \mathsf{N}.$$

Letting $n \to \infty$ and using Lemma 1.15, we get $\lambda B = \mu B$ for all $B \in \mathcal{B}$, as required. □

Before proceeding to a more detailed study of Lebesgue measure, we state a general extension theorem that can be proved by essentially the same arguments. Here a non-empty class $\mathcal{I}$ of subsets of a space S is called a *semi-ring*, if for any $I, J \in \mathcal{I}$ we have $I \cap J \in \mathcal{I}$, and the set $I \setminus J$ is a finite union of disjoint sets $I_1, \ldots, I_n \in \mathcal{I}$.

Theorem 2.5 *(extension, Carathéodory) Let μ be a finitely additive, countably sub-additive set function on a semi-ring $\mathcal{I} \subset 2^S$, such that $\mu\emptyset = 0$. Then μ extends to a measure on $\sigma(\mathcal{I})$.*

Proof: Define a set function μ^* on 2^S by

$$\mu^* A = \inf_{\{I_k\}} \textstyle\sum_k \mu I_k, \quad A \subset S,$$

where the infimum extends over all covers of A by sets $I_1, I_2, \ldots \in \mathcal{I}$, and take $\mu^* A = \infty$ when no such cover exists. Proceeding as in the proof of Lemma 2.3, we see that μ^* is an outer measure on S. To see that μ^* extends μ, fix any $I \in \mathcal{I}$, and consider an arbitrary cover $I_1, I_2, \ldots \in \mathcal{I}$ of I. Using the sub-additivity and finite additivity of μ, we get

$$\mu^* I \leq \mu I \leq \textstyle\sum_k \mu(I \cap I_k)$$
$$\leq \textstyle\sum_k \mu I_k,$$

which implies $\mu^* I = \mu I$. By Theorem 2.1, it remains to show that every set $I \in \mathcal{I}$ is μ^*-measurable. Then let $A \subset S$ be covered by some sets $I_1, I_2, \ldots \in \mathcal{I}$ with $\mu^* A \geq \sum_k \mu I_k - \varepsilon$, and proceed as in the proof of Lemma 2.4, noting that $I_n \setminus I$ is a finite disjoint union of sets $I_{nj} \in \mathcal{I}$, and hence $\mu(I_n \setminus I) = \sum_j \mu I_{nj}$ by the finite additivity of μ. □

For every $d \in \mathsf{N}$, we may use Theorem 1.29 to form the *d-dimensional Lebesgue measure* $\lambda^d = \lambda \otimes \cdots \otimes \lambda$ on R^d, generalizing the elementary notions of area and volume. We show that λ^d is invariant under *translations* (or *shifts*), as well as under arbitrary *rotations*, and that either invariance characterizes λ^d up to a constant factor.

Theorem 2.6 *(invariance of Lebesgue measure) Let μ be a locally finite measure on a product space $\mathsf{R}^d \times S$, where $(S, \mathcal{S})$ is a localized Borel space. Then these conditions are equivalent:*

(i) *$\mu = \lambda^d \otimes \nu$ for some locally finite measure ν on S,*
(ii) *μ is invariant under translations of R^d,*
(iii) *μ is invariant under rigid motions of R^d.*

Proof, (ii) $\Rightarrow$ (i): Assume (ii). Write $\mathcal{I}$ for the class of intervals $I = (a, b]$ with rational endpoints. Then for any $I_1, \ldots, I_d \in \mathcal{I}$ and $C \in \mathcal{S}$ with $\nu C < \infty$,

$$\begin{aligned}\mu\big(I_1 \times \cdots \times I_d \times C\big) &= |I_1| \cdots |I_d|\, \nu C \\ &= (\lambda^d \otimes \nu)\big(I_1 \times \cdots \times I_d \times C\big).\end{aligned}$$

For fixed $I_2, \ldots, I_d$ and C, this extends by monotonicity to arbitrary intervals I_1, and then, by the uniqueness in Theorem 2.2, to any $B_1 \in \mathcal{B}$. Proceeding recursively in d steps, we get for arbitrary $B_1, \ldots, B_d \in \mathcal{B}$

$$\mu\big(B_1 \times \cdots \times B_d \times C\big) = (\lambda^d \otimes \nu)\big(B_1 \times \cdots \times B_d \times C\big),$$

which yields (i) by the uniqueness in Theorem 1.29.

(i) $\Rightarrow$ (ii)–(iii): Let $\mu = \lambda^d \otimes \nu$. For any $h = (h_1, \ldots, h_d) \in \mathsf{R}^d$, define the *shift* $\theta_h : \mathsf{R}^d \to \mathsf{R}^d$ by $\theta_h x = x + h$ for all $x \in \mathsf{R}^d$. Then for any intervals $I_1, \ldots, I_d$ and sets $C \in \mathcal{S}$,

$$\begin{aligned}\mu\big(I_1 \times \cdots \times I_d \times C\big) &= |I_1| \cdots |I_d|\, \nu C \\ &= \mu \circ \theta_h^{-1}\big(I_1 \times \cdots \times I_d \times C\big),\end{aligned}$$

where $\theta_h(x, s) = (x + h, s)$. As before, it follows that $\mu = \mu \circ \theta_h^{-1}$.

To see that μ is also invariant under orthogonal transformations ψ on R^d, we note that for any $x, h \in \mathsf{R}^d$

$$\begin{aligned}(\theta_h \circ \psi)x &= \psi x + h\\ &= \psi\big(x + \psi^{-1}h\big)\\ &= \psi(x+h')\\ &= (\psi \circ \theta_{h'})x,\end{aligned}$$

where $h' = \psi^{-1}h$. Since μ is shift-invariant, we obtain

$$\begin{aligned}\mu \circ \psi^{-1} \circ \theta_h^{-1} &= \mu \circ \theta_{h'}^{-1} \circ \psi^{-1}\\ &= \mu \circ \psi^{-1},\end{aligned}$$

where $\psi(x,s) = (\psi x, s)$. Thus, even $\mu \circ \psi^{-1}$ is shift-invariant and hence of the form $\lambda^d \otimes \nu'$. Writing B for the unit ball in R^d, we get for any $C \in \mathcal{S}$

$$\begin{aligned}\lambda^d B \cdot \nu' C &= \mu \circ \psi^{-1}(B \times C)\\ &= \mu\big(\psi^{-1}B \times C\big)\\ &= \mu(B \times C)\\ &= \lambda^d B \cdot \nu C.\end{aligned}$$

Dividing by $\lambda^d B$ gives $\nu' C = \nu C$, and so $\nu' = \nu$, which implies $\mu \circ \psi^{-1} = \mu$. □

We show that any integrable function on R^d is continuous in a suitable average sense.

Lemma 2.7 *(mean continuity) Let f be a measurable function on R^d with $\lambda^d|f| < \infty$. Then*

$$\lim_{h\to 0} \int \big| f(x+h) - f(x) \big|\, dx = 0.$$

Proof: By Lemma 1.37 and a simple truncation, we may choose some continuous functions $f_1, f_2, \ldots$ with bounded supports such that $\lambda^d|f_n - f| \to 0$. By the triangle inequality, we get for $n \in \mathsf{N}$ and $h \in \mathsf{R}^d$

$$\int \big| f(x+h) - f(x) \big|\, dx \le \int \big| f_n(x+h) - f_n(x) \big|\, dx + 2\lambda^d|f_n - f|.$$

Since the f_n are bounded, the right-hand side tends to 0 by dominated convergence, as $h \to 0$ and then $n \to \infty$. □

By a *signed measure* on a localized Borel space $(S, \mathcal{S})$ we mean a function $\nu \colon \hat{\mathcal{S}} \to \mathsf{R}$, such that $\nu \bigcup_n B_n = \sum_n \nu B_n$ for any disjoint sets $B_1, B_2, \ldots \in \hat{\mathcal{S}}$, where the series converges absolutely. Say that the measures μ, ν on $(S, \mathcal{S})$ are (mutually) *singular* and write $\mu \perp \nu$, if there exists an $A \in \mathcal{S}$ with $\mu A = \nu A^c = 0$. Note that A may not be unique. We state the basic decomposition of a signed measure into positive components.

Theorem 2.8 *(Hahn decomposition) For any signed measure ν on S, there exist some unique, locally finite measures $\nu_\pm \ge 0$ on S, such that*

$$\nu = \nu_+ - \nu_-, \qquad \nu_+ \perp \nu_-.$$

Proof: We may take ν to be bounded. Put $c = \sup\{\nu A;\ A \in \mathcal{S}\}$, and note that if $A, A' \in \mathcal{S}$ with $\nu A \geq c - \varepsilon$ and $\nu A' \geq c - \varepsilon'$, then

$$\begin{aligned}\nu(A \cup A') &= \nu A + \nu A' - \nu(A \cap A')\\ &\geq (c-\varepsilon) + (c-\varepsilon') - c\\ &= c - \varepsilon - \varepsilon'.\end{aligned}$$

Choosing $A_1, A_2, \ldots \in \mathcal{S}$ with $\nu A_n \geq c - 2^{-n}$, we get by iteration and countable additivity

$$\begin{aligned}\nu \bigcup_{k>n} A_k &\geq c - \sum_{k>n} 2^{-k}\\ &= c - 2^{-n}, \quad n \in \mathbb{N}.\end{aligned}$$

Define $A_+ = \bigcap_n \bigcup_{k>n} A_k$ and $A_- = A_+^c$. Using the countable additivity again, we get $\nu A_+ = c$. Hence, for sets $B \in \mathcal{S}$,

$$\begin{aligned}\nu B &= \nu A_+ - \nu(A_+ \setminus B) \geq 0, \quad B \subset A_+,\\ \nu B &= \nu(A_+ \cup B) - \nu A_+ \leq 0, \quad B \subset A_-.\end{aligned}$$

We may then define the measures ν_+ and ν_- by

$$\begin{aligned}\nu_+ B &= \nu(B \cap A_+),\\ \nu_- B &= -\nu(B \cap A_-), \qquad B \in \mathcal{S}.\end{aligned}$$

To prove the uniqueness, suppose that also $\nu = \mu_+ - \mu_-$ for some positive measures $\mu_+ \perp \mu_-$. Choose a set $B_+ \in \mathcal{S}$ with $\mu_- B_+ = \mu_+ B_+^c = 0$. Then ν is both positive and negative on the sets $A_+ \setminus B_+$ and $B_+ \setminus A_+$, and therefore $\nu = 0$ on $A_+ \triangle B_+$. Hence, for any $C \in \mathcal{S}$,

$$\begin{aligned}\mu_+ C &= \mu_+(B_+ \cap C)\\ &= \nu(B_+ \cap C)\\ &= \nu(A_+ \cap C) = \nu_+ C,\end{aligned}$$

which shows that $\mu_+ = \nu_+$. Then also

$$\begin{aligned}\mu_- &= \mu_+ - \nu\\ &= \nu_+ - \nu = \nu_-.\end{aligned}$$

□

The last result yields the existence of the *maximum* $\mu \vee \nu$ and *minimum* $\mu \wedge \nu$ of two σ-finite measures μ and ν.

Corollary 2.9 *(maximum and minimum) For any σ-finite measures μ, ν on S,*

(i) *there exist a largest measure $\mu \wedge \nu$ and a smallest one $\mu \vee \nu$ satisfying*

$$\mu \wedge \nu \leq \mu, \nu \leq \mu \vee \nu,$$

(ii) *the measures in* (i) *satisfy*

$$\begin{aligned}(\mu - \mu \wedge \nu) \perp (\nu - \mu \wedge \nu),\\ \mu \wedge \nu + \mu \vee \nu = \mu + \nu.\end{aligned}$$

Proof: We may take μ and ν to be bounded. Writing $\rho_+ - \rho_-$ for the Hahn decomposition of $\mu - \nu$, we put

$$\mu \wedge \nu = \mu - \rho_+,$$
$$\mu \vee \nu = \mu + \rho_-. \qquad \square$$

For any measures μ, ν on $(S, \mathcal{S})$, we say that ν is *absolutely continuous* with respect to μ and write $\nu \ll \mu$, if $\mu A = 0$ implies $\nu A = 0$ for all $A \in \mathcal{S}$. We show that any σ-finite measure has a unique decomposition into an absolutely continuous and a singular component, where the former has a basic integral representation.

Theorem 2.10 *(Lebesgue decomposition, Radon–Nikodym theorem) Let μ, ν be σ-finite measures on S. Then*

(i) *$\nu = \nu_a + \nu_s$ for some unique measures $\nu_a \ll \mu$ and $\nu_s \perp \mu$,*
(ii) *$\nu_a = f \cdot \mu$ for a μ-a.e. unique measurable function $f \geq 0$ on S.*

Two lemmas will be needed for the proof.

Lemma 2.11 *(closure) Consider some measures μ, ν and measurable functions $f_1, f_2, \ldots \geq 0$ on S, and put $f = \sup_n f_n$. Then*

$$f_n \cdot \mu \leq \nu, \;\; n \in \mathsf{N} \quad \Leftrightarrow \quad f \cdot \mu \leq \nu.$$

Proof: First let $f \cdot \mu \leq \nu$ and $g \cdot \mu \leq \nu$, and put $h = f \vee g$. Writing $A = \{f \geq g\}$, we get

$$\begin{aligned} h \cdot \mu &= 1_A h \cdot \mu + 1_{A^c} h \cdot \mu \\ &= 1_A f \cdot \mu + 1_{A^c} g \cdot \mu \\ &\leq 1_A \nu + 1_{A^c} \nu = \nu. \end{aligned}$$

Thus, we may assume that $f_n \uparrow f$. But then $\nu \geq f_n \cdot \mu \uparrow f \cdot \mu$ by monotone convergence, and so $f \cdot \mu \leq \nu$. $\square$

Lemma 2.12 *(partial density) Let μ, ν be bounded measures on S with $\mu \not\perp \nu$. Then there exists a measurable function $f \geq 0$ on S, such that*

$$\mu f > 0, \qquad f \cdot \mu \leq \nu.$$

Proof: For any $n \in \mathsf{N}$, we introduce the signed measure $\chi_n = \nu - n^{-1}\mu$. By Theorem 2.8 there exists a set $A_n^+ \in \mathcal{S}$ with complement A_n^- such that $\pm\chi_n \geq 0$ on $A_n^\pm$. Since the χ_n are non-decreasing, we may assume that $A_1^+ \subset A_2^+ \subset \cdots$. Writing $A = \bigcup_n A_n^+$ and noting that $A^c = \bigcap_n A_n^- \subset A_n^-$, we obtain

$$\begin{aligned} \nu A^c &\leq \nu A_n^- \\ &= \chi_n A_n^- + n^{-1} \mu A_n^- \\ &\leq n^{-1} \mu S \to 0, \end{aligned}$$

and so $\nu A^c = 0$. Since $\mu \not\perp \nu$, we get $\mu A > 0$. Furthermore, $A_n^+ \uparrow A$ implies $\mu A_n^+ \uparrow \mu A > 0$, and we may choose n so large that $\mu A_n^+ > 0$. Putting $f = n^{-1} 1_{A_n^+}$, we obtain $\mu f = n^{-1} \mu A_n^+ > 0$ and

$$\begin{aligned} f \cdot \mu &= n^{-1} 1_{A_n^+} \cdot \mu \\ &= 1_{A_n^+} \cdot \nu - 1_{A_n^+} \cdot \chi_n \le \nu. \end{aligned}$$ □

Proof of Theorem 2.10: We may take μ and ν to be bounded. Let $\mathcal{C}$ be the class of measurable functions $f \ge 0$ on S with $f \cdot \mu \le \nu$, and define $c = \sup\{\mu f;\ f \in \mathcal{C}\}$. Choose $f_1, f_2, \ldots \in \mathcal{C}$ with $\mu f_n \to c$. Then $f \equiv \sup_n f_n \in \mathcal{C}$ by Lemma 2.11, and $\mu f = c$ by monotone convergence. Define $\nu_a = f \cdot \mu$ and $\nu_s = \nu - \nu_a$, and note that $\nu_a \ll \mu$. If $\nu_s \not\perp \mu$, then Lemma 2.12 yields a measurable function $g \ge 0$ with $\mu g > 0$ and $g \cdot \mu \le \nu_s$. But then $f + g \in \mathcal{C}$ with $\mu(f+g) > c$, which contradicts the definition of c. Thus, $\nu_s \perp \mu$.

To prove the uniqueness of ν_a and ν_s, suppose that also $\nu = \nu_a' + \nu_s'$ for some measures $\nu_a' \ll \mu$ and $\nu_s' \perp \mu$. Choose $A, B \in \mathcal{S}$ with $\nu_s A = \mu A^c = \nu_s' B = \mu B^c = 0$. Then clearly

$$\begin{aligned} \nu_s(A \cap B) &= \nu_s'(A \cap B) \\ &= \nu_a(A^c \cup B^c) \\ &= \nu_a'(A^c \cup B^c) = 0, \end{aligned}$$

and so

$$\begin{aligned} \nu_a &= 1_{A \cap B} \cdot \nu_a \\ &= 1_{A \cap B} \cdot \nu \\ &= 1_{A \cap B} \cdot \nu_a' = \nu_a', \\ \nu_s &= \nu - \nu_a \\ &= \nu - \nu_a' = \nu_s'. \end{aligned}$$

To see that f is a.e. unique, suppose that also $\nu_a = g \cdot \mu$ for some measurable function $g \ge 0$. Writing $h = f - g$ and noting that $h \cdot \mu = 0$, we get

$$\mu|h| = \int_{\{h>0\}} h\, d\mu - \int_{\{h<0\}} h\, d\mu = 0,$$

and so $h = 0$ a.e. by Lemma 1.26. □

We insert a simple corollary needed in Chapter 25.

Corollary 2.13 *(splitting) Consider some finite measure spaces $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \nu)$ and a measurable map $f: S \to T$, such that $\nu \le \mu \circ f^{-1}$. Then there exists a measure $\mu' \le \mu$ on S with $\nu = \mu' \circ f^{-1}$.*

Proof: Put $\mu' = (g \circ f) \cdot \mu$ with $g = d\nu / d(\mu \circ f^{-1})$, and use Lemma 1.24. □

Next we extend Theorem 2.2 to a basic correspondence between locally finite measures and non-decreasing functions on R.

Theorem 2.14 *(Lebesgue–Stieltjes measures) There is a* $1-1$ *correspondence between all locally finite measures* μ *on* R *and all non-decreasing, right-continuous functions* F *with* $F(0) = 0$*, given by*

$$\mu(a, b] = F(b) - F(a), \quad -\infty < a < b < \infty. \tag{5}$$

Proof: For any locally finite measure μ on R, we define a function F on R by

$$F(x) = \begin{cases} \mu(0, x], & x \geq 0, \\ -\mu(x, 0], & x < 0. \end{cases}$$

Then F is right-continuous and non-decreasing with $F(0) = 0$, and as such it is clearly uniquely determined by (5).

Conversely, given any F as stated, we define its left-continuous inverse $g \colon \mathsf{R} \to \bar{\mathsf{R}}$ by

$$g(t) = \inf\left\{s \in \mathsf{R};\ F(s) \geq t\right\}, \quad t \in \mathsf{R}.$$

Since g is again non-decreasing, the set $g^{-1}(-\infty, s]$ is an extended interval for each $s \in \mathsf{R}$, and so g is measurable by Lemma 1.4. We may then define a measure μ on $\bar{\mathsf{R}}$ by $\mu = \lambda \circ g^{-1}$, where λ denotes Lebesgue measure on R. Noting that $g(t) \leq x$ iff $t \leq F(x)$, we get for any $a < b$

$$\begin{aligned} \mu(a, b] &= \lambda\left\{t;\ g(t) \in (a, b]\right\} \\ &= \lambda\Big(F(a), F(b)\Big] \\ &= F(b) - F(a). \end{aligned}$$

Thus, the restriction of μ to R satisfies (5). The uniqueness of μ may be proved in the same way as for λ in Theorem 2.2. □

We now specialize Theorem 2.10 to the case where μ equals Lebesgue measure and ν is any locally finite measure on R, defined as in Theorem 2.14 in terms of a non-decreasing, right-continuous function F. The Lebesgue decomposition and Radon–Nikodym property may be expressed in terms of F as

$$F = F_a + F_s = f \cdot \lambda + F_s, \tag{6}$$

where F_a and F_s correspond to the absolutely continuous and singular components of ν, respectively, and we assume that $F_a(0) = 0$. Here $f \cdot \lambda$ denotes the function $\int_0^x f(t)\,dt$, where the *Lebesgue density* $f \colon \mathsf{R} \to \mathsf{R}_+$ is locally integrable. The following result extends the *fundamental theorem of calculus* for Riemann integrals of continuously differentiable functions—the fact that differentiation and integration are mutually inverse operations.

Theorem 2.15 *(differentiation, Lebesgue) Let* F *be a non-decreasing, right-continuous function on* R*. Then* F *is a.e. differentiable, and the Lebesgue decomposition and derivative of* F *are related by*

$$F = f \cdot \lambda + F_s \quad \Leftrightarrow \quad F' = f \ \text{ a.e. } \lambda.$$

Thus, the two parts of the fundamental theorem generalize to

$$(f \cdot \lambda)' = f \ \text{ a.e.}, \qquad F' \cdot \lambda = F_a.$$

In other words, the density of an integral can still be a.e. recovered through differentiation, whereas integration of a derivative yields only the absolutely continuous component of the underlying function. In particular, F is absolutely continuous iff $F' \cdot \lambda = F - F(0)$ and singular iff $F' = 0$ a.e.

The last result extends trivially to any difference $F = F_+ - F_-$ between non-decreasing, right-continuous functions F_+ and F_-. However, it fails for more general functions, where the derivative may not even exist. A prime example is given by the paths of a Brownian motion, as will be seen in Corollary 14.10.

Two lemmas will be helpful to prove the last theorem.

Lemma 2.16 *(interval selection) For any class $\mathcal{I}$ of open intervals with union G satisfying $\lambda G < \infty$, there exist some disjoint sets $I_1, \ldots, I_n \in \mathcal{I}$ with*

$$\textstyle\sum_k |I_k| \geq \lambda G/4.$$

Proof: By Lemma 1.36 we may choose a compact set $K \subset G$ with $\lambda K \geq 3\,\lambda G/4$. By compactness, we may next cover K by finitely many intervals $J_1, \ldots, J_m \in \mathcal{I}$. Now define recursively some $I_1, I_2, \ldots$, where I_k is the longest interval J_r not yet chosen, such that $J_r \cap I_j = \emptyset$ for all $j < k$. The selection terminates when no such interval exists.

If an interval J_r is not selected, it must intersect a longer interval I_k. Writing $\hat{I}_k$ for the interval centered at I_k with length $3|I_k|$, we obtain

$$K \subset \textstyle\bigcup_r J_r \subset \bigcup_k \hat{I}_k,$$

and so

$$\begin{aligned} 3\,\lambda G/4 &\leq \lambda K \leq \lambda \textstyle\bigcup_k \hat{I}_k \\ &\leq \textstyle\sum_k |\hat{I}_k| \\ &= 3 \textstyle\sum_k |I_k|. \end{aligned}$$

□

Lemma 2.17 *(differentiation on null sets) Let $F(x) \equiv \mu(0,x]$ for a locally finite measure μ on R. Then for any set $A \in \mathcal{B}$,*

$$\mu A = 0 \quad \Rightarrow \quad F' = 0 \ \text{ a.e. } 1_A \lambda.$$

Proof: For any $\delta > 0$, Lemma 1.36 yields an open set $G_\delta \supset A$ with $\mu G_\delta < \delta$. Define

$$A_\varepsilon = \Big\{ x \in A;\ \limsup_{h \to 0} \frac{\mu(x-h, x+h)}{h} > \varepsilon \Big\}, \quad \varepsilon > 0,$$

and note that the A_ε are measurable, since the lim sup may be taken along the rationals. For every $x \in A_\varepsilon$ there exists an interval $I = (x-h, x+h) \subset G_\delta$ with $2\mu I > \varepsilon |I|$, and we note that the class $\mathcal{I}_{\varepsilon,\delta}$ of such intervals covers A_ε. Hence, Lemma 2.16 yields some disjoint sets $I_1, \ldots, I_n \in \mathcal{I}_{\varepsilon,\delta}$ satisfying $\sum_k |I_k| \geq \lambda A_\varepsilon/4$. Then

$$\begin{aligned}\lambda A_\varepsilon &\le 4 \textstyle\sum_k |I_k| \\ &\le 8\,\varepsilon^{-1} \textstyle\sum_k \mu I_k \\ &\le 8\,\mu G_\delta/\varepsilon < 8\,\delta/\varepsilon,\end{aligned}$$

and as $\delta \to 0$ we get $\lambda A_\varepsilon = 0$. Thus, $\limsup \mu(x-h, x+h)/h \le \varepsilon$ a.e. λ on A, and the assertion follows since ε is arbitrary. □

Proof of Theorem 2.15: Since $F_s' = 0$ a.e. λ by Lemma 2.17, we may assume that $F = f \cdot \lambda$. Define

$$\begin{aligned}F_+'(x) &= \limsup_{h\to 0} h^{-1}\big\{F(x+h) - F(x)\big\}, \\ F_-'(x) &= \liminf_{h\to 0} h^{-1}\big\{F(x+h) - F(x)\big\},\end{aligned}$$

and note that $F_+' = 0$ a.e. on the set $\{f = 0\} = \{x;\, f(x) = 0\}$ by Lemma 2.17. Applying this result to the function $F_r = (f - r)_+ \cdot \lambda$ for arbitrary $r \in \mathsf{R}$, and noting that $f \le (f-r)_+ + r$, we get $F_+' \le r$ a.e. on $\{f \le r\}$. Thus, for r restricted to the rationals,

$$\begin{aligned}\lambda\{f < F_+'\} &= \lambda \textstyle\bigcup_r \big\{f \le r < F_+'\big\} \\ &\le \textstyle\sum_r \lambda\big\{f \le r < F_+'\big\} = 0,\end{aligned}$$

which shows that $F_+' \le f$ a.e. Applying this to the function $-F = (-f)\cdot\lambda$ yields $F_-' = -(-F)_+' \ge f$ a.e. Thus, $F_+' = F_-' = f$ a.e., and so F' exists a.e. and equals f. □

For a localized Borel space $(S, \mathcal{S})$, let $\mathcal{M}_S$ be the class of locally finite measures on S. It becomes a measurable space in its own right, when endowed with the σ-field induced by the *evaluation maps* $\pi_B : \mu \mapsto \mu B$, $B \in \mathcal{S}$. In particular, the class of probability measures on S is a measurable subset of $\mathcal{M}_S$.

To state the basic atomic decomposition of measures $\mu \in \mathcal{M}_S$, recall that $\delta_s B = 1_B(s)$, and write $\mathcal{M}_S^c$ for the class of *diffuse*[2] measures $\mu \in \mathcal{M}_S$, where $\mu\{s\} = 0$ for all $s \in S$. Put $\bar{\mathsf{Z}}_+ = \{0, 1, \dots; \infty\}$.

Theorem 2.18 *(atomic decomposition) For any measures $\mu \in \mathcal{M}_S$ with S a localized Borel space, we have*

(i)
$$\mu = \alpha + \sum_{k \le \kappa} \beta_k\, \delta_{\sigma_k},$$

for some $\alpha \in \mathcal{M}_S^c$, $\kappa \in \bar{\mathsf{Z}}_+$, $\beta_1, \beta_2, \ldots > 0$, and distinct $\sigma_1, \sigma_2, \ldots \in S$,

(ii) *μ is Z_+-valued iff* (i) *holds with $\alpha = 0$ and $\beta_k \in \mathsf{N}$ for all k,*

(iii) *the representation in* (i) *is unique apart from the order of terms.*

[2] or *non-atomic*

Proof: (i) By localization we may assume that $\mu S < \infty$. Since S is Borel we may let $S \in \mathcal{B}_{[0,1]}$, and by an obvious embedding we may even take $S = [0,1]$. Define $F(x) = \mu[0,x]$ as in Theorem 2.14, and note that the possible atoms of μ correspond uniquely to the jumps of F. The latter may be listed in non-increasing order as a sequence of pairs $\{x, \Delta F(x)\}$, equivalent to the set of atom locations and sizes (σ_k, β_k) of μ. By Corollary 1.17, subtraction of the atoms yields a measure α, which is clearly diffuse.

(ii) Let μ be Z_+-valued, and suppose that $\beta_k \notin \mathsf{N}$ for some k. Assuming $S = [0,1]$, we may choose some open balls $B_n \downarrow \{\sigma_k\}$. Then $\mu B_n \downarrow \beta_k$, and so $\mu B_n \notin \mathsf{Z}_+$ for large n. The contradiction shows that $\beta_k \in \mathsf{N}$ for all k. To see that even $\alpha = 0$, suppose that instead $\alpha \neq 0$. Writing $A = S \setminus \bigcup_k \{\sigma_k\}$, we get $\mu A > 0$, and so we may choose an $s \in \operatorname{supp}(1_A \mu)$ and some balls $B_n \downarrow \{s\}$. Then $0 < \mu(A \cap B_n) \downarrow 0$, which yields another contradiction.

(iii) Consider two such decompositions

$$\begin{aligned}\mu &= \alpha + \sum_{k \le \kappa} \beta_k \, \delta_{\sigma_k} \\ &= \alpha' + \sum_{k \le \kappa'} \beta'_k \, \delta_{\sigma'_k}.\end{aligned}$$

Pairing off the atoms in the two sums until no such terms remain, we are left with the equality $\alpha = \alpha'$. □

Let $\mathcal{N}_S$ be the class of Z_+-valued measures $\mu \in \mathcal{M}_S$, also known as *point measures* on S. Say that μ is *simple* if all coefficients β_k equal 1, so that μ is the counting measure on its support. The class of simple point measures on S is denoted by $\mathcal{N}_S^*$. Given a measure $\mu \in \mathcal{M}_S$, we write μ_c for the diffuse component of μ. For point measures $\mu \in \mathcal{N}_S$, we write μ^* for the simple measure on the support of μ, obtained by reducing all β_k to 1.

Theorem 2.19 *(measurability) For any localized Borel space S and measures $\mu \in \mathcal{M}_S$, we have*

(i) *the coefficients α, κ, (β_k), (σ_k) in Theorem 2.18 can be chosen to be measurable functions of μ,*

(ii) *the classes $\mathcal{M}_S^c$, $\mathcal{N}_S$, $\mathcal{N}_S^*$ are measurable subsets of $\mathcal{M}_S$,*

(iii) *the maps $\mu \mapsto \mu_c$ and $\mu \mapsto \mu^*$ are measurable on $\mathcal{M}_S$ and $\mathcal{N}_S$,*

(iv) *the class of degenerate measures $\beta \, \delta_\sigma$ is measurable,*

(v) *the mapping $(\mu, \nu) \mapsto \mu \otimes \nu$ is a measurable function on $\mathcal{M}_S^2$.*

Proof: (i) We may take $S = [0,1)$ and assume that $\mu S < \infty$. Put $I_{nj} = 2^{-n}[j-1, j)$ for $j \le 2^n$, $n \in \mathsf{N}$, and define $\gamma_{nj} = \mu I_{nj}$ and $\tau_{nj} = 2^{-n} j$. For any constants $b > a > 0$, let $(\beta_{nk}, \sigma_{nk})$ with $k \le \kappa_n$ be the pairs (γ_{nj}, τ_{nj}) with $a \le \gamma_{nj} < b$, listed in the order of increasing τ_{nj}. Since $\mu B_n \downarrow \mu\{x\}$ for any open balls $B_n \downarrow \{x\}$, we note that

$$\Big\{\beta_{n1}, \beta_{n2}, \dots ; \, \sigma_{n1}, \sigma_{n2}, \dots\Big\} \to \Big\{\beta_1, \beta_2, \dots ; \, \sigma_1, \sigma_2, \dots\Big\},$$

where the β_k and σ_k are atom sizes and positions of μ with $a \le \beta_k < b$ and increasing σ_k. The measurability on the right now follows by Lemma 1.11. To see that the atomic component of μ is measurable, it remains to partition $(0,\infty)$ into countably many intervals $[a,b)$. The measurability of the diffuse part α now follows by subtraction.

(ii)−(iv): This is clear from (i), since the mentioned classes and operations can be measurably expressed in terms of the coefficients.

(v) Note that $(\mu\otimes\nu)(B\times C)$ is measurable for any $B, C \in \mathcal{S}$, and extend by a monotone-class argument. □

Let $S^{(n)}$ denote the *non-diagonal* part of S^n, consisting of all n-tuples $(s_1,\ldots,s_n)$ with distinct components s_k. When $\mu \in \mathcal{N}_S^*$, we may define the *factorial measure* $\mu^{(n)}$ as the restriction of the product measure μ^n to $S^{(n)}$. The definition for general $\mu \in \mathcal{N}_S$ is more subtle. Write $\hat{\mathcal{N}}_S$ for the class of bounded measures in $\mathcal{N}_S$, and put $\mu^{(1)} = \mu$.

Lemma 2.20 *(factorial measures) For any measure $\mu = \sum_{i\in I}\delta_{\sigma_i}$ in $\hat{\mathcal{N}}_S$, these conditions are equivalent and define some measures $\mu^{(n)}$ on S^n:*

(i) $\displaystyle \mu^{(n)} = \sum_{i\in I^{(n)}} \delta_{\sigma_{i_1},\ldots,\sigma_{i_n}},$

(ii) $\displaystyle \mu^{(m+n)}f = \int \mu^{(m)}(ds)\int\Big(\mu - \sum_{i\le m}\delta_{s_i}\Big)^{(n)}(dt)\, f(s,t),$

(iii) $\displaystyle \sum_{n\ge 0}\frac{1}{n!}\int \mu^{(n)}(ds)\, f\Big(\sum_{i\le n}\delta_{s_i}\Big) = \sum_{\nu\le\mu} f(\nu).$

Proof: When μ is simple, (i) agrees with the elementary definition. Since also $\mu^{(1)} = \mu$, we may take (i) as our definition. We also put $\mu^{(0)} = 1$ for consistency. Since the $\mu^{(n)}$ are uniquely determined by both (ii) and (iii), it remains to show that each of those relations follows from (i).

(ii) Write $I_i = I\setminus\{i_1,\ldots,i_m\}$ for any $i = (i_1,\ldots,i_m) \in I^{(m)}$. Then by (i),

$$\begin{aligned}
&\int \mu^{(m)}(ds)\int\Big(\mu - \sum_{i\le m}\delta_{s_i}\Big)^{(n)}(dt)\, f(s,t)\\
&\quad = \sum_{i\in I^{(m)}}\int\Big(\mu - \sum_{k\le m}\delta_{\sigma_{i_k}}\Big)^{(n)}(dt)\, f\big(\sigma_{i_1},\ldots,\sigma_{i_k},t\big)\\
&\quad = \sum_{i\in I^{(m)}}\int\Big(\sum_{j\in I_i}\delta_{\sigma_j}\Big)^{(n)}(dt)\, f\big(\sigma_{i_1},\ldots,\sigma_{i_k},t\big)\\
&\quad = \sum_{i\in I^{(m)}}\sum_{j\in I_i^{(n)}} f\big(\sigma_{i_1},\ldots,\sigma_{i_m};\sigma_{j_1},\ldots,\sigma_{j_n}\big)\\
&\quad = \sum_{i\in I^{(m+n)}} f\big(\sigma_{i_1},\ldots,\sigma_{i_{m+n}}\big) = \mu^{(m+n)}f.
\end{aligned}$$

(iii) Given an enumeration $\mu = \sum_{i\in I}\delta_{\sigma_i}$ with $I \subset \mathsf{N}$, put $\hat{I}^{(n)} = \{i \in I^n;\ i_1 < \cdots < i_n\}$ for $n > 0$. Let $\mathcal{P}_n$ be the class of permutations of $1,\ldots,n$. Using the atomic representation of μ and the tetrahedral decomposition of $I^{(n)}$, we get

$$\begin{aligned}
\int \mu^{(n)}(ds)\, f\Big(\sum_{i\le n}\delta_{s_i}\Big) &= \sum_{i\in I^{(n)}} f\Big(\sum_{k\le n}\delta_{\sigma_{i_k}}\Big)\\
&= \sum_{i\in \hat I^{(n)}}\sum_{\pi\in\mathcal{P}_n} f\Big(\sum_{k\le n}\delta_{\sigma_{\pi\circ i_k}}\Big)\\
&= n!\sum_{i\in \hat I^{(n)}} f\Big(\sum_{k\le n}\delta_{\sigma_{i_k}}\Big)\\
&= n!\sum_{\nu\le\mu}\{f(\nu);\ \|\nu\|=n\}.
\end{aligned}$$

Now divide by $n!$ and sum over n. □

We now explore the relationship between signed measures and functions of bounded variation. For any function F on R, we define the *total variation* on the interval $[a,b]$ as

$$\|F\|_a^b = \sup_{\{t_k\}} \textstyle\sum_k \Big|F(t_k)-F(t_{k-1})\Big|,$$

where the supremum extends over all finite partitions $a=t_0<t_1<\cdots<t_n=b$. The *positive* and *negative variations* of F are defined by similar expressions, though with $|x|$ replaced by $x^\pm=(\pm x)\vee 0$, so that $x=x^+-x^-$ and $|x|=x^++x^-$. We also write $\Delta_a^bF=F(b)-F(a)$.

We begin with a basic decomposition of functions of locally finite variation, corresponding to the Hahn decomposition in Theorem 2.8.

Theorem 2.21 *(Jordan decomposition) For any function F on R of locally finite variation,*

(i) *$F=F_+-F_-$ for some non-decreasing functions F_+ and F_-, and*

$$\|F\|_s^t \le \Delta_s^tF_+ + \Delta_s^tF_-,\quad s<t,$$

(ii) *equality holds in* (i) *iff $\Delta_s^tF_\pm$ agree with the positive and negative variations of F on $(s,t]$.*

Proof: For any $s<t$, we have

$$\begin{aligned}
(\Delta_s^tF)^+ &= (\Delta_s^tF)^- + \Delta_s^tF,\\
|\Delta_s^tF| &= (\Delta_s^tF)^+ + (\Delta_s^tF)^-\\
&= 2\,(\Delta_s^tF)^- + \Delta_s^tF.
\end{aligned}$$

Summing over the intervals in an arbitrary partition $s=t_0<t_1<\cdots<t_n=t$ and taking the supremum of each side, we obtain

$$\begin{aligned}
\Delta_s^tF_+ &= \Delta_s^tF_- + \Delta_s^tF,\\
\|F\|_s^t &= 2\,\Delta_s^tF_- + \Delta_s^tF\\
&= \Delta_s^tF_+ + \Delta_s^tF_-,
\end{aligned}$$

where $F_\pm(x)$ denote the positive and negative variations of F on $[0,x]$ (apart from the sign when $x<0$). Thus, $F=F(0)+F_+-F_-$, and the stated relation holds with equality. If also $F = G_+ - G_-$ for some non-decreasing functions $G_\pm$, then $(\Delta_s^t F)^\pm \le \Delta_s^t G_\pm$, and so $\Delta_s^t F_\pm \le \Delta_s^t G_\pm$. Thus, $\|F\|_s^t \le \Delta_s^t G_+ + \Delta_s^t G_-$, with equality iff $\Delta_s^t F_\pm = \Delta_s^t G_\pm$. □

We turn to another useful decomposition of finite-variation functions.

Theorem 2.22 *(left and right continuity) For any function F on R of locally finite variation,*

(i) *$F = F_r + F_l$, where F_r is right-continuous with left-hand limits and F_l is left-continuous with right-hand limits,*

(ii) *if F is right-continuous, then so are the minimal components $F_\pm$ in Theorem* 2.21.

Proof: (i) By Proposition 2.21 we may take F to be non-decreasing. The right- and left-hand limits $F^\pm(s)$ then exist at every point s, and we note that $F^-(s) \le F(s) \le F^+(s)$. Further note that F has at most countably many jump discontinuities. For $t > 0$, we define

$$\begin{aligned} F_l(t) &= \sum_{s\in[0,t)} \big\{F^+(s) - F(s)\big\}, \\ F_r(t) &= F(t) - F_l(t). \end{aligned}$$

When $t \le 0$ we need to take the negative of the corresponding sum on $(t, 0]$. It is easy to check that F_l is left-continuous while F_r is right-continuous, and that both functions are non-decreasing.

(ii) Let F be right-continuous at a point s. If $\|F\|_s^t \to c > 0$ as $t \downarrow s$, we may choose $t - s$ so small that $\|F\|_s^t < 4c/3$. Next we may choose a partition $s = t_0 < t_1 < \cdots < t_n = t$ of $[s,t]$, such that the corresponding F-increments δ_k satisfy $\sum_k |\delta_k| > 2c/3$. By the right continuity of F at s, we may assume $t_1 - s$ to be small enough that $\delta_1 = |F(t_1) - F(s)| < c/3$. Then $\|F\|_{t_1}^t > c/3$, and so

$$\begin{aligned} 4c/3 &> \|F\|_s^t \\ &= \|F\|_s^{t_1} + \|F\|_{t_1}^t \\ &> c + c/3 = 4c/3, \end{aligned}$$

a contradiction. Hence $c = 0$. Assuming $F_\pm$ to be minimal, we obtain

$$\Delta_s^t F_\pm \le \|F\|_s^t \to 0, \quad t \downarrow s.$$

□

Justified by the last theorem, we may choose our finite-variation functions F to be right-continuous, which enables us to prove a basic correspondence with locally finite signed measures. We also consider the *jump part* and *atomic decompositions* $F = F^c + F^d$ and $\nu = \nu^d + \nu^a$, where for functions F or signed measures ν on R_+ we define

$$F_t^d = \sum_{s\le t} \Delta F_s, \qquad \nu^a = \sum_{t\ge 0} \nu\{t\}\,\delta_t.$$

Theorem 2.23 *(signed measures and finite-variation functions) Let F be a right-continuous function on R of locally finite variation. Then*

(i) *there exists a unique signed measure ν on R, such that*

$$\nu(s,t] = \Delta_s^t F, \quad s < t,$$

(ii) *the Hahn decomposition $\nu = \nu_+ - \nu_-$ and Jordan decomposition $F = F_+ - F_-$ into minimal components are related by*

$$\nu_\pm(s,t] \equiv \Delta_s^t F_\pm, \quad s < t,$$

(iii) *the atomic decomposition $\nu = \nu^d + \nu^a$ and jump part decomposition $F = F^c + F^d$ are related by*

$$\nu^d(s,t] \equiv \Delta_s^t F^c, \qquad \nu^a(s,t] \equiv \Delta_s^t F^d.$$

Proof: (i) The positive and negative variations $F_\pm$ are right-continuous by Proposition 2.22. Hence, Proposition 2.14 yields some locally finite measures $\mu_\pm$ on R with $\mu_\pm(s,t] \equiv \Delta_s^t F_\pm$, and we may take $\nu = \mu_+ - \mu_-$.

(ii) Choose an $A \in \mathcal{S}$ with $\nu_+ A^c = \nu_- A = 0$. For any $B \in \mathcal{B}$, we get

$$\begin{aligned} \mu_+ B &\geq \mu_+(B \cap A) \\ &\geq \nu(B \cap A) \\ &= \nu_+(B \cap A) = \nu_+ B, \end{aligned}$$

which shows that $\mu_+ \geq \nu_+$. Then also $\mu_- \geq \nu_-$. If the equality fails on some interval $(s,t]$, then

$$\begin{aligned} \|F\|_s^t &= \mu_+(s,t] + \mu_-(s,t] \\ &> \nu_+(s,t] + \nu_-(s,t], \end{aligned}$$

contradicting Proposition 2.21. Hence, $\mu_\pm = \nu_\pm$.

(iii) This is elementary, given the atomic decomposition in Theorem 2.18. □

A function $F : \mathsf{R} \to \mathsf{R}$ is said to be *absolutely continuous*, if for any $a < b$ and $\varepsilon > 0$ there exists a $\delta > 0$, such that for any finite collection of disjoint intervals $(a_k, b_k] \subset (a, b]$ with $\sum_k |b_k - a_k| < \delta$, we have $\sum_k |F(b_k) - F(a_k)| < \varepsilon$. In particular, every absolutely continuous function is continuous with locally finite variation.

Given a function F of locally finite variation, we say that F is *singular*, if for any $a < b$ and $\varepsilon > 0$ there exists a finite set of disjoint intervals $(a_k, b_k] \subset (a, b]$, such that

$$\textstyle\sum_k |b_k - a_k| < \varepsilon, \qquad \|F\|_a^b < \sum_k |F(b_k) - F(a_k)| + \varepsilon.$$

A locally finite, signed measure ν on R is said to be *absolutely continuous* or *singular*, if the components $\nu_\pm$ in the associated Hahn decomposition satisfy $\nu_\pm \ll \lambda$ or $\nu_\pm \perp \lambda$, respectively. We proceed to relate the notions of absolute continuity and singularity for functions and measures.

Theorem 2.24 *(absolutely continuous and singular functions) Let F be a right-continuous function of locally finite variation on* R*, and let ν be the signed measure on* R *with $\nu(s,t] \equiv \Delta_s^t F$. Then*

(i) *F is absolutely continuous iff $\nu \ll \lambda$,*

(ii) *F is singular iff $\nu \perp \lambda$.*

Proof: If F is absolutely continuous or singular, then the corresponding property holds for the total variation function $\|F\|_a^x$ with arbitrary a, and hence also for the minimal components $F_\pm$ in Proposition 2.23. Thus, we may take F to be non-decreasing, so that ν is a positive and locally finite measure on R.

First let F be absolutely continuous. If $\nu \not\ll \lambda$, there exists a bounded interval $I = (a,b)$ with subset $A \in \mathcal{B}$, such that $\lambda A = 0$ but $\nu A > 0$. Taking $\varepsilon = \nu A/2$, choose a corresponding $\delta > 0$, as in the definition of absolute continuity. Since A is measurable with outer Lebesgue measure 0, we may next choose an open set G with $A \subset G \subset I$ such that $\lambda G < \delta$. But then $\nu A \le \nu G < \varepsilon = \nu A/2$, a contradiction. This shows that $\nu \ll \lambda$.

Next let F be singular, and fix any bounded interval $I = (a,b]$. Given any $\varepsilon > 0$, there exist some Borel sets $A_1, A_2, \ldots \subset I$, such that $\lambda A_n < \varepsilon\, 2^{-n}$ and $\nu A_n \to \nu I$. Then $B = \bigcup_n A_n$ satisfies $\lambda B < \varepsilon$ and $\nu B = \nu I$. Next we may choose some Borel sets $B_n \subset I$ with $\lambda B_n \to 0$ and $\nu B_n = \nu I$. Then $C = \bigcap_n B_n$ satisfies $\lambda C = 0$ and $\nu C = \nu I$, which shows that $\nu \perp \lambda$ on I.

Conversely, let $\nu \ll \lambda$, so that $\nu = f \cdot \lambda$ for some locally integrable function $f \ge 0$. Fix any bounded interval I, put $A_n = \{x \in I;\, f(x) > n\}$, and let $\varepsilon > 0$ be arbitrary. Since $\nu A_n \to 0$ by Lemma 1.15, we may choose n so large that $\nu A_n < \varepsilon/2$. Put $\delta = \varepsilon/2n$. For any Borel set $B \subset I$ with $\lambda B < \delta$, we obtain

$$\begin{aligned} \nu B &= \nu(B \cap A_n) + \nu(B \cap A_n^c) \\ &\le \nu A_n + n\,\lambda B \\ &< \tfrac{1}{2}\varepsilon + n\,\delta = \varepsilon. \end{aligned}$$

In particular, this applies to any finite union B of intervals $(a_k, b_k] \subset I$, and so F is absolutely continuous.

Finally, let $\nu \perp \lambda$. Fix any finite interval $I = (a,b]$, and choose a Borel set $A \subset I$ such that $\lambda A = 0$ and $\nu A = \nu I$. For any $\varepsilon > 0$ there exists an open set $G \supset A$ with $\lambda G < \varepsilon$. Letting (a_n, b_n) denote the connected components of G and writing $I_n = (a_n, b_n]$, we get $\sum_n |I_n| < \varepsilon$ and $\sum_n \nu(I \cap I_n) = \nu I$. This shows that F is singular. □

We conclude with yet another basic extension theorem.[3] Here the underlying space S is taken to be a *locally compact, second countable, Hausdorff* topological space (abbreviated as *lcscH*). Let $\mathcal{G}$, $\mathcal{F}$, $\mathcal{K}$ be the classes of open, closed, and compact sets in S, and put $\hat{\mathcal{G}} = \{G \in \mathcal{G};\, \bar{G} \in \mathcal{K}\}$. Let $\hat{C}_+ = \hat{C}_+(S)$ be the class of continuous functions $f : S \to \mathsf{R}_+$ with compact support, defined

[3]Further basic measure extensions appear in Chapters 3 and 8.

as the closure of the set $\{x \in S;\ f(x) > 0\}$. By $U \prec f \prec V$ we mean that $f \in \hat{C}_+$ with $0 \le f \le 1$, such that

$$f = 1 \ \text{ on } U, \qquad \operatorname{supp} f \subset V^o.$$

A *positive linear functional* on $\hat{C}_+$ is defined as a mapping $\mu\colon \hat{C}_+ \to \mathsf{R}_+$, such that

$$\mu(f+g) = \mu f + \mu g, \quad f, g \in \hat{C}_+.$$

This clearly implies the homogeneity $\mu(cf) = c\,\mu f$ for any $f \in \hat{C}_+$ and $c \in \mathsf{R}_+$. A measure μ on the Borel σ-field $\mathcal{S} = \mathcal{B}_S$ is said to be *locally finite*[4], if $\mu K < \infty$ for every $K \in \mathcal{K}$. We show how any positive linear functional can be extended to a measure.

Theorem 2.25 *(Riesz representation) Let μ be a positive linear functional on $\hat{C}_+(S)$, where S is lcscH. Then μ extends uniquely to a locally finite measure on S.*

Several lemmas are needed for the proof, beginning with a simple topological fact.

Lemma 2.26 *(partition of unity) For a compact set $K \subset S$ covered by the open sets $G_1, \dots, G_n$, we may choose some functions $f_1, \dots, f_n \in \hat{C}_+(S)$ with $f_k \prec G_k$, such that $\sum_k f_k = 1$ on K.*

Proof: For any $x \in K$, we may choose some $k \le n$ and $V \in \hat{\mathcal{G}}$ with $x \in V$ and $\bar{V} \subset G_k$. By compactness, K is covered by finitely many such sets $V_1, \dots, V_m$. For every $k \le n$, let U_k be the union of all sets V_j with $\bar{V}_j \subset G_k$. Then $\bar{U}_k \subset G_k$, and so we may choose $g_1, \dots, g_n \in \hat{C}_+$ with $U_k \prec g_k \prec G_k$. Define

$$f_k = g_k(1-g_1)\cdots(1-g_{k-1}), \quad k = 1, \dots, n.$$

Then $f_k \prec G_k$ for all k, and by induction

$$f_1 + \cdots + f_n = 1 - (1-g_1)\cdots(1-g_n).$$

It remains to note that $\prod_k (1-g_k) = 0$ on K since $K \subset \bigcup_k U_k$. □

By an *inner content* on an lcscH space S we mean a non-decreasing function $\mu : \mathcal{G} \to \bar{\mathsf{R}}_+$, finite on $\hat{\mathcal{G}}$, such that μ is finitely additive and countably subadditive, and also satisfies the *inner continuity*

$$\mu G = \sup\Big\{\mu U;\ U \in \hat{\mathcal{G}},\ \bar{U} \subset G\Big\}, \quad G \in \mathcal{G}. \tag{7}$$

Lemma 2.27 *(inner approximation) For a positive linear functional μ on $\hat{C}_+(S)$, an inner content ν on S is given by*

$$\nu G = \sup\Big\{\mu f;\ f \prec G\Big\}, \quad G \in \mathcal{G}.$$

[4] also called a *Radon measure*

Proof: Note that ν is non-decreasing with $\nu\emptyset = 0$, and that $\nu G < \infty$ for bounded G. Moreover, ν is inner continuous in the sense of (7).

To see that ν is countably sub-additive, fix any $G_1, G_2, \ldots \in \mathcal{G}$, and let $f \prec \bigcup_k G_k$. The compactness implies $f \prec \bigcup_{k\le n} G_k$ for a finite n, and Lemma 2.26 yields some functions $g_k \prec G_k$, such that $\sum_k g_k = 1$ on $\operatorname{supp} f$. The products $f_k = g_k f$ satisfy $f_k \prec G_k$ and $\sum_k f_k = f$, and so

$$\mu f = \sum_{k\le n} \mu f_k \le \sum_{k\le n} \nu G_k \le \sum_{k\ge 1} \nu G_k.$$

Since $f \prec \bigcup_k G_k$ was arbitrary, we obtain $\nu \bigcup_k G_k \le \sum_k \nu G_k$, as required.

To see that ν is finitely additive, fix any disjoint sets $G, G' \in \mathcal{G}$. If $f \prec G$ and $f' \prec G'$, then $f + f' \prec G \cup G'$, and so

$$\begin{aligned} \mu f + \mu f' &= \mu(f + f') \\ &\le \nu(G \cup G') \\ &\le \nu G + \nu G'. \end{aligned}$$

Taking the supremum over all f and f' gives $\nu G + \nu G' = \nu(G \cup G')$, as required. □

An outer measure μ on S is said to be *regular*, if it is finitely additive on $\mathcal{G}$ and satisfies the *outer* and *inner regularity*

$$\mu A = \inf\Big\{\mu G;\ G \in \mathcal{G},\ G \supset A\Big\}, \qquad A \subset S, \tag{8}$$

$$\mu G = \sup\Big\{\mu K;\ K \in \mathcal{K},\ K \subset G\Big\}, \quad G \in \mathcal{G}. \tag{9}$$

Lemma 2.28 *(outer approximation) Any inner content μ on S can be extended to a regular outer measure.*

Proof: We may define the extension by (8), since the right-hand side equals μA when $A \in \mathcal{G}$. By the finite additivity on $\mathcal{G}$, we have $2\,\mu\emptyset = \mu\emptyset < \infty$, which implies $\mu\emptyset = 0$. To prove the countable sub-additivity, fix any $A_1, A_2, \ldots \subset S$. For any $\varepsilon > 0$, we may choose some $G_1, G_2, \ldots \in \mathcal{G}$ with $G_n \supset A_n$ and $\mu G_n \le \mu A_n + \varepsilon 2^{-n}$. Since μ is sub-additive on $\mathcal{G}$, we get

$$\begin{aligned} \mu \textstyle\bigcup_n A_n &\le \mu \textstyle\bigcup_n G_n \\ &\le \textstyle\sum_n \mu G_n \\ &\le \textstyle\sum_n \mu A_n + \varepsilon. \end{aligned}$$

The desired relation follows since ε was arbitrary. Thus, the extension is an outer measure on S. Finally, the inner regularity in (9) follows from (7) and the monotonicity of μ. □

Lemma 2.29 *(measurability) If μ is a regular outer measure on S, then every Borel set in S is μ-measurable.*

Proof: Fix any $F \in \mathcal{F}$ and $A \subset G \in \mathcal{G}$. By the inner regularity in (9), we may choose $G_1, G_2, \ldots \in \mathcal{G}$ with $\bar{G}_n \subset G \setminus F$ and $\mu G_n \to \mu(G \setminus F)$. Since μ is non-decreasing and finitely additive on $\mathcal{G}$, we get

$$\begin{aligned}\mu G &\geq \mu(G \setminus \partial G_n)\\ &= \mu G_n + \mu(G \setminus \bar{G}_n)\\ &\geq \mu G_n + \mu(G \cap F)\\ &\to \mu(G \setminus F) + \mu(G \cap F)\\ &\geq \mu(A \setminus F) + \mu(A \cap F).\end{aligned}$$

The outer regularity in (8) gives

$$\mu A \geq \mu(A \setminus F) + \mu(A \cap F), \quad F \in \mathcal{F},\ A \subset S.$$

Hence, every closed set is measurable, and the measurability extends to $\sigma(\mathcal{F}) = \mathcal{B}_S = \mathcal{S}$ by Theorem 2.1. □

Proof of Theorem 2.25: Construct an inner content ν as in Lemma 2.27, and conclude from Lemma 2.28 that ν admits an extension to a regular outer measure on S. By Theorem 2.1 and Lemma 2.29, the restriction of the latter to $\mathcal{S} = \mathcal{B}_S$ is a Radon measure on S, here still denoted by ν.

To see that $\mu = \nu$ on $\hat{C}_+$, fix any $f \in \hat{C}_+$. For any $n \in \mathsf{N}$ and $k \in \mathsf{Z}_+$, let

$$\begin{aligned}f_k^n(x) &= \big(nf(x) - k\big)_+ \wedge 1,\\ G_k^n &= \{nf > k\} = \{f_k^n > 0\}.\end{aligned}$$

Noting that $\bar{G}_{k+1}^n \subset \{f_k^n = 1\}$ and using the definition of ν and the outer regularity in (8), we get for appropriate k

$$\begin{aligned}\nu f_{k+1}^n \leq \nu G_{k+1}^n &\leq \mu f_k^n\\ &\leq \nu \bar{G}_k^n \leq \nu f_{k-1}^n.\end{aligned}$$

Writing $G_0 = G_0^n = \{f > 0\}$ and noting that $nf = \sum_k f_k^n$, we obtain

$$\begin{aligned}n\,\nu f - \nu G_0 &\leq n\,\mu f\\ &\leq n\,\nu f + \nu \bar{G}_0.\end{aligned}$$

Here $\nu \bar{G}_0 < \infty$ since G_0 is bounded. Dividing by n and letting $n \to \infty$ gives $\mu f = \nu f$.

To prove the asserted uniqueness, let μ and ν be locally finite measures on S with $\mu f = \nu f$ for all $f \in \hat{C}_+$. By an inner approximation, we have $\mu G = \nu G$ for every $G \in \mathcal{G}$, and so a monotone-class argument yields $\mu = \nu$. □

Exercises

1. Show that any countably additive set function $\mu \geq 0$ on a field $\mathcal{F}$ with $\mu\emptyset = 0$ extends to a measure on $\sigma(\mathcal{F})$. Further show that the extension is unique whenever μ is bounded.

2. Construct d-dimensional Lebesgue measure λ_d directly, by the method of Theorem 2.2. Then show that $\lambda_d = \lambda^d$.

3. Derive the existence of d-dimensional Lebesgue measure from Riesz' representation theorem, using basic properties of the Riemann integral.

4. Let λ denote Lebesgue measure on R_+, and fix any $p > 0$. Show that the class of step functions with bounded support and finitely many jumps is dense in $L^p(\lambda)$. Generalize to R_+^d.

5. Show that if $\mu_1 = f_1 \cdot \mu$ and $\mu_2 = f_2 \cdot \mu$, then $\mu_1 \vee \mu_2 = (f_1 \vee f_2) \cdot \mu$ and $\mu_1 \wedge \mu_2 = (f_1 \wedge f_2) \cdot \mu$. In particular, we may take $\mu = \mu_1 + \mu_2$. Extend the result to sequences $\mu_1, \mu_2, \ldots$.

6. Given any family μ_i, $i \in I$, of σ-finite measures on a measurable space S, prove the existence of a largest measure $\mu = \bigwedge_n \mu_n$, such that $\mu \leq \mu_i$ for all $i \in I$. Further show that if the μ_i are bounded by some σ-finite measure ν, there exists a smallest measure $\hat{\mu} = \bigvee_n \mu_i$, such that $\mu_i \leq \hat{\mu}$ for all i. (*Hint:* Use Zorn's lemma.)

7. For any bounded, signed measure ν on $(S, \mathcal{S})$, prove the existence of a smallest measure $|\nu|$ such that $|\nu A| \leq |\nu| A$ for all $A \in \mathcal{S}$. Further show that $|\nu| = \nu_+ + \nu_-$, where $\nu_\pm$ are the components in the Hahn decomposition of ν. Finally, for any bounded, measurable function f on S, show that $|\nu f| \leq |\nu|\,|f|$.

8. Extend the last result to complex-valued measures $\chi = \mu + i\nu$, where μ and ν are bounded, signed measures on $(S, \mathcal{S})$. Introducing the complex-valued Radon–Nikodym density $f = d\chi/d(|\mu| + |\nu|)$, show that $|\chi| = |f| \cdot (|\mu| + |\nu|)$.

Chapter 3

Kernels, Disintegration, and Invariance

Kernel criteria and properties, composition and product, disintegration of measures and kernels, partial and iterated disintegration, Haar measure, factorization and modularity, orbit measures, invariant measures and kernels, invariant disintegration, asymptotic and local invariance

So far we have mostly dealt with standard measure theory, typically covered by textbooks in real analysis, though here presented with a slightly different emphasis to suit our subsequent needs. Now we come to some special topics of fundamental importance in probability theory, which are rarely treated even in more ambitious texts.

The hurried reader may again feel tempted to bypass this material and move on to the probabilistic parts of the book. This may be fine as far as the classical probability is concerned. However, already when coming to conditioning and compensation, and to discussions of the general Markov property and continuous-time chains, the serious student may feel the need to return for reference and to catch up with some basic ideas. The need may be even stronger as he/she moves on to the later and more advanced chapters.

Our plans are to begin with a discussion of kernels, which are indispensable in the contexts of conditioning, Markov processes, random measures, and many other areas. They often arise by disintegration of suitable measures on a product space. A second main theme is the theory of invariant measures and disintegrations, of importance in the context of stationary distributions and associated Palm measures. We conclude with a discussion of locally and asymptotically invariant measures, needed in connection with Poisson and Cox convergence of particle systems.

Given some measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, we define a *kernel* $\mu\colon S \to T$ as a measurable function μ from S to the space of measures on T. In other words, μ is a function of two variables $s \in S$ and $B \in \mathcal{T}$, such that $\mu(s, B)$ is $\mathcal{S}$-measurable in s for fixed B and a measure in B for fixed s. For a simple example, we note that the set of Dirac measures δ_s, $s \in S$, can be regarded as a kernel $\delta\colon S \to S$.

Just as every measure μ on S can be identified with a linear functional $f \mapsto \mu f$ on $\mathcal{S}_+$, we may identify a kernel $\mu : S \to T$ with a linear operator $A\colon \mathcal{T}_+ \to \mathcal{S}_+$, given by

$$\begin{aligned} Af(s) &= \mu_s f, \\ \mu_s B &= A1_B(s), \qquad s \in S, \end{aligned} \tag{1}$$

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_4

for any $f \in \mathcal{T}_+$ and $B \in \mathcal{T}$. When $\mu = \delta$, the associated operator is simply the identity map I on $\mathcal{S}_+$. In general, we often denote a kernel and its associated operator by the same letter.

When T is a localized Borel space, we usually require kernels $\mu\colon S \to T$ to be *locally finite*, in the sense that μ_s is locally finite for every s. The following result characterizes the locally finite kernels from S to T.

Lemma 3.1 *(kernel criteria) For any Borel spaces S, T, consider a function $\mu : S \to \mathcal{M}_T$ with associated operator A as in* (1)*. Then for any dissecting semi-ring $\mathcal{I} \subset \hat{\mathcal{T}}$, these conditions are equivalent:*

(i) *μ is a kernel from S to T,*

(ii) *μ is a measurable function from S to $\mathcal{M}_T$,*

(iii) *$\mu_s I$ is $\mathcal{S}$-measurable for every $I \in \mathcal{I}$,*

(iv) *the operator A maps $\mathcal{T}_+$ into $\mathcal{S}_+$.*

Furthermore,

(v) *for any function $f\colon S \to T$, the mapping $s \mapsto \delta_{f(s)}$ is a kernel from S to T iff f is measurable,*

(vi) *the identity map on $\mathcal{M}_S$ is a kernel from $\mathcal{M}_S$ to S.*

Proof, (i) $\Leftrightarrow$ (ii): Since $\mathcal{M}_T$ is generated by all projections $\pi_B : \mu \mapsto \mu B$ with $B \in \mathcal{T}$, we see that (ii) $\Rightarrow$ (i). The converse holds by Lemma 1.4.

(i) $\Leftrightarrow$ (iii): Since trivially (i) $\Rightarrow$ (iii), it remains to show that (iii) $\Rightarrow$ (i). Then note that $\mathcal{I}$ is π-system, whereas the sets $B \in \hat{\mathcal{T}}$ with measurable $\mu_s B$ form a λ-system containing $\mathcal{I}$. The assertion then follows by a routine application of the monotone-class Theorem 1.1.

(i) $\Leftrightarrow$ (iv): For $f = 1_B$, we note that $Af(s) = \mu_s B$. Thus, (i) is simply the property (iv) restricted to indicator functions. The extension to general $f \in \mathcal{T}_+$ follows immediately by linearity and monotone convergence.

(v) This holds since $\big\{s;\, \delta_{f(s)} B = 1\big\} = f^{-1}B$ for all $B \in \mathcal{T}$.

(vi) The identity map on $\mathcal{M}_S$ trivially satisfies (ii). □

A kernel $\mu\colon S \to T$ is said to be *finite* if $\mu_s T < \infty$ for all $s \in S$, *s-finite* if it is a countable sum of finite kernels, and *σ-finite* if it satisfies $\mu_s f_s < \infty$ for some measurable function $f > 0$ on $S \times T$, where $f_s = f(s, \cdot)$. It is called a *probability kernel* if $\mu_s T \equiv 1$ and a *sub-probability kernel* when $\mu_s T \le 1$. We list some basic properties of s- and σ-finite kernels.

Lemma 3.2 *(kernel properties) Let the kernel $\mu\colon S \to T$ be s-finite. Then*

(i) *for any measurable function $f \ge 0$ on $S \otimes T$, the function $\mu_s f_s$ is $\mathcal{S}$-measurable, and $\nu_s = f_s \cdot \mu_s$ is again an s-finite kernel from S to T,*

(ii) *for any measurable function $f\colon S \times T \to U$, the function $\nu_s = \mu_s \circ f_s^{-1}$ is an s-finite kernel from S to U,*

(iii) $\mu = \sum_n \mu_n$ *for some sub-probability kernels* $\mu_1, \mu_2, \ldots : S \to T$,

(iv) μ *is σ-finite iff* $\mu = \sum_n \mu_n$ *for some finite, mutually singular kernels* $\mu_1, \mu_2, \ldots : S \to T$,

(v) μ *is σ-finite iff there exists a measurable function* $f > 0$ *on* $S \times T$ *with* $\mu_s f_s = 1\{\mu_s \neq 0\}$ *for all* $s \in S$,

(vi) *if* μ *is uniformly σ-finite, it is also σ-finite, and* (v) *holds with* $f(s,t) = h(\mu_s, t)$ *for some measurable function* $h > 0$ *on* $\mathcal{M}_S \times T$,

(vii) *if* μ *is a probability kernel and* T *is Borel, there exists a measurable function* $f: S \times [0,1] \to T$ *with* $\mu_s = \lambda \circ f_s^{-1}$ *for all* $s \in S$.

A function f as in (v) is called a *normalizing function* of μ.

Proof: (i) Since μ is s-finite and all measurability and measure properties are preserved by countable summations, we may assume that μ is finite. The measurability of $\mu_s f_s$ follows from the kernel property when $f = 1_B$ for some measurable rectangle $B \subset S \times T$. It extends by Theorem 1.1 to arbitrary $B \in \mathcal{S} \times \mathcal{T}$, and then by linearity and monotone convergence to any $f \in (\mathcal{S} \otimes \mathcal{T})_+$. Fixing f and applying the stated measurability to the functions $1_B f$ for arbitrary $B \in \mathcal{T}$, we see that $f_s \cdot \mu_s$ is again measurable, hence a kernel from S to T. The s-finiteness is clear if we write $f = \sum_n f_n$ for some bounded $f_1, f_2, \ldots \in (\mathcal{S} \otimes \mathcal{T})_+$.

(ii) Since μ is s-finite and the measurability, measure property, and inverse mapping property are all preserved by countable summations, we may again take μ to be finite. The measurability of f yields $f^{-1}B \in \mathcal{S} \otimes \mathcal{T}$ for every $B \in \mathcal{U}$. Since $f_s^{-1}B = \{t \in T;\ (s,t) \in f^{-1}B\} = (f^{-1}B)_s$, we see from (i) that $(\mu_s \circ f_s^{-1})B = \mu_s(f^{-1}B)_s$ is measurable, which means that $\nu_s = \mu_s \circ f_s^{-1}$ is a kernel. Since $\|\nu_s\| = \|\mu_s\| < \infty$ for all $s \in S$, the kernel ν is again finite.

(iii) Since μ is s-finite and N^2 is countable, we may assume that μ is finite. Putting $k_s = [\|\mu_s\|] + 1$, we note that $\mu_s = \sum_n k_s^{-1} 1\{n \le k_s\}\, \mu_s$, where each term is clearly a sub-probability kernel.

(iv) First let μ be σ-finite, so that $\mu_s f_s < \infty$ for some measurable function $f > 0$ on $S \times T$. Then (i) shows that $\nu_s = f_s \cdot \mu_s$ is a finite kernel from S to T, and so $\mu_s = g_s \cdot \nu_s$ with $g = 1/f$. Putting $B_n = 1\{n-1 < g \le n\}$ for $n \in \mathsf{N}$, we get $\mu_s = \sum_n (1_{B_n} g)_s \cdot \nu_s$, where each term is a finite kernel from S to T. The terms are mutually singular since $B_1, B_2, \ldots$ are disjoint.

Conversely, let $\mu = \sum_n \mu_n$ for some mutually singular, finite kernels $\mu_1, \mu_2, \ldots$. By singularity we may choose a measurable partition $B_1, B_2, \ldots$ of $S \times T$, such that μ_n is supported by B_n for each n. Since the μ_n are finite, we may further choose some measurable functions $f_n > 0$ on $S \times T$, such that $\mu_n f_n \le 2^{-n}$ for all n. Then $f = \sum_n 1_{B_n} f_n > 0$, and

$$\begin{aligned} \mu f &= \textstyle\sum_{m,n} \mu_m 1_{B_n} f_n \\ &= \textstyle\sum_n \mu_n f_n \end{aligned}$$

$$\leq \sum_n 2^{-n} = 1,$$

which shows that μ is σ-finite.

(v) If μ is σ-finite, then $\mu_s g_s < \infty$ for some measurable function $g > 0$ on $S \times T$. Putting $f(s,t) = g(s,t)/\mu_s g_s$ with $0/0 = 1$, we note that $\mu_s f_s = 1\{\mu_s \neq 0\}$. The converse assertion is clear from the definitions.

(vi) The first assertion holds by (iv). Now choose a measurable partition $B_1, B_2, \ldots$ of T with $\mu_s B_n < \infty$ for all s and n, and define

$$g(s,t) = \sum_n 2^{-n} \Big(\mu_s B_n \vee 1\Big)^{-1} 1_{B_n}(t), \quad s \in S, \ t \in T.$$

The function $f(s,t) = g(s,t)/\mu_s g_s$ has the desired properties.

(vii) Since T is Borel, we may assume that $T \in \mathcal{B}_{[0,1]}$. The function

$$f(s,r) = \inf\Big\{x \in [0,1];\ \mu_s[0,x] \geq r\Big\}, \quad s \in S, \ r \in [0,1],$$

is product measurable on $S \times [0,1]$, since the set $\{(s,r);\ \mu_s[0,x] \geq r\}$ is measurable for each x by Lemma 1.13, and the infimum can be restricted to rational x. For $s \in S$ and $x \in [0,1]$, we have

$$\begin{aligned} \lambda \circ f_s^{-1}[0,x] &= \lambda\Big\{r \in [0,1];\ f(s,r) \leq x\Big\} \\ &= \mu_s[0,x], \end{aligned}$$

and so $\lambda \circ f_s^{-1} = \mu_s$ on $[0,1]$ for all s by Lemma 4.3 below. In particular, $f_s(r) \in T$ a.e. λ for every $s \in S$. On the exceptional set $\{f_s(r) \notin T\}$, we may redefine $f_s(r) = t_0$ for any fixed $t_0 \in T$. □

For any s-finite kernels $\mu : S \to T$ and $\nu : S \times T \to U$, we define their *composition* and *product* as the kernels $\mu \otimes \nu : S \to T \times U$ and $\mu\nu : S \to U$, given by

$$\begin{aligned} (\mu \otimes \nu)_s f &= \int \mu_s(dt) \int \nu_{s,t}(du)\, f(t,u), \qquad (2) \\ (\mu\nu)_s f &= \int \mu_s(dt) \int \nu_{s,t}(du)\, f(u) \\ &= (\mu \otimes \nu)_s (1_T \otimes f). \end{aligned}$$

Note that $\mu\nu$ equals the projection of $\mu \otimes \nu$ onto U. Define $A_\mu f(s) = \mu_s f$, and write $\hat{\mu} = \delta \otimes \mu$.

Lemma 3.3 *(composition and product) For any s-finite kernels $\mu : S \to T$ and $\nu : S \times T \to U$,*

(i) *$\mu \otimes \nu$ and $\mu\nu$ are s-finite kernels from S to $T \times U$ and U, respectively,*

(ii) *$\mu \otimes \nu$ is σ-finite whenever this holds for μ, ν,*

(iii) *the kernel operations $\mu \otimes \nu$ and $\mu\nu$ are associative,*

(iv) *$\mu\hat{\nu} = \mu \otimes \nu$ and $\hat{\mu}\hat{\nu} = (\mu \otimes \nu)\hat{}$,*

(v) *for any kernels $\mu : S \to T$ and $\nu : T \to U$, we have $A_{\mu\nu} = A_\mu A_\nu$.*

Proof: (i) By Lemma 3.2 (i) applied twice, the inner integral in (2) is $\mathcal{S}\otimes\mathcal{T}$-measurable and the outer integral is $\mathcal{S}$-measurable. The countable additivity holds by repeated monotone convergence. This proves the kernel property of $\mu\otimes\nu$, and the result for $\mu\nu$ is an immediate consequence.

(ii) If μ and ν are σ-finite, we may choose some measurable functions $f>0$ on $S\times T$ and $g>0$ on $S\times T\times U$ with $\mu_s f_s\le 1$ and $\nu_{s,t}g_{s,t}\le 1$ for all s and t. Then (2) yields $(\mu\otimes\nu)_s(fg)_s\le 1$.

(iii) Consider the kernels $\mu\colon S\to T$, $\nu\colon S\times T\to U$, and $\rho\colon S\times T\times U\to V$. Letting $f\in(\mathcal{T}\otimes\mathcal{U}\otimes\mathcal{V})_+$ be arbitrary and using (2) repeatedly, we get

$$\begin{aligned}
\big\{\mu\otimes(\nu\otimes\rho)\big\}_s f &= \int\mu_s(dt)\iint(\nu\otimes\rho)_{s,t}(du\,dv)\,f(t,u,v)\\
&= \int\mu_s(dt)\int\nu_{s,t}(du)\int\rho_{s,t,u}(dv)\,f(t,u,v)\\
&= \iint(\mu\otimes\nu)_s(dt\,du)\int\rho_{s,t,u}(dv)\,f(t,u,v)\\
&= \big\{(\mu\otimes\nu)\otimes\rho\big\}_s f,
\end{aligned}$$

which shows that $\mu\otimes(\nu\otimes\rho)=(\mu\otimes\nu)\otimes\rho$. Projecting onto V yields $\mu(\nu\rho)=(\mu\nu)\rho$.

(iv) Consider any kernels $\mu\colon S\to T$ and $\nu\colon S\times T\to U$, and note that $\mu\hat\nu$ and $\mu\otimes\nu$ are both kernels from S to $T\times U$. For any $f\in(\mathcal{T}\otimes\mathcal{U})_+$, we have

$$\begin{aligned}
(\mu\hat\nu)_s f &= \int\mu_s(dt)\int\delta_{s,t}(ds'dt')\int\nu_{s',t'}(du)\,f(t',u)\\
&= \int\mu_s(dt)\int\nu_{s,t}(du)\,f(t,u)\\
&= (\mu\otimes\nu)_s f,
\end{aligned}$$

which shows that $\mu\hat\nu=\mu\otimes\nu$. Hence, by (ii)

$$\begin{aligned}
\hat\mu\hat\nu &= (\delta\otimes\mu)\otimes\nu\\
&= \delta\otimes(\mu\otimes\nu)\\
&= (\mu\otimes\nu)\hat{}.
\end{aligned}$$

(v) For any $f\in\mathcal{U}_+$, we have

$$\begin{aligned}
A_{\mu\nu}f(s) = (\mu\nu)_s f &= \int(\mu\nu)_s(du)\,f(u)\\
&= \int\mu_s(dt)\int\nu_t(du)\,f(u)\\
&= \int\mu_s(dt)\,A_\nu f(t)\\
&= A_\mu(A_\nu f)(s)\\
&= (A_\mu A_\nu)f(s),
\end{aligned}$$

which implies $A_{\mu\nu}f=(A_\mu A_\nu)f$, and hence $A_{\mu\nu}=A_\mu A_\nu$. □

We turn to the reverse problem of representing a given measure ρ on $S\times T$ in the form $\nu\otimes\mu$, for some measure ν on S and kernel $\mu: S\to T$. The resulting *disintegration* may be regarded as an infinitesimal decomposition of ρ into measures $\delta_s\otimes\mu_s$. Any σ-finite measure $\nu\sim\rho(\cdot\times T)$ is called a *supporting measure* of ρ, and we refer to μ as the associated *disintegration kernel.*

Theorem 3.4 *(disintegration) Let ρ be a σ-finite measure on $S\times T$, where T is Borel. Then*

(i) *$\rho=\nu\otimes\mu$ for a σ-finite measure $\nu\sim\rho(\cdot\times T)\equiv\hat\rho_S$ and a σ-finite kernel $\mu\colon S\to T$,*

(ii) *the μ_s are ν-a.e. unique up to normalizations, and they are a.e. bounded iff $\hat\rho_S$ is σ-finite,*

(iii) *when $\hat\rho$ is σ-finite and $\nu=\hat\rho_S$, we may choose the μ_s to be probability measures on T.*

Proof: (i) If ρ is σ-finite and $\neq 0$, we may choose $g\in(S\times T)_+$ such that $\rho=g\cdot\tilde\rho$ for some probability measure $\tilde\rho$ on $S\times T$. If $\tilde\rho=\nu\otimes\tilde\mu$ for some measure ν on S and kernel $\tilde\mu\colon S\times T$, then Lemma 3.2 (i) shows that $\mu=g\cdot\tilde\mu$ is a σ-finite kernel from S to T, and so for any $f\geq 0$,

$$\begin{aligned}(\nu\otimes\mu)f &= \Big\{\mu\otimes(g\cdot\tilde\mu)\Big\}f\\ &= (\mu\otimes\tilde\mu)(fg)\\ &= \tilde\rho(fg)=\rho f,\end{aligned}$$

which shows that $\rho=\nu\otimes\mu$. We may then assume that ρ is a probability measure on $S\times T$. Since T is Borel, we may further take $T=\mathsf{R}$.

Choosing $\nu=\hat\rho$ and putting $\nu_r=\rho(\cdot\times(-\infty,r])$, we get $\nu_r\leq\nu$ for all $r\in\mathsf{R}$, and so Theorem 2.10 yields $\nu_r=f_r\cdot\nu$ for some measurable functions $f_r\colon S\to[0,1]$. Here f_r is non-decreasing in $r\in\mathsf{Q}$ with limits 0 and 1 at $\pm\infty$, outside a fixed ν-null set $N\subset S$. Redefining $f_r=0$ on N, we can make those properties hold identically. For the right-continuous versions $F_{s,t}=\inf_{r>t}f_r$ with $t\in\mathsf{R}$, we see by monotone convergence that $F_r=f_r$ a.e. for all $r\in\mathsf{Q}$.

Now Proposition 2.14 yields some probability measures $\mu(s,\cdot)$ on R with distribution functions $F(s,t)$. Here $\mu(s,B)$ is measurable in $s\in S$ for each $B\in\mathcal{T}$ by a monotone-class argument, which means that μ is a kernel from S to T. Furthermore, we have for any $A\in\mathcal{S}$ and $r\in\mathsf{Q}$

$$\begin{aligned}\rho\Big(A\times(-\infty,r]\Big) &= \int_A\nu(ds)\,F_r(s)\\ &= \int_A\nu(ds)\,\mu_s(-\infty,r],\end{aligned}$$

which extends by a monotone-class argument and monotone convergence to the general disintegration formula $\rho f=(\nu\otimes\mu)f$, for arbitrary $f\in(\mathcal{S}\times\mathcal{T})_+$.

(ii) Suppose that $\rho=\nu\otimes\mu=\nu'\otimes\mu'$ with $\nu\sim\nu'\sim\rho(\cdot\times T)$. Then

$$\begin{aligned} g \cdot \rho &= \nu \otimes (g \cdot \mu) \\ &= \nu' \otimes (g \cdot \mu'), \quad g \geq 0, \end{aligned}$$

which allows us to consider only bounded measures ρ. Since also $(h \cdot \nu) \otimes \mu = \nu' \otimes (h \cdot \mu')$ for any $h \in \mathcal{S}_+$, we may further assume that $\nu = \nu'$. Then

$$\int_A \nu(ds) \left(\mu_s B - \mu'_s B\right) = 0, \quad A \in \mathcal{S},\ B \in \mathcal{T},$$

which implies $\mu_s B = \mu'_s B$ a.e. ν. Since T is Borel, a monotone-class argument yields $\mu_s = \mu'_s$ a.e. ν. □

The following partial disintegration will be needed in Chapter 31.

Corollary 3.5 *(partial disintegration) Let ν, ρ be σ-finite measures on S and $S \times T$, respectively, where T is Borel. Then there exists an a.e. unique maximal kernel*

$$\mu \colon S \to T, \qquad \nu \otimes \mu \leq \rho.$$

Here the *maximality* of μ means that, whenever μ' is a kernel satisfying $\nu \otimes \mu' \leq \rho$, we have $\mu'_s \leq \mu_s$ for $s \in S$ a.e. ν.

Proof: Since ρ is σ-finite, we may choose a σ-finite measure γ on S with $\gamma \sim \rho(\cdot \times T)$. Consider its Lebesgue decomposition $\gamma = \gamma_a + \gamma_s$ with respect to ν, as in Theorem 2.10, so that $\gamma_a \ll \nu$ and $\gamma_s \perp \nu$. Choose $A \in \mathcal{S}$ with $\nu A^c = \gamma_s A = 0$, and let ρ' denote the restriction of ρ to $A \times T$. Then $\rho'(\cdot \times T) \ll \nu$, and so Theorem 3.4 yields a σ-finite kernel $\mu : S \to T$ with $\nu \otimes \mu = \rho' \leq \rho$.

Now consider any kernel $\mu' : S \to T$ with $\nu \otimes \mu' \leq \rho$. Since $\nu A^c = 0$, we have $\nu \otimes \mu' \leq \rho' = \nu \otimes \mu$. Since μ and μ' are σ-finite, there exists a measurable function $g > 0$ on $S \times T$, such that the kernels $\tilde{\mu} = g \cdot \mu$ and $\tilde{\mu}' = g \cdot \mu'$ satisfy $\|\tilde{\mu}_s\| \vee \|\tilde{\mu}'_s\| \leq 1$. Then $\nu \otimes \tilde{\mu}' \leq \nu \otimes \tilde{\mu}$, and so $\tilde{\mu}'_s B \leq \tilde{\mu}_s B$, $s \in S$ a.e. ν, for every $B \in \mathcal{S}$. This extends by a monotone-class argument to $\mu'_s \leq \mu_s$ a.e. ν, which shows that μ is maximal. In particular, μ is a.e. unique. □

We often need to disintegrate kernels. Here the previous construction still applies, except that now we need a product-measurable version of the Radon–Nikodym theorem, and we must check that the required measurability is preserved throughout the construction. Since the simplest approach is based on martingale methods, we postpone the proof until Theorem 9.27.

Corollary 3.6 *(disintegration of kernels) Consider some σ-finite kernels $\nu \colon S \to T$ and $\rho \colon S \to T \times U$ with $\nu_s \sim \rho_s(\cdot \times U)$ for all $s \in S$, where T, U are Borel. Then there exists a σ-finite kernel*

$$\mu \colon S \times T \to U, \qquad \rho = \nu \otimes \mu.$$

We turn to the subject of iterated disintegration, needed for some conditional constructions in Chapters 8 and 31. To explain the notation, we consider some σ-finite measures μ_1, μ_2, μ_3, μ_{12}, μ_{13}, μ_{23}, μ_{123} on products of the Borel spaces S, T, U. They are said to form a *projective* family, if all relations of the form $\mu_1 \sim \mu_{12}(\cdot \times T)$ or $\mu_{12} \sim \mu_{123}(\cdot \times U)$ are satisfied[1]. Then Theorem 3.4 yields some disintegrations

$$\begin{aligned}
\mu_{12} &= \mu_1 \otimes \mu_{2|1} \cong \mu_2 \otimes \mu_{1|2},\\
\mu_{13} &= \mu_1 \otimes \mu_{3|1},\\
\mu_{123} &= \mu_{12} \otimes \mu_{3|12} = \mu_1 \otimes \mu_{23|1}\\
&\cong \mu_2 \otimes \mu_{13|2},
\end{aligned}$$

for some σ-finite kernels $\mu_{2|1}$, $\mu_{3|12}$, $\mu_{12|3}$, ... between appropriate spaces, where $\cong$ denotes equality apart from the order of component spaces. If $\mu_{2|1} \sim \mu_{23|1}(\cdot \times U)$ a.e. μ_1, we may proceed to form some iterated disintegration kernels such as $\mu_{3|2|1}$, where $\mu_{23|1} = \mu_{2|1} \otimes \mu_{3|2|1}$ a.e. μ_1. We show that the required support properties hold automatically, and that the iterated disintegration kernels $\mu_{3|2|1}$ and $\mu_{3|1|2}$ are a.e. equal and agree with the single disintegration kernel $\mu_{3|12}$.

Theorem 3.7 *(iterated disintegration) Consider a projective set of σ-finite measures μ_1, μ_2, μ_3, μ_{12}, μ_{13}, μ_{23}, μ_{123} on products of Borel spaces S, T, U, and form the associated disintegration kernels $\mu_{1|2}$, $\mu_{2|1}$, $\mu_{3|1}$, $\mu_{3|12}$. Then*

(i) $\mu_{2|1} \sim \mu_{23|1}(\cdot \times U)$ *a.e.* μ_1,
$\mu_{1|2} \sim \mu_{13|2}(\cdot \times U)$ *a.e.* μ_2,

(ii) $\mu_{3|12} = \mu_{3|2|1} \cong \mu_{3|1|2}$ *a.e.* μ_{12},

(iii) *if* $\mu_{13} = \mu_{123}(\cdot \times T \times \cdot)$, *then also*
$\mu_{3|1} = \mu_{2|1}\, \mu_{3|1|2}$ *a.e.* μ_1.

Proof: (i) We may assume that $\mu_{123} \neq 0$. Fixing a probability measure $\tilde\mu_{123} \sim \mu_{123}$, we form the projections $\tilde\mu_1$, $\tilde\mu_2$, $\tilde\mu_3$, $\tilde\mu_{12}$, $\tilde\mu_{13}$, $\tilde\mu_{23}$ of $\tilde\mu_{123}$ onto S, T, U, $S \times T$, $S \times U$, $T \times U$, respectively. Then for any $A \in \mathcal{S} \otimes \mathcal{T}$,

$$\begin{aligned}
\tilde\mu_{12} A &= \tilde\mu_{123}(A \times U)\\
&\sim \mu_{123}(A \times U) \sim \mu_{12} A,
\end{aligned}$$

which shows that $\tilde\mu_{12} \sim \mu_{12}$. A similar argument gives $\tilde\mu_1 \sim \mu_1$ and $\tilde\mu_2 \sim \mu_2$. Hence, Theorem 2.10 yields some measurable functions p_1, p_2, p_{12}, $p_{123} > 0$ on S, T, $S \times T$, $S \times T \times U$, satisfying

$$\begin{aligned}
\tilde\mu_1 &= p_1 \cdot \mu_1, & \tilde\mu_2 &= p_2 \cdot \mu_2,\\
\tilde\mu_{12} &= p_{12} \cdot \mu_{12}, & \tilde\mu_{123} &= p_{123} \cdot \mu_{123}.
\end{aligned}$$

Inserting those densities into the disintegrations

[1]For functions or constants $a, b \geq 0$, the relation $a \sim b$ means that $a = 0$ iff $b = 0$.

$$\begin{aligned}\tilde{\mu}_{12} &= \tilde{\mu}_1 \otimes \tilde{\mu}_{2|1} \overset{\sim}{=} \tilde{\mu}_2 \otimes \tilde{\mu}_{1|2},\\ \tilde{\mu}_{123} &= \tilde{\mu}_1 \otimes \tilde{\mu}_{23|1} \overset{\sim}{=} \tilde{\mu}_2 \otimes \tilde{\mu}_{13|2},\end{aligned} \tag{3}$$

we obtain

$$\begin{aligned}\mu_{12} &= \mu_1 \otimes \big(p_{2|1} \cdot \tilde{\mu}_{2|1}\big)\\ &\overset{\sim}{=} \mu_2 \otimes \big(p_{1|2} \cdot \tilde{\mu}_{1|2}\big),\\ \mu_{123} &= \mu_1 \otimes \big(p_{23|1} \cdot \tilde{\mu}_{23|1}\big)\\ &\overset{\sim}{=} \mu_2 \otimes \big(p_{13|2} \cdot \tilde{\mu}_{13|2}\big),\end{aligned}$$

where

$$p_{2|1} = \frac{p_1}{p_{12}}, \qquad p_{1|2} = \frac{p_2}{p_{12}}, \qquad p_{23|1} = \frac{p_1}{p_{123}}, \qquad p_{13|2} = \frac{p_2}{p_{123}}.$$

Comparing with the disintegrations of μ_{12} and μ_{123}, and invoking the uniqueness in Theorem 3.4, we obtain a.e.

$$\begin{aligned}\tilde{\mu}_{2|1} &\sim \mu_{2|1}, \qquad \tilde{\mu}_{1|2} \sim \mu_{1|2},\\ \tilde{\mu}_{23|1} &\sim \mu_{23|1}, \qquad \tilde{\mu}_{13|2} \sim \mu_{13|2}.\end{aligned} \tag{4}$$

Furthermore, we get from (3)

$$\begin{aligned}\tilde{\mu}_1 \otimes \tilde{\mu}_{2|1} &= \tilde{\mu}_{12} = \tilde{\mu}_{123}(\cdot \times U)\\ &= \tilde{\mu}_1 \otimes \tilde{\mu}_{23|1}(\cdot \times U),\end{aligned}$$

and similarly with subscripts 1 and 2 interchanged, and so the mentioned uniqueness yields a.e.

$$\begin{aligned}\tilde{\mu}_{2|1} &= \tilde{\mu}_{23|1}(\cdot \times U),\\ \tilde{\mu}_{1|2} &= \tilde{\mu}_{13|2}(\cdot \times U).\end{aligned} \tag{5}$$

Combining (4) and (5), we get a.e.

$$\begin{aligned}\mu_{2|1} &\sim \tilde{\mu}_{2|1} = \tilde{\mu}_{23|1}(\cdot \times U)\\ &\sim \mu_{23|1}(\cdot \times U),\end{aligned}$$

and similarly with 1 and 2 interchanged.

(ii) By (i) and Theorem 3.6, we have a.e.

$$\begin{aligned}\mu_{23|1} &= \mu_{2|1} \otimes \mu_{3|2|1},\\ \mu_{13|2} &= \mu_{1|2} \otimes \mu_{3|1|2},\end{aligned}$$

for some product-measurable kernels $\mu_{3|2|1}$ and $\mu_{3|1|2}$. Combining the various disintegrations and using the commutativity in Lemma 3.3 (ii), we get

$$\begin{aligned}\mu_{12} \otimes \mu_{3|12} &= \mu_{123} = \mu_1 \otimes \mu_{23|1}\\ &= \mu_1 \otimes \mu_{2|1} \otimes \mu_{3|2|1}\\ &= \mu_{12} \otimes \mu_{3|2|1},\end{aligned}$$

and similarly with 1 and 2 interchanged. It remains to use the a.e. uniqueness in Theorem 3.4.

(iii) The stated hypothesis yields

$$\begin{aligned}\mu_1 \otimes \mu_{3|1} &= \mu_{13} = \mu_{123}(\cdot \times T \times \cdot) \\ &= \mu_1 \otimes \mu_{23|1}(T \times \cdot),\end{aligned}$$

and so by (ii) and the uniqueness in Theorem 3.4, we have a.e.

$$\begin{aligned}\mu_{3|1} &= \mu_{23|1}(T \times \cdot) \\ &= \mu_{2|1} \otimes \mu_{3|2|1}(T \times \cdot) \\ &= \mu_{2|1}\, \mu_{3|2|1} \\ &= \mu_{2|1}\, \mu_{3|1|2}.\end{aligned}$$

□

A group G with associated σ-field $\mathcal{G}$ is said to be *measurable*, if the group operations $r \mapsto r^{-1}$ and $(r,s) \mapsto rs$ on G are $\mathcal{G}$-measurable. Defining the left and right *shifts* θ_r and $\tilde\theta_r$ on G by $\theta_r s = \tilde\theta_s r = rs$, $r, s \in G$, we say that a measure λ on G is *left-invariant* if $\lambda \circ \theta_r^{-1} = \lambda$ for all $r \in G$. A σ-finite, left-invariant measure $\lambda \neq 0$ is called a *Haar measure* on G. Writing $\tilde f(r) = f(r^{-1})$, we may define a corresponding right-invariant measure $\tilde\lambda$ by $\tilde\lambda f = \lambda \tilde f$.

We call G a *topological group* if it is endowed with a topology rendering the group operations continuous. When G is *lcscH*[2], it becomes a measurable group when endowed with the Borel σ-field $\mathcal{G}$. We state the basic existence and uniqueness theorem for Haar measures.

Theorem 3.8 *(Haar measure) For any lcscH group G,*

(i) *there exists a left-invariant, locally finite measure $\lambda \neq 0$ on G,*

(ii) *λ is unique up to a normalization,*

(iii) *when G is compact, λ is also right-invariant.*

Proof (Weil): For any $f, g \in \hat C_+$ we define $|f|_g = \inf \sum_k c_k$, where the infimum extends over all finite sets of constants $c_1, \dots, c_n \geq 0$, such that

$$f(x) \leq \sum_{k \leq n} c_k\, g(s_k x), \quad x \in G,$$

for some $s_1, \dots, s_n \in G$. By compactness we have $|f|_g < \infty$ when $g \neq 0$, and since $|f|_g$ is non-decreasing and translation invariant in f, it satisfies the sub-additivity and homogeneity properties

$$|f + f'|_g \leq |f|_g + |f'|_g, \qquad |cf|_g = c|f|_g, \tag{6}$$

as well as the inequalities

$$\frac{\|f\|}{\|g\|} \leq |f|_g \leq |f|_h |h|_g. \tag{7}$$

We may normalize $|f|_g$ by fixing an $f_0 \in \hat C_+ \setminus \{0\}$ and putting

$$\lambda_g f = |f|_g / |f_0|_g, \quad f, g \in \hat C_+, \; g \neq 0.$$

[2] locally compact, second countable, Hausdorff

From (6) and (7) we note that

$$\begin{aligned}\lambda_g(f+f') &\leq \lambda_g f + \lambda_g f',\\ \lambda_g(cf) &= c\lambda_g f,\\ |f_0|_f^{-1} &\leq \lambda_g f \leq |f|_{f_0}.\end{aligned} \tag{8}$$

Conversely, λ_g is nearly super-additive in the following sense:

Lemma 3.9 *(near super-additivity) For any $f, f' \in \hat{C}_+$ and $\varepsilon > 0$, there exists an open set $U \neq \emptyset$ with*

$$\lambda_g f + \lambda_g f' \leq \lambda_g(f+f') + \varepsilon, \quad 0 \neq g \prec U.$$

Proof: Fix any $h \in \hat{C}_+$ with $h = 1$ on $\operatorname{supp}(f+f')$, and define for $\delta > 0$

$$\begin{aligned}f_\delta &= f + f' + \delta h,\\ h_\delta &= f/f_\delta, \quad h'_\delta = f'/f_\delta,\end{aligned}$$

so that $h_\delta, h'_\delta \in \hat{C}_+$. By compactness, we may choose a neighborhood U of the identity element $\iota \in G$ satisfying

$$\begin{aligned}\big|h_\delta(x) - h_\delta(y)\big| &< \delta,\\ \big|h'_\delta(x) - h'_\delta(y)\big| &< \delta, \qquad x^{-1}y \in U.\end{aligned} \tag{9}$$

Now assume $0 \neq g \prec U$, and let $f_\delta(x) \leq \sum_k c_k\, g(s_k x)$ for some $s_1, \dots, s_n \in G$ and $c_1, \dots, c_n \geq 0$. Since $g(s_k x) \neq 0$ implies $s_k x \in U$, (9) yields

$$\begin{aligned}f(x) &= f_\delta(x)\, h_\delta(x)\\ &\leq \textstyle\sum_k c_k\, g(s_k x)\, h_\delta(x)\\ &\leq \textstyle\sum_k c_k\, g(s_k x)\, \big\{h_\delta(s_k^{-1}) + \delta\big\},\end{aligned}$$

and similarly for f'. Noting that $h_\delta + h'_\delta \leq 1$, we get

$$|f|_g + |f'|_g \leq \textstyle\sum_k c_k(1+2\delta).$$

Taking the infimum over all dominating sums for f_δ and using (6), we obtain

$$\begin{aligned}|f|_g + |f'|_g &\leq |f_\delta|_g(1+2\delta)\\ &\leq \big\{|f+f'|_g + \delta|h|_g\big\}(1+2\delta).\end{aligned}$$

Now divide by $|f_0|_g$, and use (8) to obtain

$$\begin{aligned}\lambda_g f + \lambda_g f' &\leq \big\{\lambda_g(f+f') + \delta\lambda_g h\big\}(1+2\delta)\\ &\leq \lambda_g(f+f') + 2\delta|f+f'|_{f_0} + \delta(1+2\delta)|h|_{f_0},\end{aligned}$$

which tends to $\lambda_g(f+f')$ as $\delta \to 0$. □

End of proof of Theorem 3.8: We may regard the functionals λ_g as elements of the product space $\Lambda = (\mathsf{R}_+)^{\hat{C}_+}$. For any neighborhood U of ι, let Λ_U be the closure in Λ of the set $\{\lambda_g;\, 0 \neq g \prec U\}$. Since $\lambda_g f \leq |f|_{f_0} < \infty$ for all $f \in \hat{C}_+$ by (8), the Λ_U are compact by Tychonov's theorem[3]. Furthermore, the finite intersection property holds for the family $\{\Lambda_U;\, \iota \in U\}$, since $U \subset V$ implies $\Lambda_U \subset \Lambda_V$. We may then choose an element $\lambda \in \bigcap_U \Lambda_U$, here regarded as a functional on $\hat{C}_+$. From (8) we note that $\lambda \neq 0$.

To see that λ is linear, fix any $f, f' \in \hat{C}_+$ and $a, b \geq 0$, and choose some $g_1, g_2, \ldots \in \hat{C}_+$ with $\operatorname{supp} g_n \downarrow \{\iota\}$, such that

$$\lambda_{g_n} f \to \lambda f, \quad \lambda_{g_n} f' \to \lambda f',$$
$$\lambda_{g_n}(\, af + bf'\,) \to \lambda(\, af + bf'\,).$$

By (8) and Lemma 3.9 we obtain $\lambda(af + bf') = a\,\lambda f + b\,\lambda f'$. Thus, λ is a non-trivial, positive linear functional on $\hat{C}_+$, and so by Theorem 2.25 it extends uniquely to a locally finite measure on S. The invariance of the functionals λ_g clearly carries over to λ.

Now consider any left-invariant, locally finite measure $\lambda \neq 0$ on G. Fixing a right-invariant, locally finite measure $\mu \neq 0$ and a function $h \in \hat{C}_+ \setminus \{0\}$, we define

$$p(x) = \int h(y^{-1}x)\, \mu(dy), \quad x \in G,$$

and note that $p > 0$ on G. Using the invariance of λ and μ along with Fubini's theorem, we get for any $f \in \hat{C}_+$

$$\begin{aligned}
(\lambda h)(\mu f) &= \int h(x)\, \lambda(dx) \int f(y)\, \mu(dy) \\
&= \int h(x)\, \lambda(dx) \int f(yx)\, \mu(dy) \\
&= \int \mu(dy) \int h(x) f(yx)\, \lambda(dx) \\
&= \int \mu(dy) \int h(y^{-1}x) f(x)\, \lambda(dx) \\
&= \int f(x)\, \lambda(dx) \int h(y^{-1}x)\, \mu(dy) = \lambda(fp).
\end{aligned}$$

Since f was arbitrary, we obtain $(\lambda h)\mu = p \cdot \lambda$ or equivalently $\lambda/\lambda h = p^{-1} \cdot \mu$. The asserted uniqueness now follows, since the right-hand side is independent of λ. If S is compact, we may choose $h \equiv 1$ to obtain $\lambda/\lambda S = \mu/\mu S$. □

From now on, we make no topological or other assumptions on G, beyond the existence of a Haar measure. The uniqueness and basic properties of such measures are covered by the following result, which also provides a basic factorization property. For technical convenience we consider *s-finite* (rather than σ-finite) measures, defined as countable sums of bounded measures. Note that most of the basic computational rules, including Fubini's theorem, remain valid

[3]Products of compact spaces are again compact.

for s-finite measures, and that s-finiteness has the advantage of being preserved by measurable maps.

Theorem 3.10 *(factorization and modularity) Consider a measurable space S and a measurable group G with Haar measure λ. Then*

(i) *for any s-finite, G-invariant measure μ on $G \times S$, there exists a unique, s-finite measure ν on S satisfing $\mu = \lambda \otimes \nu$,*

(ii) *μ, ν are simultaneously σ-finite,*

(iii) *there exists a measurable homomorphism $\Delta\colon G \to (0, \infty)$ with*

$$\tilde{\lambda} = \Delta \cdot \lambda, \qquad \lambda \circ \tilde{\theta}_r^{-1} = \Delta_r \lambda, \quad r \in G,$$

(iv) *when $\|\lambda\| < \infty$ we have $\Delta \equiv 1$, and λ is also right-invariant with $\tilde{\lambda} = \lambda$.*

In particular, (i) shows that any s-finite, left-invariant measure on G is proportional to λ. The dual mapping $\tilde{\Delta}$ is known as the *modular function* of G. The use of s-finite measures leads to some significant simplifications in subsequent proofs.

Proof: (i) Since λ is σ-finite, we may choose a measurable function $h > 0$ on G with $\lambda h = 1$, and define

$$\Delta_r = \lambda(h \circ \tilde{\theta}_r) \in (0, \infty], \quad r \in G.$$

Using Fubini's theorem (three times), the G-invariance of μ and λ, and the definitions of Δ and $\tilde{\lambda}$, we get for any $f \geq 0$

$$\begin{aligned}
\mu(f\Delta) &= \iint \mu(dr\, ds)\, f(r,s) \int \lambda(dp)\, h(pr) \\
&= \int \lambda(dp) \iint \mu(dr\, ds)\, f\big(p^{-1}r, s\big)\, h(r) \\
&= \iint \mu(dr\, ds)\, h(r) \int \lambda(dp)\, f(p^{-1}, s) \\
&= \int \tilde{\lambda}(dp) \iint \mu(dr\, ds)\, h(r)\, f(p, s).
\end{aligned}$$

Choosing $\mu = \lambda$ for a singleton S, we get $\tilde{\lambda} f = \lambda(f\Delta)$, and so $\tilde{\lambda} = \Delta \cdot \lambda$. Therefore $\lambda = \tilde{\Delta} \cdot \tilde{\lambda} = \tilde{\Delta}\Delta \cdot \lambda$, hence $\tilde{\Delta}\Delta = 1$ a.e. λ, and finally $\Delta \in (0, \infty)$ a.e. λ. Thus, for general S,

$$\mu(f\Delta) = \int \lambda(dp)\, \Delta(p) \iint \mu(dr\, ds)\, h(r)\, f(p, s).$$

Since $\Delta > 0$, we have $\mu(\cdot \times S) \ll \lambda$, and so $\Delta \in (0, \infty)$ a.e. $\mu(\cdot \times S)$, which yields the simplified formula

$$\mu f = \int \lambda(dp) \iint \mu(dr\, ds)\, h(r)\, f(p, s),$$

showing that $\mu = \lambda \otimes \nu$ with $\nu f = \mu(h \otimes f)$.

(ii) When μ is σ-finite, we have $\mu f < \infty$ for some measurable function $f > 0$ on $G \times S$, and so by Fubini's theorem $\nu f(r, \cdot) < \infty$ for $r \in G$ a.e. λ, which shows that even ν is σ-finite. The reverse implication is obvious.

(iii) For every $r \in G$, the measure $\lambda \circ \tilde{\theta}_r^{-1}$ is left-invariant since θ_p and $\tilde{\theta}_r$ commute for all $p, r \in G$. Hence, (i) yields $\lambda \circ \tilde{\theta}_r^{-1} = c_r \lambda$ for some constants $c_r > 0$. Applying both sides to h gives $c_r = \Delta(r)$. The homomorphism relation $\Delta(pq) = \Delta(p)\Delta(q)$ follows from the reverse semi-group property $\tilde{\theta}_{pq} = \tilde{\theta}_q \circ \tilde{\theta}_p$, and the measurability holds by the measurability of the group operation and Fubini's theorem.

(iv) Statement (iii) yields $\|\lambda\| = \Delta_r \|\lambda\|$ for all $r \in G$, and so $\Delta \equiv 1$ when $\|\lambda\| < \infty$. Then (iii) gives $\tilde{\lambda} = \lambda$ and $\lambda \circ \tilde{\theta}_r^{-1} = \lambda$ for all $r \in G$, which shows that λ is also right-invariant. □

Given a group G and a space S, we define an *action* of G on S as a mapping $(r, s) \mapsto rs$ from $G \times S$ to S, such that $p(rs) = (pr)s$ and $\iota s = s$ for all $p, r \in G$ and $s \in S$, where ι denotes the identity element of G. If G and S are measurable spaces, we say that G acts *measurably* on S if the action map is product-measurable. The *shifts* θ_r and *projections* π_s are defined by $rs = \theta_r s = \pi_s r$, and the *orbit* containing s is given by $\pi_s G = \{rs;\, r \in G\}$. The orbits form a partition of S, and we say that the action is *transitive* if $\pi_s G \equiv S$. For $s, s' \in S$, we write $s \sim s'$ if s and s' belong to the same orbit, so that $s = rs'$ for some $r \in G$. If G acts on both S and T, its *joint action* on $S \times T$ is given by $r(s, t) = (rs, rt)$.

For any group G acting on S, we say that a subset $B \subset S$ is *G-invariant* if $B \circ \theta_r^{-1} = B$ for all $r \in G$, and a function f on S is *G-invariant* if $f \circ \theta_r = f$. When S is measurable with σ-field $\mathcal{S}$, the class of G-invariant sets in $\mathcal{S}$ is again a σ-field, denoted by $\mathcal{I}_S$. For a transitive group action we have $\mathcal{I}_S = \{\emptyset, S\}$, and every invariant function is a constant. When the group action is measurable, a measure ν on S is said to be *G-invariant* if $\nu \circ \theta_r^{-1} = \nu$ for all $r \in G$.

When G is a measurable group with Haar measure λ, acting measurably on S, we say that G acts *properly*[4] on S, if there exists a *normalizing* function $g > 0$ on S, such that g is measurable and satisfies $\lambda(g \circ \pi_s) < \infty$ for all $s \in S$. We may then define a kernel φ on S by

$$\varphi_s = \frac{\lambda \circ \pi_s^{-1}}{\lambda(g \circ \pi_s)}, \quad s \in S. \tag{10}$$

Lemma 3.11 *(orbit measures) Let G be a measurable group with Haar measure λ, acting properly on S. Then the measures φ_s are σ-finite and satisfy*

(i) $\varphi_s = \varphi_{s'}$, $s \sim s'$,
$\varphi_s \perp \varphi_{s'}$, $s \not\sim s'$,

(ii) $\varphi_s \circ \theta_r^{-1} = \varphi_s = \varphi_{rs}$, $r \in G$, $s \in S$.

[4]Not to be confused with the topological notion of proper group action.

Proof: The φ_s are σ-finite since $\varphi_s g \equiv 1$. The first equality in (ii) holds by the left-invariance of λ, and the second holds since $\lambda \circ \pi_{rs}^{-1} = \Delta_r \lambda \circ \pi_s^{-1}$ by Lemma 3.10 (ii). The latter equation yields $\varphi_s = \varphi_{s'}$ when $s \sim s'$. The orthogonality holds for $s \not\sim s'$, since φ_s is confined to the orbit $\pi_s(G)$, which is universally measurable by Theorem A1.2 (i). $\square$

We show that any s-finite, G-invariant measure on S is a unique mixture of orbit measures. An invariant measure ν on S is said to be *ergodic* if $\nu I \wedge \nu I^c = 0$ for all $I \in \mathcal{I}_S$. It is further said to be *extreme*, if for any decomposition $\nu = \nu' + \nu''$ with invariant ν' and ν'', all three measures are proportional.

Theorem 3.12 *(invariant measures) Let G be a measurable group with Haar measure λ, acting properly on S, and define φ by* (10) *in terms of a normalizing function $g > 0$ on S. Then there is a $1{-}1$ correspondence between all s-finite, invariant measures ν on S and all s-finite measures μ on the range $\varphi(S)$, given by*

(i) $\quad \nu = \int \nu(ds)\, g(s)\, \varphi_s = \int m\, \mu(dm),$

(ii) $\quad \mu f = \int \nu(ds)\, g(s)\, f(\varphi_s), \quad f \geq 0.$

For such μ, ν, we have

(iii) *μ, ν are simultaneously σ-finite,*

(iv) *ν is ergodic or extreme iff μ is degenerate,*

(v) *there exists a σ-finite, G-invariant measure $\tilde{\nu} \sim \nu$.*

Proof, (i)−(ii): Let ν be s-finite and G-invariant. Using (10), Fubini's theorem, and Theorem 3.10 (iii), we get for any $f \in \mathcal{S}_+$

$$\begin{aligned}
\int \nu(ds)\, g(s)\, \varphi_s f &= \int \nu(ds)\, g(s)\, \frac{\lambda(f \circ \pi_s)}{\lambda(g \circ \pi_s)} \\
&= \int \lambda(dr) \int \nu(ds)\, \frac{g(s)\, f(rs)}{\lambda(g \circ \pi_s)} \\
&= \int \lambda(dr) \int \nu(ds)\, \frac{g(r^{-1}s)\, f(s)}{\lambda(g \circ \pi_{r^{-1}s})} \\
&= \int \lambda(dr)\, \Delta_r \int \nu(ds)\, \frac{g(r^{-1}s)\, f(s)}{\lambda(g \circ \pi_s)} \\
&= \int \lambda(dr) \int \nu(ds)\, \frac{g(rs)\, f(s)}{\lambda(g \circ \pi_s)} \\
&= \int \nu(ds)\, f(s) = \nu f,
\end{aligned}$$

which proves the first relation in (i). The second relation follows with μ as in (ii), by the substitution rule for integrals.

Conversely, let $\nu = \int m\,\mu(dm)$ for an s-finite measure μ on $\varphi(S)$. For measurable $M \subset \mathcal{M}_S$ we have $A \equiv \varphi^{-1}M \in \mathcal{I}_S$ by Lemma 3.11, and so

$$\begin{aligned}\varphi_s(g;A) &= \frac{(\lambda\circ\pi_s^{-1})(1_A g)}{\lambda(g\circ\pi_s)}\\ &= 1_A(s) = 1_M(\varphi_s).\end{aligned}$$

Hence,

$$\begin{aligned}(g\cdot\nu)\circ\varphi^{-1}M &= \nu(g;A) = \int \mu(dm)\,m(g;A)\\ &= \int_M \mu(dm) = \mu M,\end{aligned}$$

which shows that μ is uniquely given by (ii) and is therefore s-finite.

(iii) Use (i)−(ii) and the facts that $g > 0$ and $\varphi_s \neq 0$.

(iv) Since μ is unique, ν is extreme iff μ is degenerate. For any $I \in \mathcal{I}_S$,

$$\varphi_s I \wedge \varphi_s I^c = \frac{\lambda G}{\lambda(g\circ\pi_s)}\Big\{1_I(s)\wedge 1_{I^c}(s)\Big\} = 0, \quad s \in S,$$

which shows that the φ_s are ergodic. Conversely, let $\nu = \int m\,\mu(dm)$ with μ non-degenerate, so that $\mu M \wedge \mu M^c \neq 0$ for a measurable subset $M \subset \mathcal{M}_S$. Then $I = \varphi^{-1}M \in \mathcal{I}_S$ with $\nu I \wedge \nu I^c \neq 0$, which shows that ν is non-ergodic. Hence, ν is ergodic iff μ is degenerate.

(v) Choose a bounded measure $\hat\nu \sim \nu$, and define $\tilde\nu = \hat\nu\varphi$. Then $\tilde\nu$ is invariant by Lemma 3.11, and $\tilde\nu \sim (g\cdot\nu)\varphi = \nu$ by (ii). It is also σ-finite, since $\tilde\nu g = \hat\nu\varphi g = \hat\nu S < \infty$ by (i). □

When G acts measurably on S and T, we say that a kernel $\mu\colon S \to T$ is *invariant*[5] if

$$\mu_{rs} = \mu_s \circ \theta_r^{-1}, \quad r \in G,\ s \in S.$$

The definition is motivated by the following observation:

Lemma 3.13 *(invariant kernels) Let the group G act measurably on the Borel spaces S, T, and consider a σ-finite, invariant measure ν on S and a kernel $\mu\colon S \to T$. Then these conditions are equivalent:*

(i) *the measure $\rho = \nu \otimes \mu$ is jointly invariant on $S \times T$,*

(ii) *for any $r \in G$, the kernel μ is r-invariant in $s \in S$ a.e. ν.*

Proof, (ii) ⇒ (i): Assuming (ii), we get for any measurable $f \geq 0$ on $S \times T$

$$\begin{aligned}\Big\{(\nu\otimes\mu)\circ\theta_r^{-1}\Big\}f &= (\nu\otimes\mu)(f\circ\theta_r)\\ &= \int \nu(ds)\int \mu_s(dt)\,f(rs,rt)\\ &= \int \nu(ds)\int \Big(\mu_s\circ\theta_r^{-1}\Big)(dt)\,f(rs,t)\end{aligned}$$

[5]If we write $\mu\circ\theta_r^{-1} = \theta_r\mu$, the defining condition simplifies to $\mu_{rs} = \theta_r\mu_s$.

$$\begin{aligned}&= \int \nu(ds) \int \mu_{rs}(dt)\, f(rs,t)\\ &= \int \nu(ds) \int \mu_s(dt)\, f(s,t)\\ &= (\nu \otimes \mu) f.\end{aligned}$$

(i) ⇒ (ii): Assuming (i) and fixing $r \in G$, we get by the same calculation

$$\int \left(\mu_s \circ \theta_r^{-1}\right)(dt)\, f(rs,t) = \int \mu_{rs}(dt)\, f(rs,t), \quad s \in S \text{ a.e. } \nu,$$

which extends to (ii) since f is arbitrary. □

We turn to the reverse problem of invariant disintegration.

Theorem 3.14 *(invariant disintegration) Let G be a measurable group with Haar measure λ, acting properly on S and measurably on T, where S, T are Borel, and consider a σ-finite, jointly G-invariant measure ρ on $S \times T$. Then*

(i) *$\rho = \nu \otimes \mu$ for a σ-finite, invariant measure ν on S and an invariant kernel $\mu\colon S \to T$,*

(ii) *when $S = G$ we may choose $\nu = \lambda$, in which case μ is given uniquely, for any $B \in \mathcal{G}$ with $\lambda B = 1$, by*

$$\mu_r f = \iint_{B\times S} \rho(dp\, ds)\, f(rp^{-1}s), \quad r \in G,\ f \geq 0.$$

We begin with part (ii), which may be proved by a simple *skew factorization.*

Proof of (ii): On $G \times S$ we define the mappings

$$\begin{aligned}\vartheta(r,s) &= (r, rs),\\ \theta_r(p,s) &= (rp, rs),\\ \theta'_r(p,s) &= (rp, s),\end{aligned}$$

where $p, r \in G$ and $s \in S$. Since clearly $\vartheta^{-1} \circ \theta_r = \theta'_r \circ \vartheta^{-1}$, the joint G-invariance of ρ gives

$$\begin{aligned}\rho \circ \vartheta \circ \theta_r'^{-1} &= \rho \circ \theta_r^{-1} \circ \vartheta\\ &= \rho \circ \vartheta, \quad r \in G,\end{aligned}$$

where $\vartheta = (\vartheta^{-1})^{-1}$. Hence, $\rho \circ \vartheta$ is invariant under shifts of G alone, and so $\rho \circ \vartheta = \lambda \otimes \nu$ by Theorem 3.10 for a σ-finite measure ν on S, which yields the stated formula for the invariant kernel $\mu_r = \nu \circ \theta_r^{-1}$. To see that $\rho = \lambda \otimes \mu$, let $f \geq 0$ be measurable on $G \times S$, and write

$$\begin{aligned}(\lambda \otimes \mu) f &= \int \lambda(dr) \int \nu \circ \theta_r^{-1}(ds)\, f(r,s)\\ &= \int \lambda(dr) \int \nu(ds)\, f(r, rs)\\ &= \left\{(\lambda \otimes \nu) \circ \vartheta^{-1}\right\} f = \rho f.\end{aligned}$$

□

To prepare for the proof of part (i), we begin with the special case where S contains G as a factor.

Corollary 3.15 *(repeated disintegration) Let G be a measurable group with Haar measure λ, acting measurably on the Borel spaces S, T, and consider a σ-finite, jointly invariant measure ρ on $G \times S \times T$. Then*

(i) *$\rho = \lambda \otimes \nu \otimes \mu$ for some invariant kernels $\nu: G \to S$ and $\mu: G \times S \to T$,*

(ii) *$\nu_r \equiv \hat\nu \circ \theta_r^{-1}$ and $\mu_{r,s} \equiv \hat\mu_{r^{-1}s} \circ \theta_r^{-1}$ for some measure $\hat\nu$ on S and kernel $\hat\mu: S \to T$,*

(iii) *the measure $\hat\nu \otimes \hat\mu$ on $S \times T$ is unique,*

(iv) *we may choose ν as any invariant kernel $G \to S$ with $\lambda \otimes \nu \sim \rho(\cdot \times T)$.*

When G acts measurably on S, its action on $G \times S$ is proper, and so for any s-finite, jointly G-invariant measure β on $G \times S$, Theorem 3.12 yields a σ-finite, G-invariant measure $\tilde\beta \sim \beta$. Hence, $\tilde\beta = \lambda \otimes \nu$ by Theorem 3.14 (ii) for an invariant kernel $\nu: G \to S$.

Proof: By Theorem 3.14 (ii) we have $\rho = (\lambda \otimes \chi) \circ \vartheta^{-1}$ for a σ-finite measure χ on $S \times T$, where $\vartheta(r,s,t) = (r, rs, rt)$ for all $(r,s,t) \in G \times S \times T$. Since T is Borel, Theorem 3.4 yields a further disintegration $\chi = \hat\nu \otimes \hat\mu$ in terms of a measure $\hat\nu$ on S and a kernel $\hat\mu: S \to T$, generating some invariant kernels ν and μ as in (ii). Then (i) may be verified as before, and (iii) holds by the uniqueness of χ.

(iv) For any kernel ν' as stated, write $\hat\nu'$ for the generating measure on S. Since $\lambda \times \nu' \sim \lambda \times \mu$, we have $\hat\nu' \sim \hat\nu$, and so $\chi = \hat\nu' \otimes \hat\mu'$ for some kernel $\hat\mu': S \to T$. Now continue as before. $\square$

We also need some technical facts:

Lemma 3.16 *(invariance sets) Let G be a group with Haar measure λ, acting measurably on S, T. Fix a kernel $\mu: S \to T$ and a measurable, jointly G-invariant function f on $S \times T$. Then these subsets of S and $S \times T$ are G-invariant:*

$$\begin{aligned} A &= \Big\{ s \in S;\ \mu_{rs} = \mu_s \circ \theta_r^{-1},\ r \in G \Big\}, \\ B &= \Big\{ (s,t) \in S \times T;\ f(rs,t) = f(ps,t),\ (r,p) \in G^2 \ \textit{a.e.}\ \lambda^2 \Big\}. \end{aligned}$$

Proof: For any $s \in A$ and $r, p \in G$, we get

$$\begin{aligned} \mu_{rs} \circ \theta_p^{-1} &= \mu_s \circ \theta_r^{-1} \circ \theta_p^{-1} \\ &= \mu_s \circ \theta_{pr}^{-1} = \mu_{prs}, \end{aligned}$$

which shows that even $rs \in A$. Conversely, $rs \in A$ implies $s = r^{-1}(rs) \in A$, and so $\theta_r^{-1} A = A$.

Now let $(s,t) \in B$. Then Theorem 3.10 (ii) yields $(qs,t) \in B$ for any $q \in G$, and the invariance of λ gives

$$f\big(q^{-1} r q s, t\big) = f\big(q^{-1} p\, q s, t\big), \quad (r,p) \in G^2 \ \text{a.e.}\ \lambda^2,$$

which implies $(qs, qt) \in B$ by the joint G-invariance of f. Conversely, $(qs, qt) \in B$ implies $(s, t) = q^{-1}(qs, qt) \in B$, and so $\theta_q^{-1} B = B$. $\square$

Proof of Theorem 3.14 (i): The measure $\rho(\cdot \times T)$ is clearly s-finite and G-invariant on S. Since G acts properly on S, Theorem 3.12 yields a σ-finite, G-invariant measure ν on S. Applying Corollary 3.15 to the G-invariant measures $\lambda \otimes \rho$ on $G \times S \times T$ and $\lambda \otimes \nu$ on $G \times S$, we obtain a G-invariant kernel $\beta \colon G \times S \to T$ satisfying $\lambda \otimes \rho = \lambda \otimes \nu \otimes \beta$.

Now introduce the kernels $\beta_r(p, s) = \beta(rp, s)$ and write $f_r(p, s, t) = f(r^{-1}p, s, t)$ for any measurable function $f \geq 0$ on $G \times S \times T$. By the invariance of λ, we have for any $r \in G$

$$\begin{aligned}\big(\lambda \otimes \nu \otimes \beta_r\big) f &= \big(\lambda \otimes \nu \otimes \beta\big) f_r \\ &= (\lambda \otimes \rho) f_r = (\lambda \otimes \rho) f,\end{aligned}$$

and so $\lambda \otimes \rho = \lambda \otimes \nu \otimes \beta_r$. Hence, by the a.e. uniqueness

$$\beta(rp, s) = \beta(p, s), \quad (p, s) \in G \times S \ \text{ a.e. } \lambda \otimes \nu, \ \ r \in G.$$

Let A be the set of all $s \in S$ satisfying

$$\beta(rp, s) = \beta(p, s), \quad (r, p) \in G^2 \ \text{ a.e. } \lambda^2,$$

and note that $\nu A^c = 0$ by Fubini's theorem. By Lemma 3.10 (ii), the defining condition is equivalent to

$$\beta(r, s) = \beta(p, s), \quad (r, p) \in G^2 \ \text{ a.e. } \lambda^2,$$

and so for any $g, h \in \mathcal{G}_+$ with $\lambda g = \lambda h = 1$,

$$(g \cdot \lambda)\beta(\cdot, s) = (h \cdot \lambda)\beta(\cdot, s) = \mu(s), \quad s \in A. \tag{11}$$

To extend this to an identity, we may redefine $\beta(r, s) = 0$ when $s \in A^c$. This will not affect the disintegration of $\lambda \otimes \rho$, and by Lemma 3.16 it even preserves the G-invariance of β. Fixing a $g \in \mathcal{G}_+$ with $\lambda g = 1$, we get for any $f \geq 0$ on $S \times T$

$$\begin{aligned}\rho f &= (\lambda \otimes \rho)(g \otimes f) \\ &= \big(\lambda \otimes \nu \otimes \beta\big)(g \otimes f) \\ &= \big\{\nu \otimes (g \cdot \lambda)\beta\big\} f \\ &= (\nu \otimes \mu) f,\end{aligned}$$

and so $\rho = \nu \otimes \mu$. Now combine (11) with the G-invariance of β and λ to get

$$\begin{aligned}\mu_s \circ \theta_r^{-1} &= \int \lambda(dp)\, g(p)\, \beta(p, s) \circ \theta_r^{-1} \\ &= \int \lambda(dp)\, g(p)\, \beta(rp, rs) \\ &= \int \lambda(dp)\, g(r^{-1}p)\, \beta(p, rs) = \mu_{rs},\end{aligned}$$

which shows that μ is G-invariant. □

We turn to some notions of asymptotic and local invariance, of importance in Chapters 25, 27, and 30–31. The uniformly bounded measures μ_n on R^d are said to be *asymptotically invariant* if

$$\left\| \mu_n - \mu_n * \delta_x \right\| \to 0, \quad x \in \mathsf{R}^d. \tag{12}$$

Lemma 3.17 *(asymptotic invariance) Let $\mu_1, \mu_2, \ldots$ be uniformly bounded, asymptotically invariant measures on R^d. Then*

(i) (12) *holds uniformly for bounded x,*

(ii) $\|\mu_n - \mu_n * \nu\| \to 0$ *for any distribution ν on R^d,*

(iii) *the singular components μ_n'' of μ_n satisfy $\|\mu_n''\| \to 0$.*

Proof: (ii) For any distribution ν, we get by dominated convergence

$$\begin{aligned}\left\| \mu_n - \mu_n * \nu \right\| &= \left\| \int \left(\mu_n - \mu_n * \delta_x\right) \nu(dx) \right\| \\ &\le \int \left\| \mu_n - \mu_n * \delta_x \right\| \nu(dx) \to 0.\end{aligned}$$

(iii) When ν is absolutely continuous, so is $\mu_n * \nu$ for all n, and we get

$$\|\mu_n''\| \le \left\| \mu_n - \mu_n * \nu \right\| \to 0.$$

(i) Writing λ_h for the uniform distribution on $[0,h]^d$ and noting that $\|\mu * \nu\| \le \|\mu\| \, \|\nu\|$, we get for any $x \in \mathsf{R}^d$ with $|x| \le r$

$$\begin{aligned}\left\| \mu_n - \mu_n * \delta_x \right\| &\le \left\| \mu_n - \mu_n * \lambda_h \right\| + \left\| \mu_n * \lambda_h - \mu_n * \lambda_h * \delta_x \right\| \\ &\quad + \left\| \mu_n * \lambda_h * \delta_x - \mu_n * \delta_x \right\| \\ &\le 2 \left\| \mu_n - \mu_n * \lambda_h \right\| + \left\| \lambda_h - \lambda_h * \delta_x \right\| \\ &\le 2 \left\| \mu_n - \mu_n * \lambda_h \right\| + 2\, r h^{-1} d^{1/2},\end{aligned}$$

which tends to 0 for fixed r, as $n \to \infty$ and then $h \to \infty$. □

Next we say that the measures μ_n on R^d are *weakly asymptotically invariant*, if they are uniformly bounded and such that the convolutions $\mu_n * \nu$ are asymptotically invariant for every bounded measure $\nu \ll \lambda^d$ on R^d, so that

$$\left\| \mu_n * \nu - \mu_n * \nu * \delta_x \right\| \to 0, \quad x \in \mathsf{R}^d. \tag{13}$$

Lemma 3.18 *(weak asymptotic invariance) For any uniformly bounded measures $\mu_1, \mu_2, \ldots$ on R^d, these conditions are equivalent:*

(i) *the μ_n are weakly asymptotically invariant,*

(ii) $\displaystyle\int \left| \mu_n(I_h + x) - \mu_n(I_h + x + he_i) \right| dx \to 0, \quad h > 0, \; i \le d.$

(iii) $\left\| \mu_n * \nu - \mu_n * \nu' \right\| \to 0$ *for all probability measures* $\nu, \nu' \ll \lambda^d$.

In (ii) *it suffices to consider dyadic* $h = 2^{-k}$ *and to replace the integration by summation over* $(h\mathsf{Z})^d$.

Proof, (i) $\Leftrightarrow$ (ii): Letting $\nu = f \cdot \lambda^d$ and $\nu' = f' \cdot \lambda^d$, we note that

$$\begin{aligned}\left\| \mu * (\nu - \nu') \right\| &\le \|\mu\| \, \|\nu - \nu'\| \\ &= \|\mu\| \, \|f - f'\|_1.\end{aligned}$$

By Lemma 1.37 it is then enough in (13) to consider measures $\nu = f \cdot \lambda^d$, where f is continuous with bounded support. We may also take $r = he_i$ for some $h > 0$ and $i \le d$, and by uniform continuity we may restrict h to dyadic values 2^{-k}. By uniform continuity, we may next approximate f in L^1 by simple functions f_n over the cubic grids $\mathcal{I}_n$ in R^d of mesh size 2^{-n}, which implies the equivalence with (ii). The last assertion is yet another consequence of the uniform continuity.

(i) $\Leftrightarrow$ (iii): Condition (13) is clearly a special case of (iii). To show that (13) $\Rightarrow$ (iii), we may approximate as in (ii) to reduce to the case of finite sums $\nu = \sum_j a_j \lambda_{mj}$ and $\nu' = \sum_j b_j \lambda_{mj}$, where $\sum_j a_j = \sum_j b_j = 1$, and λ_{mj} denotes the uniform distribution over the cube $I_{mj} = 2^{-m}[j-1, j]$ in R^d. Assuming (i) and writing $h = 2^{-m}$, we get

$$\begin{aligned}\left\| \mu_n * \nu - \mu_n * \nu' \right\| &\le \left\| \mu_n * \nu - \mu_n * \lambda_{m0} \right\| + \left\| \mu_n * \nu' - \mu_n * \lambda_{m0} \right\| \\ &\le \textstyle\sum_j (a_j + b_j) \left\| \mu_n * \lambda_{mj} - \mu_n * \lambda_{m0} \right\| \\ &= \textstyle\sum_j (a_j + b_j) \left\| \mu_n * \lambda_{m0} * \delta_{jh} - \mu_n * \lambda_{m0} \right\| \to 0,\end{aligned}$$

by the criterion in (ii). □

Now let γ be a random vector in R^d with distribution μ, and write $\mu_t = \mathcal{L}(\gamma_t)$, where $\gamma_t = t\gamma$. Then μ is said to be *locally invariant*, if the μ_t are weakly asymptotically invariant as $t \to \infty$. We might also say that μ is *strictly* locally invariant, if the μ_t are asymptotically invariant in the strict sense. However, the latter notion gives nothing new:

Theorem 3.19 *(local invariance) Let* $\mu' \ll \mu$ *be locally finite measures on* R^d. *Then*

(i) μ *strictly locally invariant* $\Leftrightarrow$ $\mu \ll \lambda^d$,

(ii) μ *locally invariant* $\Rightarrow$ μ' *locally invariant.*

Proof: (i) Let μ be strictly locally invariant with singular component μ''. Then Lemma 3.17 yields $\|\mu''\| = \|\mu'' \circ S_t^{-1}\| \to 0$ as $t \to \infty$, where $S_t x = tx$, which implies $\mu'' = 0$ and hence $\mu \ll \lambda^d$.

Conversely, let $\mu = f\cdot\lambda^d$ for some measurable function $f \geq 0$. Then Lemma 1.37 yields some continuous functions $f_n \geq 0$ with bounded supports such that $\|f - f_n\|_1 \to 0$. Writing $\mu_t = \mu \circ S_t^{-1}$ and $h = 1/t$, we get for any $r \in \mathsf{R}^d$

$$\begin{aligned}\big\|\, \mu_t - \mu_t * \delta_r \big\| &= \int \big| f(x) - f(x+hr)\big|\, dx \\ &\leq \int \big| f_n(x) - f_n(x+hr)\big|\, dx + 2\, \|f - f_n\|_1,\end{aligned}$$

which tends to 0 as $h \to 0$ and then $n \to \infty$. Since r was arbitrary, the strict local invariance follows.

(ii) We may take μ to be bounded, so that its local invariance reduces to weak asymptotic invariance of the measures $\mu_t = \mu \circ S_t^{-1}$. Since $\mu' \ll \mu$ is locally finite, we may further choose $\mu' = f \cdot \mu$ for some μ-integrable function $f \geq 0$ on R^d. Then Lemma 1.37 yields some continuous functions f_n with bounded supports, such that $f_n \to f$ in $L^1(\mu)$. Writing $\mu'_n = f_n \cdot \mu$ and $\mu'_{n,t} = \mu'_n \circ S_t^{-1}$, we get

$$\begin{aligned}\big\|\, \mu'_{n,t} * \nu - \mu'_t * \nu \,\big\| &\leq \big\|\, \mu'_{n,t} - \mu'_t \big\| + \big\|\, \mu'_n - \mu' \big\| \\ &= \big\|(f_n - f)\cdot \mu\big\| \\ &= \mu|f_n - f| \to 0,\end{aligned}$$

and so it suffices to verify (13) for the measures $\mu'_{n,t}$. Thus, we may henceforth choose $\mu' = f \cdot \mu$ for a continuous function f with bounded support.

To verify (13), let r and ν be such that the measures ν and $\nu * \delta_r$ are both supported by the ball B_0^a. Writing

$$\begin{aligned}\mu'_t &= (f\cdot\mu)_t \\ &= (f\cdot\mu)\circ S_t^{-1} \\ &= (f \circ S_t)\cdot(\mu \circ S_t^{-1}) \\ &= f_t \cdot \mu_t, \\ \nu - \nu * \delta_r &= g\cdot\lambda^d - (g\circ\theta_{-r})\cdot\lambda^d \\ &= \Delta g \cdot \lambda^d,\end{aligned}$$

we get for any bounded, measurable function h on R^d

$$\begin{aligned}\Big(\mu'_t * \nu - \mu'_t * \nu * \delta_r\Big) h &= \Big\{(f_t\cdot\mu_t) * (\Delta g \cdot \lambda^d)\Big\} h \\ &= \int f_t(x)\, \mu_t(dx) \int \Delta g(y)\, h(x+y)\, dy \\ &= \int f_t(x)\, \mu_t(dx) \int \Delta g(y-x)\, h(y)\, dy \\ &= \int h(y)\, dy \int f_t(x)\, \Delta g(y-x)\, \mu_t(dx) \\ &\approx \int h(y)\, f_t(y)\, dy \int \Delta g(y-x)\, \mu_t(dx) \\ &= \Big\{\mu_t * (\Delta g\cdot\lambda^d)\Big\} f_t h.\end{aligned}$$

Letting m be the modulus of continuity of f and putting $m_t = m(r/t)$, we may estimate the approximation error in the fifth step by

$$\begin{aligned}\int |h(y)|\,dy \int \Big| f_t(x) - f_t(y) \Big| \, \Big| \Delta g(y-x) \Big| \, \mu_t(dx) & \\ \le m_t \int \mu_t(dx) \int \Big| h(y)\, \Delta g(y-x) \Big| \, dy & \\ \le m_t \, \|h\| \, \|\mu_t\| \int |\Delta g(y)| \, dy & \\ \le 2\, m_t \, \|h\| \, \|\mu\| \, \|\nu\| \to 0, &\end{aligned}$$

by the uniform continuity of f. It remains to note that

$$\Big\| f_t \cdot \Big\{ \mu_t * (\Delta g \cdot \lambda^d) \Big\} \Big\| \le \|f\| \, \Big\| \mu_t * \nu - \mu_t * \nu * \delta_t \Big\| \to 0,$$

by the local invariance of μ. $\square$

Exercises

1. Let $\mu_1, \mu_2, \ldots$ be kernels between two measurable spaces S, T. Show that the function $\mu = \sum_n \mu_n$ is again a kernel.

2. Fix a function f between two measurable spaces S, T, and define $\mu(s, B) = 1_B \circ f(s)$. Show that μ is a kernel iff f is measurable.

3. Prove the existence and uniqueness of Lebesgue measure on R^d from the corresponding properties of Haar measures.

4. Show that if the group G is compact, then every measurable group action is proper. Then show how the formula for the orbit measures simplifies in this case, and give a corresponding version of Theorem 3.12.

5. Use the previous exercise to find the general form of the invariant measures on a sphere and on a spherical cylinder.

6. Extend Lemma 3.10 (i) to the context of general invariant measures.

7. Extend the mean continuity in Lemma 2.7 to general invariant measures.

8. Give an example of some measures $\mu_1, \mu_2, \ldots$ that are weakly but not strictly asymptotically invariant.

9. Give an example of a bounded measure μ on R that is locally invariant but not absolutely continuous.

II. Some Classical Probability Theory

Armed with the basic measure-theoretic machinery from Part I, we proceed with the foundations of general probability theory. After giving precise definitions of processes, distributions, expectations, and independence, we introduce the various notions of probabilistic convergence, and prove some of the classical limit theorems for sums and averages, including the law of large numbers and the central limit theorem. Finally, a general compound Poisson approximation will enable us to give a short, probabilistic proof of the classical limit theorem for general null arrays. The material in Chapter 4, along with at least the beginnings of Chapters 5−6, may be regarded as essential core material, constantly needed for the continued study, whereas a detailed reading of the remaining material might be postponed until a later stage.

— — —

4. Processes, distributions, and independence. Here the probabilistic notions of processes, distributions, expected values, and independence are defined in terms of standard notions of measure theory from Chapter 1. We proceed with a detailed study of independence criteria, 0−1 laws, and regularity conditions for processes. Finally, we prove the existence of sequences of independent random variables, and show how distributions on R^d are induced by their distribution functions.

5. Random sequences, series, and averages. Here we examine the relationship between convergence a.s., in probability, and in L^p, and give conditions for uniform integrability. Next, we develop convergence criteria for random series and averages, including the strong laws of large numbers. Finally, we study the basic properties of convergence in distribution, with special emphasis on the continuous mapping and approximation theorems as well as the Skorohod coupling.

6. Gaussian and Poisson convergence. Here we prove the classical Poisson and Gaussian limit theorems for null arrays of random variables or vectors, including the celebrated central limit theorem. The theory of regular variation enables us to identify the domain of Gaussian attraction. The mentioned results also yield precise criteria for the weak law of large numbers. Finally, we study the relationship between weak compactness and tightness, in the special case of distributions on Euclidean spaces.

7. Infinite divisibility and general null arrays. Here we establish the representation of infinitely divisible distributions on R^d, involving a Gaussian component and a compound Poisson integral, material needed for the theory of Lévy processes and essential for a good understanding of Feller processes and general semi-martingales. The associated convergence criteria, along with a basic compound Poisson approximation, yield a short, non-technical approach to the classical limit theorem for general null arrays.

Chapter 4

Processes, Distributions, and Independence

Random elements and processes, finite-dimensional distributions, expectation and covariance, moments and tails, Jensen's inequality, independence, pairwise independence and grouping, product measures and convolution, iterated expectation, Kolmogorov and Hewitt–Savage $0-1$ *laws, Borel–Cantelli lemma, replication and Bernoulli sequences, kernel representation, Hölder continuity, multi-variate distributions*

Armed with the basic notions and results of measure theory from previous chapters, we may now embark on our study of proper probability theory. The dual purposes of this chapter are to introduce some basic terminology and notation and to prove some fundamental results, many of which will be needed throughout the remainder of the book.

In modern probability theory, it is customary to relate all objects of study to a basic *probability space* $(\Omega, \mathcal{A}, P)$, which is simply a normalized measure space. *Random variables* may then be defined as measurable functions ξ on Ω, and their *expected values* as integrals $E\xi = \int \xi \, dP$. Furthermore, *independence* between random quantities reduces to a kind of orthogonality between the induced sub-σ-fields. The reference space Ω is introduced only for technical convenience, to provide a consistent mathematical framework, and its actual choice plays no role[1]. Instead, our interest focuses on the induced distributions $\mathcal{L}(\xi) = P \circ \xi^{-1}$ and their associated characteristics.

The notion of independence is fundamental for all areas of probability theory. Despite its simplicity, it has some truly remarkable consequences. A particularly striking result is *Kolmogorov's* $0-1$ *law*, stating that every tail event associated with a sequence of independent random elements has probability 0 or 1. Thus, any random variable that depends only on the 'tail' of the sequence is a.s. a constant. This result and the related *Hewitt–Savage* $0-1$ *law* convey much of the flavor of modern probability: though the individual elements of a random sequence are erratic and unpredictable, the long-term behavior may often conform to deterministic laws and patterns. A major objective is then to uncover the latter. Here the classical *Borel–Cantelli lemma* is one of the simplest but most powerful tools available.

To justify our study, we need to ensure the existence[2] of the random objects

[1]except in some special contexts, such as in the modern theory of Markov processes

[2]The entire theory relies in a subtle way on the existence of non-trivial, countably additive measures. Assuming only finite additivity wouldn't lead us very far.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_5

under discussion. For most purposes it suffices to use the Lebesgue unit interval $([0,1],\mathcal{B},\lambda)$ as our basic probability space. In this chapter, the existence will be proved only for independent random variables with prescribed distributions, whereas a more general discussion is postponed until Chapter 8. As a key step, we may use the binary expansion of real numbers to construct a so-called *Bernoulli sequence*, consisting of independent random digits 1 or 0 with probabilities p and $1-p$, respectively.[3] The latter may be regarded as a discrete-time counterpart of the fundamental *Poisson processes*, to be introduced and studied in Chapter 15.

The distribution of a random process X is determined by its finite-dimensional distributions, which are not affected by a change on a null set of each variable X_t. It is then natural to look for versions of X with suitable regularity properties. As another striking result, we provide a moment condition ensuring the existence of a Hölder continuous modification of the process. Regularizations of various kind are important throughout modern probability theory, as they may enable us to deal with events depending on the values of a process at uncountably many times.

To begin our systematic exposition of the theory, we fix an arbitrary *probability space* $(\Omega,\mathcal{A},P)$, where P is a *probability measure* with total mass 1. In the probabilistic context, the sets $A\in\mathcal{A}$ are called *events*, and $PA=P(A)$ is called the *probability* of A. In addition to results valid for all measures, there are properties requiring the boundedness or normalization of P, such as the relation $PA^c=1-PA$ and the fact that $A_n\downarrow A$ implies $PA_n\to PA$.

Some infinite set operations have a special probabilistic significance. Thus, for any sequence of events $A_1,A_2,\ldots\in\mathcal{A}$, we may consider the sets $\{A_n \text{ i.o.}\}$ where A_n happens *infinitely often*, and $\{A_n \text{ ult.}\}$ where A_n happens *ultimately* (for all but finitely many n). Those occurrences are events in their own right, formally expressible in terms of the A_n as

$$\{A_n \text{ i.o.}\} = \Big\{\textstyle\sum_n 1_{A_n}=\infty\Big\} = \bigcap_n \bigcup_{k\ge n} A_k, \tag{1}$$

$$\{A_n \text{ ult.}\} = \Big\{\textstyle\sum_n 1_{A_n^c}<\infty\Big\} = \bigcup_n \bigcap_{k\ge n} A_k. \tag{2}$$

The argument $\omega\in\Omega$ is usually omitted from our notation, when there is no risk for confusion. Thus, for example, the expression $\{\sum_n 1_{A_n}=\infty\}$ is a convenient shorthand form of the unwieldy $\{\omega\in\Omega;\ \sum_n 1_{A_n}(\omega)=\infty\}$.

The indicator functions[4] of the events in (1) and (2) may be expressed as

$$\begin{aligned} 1\{A_n \text{ ult.}\} &= \liminf_{n\to\infty} 1_{A_n} \\ &\le \limsup_{n\to\infty} 1_{A_n} = 1\{A_n \text{ i.o.}\}. \end{aligned}$$

Applying Fatou's lemma to the functions 1_{A_n} and $1_{A_n^c}$ yields

$$P\{A_n \text{ ult.}\} \le \liminf_{n\to\infty} PA_n$$

[3]Its existence is equivalent to that of Lebesgue measure, hence highly non-trivial, as we have seen in Chapter 2.

[4]For typographical convenience, we are often writing $1\{\cdot\}$ instead of $1_{\{\cdot\}}$.

$$\begin{aligned} &\le \limsup_{n\to\infty} PA_n \\ &\le P\{A_n \text{ i.o.}\}. \end{aligned}$$

Using the continuity and sub-additivity of P, we further see from (1) that

$$P\{A_n \text{ i.o.}\} = \lim_{n\to\infty} P\bigcup_{k\ge n} A_k \le \lim_{n\to\infty} \sum_{k\ge n} PA_k.$$

When $\sum_n PA_n < \infty$, the right-hand side becomes 0, and we get $P\{A_n \text{ i.o.}\} = 0$. The resulting implication constitutes the easy part of the *Borel–Cantelli lemma*, to be revisited in Theorem 4.18.

Any measurable mapping ξ of Ω into some measurable space $(S, \mathcal{S})$ is called a *random element* in S. If $B \in \mathcal{S}$, then $\{\xi \in B\} = \xi^{-1}B \in \mathcal{A}$, and we may consider the associated probabilities

$$\begin{aligned} P\{\xi \in B\} &= P(\xi^{-1}B) \\ &= (P \circ \xi^{-1})B, \quad B \in \mathcal{S}. \end{aligned}$$

The set function $\mathcal{L}(\xi) = P \circ \xi^{-1}$ is a probability measure on the range space S of ξ, called the *distribution* or *law* of ξ. We often use the term *distribution* as synonymous to probability measure, even when no generating random element has been introduced.

Random elements are of interest in a wide variety of spaces. A random element in S is called a *random variable* when $S = \mathsf{R}$, a *random vector* when $S = \mathsf{R}^d$, a *random sequence* when $S = \mathsf{R}^\infty$, a *random* (or *stochastic*[5]) *process* when S is a function space, and a *random measure* or *set* when S is a class of measures or sets, respectively. A metric or topological space S will be endowed with its Borel σ-field $\mathcal{B}_S$, unless a σ-field is otherwise specified. For any separable metric space S, Lemma 1.2 shows that $\xi = (\xi_1, \xi_2, \ldots)$ is a random element in S^∞ iff $\xi_1, \xi_2, \ldots$ are random elements in S.

For any measurable space $(S, \mathcal{S})$, a subset $A \subset S$ becomes a measurable space in its own right, when endowed with the σ-field $A \cap \mathcal{S} = \{A \cap B;\, B \in \mathcal{S}\}$. In particular, Lemma 1.6 shows that if S is a metric space with Borel σ-field $\mathcal{S}$, then $A \cap \mathcal{S}$ is the Borel σ-field in A. Any random element in $(A, A \cap \mathcal{S})$ may be regarded, alternatively, as a random element in S. Conversely, if ξ is a random element in S with $\xi \in A$ a.s.[6] for some $A \in \mathcal{S}$, then $\xi = \eta$ a.s. for some random element η in A.

Fixing a measurable space $(S, \mathcal{S})$ and an index set T, we write S^T for the class of functions $f\colon T \to S$, and let $\mathcal{S}^T$ denote the σ-field in S^T generated by all *evaluation maps* $\pi_t\colon S^T \to S$, $t \in T$, given by $\pi_t f = f(t)$. If $X\colon \Omega \to U \subset S^T$, then clearly $X_t = \pi_t \circ X$ maps Ω into S. Thus, X may also be regarded as a function $X(t, \omega) = X_t(\omega)$ from $T \times \Omega$ to S.

Lemma 4.1 *(measurability) For any measurable space $(S, \mathcal{S})$, index set T, subset $U \subset S^T$, and function $X\colon \Omega \to U$, these conditions are equivalent:*

[5] We regard the words *random* and *stochastic* as synonymous.

[6] almost surely or with probability 1

(i) *X is $U \cap \mathcal{S}^T$-measurable,*

(ii) *$X_t \colon \Omega \to S$ is $\mathcal{S}$-measurable for every $t \in T$.*

Proof: Since X is U-valued, its $U \cap \mathcal{S}^T$-measurability is equivalent to measurability with respect to $\mathcal{S}^T$. The result now follows, by Lemma 1.4, from the fact that $\mathcal{S}^T$ is generated by the mappings π_t. □

A mapping X with the properties in Lemma 4.1 is called an *S-valued (random) process* on T with *paths* in U. By the lemma it is equivalent to regard X as a collection of random elements X_t in the *state space* S.

For any random elements ξ and η in a common measurable space, the equality $\xi \overset{d}{=} \eta$ means that ξ and η have the same distribution, or $\mathcal{L}(\xi) = \mathcal{L}(\eta)$. If X is a random process on an index set T, the associated *finite-dimensional distributions* are given by

$$\mu_{t_1,\dots,t_n} = \mathcal{L}(X_{t_1}, \dots, X_{t_n}), \quad t_1, \dots, t_n \in T,\ n \in \mathsf{N}.$$

The distribution of a process is determined by the set of finite-dimensional distributions:

Proposition 4.2 *(finite-dimensional distributions) For S, T, U as in Lemma* 4.1*, let X, Y be processes on T with paths in U. Then $X \overset{d}{=} Y$ iff*

$$(X_{t_1}, \dots, X_{t_n}) \overset{d}{=} (Y_{t_1}, \dots, Y_{t_n}), \quad t_1, \dots, t_n \in T,\ n \in \mathsf{N}. \tag{3}$$

Proof: Assume (3), let $\mathcal{D}$ be the class of sets $A \in \mathcal{S}^T$ with $P\{X \in A\} = P\{Y \in A\}$, and let $\mathcal{C}$ consist of all sets

$$A = \{f \in S^T;\ (f_{t_1}, \dots, f_{t_n}) \in B\}, \quad t_1, \dots, t_n \in T,\ B \in \mathcal{S}^n,\ n \in \mathsf{N}.$$

Then $\mathcal{C}$ is a π-system while $\mathcal{D}$ a λ-system, and $\mathcal{C} \subset \mathcal{D}$ by hypothesis. Hence, $\mathcal{S}^T = \sigma(\mathcal{C}) \subset \mathcal{D}$ by Theorem 1.1, which means that $X \overset{d}{=} Y$. □

For any random vector $\xi = (\xi_1, \dots, \xi_d)$ in R^d, we define the associated *distribution function* F by

$$F(x_1, \dots, x_d) = P \bigcap_{k \le d} \{\xi_k \le x_k\}, \quad x_1, \dots, x_d \in \mathsf{R}.$$

We note that F determines the distribution of ξ.

Lemma 4.3 *(distribution functions) Let ξ, η be random vectors in R^d with distribution functions F, G. Then*

$$\xi \overset{d}{=} \eta \quad \Leftrightarrow \quad F = G.$$

Proof: Use Theorem 1.1. □

The *expected value, expectation*, or *mean* of a random variable ξ, written as[7] $E\,\xi$, is defined as

$$E\,\xi = \int_\Omega \xi\, dP = \int_{\mathsf{R}} x\,(P\circ\xi^{-1})(dx), \tag{4}$$

whenever either integral exists. The last equality then holds by Lemma 1.24. By the same result, we get for any random elements ξ in a measurable space S, and for measurable functions $f\colon S\to\mathsf{R}$

$$\begin{aligned} Ef(\xi) &= \int_\Omega f(\xi)\, dP = \int_S f(s)\,\{P\circ\xi^{-1}\}(ds) \\ &= \int_{\mathsf{R}} x\,\big\{P\circ(f\circ\xi)^{-1}\big\}(dx), \end{aligned} \tag{5}$$

provided that at least one of the three integrals exists. Integrals over a measurable subset $A\subset\Omega$ are often denoted by

$$E(\xi;A) = E(\xi\,1_A) = \int_A \xi\, dP, \quad A\in\mathcal{A}.$$

For any random variable ξ and constant $p>0$, the integral $E|\xi|^p = \|\xi\|_p^p$ is called the p-th *absolute moment* of ξ. By Hölder's inequality (or by Jensen's inequality in Lemma 4.5) we have $\|\xi\|_p \le \|\xi\|_q$ for $p\le q$, so that the corresponding L^p-spaces are non-increasing in p. If $\xi\in L^p$ and either $p\in\mathsf{N}$ or $\xi\ge 0$, we may further define the p-th *moment* of ξ as $E\,\xi^p$.

We give a useful relationship between moments and tail probabilities.

Lemma 4.4 *(moments and tails) For any random variable $\xi\ge 0$ and constant $p>0$, we have*

$$\begin{aligned} E\,\xi^p &= p\int_0^\infty P\{\xi>t\}\,t^{p-1}dt \\ &= p\int_0^\infty P\{\xi\ge t\}\,t^{p-1}dt. \end{aligned}$$

Proof: By Fubini's theorem and calculus,

$$\begin{aligned} E\,\xi^p &= p\,E\int_0^\xi t^{p-1}dt \\ &= p\,E\int_0^\infty 1\{\xi>t\}\,t^{p-1}dt \\ &= p\int_0^\infty P\{\xi>t\}\,t^{p-1}dt. \end{aligned}$$

The proof of the second expression is similar. □

A random vector $\xi=(\xi_1,\dots,\xi_d)$ or process $X=(X_t)$ is said to be *integrable*, if integrability holds for every component ξ_k or value X_t, in which case

[7] We omit the customary square brackets, as in $E[\xi]$, when there is no risk for confusion.

we may write $E\,\xi = (E\,\xi_1, \dots, E\,\xi_d)$ or $EX = (EX_t)$. Recall that a function $f\colon \mathsf{R}^d \to \mathsf{R}$ is said to be *convex* if

$$f\Big\{cx + (1-c)y\Big\} \le cf(x) + (1-c)f(y), \quad x, y \in \mathsf{R}^d,\ c \in [0,1]. \tag{6}$$

This may be written as $f(E\,\xi) \le Ef(\xi)$, where ξ is a random vector in R^d with $P\{\xi = x\} = 1 - P\{\xi = y\} = c$. The following extension to integrable random vectors is known as *Jensen's inequality.*

Lemma 4.5 *(convex maps, Hölder, Jensen) For any integrable random vector ξ in R^d and convex function $f\colon \mathsf{R}^d \to \mathsf{R}$, we have*

$$Ef(\xi) \ge f(E\,\xi).$$

Proof: By a version of the *Hahn–Banach theorem*[8], the convexity condition (6) is equivalent to the existence for every $s \in \mathsf{R}^d$ of a *supporting* affine function $h_s(x) = ax + b$ with $f \ge h_s$ and $f(s) = h_s(s)$. Taking $s = E\,\xi$ gives

$$\begin{aligned} Ef(\xi) &\ge E\,h_s(\xi) \\ &= h_s(E\,\xi) \\ &= f(E\,\xi). \end{aligned}$$

□

The *covariance* of two random variables $\xi, \eta \in L^2$ is given by

$$\begin{aligned} \mathrm{Cov}(\xi, \eta) &= E\Big\{(\xi - E\,\xi)(\eta - E\,\eta)\Big\} \\ &= E\,\xi\eta - (E\,\xi)(E\,\eta). \end{aligned}$$

It is clearly *bi-linear*, in the sense that

$$\mathrm{Cov}\Big(\sum_{j \le m} a_j\,\xi_j,\ \sum_{k \le n} b_k\,\eta_k\Big) = \sum_{j \le m}\sum_{k \le n} a_j b_k\,\mathrm{Cov}(\xi_j, \eta_k).$$

Taking $\xi = \eta \in L^2$ gives the *variance*

$$\begin{aligned} \mathrm{Var}(\xi) &= \mathrm{Cov}(\xi, \xi) \\ &= E(\xi - E\xi)^2 \\ &= E\,\xi^2 - (E\,\xi)^2, \end{aligned}$$

and Cauchy's inequality yields

$$\Big|\,\mathrm{Cov}(\xi, \eta)\Big| \le \Big\{\,\mathrm{Var}(\xi)\,\mathrm{Var}(\eta)\Big\}^{1/2}.$$

Two random variables $\xi, \eta \in L^2$ are said to be *uncorrelated* if $\mathrm{Cov}(\xi, \eta) = 0$.

For any collection of random variables $\xi_t \in L^2$, $t \in T$, the associated *covariance function* $\rho_{s,t} = \mathrm{Cov}(\xi_s, \xi_t)$, $s, t \in T$, is *non-negative definite*, in the sense that $\sum_{ij} a_i a_j\,\rho_{t_i,t_j} \ge 0$ for any $n \in \mathsf{N}$, $t_1, \dots t_n \in T$, and $a_1, \dots, a_n \in \mathsf{R}$. This is clear if we write

[8] Any two disjoint, convex sets in R^d can be separated by a hyperplane.

$$\begin{aligned}\textstyle\sum_{i,j} a_i a_j \,\rho_{t_i,t_j} &= \textstyle\sum_{i,j} a_i a_j \operatorname{Cov}(\xi_{t_i}, \xi_{t_j})\\ &= \operatorname{Var}\Big(\textstyle\sum_i a_i \,\xi_{t_i}\Big) \geq 0.\end{aligned}$$

The events $A_t \in \mathcal{A}$, $t \in T$, are said to be *independent*, if for any distinct indices $t_1, \dots, t_n \in T$,

$$P \bigcap_{k \leq n} A_{t_k} = \prod_{k \leq n} P A_{t_k}. \tag{7}$$

More generally, we say that the families $\mathcal{C}_t \subset \mathcal{A}$, $t \in T$, are independent[9], if independence holds between the events A_t for any $A_t \in \mathcal{C}_t$, $t \in T$. Finally, the random elements ξ_t, $t \in T$, are said to be independent if independence holds between the generated σ-fields $\sigma(\xi_t)$, $t \in T$. Pairwise independence between the objects A and B, ξ and η, or $\mathcal{B}$ and $\mathcal{C}$ is often denoted by $A \perp\!\!\!\perp B$, $\xi \perp\!\!\!\perp \eta$, or $\mathcal{B} \perp\!\!\!\perp \mathcal{C}$, respectively.

The following result is often useful to extend the independence property.

Lemma 4.6 *(extension) Let $\mathcal{C}_t$, $t \in T$, be independent π-systems. Then the independence extends to the σ-fields*

$$\mathcal{F}_t = \sigma(\mathcal{C}_t), \quad t \in T.$$

Proof: We may assume that $\mathcal{C}_t \neq \emptyset$ for all t. Fix any distinct indices $t_1, \dots, t_n \in T$, and note that (7) holds for arbitrary $A_{t_k} \in \mathcal{C}_{t_k}$, $k = 1, \dots, n$. For fixed $A_{t_2}, \dots, A_{t_n}$, we introduce the class $\mathcal{D}$ of sets $A_{t_1} \in \mathcal{A}$ satisfying (7). Then $\mathcal{D}$ is a λ-system containing $\mathcal{C}_{t_1}$, and so $\mathcal{D} \supset \sigma(\mathcal{C}_{t_1}) = \mathcal{F}_{t_1}$ by Theorem 1.1. Thus, (7) holds for arbitrary $A_{t_1} \in \mathcal{F}_{t_1}$ and $A_{t_k} \in \mathcal{C}_{t_k}$, $k = 2, \dots, n$. Proceeding recursively in n steps, we obtain the desired extension to arbitrary $A_{t_k} \in \mathcal{F}_{t_k}$, $k = 1, \dots, n$. $\square$

An immediate consequence is the following basic *grouping* property. Here and below, we write

$$\mathcal{F} \vee \mathcal{G} = \sigma(\mathcal{F}, \mathcal{G}), \qquad \mathcal{F}_S = \bigvee_{t \in S} \mathcal{F}_t = \sigma\{\mathcal{F}_t;\, t \in S\}.$$

Corollary 4.7 *(grouping) Let $\mathcal{F}_t$, $t \in T$, be independent σ-fields, and let $\mathcal{T}$ be a disjoint partition of T. Then the independence extends to the σ-fields*

$$\mathcal{F}_S = \bigvee_{t \in S} \mathcal{F}_t, \quad S \in \mathcal{T}.$$

Proof: For any $S \in \mathcal{T}$, let $\mathcal{C}_S$ be the class of finite intersections of sets in $\bigcup_{t \in S} \mathcal{F}_t$. Then the classes $\mathcal{C}_S$ are independent π-systems, and the independence extends by Lemma 4.6 to the generated σ-fields $\mathcal{F}_S$. $\square$

Though independence between more than two σ-fields is clearly stronger than pairwise independence, the full independence may be reduced in various ways to the pairwise version. Given a set T, we say that a class $\mathcal{T} \subset 2^T$ is *separating*, if for any $s \neq t$ in T there exists a set $S \in \mathcal{T}$, such that exactly one of the elements s and t lies in S.

[9]not to be confused with independence *within* each class

Lemma 4.8 *(pairwise independence) Let the $\mathcal{F}_n$ or $\mathcal{F}_t$ be σ-fields on Ω indexed by* N *or* T. *Then*

(i) $\mathcal{F}_1, \mathcal{F}_2, \ldots$ *are independent iff*

$$\sigma(\mathcal{F}_1, \ldots, \mathcal{F}_n) \perp\!\!\!\perp \mathcal{F}_{n+1}, \quad n \in \mathsf{N},$$

(ii) *for any separating class* $\mathcal{T} \subset 2^T$, *the* $\mathcal{F}_t$ *are independent iff*

$$\mathcal{F}_S \perp\!\!\!\perp \mathcal{F}_{S^c}, \quad S \in \mathcal{T}.$$

Proof: The necessity of the two conditions follows from Corollary 4.7. As for the sufficiency, we consider only part (ii), the proof for (i) being similar. Under the stated condition, we need to show that, for any finite subset $S \subset T$, the σ-fields $\mathcal{F}_s$, $s \in S$, are independent. Assume the statement to be true for $|S| \le n$, where $|S|$ denote the cardinality of S. Proceeding to the case where $|S| = n+1$, we may choose $U \in \mathcal{T}$ such that $S' = S \cap U$ and $S'' = S \setminus U$ are non-empty. Since $\mathcal{F}_{S'} \perp\!\!\!\perp \mathcal{F}_{S''}$, we get for any sets $A_s \in \mathcal{F}_s$, $s \in S$,

$$P \bigcap_{s\in S} A_s = \Big(P \bigcap_{s \in S'} A_s\Big)\Big(P \bigcap_{s\in S''} A_s\Big) = \prod_{s\in S} PA_s,$$

where the last relation follows from the induction hypothesis. □

A σ-field $\mathcal{F}$ is said to be P-*trivial* if $PA = 0$ or 1 for every $A \in \mathcal{F}$. Note that a random element is a.s. a constant iff its distribution is a degenerate probability measure.

Lemma 4.9 *(triviality and degeneracy) For a σ-field $\mathcal{F}$ on Ω, these conditions are equivalent:*

(i) $\mathcal{F}$ *is* P*-trivial,*

(ii) $\mathcal{F} \perp\!\!\!\perp \mathcal{F}$,

(iii) *every* $\mathcal{F}$*-measurable random variable is a.s. a constant.*

Proof, (i) ⇔ (ii): If $\mathcal{F} \perp\!\!\!\perp \mathcal{F}$, then for any $A \in \mathcal{F}$ we have $PA = P(A \cap A) = (PA)^2$, and so $PA = 0$ or 1. Conversely, if $\mathcal{F}$ is P-trivial, then for any $A, B \in \mathcal{F}$ we have $P(A \cap B) = PA \wedge PB = PA \cdot PB$, which means that $\mathcal{F} \perp\!\!\!\perp \mathcal{F}$.

(i) ⇔ (iii): Assume (iii). Then 1_B is a.s. a constant for every $B \in \mathcal{F}$, and so $PB = 0$ or 1, proving (i). Conversely, (i) yields $F_t \equiv P\{\xi \le t\} = 0$ or 1 for every $\mathcal{F}$-measurable random variable ξ, and so $\xi = \sup\{t;\ F_t = 0\}$ a.s., proving (iii). □

We proceed with a basic relationship between independence and product measures.

Lemma 4.10 *(product measures and independence) Let $\xi_1, \ldots, \xi_n$ be random elements with distributions $\mu_1, \ldots, \mu_n$, in some measurable spaces $S_1, \ldots, S_n$. Then these conditions are equivalent:*

(i) *the* ξ_k *are independent,*

(ii) $\xi = (\xi_1, \ldots, \xi_n)$ *has distribution* $\mu_1 \otimes \cdots \otimes \mu_n$.

Proof: Assuming (i), we get for any measurable product set $B = B_1 \times \cdots \times B_n$

$$\begin{aligned} P\{\xi \in B\} &= \prod_{k \le n} P\{\xi_k \in B_k\} \\ &= \prod_{k \le n} \mu_k B_k = \bigotimes_{k \le n} \mu_k \, B. \end{aligned}$$

This extends by Theorem 1.1 to arbitrary sets in the product σ-field, proving (ii). The converse is obvious. □

Combining the last result with Fubini's theorem, we obtain a useful method of calculating expected values, in the context of independence. A more general version is given in Theorem 8.5.

Lemma 4.11 *(iterated expectation) Let* ξ, η *be independent random elements in some measurable spaces* S, T, *and consider a measurable function* $f : S \times T \to \mathsf{R}$ *with* $E\{E|f(s,\eta)|\}_{s=\xi} < \infty$. *Then*

$$Ef(\xi,\eta) = E\Big\{Ef(s,\eta)\Big\}_{s=\xi}.$$

Proof: Let μ and ν be the distributions of ξ and η, respectively. Assuming $f \ge 0$ and writing $g(s) = Ef(s,\eta)$, we get by Lemma 1.24 and Fubini's theorem

$$\begin{aligned} Ef(\xi,\eta) &= \int f(s,t)\,(\mu \otimes \nu)(ds\,dt) \\ &= \int \mu(ds) \int f(s,t)\,\nu(dt) \\ &= \int g(s)\,\mu(ds) = Eg(\xi). \end{aligned}$$

For general f, this applies to the function $|f|$, and so $E|f(\xi,\eta)| < \infty$. The desired relation then follows as before. □

In particular, we get for any independent random variables $\xi_1, \ldots, \xi_n$

$$E \prod\nolimits_k \xi_k = \prod\nolimits_k E\,\xi_k, \qquad \mathrm{Var}\Big(\sum\nolimits_k \xi_k\Big) = \sum\nolimits_k \mathrm{Var}(\xi_k),$$

whenever the expressions on the right exist.

For any random elements ξ, η in a measurable group G, the product $\xi\eta$ is again a random element in G. We give a connection between independence and the convolutions of Lemma 1.30.

Corollary 4.12 *(convolution) Let* ξ, η *be independent random elements in a measurable group* G *with distributions* μ, ν. *Then*

$$\mathcal{L}(\xi\eta) = \mu * \nu.$$

Proof: For any measurable set $B \subset G$, we get by Lemma 4.10 and the definition of convolution

$$P\{\xi\eta \in B\} = (\mu \otimes \nu)\Big\{(x,y) \in G^2;\, xy \in B\Big\}$$
$$= (\mu * \nu)B. \qquad \square$$

For any σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots$, we introduce the associated *tail σ-field*

$$\mathcal{T} = \bigcap_n \bigvee_{k>n} \mathcal{F}_k = \bigcap_n \sigma\{\mathcal{F}_k;\, k > n\}.$$

The following remarkable result shows that $\mathcal{T}$ is trivial when the $\mathcal{F}_n$ are independent. An extension appears in Corollary 9.26.

Theorem 4.13 *(Kolmogorov's $0-1$ law) Let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be σ-fields with associated tail σ-field $\mathcal{T}$. Then*

$$\mathcal{F}_1, \mathcal{F}_2, \ldots \text{ independent} \quad \Rightarrow \quad \mathcal{T} \text{ trivial.}$$

Proof: For each $n \in \mathsf{N}$, define $\mathcal{T}_n = \bigvee_{k>n} \mathcal{F}_k$, and note that $\mathcal{F}_1, \ldots, \mathcal{F}_n, \mathcal{T}_n$ are independent by Corollary 4.7. Hence, so are the σ-fields $\mathcal{F}_1, \ldots, \mathcal{F}_n, \mathcal{T}$, and then also $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{T}$. By the same theorem, we obtain $\mathcal{T}_0 \perp\!\!\!\perp \mathcal{T}$, and so $\mathcal{T} \perp\!\!\!\perp \mathcal{T}$. Thus, $\mathcal{T}$ is P-trivial by Lemma 4.9. $\square$

We consider some simple illustrations:

Corollary 4.14 *(sums and averages) Let $\xi_1, \xi_2, \ldots$ be independent random variables, and put $S_n = \xi_1 + \cdots + \xi_n$. Then*

(i) *the sequence (S_n) is a.s. convergent or a.s. divergent,*

(ii) *the sequence (S_n/n) is a.s. convergent or a.s. divergent, and its possible limit is a.s. a constant.*

Proof: Define $\mathcal{F}_n = \sigma\{\xi_n\}$, $n \in \mathsf{N}$, and note that the associated tail σ-field $\mathcal{T}$ is P-trivial by Theorem 4.13. Since the sets where (S_n) and (S_n/n) converge are $\mathcal{T}$-measurable by Lemma 1.10, the first assertions follows. The last one is clear from Lemma 4.9. $\square$

By a *finite permutation* of N we mean a bijective map $p\colon \mathsf{N} \to \mathsf{N}$, such that $p_n = n$ for all but finitely many n. For any space S, a finite permutation p of N induces a permutation T_p on S^∞, given by

$$T_p(s) = s \circ p = (s_{p_1}, s_{p_2}, \ldots), \quad s = (s_1, s_2, \ldots) \in S^\infty.$$

A set $I \subset S^\infty$ is said to be *symmetric* or *permutation invariant* if

$$T_p^{-1} I \equiv \{s \in S^\infty;\, s \circ p \in I\} = I,$$

for every finite permutation p of N. For a measurable space $(S, \mathcal{S})$, the symmetric sets $I \in \mathcal{S}^\infty$ form a σ-field $\mathcal{I} \subset \mathcal{S}^\infty$, called the *permutation invariant σ-field* in S^∞.

We may now state the second basic $0-1$ law, which applies to sequences of random elements that are independent and identically distributed, abbreviated as *i.i.d.*.

Theorem 4.15 *(Hewitt–Savage 0–1 law) Let $\xi = (\xi_k)$ be an infinite sequence of random elements in a measurable space S, and let $\mathcal{I}$ denote the permutation invariant σ-field in S^∞. Then*

$$\xi_1, \xi_2, \ldots \ \text{i.i.d.} \quad \Rightarrow \quad \xi^{-1}\mathcal{I} \ \text{trivial.}$$

Our proof is based on a simple approximation. Write

$$A \triangle B = (A \setminus B) \cup (B \setminus A),$$

and note that

$$\begin{aligned} P(A \triangle B) &= P(A^c \triangle B^c) \\ &= E|1_A - 1_B|, \quad A, B \in \mathcal{A}. \end{aligned} \tag{8}$$

Lemma 4.16 *(approximation) For any σ-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ and set $A \in \bigvee_n \mathcal{F}_n$, we may choose*

$$A_1, A_2, \ldots \in \bigcup_n \mathcal{F}_n, \qquad P(A \triangle A_n) \to 0.$$

Proof: Define $\mathcal{C} = \bigcup_n \mathcal{F}_n$, and let $\mathcal{D}$ be the class of sets $A \in \bigvee_n \mathcal{F}_n$ with the stated property. Since $\mathcal{C}$ is a π-system and $\mathcal{D}$ is a λ-system containing $\mathcal{C}$, Theorem 1.1 yields $\bigvee_n \mathcal{F}_n = \sigma(\mathcal{C}) \subset \mathcal{D}$. □

Proof of Theorem 4.15: Define $\mu = \mathcal{L}(\xi)$, put $\mathcal{F}_n = \mathcal{S}^n \times S^\infty$, and note that $\mathcal{I} \subset \mathcal{S}^\infty = \bigvee_n \mathcal{F}_n$. For any $I \in \mathcal{I}$, Lemma 4.16 yields some $B_n \in \mathcal{S}^n$, such that the corresponding cylinder sets $I_n = B_n \times S^\infty$ satisfy $\mu(I \triangle I_n) \to 0$. Put $\tilde{I}_n = S^n \times B_n \times S^\infty$. Then the symmetry of μ and I yields $\mu\tilde{I}_n = \mu I_n \to \mu I$ and $\mu(I \triangle \tilde{I}_n) = \mu(I \triangle I_n) \to 0$, and so by (8)

$$\mu\big\{I \triangle (I_n \cap \tilde{I}_n)\big\} \le \mu(I \triangle I_n) + \mu(I \triangle \tilde{I}_n) \to 0.$$

Since moreover $I_n \perp\!\!\!\perp \tilde{I}_n$ under μ, we get

$$\begin{aligned} \mu I &\leftarrow \mu(I_n \cap \tilde{I}_n) \\ &= (\mu I_n)(\mu \tilde{I}_n) \to (\mu I)^2. \end{aligned}$$

Thus, $\mu I = (\mu I)^2$, and so $P \circ \xi^{-1} I = \mu I = 0$ or 1. □

Again we list some easy consequences. Here a random variable ξ is said to be *symmetric* if $\xi \stackrel{d}{=} -\xi$.

Corollary 4.17 *(random walk) Let $\xi_1, \xi_2, \ldots$ be i.i.d., non-degenerate random variables, and put $S_n = \xi_1 + \cdots + \xi_n$. Then*

(i) $P\{S_n \in B \text{ i.o.}\} = 0$ *or* 1 *for every* $B \in \mathcal{B}$,

(ii) $\limsup_n S_n = \infty$ *a.s. or* $= -\infty$ *a.s.*,

(iii) $\limsup_n (\pm S_n) = \infty$ *a.s. when the* ξ_n *are symmetric.*

Proof: (i) This holds by Theorem 4.15, since for any finite permutation p of N we have $x_{p_1} + \cdots + x_{p_n} = x_1 + \cdots + x_n$ for all but finitely many n.

(ii) By Theorem 4.15 and Lemma 4.9, we have $\limsup_n S_n = c$ a.s. for some constant $c \in \bar{\mathsf{R}} = [-\infty, \infty]$, and so a.s.

$$\begin{aligned} c &= \limsup_{n\to\infty} S_{n+1} \\ &= \limsup_{n\to\infty} (S_{n+1} - \xi_1) + \xi_1 \\ &= c + \xi_1. \end{aligned}$$

Since $|c| < \infty$ would imply $\xi_1 = 0$ a.s., contradicting the non-degeneracy of ξ_1, we have $|c| = \infty$.

(iii) Writing

$$\begin{aligned} c &= \limsup_{n\to\infty} S_n \\ &\geq \liminf_{n\to\infty} S_n \\ &= -\limsup_{n\to\infty} (-S_n) = -c, \end{aligned}$$

we get $-c \leq c \in \{\pm\infty\}$, which implies $c = \infty$. □

From a suitable 0−1 law it is often easy to see that a given event has probability 0 or 1. To distinguish between the two cases is often much harder. The following classical result, known as the *Borel–Cantelli lemma*, is sometimes helpful, especially when the events are independent. A more general version appears in Corollary 9.21.

Theorem 4.18 *(Borel, Cantelli) For any $A_1, A_2, \ldots \in \mathcal{A}$,*

$$\textstyle\sum_n PA_n < \infty \quad \Rightarrow \quad P\{A_n \text{ i.o.}\} = 0,$$

and equivalence holds when the A_n are independent.

The first assertion was proved earlier as an application of Fatou's lemma. Using expected values yields a more transparent argument.

Proof: If $\sum_n PA_n < \infty$, we get by monotone convergence

$$\begin{aligned} E \textstyle\sum_n 1_{A_n} &= \textstyle\sum_n E1_{A_n} \\ &= \textstyle\sum_n PA_n < \infty. \end{aligned}$$

Thus, $\sum_n 1_{A_n} < \infty$ a.s., which means that $P\{A_n \text{ i.o.}\} = 0$.

Now let the A_n be independent with $\sum_n PA_n = \infty$. Since $1 - x \leq e^{-x}$ for all x, we get

$$\begin{aligned} P \bigcup_{k\geq n} A_k &= 1 - P \bigcap_{k\geq n} A_k^c \\ &= 1 - \prod_{k\geq n} PA_k^c \\ &= 1 - \prod_{k\geq n} (1 - PA_k) \\ &\geq 1 - \prod_{k\geq n} \exp(-PA_k) \\ &= 1 - \exp\Big(-\sum_{k\geq n} PA_k\Big) = 1. \end{aligned}$$

Hence, as $n \to \infty$,

$$\begin{aligned} 1 &= P \bigcup_{k \ge n} A_k \downarrow P \bigcap_n \bigcup_{k \ge n} A_k \\ &= P\{A_n \text{ i.o.}\}, \end{aligned}$$

and so the probability on the right equals 1. □

For many purposes, it suffices to use the *Lebesgue unit interval* $([0,1], \mathcal{B}[0,1], \lambda)$ as our basic probability space. In particular, the following result ensures the existence on $[0,1]$ of some independent random variables $\xi_1, \xi_2, \ldots$ with arbitrarily prescribed distributions. The present statement is only preliminary. Thus, we will remove the independence assumption in Theorem 8.21, prove an extension to arbitrary index sets in Theorem 8.23, and eliminate the restriction on the spaces in Theorem 8.24.

Theorem 4.19 *(existence, Borel) For any probability measures $\mu_1, \mu_2, \ldots$ on the Borel spaces $S_1, S_2, \ldots$, there exist some independent random elements $\xi_1, \xi_2, \ldots$ on $([0,1], \lambda)$ with distributions $\mu_1, \mu_2, \ldots$.*

As a consequence, there exists a probability measure μ on $S_1 \times S_2 \times \cdots$ satisfying

$$\mu \circ (\pi_1, \ldots, \pi_n)^{-1} = \mu_1 \otimes \cdots \otimes \mu_n, \quad n \in \mathbb{N}.$$

For the proof, we begin with two special cases of independent interest.

By a *Bernoulli sequence* with *rate* p we mean a sequence of i.i.d. random variables $\xi_1, \xi_2, \ldots$ with $P\{\xi_n = 1\} = p = 1 - P\{\xi_n = 0\}$. We further say that a random variable ξ is *uniformly distributed* on $[0,1]$, written as $U(0,1)$, if its distribution $\mathcal{L}(\xi)$ equals Lebesgue measure λ on $[0,1]$. Every number $x \in [0,1]$ has a *binary expansion* $r_1, r_2, \ldots \in \{0,1\}$ satisfying $x = \sum_n r_n 2^{-n}$, where for uniqueness we require that $\sum_n r_n = \infty$ when $x > 0$. We give a simple construction of Bernoulli sequences on the Lebesgue unit interval.

Lemma 4.20 *(Bernoulli sequence) Let ξ be a random variable in $[0,1]$ with binary expansion $\eta_1, \eta_2, \ldots$. Then these conditions are equivalent:*
(i) *ξ is $U(0,1)$,*
(ii) *the η_n form a Bernoulli sequence with rate $\frac{1}{2}$.*

Proof, (i) $\Rightarrow$ (ii): If ξ is $U(0,1)$, then $P \bigcap_{j \le n} \{\eta_j = k_j\} = 2^{-n}$ for all $k_1, \ldots, k_n \in \{0,1\}$. Summing over $k_1, \ldots, k_{n-1}$ gives $P\{\eta_n = k\} = \frac{1}{2}$ for $k = 0$ or 1. A similar calculation yields the asserted independence.

(ii) $\Rightarrow$ (i): Assume (ii). Letting ξ' be $U(0,1)$ with binary expansion $\eta'_1, \eta'_2, \ldots$, we get $(\eta_n) \stackrel{d}{=} (\eta'_n)$, and so

$$\begin{aligned} \xi &= \textstyle\sum_n \eta_n 2^{-n} \\ &\stackrel{d}{=} \textstyle\sum_n \eta'_n 2^{-n} = \xi'. \end{aligned}$$ □

Next we show how a single $U(0,1)$ random variable can be used to generate a whole sequence.

Lemma 4.21 *(replication) There exist some measurable functions* $f_1, f_2, \dots$ *on* $[0,1]$ *such that* (i) $\Rightarrow$ (ii)*, where*

(i) ξ *is* $U(0,1)$*,*
(ii) $\xi_n = f_n(\xi)$*,* $n \in \mathsf{N}$*, are i.i.d.* $U(0,1)$*.*

Proof: Let $b_1(x), b_2(x), \dots$ be the binary expansion of $x \in [0,1]$, and note that the b_k are measurable. Rearranging the b_k into a two-dimensional array h_{nj}, $n, j \in \mathsf{N}$, we define

$$f_n(x) = \textstyle\sum_j 2^{-j}\, h_{nj}(x), \quad x \in [0,1],\ n \in \mathsf{N}.$$

By Lemma 4.20, the random variables $b_k(\xi)$ form a Bernoulli sequence with rate $\frac{1}{2}$, and the same result shows that the variables $\xi_n = f_n(\xi)$ are $U(0,1)$. They are further independent by Corollary 4.7. □

We finally need to construct a random element with specified distribution from a given randomization variable. The result is stated in an extended form for kernels, to meet our needs in Chapters 8, 11, 22, and 27–28.

Lemma 4.22 *(kernel representation) Consider a probability kernel* $\mu\colon S \to T$ *with* T *Borel, and let* ξ *be* $U(0,1)$*. Then there exists a measurable function* $f\colon S \times [0,1] \to T$*, such that*

$$\mathcal{L}\Big\{f(s,\xi)\Big\} = \mu(s,\cdot), \quad s \in S.$$

Proof: We may choose $T \in \mathcal{B}_{[0,1]}$, from where we can easily reduce to the case of $T = [0,1]$. Define

$$f(s,t) = \sup\Big\{x \in [0,1];\ \mu(s,[0,x]) < t\Big\}, \quad s \in S,\ t \in [0,1], \tag{9}$$

and note that f is product measurable on $S \times [0,1]$, since the set $\{(s,t);\ \mu(s,[0,x]) < t\}$ is measurable for each x by Lemma 1.13, and the supremum in (9) can be restricted to rational x. If ξ is $U(0,1)$, we get for $x \in [0,1]$

$$\begin{aligned} P\Big\{f(s,\xi) \le x\Big\} &= P\Big\{\xi \le \mu(s,[0,x])\Big\} \\ &= \mu(s,[0,x]), \end{aligned}$$

and so by Lemma 4.3 $f(s,\xi)$ has distribution $\mu(s,\cdot)$. □

Proof of Theorem 4.19: By Lemma 4.22 there exist some measurable functions $f_n : [0,1] \to S_n$ with $\lambda \circ f_n^{-1} = \mu_n$. Letting ξ be the identity map on $[0,1]$ and choosing $\xi_1, \xi_2, \dots$ as in Lemma 4.21, we note that the functions $\eta_n = f_n(\xi_n)$, $n \in \mathsf{N}$, have the desired joint distribution. □

We turn to some regularizations and path properties of random processes. Say that two processes X, Y on a common index set T are *versions* of one another if $X_t = Y_t$ a.s. for every $t \in T$. When $T = \mathsf{R}^d$ or R_+, two continuous or right-continuous versions X, Y of a given process are clearly *indistinguishable*, in the sense that $X \equiv Y$ a.s., so that the entire paths agree outside a fixed null set.

For any mapping f between two metric spaces (S, ρ) and (S', ρ'), we define the *modulus of continuity* $w_f = w(f, \cdot)$ by

$$w_f(r) = \sup\Big\{\rho'(f_s, f_t);\ s, t \in S,\ \rho(s,t) \le r\Big\}, \quad r > 0,$$

so that f is uniformly continuous iff $w_f(r) \to 0$ as $r \to 0$. Say that f is *Hölder continuous of order* p, if[10] $w_f(r) \lesssim r^p$ as $r \to 0$. The stated property is said to hold *locally* if it is valid on every bounded set.

A simple moment condition ensures the existence of a Hölder-continuous version of a given process on R^d. Important applications are given in Theorems 14.5, 29.4, and 32.3, and a related tightness criterion appears in Corollary 23.7.

Theorem 4.23 *(moments and Hölder continuity, Kolmogorov, Loève, Chentsov) Let X be a process on R^d with values in a complete metric space (S, ρ), such that*

$$E\Big\{\rho(X_s, X_t)\Big\}^a \lesssim |s - t|^{d+b}, \quad s, t \in \mathsf{R}^d, \tag{10}$$

for some constants $a, b > 0$. Then a version of X is locally Hölder continuous of order p, for every $p \in (0, b/a)$.

Proof: It is clearly enough to consider the restriction of X to $[0,1]^d$. Define

$$D_n = \Big\{2^{-n}(k_1, \ldots, k_d);\ k_1, \ldots, k_n \in \{1, \ldots, 2^n\}\Big\}, \quad n \in \mathsf{N},$$

and put

$$\xi_n = \max\Big\{\rho(X_s, X_t);\ s, t \in D_n,\ |s - t| = 2^{-n}\Big\}, \quad n \in \mathsf{N}.$$

Since

$$\Big|\Big\{(s,t) \in D_n^2;\ |s-t| = 2^{-n}\Big\}\Big| \le d\,2^{dn}, \quad n \in \mathsf{N},$$

we get by (10) for any $p \in (0, b/a)$

$$\begin{aligned} E \textstyle\sum_n (2^{pn}\,\xi_n)^a &= \textstyle\sum_n 2^{apn}\, E\,\xi_n^a \\ &\lesssim \textstyle\sum_n 2^{apn}\, 2^{dn}\, (2^{-n})^{d+b} \\ &= \textstyle\sum_n 2^{(ap-b)\,n} < \infty. \end{aligned}$$

The sum on the left is then a.s. convergent, and so $\xi_n \lesssim 2^{-pn}$ a.s. Now any points $s, t \in \bigcup_n D_n$ with $|s - t| \le 2^{-m}$ can be connected by a piecewise linear path, which for every $n \ge m$ involves at most $2d$ steps between nearest neighbors in D_n. Thus, for any $r \in [2^{-m-1}, 2^{-m}]$,

$$\begin{aligned} &\sup\Big\{\rho(X_s, X_t);\ s, t \in \textstyle\bigcup_n D_n,\ |s - t| \le r\Big\} \\ &\qquad \lesssim \sum_{n \ge m} \xi_n \lesssim \sum_{n \ge m} 2^{-pn} \\ &\qquad \lesssim 2^{-pm} \lesssim r^p, \end{aligned}$$

[10]For functions $f, g > 0$, we mean by $f \lesssim g$ that $f \le c\,g$ for some constant $c < \infty$.

showing that X is a.s. Hölder continuous on $\bigcup_n D_n$ of order p.

In particular, X agrees a.s. on $\bigcup_n D_n$ with a continuous process Y on $[0,1]^d$, and we note that the Hölder continuity of Y on $\bigcup_n D_n$ extends with the same order p to the entire cube $[0,1]^d$. To show that Y is a version of X, fix any $t \in [0,1]^d$, and choose $t_1, t_2, \ldots \in \bigcup_n D_n$ with $t_n \to t$. Then $X_{t_n} = Y_{t_n}$ a.s. for each n. Since also $X_{t_n} \xrightarrow{P} X_t$ by (10) and $Y_{t_n} \to Y_t$ a.s. by continuity, we get $X_t = Y_t$ a.s. $\square$

Path regularity can sometimes be established by comparison with a regular process:

Lemma 4.24 *(transfer of regularity) Let $X \overset{d}{=} Y$ be random processes on T with values in a separable metric space S, where the paths of Y belong to a set $U \subset S^T$ that is Borel for the σ-field $\mathcal{U} = (\mathcal{B}_S)^T \cap U$. Then X has a version with paths in U.*

Proof: For clarity we may write $\tilde{Y}$ for the path of Y, regarded as a random element in U. Then $\tilde{Y}$ is Y-measurable, and Lemma 1.14 yields a measurable mapping $f : S^T \to U$ with $\tilde{Y} = f(Y)$ a.s. Define $\tilde{X} = f(X)$, and note that $(\tilde{X}, X) \overset{d}{=} (\tilde{Y}, Y)$. Since the diagonal in S^2 is measurable, we get in particular

$$P\{\tilde{X}_t = X_t\} = P\{\tilde{Y}_t = Y_t\} = 1, \quad t \in T. \qquad \square$$

We conclude with a characterization of distribution functions on R^d, required in Chapter 6. For any vectors $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$, write $x \le y$ for the component-wise inequality $x_k \le y_k$, $k = 1, \ldots, d$, and similarly for $x < y$. In particular, the distribution function F of a probability measure μ on R^d is given by $F(x) = \mu\{y;\, y \le x\}$. Similarly, let $x \vee y$ denote the componentwise maximum. Put $\mathbf{1} = (1, \ldots, 1)$ and $\infty = (\infty, \ldots, \infty)$.

For any rectangular box

$$\begin{aligned} (x, y] &= \{u;\, x < u \le y\} \\ &= (x_1, y_1] \times \cdots \times (x_d, y_d], \end{aligned}$$

we note that $\mu(x, y] = \sum_u s(u) F(u)$, where $s(u) = (-1)^p$ with $p = \sum_k 1\{u_k = y_k\}$, and the summation extends over all vertices u of $(x, y]$. Writing $F(x, y]$ for the stated sum, we say that F has *non-negative increments* if $F(x, y] \ge 0$ for all pairs $x < y$. We further say that F is *right-continuous* if $F(x_n) \to F(x)$ as $x_n \downarrow x$, and *proper* if $F(x) \to 1$ or 0 as $\min_k x_k \to \pm\infty$, respectively.

We show that any function with the stated properties determines a probability measure on R^d.

Theorem 4.25 *(distribution functions) For functions $F : \mathsf{R}^d \to [0,1]$, these conditions are equivalent:*

(i) *F is the distribution function of a probability measure μ on R^d,*

(ii) *F is right-continuous and proper with non-negative increments.*

Proof: For any F as in (ii), the set function $F(x,y]$ is clearly finitely additive. Since F is proper, we have also $F(x,y] \to 1$ as $x \to -\infty$ and $y \to \infty$, i.e., as $(x,y] \uparrow (-\infty, \infty) = \mathsf{R}^d$. Hence, for every $n \in \mathsf{N}$ there exists a probability measure μ_n on $(2^{-n}\mathsf{Z})^d$ with $\mathsf{Z} = \{\dots, -1, 0, 1, \dots\}$, such that

$$\mu_n\{2^{-n}k\} = F\Big(2^{-n}(k-1),\, 2^{-n}k\Big], \quad k \in \mathsf{Z}^d,\ n \in \mathsf{N}.$$

The finite additivity of $F(x,y]$ yields

$$\mu_m\Big\{2^{-m}(k-1,k]\Big\} = \mu_n\Big\{2^{-m}(k-1,k]\Big\}, \quad k \in \mathsf{Z}^d,\ m < n \text{ in } \mathsf{N}. \tag{11}$$

By (11) we can split the Lebesgue unit interval $([0,1], \mathcal{B}[0,1], \lambda)$ recursively to construct some random vectors $\xi_1, \xi_2, \dots$ with distributions $\mu_1, \mu_2, \dots$, satisfying

$$\xi_m - 2^{-m} < \xi_n \le \xi_m, \quad m < n.$$

In particular, $\xi_1 \ge \xi_2 \ge \cdots \ge \xi_1 - 1$, and so ξ_n converges pointwise to some random vector ξ. Define $\mu = \lambda \circ \xi^{-1}$.

To see that μ has distribution function F, we conclude from the properness of F that

$$\begin{aligned}\lambda\Big\{\xi_n \le 2^{-n}k\Big\} &= \mu_n(-\infty,\, 2^{-n}k] \\ &= F(2^{-n}k), \quad k \in \mathsf{Z}^d,\ n \in \mathsf{N}.\end{aligned}$$

Since also $\xi_n \downarrow \xi$ a.s., Fatou's lemma yields for dyadic $x \in \mathsf{R}^d$

$$\begin{aligned}\lambda\{\xi < x\} &= \lambda\{\xi_n < x \text{ ult.}\} \\ &\le \liminf_{n\to\infty} \lambda\{\xi_n < x\} \le F(x) \\ &= \limsup_{n\to\infty} \lambda\{\xi_n \le x\} \\ &\le \lambda\{\xi_n \le x \text{ i.o.}\} \\ &\le \lambda\{\xi \le x\},\end{aligned}$$

and so

$$\begin{aligned}F(x) &\le \lambda\{\xi \le x\} \\ &\le F(x + 2^{-n}\mathbf{1}), \quad n \in \mathsf{N}.\end{aligned}$$

Letting $n \to \infty$ and using the right-continuity of F, we get $\lambda\{\xi \le x\} = F(x)$, which extends to any $x \in \mathsf{R}^d$, by the right-continuity of both sides. $\square$

We also need a version for unbounded measures, extending the one-dimensional Theorem 2.14:

Corollary 4.26 *(unbounded measures) For any right-continuous function F on R^d with non-negative increments, there exists a measure μ on R^d with*

$$\mu(x,y] = F(x,y], \quad x \le y \textit{ in } \mathsf{R}^d.$$

Proof: For any $a \in \mathbb{R}^d$, we may apply Theorem 4.25 to suitably normalized versions of the function $F_a(x) = F(a, a \vee x)$ to obtain a measure μ_a on $[a, \infty)$ with $\mu_a(a, x] = F(a, x]$ for all $x > a$. Then clearly $\mu_a = \mu_b$ on $(a \vee b, \infty)$ for any a, b, and so the set function $\mu = \sup_a \mu_a$ is a measure with the required property. $\square$

Exercises

1. Give an example of two processes X, Y with different distributions, such that $X_t \stackrel{d}{=} Y_t$ for all t.

2. Let X, Y be $\{0,1\}$-valued processes on an index set T. Show that $X \stackrel{d}{=} Y$ iff $P\{X_{t_1} + \cdots + X_{t_n} > 0\} = P\{Y_{t_1} + \cdots + Y_{t_n} > 0\}$ for all $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in T$.

3. Let F be a right-continuous function of bounded variation with $F(-\infty) = 0$. For any random variable ξ, show that $EF(\xi) = \int P\{\xi \geq t\}\, F(dt)$. (*Hint:* First let F be the distribution function of a random variable $\eta \perp\!\!\!\perp \xi$, and use Lemma 4.11.)

4. Given a random variable $\xi \in L^1$ and a strictly convex function f on $\mathbb{R}$, show that $Ef(\xi) = f(E\xi)$ iff $\xi = E\xi$ a.s.

5. Let $\xi = \sum_j a_j\, \xi_j$ and $\eta = \sum_j b_j\, \eta_j$, where the sums converge in L^2. Show that $\mathrm{Cov}(\xi, \eta) = \sum_{ij} a_i b_j\, \mathrm{Cov}(\xi_i, \eta_j)$, where the double series is absolutely convergent.

6. Let the σ-fields $\mathcal{F}_{t,n}$, $t \in T$, $n \in \mathbb{N}$, be non-decreasing in n for fixed t and independent in t for fixed n. Show that the independence extends to the σ-fields $\mathcal{F}_t = \bigvee_n \mathcal{F}_{t,n}$.

7. For every $t \in T$, let $\xi^t, \xi_1^t, \xi_2^t, \ldots$ be random elements in a metric space S_t with $\xi_n^t \to \xi^t$ a.s., such that the ξ_n^t are independent in t for fixed $n \in \mathbb{N}$. Show that the independence extends to the limits ξ^t. (*Hint:* First show that $E \prod_{t \in S} f_t(\xi^t) = \prod_{t \in S} E f_t(\xi^t)$ for any bounded, continuous functions f_t on S_t, and for finite subsets $S \subset T$.)

8. Give an example of three events that are pairwise independent but not independent.

9. Give an example of two random variables that are uncorrelated but not independent.

10. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random elements with distribution μ in a measurable space $(S, \mathcal{S})$. Fix a set $A \in \mathcal{S}$ with $\mu A > 0$, and put $\tau = \inf\{k;\, \xi_k \in A\}$. Show that ξ_τ has distribution $\mu(\cdot\,|A) = \mu(\cdot \cap A)/\mu A$.

11. Let $\xi_1, \xi_2, \ldots$ be independent random variables with values in $[0,1]$. Show that $E \prod_n \xi_n = \prod_n E\xi_n$. In particular, $P \bigcap_n A_n = \prod_n PA_n$ for any independent events $A_1, A_2, \ldots$.

12. For any random variables $\xi_1, \xi_2, \ldots$, prove the existence of some constants $c_1, c_2, \ldots > 0$, such that the series $\sum_n c_n\, \xi_n$ converges a.s.

13. Let $\xi_1, \xi_2, \ldots$ be random variables with $\xi_n \to 0$ a.s. Prove the existence of a measurable function $f \geq 0$ with $f > 0$ outside 0, such that $\sum_n f(\xi_n) < \infty$ a.s. Show that the conclusion fails if we assume only L^1-convergence.

14. Give an example of events $A_1, A_2, \ldots$, such that $P\{A_n \text{ i.o.}\} = 0$ while $\sum_n PA_n = \infty$.

15. Extend Lemma 4.20 to a correspondence between $U(0,1)$ random variables ϑ and Bernoulli sequences $\xi_1, \xi_2, \ldots$ with rate $p \in (0,1)$.

16. Give an elementary proof of Theorem 4.25 for $d = 1$. (*Hint:* Define $\xi = F^{-1}(\vartheta)$ where ϑ is $U(0,1)$, and note that ξ has distribution function F.)

17. Let $\xi_1, \xi_2, \ldots$ be random variables with $P\{\xi_n \neq 0 \text{ i.o.}\} = 1$. Prove the existence of some constants $c_n \in \mathsf{R}$, such that $P\{|c_n \xi_n| > 1 \text{ i.o.}\} = 1$. (*Hint:* Note that $P\left\{\sum_{k \le n} |\xi_k| > 0\right\} \to 1$.)

Chapter 5

Random Sequences, Series, and Averages

Moments and tails, convergence a.s. and in probability, sub-sequence criterion, continuity and completeness, convergence in distribution, tightness and uniform integrability, convergence of means, L^p-convergence, random series and averages, positive and symmetric terms, variance criteria, three-series criterion, strong laws of large numbers, empirical distributions, portmanteau theorem, continuous mapping and approximation, Skorohod coupling, representation and measurability of limits

Here our first aim is to introduce and compare the basic modes of convergence of random quantities. For random elements ξ and $\xi_1, \xi_2, \ldots$ in a metric or topological space S, the most commonly used notions are those of almost sure convergence ($\xi_n \to \xi$ a.s.) and convergence in probability ($\xi_n \xrightarrow{P} \xi$), corresponding to the general notions of convergence a.e. and in measure, respectively. When $S = \mathsf{R}$ we also have the concept of L^p-convergence, familiar from Chapter 1. Those three notions are used throughout this book. For a special purpose in Chapter 10, we also need the notion of weak L^1-convergence.

Our second major theme is to study the very different notion of convergence in distribution ($\xi_n \xrightarrow{d} \xi$), defined by the condition $Ef(\xi_n) \to Ef(\xi)$ for all bounded, continuous functions f on S. This is clearly equivalent to weak convergence of the associated distributions $\mu_n = \mathcal{L}(\xi_n)$ and $\mu = \mathcal{L}(\xi)$, written as $\mu_n \xrightarrow{w} \mu$ and defined by the condition $\mu_n f \to \mu f$ for every f as above. In this chapter we will establish only the most basic results of weak convergence theory, such as the 'portmanteau' theorem, the continuous mapping and approximation theorems, and the Skorohod coupling. Our development of the general theory continues in Chapters 6 and 23, and further distributional limit theorems of various kind appear throughout the remainder of the book.

Our third main theme is to characterize the convergence of series $\sum_k \xi_k$ and averages $n^{-c} \sum_{k \le n} \xi_k$, where $\xi_1, \xi_2, \ldots$ are independent random variables and c is a positive constant. The two problems are related by the elementary Kronecker lemma, and the main results are the basic three-series criterion and the strong law of large numbers. The former result is extended in Chapter 9 to the powerful martingale convergence theorem, whereas extensions and refinements of the latter result are proved in Chapters 22 and 25. The mentioned theorems are further related to certain weak convergence results presented in Chapters 6–7.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_6

Before embarking on our systematic study of the various notions of convergence, we consider a couple of elementary but useful inequalities.

Lemma 5.1 *(moments and tails, Bienaymé, Chebyshev, Paley & Zygmund) For any random variable $\xi \geq 0$ with $0 < E\xi < \infty$, we have*

$$(1-r)_+^2 \frac{(E\xi)^2}{E\xi^2} \leq P\Big\{\xi > rE\xi\Big\} \leq \frac{1}{r}, \quad r > 0.$$

The upper bound is often referred to as the *Chebyshev* or *Markov inequality*. When $E\xi^2 < \infty$, we get in particular the classical estimate

$$P\Big\{|\xi - E\xi| > \varepsilon\Big\} \leq \varepsilon^{-2}\operatorname{Var}(\xi), \quad \varepsilon > 0.$$

Proof: We may clearly assume that $E\xi = 1$. The upper bound then follows as we take expectations in the inequality $r1\{\xi > r\} \leq \xi$. To get the lower bound, we note that for any $r, t > 0$,

$$\begin{aligned} t^2 1\{\xi > r\} &\geq (\xi - r)(2t + r - \xi) \\ &= 2\xi(r+t) - r(2t+r) - \xi^2. \end{aligned}$$

Taking expected values, we get for $r \in (0,1)$

$$\begin{aligned} t^2 P\{\xi > r\} &\geq 2(r+t) - r(2t+r) - E\xi^2 \\ &\geq 2t(1-r) - E\xi^2. \end{aligned}$$

Now choose $t = E\xi^2/(1-r)$. □

For any random elements ξ and $\xi_1, \xi_2, \ldots$ in a separable metric space (S, ρ), we say that ξ_n *converges in probability to* ξ and write $\xi_n \xrightarrow{P} \xi$ if

$$\lim_{n\to\infty} P\Big\{\rho(\xi_n, \xi) > \varepsilon\Big\} = 0, \quad \varepsilon > 0.$$

Lemma 5.2 *(convergence in probability) For any random elements $\xi, \xi_1, \xi_2, \ldots$ in a separable metric space (S, ρ), these conditions are equivalent:*

(i) $\xi_n \xrightarrow{P} \xi$,

(ii) $E\Big\{\rho(\xi_n, \xi) \wedge 1\Big\} \to 0$,

(iii) *for any sub-sequence $N' \subset \mathsf{N}$, we have $\xi_n \to \xi$ a.s. along a further sub-sequence $N'' \subset N'$.*

In particular, $\xi_n \to \xi$ a.s. implies $\xi_n \xrightarrow{P} \xi$, and the notion of convergence in probability depends only on the topology, regardless of the metrization ρ.

Proof: The implication (i) ⇒ (ii) is obvious, and the converse holds by Chebyshev's inequality. Now assume (ii), and fix an arbitrary sub-sequence $N' \subset \mathsf{N}$. We may then choose a further sub-sequence $N'' \subset N'$ such that

$$E\sum_{n\in N''}\Big\{\rho(\xi_n, \xi) \wedge 1\Big\} = \sum_{n\in N''} E\Big\{\rho(\xi_n, \xi) \wedge 1\Big\} < \infty,$$

where the equality holds by monotone convergence. The series on the left then converges a.s., which implies $\xi_n \to \xi$ a.s. along N'', proving (iii).

Now assume (iii). If (ii) fails, there exists an $\varepsilon > 0$ such that $E\{\rho(\xi_n, \xi) \wedge 1\} > \varepsilon$ along a sub-sequence $N' \subset \mathsf{N}$. Then (iii) yields $\xi_n \to \xi$ a.s. along a further sub-sequence $N'' \subset N'$, and so by dominated convergence $E\{\rho(\xi_n, \xi) \wedge 1\} \to 0$ along N'', a contradiction proving (ii). □

For a first application, we show that convergence in probability is preserved by continuous mappings.

Lemma 5.3 *(continuous mapping) For any separable metric spaces S, T, let $\xi, \xi_1, \xi_2, \ldots$ be random elements in S, and let the mapping $f : S \to T$ be measurable and a.s. continuous at ξ. Then*

$$\xi_n \xrightarrow{P} \xi \quad \Rightarrow \quad f(\xi_n) \xrightarrow{P} f(\xi).$$

Proof: Fix any sub-sequence $N' \subset \mathsf{N}$. By Lemma 5.2 we have $\xi_n \to \xi$ a.s. along a further sub-sequence $N'' \subset N'$, and so by continuity $f(\xi_n) \to f(\xi)$ a.s. along N''. Hence, $f(\xi_n) \xrightarrow{P} f(\xi)$ by Lemma 5.2. □

For any separable metric spaces (S_k, ρ_k), we may introduce the product space $S = S_1 \times S_2 \times \cdots$ endowed with the product topology, admitting the metrization

$$\rho(x, y) = \textstyle\sum_k 2^{-k} \Big\{ \rho_k(x_k, y_k) \wedge 1 \Big\}, \quad x, y \in S. \tag{1}$$

Since $\mathcal{B}_S = \bigotimes_k \mathcal{B}_{S_k}$ by Lemma 1.2, a random element in S is simply a sequence of random elements in S_k, $k \in \mathsf{N}$.

Lemma 5.4 *(random sequences) Let $\xi = (\xi_1, \xi_2, \ldots)$ and $\xi^n = (\xi_1^n, \xi_2^n, \ldots)$, $n \in \mathsf{N}$, be random elements in $S_1 \times S_2 \times \cdots$, for some separable metric spaces $S_1, S_2, \ldots$. Then*

$$\xi^n \xrightarrow{P} \xi \quad \Leftrightarrow \quad \xi_k^n \xrightarrow{P} \xi_k \;\; in \;\; S_k, \;\; k \in \mathsf{N}.$$

Proof: With ρ as in (1), we get for any $n \in \mathsf{N}$

$$\begin{aligned} E\Big\{ \rho(\xi^n, \xi) \wedge 1 \Big\} &= E\,\rho(\xi^n, \xi) \\ &= \textstyle\sum_k 2^{-k} E\Big\{ \rho_k(\xi_k^n, \xi_k) \wedge 1 \Big\}. \end{aligned}$$

Thus, by dominated convergence,

$$E\Big\{ \rho(\xi^n, \xi) \wedge 1 \Big\} \to 0 \quad \Leftrightarrow \quad E\Big\{ \rho_k(\xi_k^n, \xi_k) \wedge 1 \Big\} \to 0, \quad k \in \mathsf{N}. \qquad \square$$

Combining the last two lemmas, we may show how convergence in probability is preserved by the basic arithmetic operations.

Corollary 5.5 *(elementary operations) For any random variables $\xi_n \xrightarrow{P} \xi$ and $\eta_n \xrightarrow{P} \eta$, we have*

(i) $a\,\xi_n + b\,\eta_n \xrightarrow{P} a\,\xi + b\,\eta,\ \ a, b \in \mathsf{R}$,

(ii) $\xi_n\eta_n \xrightarrow{P} \xi\eta$,

(iii) $\xi_n/\eta_n \xrightarrow{P} \xi/\eta$ *when* $\eta, \eta_1, \eta_2, \ldots \neq 0$ *a.s.*

Proof: By Lemma 5.4 we have $(\xi_n, \eta_n) \xrightarrow{P} (\xi, \eta)$ in R^2, and so (i) and (ii) follow by Lemma 5.3. To prove (iii), we may apply Lemma 5.3 to the function $f\colon (x, y) \mapsto (x/y)1\{y \neq 0\}$, which is clearly a.s. continuous at (ξ, η). □

We turn to some associated completeness properties. For any random elements $\xi_1, \xi_2, \ldots$ in a separable metric space (S, ρ), we say that (ξ_n) is *Cauchy (convergent) in probability* if $\rho(\xi_m, \xi_n) \xrightarrow{P} 0$ as $m, n \to \infty$, in the sense that $E\{\rho(\xi_m, \xi_n) \wedge 1\} \to 0$.

Lemma 5.6 *(completeness) Let $\xi_1, \xi_2, \ldots$ be random elements in a separable, complete metric space (S, ρ). Then these conditions are equivalent:*

(i) *(ξ_n) is Cauchy in probability,*

(ii) *$\xi_n \xrightarrow{P} \xi$ for a random element ξ in S.*

Similar results hold for a.s. convergence.

Proof: The a.s. case is immediate from Lemma 1.11. Assuming (ii), we get

$$E\Big\{\rho(\xi_m, \xi_n) \wedge 1\Big\} \leq E\Big\{\rho(\xi_m, \xi) \wedge 1\Big\} + E\Big\{\rho(\xi_n, \xi) \wedge 1\Big\} \to 0,$$

proving (i).

Now assume (i), and define

$$n_k = \inf\Big\{n \geq k;\ \sup_{m \geq n} E\{\rho(\xi_m, \xi_n) \wedge 1\} \leq 2^{-k}\Big\}, \quad k \in \mathsf{N}.$$

The n_k are finite and satisfy

$$E\,\textstyle\sum_k \Big\{\,\rho(\xi_{n_k}, \xi_{n_{k+1}}) \wedge 1\Big\} \leq \sum_k 2^{-k} < \infty,$$

and so $\sum_k \rho(\xi_{n_k}, \xi_{n_{k+1}}) < \infty$ a.s. The sequence (ξ_{n_k}) is then a.s. Cauchy and converges a.s. toward a measurable limit ξ. To prove (ii), write

$$E\Big\{\rho(\xi_m, \xi) \wedge 1\Big\} \leq E\Big\{\rho(\xi_m, \xi_{n_k}) \wedge 1\Big\} + E\Big\{\rho(\xi_{n_k}, \xi) \wedge 1\Big\},$$

and note that the right-hand side tends to zero as $m, k \to \infty$, by the Cauchy convergence of (ξ_n) and dominated convergence. □

For any probability measures μ and $\mu_1, \mu_2, \ldots$ on a metric space (S, ρ) with Borel σ-field $\mathcal{S}$, we say that μ_n *converges weakly to* μ and write $\mu_n \xrightarrow{w} \mu$, if $\mu_n f \to \mu f$ for every $f \in \hat{C}_S$, defined as the class of bounded, continuous functions $f\colon S \to \mathsf{R}$. For any random elements ξ and $\xi_1, \xi_2, \ldots$ in S, we further say that ξ_n *converges in distribution to* ξ and write $\xi_n \xrightarrow{d} \xi$ if $\mathcal{L}(\xi_n) \xrightarrow{w} \mathcal{L}(\xi)$, i.e., if $Ef(\xi_n) \to Ef(\xi)$ for all $f \in \hat{C}_S$. The latter mode of convergence clearly depends only on the distributions, and the ξ_n and ξ need not even be defined on the same probability space. To motivate the definition, note that $x_n \to x$ in a metric space S iff $f(x_n) \to f(x)$ for all continuous functions $f: S \to \mathsf{R}$, and that $\mathcal{L}(\xi)$ is determined by the integrals $Ef(\xi)$ for all $f \in \hat{C}_S$.

We give a connection between convergence in probability and distribution.

Lemma 5.7 *(convergence in probability and distribution) Let $\xi, \xi_1, \xi_2, \ldots$ be random elements in a separable metric space (S, ρ). Then*

$$\xi_n \xrightarrow{P} \xi \quad \Rightarrow \quad \xi_n \xrightarrow{d} \xi,$$

with equivalence when ξ is a.s. a constant.

Proof: Let $\xi_n \xrightarrow{P} \xi$. For any $f \in \hat{C}_S$ we need to show that $Ef(\xi_n) \to Ef(\xi)$. If the convergence fails, there exists a sub-sequence $N' \subset \mathsf{N}$ such that $\inf_{n \in N'} |Ef(\xi_n) - Ef(\xi)| > 0$. Then Lemma 5.2 yields $\xi_n \to \xi$ a.s. along a further sub-sequence $N'' \subset N'$. By continuity and dominated convergence, we get $Ef(\xi_n) \to Ef(\xi)$ along N'', a contradiction.

Conversely, let $\xi_n \xrightarrow{d} s \in S$. Since $\rho(x,s) \wedge 1$ is a bounded and continuous function of x, we get $E\{\rho(\xi_n, s) \wedge 1\} \to E\{\rho(s,s) \wedge 1\} = 0$, and so $\xi_n \xrightarrow{P} s$. □

A T-indexed family of random vectors ξ_t in R^d is said to be *tight*, if

$$\lim_{r \to \infty} \sup_{t \in T} P\Big\{|\xi_t| > r\Big\} = 0.$$

For sequences (ξ_n), this is clearly equivalent to

$$\lim_{r \to \infty} \limsup_{n \to \infty} P\Big\{|\xi_n| > r\Big\} = 0, \tag{2}$$

which is often easier to verify. Tightness plays an important role for the compactness methods developed in Chapters 6 and 23. For the moment, we note only the following simple connection with weak convergence.

Lemma 5.8 *(weak convergence and tightness) For random vectors $\xi, \xi_1, \xi_2, \ldots$ in R^d, we have*

$$\xi_n \xrightarrow{d} \xi \quad \Rightarrow \quad (\xi_n) \text{ is tight.}$$

Proof: Fix any $r > 0$, and define $f(x) = \{1 - (r - |x|)_+\}_+$. Then

$$\begin{aligned} \limsup_{n \to \infty} P\Big\{|\xi_n| > r\Big\} &\le \lim_{n \to \infty} Ef(\xi_n) = Ef(\xi) \\ &\le P\Big\{|\xi| > r - 1\Big\}. \end{aligned}$$

Here the right-hand side tends to 0 as $r \to \infty$, and (2) follows. □

We further note the following simple relationship between tightness and convergence in probability.

Lemma 5.9 *(tightness and convergence in probability) For random vectors $\xi_1, \xi_2, \ldots$ in R^d, these conditions are equivalent:*

(i) *(ξ_n) is tight,*

(ii) *$c_n\,\xi_n \xrightarrow{P} 0$ for any constants $c_n \geq 0$ with $c_n \to 0$.*

Proof: Assume (i), and let $c_n \to 0$. Fixing any $r, \varepsilon > 0$, and noting that $c_n\, r \leq \varepsilon$ for all but finitely many $n \in \mathsf{N}$, we get

$$\limsup_{n\to\infty} P\Big\{|c_n\,\xi_n| > \varepsilon\Big\} \leq \limsup_{n\to\infty} P\Big\{|\xi_n| > r\Big\}.$$

Here the right-hand side tends to 0 as $r \to \infty$, and so $P\{|c_n\,\xi_n| > \varepsilon\} \to 0$. Since ε was arbitrary, we get $c_n\,\xi_n \xrightarrow{P} 0$, proving (ii). If instead (i) is false, we have $\inf_k P\{|\xi_{n_k}| > k\} > 0$ for a sub-sequence $(n_k) \subset \mathsf{N}$. Putting $c_n = \sup\{k^{-1};\, n_k \geq n\}$, we note that $c_n \to 0$, and yet $P\{|c_{n_k}\xi_{n_k}| > 1\} \not\to 0$. Thus, even (ii) fails. □

We turn to a related notion for expected values. The random variables ξ_t, $t \in T$, are said to be *uniformly integrable* if

$$\lim_{r\to\infty} \sup_{t\in T} E\Big(|\xi_t|;\ |\xi_t| > r\Big) = 0, \tag{3}$$

where $E(\xi; A) = E(1_A \xi)$. For sequences (ξ_n) in L^1, this is clearly equivalent to

$$\lim_{r\to\infty} \limsup_{n\to\infty} E\Big(|\xi_n|;\ |\xi_n| > r\Big) = 0. \tag{4}$$

Condition (3) holds in particular if the ξ_t are *L^p-bounded* for some $p > 1$, in the sense that $\sup_t E|\xi_t|^p < \infty$. To see this, it suffices to write

$$E\Big(|\xi_t|;\ |\xi_t| > r\Big) \leq r^{-p+1} E|\xi_t|^p, \quad r, p > 0.$$

We give a useful criterion for uniform integrability. For motivation, we note that if ξ is an integrable random variable, then $E(|\xi|; A) \to 0$ as $PA \to 0$, by Lemma 5.2 and dominated convergence, in the sense that

$$\lim_{\varepsilon\to 0} \sup_{PA<\varepsilon} E(|\xi|; A) = 0.$$

Lemma 5.10 *(uniform integrability) For any family of random variables ξ_t, $t \in T$, these conditions are equivalent:*

(i) *the ξ_t are uniformly integrable,*

(ii) $\displaystyle\sup_{t\in T} E|\xi_t| < \infty$ *and* $\displaystyle\lim_{PA\to 0} \sup_{t\in T} E(|\xi_t|; A) = 0.$

Proof: Assuming (i), we may write

$$E(|\xi_t|; A) \leq r\,PA + E\Big(|\xi_t|;\ |\xi_t| > r\Big), \quad r > 0.$$

Here the second part of (ii) follows as we let $PA \to 0$ and then $r \to \infty$. To prove the first part, we may take $A = \Omega$ and choose $r > 0$ large enough.

Conversely, assume (ii). By Chebyshev's inequality, we get as $r \to \infty$

$$\sup_t P\{|\xi_t| > r\} \le r^{-1} \sup_t E|\xi_t| \to 0,$$

and (i) follows from the second part of (ii) with $A = \{|\xi_t| > r\}$. $\square$

The relevance of uniform integrability for the convergence of moments is clear from the following result, which also provides a weak convergence version of Fatou's lemma.

Lemma 5.11 *(convergence of means) For any random variables* $\xi, \xi_1, \xi_2, \ldots \ge 0$ *with* $\xi_n \xrightarrow{d} \xi$, *we have*

(i) $E\,\xi \le \liminf_n E\,\xi_n$,

(ii) $E\,\xi_n \to E\,\xi < \infty \Leftrightarrow$ *the* ξ_n *are uniformly integrable.*

Proof: For any $r > 0$, the function $x \mapsto x \wedge r$ is bounded and continuous on R_+. Thus,

$$\begin{aligned} \liminf_{n\to\infty} E\,\xi_n &\ge \lim_{n\to\infty} E(\xi_n \wedge r) \\ &= E(\xi \wedge r), \end{aligned}$$

and the first assertion follows as we let $r \to \infty$. Next assume (4), and note in particular that $E\,\xi \le \liminf_n E\,\xi_n < \infty$. For any $r > 0$, we get

$$\begin{aligned} \Big|E\,\xi_n - E\,\xi\Big| \le \Big|E\,\xi_n - E(\xi_n \wedge r)\Big| + \Big|E(\xi_n \wedge r) - E(\xi \wedge r)\Big| \\ + \Big|E(\xi \wedge r) - E\,\xi\Big|. \end{aligned}$$

Letting $n \to \infty$ and then $r \to \infty$, we obtain $E\,\xi_n \to E\,\xi$. Now assume instead that $E\,\xi_n \to E\,\xi < \infty$. Keeping $r > 0$ fixed, we get as $n \to \infty$

$$\begin{aligned} E\Big(\xi_n;\, \xi_n > r\Big) &\le E\Big\{\xi_n - \xi_n \wedge (r - \xi_n)_+\Big\} \\ &\to E\Big\{\xi - \xi \wedge (r - \xi)_+\Big\}. \end{aligned}$$

Since $x \wedge (r - x)_+ \uparrow x$ as $r \to \infty$, the right-hand side tends to zero by dominated convergence, and (4) follows. $\square$

We may now prove some useful criteria for convergence in L^p.

Theorem 5.12 *(L^p-convergence) For a fixed* $p > 0$, *let* $\xi_1, \xi_2, \ldots \in L^p$ *with* $\xi_n \xrightarrow{P} \xi$. *Then these conditions are equivalent:*

(i) $\|\xi_n - \xi\|_p \to 0$,

(ii) $\|\xi_n\|_p \to \|\xi\|_p < \infty$,

(iii) *the variables* $|\xi_n|^p$ *are uniformly integrable.*

Conversely, (i) *implies* $\xi_n \xrightarrow{P} \xi$.

Proof: First let $\xi_n \to \xi$ in L^p. Then $\|\xi_n\|_p \to \|\xi\|_p$ by Lemma 1.31, and Lemma 5.1 yields for any $\varepsilon > 0$

$$\begin{aligned} P\Big\{|\xi_n - \xi| > \varepsilon\Big\} &= P\Big\{|\xi_n - \xi|^p > \varepsilon^p\Big\} \\ &\le \varepsilon^{-p}\,\|\xi_n - \xi\|_p^p \to 0, \end{aligned}$$

which shows that $\xi_n \xrightarrow{P} \xi$. We may henceforth assume that $\xi_n \xrightarrow{P} \xi$. In particular, $|\xi_n|^p \xrightarrow{d} |\xi|^p$ by Lemmas 5.3 and 5.7, and so (ii) $\Leftrightarrow$ (iii) by Lemma 5.11. Next assume (ii). If (i) fails, there exists a sub-sequence $N' \subset \mathsf{N}$ with $\inf_{n\in N'} \|\xi_n - \xi\|_p > 0$. Then Lemma 5.2 yields $\xi_n \to \xi$ a.s. along a further sub-sequence $N'' \subset N'$, and so Lemma 1.34 gives $\|\xi_n - \xi\|_p \to 0$ along N'', a contradiction. Thus, (ii) $\Rightarrow$ (i), and so all three conditions are equivalent. □

We consider yet another mode of convergence for random variables. Letting $\xi, \xi_1, \ldots \in L^p$ for a $p \in [1,\infty)$, we say that $\xi_n \to \xi$ *weakly in* L^p if $E\,\xi_n\eta \to E\,\xi\eta$ for every $\eta \in L^q$, where $p^{-1} + q^{-1} = 1$. Taking $\eta = |\xi|^{p-1}\mathrm{sgn}\,\xi$ gives $\|\eta\|_q = \|\xi\|_p^{p-1}$, and so by Hölder's inequality,

$$\begin{aligned} \|\xi\|_p^p &= E\,\xi\eta = \lim_{n\to\infty} E\,\xi_n\eta \\ &\le \|\xi\|_p^{p-1} \liminf_{n\to\infty} \|\xi_n\|_p, \end{aligned}$$

which implies $\|\xi\|_p \le \liminf_n \|\xi_n\|_p$.

Now recall that any L^2-bounded sequence has a sub-sequence converging weakly in L^2. The following related criterion for weak compactness in L^1 will be needed in Chapter 10.

Lemma 5.13 *(weak L^1-compactness*[1]*, Dunford) Every uniformly integrable sequence of random variables has a sub-sequence converging weakly in L^1.*

Proof: Let (ξ_n) be uniformly integrable. Define $\xi_n^k = \xi_n 1\{|\xi_n| \le k\}$, and note that (ξ_n^k) is L^2-bounded in n for fixed k. By the compactness in L^2 and a diagonal argument, there exist a sub-sequence $N' \subset \mathsf{N}$ and some random variables $\eta_1, \eta_2, \ldots$ such that $\xi_n^k \to \eta_k$ holds weakly in L^2 and then also in L^1, as $n \to \infty$ along N' for fixed k.

Now $\|\eta_k - \eta_l\|_1 \le \liminf_n \|\xi_n^k - \xi_n^l\|_1$, and by uniform integrability the right-hand side tends to zero as $k, l \to \infty$. Thus, the sequence (η_k) is Cauchy in L^1, and so it converges in L^1 toward some ξ. By approximation it follows easily that $\xi_n \to \xi$ weakly in L^1 along N'. □

We turn to some convergence criteria for random series, beginning with an important special case.

Proposition 5.14 *(series with positive terms) For any independent random variables $\xi_1, \xi_2, \ldots \ge 0$,*

$$\textstyle\sum_n \xi_n < \infty \ \text{ a.s.} \quad \Leftrightarrow \quad \sum_n E(\xi_n \wedge 1) < \infty.$$

[1]The converse statement is also true, but it is not needed in this book.

Proof: The right-hand condition yields $E\sum_n(\xi_n \wedge 1) < \infty$ by Fubini's theorem, and so $\sum_n(\xi_n \wedge 1) < \infty$ a.s. In particular, $\sum_n 1\{\xi_n > 1\} < \infty$ a.s., and so the series $\sum_n(\xi_n \wedge 1)$ and $\sum_n \xi_n$ differ by at most finitely many terms, which implies $\sum_n \xi_n < \infty$ a.s.

Conversely, let $\sum_n \xi_n < \infty$ a.s. Then also $\sum_n(\xi_n \wedge 1) < \infty$ a.s., and so we may assume that $\xi_n \le 1$ for all n. Since $1 - x \le e^{-x} \le 1 - a\,x$ for $x \in [0,1]$ with $a = 1 - e^{-1}$, we get

$$\begin{aligned} 0 &< E\exp\Big(-\textstyle\sum_n \xi_n\Big) \\ &= \textstyle\prod_n Ee^{-\xi_n} \\ &\le \textstyle\prod_n (1 - a\,E\xi_n) \\ &\le \textstyle\prod_n \exp(-a\,E\xi_n) \\ &= \exp\Big(-\,a\textstyle\sum_n E\xi_n\Big), \end{aligned}$$

and so $\sum_n E\xi_n < \infty$. □

To deal with more general series, we need the following strengthened version of Chebyshev's inequality. A further extension appears as Proposition 9.16.

Lemma 5.15 *(maximum inequality, Kolmogorov) Let $\xi_1, \xi_2, \ldots$ be independent random variables with mean* 0*, and put $S_n = \sum_{k \le n} \xi_k$. Then*

$$P\Big\{\sup_n |S_n| > r\Big\} \le r^{-2} \textstyle\sum_n \mathrm{Var}(\xi_n), \quad r > 0.$$

Proof: We may assume that $\sum_n E\,\xi_n^2 < \infty$. Writing $\tau = \inf\{n;\, |S_n| > r\}$ and noting that $S_k 1\{\tau = k\} \perp\!\!\!\perp (S_n - S_k)$ for $k \le n$, we get

$$\begin{aligned} \sum_{k \le n} E\,\xi_k^2 = E\,S_n^2 &\ge \sum_{k \le n} E\Big(S_n^2;\, \tau = k\Big) \\ &\ge \sum_{k \le n} \Big(E\Big\{S_k^2;\, \tau = k\Big\} + 2E\Big\{S_k(S_n - S_k);\, \tau = k\Big\}\Big) \\ &= \sum_{k \le n} E\Big(S_k^2;\, \tau = k\Big) \ge r^2 P\{\tau \le n\}. \end{aligned}$$

As $n \to \infty$, we obtain

$$\begin{aligned} \textstyle\sum_k E\,\xi_k^2 &\ge r^2 P\{\tau < \infty\} \\ &= r^2 P\Big\{\sup_k |S_k| > r\Big\}. \end{aligned}$$ □

The last result yields a sufficient condition for a.s. convergence of random series with independent terms. Precise criteria will be given in Theorem 5.18.

Lemma 5.16 *(variance criterion for series, Khinchin & Kolmogorov) Let $\xi_1, \xi_2, \ldots$ be independent random variables with mean* 0*. Then*

$$\textstyle\sum_n \mathrm{Var}(\xi_n) < \infty \quad \Rightarrow \quad \sum_n \xi_n \ \text{ converges a.s.}$$

Proof: Write $S_n = \xi_1 + \cdots + \xi_n$. By Lemma 5.15 we get for any $\varepsilon > 0$

$$P\Big\{\sup_{k\geq n}|S_n - S_k| > \varepsilon\Big\} \leq \varepsilon^{-2}\sum_{k\geq n} E\,\xi_k^2.$$

Hence, as $n \to \infty$,

$$\sup_{h,k\geq n}|S_h - S_k| \leq 2\sup_{k\geq n}|S_n - S_k| \xrightarrow{P} 0,$$

and Lemma 5.2 yields the corresponding a.s. convergence along a sub-sequence. Since the supremum on the left is non-increasing in n, the a.s. convergence extends to the entire sequence, which means that (S_n) is a.s. Cauchy convergent. Thus, S_n converges a.s. by Lemma 5.6. □

We turn to a basic connection between series with positive and symmetric terms. By $\xi_n \xrightarrow{P} \infty$ we mean that $P\{\xi_n > r\} \to 1$ for every $r > 0$.

Theorem 5.17 *(positive and symmetric terms) Let $\xi_1, \xi_2, \ldots$ be independent, symmetric random variables. Then these conditions are equivalent:*

(i) $\sum_n \xi_n$ *converges a.s.,*

(ii) $\sum_n \xi_n^2 < \infty$ *a.s.,*

(iii) $\sum_n E\big(\xi_n^2 \wedge 1\big) < \infty$.

If the conditions fail, then $\Big|\sum_{k\leq n}\xi_k\Big| \xrightarrow{P} \infty$.

Proof: The equivalence (ii) $\Leftrightarrow$ (iii) holds by Proposition 5.14. Next assume (iii), and conclude from Lemma 5.16 that $\sum_n \xi_n 1\{|\xi_n| \leq 1\}$ converges a.s. From (iii) and Fubini's theorem we have also $\sum_n 1\{|\xi_n| > 1\} < \infty$ a.s. Hence, the series $\sum_n \xi_n 1\{|\xi_n| \leq 1\}$ and $\sum_n \xi_n$ differ by at most finitely many terms, and so even the latter series converges a.s. This shows that (i) $\Leftarrow$ (ii) $\Leftrightarrow$ (iii). It remains to show that $|S_n| \xrightarrow{P} \infty$ when (ii) fails, since the former condition yields $|S_n| \to \infty$ a.s. along a sub-sequence, contradicting (i). Here $S_n = \xi_1 + \cdots + \xi_n$ as before.

Thus, suppose that (ii) fails. Then Kolmogorov's 0−1 law yields $\sum_n \xi_n^2 = \infty$ a.s. First let $|\xi_n| = c_n$ be non-random for every n. If the c_n are unbounded, then for every $r > 0$ we may choose a sub-sequence $n_1, n_2, \ldots \in \mathsf{N}$ such that $c_{n_1} > r$ and $c_{n_{k+1}} > 2c_{n_k}$ for all k. Then clearly $P\Big\{\sum_{j\leq k}\xi_{n_j} \in I\Big\} \leq 2^{-k}$ for every interval I of length $2r$, and so by convolution $P\{|S_n| \leq r\} \leq 2^{-k}$ for all $n \geq n_k$, which implies $P\{|S_n| \leq r\} \to 0$.

Next let $c_n \leq c < \infty$ for all n. Choosing $a > 0$ so small that $\cos x \leq e^{-a\,x^2}$ for $|x| \leq 1$, we get for $0 < |t| \leq c^{-1}$

$$\begin{aligned} 0 \leq E e^{itS_n} &= \prod_{k\leq n}\cos(t\,c_k) \\ &\leq \prod_{k\leq n}\exp\Big(-a\,t^2 c_k^2\Big) \\ &= \exp\Big(-\,a\,t^2\sum_{k\leq n} c_k^2\Big) \to 0. \end{aligned}$$

Anticipating the elementary Lemma 6.1 below, we get again $P\{|S_n| \le r\} \to 0$ for each $r > 0$.

For general ξ_n, choose some independent i.i.d. random variables ϑ_n with $P\{\vartheta_n = \pm 1\} = \frac{1}{2}$, and note that the sequences (ξ_n) and $(\vartheta_n |\xi_n|)$ have the same distribution. Letting μ be the distribution of the sequence $(|\xi_n|)$, we get by Lemma 4.11

$$P\Big\{|S_n| > r\Big\} = \int P\Big\{\Big|\sum_{k \le n} \vartheta_k\, x_k\Big| > r\Big\}\, \mu(dx), \quad r > 0.$$

Here the integrand tends to 0 for μ-almost every sequence $x = (x_n)$, by the result for constant $|\xi_n|$, and so the integral tends to 0 by dominated convergence. □

We may now give precise criteria for convergence, a.s. or in distribution, of a series of independent random variables. Write $\mathrm{Var}(\xi; A) = \mathrm{Var}(\xi 1_A)$.

Theorem 5.18 *(three-series criterion, Kolmogorov, Lévy) Let $\xi_1, \xi_2, \ldots$ be independent random variables. Then $\sum_n \xi_n$ converges a.s. iff it converges in distribution, which holds iff*

(i) $\sum_n P\Big\{|\xi_n| > 1\Big\} < \infty$,

(ii) $\sum_n E\Big(\xi_n;\, |\xi_n| \le 1\Big)$ *converges,*

(iii) $\sum_n \mathrm{Var}\Big(\xi_n;\, |\xi_n| \le 1\Big) < \infty$.

Our proof requires some simple symmetrization inequalities. Say that m is a *median* of the random variable ξ if $P\{\xi > m\} \vee P\{\xi < m\} \le \frac{1}{2}$. A *symmetrization* of ξ is defined as a random variable of the form $\tilde{\xi} = \xi - \xi'$, where $\xi' \perp\!\!\!\perp \xi$ with $\xi' \overset{d}{=} \xi$. For symmetrizations of the random variables $\xi_1, \xi_2, \ldots$, we require the same properties for the whole sequences (ξ_n) and (ξ'_n).

Lemma 5.19 *(symmetrization) Let $\tilde{\xi}$ be a symmetrization of a random variable ξ with median m. Then for any $r > 0$,*

$$\tfrac{1}{2} P\Big\{|\xi - m| > r\Big\} \le P\Big\{|\tilde{\xi}| > r\Big\} \le 2\,P\Big\{|\xi| > r/2\Big\}.$$

Proof: Let $\tilde{\xi} = \xi - \xi'$ as above, and write

$$\begin{aligned}\Big\{\xi - m > r,\ \xi' \le m\Big\} \cup \Big\{\xi - m < -r,\ \xi' \ge m\Big\} &\subset \Big\{|\tilde{\xi}| > r\Big\} \\ &\subset \Big\{|\xi| > r/2\Big\} \cup \Big\{|\xi'| > r/2\Big\}.\end{aligned}$$

□

We also need a simple centering lemma.

Lemma 5.20 *(centering) For any random variables $\xi, \eta, \xi_1, \xi_2, \ldots$ and constants $c_1, c_2, \ldots$, we have*

$$\left.\begin{array}{r} \xi_n \xrightarrow{d} \xi \\ \xi_n + c_n \xrightarrow{d} \eta \end{array}\right\} \quad \Rightarrow \quad c_n \to \ \textit{some } c.$$

Proof: Let $\xi_n \xrightarrow{d} \xi$. If $c_n \to \pm\infty$ along a sub-sequence $N' \subset \mathsf{N}$, then clearly $\xi_n + c_n \xrightarrow{P} \pm\infty$ along N', which contradicts the tightness of $\xi_n + c_n$. Thus, the c_n are bounded. Now assume that $c_n \to a$ and $c_n \to b$ along two sub-sequences $N_1, N_2 \subset \mathsf{N}$. Then $\xi_n + c_n \xrightarrow{d} \xi + a$ along N_1, while $\xi_n + c_n \xrightarrow{d} \xi + b$ along N_2, and so $\xi + a \stackrel{d}{=} \xi + b$. Iterating this yields $\xi + n(b-a) \stackrel{d}{=} \xi$ for every $n \in \mathsf{Z}$, which is only possible when $a = b$. Thus, all limit points of (c_n) agree, and c_n converges. □

Proof of Theorem 5.18: Assume (i)−(iii), and define $\xi'_n = \xi_n 1\{|\xi_n| \le 1\}$. By (iii) and Lemma 5.16 the series $\sum_n(\xi'_n - E\,\xi'_n)$ converges a.s., and so by (ii) the same thing is true for $\sum_n \xi'_n$. Finally, $P\{\xi_n \ne \xi'_n \text{ i.o.}\} = 0$ by (i) and the Borel–Cantelli lemma, and so $\sum_n(\xi_n - \xi'_n)$ has a.s. finitely many non-zero terms. Hence, even $\sum_n \xi_n$ converges a.s.

Conversely, suppose that $\sum_n \xi_n$ converges in distribution. By Lemma 5.19 the sequence of symmetrized partial sums $\sum_{k\le n} \tilde{\xi}_k$ is tight, and so $\sum_n \tilde{\xi}_n$ converges a.s. by Theorem 5.17. In particular, $\tilde{\xi}_n \to 0$ a.s. For any $\varepsilon > 0$ we obtain $\sum_n P\{|\tilde{\xi}_n| > \varepsilon\} < \infty$ by the Borel–Cantelli lemma. Hence, $\sum_n P\{|\xi_n - m_n| > \varepsilon\} < \infty$ by Lemma 5.19, where $m_1, m_2, \ldots$ are medians of $\xi_1, \xi_2, \ldots$. Using the Borel–Cantelli lemma again, we get $\xi_n - m_n \to 0$ a.s.

Now let $c_1, c_2, \ldots$ be arbitrary with $m_n - c_n \to 0$. Then even $\xi_n - c_n \to 0$ a.s. Putting $\eta_n = \xi_n 1\{|\xi_n - c_n| \le 1\}$, we get a.s. $\xi_n = \eta_n$ for all but finitely many n, and similarly for the symmetrized variables $\tilde{\xi}_n$ and $\tilde{\eta}_n$. Thus, even $\sum_n \tilde{\eta}_n$ converges a.s. Since the $\tilde{\eta}_n$ are bounded and symmetric, Theorem 5.17 yields $\sum_n \operatorname{Var}(\eta_n) = \frac{1}{2}\sum_n \operatorname{Var}(\tilde{\eta}_n) < \infty$. Thus, $\sum_n (\eta_n - E\eta_n)$ converges a.s. by Lemma 5.16, as does the series $\sum_n (\xi_n - E\eta_n)$. Comparing with the distributional convergence of $\sum_n \xi_n$, we conclude from Lemma 5.20 that $\sum_n E\eta_n$ converges. In particular, $E\eta_n \to 0$ and $\eta_n - E\eta_n \to 0$ a.s., and so $\eta_n \to 0$ a.s., whence $\xi_n \to 0$ a.s. Then $m_n \to 0$, and so we may take $c_n = 0$ in the previous argument, and (i)−(iii) follow. □

A sequence of random variables $\xi_1, \xi_2, \ldots$ with partial sums S_n is said to obey the *strong law of large numbers*, if S_n/n converges a.s. to a constant. The *weak law* is the corresponding property with convergence in probability. The following elementary proposition enables us to convert convergence results for random series into laws of large numbers.

Lemma 5.21 *(series and averages, Kronecker) For any $a_1, a_2, \ldots \in \mathsf{R}$ and $c > 0$,*

$$\sum_{n\ge1} n^{-c} a_n \ \text{ converges } \quad \Rightarrow \quad n^{-c}\sum_{k\le n} a_k \to 0$$

Proof: Let $\sum_n b_n = b$ with $b_n = n^{-c} a_n$. By dominated convergence as $n \to \infty$,

$$\sum_{k\le n} b_k - n^{-c}\sum_{k\le n} a_k \;=\; \sum_{k\le n} \left\{1 - (k/n)^c\right\} b_k$$

$$\begin{aligned}
&= c \sum_{k\le n} b_k \int_{k/n}^1 x^{c-1} dx \\
&= c \int_0^1 x^{c-1} dx \sum_{k\le nx} b_k \\
&\to b\, c \int_0^1 x^{c-1} dx = b,
\end{aligned}$$

and the assertion follows since the first term on the left tends to b. □

The following simple result illustrates the method.

Corollary 5.22 *(variance criterion for averages, Kolmogorov) Let $\xi_1, \xi_2, \ldots$ be independent random variables with mean* 0*, and fix any $c > 0$. Then*

$$\sum_{n\ge 1} n^{-2c}\, \mathrm{Var}(\xi_n) < \infty \quad \Rightarrow \quad n^{-c} \sum_{k\le n} \xi_k \to 0 \;\; a.s.$$

Proof: Since $\sum_n n^{-c}\, \xi_n$ converges a.s. by Lemma 5.16, the assertion follows by Lemma 5.21. □

In particular, we see that if $\xi, \xi_1, \xi_2, \ldots$ are i.i.d. with $E\,\xi = 0$ and $E\,\xi^2 < \infty$, then $n^{-c} \sum_{k\le n} \xi_k \to 0$ a.s. for any $c > \frac{1}{2}$. The statement fails for $c = \frac{1}{2}$, as we see by taking ξ to be $N(0,1)$. The best possible normalization will be given in Corollary 22.8. The next result characterizes the stated convergence for arbitrary $c > \frac{1}{2}$. For $c = 1$ we recognize the strong law of large numbers. Corresponding criteria for the weak law are given in Theorem 6.17.

Theorem 5.23 *(strong laws of large numbers, Kolmogorov, Marcinkiewicz & Zygmund) Let ξ, $\xi_1, \xi_2, \ldots$ be i.i.d. random variables, put $S_n = \sum_{k\le n} \xi_k$, and fix a $p \in (0,2)$. Then $n^{-1/p}\, S_n$ converges a.s. iff these conditions hold, depending on the value of p:*

- *for $p \in (0,1]$: $\xi \in L^p$,*
- *for $p \in (1,2)$: $\xi \in L^p$ and $E\xi = 0$.*

In that case, the limit equals $E\xi$ when $p = 1$ and is otherwise equal to 0.

Proof: Assume $E|\xi|^p < \infty$ and also, for $p \ge 1$, that $E\,\xi = 0$. Define $\xi_n' = \xi_n 1\{|\xi_n| \le n^{1/p}\}$, and note that by Lemma 4.4

$$\begin{aligned}
\textstyle\sum_n P\Big\{\xi_n' \ne \xi_n\Big\} &= \textstyle\sum_n P\Big\{|\xi|^p > n\Big\} \\
&\le \int_0^\infty P\Big\{|\xi|^p > t\Big\}\, dt \\
&= E|\xi|^p < \infty.
\end{aligned}$$

Hence, the Borel–Cantelli lemma yields $P\{\xi_n' \ne \xi_n \text{ i.o.}\} = 0$, and so $\xi_n' = \xi_n$ for all but finitely many $n \in \mathsf{N}$ a.s. It is then equivalent to show that $n^{-1/p} \sum_{k\le n} \xi_k' \to 0$ a.s. By Lemma 5.21 it suffices to prove instead that $\sum_n n^{-1/p}\, \xi_n'$ converges a.s.

For $p < 1$, this is clear if we write

$$\begin{aligned} E \sum_n n^{-1/p} |\xi'_n| &= \sum_n n^{-1/p} E\Big(|\xi|;\, |\xi| \le n^{1/p}\Big) \\ &\lesssim \int_0^\infty t^{-1/p} E\Big(|\xi|;\, |\xi| \le t^{1/p}\Big)\, dt \\ &= E\Big(|\xi| \int_{|\xi|^p}^\infty t^{-1/p}\, dt\Big) \\ &\lesssim E|\xi|^p < \infty. \end{aligned}$$

If instead $p > 1$, then by Theorem 5.18 it suffices to prove that $\sum_n n^{-1/p} E\,\xi'_n$ converges and $\sum_n n^{-2/p}\, \mathrm{Var}(\xi'_n) < \infty$. Since $E\,\xi'_n = -E(\xi;\, |\xi| > n^{1/p})$, we have for the former series

$$\begin{aligned} \sum_n n^{-1/p} |E\,\xi'_n| &\le \sum_n n^{-1/p} E\Big(|\xi|;\, |\xi| > n^{1/p}\Big) \\ &\le \int_0^\infty t^{-1/p} E\Big(|\xi|;\, |\xi| > t^{1/p}\Big)\, dt \\ &= E\Big(|\xi| \int_0^{|\xi|^p} t^{-1/p}\, dt\Big) \\ &\lesssim E|\xi|^p < \infty. \end{aligned}$$

For the latter series, we obtain

$$\begin{aligned} \sum_n n^{-2/p}\, \mathrm{Var}(\xi'_n) &\le \sum_n n^{-2/p}\, E(\xi'_n)^2 \\ &= \sum_n n^{-2/p}\, E\Big(\xi^2;\, |\xi| \le n^{1/p}\Big) \\ &\lesssim \int_0^\infty t^{-2/p}\, E\Big(\xi^2;\, |\xi| \le t^{1/p}\Big)\, dt \\ &= E\Big(\xi^2 \int_{|\xi|^p}^\infty t^{-2/p}\, dt\Big) \\ &\lesssim E|\xi|^p < \infty. \end{aligned}$$

When $p = 1$, we have $E\,\xi'_n = E(\xi;\, |\xi| \le n) \to 0$ by dominated convergence. Thus, $n^{-1} \sum_{k \le n} E\,\xi'_k \to 0$, and we may prove instead that $n^{-1} \sum_{k \le n} \xi''_k \to 0$ a.s., where $\xi''_n = \xi'_n - E\,\xi'_n$. By Lemma 5.21 and Theorem 5.18 it is then enough to show that $\sum_n n^{-2}\, \mathrm{Var}(\xi'_n) < \infty$, which may be seen as before.

Conversely, suppose that $n^{-1/p} S_n = n^{-1/p} \sum_{k \le n} \xi_k$ converges a.s. Then

$$\frac{\xi_n}{n^{1/p}} = \frac{S_n}{n^{1/p}} - \Big(\frac{n-1}{n}\Big)^{1/p} \frac{S_{n-1}}{(n-1)^{1/p}} \to 0 \ \text{ a.s.},$$

and in particular $P\Big\{|\xi_n|^p > n \text{ i.o.}\Big\} = 0$. Hence, by Lemma 4.4 and the Borel–Cantelli lemma,

$$\begin{aligned} E|\xi|^p &= \int_0^\infty P\Big\{|\xi|^p > t\Big\}\, dt \\ &\le 1 + \sum_{n \ge 1} P\Big\{|\xi|^p > n\Big\} < \infty. \end{aligned}$$

For $p > 1$, the direct assertion yields $n^{-1/p}(S_n - n\,E\xi) \to 0$ a.s., and so $n^{1-1/p} E\,\xi$ converges, which implies $E\,\xi = 0$. □

For a simple application of the law of large numbers, consider an arbitrary sequence of random variables $\xi_1, \xi_2, \ldots$, and define the associated *empirical distributions* as the random probability measures $\hat{\mu}_n = n^{-1} \sum_{k \le n} \delta_{\xi_k}$. The corresponding *empirical distribution functions* $\hat{F}_n$ are given by

$$\begin{aligned}\hat{F}_n(x) &= \hat{\mu}_n(-\infty, x] \\ &= n^{-1} \sum_{k \le n} 1\{\xi_k \le x\}, \quad x \in \mathsf{R},\ n \in \mathsf{N}.\end{aligned}$$

Proposition 5.24 *(empirical distribution functions, Glivenko, Cantelli) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with distribution function F and empirical distribution functions $\hat{F}_1, \hat{F}_2, \ldots$. Then*

$$\lim_{n \to \infty} \sup\nolimits_x \left| \hat{F}_n(x) - F(x) \right| = 0 \quad a.s. \tag{5}$$

Proof: The law of large numbers yields $\hat{F}_n(x) \to F(x)$ a.s. for every $x \in \mathsf{R}$. Now fix any finite partition $-\infty = x_1 < x_2 < \cdots < x_m = \infty$. By the monotonicity of F and $\hat{F}_n$,

$$\begin{aligned}\sup\nolimits_x \left| \hat{F}_n(x) - F(x) \right| \le{}& \max\nolimits_k \left| \hat{F}_n(x_k) - F(x_k) \right| \\ &+ \max\nolimits_k \left| F(x_{k+1}) - F(x_k) \right|.\end{aligned}$$

Letting $n \to \infty$ and refining the partition indefinitely, we get in the limit

$$\limsup_{n \to \infty} \sup\nolimits_x \left| \hat{F}_n(x) - F(x) \right| \le \sup\nolimits_x \Delta F(x) \quad \text{a.s.},$$

which proves (5) when F is continuous.

For general F, let $\vartheta_1, \vartheta_2, \ldots$ be i.i.d. $U(0,1)$, and define $\eta_n = g(\vartheta_n)$ for each n, where $g(t) = \sup\{x;\ F(x) < t\}$. Then $\eta_n \le x$ iff $\vartheta_n \le F(x)$, and so $(\eta_n) \stackrel{d}{=} (\xi_n)$. We may then assume that $\xi_n \equiv \eta_n$. Writing $\hat{G}_1, \hat{G}_2, \ldots$ for the empirical distribution functions of $\vartheta_1, \vartheta_2, \ldots$, we see that also $\hat{F}_n = \hat{G}_n \circ F$. Writing $A = F(\mathsf{R})$ and using the result for continuous F, we get a.s.

$$\begin{aligned}\sup\nolimits_x \left| \hat{F}_n(x) - F(x) \right| &= \sup_{t \in A} \left| \hat{G}_n(t) - t \right| \\ &\le \sup_{t \in [0,1]} \left| \hat{G}_n(t) - t \right| \to 0.\end{aligned}$$

□

We turn to a more systematic study of convergence in distribution. Though for the moment we are mostly interested in distributions on Euclidean spaces, it is crucial for future applications to consider the more general setting of an abstract metric space. In particular, the theory is applied in Chapter 23 to random elements in various function spaces. For a random elements ξ in a metric space S with Borel σ-field $\mathcal{S}$, let $\mathcal{S}_\xi$ denote the class of sets $B \in \mathcal{S}$ with $\xi \notin \partial B$ a.s., called the ξ-*continuity sets*.

Theorem 5.25 *(portmanteau[2] theorem, Alexandrov) Let $\xi, \xi_1, \xi_2, \ldots$ be random elements in a metric space $(S, \mathcal{S})$ with classes $\mathcal{G}$, $\mathcal{F}$ of open and closed sets. Then these conditions are equivalent:*

(i) $\xi_n \xrightarrow{d} \xi$,

(ii) $\liminf_{n\to\infty} P\{\xi_n \in G\} \geq P\{\xi \in G\}, \quad G \in \mathcal{G}$,

(iii) $\limsup_{n\to\infty} P\{\xi_n \in F\} \leq P\{\xi \in F\}, \quad F \in \mathcal{F}$,

(iv) $P\{\xi_n \in B\} \to P\{\xi \in B\}, \quad B \in \mathcal{S}_\xi$.

Proof: Assume (i), and fix any $G \in \mathcal{G}$. Letting f be continuous with $0 \leq f \leq 1_G$, we get $Ef(\xi_n) \leq P\{\xi_n \in G\}$, and (ii) follows as we let $n \to \infty$ and then $f \uparrow 1_G$. The equivalence (ii) $\Leftrightarrow$ (iii) is clear from taking complements. Now assume (ii)$-$(iii). For any $B \in \mathcal{S}$,

$$\begin{aligned} P\{\xi \in B^o\} &\leq \liminf_{n\to\infty} P\{\xi_n \in B\} \\ &\leq \limsup_{n\to\infty} P\{\xi_n \in B\} \\ &\leq P\{\xi \in \bar{B}\}. \end{aligned}$$

Here the extreme members agree when $\xi \notin \partial B$ a.s., and (iv) follows.

Conversely, assume (iv), and fix any $F \in \mathcal{F}$. Write $F^\varepsilon = \{s \in S;\, \rho(s, F) \leq \varepsilon\}$. Then the sets $\partial F^\varepsilon \subset \{s;\, \rho(s, F) = \varepsilon\}$ are disjoint, and so $\xi \notin \partial F^\varepsilon$ for almost every $\varepsilon > 0$. For such an ε, we may write $P\{\xi_n \in F\} \leq P\{\xi_n \in F^\varepsilon\}$, and (iii) follows as we let $n \to \infty$ and then $\varepsilon \to 0$. Finally, assume (ii), and let $f \geq 0$ be continuous. By Lemma 4.4 and Fatou's lemma,

$$\begin{aligned} Ef(\xi) &= \int_0^\infty P\Big\{f(\xi) > t\Big\}\, dt \\ &\leq \int_0^\infty \liminf_{n\to\infty} P\Big\{f(\xi_n) > t\Big\}\, dt \\ &\leq \liminf_{n\to\infty} \int_0^\infty P\Big\{f(\xi_n) > t\Big\}\, dt \\ &= \liminf_{n\to\infty} Ef(\xi_n). \end{aligned} \tag{6}$$

Now let f be continuous with $|f| \leq c < \infty$. Applying (6) to the functions $c \pm f$ yields $Ef(\xi_n) \to Ef(\xi)$, which proves (i). $\square$

We insert an easy application to subspaces, needed in Chapter 23.

Corollary 5.26 *(sub-spaces) Let $\xi, \xi_1, \xi_2, \ldots$ be random elements in a subspace A of a metric space (S, ρ). Then*

$$\xi_n \xrightarrow{d} \xi \ \textit{ in } (A, \rho) \quad \Leftrightarrow \quad \xi_n \xrightarrow{d} \xi \ \textit{ in } (S, \rho)$$

[2] From the French word *portemanteau* = 'big suitcase'. Here it is just a collection of criteria, traditionally grouped together.

Proof: Since $\xi, \xi_1, \xi_2, \ldots \in A$, condition (ii) of Theorem 5.25 is equivalent to

$$\liminf_{n\to\infty} P\Big\{\xi_n \in A\cap G\Big\} \geq P\Big\{\xi \in A\cap G\Big\}, \quad G \subset S \text{ open}.$$

By Lemma 1.6, this agrees with condition (ii) of Theorem 5.25 for the subspace A. □

Directly from the definitions, it is clear that convergence in distribution is preserved by continuous mappings. The following more general statement is a key result of weak convergence theory.

Theorem 5.27 *(continuous mapping, Mann & Wald, Prohorov, Rubin) For any metric spaces S, T and set $C \in \mathcal{S}$, consider some measurable functions $f, f_1, f_2, \ldots : S \to T$ satisfying*

$$s_n \to s \in C \quad \Rightarrow \quad f_n(s_n) \to f(s).$$

Then for any random elements $\xi, \xi_1, \xi_2, \ldots$ in S,

$$\xi_n \xrightarrow{d} \xi \in C \;\; a.s. \quad \Rightarrow \quad f_n(\xi_n) \xrightarrow{d} f(\xi).$$

In particular, we see that if $f : S \to T$ is a.s. continuous at ξ, then

$$\xi_n \xrightarrow{d} \xi \quad \Rightarrow \quad f(\xi_n) \xrightarrow{d} f(\xi).$$

Proof: Fix any open set $G \subset T$, and let $s \in f^{-1}G \cap C$. By hypothesis there exist an integer $m \in \mathsf{N}$ and a neighborhood B of s, such that $f_k(s') \in G$ for all $k \geq m$ and $s' \in B$. Thus, $B \subset \bigcap_{k\geq m} f_k^{-1}G$, and so

$$f^{-1}G \cap C \subset \bigcup_m \Big(\bigcap_{k\geq m} f_k^{-1}G\Big)^o.$$

Writing $\mu, \mu_1, \mu_2, \ldots$ for the distributions of $\xi, \xi_2, \xi_2, \ldots$, we get by Theorem 5.25

$$\begin{aligned} \mu(f^{-1}G) &\leq \mu \bigcup_m \Big(\bigcap_{k\geq m} f_k^{-1}G\Big)^o \\ &= \sup_m \mu \Big(\bigcap_{k\geq m} f_k^{-1}G\Big)^o \\ &\leq \sup_m \liminf_{n\to\infty} \mu_n \Big(\bigcap_{k\geq m} f_k^{-1}G\Big)^o \\ &\leq \liminf_{n\to\infty} \mu_n(f_n^{-1}G). \end{aligned}$$

Then the same theorem yields $\mu_n \circ f_n^{-1} \xrightarrow{w} \mu \circ f^{-1}$, which means that $f_n(\xi_n) \xrightarrow{d} f(\xi)$. □

For a simple consequence, we note a useful randomization principle:

Corollary 5.28 *(randomization) For any metric spaces S, T, consider some probability kernels $\mu, \mu_1, \mu_2, \ldots : S \to T$ satisfying*

$$s_n \to s \;\text{ in } S \quad \Rightarrow \quad \mu_n(s_n, \cdot) \xrightarrow{w} \mu(s, \cdot) \;\text{ in } \mathcal{M}_T.$$

Then for any random elements $\xi, \xi_1, \xi_2, \ldots$ in S,

$$\xi_n \xrightarrow{d} \xi \quad \Rightarrow \quad E\,\mu_n(\xi_n, \cdot) \xrightarrow{w} E\,\mu(\xi, \cdot) \;\text{ in } \mathcal{M}_T.$$

Proof: For any bounded, continuous function f on T, the integrals μf and $\mu_n f$ are bounded, measurable functions on S, such that $s_n \to s$ implies $\mu_n f(s_n) \to \mu f(s)$. Hence, Theorem 5.27 yields $\mu_n f(\xi_n) \xrightarrow{d} \mu f(\xi)$, and so $E\,\mu_n f(\xi_n) \to E\,\mu f(\xi)$. The assertion follows since f was arbitrary. □

We turn to an equally important approximation theorem. Here the idea is to prove $\xi_n \xrightarrow{d} \xi$ by approximating $\xi_n \approx \eta_n$ and $\xi \approx \eta$, where $\eta_n \xrightarrow{d} \eta$. The required convergence will then follow, provided the former approximations hold uniformly in n.

Theorem 5.29 *(approximation) Let ξ, ξ_n and η^k, η_n^k be random elements in a separable metric space (S,ρ). Then $\xi_n \xrightarrow{d} \xi$, whenever*

(i) $\eta^k \xrightarrow{d} \xi$,

(ii) $\eta_n^k \xrightarrow{d} \eta^k$ *as* $n \to \infty$, $k \in \mathsf{N}$,

(iii) $\lim_{k\to\infty} \limsup_{n\to\infty} E\Big\{\rho(\eta_n^k, \xi_n) \wedge 1\Big\} = 0.$

Proof: For any closed set $F \subset S$ and constant $\varepsilon > 0$, we have

$$P\{\xi_n \in F\} \le P\Big\{\eta_n^k \in F^\varepsilon\Big\} + P\Big\{\rho(\eta_n^k, \xi_n) > \varepsilon\Big\},$$

where $F^\varepsilon = \{s \in S;\, \rho(s,F) \le \varepsilon\}$. By Theorem 5.25, we get as $n \to \infty$

$$\limsup_{n\to\infty} P\{\xi_n \in F\} \le P\{\eta^k \in F^\varepsilon\} + \limsup_{n\to\infty} P\Big\{\rho(\eta_n^k, \xi_n) > \varepsilon\Big\}.$$

Letting $k \to \infty$, we conclude from Theorem 5.25 and (iii) that

$$\limsup_{n\to\infty} P\{\xi_n \in F\} \le P\{\xi \in F^\varepsilon\}.$$

Here the right-hand side tends to $P\{\xi \in F\}$ as $\varepsilon \to 0$. Since F was arbitrary, Theorem 5.25 yields $\xi_n \xrightarrow{d} \xi$. □

We may now consider convergence in distribution of random sequences.

Theorem 5.30 *(random sequences) Let $\xi = (\xi^1, \xi^2, \dots)$ and $\xi_n = (\xi_n^1, \xi_n^2, \dots)$, $n \in \mathsf{N}$, be random elements in $S_1 \times S_2 \times \cdots$, for some separable metric spaces $S_1, S_2, \dots$. Then $\xi_n \xrightarrow{d} \xi$ iff for any functions $f_k \in \hat{C}_{S_k}$,*

$$E\Big\{f_1(\xi_n^1) \cdots f_m(\xi_n^m)\Big\} \to E\Big\{f_1(\xi^1) \cdots f_m(\xi^m)\Big\}, \quad m \in \mathsf{N}. \tag{7}$$

In particular, $\xi_n \xrightarrow{d} \xi$ follows from the finite-dimensional convergence

$$\Big(\xi_n^1, \dots, \xi_n^m\Big) \xrightarrow{d} \Big(\xi^1, \dots, \xi^m\Big), \quad m \in \mathsf{N}. \tag{8}$$

If the sequences ξ and ξ_n have independent components, it suffices that $\xi_n^k \xrightarrow{d} \xi^k$ for each k.

Proof: The necessity is clear from the continuity of the projections $s \mapsto s_k$. To prove the sufficiency, we first assume that (7) holds for a fixed m. Writing $\mathcal{S}'_k = \{B \in \mathcal{B}_{S_k};\ \xi^k \notin \partial B \text{ a.s.}\}$ and applying Theorem 5.25 m times, we obtain

$$P\Big\{(\xi_n^1, \dots, \xi_n^m) \in B\Big\} \to P\Big\{(\xi^1, \dots, \xi^m) \in B\Big\}, \tag{9}$$

for any set $B = B^1 \times \cdots \times B^m$ with $B^k \in \mathcal{S}'_k$ for all k. Since the S_k are separable, we may choose some countable bases $\mathcal{C}_k \subset \mathcal{S}'_k$, so that $\mathcal{C}_1 \times \cdots \times \mathcal{C}_m$ becomes a countable base in $S_1 \times \cdots \times S_m$. Hence, any open set $G \subset S_1 \times \cdots \times S_m$ is a countable union of measurable rectangles $B_j = B_j^1 \times \cdots \times B_j^m$ with $B_j^k \in \mathcal{S}'_k$ for all k. Since the $\mathcal{S}'_k$ are fields, we may easily reduce to the case of disjoint sets B_j. By Fatou's lemma and (9),

$$\begin{aligned}
\liminf_{n\to\infty} P\Big\{(\xi_n^1, \dots, \xi_n^m) \in G\Big\} &= \liminf_{n\to\infty} \textstyle\sum_j P\Big\{(\xi_n^1, \dots, \xi_n^m) \in B_j\Big\} \\
&\ge \textstyle\sum_j P\Big\{(\xi^1, \dots, \xi^m) \in B_j\Big\} \\
&= P\Big\{(\xi^1, \dots, \xi^m) \in G\Big\},
\end{aligned}$$

and so (8) holds by Theorem 5.25.

To see that (8) implies $\xi_n \xrightarrow{d} \xi$, fix any $a_k \in S_k$, $k \in \mathsf{N}$, and note that the mapping $(s_1, \dots, s_m) \mapsto (s_1, \dots, s_m, a_{m+1}, a_{m+2}, \dots)$ is continuous on $S_1 \times \cdots \times S_m$ for each $m \in \mathsf{N}$. By (8) it follows that

$$\Big(\xi_n^1, \dots, \xi_n^m, a_{m+1}, \dots\Big) \xrightarrow{d} \Big(\xi^1, \dots, \xi^m, a_{m+1}, \dots\Big), \quad m \in \mathsf{N}. \tag{10}$$

Writing η_n^m and η^m for the sequences in (10) and letting ρ be the metric in (1), we also note that $\rho(\xi, \eta^m) \le 2^{-m}$ and $\rho(\xi_n, \eta_n^m) \le 2^{-m}$ for all m and n. The convergence $\xi_n \xrightarrow{d} \xi$ now follows by Theorem 5.29. □

For distributional convergence of some random objects $\xi_1, \xi_2, \dots$, the joint distribution of the elements ξ_n is clearly irrelevant. This suggests that we look for a more useful dependence, which may lead to simpler and more transparent proofs.

Theorem 5.31 *(coupling, Skorohod, Dudley) Let $\xi, \xi_1, \xi_2, \dots$ be random elements in a separable metric space (S, ρ), such that $\xi_n \xrightarrow{d} \xi$. Then there exist some random elements $\tilde\xi, \tilde\xi_1, \tilde\xi_2, \dots$ in S with*

$$\tilde\xi \stackrel{d}{=} \xi, \quad \tilde\xi_n \stackrel{d}{=} \xi_n; \qquad \tilde\xi_n \to \tilde\xi \ \text{a.s.}$$

Our proof involves families of independent random elements with specified distributions, whose existence is ensured in general by Corollary 8.25 below. When S is complete, we may rely on the more elementary Theorem 4.19.

Proof: First take $S = \{1, \dots, m\}$, and put $p_k = P\{\xi = k\}$ and $p_k^n = P\{\xi_n = k\}$. Letting $\vartheta \perp\!\!\!\perp \xi$ be $U(0,1)$, we may construct some random elements $\tilde\xi_n \stackrel{d}{=} \xi_n$, such that $\tilde{\xi}_n = k$ whenever $\xi = k$ and $\vartheta \le p_k^n/p_k$. Since $p_k^n \to p_k$ for each k, we get $\tilde{\xi}_n \to \xi$ a.s.

For general S, fix any $p \in \mathsf{N}$, and choose a partition of S into sets $B_1, B_2, \ldots \in \mathcal{S}_\xi$ of diameter $< 2^{-p}$. Next choose m so large that $P\{\xi \notin \bigcup_{k \le m} B_k\} < 2^{-p}$, and put $B_0 = \bigcap_{k \le m} B_k^c$. For $k = 0, \ldots, m$, define $\kappa = k$ when $\xi \in B_k$ and $\kappa_n = k$ when $\xi_n \in B_k$, $n \in \mathsf{N}$. Then $\kappa_n \xrightarrow{d} \kappa$, and the result for finite S yields some $\tilde{\kappa}_n \stackrel{d}{=} \kappa_n$ with $\tilde{\kappa}_n \to \kappa$ a.s. We further introduce some independent random elements ζ_n^k in S with distributions $\mathcal{L}(\xi_n \,|\, \xi_n \in B_k)$, and define $\tilde{\xi}_n^p = \sum_k \zeta_n^k 1\{\tilde{\kappa}_n = k\}$, so that $\tilde{\xi}_n^p \stackrel{d}{=} \xi_n$ for each n.

By construction, we have

$$\left\{\rho(\tilde{\xi}_n^p, \xi) > 2^{-p}\right\} \subset \{\tilde{\kappa}_n \ne \kappa\} \cup \{\xi \in B_0\}, \quad n, p \in \mathsf{N}.$$

Since $\tilde{\kappa}_n \to \kappa$ a.s. and $P\{\xi \in B_0\} < 2^{-p}$, there exists for every p some $n_p \in \mathsf{N}$ with

$$P \bigcup_{n \ge n_p} \left\{\rho(\tilde{\xi}_n^p, \xi) > 2^{-p}\right\} < 2^{-p}, \quad p \in \mathsf{N},$$

and we may further assume that $n_1 < n_2 < \cdots$. Then the Borel–Cantelli lemma yields a.s. $\sup_{n \ge n_p} \rho(\tilde{\xi}_n^p, \xi) \le 2^{-p}$ for all but finitely many p. Defining $\eta_n = \tilde{\xi}_n^p$ for $n_p \le n < n_{p+1}$, we note that $\xi_n \stackrel{d}{=} \eta_n \to \xi$ a.s. □

We conclude with a result on the functional representation of limits, needed in Chapters 18 and 32. To motivate the problem, recall from Lemma 5.6 that, if $\xi_n \xrightarrow{P} \eta$ for some random elements in a complete metric space S, then $\eta = f(\xi)$ a.s. for some measurable function $f : S^\infty \to S$, where $\xi = (\xi_n)$. Since f depends on the distribution μ of ξ, a universal representation would be of the form $\eta = f(\xi, \mu)$. For certain purposes, we need to choose a measurable version of the latter function. To allow constructions by repeated approximation in probability, we consider the more general case where $\eta_n \xrightarrow{P} \eta$ for some random elements $\eta_n = f_n(\xi, \mu)$.

For a precise statement, let $\hat{\mathcal{M}}_S$ be the space of probability measures μ on S, endowed with the σ-field induced by the evaluation maps $\mu \mapsto \mu B$, $B \in \mathcal{B}_S$.

Proposition 5.32 *(representation of limits) Consider a complete metric space (S, ρ), a measurable space U, and some measurable functions $f_1, f_2, \ldots : U \times \hat{\mathcal{M}}_U \to S$. Then there exist a measurable set $A \subset \hat{\mathcal{M}}_U$ and a function $f : U \times A \to S$ such that, for any random element ξ in U with $\mathcal{L}(\xi) = \mu$,*

$$f_n(\xi, \mu) \xrightarrow{P} \text{some } \eta \quad \Leftrightarrow \quad \mu \in A,$$

in which case we can choose $\eta = f(\xi, \mu)$.

Proof: For sequences $s = (s_1, s_2, \ldots)$ in S, define $l(s) = \lim_k s_k$ when the limit exists and put $l(s) = s_\infty$ otherwise, where $s_\infty \in S$ is arbitrary. By Lemma 1.11, l is a measurable mapping from S^∞ to S. Next consider a sequence $\eta = (\eta_1, \eta_2, \ldots)$ of random elements in S, and put $\nu = \mathcal{L}(\eta)$. Define

$n_1, n_2, \ldots$ as in the proof of Lemma 5.6, and note that each $n_k = n_k(\nu)$ is a measurable function of ν. Let C be the set of measures ν with $n_k(\nu) < \infty$ for all k, and note that η_n converges in probability iff $\nu \in C$. Introduce the measurable function

$$g(s, \nu) = l\left\{s_{n_1(\nu)}, s_{n_2(\nu)}, \ldots\right\}, \quad s = (s_1, s_2, \ldots) \in S^\infty, \ \nu \in \hat{\mathcal{M}}_{S^\infty}.$$

If $\nu \in C$, the proof of Lemma 5.6 shows that $\eta_{n_k(\nu)}$ converges a.s., and so $\eta_n \xrightarrow{P} g(\eta, \nu)$.

Now let $\eta_n = f_n(\xi, \mu)$ for a random element ξ in U with distribution μ and some measurable functions f_n. We need to show that ν is a measurable function of μ. But this is clear from Lemma 3.2 (ii), applied to the kernel $\kappa(\mu, \cdot) = \mu$ from $\hat{\mathcal{M}}_U$ to U and the function $F = (f_1, f_2, \ldots) \colon U \times \hat{\mathcal{M}}_U \to S^\infty$. $\square$

For a simple consequence, we consider limits in probability of measurable processes. The resulting statement will be needed in Chapter 18.

Corollary 5.33 *(measurability of limits, Stricker & Yor) Let $X^1, X^2, \ldots$ be S-valued, measurable processes on T, for a complete metric space S and a measurable space T. Then there exists a measurable set $A \subset T$, such that*

$$X_t^n \xrightarrow{P} \text{some } X_t \quad \Leftrightarrow \quad t \in A,$$

in which case we can choose X to be product measurable on $A \times \Omega$.

Proof: Define $\xi_t = (X_t^1, X_t^2, \ldots)$ and $\mu_t = \mathcal{L}(\xi_t)$. By Proposition 5.32 there exist a measurable set $C \subset \mathcal{P}_{S^\infty}$ and a measurable function $f \colon S^\infty \times C \to S$, such that X_t^n converges in probability iff $\mu_t \in C$, in which case $X_t^n \xrightarrow{P} f(\xi_t, \mu_t)$. It remains to note that the mapping $t \mapsto \mu_t$ is measurable, which is clear from Lemmas 1.4 and 1.28. $\square$

Exercises

1. Let $\xi_1, \ldots, \xi_n$ be independent, symmetric random variables. Show that $P\{(\sum_k \xi_k)^2 \geq r \sum_k \xi_k^2\} \geq (1-r)^2/3$ for all $r \in (0,1)$. (*Hint:* Reduce by Lemma 4.11 to the case of non-random $|\xi_k|$, and use Lemma 5.1.)

2. Let $\xi_1, \ldots, \xi_n$ be independent, symmetric random variables. Show that $P\{\max_k |\xi_k| > r\} \leq 2\,P\{|S| > r\}$ for all $r > 0$, where $S = \sum_k \xi_k$. (*Hint:* Let η be the first term ξ_k where $\max_k |\xi_k|$ is attained, and check that $(\eta, S - \eta) \stackrel{d}{=} (\eta, \eta - S)$.)

3. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with $P\{|\xi_n| > t\} > 0$ for all $t > 0$. Prove the existence of some constants $c_1, c_2, \ldots$, such that $c_n \xi_n \to 0$ in probability but not a.s.

4. Show that a family of random variables ξ_t is tight, iff $\sup_t Ef(|\xi_t|) < \infty$ for some increasing function $f \colon \mathsf{R}_+ \to \mathsf{R}_+$ with $f(\infty) = \infty$.

5. Consider some random variables ξ_n, η_n, such that (ξ_n) is tight and $\eta_n \xrightarrow{P} 0$. Show that even $\xi_n \eta_n \xrightarrow{P} 0$.

6. Show that the random variables ξ_t are uniformly integrable, iff $\sup_t Ef(|\xi_t|) < \infty$ for some increasing function $f\colon \mathsf{R}_+ \to \mathsf{R}_+$ with $f(x)/x \to \infty$ as $x \to \infty$.

7. Show that the condition $\sup_t E|\xi_t| < \infty$ in Lemma 5.10 can be omitted if $\mathcal{A}$ is non-atomic.

8. Let $\xi_1, \xi_2, \ldots \in L^1$. Show that the ξ_n are uniformly integrable, iff the condition in Lemma 5.10 holds with $\sup_n$ replaced by $\limsup_n$.

9. Deduce the dominated convergence theorem from Lemma 5.11.

10. Show that if $(|\xi_t|^p)$ and $(|\eta_t|^p)$ are uniformly integrable for some $p > 0$, then so is $(|a\xi_t + b\eta_t|^p)$ for any $a, b \in \mathsf{R}$. (*Hint:* Use Lemma 5.10.) Use this fact to derive Proposition 5.12 from Lemma 5.11.

11. Give examples of random variables $\xi, \xi_1, \xi_2, \ldots \in L^2$, such that $\xi_n \to \xi$ a.s. but not in L^2, in L^2 but not a.s., or in L^1 but not in L^2.

12. Let $\xi_1, \xi_2, \ldots$ be independent random variables in L^2. Show that $\sum_n \xi_n$ converges in L^2, iff $\sum_n E\xi_n$ and $\sum_n \mathrm{Var}(\xi_n)$ both converge.

13. Give an example of some independent, symmetric random variables $\xi_1, \xi_2, \ldots$, such that $\sum_n \xi_n$ converges a.s. but $\sum_n |\xi_n| = \infty$ a.s.

14. Let ξ_n, η_n be symmetric random variables with $|\xi_n| \le |\eta_n|$, such that the pairs (ξ_n, η_n) are independent. Show that $\sum_n \xi_n$ converges whenever $\sum_n \eta_n$ does.

15. Let $\xi_1, \xi_2, \ldots$ be independent, symmetric random variables. Show that $E\{(\sum_n \xi_n)^2 \wedge 1\} \le \sum_n E(\xi_n^2 \wedge 1)$ whenever the latter series converges. (*Hint:* Integrate over the sets where $\sup_n |\xi_n| \le 1$ or > 1, respectively.)

16. Consider some independent sequences of symmetric random variables $\xi_k, \eta_k^1, \eta_k^2, \ldots$ with $|\eta_k^n| \le |\xi_k|$ such that $\sum_k \xi_k$ converges, and suppose that $\eta_k^n \xrightarrow{P} \eta_k$ for each k. Show that $\sum_k \eta_k^n \xrightarrow{P} \sum_k \eta_k$. (*Hint:* Use a truncation based on the preceding exercise.)

17. Let $\sum_n \xi_n$ be a convergent series of independent random variables. Show that the sum is a.s. independent of the order of terms iff $\sum_n |E(\xi_n;\, |\xi_n| \le 1)| < \infty$.

18. Let the random variables ξ_{nj} be symmetric and independent for each n. Show that $\sum_j \xi_{nj} \xrightarrow{P} 0$ iff $\sum_j E(\xi_{nj}^2 \wedge 1) \to 0$.

19. Let $\xi_n \xrightarrow{d} \xi$ and $a_n \xi_n \xrightarrow{d} \xi$ for a non-degenerate random variable ξ and some constants $a_n > 0$. Show that $a_n \to 1$. (*Hint:* Turning to sub-sequences, we may assume that $a_n \to a$.)

20. Let $\xi_n \xrightarrow{d} \xi$ and $a_n \xi_n + b_n \xrightarrow{d} \xi$ for some non-degenerate random variable ξ, where $a_n > 0$. Show that $a_n \to 1$ and $b_n \to 0$. (*Hint:* Symmetrize.)

21. Let $\xi_1, \xi_2, \ldots$ be independent random variables such that $a_n \sum_{k \le n} \xi_k$ converges in probability for some constants $a_n \to 0$. Show that the limit is degenerate.

22. Show that Theorem 5.23 fails for $p = 2$. (*Hint:* Choose the ξ_k to be independent and $N(0,1)$.)

23. Let $\xi_1, \xi_2, \ldots$ be i.i.d. and such that $n^{-1/p} \sum_{k \le n} \xi_k$ is a.s. bounded for some $p \in (0, 2)$. Show that $E|\xi_1|^p < \infty$. (*Hint:* Argue as in the proof of Theorem 5.23.)

24. For any $p \le 1$, show that the a.s. convergence in Theorem 5.23 remains valid in L^p. (*Hint:* Truncate the ξ_k.)

25. Give an elementary proof of the strong law of large numbers when $E|\xi|^4 < \infty$. (*Hint:* Assuming $E\xi = 0$, show that $E\sum_n (S_n/n)^4 < \infty$.)

26. Show by examples that Theorem 5.25 fails without the stated restrictions on the sets G, F, B.

27. Use Theorem 5.31 to give a simple proof of Theorem 5.27 when S is separable. Generalize to random elements ξ, ξ_n in Borel sets C, C_n, respectively, assuming only $f_n(x_n) \to f(x)$ for $x_n \in C_n$ and $x \in C$ with $x_n \to x$. Extend the original proof to that case.

28. Give a short proof of Theorem 5.31 when $S = \mathsf{R}$. (*Hint:* Note that the distribution functions F_n, F satisfy $F_n^{-1} \to F^{-1}$ a.e. on $[0, 1]$.)

Chapter 6

Gaussian and Poisson Convergence

Characteristic functions and Laplace transforms, equi-continuity and tightness, linear projections, null arrays, Poisson convergence, positive and symmetric terms, central limit theorem, local CLT, Lindeberg condition, Gaussian convergence, weak laws of large numbers, domain of Gaussian attraction, slow variation, Helly's selection theorem, vague and weak convergence, tightness and weak compactness, extended continuity theorem

This is yet another key chapter, dealing with issues of fundamental importance throughout modern probability. Our main focus will be on the Poisson and Gaussian distributions, along with associated limit theorems, extending the classical central limit theorem and related Poisson approximation.

The importance of the mentioned distributions extends far beyond the present context, as will gradually become clear throughout later chapters. Indeed, the Gaussian distributions underlie the construction of Brownian motion, arguably the most important process of the entire subject, whose study leads in turn into stochastic calculus. They also form a basis for the multiple Wiener–Itô integrals, of crucial importance in Malliavin calculus and other subjects. Similarly, the Poisson distributions underlie the construction of Poisson processes, which appear as limit laws for a wide variety of particle systems.

Throughout this chapter, we will use the powerful machinery of characteristic functions and Laplace transforms, which leads to short and transparent proofs of all the main results. Methods based on such functions also extend far beyond the present context. Thus, characteristic functions are used to derive the basic time-change theorems for continuous martingales, and they underlie the use of exponential martingales, which are basic for the Girsanov theorems and related applications of stochastic calculus. Likewise, Laplace transforms play a basic role throughout random-measure theory, and underlie the notion of potentials, of utmost importance in advanced Markov-process theory.

To begin with the basic definitions, consider a random vector ξ in R^d with distribution μ. The associated *characteristic function* $\hat{\mu}$ is given by

$$\hat{\mu}(t) = \int e^{itx}\mu(dx) = E\,e^{it\xi}, \quad t \in \mathsf{R}^d,$$

where tx denotes the inner product $t_1x_1 + \cdots + t_dx_d$. For distributions μ on R^d_+, it is often more convenient to consider the *Laplace transform* $\tilde{\mu}$, given by

$$\tilde{\mu}(u) = \int e^{-ux}\mu(dx) = E\,e^{-u\xi}, \quad u \in \mathsf{R}^d_+.$$

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_7

Finally, for distributions μ on Z_+, it may be preferable to use the (probability) *generating function* ψ, given by

$$\psi(s) = \sum_{n\geq 0} s^n P\{\xi = n\} = E\, s^{\xi}, \quad s \in [0,1].$$

Formally, $\tilde\mu(u) = \hat\mu(iu)$ and $\hat\mu(t) = \tilde\mu(-it)$, and so the functions $\hat\mu$ and $\tilde\mu$ essentially agree apart from domain. Furthermore, the generating function ψ is related to the Laplace transform $\tilde\mu$ by $\tilde\mu(u) = \psi(e^{-u})$ or $\psi(s) = \tilde\mu(-\log s)$. Though the characteristic function always exists, it can't always be extended to an analytic function on the complex plane.

For any distribution μ on R^d, the characteristic function $\varphi = \hat\mu$ is clearly uniformly continuous with $|\varphi(t)| \leq \varphi(0) = 1$. It is also *Hermitian* in the sense that $\varphi(-t) = \bar\varphi(t)$, where the bar denotes complex conjugation. If ξ has characteristic function φ, then the linear combination $a\,\xi = a_1\xi_1 + \cdots + a_d\,\xi_d$ has characteristic function $t \mapsto \varphi(ta)$. Also note that, if ξ and η are independent random vectors with characteristic functions φ and ψ, then the characteristic function of the pair (ξ,η) equals the tensor product $\varphi \otimes \psi\colon (s,t) \mapsto \varphi(s)\psi(t)$. In particular, $\xi + \eta$ has characteristic function $\varphi\,\psi$, and for the symmetrized variable $\xi - \xi'$ it equals $|\varphi|^2$.

Whenever applicable, the quoted statements carry over to Laplace transforms and generating functions. The latter functions have the further advantage of being positive, monotone, convex, and analytic—properties that simplify many arguments.

We list some simple but useful estimates involving characteristic functions. The second inequality was used already in the proof of Theorem 5.17, and the remaining relations will be useful in the sequel to establish tightness.

Lemma 6.1 *(tail estimates) For probability measures μ on R,*

(i) $$\mu\Big\{x;\, |x| \geq r\Big\} \leq \frac{r}{2}\int_{-2/r}^{2/r} (1 - \hat\mu_t)\, dt, \quad r > 0,$$

(ii) $$\mu[-r,r] \leq 2\,r \int_{-1/r}^{1/r} |\hat\mu_t|\, dt, \quad r > 0.$$

If μ is supported by R_+, then also

(iii) $$\mu[r,\infty) \leq 2\Big\{1 - \tilde\mu(r^{-1})\Big\}, \quad r > 0.$$

Proof: (i) Using Fubini's theorem and noting that $\sin x \leq x/2$ for $x \geq 2$, we get for any $c > 0$

$$\begin{aligned}\int_{-c}^{c} (1 - \hat\mu_t)\, dt &= \int \mu(dx) \int_{-c}^{c} (1 - e^{itx})\, dt \\ &= 2c \int \left(1 - \frac{\sin cx}{cx}\right) \mu(dx) \\ &\geq c\,\mu\Big\{x;\, |cx| \geq 2\Big\},\end{aligned}$$

and it remains to take $c = 2/r$.

(ii) Write

$$\begin{aligned}\tfrac{1}{2}\,\mu[-r,r] &\le 2\int \frac{1-\cos(x/r)}{(x/r)^2}\mu(dx)\\ &= r\int \mu(dx)\int (1-r|t|)_+\, e^{ixt}dt\\ &= r\int (1-r|t|)_+\,\hat\mu_t\,dt\\ &\le r\int_{-1/r}^{1/r} |\hat\mu_t|\,dt.\end{aligned}$$

(iii) Noting that $e^{-x} < \frac{1}{2}$ when $x \ge 1$, we get for any $t > 0$

$$\begin{aligned}1-\tilde\mu_t &= \int (1-e^{-tx})\,\mu(dx)\\ &\ge \tfrac{1}{2}\,\mu\Big\{x;\ tx\ge 1\Big\}.\end{aligned}$$

□

Recall that a family of probability measures μ_a on R^d is said to be *tight* if

$$\lim_{r\to\infty}\sup_a \mu_a\Big\{x;\ |x|>r\Big\} = 0.$$

We may characterize tightness in terms of characteristic functions.

Lemma 6.2 *(equi-continuity and tightness) For a family of probability measures μ_a on R^d or R^d_+ with characteristics functions $\hat\mu_a$ or Laplace transforms $\tilde\mu_a$, we have*

(i) (μ_a) *is tight in* R^d $\Leftrightarrow$ $(\hat\mu_a)$ *is equi-continuous at* 0,

(ii) (μ_a) *is tight in* R^d_+ $\Leftrightarrow$ $(\tilde\mu_a)$ *is equi-continuous at* 0.

In that case, $(\hat\mu_a)$ *or* $(\tilde\mu_a)$ *is uniformly equi-continuous on* R^d *or* R^d_+.

Proof: The sufficiency is immediate from Lemma 6.1, applied separately in each coordinate. To prove the necessity, let ξ_a denote a random vector with distribution μ_a, and write for any $s,t\in\mathsf{R}^d$

$$\begin{aligned}|\,\hat\mu_a(s)-\hat\mu_a(t)\,| &\le E\Big|\,e^{is\,\xi_a}-e^{it\,\xi_a}\Big|\\ &= E\Big|\,1-e^{i(t-s)\,\xi_a}\Big|\\ &\le 2\,E\Big(\Big|(t-s)\,\xi_a\Big|\wedge 1\Big).\end{aligned}$$

If $\{\xi_a\}$ is tight, then by Lemma 5.9 the right-hand side tends to 0 as $t-s\to 0$, uniformly in a, and the asserted uniform equi-continuity follows. The proof for Laplace transforms is similar. □

For any probability measures $\mu,\mu_1,\mu_2,\dots$ on R^d, recall that the weak convergence $\mu_n \stackrel{w}{\to} \mu$ is defined by $\mu_n f\to\mu f$ for any bounded, continuous function f on R^d, where μf denotes the integral $\int f d\mu$. The usefulness of characteristic functions is mainly due to the following basic result.

Theorem 6.3 *(characteristic functions, Lévy) Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures on R^d or R_+^d with characteristic functions $\hat\mu, \hat\mu_1, \hat\mu_2, \ldots$ or Laplace transforms $\tilde\mu, \tilde\mu_1, \tilde\mu_2, \ldots$. Then*

(i) $\mu_n \xrightarrow{w} \mu$ *on* $\mathsf{R}^d \iff \hat\mu_n(t) \to \hat\mu(t),\ t \in \mathsf{R}^d,$

(ii) $\mu_n \xrightarrow{w} \mu$ *on* $\mathsf{R}_+^d \iff \tilde\mu_n(u) \to \tilde\mu(u),\ u \in \mathsf{R}_+^d,$

in which case $\hat\mu_n \to \hat\mu$ or $\tilde\mu_n \to \tilde\mu$ uniformly on every bounded set.

In particular, a probability measure on R^d or R_+^d is uniquely determined by its characteristic function or Laplace transform, respectively. For the proof of Theorem 6.3, we need some special cases of the *Stone–Weierstrass approximation theorem*. Here $[0,\infty]$ denotes the compactification of R_+.

Lemma 6.4 *(uniform approximation, Weierstrass)*

(i) *Any continuous function $f\colon \mathsf{R}^d \to \mathsf{R}$ with period 2π in each coordinate admits a uniform approximation by linear combinations of* $\cos kx$ *and* $\sin kx$ *with $k \in \mathsf{Z}_+^d$.*

(ii) *Any continuous function $g\colon [0,\infty]^d \to \mathsf{R}_+$ admits a uniform approximation by linear combinations of functions e^{-kx} with $k \in \mathsf{Z}_+^d$.*

Proof of Theorem 6.3: We consider only (i), the proof of (ii) being similar. If $\mu_n \xrightarrow{w} \mu$, then $\hat\mu_n(t) \to \hat\mu(t)$ for every t, by the definition of weak convergence. By Lemmas 5.8 and 6.2, the latter convergence is uniform on every bounded set.

Conversely, let $\hat\mu_n(t) \to \hat\mu(t)$ for every t. By Lemma 6.1 and dominated convergence, we get for any $a \in \mathsf{R}^d$ and $r > 0$

$$\begin{aligned}\limsup_{n\to\infty} \mu_n\Big\{x;\ |ax| > r\Big\} &\le \lim_{n\to\infty} \frac{r}{2}\int_{-2/r}^{2/r}\Big\{1 - \hat\mu_n(ta)\Big\}\,dt \\ &= \frac{r}{2}\int_{-2/r}^{2/r}\Big\{1 - \hat\mu(ta)\Big\}\,dt.\end{aligned}$$

Since $\hat\mu$ is continuous at 0, the right-hand side tends to 0 as $r \to \infty$, which shows that the sequence (μ_n) is tight. Given any $\varepsilon > 0$, we may then choose $r > 0$ so large that $\mu_n\{|x| > r\} \le \varepsilon$ for all n and $\mu\{|x| > r\} \le \varepsilon$.

Now fix any bounded, continuous function $f\colon \mathsf{R}^d \to \mathsf{R}$, say with $|f| \le c < \infty$. Let f_r denote the restriction of f to the ball $\{|x| \le r\}$, and extend f_r to a continuous function $\tilde f$ on R^d with $|\tilde f| \le c$ and period $2\pi r$ in each coordinate. By Lemma 6.4 there exists a linear combination g of the functions $\cos(kx/r)$ and $\sin(kx/r)$, $k \in \mathsf{Z}_+^d$, such that $|\tilde f - g| \le \varepsilon$. Writing $\|\cdot\|$ for the supremum norm, we get for any $n \in \mathsf{N}$

$$\begin{aligned}\Big|\mu_n f - \mu_n g\Big| &\le \mu_n\{|x| > r\}\|f - \tilde f\| + \|\tilde f - g\| \\ &\le (2c+1)\,\varepsilon,\end{aligned}$$

and similarly for μ. Thus,

$$\left|\mu_n f - \mu f\right| \le \left|\mu_n g - \mu g\right| + 2\,(2c+1)\,\varepsilon, \quad n \in \mathsf{N}.$$

Letting $n \to \infty$ and then $\varepsilon \to 0$, we obtain $\mu_n f \to \mu f$. Since f was arbitrary, this proves that $\mu_n \xrightarrow{w} \mu$. □

The next result often enables us to reduce the d-dimensional case to that of dimension 1.

Corollary 6.5 *(one-dimensional projections, Cramér & Wold) For any random vectors $\xi, \xi_1, \xi_2, \ldots$ in R^d or R_+^d, we have*

(i) $\xi_n \xrightarrow{d} \xi$ *in* $\mathsf{R}^d \Leftrightarrow t\,\xi_n \xrightarrow{d} t\,\xi,\ \ t \in \mathsf{R}^d$,

(ii) $\xi_n \xrightarrow{d} \xi$ *in* $\mathsf{R}_+^d \Leftrightarrow u\,\xi_n \xrightarrow{d} u\,\xi,\ \ u \in \mathsf{R}_+^d$.

In particular, the distribution of a random vector ξ in R^d or R_+^d is uniquely determined by those of all linear combinations $t\xi$ with $t \in \mathsf{R}^d$ or R_+^d, respectively.

Proof: If $t\,\xi_n \xrightarrow{d} t\,\xi$ for all $t \in \mathsf{R}^d$, then $Ee^{it\,\xi_n} \to Ee^{it\,\xi}$ by the definition of weak convergence, and so $\xi_n \xrightarrow{d} \xi$ by Theorem 6.3, proving (i). The proof of (ii) is similar. □

We now apply the continuity Theorem 6.3 to prove some classical limit theorems, beginning with the case of Poisson convergence. To motivate the introduction of the associated distribution, consider for each $n \in \mathsf{N}$ some i.i.d. random variables $\xi_{n1}, \ldots, \xi_{nn}$ with distribution

$$P\{\xi_{nj}=1\} = 1 - P\{\xi_{nj}=0\} = c_n, \quad n \in \mathsf{N},$$

and assume that $n\,c_n \to c < \infty$. Then the sums $S_n = \xi_{n1} + \ldots + \xi_{nn}$ have generating functions

$$\begin{aligned} \psi_n(s) &= \left\{1 - (1-s)\,c_n\right\}^n \\ &\to e^{-c\,(1-s)} \\ &= e^{-c} \sum_{n \ge 0} \frac{c^n s^n}{n!}, \quad s \in [0,1]. \end{aligned}$$

The limit $\psi(s) = e^{-c(1-s)}$ is the generating function of the *Poisson distribution* with parameter c, the distribution of a random variable η with probabilities $P\{\eta = n\} = e^{-c}c^n/n!$ for $n \in \mathsf{Z}_+$. Note that η has expected value $E\eta = \psi'(1) = c$. Since $\psi_n \to \psi$, Theorem 6.3 yields $S_n \xrightarrow{d} \eta$.

To state some more general instances of Poisson convergence, we need to introduce the notion of a *null array*[1]. By this we mean a triangular array of

[1] Its elements are also said to be *uniformly asymptotically negligible*, abbreviated as *u.a.n.*

random variables or vectors ξ_{nj}, $1 \le j \le m_n$, $n \in \mathsf{N}$, that are independent for each n and satisfy

$$\sup_j E\Big(|\xi_{nj}| \wedge 1\Big) \to 0. \tag{1}$$

Informally, this means that $\xi_{nj} \xrightarrow{P} 0$ as $n \to \infty$, uniformly in j. When $\xi_{nj} \ge 0$ for all n and j, we may allow the m_n to be infinite.

The 'null' property can be described as follows in terms of characteristic functions or Laplace transforms:

Lemma 6.6 *(null arrays) Let (ξ_{nj}) be a triangular array in R^d or R^d_+ with characteristic functions $\hat\mu_{nj}$ or Laplace transforms $\tilde\mu_{nj}$. Then*

(i) *(ξ_{nj}) is null in R^d* $\Leftrightarrow$ $\sup_j|1-\hat\mu_{nj}(t)| \to 0,\ t \in \mathsf{R}^d$,

(ii) *(ξ_{nj}) is null in R^d_+* $\Leftrightarrow$ $\inf_j \tilde\mu_{nj}(u) \to 1,\ u \in \mathsf{R}^d_+$.

Proof: Relation (1) holds iff $\xi_{n,j_n} \xrightarrow{P} 0$ for all sequences (j_n). By Theorem 6.3 this is equivalent to $\hat\mu_{n,j_n}(t) \to 1$ for all t and (j_n), which is in turn equivalent to the condition in (i). The proof of (ii) is similar. □

We may now give a general criterion for Poisson convergence of the row sums in a null array of integer-valued random variables. The result will be extended in Theorem 7.14 to more general limiting distributions and in Theorem 30.1 to the context of point processes.

Theorem 6.7 *(Poisson convergence) Let (ξ_{nj}) be a null array of Z_+-valued random variables, and let ξ be a Poisson variable with mean c. Then $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff*

(i) $\sum_j P\{\xi_{nj} > 1\} \to 0$,

(ii) $\sum_j P\{\xi_{nj} = 1\} \to c$.

Here (i) *is equivalent to*

(i′) $\sup_j \xi_{nj} \vee 1 \xrightarrow{P} 1$,

and when $\sum_j \xi_{nj}$ is tight, (i) *holds iff every limit is Poisson.*

We need an elementary lemma. Related approximations will also be used in Chapters 7 and 30.

Lemma 6.8 *(sums and products) Consider a null array of constants $c_{nj} \ge 0$, and fix any $c \in [0,\infty]$. Then*

$$\prod\nolimits_j (1-c_{nj}) \to e^{-c} \quad \Leftrightarrow \quad \sum\nolimits_j c_{nj} \to c.$$

Proof: Since $\sup_j c_{nj} < 1$ for large n, the first relation is equivalent to $\sum_j \log(1-c_{nj}) \to -c$, and the assertion follows since $\log(1-x) = -x + o(x)$ as $x \to 0$. □

Proof of Theorem 6.7: Let ψ_{nj} be the generating function of ξ_{nj}. By Theorem 6.3, the convergence $\sum_j \xi_{nj} \xrightarrow{d} \xi$ is equivalent to $\prod_j \psi_{nj}(s) \to e^{-c(1-s)}$ for arbitrary $s \in [0,1]$, which holds by Lemmas 6.6 and 6.8 iff

$$\sum\nolimits_j \Big\{1 - \psi_{nj}(s)\Big\} \to c(1-s), \quad s \in [0,1]. \tag{2}$$

By an easy computation, the sum on the left equals

$$(1-s)\sum\nolimits_j P\{\xi_{nj} > 0\} + \sum_{k>1}(s - s^k)\sum\nolimits_j P\{\xi_{nj} = k\} = T_1 + T_2, \tag{3}$$

and we also note that

$$s(1-s)\sum\nolimits_j P\{\xi_{nj} > 1\} \le T_2 \le s\sum\nolimits_j P\{\xi_{nj} > 1\}. \tag{4}$$

Assuming (i)−(ii), we note that (2) follows from (3) and (4). Now assume (2). Then for $s = 0$ we get $\sum_j P\{\xi_{nj} > 0\} \to c$, and so in general $T_1 \to c(1-s)$. But then (2) implies $T_2 \to 0$, and (i) follows by (4). Finally, (ii) follows by subtraction.

To see that (i) ⇔ (i′), we note that

$$\begin{aligned} P\Big\{\sup\nolimits_j \xi_{nj} \le 1\Big\} &= \prod\nolimits_j P\{\xi_{nj} \le 1\} \\ &= \prod\nolimits_j \Big(1 - P\{\xi_{nj} > 1\}\Big). \end{aligned}$$

By Lemma 6.8, the right-hand side tends to 1 iff $\sum_j P\{\xi_{nj} > 1\} \to 0$, which is the stated equivalence.

To prove the last assertion, put $c_{nj} = P\{\xi_{nj} > 0\}$, and write

$$\begin{aligned} E\exp\Big(-\sum\nolimits_j \xi_{nj}\Big) - P\Big\{\sup\nolimits_j \xi_{nj} > 1\Big\} &\le E\exp\Big\{-\sum\nolimits_j (\xi_{nj} \wedge 1)\Big\} \\ &= \prod\nolimits_j E\exp\Big\{-(\xi_{nj} \wedge 1)\Big\} \\ &= \prod\nolimits_j \Big\{1 - (1 - e^{-1})\, c_{nj}\Big\} \\ &\le \prod\nolimits_j \exp\Big\{-(1 - e^{-1})\, c_{nj}\Big\} \\ &= \exp\Big\{-(1 - e^{-1})\sum\nolimits_j c_{nj}\Big\}. \end{aligned}$$

If (i) holds and $\sum_j \xi_{nj} \xrightarrow{d} \eta$, then the left-hand side tends to $Ee^{-\eta} > 0$, and so the sums $c_n = \sum_j c_{nj}$ are bounded. Hence, c_n converges along a sub-sequence $N' \subset \mathbb{N}$ toward a constant c. But then (i)−(ii) hold along N', and the first assertion shows that η is Poisson with mean c. □

Next we consider some i.i.d. random variables $\xi_1, \xi_2, \ldots$ with $P\{\xi_k = \pm 1\} = \frac{1}{2}$, and write $S_n = \xi_1 + \cdots + \xi_n$. Then $n^{-1/2}S_n$ has characteristic function

$$\begin{aligned} \varphi_n(t) &= \cos^n(n^{-1/2}\,t) \\ &= \Big\{1 - \tfrac{1}{2}\,t^2 n^{-1} + O(n^{-2})\Big\}^n \\ &\to e^{-t^2/2} = \varphi(t). \end{aligned}$$

By a classical computation, the function $e^{-x^2/2}$ has Fourier transform

$$\int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = (2\pi)^{1/2} e^{-t^2/2}, \quad t \in \mathsf{R}.$$

Hence, φ is the characteristic function of a probability measure on R with density $(2\pi)^{-1/2} e^{-x^2/2}$. This is the standard *normal* or *Gaussian* distribution $N(0,1)$, and Theorem 6.3 shows that $n^{-1/2} S_n \xrightarrow{d} \zeta$, where ζ is $N(0,1)$. The notation is motivated by the facts that $E\zeta = 0$ and $\mathrm{Var}(\zeta) = 1$, where the former relation is obvious by symmetry and latter follows from Lemma 6.9 below. The general Gaussian law $N(m, \sigma^2)$ is defined as the distribution of the random variable $\eta = m + \sigma\zeta$, which has clearly mean m and variance σ^2. From the form of the characteristic functions together with the uniqueness property, we see that any linear combination of independent Gaussian random variables is again Gaussian.

More general limit theorems may be derived from the following technical result.

Lemma 6.9 *(Taylor expansion) Let φ be the characteristic function of a random variable ξ with $E|\xi|^n < \infty$. Then*

$$\varphi(t) = \sum_{k=0}^{n} \frac{(it)^k E\xi^k}{k!} + o(t^n), \quad t \to 0.$$

Proof: Noting that $|e^{it} - 1| \le |t|$ for all $t \in \mathsf{R}$, we get recursively by dominated convergence

$$\varphi^{(k)}(t) = E(i\xi)^k e^{it\xi}, \quad t \in \mathsf{R},\ 0 \le k \le n.$$

In particular, $\varphi^{(k)}(0) = E(i\xi)^k$ for $k \le n$, and the result follows from Taylor's formula. □

The following classical result, known as the *central limit theorem*, explains the importance of the normal distributions. The present statement is only preliminary, and more general versions will be obtained by different methods in Theorems 6.13, 6.16, and 6.18.

Theorem 6.10 *(central limit theorem, Lindeberg, Lévy) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with mean 0 and variance 1, and let ζ be $N(0,1)$. Then*

$$n^{-1/2} \sum_{k \le n} \xi_k \xrightarrow{d} \zeta.$$

Proof: Let the ξ_k have characteristic function φ. By Lemma 6.9, the characteristic function of $n^{-1/2} S_n$ equals

$$\begin{aligned} \varphi_n(t) &= \left\{\varphi(n^{-1/2}t)\right\}^n \\ &= \left\{1 - \tfrac{1}{2} t^2 n^{-1} + o(n^{-1})\right\}^n \\ &\to e^{-t^2/2}, \end{aligned}$$

where the convergence holds as $n \to \infty$ for fixed t. □

We will also establish a stronger density version, needed in Chapter 30. Here we introduce the smoothing densities

$$p_h(x) = (\pi h)^{-d} \prod_{i \le d} \Big(1 - \cos h x_i\Big)/x_i^2, \quad x = (x_1, \ldots, x_d) \in \mathsf{R}^d,$$

with characteristic functions

$$\hat{p}_h(t) = \prod_{i \le d} \Big(1 - |t_i|/h\Big)_+, \quad t = (t_1, \ldots, t_d) \in \mathsf{R}^d.$$

Say that a distribution on R^d is *non-lattice* if it is not supported by any shifted, proper sub-group of R^d.

Theorem 6.11 *(density convergence, Stone) Let μ be a non-lattice distribution on R^d with mean 0 and covariances δ_{ij}, and write φ for the standard normal density on R^d. Then*

$$\lim_{n \to \infty} n^{d/2} \Big\| \mu^{*n} * p_h - \varphi^{*n} \Big\| = 0, \quad h > 0.$$

Proof: Since $|\partial_i \varphi^{*n}| \lesssim n^{-1-d/2}$ for all $i \le d$, the assertion is trivially fulfilled for the standard normal distribution. It is then enough to prove that, for any measures μ and ν as stated,

$$\lim_{n \to \infty} n^{d/2} \Big\| (\mu^{*n} - \nu^{*n}) * p_h \Big\| = 0, \quad h > 0.$$

By Fourier inversion,

$$\Big(\mu^{*n} - \nu^{*n}\Big) * p_h(x) = (2\pi)^{-d} \int e^{ixt}\, \hat{p}_h(t) \Big\{ \hat{\mu}^n(t) - \hat{\nu}^n(t) \Big\}\, dt,$$

and so we get with $I_h = [-h, h]^d$

$$\Big\| (\mu^{*n} - \nu^{*n}) * p_h \Big\| \le (2\pi)^{-d} \int_{I_h} \Big| \hat{\mu}^n(t) - \hat{\nu}^n(t) \Big|\, dt.$$

It remains to show that the integral on the right declines faster than $n^{-d/2}$. For notational convenience we may take $d = 1$, the general case being similar.

By a standard Taylor expansion, as in Lemma 6.9, we have

$$\begin{aligned} \hat{\mu}_n(t) &= \Big\{1 - \tfrac{1}{2}\, t^2 + o(t^2)\Big\}^n \\ &= \exp\Big\{-\tfrac{1}{2}\, n\, t^2 (1 + o(1))\Big\} \\ &= e^{-nt^2/2} \Big\{1 + n\, t^2 o(1)\Big\}, \end{aligned}$$

and similarly for $\hat{\nu}^n(t)$, and so

$$\Big| \hat{\mu}^n(t) - \hat{\nu}^n(t) \Big| = e^{-nt^2/2}\, n\, t^2 o(1).$$

Here clearly

$$n^{1/2}\int e^{-nt^2/2}\,n\,t^2\,dt = \int t^2 e^{-t^2/2}\,dt < \infty,$$

and since μ and ν are non-lattice, we have $|\hat\mu| \vee |\hat\nu| < 1$, uniformly on compacts in $\mathsf{R} \setminus \{0\}$. Writing $I_h = I_\varepsilon \cup (I_h \setminus I_\varepsilon)$, we get for any $\varepsilon > 0$

$$n^{1/2}\int_{I_h} \left|\hat\mu^n(t) - \hat\nu^n(t)\right| dt \lesssim r_\varepsilon + h\,n^{1/2} m_\varepsilon^n,$$

where $r_\varepsilon \to 0$ as $\varepsilon \to 0$, and $m_\varepsilon < 1$ for all $\varepsilon > 0$. The desired convergence now follows, as we let $n \to \infty$ and then $\varepsilon \to 0$. $\square$

We now examine the relationship between null arrays of symmetric and positive random variables. In this context, we also derive criteria for convergence toward Gaussian and degenerate limits, respectively.

Theorem 6.12 *(positive or symmetric terms) Let (ξ_{nj}) be a null array of symmetric random variables, and let ζ be $N(0,c)$ for some $c \geq 0$. Then*

$$\textstyle\sum_j \xi_{nj} \xrightarrow{d} \zeta \quad \Leftrightarrow \quad \sum_j \xi_{nj}^2 \xrightarrow{P} c,$$

where convergence holds iff

(i) $\sum_j P\Big\{|\xi_{nj}| > \varepsilon\Big\} \to 0, \;\; \varepsilon > 0,$

(ii) $\sum_j E\Big(\xi_{nj}^2 \wedge 1\Big) \to c.$

Here (i) *is equivalent to*

(i′) $\sup_j |\xi_{nj}| \xrightarrow{P} 0,$

and when $\sum_j \xi_{nj}$ or $\sum_j \xi_{nj}^2$ is tight, (i) *holds iff every limit is Gaussian or degenerate, respectively.*

The necessity of (i) is remarkable and plays a crucial role in our proof of the more general Theorem 6.16. It is instructive to compare the present statement with the corresponding result for random series in Theorem 5.17. Further note the extended version in Proposition 7.10.

Proof: First let $\sum_j \xi_{nj} \xrightarrow{d} \zeta$. By Theorem 6.3 and Lemmas 6.6 and 6.8, it is equivalent that

$$\textstyle\sum_j E\Big\{1 - \cos(t\,\xi_{nj})\Big\} \to \frac{1}{2}\,c\,t^2, \quad t \in \mathsf{R}, \tag{5}$$

where the convergence is uniform on every bounded interval. Comparing the integrals of (5) over $[0,1]$ and $[0,2]$, we get $\sum_j Ef(\xi_{nj}) \to 0$, where $f(0) = 0$ and

$$f(x) = 3 - \frac{4\sin x}{x} + \frac{\sin 2x}{2x}, \quad x \in \mathsf{R} \setminus \{0\}.$$

Here f is continuous with $f(x) \to 3$ as $|x| \to \infty$, and $f(x) > 0$ for $x \neq 0$. Indeed, the latter relation is equivalent to $8\sin x - \sin 2x < 6\,x$ for $x > 0$, which is obvious when $x \geq \pi/2$ and follows by differentiation twice when $x \in (0, \pi/2)$. Writing $g(x) = \inf_{y>x} f(y)$ and letting $\varepsilon > 0$ be arbitrary, we get

$$\begin{aligned} \textstyle\sum_j P\Big\{|\xi_{nj}| > \varepsilon\Big\} &\leq \textstyle\sum_j P\Big\{f(\xi_{nj}) > g(\varepsilon)\Big\} \\ &\leq \textstyle\sum_j Ef(\xi_{nj})\Big/g(\varepsilon) \to 0, \end{aligned}$$

which proves (i).

If instead $\sum_j \xi_{nj}^2 \xrightarrow{P} c$, the corresponding symmetrized variables η_{nj} satisfy $\sum_j \eta_{nj} \xrightarrow{P} 0$, and so $\sum_j P\{|\eta_{nj}| > \varepsilon\} \to 0$ as before. By Lemma 5.19 it follows that $\sum_j P\{|\xi_{nj}^2 - m_{nj}| > \varepsilon\} \to 0$, where the m_{nj} are medians of ξ_{nj}^2, and since $\sup_j m_{nj} \to 0$, condition (i) follows again. Using Lemma 6.8, we further note that (i) $\Leftrightarrow$ (i′). Thus, we may henceforth assume that (i) is fulfilled.

Next we note that, for any $t \in \mathsf{R}$ and $\varepsilon > 0$,

$$\begin{aligned} &\textstyle\sum_j E\Big\{1 - \cos(t\,\xi_{nj});\ |\xi_{nj}| \leq \varepsilon\Big\} \\ &\qquad = \tfrac{1}{2}\, t^2 \Big\{1 - O(t^2\varepsilon^2)\Big\} \textstyle\sum_j E\Big(\xi_{nj}^2;\ |\xi_{nj}| \leq \varepsilon\Big). \end{aligned}$$

Assuming (i), the equivalence between (5) and (ii) now follows as we let $n \to \infty$ and then $\varepsilon \to 0$. To prove the corresponding result for the variables ξ_{nj}^2, we may write instead, for any $t, \varepsilon > 0$,

$$\textstyle\sum_j E\Big(1 - e^{-t\,\xi_{nj}^2};\ \xi_{nj}^2 \leq \varepsilon\Big) = t\,\Big\{1 - O(t\,\varepsilon)\Big\} \textstyle\sum_j E\Big(\,\xi_{nj}^2;\ \xi_{nj}^2 \leq \varepsilon\Big),$$

and proceed as before. This completes the proof of the first assertion.

Finally, assume that (i) holds and $\sum_j \xi_{nj} \xrightarrow{d} \eta$. Then the same relation holds for the truncated variables $\xi_{nj} 1\{|\xi_{nj}| \leq 1\}$, and so we may assume that $|\xi_{nj}| \leq 1$ for all j and k. Define $c_n = \sum_j E\xi_{nj}^2$. If $c_n \to \infty$ along a sub-sequence, then the distribution of $c_n^{-1/2} \sum_j \xi_{nj}$ tends to $N(0,1)$ by the first assertion, which is impossible by Lemmas 5.8 and 5.9. Thus, (c_n) is bounded and converges along a sub-sequence. By the first assertion, $\sum_j \xi_{nj}$ then tends to a Gaussian limit, and so even η is Gaussian. □

We may now prove a classical criterion for Gaussian convergence, involving normalization by second moments.

Theorem 6.13 *(Gaussian variance criteria, Lindeberg, Feller) Let* (ξ_{nj}) *be a triangular array with* $E\xi_{nj} = 0$ *and* $\sum_j \mathrm{Var}(\xi_{nj}) \to 1$, *and let* ζ *be* $N(0,1)$. *Then these conditions are equivalent:*

(i) $\sum_j \xi_{nj} \xrightarrow{d} \zeta$ *and* $\sup_j \mathrm{Var}(\xi_{nj}) \to 0$,

(ii) $\sum_j E\Big(\xi_{nj}^2;\ |\xi_{nj}| > \varepsilon\Big) \to 0, \quad \varepsilon > 0$.

Here (ii) is the celebrated *Lindeberg condition*. Our proof is based on two elementary lemmas.

Lemma 6.14 *(product comparison) For any complex numbers $z_1, \ldots, z_n$ and $z'_1, \ldots, z'_n$ of modulus ≤ 1, we have*

$$\Big| \prod_k z_k - \prod_k z'_k \Big| \le \sum_k |z_k - z'_k|.$$

Proof: For $n = 2$ we get

$$\begin{aligned} |z_1 z_2 - z'_1 z'_2| &\le |z_1 z_2 - z'_1 z_2| + |z'_1 z_2 - z'_1 z'_2| \\ &\le |z_1 - z'_1| + |z_2 - z'_2|, \end{aligned}$$

and the general result follows by induction. □

Lemma 6.15 *(Taylor expansion) For any $t \in \mathsf{R}$ and $n \in \mathsf{Z}_+$,*

$$\Big| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \Big| \le \frac{2|t|^n}{n!} \wedge \frac{|t|^{n+1}}{(n+1)!}.$$

Proof: Letting $h_n(t)$ denote the difference on the left, we get

$$h_n(t) = i \int_0^t h_{n-1}(s)ds, \quad t > 0, \;\; n \in \mathsf{Z}_+.$$

Starting from the obvious relations $|h_{-1}| \equiv 1$ and $|h_0| \le 2$, we get by induction $|h_{n-1}(t)| \le |t|^n/n!$ and $|h_n(t)| \le 2|t|^n/n!$. □

Returning to the proof of Theorem 6.13, we consider at this point only the sufficiency of condition (ii), needed for the proof of the main Theorem 6.16. The necessity will be established after the proof of that theorem.

Proof of Theorem 6.13, (ii) $\Rightarrow$ (i): Write $c_{nj} = E\,\xi_{nj}^2$ and $c_n = \sum_j c_{nj}$. First note that for any $\varepsilon > 0$,

$$\begin{aligned} \sup_j c_{nj} &\le \varepsilon^2 + \sup_j E\Big(\xi_{nj}^2;\, |\xi_{nj}| > \varepsilon\Big) \\ &\le \varepsilon^2 + \sum_j E\Big(\xi_{nj}^2;\, |\xi_{nj}| > \varepsilon\Big), \end{aligned}$$

which tends to 0 under (ii), as $n \to \infty$ and then $\varepsilon \to 0$.

Now introduce some independent random variables ζ_{nj} with distributions $N(0, c_{nj})$, and note that $\zeta_n = \sum_j \zeta_{nj}$ is $N(0, c_n)$. Hence, $\zeta_n \xrightarrow{d} \zeta$. Letting φ_{nj} and ψ_{nj} denote the characteristic functions of ξ_{nj} and ζ_{nj}, respectively, it remains by Theorem 6.3 to show that $\prod_j \varphi_{nj} - \prod_j \psi_{nj} \to 0$. Then conclude from Lemmas 6.14 and 6.15 that, for fixed $t \in \mathsf{R}$,

$$\begin{aligned} &\Big| \prod_j \varphi_{nj}(t) - \prod_j \psi_{nj}(t) \Big| \\ &\quad \le \sum_j \Big| \varphi_{nj}(t) - \psi_{nj}(t) \Big| \\ &\quad \le \sum_j \Big| \varphi_{nj}(t) - 1 + \tfrac{1}{2}\, t^2 c_{nj} \Big| + \sum_j \Big| \psi_{nj}(t) - 1 + \tfrac{1}{2}\, t^2 c_{nj} \Big| \\ &\quad \lesssim \sum_j E\,\xi_{nj}^2 (1 \wedge |\xi_{nj}|) + \sum_j E\,\zeta_{nj}^2 (1 \wedge |\zeta_{nj}|). \end{aligned}$$

For any $\varepsilon > 0$, we have

$$\sum_j E\,\xi_{nj}^2(1 \wedge |\xi_{nj}|) \le \varepsilon \sum_j c_{nj} + \sum_j E\Big(\xi_{nj}^2;\ |\xi_{nj}| > \varepsilon\Big),$$

which tends to 0 by (ii), as $n \to \infty$ and then $\varepsilon \to 0$. Further note that

$$\begin{aligned}\sum_j E\,\zeta_{nj}^2(1 \wedge |\zeta_{nj}|) &\le \sum_j E|\zeta_{nj}|^3 \\ &= \sum_j c_{nj}^{3/2} E|\xi|^3 \\ &\lesssim c_n \sup_j c_{nj}^{1/2} \to 0,\end{aligned}$$

by the first part of the proof. □

We may now derive precise criteria for convergence to a Gaussian limit. Note the striking resemblance with the three-series criterion in Theorem 5.18. A far-reaching extension of the present result is obtained by different methods in Chapter 7. As before, we write $\mathrm{Var}(\xi; A) = \mathrm{Var}(\xi 1_A)$.

Theorem 6.16 *(Gaussian convergence, Feller, Lévy) Let (ξ_{nj}) be a null array of random variables, and let ζ be $N(b,c)$ for some constants b, c. Then $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff*

(i) $\sum_j P\Big\{|\xi_{nj}| > \varepsilon\Big\} \to 0, \ \ \varepsilon > 0,$

(ii) $\sum_j E\Big(\xi_{nj};\ |\xi_{nj}| \le 1\Big) \to b,$

(iii) $\sum_j \mathrm{Var}\Big(\xi_{nj};\ |\xi_{nj}| \le 1\Big) \to c.$

Here (i) *is equivalent to*

(i′) $\sup_j |\xi_{nj}| \xrightarrow{P} 0,$

and when $\sum_j \xi_{nj}$ is tight, (i) *holds iff every limit is Gaussian.*

Proof: To see that (i) ⇔ (i′), we note that

$$P\Big\{\sup_j |\xi_{nj}| > \varepsilon\Big\} = 1 - \prod_j \Big(1 - P\{|\xi_{nj}| > \varepsilon\}\Big), \quad \varepsilon > 0.$$

Since $\sup_j P\{|\xi_{nj}| > \varepsilon\} \to 0$ under both conditions, the assertion follows by Lemma 6.8.

Now let $\sum_j \xi_{nj} \xrightarrow{d} \zeta$. Introduce medians m_{nj} and symmetrizations $\tilde\xi_{nj}$ of the variables ξ_{nj}, and note that $m_n \equiv \sup_j |m_{nj}| \to 0$ and $\sum_j \tilde\xi_{nj} \xrightarrow{d} \tilde\zeta$, where $\tilde\zeta$ is $N(0, 2c)$. By Lemma 5.19 and Theorem 6.12, we get for any $\varepsilon > 0$

$$\begin{aligned}\sum_j P\Big\{|\xi_{nj}| > \varepsilon\Big\} &\le \sum_j P\Big\{|\xi_{nj} - m_{nj}| > \varepsilon - m_n\Big\} \\ &\le 2\sum_j P\Big\{|\tilde\xi_{nj}| > \varepsilon - m_n\Big\} \to 0.\end{aligned}$$

Thus, we may henceforth assume condition (i) and hence that $\sup_j |\xi_{nj}| \xrightarrow{P} 0$. Then $\sum_j \xi_{nj} \xrightarrow{d} \eta$ is equivalent to $\sum_j \xi'_{nj} \xrightarrow{d} \eta$, where $\xi'_{nj} = \xi_{nj} 1\{|\xi_{nj}| \le 1\}$, and so we may further let $|\xi_{nj}| \le 1$ a.s. for all n and j. Then (ii) and (iii) reduce to $b_n \equiv \sum_j E\,\xi_{nj} \to b$ and $c_n \equiv \sum_j \mathrm{Var}(\xi_{nj}) \to c$, respectively.

Write $b_{nj} = E\,\xi_{nj}$, and note that $\sup_j |b_{nj}| \to 0$ by (i). Assuming (ii)$-$(iii), we get $\sum_j \xi_{nj} - b_n \xrightarrow{d} \zeta - b$ by Theorem 6.13, and so $\sum_j \xi_{nj} \xrightarrow{d} \zeta$. Conversely, $\sum_j \xi_{nj} \xrightarrow{d} \zeta$ implies $\sum_j \tilde\xi_{nj} \xrightarrow{d} \tilde\zeta$, and (iii) follows by Theorem 6.12. But then $\sum_j \xi_{nj} - b_n \xrightarrow{d} \zeta - b$, whence by Lemma 5.20 the b_n converge toward some b'. This gives $\sum_j \xi_{nj} \xrightarrow{d} \zeta + b' - b$, and so $b' = b$, which means that even (ii) is fulfilled.

It remains to prove, under condition (i), that any limiting distribution is Gaussian. Then assume $\sum_j \xi_{nj} \xrightarrow{d} \eta$, and note that $\sum_j \tilde\xi_{nj} \xrightarrow{d} \tilde\eta$, where $\tilde\eta$ is a symmetrization of η. If $c_n \to \infty$ along a sub-sequence, then $c_n^{-1/2} \sum_j \tilde\xi_{nj}$ tends to $N(0,2)$ by the first assertion, which is impossible by Lemma 5.9. Thus, (c_n) is bounded, and so the convergence $c_n \to c$ holds along a sub-sequence. But then $\sum_{nj} \xi_{nj} - b_n$ tends to $N(0,c)$, again by the first assertion, and Lemma 5.20 shows that even b_n converges toward a limit b. Hence, $\sum_{nj} \xi_{nj}$ tends to $N(b,c)$, which is then the distribution of η. □

Proof of Theorem 6.13, (i) ⇒ (ii): The second condition in (i) shows that (ξ_{nj}) is a null array. Furthermore, we have for any $\varepsilon > 0$

$$\begin{aligned}\textstyle\sum_j \mathrm{Var}\big(\xi_{nj};\, |\xi_{nj}| \le \varepsilon\big) &\le \textstyle\sum_j E\big(\xi_{nj}^2;\, |\xi_{nj}| \le \varepsilon\big)\\ &\le \textstyle\sum_j E\,\xi_{nj}^2 \to 1.\end{aligned}$$

By Theorem 6.16 even the left-hand side tends to 1, and (ii) follows. □

Using Theorem 6.16, we may derive the ultimate versions of the weak laws of large numbers. Note that the present conditions are only slightly weaker than those for the strong laws in Theorem 5.23.

Theorem 6.17 *(weak laws of large numbers) Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. random variables, put $S_n = \sum_{k\le n} \xi_k$, and fix a $p \in (0,2)$. Then $n^{-1/p} S_n$ converges in probability iff these conditions hold as $r \to \infty$, depending on the value of p:*

- *for $p \ne 1$: $r^p P\{|\xi| > r\} \to 0$,*
- *for $p = 1$: $rP\{|\xi| > r\} \to 0$ and $E(\xi;\, |\xi| \le r) \to$ some $c \in \mathrm{R}$.*

In that case, the limit equals c when $p = 1$, and is otherwise equal to 0.

Proof: By Theorem 6.16 applied to the null array of random variables $\xi_{nj} = n^{-1/p}\,\xi_j$, $j \le n$, the stated convergence is equivalent to

(i) $n\,P\big\{|\xi| > n^{1/p}\,\varepsilon\big\} \to 0,\ \ \varepsilon > 0,$

(ii) $n^{1-1/p}\,E\big(\xi;\, |\xi| \le n^{1/p}\big) \to c,$

(iii) $n^{1-2/p}\,\mathrm{Var}\big(\xi;\, |\xi| \le n^{1/p}\big) \to 0.$

Here (i) is equivalent to $r^p P\{|\xi| > r\} \to 0$, by the monotonicity of $P\{|\xi| > r^{1/p}\}$. Furthermore, Lemma 4.4 yields for any $r > 0$

$$\begin{aligned} r^{p-2}\,\mathrm{Var}\Big(\xi;\ |\xi| \le r\Big) &\le r^p E\Big((\xi/r)^2 \wedge 1\Big) \\ &= r^p \int_0^1 P\Big\{|\xi| \ge r\sqrt{t}\Big\}\, dt, \\ r^{p-1}\Big|E\Big(\xi;\ |\xi| \le r\Big)\Big| &\le r^p E\Big(|\xi/r| \wedge 1\Big) \\ &= r^p \int_0^1 P\Big\{|\xi| \ge rt\Big\}\, dt. \end{aligned}$$

Since t^{-a} is integrable on $[0,1]$ for any $a < 1$, we get by dominated convergence (i) $\Rightarrow$ (iii), and also (i) $\Rightarrow$ (ii) with $c = 0$ when $p < 1$.

If instead $p > 1$, we see from (i) and Lemma 4.4 that

$$\begin{aligned} E|\xi| \;&=\; \int_0^\infty P\Big\{|\xi| > r\Big\}\, dr \\ &\lesssim\; \int_0^\infty \Big(1 \wedge r^{-p}\Big)\, dr < \infty. \end{aligned}$$

Then $E(\xi;\ |\xi| \le r) \to E\xi$, and (ii) yields $E\xi = 0$. Moreover, (i) implies

$$r^{p-1} E\Big(|\xi|;\ |\xi| > r\Big) = r^p P\Big\{|\xi| > r\Big\} + r^{p-1} \int_r^\infty P\Big\{|\xi| > t\Big\}\, dt \to 0.$$

Assuming in addition that $E\xi = 0$, we obtain (ii) with $c = 0$.

When $p = 1$, condition (i) yields

$$E\Big(|\xi|;\ n < |\xi| \le n+1\Big) \lesssim n\, P\Big\{|\xi| > n\Big\} \to 0.$$

Hence, under (i), condition (ii) is equivalent to $E(\xi;\ |\xi| \le r) \to c$. □

For a further extension of the central limit theorem in Proposition 6.10, we characterize convergence toward a Gaussian limit of suitably normalized partial sums from a single i.i.d. sequence. Here a non-decreasing function $L \ge 0$ is said to *vary slowly at* ∞, if $\sup_x L(x) > 0$ and $L(cx) \sim L(x)$ as $x \to \infty$ for each $c > 0$. This holds in particular when L is bounded, but it is also true for many unbounded functions, such as for $\log(x \vee 1)$.

Theorem 6.18 *(domain of Gaussian attraction, Lévy, Feller, Khinchin) Let* $\xi, \xi_1, \xi_2, \ldots$ *be i.i.d., non-degenerate random variables, and let* ζ *be* $N(0,1)$. *Then these conditions are equivalent:*

(i) *there exist some constants* a_n *and* m_n*, such that*

$$a_n \sum_{k \le n} (\xi_k - m_n) \xrightarrow{d} \zeta,$$

(ii) *the function* $L(x) = E\Big(\xi^2;\ |\xi| \le x\Big)$ *varies slowly at* ∞.

We may then take $m_n \equiv E\xi$. *Finally,* (i) *holds with* $a_n \equiv n^{-1/2}$ *and* $m_n \equiv 0$ *iff* $E\xi = 0$ *and* $E\xi^2 = 1$.

Even other stable distributions may occur as limits, though the convergence criteria are then more complicated. Our proof of Theorem 6.18 is based on a technical result, where for every $m \in \mathsf{R}$ we define

$$L_m(x) = E\Big\{(\xi - m)^2;\ |\xi - m| \le x\Big\}, \quad x > 0.$$

Lemma 6.19 *(slow variation, Karamata) Let ξ be a non-degenerate random variable, such that L_0 varies slowly at ∞. Then so does the function L_m for every $m \in \mathsf{R}$, and as $x \to \infty$,*

$$\frac{x^{2-p}\, E\Big(|\xi|^p;\ |\xi| > x\Big)}{L_0(x)} \to 0, \quad p \in [0, 2). \tag{6}$$

Proof: Fix any constant $r \in (1, 2^{2-p})$, and choose $x_0 > 0$ so large that $L_0(2x) \le r L_0(x)$ for all $x \ge x_0$. For such an x, we get with summation over $n \ge 0$

$$\begin{aligned} x^{2-p} E\Big(|\xi|^p;\ |\xi| > x\Big) &= x^{2-p} \textstyle\sum_n E\Big\{|\xi|^p;\ |\xi|/x \in (2^n, 2^{n+1}]\Big\} \\ &\le \textstyle\sum_n 2^{(p-2)n} E\Big\{\xi^2;\ |\xi|/x \in (2^n, 2^{n+1}]\Big\} \\ &\le \textstyle\sum_n 2^{(p-2)n} (r-1)\, r^n L_0(x) \\ &= \frac{(r-1) L_0(x)}{1 - 2^{p-2}\, r}. \end{aligned}$$

Here (6) follows, as we divide by $L_0(x)$ and let $x \to \infty$ and then $r \to 1$.

In particular, $E|\xi|^p < \infty$ for all $p < 2$. If even $E\xi^2 < \infty$, then $E(\xi - m)^2 < \infty$, and the first assertion is obvious. If instead $E\xi^2 = \infty$, we may write

$$L_m(x) = E\Big(\xi^2;\ |\xi - m| \le x\Big) + m\, E\Big(m - 2\xi;\ |\xi - m| \le x\Big).$$

Here the last term is bounded, and the first term lies between the bounds $L_0(x \pm m) \sim L_0(x)$. Thus, $L_m(x) \sim L_0(x)$, and the slow variation of L_m follows from that of L_0. $\Box$

Proof of Theorem 6.18: Let L vary slowly at ∞. Then so does L_m with $m = E\xi$ by Lemma 6.19, and so we may take $E\xi = 0$. Now define

$$c_n = 1 \vee \sup\Big\{x > 0;\ nL(x) \ge x^2\Big\}, \quad n \in \mathsf{N},$$

and note that $c_n \uparrow \infty$. From the slow variation of L it is further clear that $c_n < \infty$ for all n, and that moreover $nL(c_n) \sim c_n^2$. In particular, $c_n \sim n^{1/2}$ iff $L(c_n) \sim 1$, i.e., iff $\mathrm{Var}(\xi) = 1$.

We shall verify the conditions of Theorem 6.16 with $b = 0$, $c = 1$, and $\xi_{nj} = \xi_j / c_n$, $j \le n$. Beginning with (i), let $\varepsilon > 0$ be arbitrary, and conclude from Lemma 6.19 that

$$n\, P\Big\{|\xi/c_n| > \varepsilon\Big\} \sim \frac{c_n^2 P\{|\xi| > c_n\varepsilon\}}{L(c_n)} \sim \frac{c_n^2 P\{|\xi| > c_n\varepsilon\}}{L(c_n\varepsilon)} \to 0.$$

Recalling that $E\xi = 0$, we get by the same lemma

$$n\left|E\big(\xi/c_n;\ |\xi/c_n| \le 1\big)\right| \le n\,c_n^{-1}E\big(|\xi|;\ |\xi| > c_n\big) \sim \frac{c_n E\big(|\xi|;\ |\xi| > c_n\big)}{L(c_n)} \to 0, \tag{7}$$

which proves (ii). To obtain (iii), we note that by (7)

$$n\,\mathrm{Var}\big(\xi/c_n;\ |\xi/c_n| \le 1\big) = n\,c_n^{-2}\,L(c_n) - n\left\{E\big(\xi/c_n;\ |\xi| \le c_n\big)\right\}^2 \to 1.$$

Hence, Theorem 6.16 yields (i) with $a_n = c_n^{-1}$ and $m_n \equiv 0$.

Now assume (i) for some constants a_n and m_n. Then a corresponding result holds for the symmetrized variables $\tilde\xi, \tilde\xi_1, \tilde\xi_2, \dots$, with constants $a_n/\sqrt{2}$ and 0, and so we may assume that $c_n^{-1}\sum_{k\le n}\tilde\xi_k \xrightarrow{d} \zeta$. Here clearly $c_n \to \infty$, and moreover $c_{n+1} \sim c_n$, since even $c_{n+1}^{-1}\sum_{k\le n}\tilde\xi_k \xrightarrow{d} \zeta$ by Theorem 5.29. Now define for $x > 0$

$$\begin{aligned}
\tilde T(x) &= P\{|\tilde\xi| > x\},\\
\tilde L(x) &= E\big(\tilde\xi^{\,2};\ |\tilde\xi| \le x\big),\\
\tilde U(x) &= E\big(\tilde\xi^{\,2} \wedge x^2\big).
\end{aligned}$$

Then Theorem 6.16 yields $n\,\tilde T(c_n\varepsilon) \to 0$ for all $\varepsilon > 0$, and also $n\,c_n^{-2}\tilde L(c_n) \to 1$. Thus, $c_n^2\,\tilde T(c_n\varepsilon)/\tilde L(c_n) \to 0$, which extends by monotonicity to

$$\frac{x^2\,\tilde T(x)}{\tilde U(x)} \le \frac{x^2\,\tilde T(x)}{\tilde L(x)} \to 0, \quad x \to \infty.$$

Next define for any $x > 0$

$$\begin{aligned}
T(x) &= P\{|\xi| > x\},\\
U(x) &= E(\,\xi^2 \wedge x^2).
\end{aligned}$$

By Lemma 5.19, we have $T(x + |m|) \le 2\,\tilde T(x)$ for any median m of ξ. Furthermore, we get by Lemmas 4.4 and 5.19

$$\begin{aligned}
\tilde U(x) &= \int_0^{x^2} P\{\tilde\xi^{\,2} > t\}\,dt\\
&\le 2\int_0^{x^2} P\{4\,\xi^2 > t\}\,dt\\
&= 8\,U(x/2).
\end{aligned}$$

Hence, as $x \to \infty$,

$$\begin{aligned}
\frac{L(2x) - L(x)}{L(x)} &\le \frac{4\,x^2\,T(x)}{U(x) - x^2\,T(x)}\\
&\le \frac{8\,x^2\,\tilde T(x - |m|)}{8^{-1}\,\tilde U(2x) - 2x^2\,\tilde T(x - |m|)} \to 0,
\end{aligned}$$

which shows that L is slowly varying.

Finally, let $n^{-1/2}\sum_{k\le n}\xi_k \xrightarrow{d} \zeta$. By the previous argument with $c_n = n^{1/2}$ we get $\tilde{L}(n^{1/2}) \to 2$, which implies $E\tilde{\xi}^2 = 2$ and hence $\mathrm{Var}(\xi) = 1$. But then $n^{-1/2}\sum_{k\le n}(\xi_k - E\,\xi) \xrightarrow{d} \zeta$, and so by comparison $E\xi = 0$. □

We return to the general problem of characterizing weak convergence of probability measures μ_n on R^d in terms of the associated characteristic functions $\hat{\mu}_n$ or Laplace transforms $\tilde{\mu}_n$. Suppose that $\hat{\mu}_n$ or $\tilde{\mu}_n$ converges toward a continuous limit φ, which is not known in advance to be a characteristic function or Laplace transform. To conclude that μ_n tends weakly toward some measure μ, we need an extended version of Theorem 6.3, whose proof requires a compactness argument.

Here let $\mathcal{M}_d$ be the space of locally finite measures on R^d, endowed with the *vague topology* generated by the evaluation maps $\pi_f\colon \mu \mapsto \mu f = \int f d\mu$ for all $f \in \hat{C}_+$, where $\hat{C}_+$ is the class of continuous functions $f \ge 0$ on R^d with bounded support. Thus, μ_n converges vaguely to μ, written as $\mu_n \xrightarrow{v} \mu$, iff $\mu_n f \to \mu f$ for all $f \in \hat{C}_+$. If the μ_n are probability measures, then clearly $\|\mu\| = \mu\mathsf{R}^d \le 1$. We prove a version of *Helly's selection theorem*, showing that the set of probability measures on R^d is vaguely relatively sequentially compact.

Theorem 6.20 *(vague compactness, Helly) Every sequence of probability measures on R^d has a vaguely convergent sub-sequence.*

Proof: Fix some probability measures $\mu_1, \mu_2, \ldots$ on R^d, and let $F_1, F_2, \ldots$ be the corresponding distribution functions. Write Q for the set of rational numbers. By a diagonal argument, F_n converges on Q^d along a sub-sequence $N' \subset \mathsf{N}$ toward a limit G, and we may define

$$F(x) = \inf\Big\{G(r);\, r \in \mathsf{Q}^d,\, r > x\Big\}, \quad x \in \mathsf{R}^d. \tag{8}$$

Since each F_n has non-negative increments, so has G and hence also F. We further see from (8) and the monotonicity of G that F is right-continuous. Hence, Corollary 4.26 yields a measure μ on R^d with $\mu(x,y] = F(x,y]$ for any bounded rectangular box $(x,y] \subset \mathsf{R}^d$, and we need to show that $\mu_n \xrightarrow{v} \mu$ along N'.

Then note that $F_n(x) \to F(x)$ at every continuity point x of F. By the monotonicity of F, there exists a countable, dense set $D \subset \mathsf{R}$, such that F is continuous on $C = (D^c)^d$. Then $\mu_n U \to \mu U$ for every finite union U of rectangular boxes with corners in C, and for any bounded Borel set $B \subset \mathsf{R}^d$, a simple approximation yields

$$\begin{aligned}\mu B^o &\le \liminf_{n\to\infty} \mu_n B \\ &\le \limsup_{n\to\infty} \mu_n B \le \mu \bar{B}.\end{aligned} \tag{9}$$

For any bounded μ-continuity set B, we may consider functions $f \in \hat{C}_+$ supported by B, and show as in Theorem 5.25 that $\mu_n f \to \mu f$, proving that $\mu_n \xrightarrow{v} \mu$. □

If $\mu_n \xrightarrow{v} \mu$ for some probability measures μ_n on R^d, we may still have $\|\mu\| < 1$, due to an escape of mass to infinity. To exclude this possibility, we need to assume that (μ_n) is tight.

Lemma 6.21 *(vague and weak convergence) Let $\mu_1, \mu_2, \ldots$ be probability measures on R^d, such that $\mu_n \xrightarrow{v} \mu$ for a measure μ. Then these conditions are equivalent:*

(i) (μ_n) *is tight,*

(ii) $\|\mu\| = 1$,

(iii) $\mu_n \xrightarrow{w} \mu$.

Proof, (i) ⇔ (ii): By a simple approximation, the vague convergence yields (9) for every bounded Borel set B, and in particular for the balls $B_r = \{x \in \mathsf{R}^d;\ |x| < r\}$, $r > 0$. If (ii) holds, then $\mu B_r \to 1$ as $r \to \infty$, and the first inequality gives (i). Conversely, (i) implies $\limsup_n \mu_n B_r \to 1$, and the last inequality yields (ii).

(i) ⇔ (iii): Assume (i), and fix any bounded continuous function $f\colon \mathsf{R}^d \to \mathsf{R}$. For any $r > 0$, we may choose a $g_r \in \hat{C}_+$ with $1_{B_r} \le g_r \le 1$, and note that

$$\begin{aligned} \big|\mu_n f - \mu f\big| &\le \big|\mu_n f - \mu_n f g_r\big| + \big|\mu_n f g_r - \mu f g_r\big| + \big|\mu f g_r - \mu f\big| \\ &\le \big|\mu_n f g_r - \mu f g_r\big| + \|f\|(\mu_n + \mu)B_0^c. \end{aligned}$$

Here the right-hand side tends to zero as $n \to \infty$ and then $r \to \infty$, and so $\mu_n f \to \mu f$, proving (iii). The converse was proved in Lemma 5.8. □

Combining the last two results, we may prove the equivalence of tightness and weak sequential compactness. A more general version appears in Theorem 23.2, which forms the starting point for a theory of weak convergence on function spaces.

Corollary 6.22 *(tightness and weak compactness) For any probability measures $\mu_1, \mu_2, \ldots$ on R^d, these conditions are equivalent:*

(i) (μ_n) *is tight,*

(ii) *every sub-sequence of (μ_n) has a weakly convergent further sub-sequence.*

Proof, (i) ⇒ (ii): By Theorem 6.20, every sub-sequence has a vaguely convergent further sub-sequence. If (μ_n) is tight, then by Lemma 6.21 the convergence holds even in the weak sense.

(ii) ⇒ (i): Assume (ii). If (i) fails, we may choose some $n_k \to \infty$ and $\varepsilon > 0$, such that $\mu_{n_k} B_k^c > \varepsilon$ for all $k \in \mathsf{N}$. By (ii) we have $\mu_{n_k} \xrightarrow{w} \mu$ along a sub-sequence $N' \subset \mathsf{N}$, for a probability measure μ. Then $(\mu_{n_k};\, k \in N')$ is tight by Lemma 5.8, and in particular there exists an $r > 0$ with $\mu_{n_k} B_r^c \le \varepsilon$ for all $k \in N'$. For $k > r$ this is a contradiction, and (i) follows. □

We may now prove the desired extension of Theorem 6.3.

Theorem 6.23 *(extended continuity theorem, Lévy, Bochner) Let $\mu_1, \mu_2, \ldots$ be probability measures on R^d with characteristics functions $\hat\mu_n$. Then these conditions are equivalent:*

(i) *$\hat\mu_n(t) \to \varphi(t)$, $t \in \mathsf{R}^d$, where φ is continuous at 0,*

(ii) *$\mu_n \xrightarrow{w} \mu$ for a probability measure μ on R^d.*

In that case, μ has characteristic function φ. Similar statements hold for distributions μ_n on R_+^d with Laplace transforms $\tilde\mu_n$.

Proof, (ii) ⇒ (i): Clear by Theorem 6.3.

(i) ⇒ (ii): Assuming (i), we see from the proof of Theorem 6.3 that (μ_n) is tight. Then Corollary 6.22 yields $\mu_n \xrightarrow{w} \mu$ along a sub-sequence $N' \subset \mathsf{N}$, for a probability measure μ. The convergence extends to N by Theorem 6.3, proving (ii). The proof for Laplace transforms is similar. □

Exercises

1. Show that if ξ, η are independent Poisson random variables, then $\xi + \eta$ is again Poisson. Also show that the Poisson property is preserved by convergence in distribution.

2. Show that any linear combination of independent Gaussian random variables is again Gaussian. Also show that the Gaussian property is preserved by convergence in distribution.

3. Show that $\varphi_r(t) = (1 - t/r)_+$ is a characteristic functions for every $r > 0$. (*Hint:* Calculate the Fourier transform $\hat\psi_r$ of the function $\psi_r(t) = 1\{|t| \le r\}$, and note that the Fourier transform $\hat\psi_r^2$ of ψ_r^{*2} is integrable. Now use Fourier inversion.)

4. Let φ be a real, even function that is convex on R_+ and satisfies $\varphi(0) = 1$ and $\varphi(\infty) \in [0, 1]$. Show that φ is the characteristic function of a symmetric distribution on R. In particular, $\varphi(t) = e^{-|t|^c}$ is a characteristic function for every $c \in [0, 1]$. (*Hint:* Approximate by convex combinations of functions φ_r as above, and use Theorem 6.23.)

5. Show that if $\hat\mu$ is integrable, then μ has a bounded, continuous density. (*Hint:* Let φ_r be the triangular density above. Then $(\hat\varphi_r)\hat{}= 2\pi\varphi_r$, and so $\int e^{-itu}\, \hat\mu_t\, \hat\varphi_r(t)\, dt = 2\pi \int \varphi_r(x - u)\, \mu(dx)$. Now let $r \to 0$.)

6. Show that a distribution μ is supported by a set $a\,\mathsf{Z} + b$ iff $|\hat\mu_t| = 1$ for some $t \ne 0$.

7. Give a direct proof of the continuity theorem for generating functions of distributions on Z_+. (*Hint:* Note that if $\mu_n \xrightarrow{v} \mu$ for some distributions on R_+, then $\tilde{\mu}_n \to \tilde{\mu}$ on $(0,\infty)$.)

8. The *moment-generating function* of a distribution μ on R is given by $\tilde{\mu}_t = \int e^{tx}\mu(dx)$. Show that if $\tilde{\mu}_t < \infty$ for all t in a non-degenerate interval I, then $\tilde{\mu}$ is analytic in the strip $\{z \in \mathsf{C};\, \Re z \in I^o\}$. (*Hint:* Approximate by measures with bounded support.)

9. Let $\mu, \mu_1, \mu_2, \ldots$ be distributions on R with moment-generating functions $\tilde{\mu}, \tilde{\mu}_1, \tilde{\mu}_2, \ldots$, such that $\tilde{\mu}_n \to \tilde{\mu} < \infty$ on a non-degenerate interval I. Show that $\mu_n \xrightarrow{w} \mu$. (*Hint:* If $\mu_n \xrightarrow{v} \nu$ along a sub-sequence N', then $\tilde{\mu}_n \to \tilde{\nu}$ on I^o along N', and so $\tilde{\nu} = \tilde{\mu}$ on I. By the preceding exercise, we get $\nu\mathsf{R} = 1$ and $\hat{\nu} = \hat{\mu}$. Thus, $\nu = \mu$.)

10. Let μ, ν be distributions on R with finite moments $\int x^n \mu(dx) = \int x^n \nu(dx) = m_n$, where $\sum_n t^n |m_n|/n! < \infty$ for some $t > 0$. Show that $\mu = \nu$. (*Hint:* The absolute moments satisfy the same relation for any smaller value of t, and so the moment-generating functions exist and agree on $(-t, t)$.)

11. For each $n \in \mathsf{N}$, let μ_n be a distribution on R with finite moments m_n^k, $k \in \mathsf{N}$, such that $\lim_n m_n^k = a_k$ for some constants a_k with $\sum_k t^k |a_k|/k! < \infty$ for a $t > 0$. Show that $\mu_n \xrightarrow{w} \mu$ for a distribution μ with moments a_k. (*Hint:* Each function x^k is uniformly integrable with respect to the measures μ_n. In particular, (μ_n) is tight. If $\mu_n \xrightarrow{w} \nu$ along a sub-sequence, then ν has moments a_k.)

12. Given a distribution μ on $\mathsf{R} \times \mathsf{R}_+$, introduce the Fourier–Laplace transform $\varphi(s,t) = \int e^{isx - ty}\mu(dx\,dy)$, where $s \in \mathsf{R}$ and $t \ge 0$. Prove versions for φ of Theorems 6.3 and 6.23.

13. Consider a null array of random vectors $\xi_{nj} = (\xi_{nj}^1, \ldots, \xi_{nj}^d)$ in Z_+^d, let $\xi^1, \ldots, \xi^d$ be independent Poisson variables with means $c_1, \ldots, c_d$, and put $\xi = (\xi^1, \ldots, \xi^d)$. Show that $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff $\sum_j P\{\xi_{nj}^k = 1\} \to c_k$ for all k and $\sum_j P\{\sum_k \xi_{nj}^k > 1\} \to 0$. (*Hint:* Introduce some independent random variables $\eta_{nj}^k \overset{d}{=} \xi_{nj}^k$, and note that $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff $\sum_j \eta_{nj} \xrightarrow{d} \xi$.)

14. Consider some random variables $\xi \perp\!\!\!\perp \eta$ with finite variance, such that the distribution of (ξ, η) is rotationally invariant. Show that ξ is centered Gaussian. (*Hint:* Let $\xi_1, \xi_2, \ldots$ be i.i.d. and distributed as ξ, and note that $n^{-1/2}\sum_{k \le n} \xi_k$ has the same distribution for all n. Now use Proposition 6.10.)

15. Prove a multi-variate version of the Taylor expansion in Lemma 6.9.

16. Let μ have a finite n-th moment m_n. Show that $\hat{\mu}$ is n times continuously differentiable and satisfies $\hat{\mu}_0^{(n)} = i^n m_n$. (*Hint:* Differentiate n times under the integral sign.)

17. For μ and m_n as above, show that $\hat{\mu}_0^{(2n)}$ exists iff $m_{2n} < \infty$. Also characterize the distributions μ where $\hat{\mu}_0^{(2n-1)}$ exists. (*Hint:* For $\hat{\mu}_0''$, proceed as in the proof of Proposition 6.10, and use Theorem 6.18. For $\hat{\mu}_0'$, use Theorem 6.17. Extend by induction to $n > 1$.)

18. Let μ be a distribution on R_+ with moments m_n. Show that $\tilde{\mu}_0^{(n)} = (-1)^n m_n$, whenever either side exists and is finite. (*Hint:* Prove the statement for $n = 1$, and extend by induction.)

19. Deduce Proposition 6.10 from Theorem 6.13.

20. Let the random variables ξ and ξ_{nj} be such as in Theorem 6.13, and assume that $\sum_j E|\xi_{nj}|^p \to 0$ for a $p > 2$. Show that $\sum_j \xi_{nj} \xrightarrow{d} \xi$.

21. Extend Theorem 6.13 to random vectors in R^d, with the condition $\sum_j E\xi_{nj}^2 \to 1$ replaced by $\sum_j \mathrm{Cov}(\xi_{nj}) \to a$, with ξ as $N(0,a)$, and with ξ_{nj}^2 replaced by $|\xi_{nj}|^2$. (*Hint:* Use Corollary 6.5 to reduce to dimension 1.)

22. Show that Theorem 6.16 remains true for random vectors in R^d, with $\mathrm{Var}(\xi_{nj};\, |\xi_{nj}| \le 1)$ replaced by the corresponding covariance matrix. (*Hint:* If $a, a_1, a_2, \ldots$ are symmetric, non-negative definite matrices, then $a_n \to a$ iff $u'a_n u \to u'au$ for all $u \in \mathsf{R}^d$. To see this, use a compactness argument.)

23. Show that Theorems 6.7 and 6.16 remain valid for possibly infinite row-sums $\sum_j \xi_{nj}$. (*Hint:* Use Theorem 5.17 or 5.18, together with Theorem 5.29.)

24. Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. random variables. Show that $n^{-1/2}\sum_{k\le n}\xi_k$ converges in probability iff $\xi = 0$ a.s. (*Hint:* Use condition (iii) in Theorem 6.16.)

25. Let $\xi_1, \xi_2, \ldots$ be i.i.d. μ, and fix any $p \in (0,2)$. Find a μ such that $n^{-1/p}\sum_{k\le n}\xi_k \to 0$ in probability but not a.s.

26. Let $\xi_1, \xi_2, \ldots$ be i.i.d., and let $p > 0$ be such that $n^{-1/p}\sum_{k\le n}\xi_k \to 0$ in probability but not a.s. Show that $\limsup_n n^{-1/p}\,|\sum_{k\le n}\xi_k| = \infty$ a.s. (*Hint:* Note that $E|\xi_1|^p = \infty$.)

27. Give an example of a distribution on R with infinite second moment, belonging to the domain of attraction of the Gaussian law. Also find the corresponding normalization.

Chapter 7

Infinite Divisibility and General Null Arrays

Compound Poisson distributions and approximation, i.i.d. and null arrays, infinitely divisible distributions, Lévy measure, characteristics, Lévy–Khinchin formula, convergence and closure, one-dimensional criteria, positive and symmetric terms, general null arrays, limit laws and convergence criteria, sums and extremes, diffuse distributions

The fundamental roles of Gaussian and Poisson distributions should be clear from Chapter 6. Here we consider the more general and equally basic family of infinitely divisible distributions, which may be defined as distributional limits of linear combinations of independent Poisson and Gaussian random variables. Such distributions constitute the most general limit laws appearing in the classical limit theorems for null arrays. They further appear as finite-dimensional distributions of Lévy processes—the fundamental processes with stationary, independent increments, treated in Chapter 16—which may be regarded as prototypes of general Markov processes.

The special criteria for convergence toward Poisson and Gaussian distributions, derived by simple analytic methods in Chapter 6, will now be extended to the case of general infinitely divisible limits. Though the use of some analytic tools is still unavoidable, the present treatment is more probabilistic in flavor, and involves as crucial steps a centering at truncated means followed by a compound Poisson approximation. This approach has the advantage of providing some better insight into even the Gaussian and Poisson results, previously obtained by elementary but more technical estimates.

Since the entire theory is based on approximations by compound Poisson distributions, we begin with some characterizations of the latter. Here the basic representation is in terms of Poisson processes, discussed more extensively in Chapter 15. Here we need only the basic definition in terms of independent, Poisson distributed increments.

Lemma 7.1 *(compound Poisson distributions) Let ξ be a random vector in* R^d*, fix any bounded measure $\nu \neq 0$ on $\mathsf{R}^d \setminus \{0\}$, and put $c = \|\nu\|$ and $\hat\nu = \nu/c$. Then these conditions are equivalent:*

(i) *$\xi \overset{d}{=} X_\kappa$ for a random walk $X = (X_n)$ in R^d based on $\hat\nu$, where $\kappa \perp\!\!\!\perp X$ is Poisson distributed with $E\kappa = c$,*

(ii) *$\xi \overset{d}{=} \int x\,\eta(dx)$ for a Poisson process η on $\mathsf{R}^d \setminus \{0\}$ with $E\eta = \nu$,*

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_8

(iii) $\log E e^{iu\xi} = \int (\,e^{iux} - 1)\,\nu(dx), \quad u \in \mathsf{R}^d.$

The measure ν is then uniquely determined by $\mathcal{L}(\xi)$.

Proof, (i) ⇒ (iii): Writing $Ee^{iuX_1} = \varphi(u)$, we get

$$\begin{aligned} Ee^{iuX_\kappa} &= \sum_{n\geq 0} e^{-c}\,\frac{c^n}{n!}\,\{\varphi(u)\}^n \\ &= e^{-c} e^{c\,\varphi(u)} \\ &= e^{c\{\varphi(u)-1\}}, \end{aligned}$$

and so

$$\begin{aligned} \log Ee^{iuX_\kappa} &= c\Big\{\varphi(u) - 1\Big\} \\ &= c\Big\{\int e^{iux}\hat{\nu}(dx) - 1\Big\} \\ &= \int (\,e^{iux} - 1)\,\nu(dx). \end{aligned}$$

(ii) ⇒ (iii): To avoid repetitions, we defer the proof until Lemma 15.2 (ii), which remains valid with $-f$ replaced by the function $f(x) = ix$.

(iii) ⇒ (i)–(ii): Clear by the uniqueness in Theorem 6.3. □

The distributions in Lemma 7.1 are said to be *compound Poisson*, and the underlying measure ν is called the *characteristic measure* of ξ. We will use the compound Poisson distributions to approximate the row sums in triangular arrays (ξ_{nj}), where the ξ_{nj} are understood to be independent in j for fixed n. Recall that (ξ_{nj}) is called a *null array* if $\xi_{nj} \xrightarrow{P} 0$ as $n \to \infty$, uniformly in j. It is further called an *i.i.d. array* if the ξ_{nj} are i.i.d. in j for fixed n, and the successive rows have lengths $m_n \to \infty$.

For any random vector ξ with distribution μ, we introduce an *associated* compound Poisson random vector $\tilde{\xi}$ with characteristic measure μ. For a triangular array ξ_{nj}, the corresponding compound Poisson vectors $\tilde{\xi}_{ij}$ are again assumed to have row-wise independent entries. By $\xi_n \overset{d}{\sim} \eta_n$ we mean that, if either side converges in distribution along a sub-sequence, then so does the other along the same sequence, and the two limits agree.

Proposition 7.2 *(compound Poisson approximation) Let (ξ_{nj}) be a triangular array in R^d with associated compound Poisson array $(\tilde{\xi}_{nj})$. Then the equivalence $\sum_j \xi_{nj} \overset{d}{\sim} \sum_j \tilde{\xi}_{nj}$ holds under each of these conditions:*

(i) *(ξ_{nj}) is a null array in R^d_+,*

(ii) *(ξ_{nj}) is a null array with $\xi_{nj} \overset{d}{=} -\xi_{nj}$,*

(iii) *(ξ_{nj}) is an i.i.d. array.*

The stated equivalence fails for general null arrays, where a modified version is given in Theorem 7.11. For the proof, we first need to show that the 'null' property holds even in case (iii). This requires a simple technical lemma:

Lemma 7.3 *(i.i.d. and null arrays) Let* (ξ_{nj}) *be an i.i.d. array in* R^d, *such that the sequence* $\xi_n = \sum_j \xi_{nj}$ *is tight. Then* $\xi_{n1} \xrightarrow{P} 0$.

Proof: It is enough to take $d = 1$. By Lemma 6.2 the functions $\hat{\mu}_n(u) = Ee^{iu\xi_n}$ are equi-continuous at 0, and so $\hat{\mu}_n = e^{\psi_n}$ on a suitable interval $I = [-a, a]$, for some continuous functions ψ_n with $|\psi_n| \le \frac{1}{2}$. Writing $\hat{\mu}_{nj}(r) = Ee^{ir\xi_{nj}}$ for $j \le m_n$, we conclude by continuity that $\hat{\mu}_{nj} = e^{\psi_n/m_n} \to 1$, uniformly on I. Proceeding as in Lemma 6.1 (i), we get for any $\varepsilon \le a^{-1}$

$$\begin{aligned}\int_{-a}^{a} \Big\{1 - \hat{\mu}_{nj}(r)\Big\}\, dr &= 2r \int \Big(1 - \frac{\sin ax}{ax}\Big)\, \mu_{nj}(dx)\\ &\ge 2r\Big(1 - \frac{\sin a\varepsilon}{a\varepsilon}\Big)\, \mu_{nj}\{|x| \ge \varepsilon\},\end{aligned}$$

and so $P\{|\xi_{nj}| \ge \varepsilon\} \to 0$ as $n \to \infty$, which implies $\xi_{nj} \xrightarrow{P} 0$. □

For the main proof, we often need the elementary fact

$$e^r - 1 \sim r \ \text{ as } \ r \to 0, \tag{1}$$

which holds when $r \uparrow 0$ or $r \downarrow 0$, and even when $r \to 0$ in the complex plane.

Proof of Proposition 7.2: Let $\hat{\mu}_{nj}$ or $\tilde{\mu}_{nj}$ be the characteristic functions or Laplace transforms of ξ_{nj}. Since the latter form a null array, even in case (iii) by Lemma 7.3, Lemma 6.6 ensures that, for every $r > 0$, there exists an $m \in \mathsf{N}$ such that $\hat{\mu}_{nj} \ne 0$ or $\tilde{\mu}_{nj} \ne 0$ on the ball B_0^r for all $n > m$. Then $\hat{\mu}_{nj} = e^{\psi_{nj}}$ or $\tilde{\mu}_{nj} = e^{-\varphi_{nj}}$, for some continuous functions ψ_{nj} and φ_{nj}, respectively.

(i) For any $u \in \mathsf{R}_+^d$,

$$\begin{aligned}-\log Ee^{-u\xi_{nj}} &= -\log \tilde{\mu}_{nj}(u)\\ &= \varphi_{nj}(u),\\ -\log Ee^{-u\tilde{\xi}_{nj}} &= 1 - \tilde{\mu}_{nj}(u)\\ &= 1 - e^{-\varphi_{nj}(u)},\end{aligned}$$

and we need to show that

$$\textstyle\sum_j \varphi_{nj}(u) \approx \sum_j \Big(1 - e^{-\varphi_{nj}(u)}\Big), \quad u \in \mathsf{R}_+^d,$$

in the sense that if either side converges to a positive limit, then so does the other and the limits are equal. Since all functions are positive, this is clear from (1).

(ii) For any $u \in \mathsf{R}^d$,

$$\begin{aligned}\log Ee^{iu\xi_{nj}} &= \log \hat{\mu}_{nj}(u)\\ &= \psi_{nj}(u),\\ \log Ee^{iu\tilde{\xi}_{nj}} &= \hat{\mu}_{nj}(u) - 1\\ &= e^{\psi_{nj}(u)} - 1,\end{aligned}$$

and we need to show that

$$\sum_j \psi_{nj}(u) \approx \sum_j \left(e^{\psi_{nj}(u)} - 1\right), \quad u \in \mathsf{R}^d,$$

in the sense of simultaneous convergence to a common finite limit. Since all functions are real and ≤ 0, this follows from (1).

(iii) Writing $\psi_{nj} = \psi_n$, we may reduce our claim to

$$\exp\left\{m_n\, \psi_n(u)\right\} \approx \exp\left\{\, m_n\left(e^{\psi_n(u)} - 1\right)\right\}, \quad u \in \mathsf{R}^d,$$

in the sense of simultaneous convergence to a common complex number. This follows from (1). □

A random vector ξ or its distribution $\mathcal{L}(\xi)$ is said to be *infinitely divisible*, if for every $n \in \mathsf{N}$ we have $\xi \overset{d}{=} \sum_j \xi_{nj}$ for some i.i.d. random vectors $\xi_{n1}, \ldots, \xi_{nn}$. Obvious examples of infinitely divisible distributions include the Gaussian and Poisson laws, which is clear by elementary calculations. More generally, we see from Lemma 7.1 that any compound Poisson distribution is infinitely divisible. The following approximation is fundamental:

Corollary 7.4 *(compound Poisson limits) For a random vector ξ in R^d, these conditions are equivalent:*

(i) *ξ is infinitely divisible,*

(ii) *$\xi_n \xrightarrow{d} \xi$ for some compound Poisson random vectors ξ_n.*

Proof, (i) ⇒ (ii): Under (i) we have $\xi \overset{d}{=} \sum_j \xi_{nj}$ for an i.i.d. array (ξ_{nj}). Choosing an associated compound Poisson array $(\tilde{\xi}_{nj})$, we get $\sum_j \tilde{\xi}_{nj} \xrightarrow{d} \xi$ by Lemma 7.2 (iii), and we note that $\xi_n = \sum_j \tilde{\xi}_{nj}$ is again compound Poisson.

(ii) ⇒ (i): Assume (ii) for some ξ_n with characteristic measures ν_n. For every $m \in \mathsf{N}$, we may write $\xi_n \overset{d}{=} \sum_{j \leq m} \xi_{nj}$ for some i.i.d. compound Poisson vectors ξ_{nj} with characteristic measures ν_n/m. Then the sequence $(\xi_{n1}, \ldots, \xi_{nm})$ is tight by Lemma 6.2, and so $(\xi_{n1}, \ldots, \xi_{nm}) \xrightarrow{d} (\zeta_1, \ldots, \zeta_m)$ as $n \to \infty$ along a sub-sequence, where the ζ_j are again i.i.d. with sum $\sum_j \zeta_j \overset{d}{=} \xi$. Since m was arbitrary, (i) follows. □

We proceed to characterize the general infinitely divisible distributions. For a special case, we may combine the Gaussian and compound Poisson distributions and use Lemma 7.1 (ii) to form a variable $\xi \overset{d}{=} b + \zeta + \int x\, \eta(dx)$, involving a constant vector b, a centered Gaussian random vector ζ, and an independent Poisson process η on R^d. For extensions to unbounded $\nu = E\eta$, we need to apply a suitable centering, to ensure convergence of the Poisson integral. This suggests the more general representation:

Theorem 7.5 *(infinitely divisible distributions, Lévy, Itô)*

(i) *A random vector ξ in R^d is infinitely divisible iff*

$$\xi \stackrel{d}{=} b + \zeta + \int_{|x|\le 1} x\,(\eta - E\eta)(dx) + \int_{|x|>1} x\,\eta(dx),$$

for a constant vector $b \in \mathsf{R}^d$, a centered Gaussian random vector ζ with $\mathrm{Cov}(\zeta) = a$, and an independent Poisson process η on $\mathsf{R}^d \setminus \{0\}$ with intensity ν satisfying $\int (|x|^2 \wedge 1)\,\nu(dx) < \infty$.

(ii) *A random vector ξ in R^d_+ is infinitely divisible iff*

$$\xi \stackrel{d}{=} a + \int x\,\eta(dx),$$

for a constant vector $a \in \mathsf{R}^d_+$ and a Poisson process η on $\mathsf{R}^d_+ \setminus \{0\}$ with intensity ν satisfying $\int (|x| \wedge 1)\,\nu(dx) < \infty$.

The triple (a, b, ν) or pair (a, ν) is then unique, and any choice with the stated properties may occur.

This shows that an infinitely divisible distribution on R^d or R^d_+ is uniquely characterized by the triple (a, b, ν) or pair (a, ν), referred to as the *characteristics* of ξ or $\mathcal{L}(\xi)$. The intensity measure ν of the Poisson component is called the *Lévy measure* of ξ or $\mathcal{L}(\xi)$. The representations of Theorem 7.5 can also be expressed analytically in terms of characteristic functions or Laplace transforms, which leads to some celebrated formulas:

Corollary 7.6 *(Lévy–Khinchin representations, Kolmogorov, Lévy) Let ξ be an infinitely divisible random vector in R^d or R^d_+ with characteristics (a, b, ν) or (a, ν). Then*

(i) *for R^d-valued ξ we have $E\,e^{iu\xi} = e^{\psi(u)}$, $u \in \mathsf{R}^d$, where*

$$\psi(u) = iu'b - \tfrac{1}{2}\,u'au + \int \Big(e^{iu'x} - 1 - iu'x\,1\{|x| \le 1\} \Big)\,\nu(dx),$$

(ii) *for R^d_+-valued ξ we have $E\,e^{-u\xi} = e^{-\varphi(u)}$, $u \in \mathsf{R}^d_+$, where*

$$\varphi(u) = ua + \int (1 - e^{-ux})\,\nu(dx).$$

The stated representations will be proved together with the associated convergence criteria, stated below. Given a characteristic triple (a, b, ν), as above, we define for any $h > 0$ the truncated versions

$$a^h = a + \int_{|x|\le h} xx'\,\nu(dx),$$
$$b^h = b - \int_{h<|x|\le 1} x\,\nu(dx),$$

where $\int_{h<|x|\le 1} = -\int_{1<|x|\le h}$ when $h > 1$. In the positive case, we put instead $a^h = a + \int_{x\le h} x\,\nu(dx)$. Let $\overline{\mathsf{R}^d}$ denote the one-point compactification of R^d.

Theorem 7.7 *(convergence and closure)*

(i) *The class of infinitely divisible distributions on* $\mathbb{R}^d$ *is closed under weak convergence.*

(ii) *Let* ξ_n *and* ξ *be infinitely divisible in* $\mathbb{R}^d$ *with characteristics* (a_n, b_n, ν_n) *and* (a, b, ν)*, and fix any* $h > 0$ *with* $\nu\, \partial B_0^h = 0$*. Then* $\xi_n \xrightarrow{d} \xi$ *iff*[1]

$$a_n^h \to a^h, \qquad b_n^h \to b^h, \qquad \nu_n \xrightarrow{v} \nu \ \text{ on } \overline{\mathbb{R}^d} \setminus \{0\}.$$

(iii) *Let* ξ_n *and* ξ *be infinitely divisible in* $\mathbb{R}_+^d$ *with characteristics* (a_n, ν_n) *and* (a, ν)*, and fix any* $h > 0$ *with* $\nu\, \partial B_0^h = 0$*. Then* $\xi_n \xrightarrow{d} \xi$ *iff*

$$a_n^h \to a^h, \qquad \nu_n \xrightarrow{v} \nu \ \text{ on } \overline{\mathbb{R}_+^d} \setminus \{0\}.$$

We first consider the one-dimensional case, which allows some significant simplifications. The characteristic exponent ψ in Theorem 7.6 may then be written as

$$\psi(r) = i\,c\,r + \int \left(e^{irx} - 1 - \frac{i\,rx}{1+x^2} \right) \frac{1+x^2}{x^2}\, \tilde{\nu}(dx), \quad r \in \mathbb{R}, \tag{2}$$

where

$$\tilde{\nu}(dx) = \sigma^2 \delta_0(dx) + \frac{x^2}{1+x^2}\, \nu(dx),$$
$$c = b + \int \left(\frac{x}{1+x^2} - x\, 1\{|x| \le 1\} \right) \nu(dx),$$

and the integrand in (2) is defined by continuity as $-\frac{1}{2} r^2$ when $x = 0$.

For infinitely divisible distributions on $\mathbb{R}_+$, we define instead

$$\tilde{\nu}(dx) = a\,\delta_0 + (1 - e^{-x})\, \nu(dx),$$

so that the characteristic exponent φ in Theorem 7.6 becomes

$$\varphi(r) = \int \frac{1 - e^{-rx}}{1 - e^{-x}}\, \tilde{\nu}(dx), \quad r \ge 0, \tag{3}$$

where the integrand is interpreted as r when $x = 0$. We refer to the pair $(c, \tilde{\nu})$ or measure $\tilde{\nu}$ as the *modified characteristics* of ξ or $\mathcal{L}(\xi)$.

Lemma 7.8 *(one-dimensional criteria)*

(i) *Let* ξ_n *and* ξ *be infinitely divisible in* $\mathbb{R}$ *with modified characteristics* $(c_n, \tilde{\nu}_n)$ *and* $(c, \tilde{\nu})$*. Then*

$$\xi_n \xrightarrow{d} \xi \quad \Leftrightarrow \quad c_n \to c, \ \ \tilde{\nu}_n \xrightarrow{w} \tilde{\nu}.$$

(ii) *Let* ξ_n *and* ξ *be infinitely divisible in* $\mathbb{R}_+$ *with modified characteristics* $\tilde{\nu}_n$ *and* $\tilde{\nu}$*. Then*

$$\xi_n \xrightarrow{d} \xi \quad \Leftrightarrow \quad \tilde{\nu}_n \xrightarrow{w} \tilde{\nu}.$$

---

[1]Here $\nu_n \xrightarrow{v} \nu$ means that $\nu_n f \to \nu f$ for all $f \in \hat{C}_+(\overline{\mathbb{R}^d} \setminus \{0\})$, the class of continuous functions $f \ge 0$ on $\mathbb{R}^d$ with a finite limit as $|x| \to \infty$, and such that $f(x) = 0$ in a neighborhood of 0. The meaning in (iii) is similar.

The last two statements will first be proved with 'infinitely divisible' replaced by *representable*, meaning that ξ_n and ξ are random vectors allowing representations as in Theorem 7.5. Some earlier lemmas will then be used to deduce the stated versions.

Proof of Lemma 7.8, *representable case:* (ii) If $\tilde{\nu}_n \xrightarrow{w} \nu$, then $\varphi_n \to \varphi$ by the continuity in (3), and so the associated Laplace transforms satisfy $\hat{\mu}_n \to \hat{\mu}$, which implies $\mu_n \xrightarrow{w} \mu$ by Theorem 6.3. Conversely, $\mu_n \xrightarrow{w} \mu$ implies $\hat{\mu}_n \to \hat{\mu}$, and so $\varphi_n \to \varphi$, which yields $\Delta\varphi_n \to \Delta\varphi$, where

$$\begin{aligned} \Delta\varphi(r) &= \varphi(r+1) - \varphi(r) \\ &= \int e^{-rx}\, \tilde{\nu}(dx). \end{aligned}$$

Using Theorem 6.3, we conclude that $\tilde{\nu}_n \xrightarrow{w} \tilde{\nu}$.

(i) If $c_n \to c$ and $\tilde{\nu}_n \xrightarrow{w} \tilde{\nu}$, then $\psi_n \to \psi$ by the boundedness and continuity of the integrand in (2), and so $\hat{\mu}_n \to \hat{\mu}$, which implies $\mu_n \xrightarrow{w} \mu$ by Theorem 6.3. Conversely, $\mu_n \xrightarrow{w} \mu$ implies $\hat{\mu}_n \to \hat{\mu}$ uniformly on bounded intervals, and so $\psi_n \to \psi$ in the same sense. Now define

$$\begin{aligned} \chi(r) &= \int_{-1}^{1} \bigl\{\psi(r) - \psi(r+s)\bigr\}\, ds \\ &= 2\int e^{irx}\Bigl(1 - \frac{\sin x}{x}\Bigr)\frac{1+x^2}{x^2}\, \tilde{\nu}(dx), \end{aligned}$$

and similarly for χ_n, where the interchange of integrations is justified by Fubini's theorem. Then $\chi_n \to \chi$, and so by Theorem 6.3

$$\Bigl(1 - \frac{\sin x}{x}\Bigr)\frac{1+x^2}{x^2}\, \tilde{\nu}_n(dx) \xrightarrow{w} \Bigl(1 - \frac{\sin x}{x}\Bigr)\frac{1+x^2}{x^2}\, \tilde{\nu}(dx).$$

Since the integrand is continuous and bounded away from 0, it follows that $\tilde{\nu}_n \xrightarrow{w} \tilde{\nu}$. This implies convergence of the integral in (2), and so by subtraction $c_n \to c$. □

Proof of Theorem 7.7, *representable case:* For bounded measures m_n and m on R, we note that $m_n \xrightarrow{w} m$ iff $m_n \xrightarrow{v} m$ on $\bar{\mathsf{R}}\setminus\{0\}$ and $m_n(-h,h) \to m(-h,h)$ for some $h > 0$ with $m\{\pm h\} = 0$. Thus, for distributions μ and μ_n on R, we have $\tilde{\nu}_n \xrightarrow{w} \tilde{\nu}$ iff $\nu_n \xrightarrow{v} \nu$ on $\bar{\mathsf{R}}\setminus\{0\}$ and $a_n^h \to a^h$ for every $h > 0$ with $\nu\{\pm h\} = 0$. Similarly, $\tilde{\nu}_n \xrightarrow{w} \tilde{\nu}$ holds for distributions μ and μ_n on R_+ iff $\nu_n \xrightarrow{v} \nu$ on $(0,\infty]$ and $a_n^h \to a^h$ for all $h > 0$ with $\nu\{h\} = 0$. Thus, (iii) follows immediately from Lemma 7.8. To obtain (ii) from the same lemma when $d = 1$, it remains to note that the conditions $b_n^h \to b^h$ and $c_n \to c$ are equivalent when $\tilde{\nu}_n \xrightarrow{w} \tilde{\nu}$ and $\nu\{\pm h\} = 0$, since $|x - x(1+x^2)^{-1}| \le |x|^3$.

Turning to the proof of (ii) when $d > 1$, we first assume that $\nu_n \xrightarrow{v} \nu$ on $\overline{\mathsf{R}^d} \setminus \{0\}$, and that $a_n^h \to a^h$ and $b_n^h \to b^h$ for some $h > 0$ with $\nu\{|x| = h\} = 0$. To prove $\mu_n \xrightarrow{w} \mu$, it suffices by Corollary 6.5 to show that, for any one-dimensional projection $\pi_u : x \mapsto u'x$ with $u \ne 0$, $\mu_n \circ \pi_u^{-1} \xrightarrow{w} \mu \circ \pi_u^{-1}$. Then fix any $k > 0$ with $\nu\{|u'x| = k\} = 0$, and note that $\mu \circ \pi_u^{-1}$ has the associated characteristics $\nu^u = \nu \circ \pi_u^{-1}$ and

$$a^{u,k} = u'a^h u + \int (u'x)^2 \Big\{ 1_{(0,k]}(|u'x|) - 1_{(0,h]}(|x|) \Big\} \nu(dx),$$
$$b^{u,k} = u'b^h + \int u'x \Big\{ 1_{(1,k]}(|u'x|) - 1_{(1,h]}(|x|) \Big\} \nu(dx).$$

Let $a_n^{u,k}$, $b_n^{u,k}$, and ν_n^u denote the corresponding characteristics of $\mu_n \circ \pi_u^{-1}$. Then $\nu_n^u \xrightarrow{v} \nu^u$ on $\bar{\mathsf{R}} \setminus \{0\}$, and furthermore $a_n^{u,k} \to a^{u,k}$ and $b_n^{u,k} \to b^{u,k}$. The desired convergence now follows from the one-dimensional result.

Conversely, let $\mu_n \xrightarrow{w} \mu$. Then $\mu_n \circ \pi_u^{-1} \xrightarrow{w} \mu \circ \pi_u^{-1}$ for every $u \neq 0$, and the one-dimensional result yields $\nu_n^u \xrightarrow{v} \nu^u$ on $\bar{\mathsf{R}} \setminus \{0\}$, as well as $a_n^{u,k} \to a^{u,k}$ and $b_n^{u,k} \to b^{u,k}$ for any $k > 0$ with $\nu\{|u'x| = k\} = 0$. In particular, the sequence $(\nu_n K)$ is bounded for every compact set $K \subset \overline{\mathsf{R}^d} \setminus \{0\}$, and so the sequences $(u'a_n^h u)$ and $(u'b_n^h)$ are bounded for any $u \neq 0$ and $h > 0$. In follows easily that (a_n^h) and (b_n^h) are bounded for every $h > 0$, and therefore all three sequences are relatively compact.

Given a sub-sequence $N' \subset \mathsf{N}$, we have $\nu_n \xrightarrow{v} \nu'$ along a further sub-sequence $N'' \subset N'$, for some measure ν' satisfying $\int (|x|^2 \wedge 1)\nu'(dx) < \infty$. Fixing any $h > 0$ with $\nu'\{|x| = h\} = 0$, we may choose yet another sub-sequence N''', such that even a_n^h and b_n^h converge toward some limits a' and b'. The direct assertion then yields $\mu_n \xrightarrow{w} \mu'$ along N''', where μ' is infinitely divisible with characteristics determined by (a', b', ν'). Since $\mu' = \mu$, we get $\nu' = \nu$, $a' = a^h$, and $b' = b^h$. Thus, the convergence remains valid along the original sequence. □

Proof of Theorem 7.5*:* Directly from the representation, or even easier by Corollary 7.6, we see that any representable random vector ξ is infinitely divisible. Conversely, if ξ is infinitely divisible, Corollary 7.4 yields $\xi_n \xrightarrow{d} \xi$ for some compound Poisson random vectors ξ_n, which are representable by Lemma 7.1. Then even ξ is representable, by the preliminary version of Theorem 7.7 (i). □

Proof of Theorem 7.7*, general case:* Since infinite divisibility and representability are equivalent by Theorem 7.5, the proof in the representable case remains valid for infinitely divisible random vectors ξ_n and ξ. □

We may now complete the classical limit theory for sums of independent random variables from Chapter 6, beginning with the case of general i.i.d. arrays (ξ_{nj}).

Corollary 7.9 *(i.i.d.-array convergence) Let (ξ_{nj}) be an i.i.d. array in R^d, let ξ be infinitely divisible in R^d with characteristics (a, b, ν), and fix any $h > 0$ with $\nu\{|x| = h\} = 0$. Then $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff*

(i) $m_n \mathcal{L}(\xi_{n1}) \xrightarrow{v} \nu$ *on* $\overline{\mathsf{R}^d} \setminus \{0\}$,

(ii) $m_n E\Big(\xi_{n1}\xi_{n1}';\ |\xi_{n1}| \le h\Big) \to a^h$,

(iii) $m_n E\Big(\xi_{n1};\ |\xi_{n1}| \le h\Big) \to b^h$.

Proof: For an i.i.d. array (ξ_{nj}) in R^d, the associated compound Poisson array $(\tilde{\xi}_{nj})$ is infinitely divisible with characteristics

$$a_n = 0, \qquad b_n = E\Big(\xi_{nj};\, |\xi_{nj}| \le 1\Big), \qquad \nu_n = \mathcal{L}(\xi_{nj}),$$

and so the row sums $\xi_n = \sum_j \xi_{nj}$ are infinitely divisible with characteristics $(0,\, m_n b_n,\, m_n \mu_n)$. The assertion now follows by Theorems 7.2 and 7.7. □

The following extension of Theorem 6.12 clarifies the connection between null arrays with positive and symmetric terms. Here we define $p_2(x) = x^2$.

Theorem 7.10 *(positive and symmetric terms) Let (ξ_{nj}) be a null array of symmetric random variables, and let ξ, η be infinitely divisible with characteristics $(a, 0, \nu)$ and $(a,\, \nu \circ p_2^{-1})$, respectively, where ν is symmetric and $a \ge 0$. Then*

$$\textstyle\sum_j \xi_{nj} \xrightarrow{d} \xi \quad \Leftrightarrow \quad \sum_j \xi_{nj}^2 \xrightarrow{d} \eta.$$

Proof: Define $\mu_{nj} = \mathcal{L}(\xi_{nj})$, and fix any $h > 0$ with $\nu\{|x| = h\} = 0$. By Lemma 7.2 and Theorem 7.7 (i), we have $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff

$$\textstyle\sum_j \mu_{nj} \xrightarrow{v} \nu \ \text{ on } \ \bar{\mathsf{R}} \setminus \{0\},$$
$$\textstyle\sum_j E\Big(\xi_{nj}^2;\, |\xi_{nj}| \le h\Big) \to a + \displaystyle\int_{|x| \le h} x^2 \nu(dx),$$

whereas $\sum_j \xi_{nj}^2 \xrightarrow{d} \eta$ iff

$$\textstyle\sum_j \mu_{nj} \circ p_2^{-1} \xrightarrow{v} \nu \circ p_2^{-1} \ \text{ on } \ (0, \infty],$$
$$\textstyle\sum_j E\Big(\xi_{nj}^2;\, \xi_{nj}^2 \le h^2\Big) \to a + \displaystyle\int_{y \le h^2} y\, (\nu \circ p_2^{-1})(dy).$$

The two sets of conditions are equivalent by Lemma 1.24. □

The convergence problem for general null arrays is more delicate, since a compound Poisson approximation as in Proposition 7.2 applies only after a centering at truncated means, as specified below.

Proposition 7.11 *(compound Poisson approximation) Let (ξ_{nj}) be a null array of random vectors in R^d, fix any $h > 0$, and define*

$$b_{nj} = E\Big(\xi_{nj};\, |\xi_{nj}| \le h\Big), \quad n, j \in \mathsf{N}. \tag{4}$$

Then

$$\textstyle\sum_j \xi_{nj} \overset{d}{\sim} \sum_j \Big\{(\xi_{nj} - b_{nj})^{\sim} + b_{nj}\Big\}. \tag{5}$$

Our proof will be based on a technical estimate:

Lemma 7.12 *(uniform summability) Let (ξ_{nj}) be a null array with truncated means b_{nj} as in (4), and let φ_{nj} be the characteristic functions of the vectors $\eta_{nj} = \xi_{nj} - b_{nj}$. Then convergence of either side of (5) implies*

$$\limsup_{n \to \infty} \textstyle\sum_j \Big|1 - \varphi_{nj}(u)\Big| < \infty, \quad u \in \mathsf{R}^d.$$

Proof: The definitions of b_{nj}, η_{nj}, and φ_{nj} yield

$$1 - \varphi_{nj}(u) = E\Big(1 - e^{iu'\eta_{nj}} + iu'\eta_{nj} 1\Big\{|\xi_{nj}| \le h\Big\}\Big) \\ - iu'b_{nj} P\Big\{|\xi_{nj}| > h\Big\}.$$

Putting

$$a_n = \sum\nolimits_j E\Big(\eta_{nj}\eta'_{nj};\ |\xi_{nj}| \le h\Big), \\ p_n = \sum\nolimits_j P\Big\{|\xi_{nj}| > h\Big\},$$

we get by Lemma 6.15

$$\sum\nolimits_j \Big|1 - \varphi_{nj}(u)\Big| \lesssim \tfrac{1}{2}\, u'a_n u + (2 + |u|)\, p_n.$$

It is then enough to show that $(u'a_n u)$ and (p_n) are bounded.

Assuming convergence on the right of (5), the stated boundedness follows easily from Theorem 7.7, together with the fact that $\max_j |b_{nj}| \to 0$. If instead $\sum_j \xi_{nj} \xrightarrow{d} \xi$, we may introduce an independent copy (ξ'_{nj}) of the array (ξ_{nj}) and apply Proposition 7.2 and Theorem 7.7 to the symmetric random variables $\zeta^u_{nj} = u'\xi_{nj} - u'\xi'_{nj}$. Then for any $h' > 0$,

$$\limsup_{n\to\infty} \sum\nolimits_j P\Big\{|\zeta^u_{nj}| > h'\Big\} < \infty, \tag{6}$$

$$\limsup_{n\to\infty} \sum\nolimits_j E\Big\{(\zeta^u_{nj})^2;\ |\zeta^u_{nj}| \le h'\Big\} < \infty. \tag{7}$$

The boundedness of p_n follows from (6) and Lemma 5.19. Next we note that (7) remains true with the condition $|\zeta^u_{nj}| \le h'$ replaced by $|\xi_{nj}| \vee |\xi'_{nj}| \le h$. Using the independence of ξ_{nj} and ξ'_{nj}, we get

$$\begin{aligned} &\tfrac{1}{2} \sum\nolimits_j E\Big\{(\zeta^u_{nj})^2;\ |\xi_{nj}| \vee |\xi'_{nj}| \le h]\Big\} \\ &\quad = \sum\nolimits_j E\Big\{(u'\eta_{nj})^2;\ |\xi_{nj}| \le h\Big\}\, P\Big\{|\xi_{nj}| \le h\Big\} \\ &\qquad - \sum\nolimits_j \Big\{E\Big(u'\eta_{nj};\ |\xi_{nj}| \le h\Big)\Big\}^2 \\ &\quad \ge u'a_n u\ \min\nolimits_j P\Big\{|\xi_{nj}| \le h\Big\} - \sum\nolimits_j \Big(u'b_{nj} P\Big\{|\xi_{nj}| > h\Big\}\Big)^2. \end{aligned}$$

Here the last sum is bounded by $p_n \max_j (u'b_{nj})^2 \to 0$, and the minimum on the right tends to 1. The boundedness of $(u'a_n u)$ now follows by (7). □

Proof of Proposition 7.11: By Lemma 6.14 it suffices to show that

$$\sum\nolimits_j \Big|\varphi_{nj}(u) - \exp\Big\{\varphi_{nj}(u) - 1\Big\}\Big| \to 0, \quad u \in \mathsf{R}^d,$$

where φ_{nj} is the characteristic function of η_{nj}. This follows from Taylor's formula, together with Lemmas 6.6 and 7.12. □

In particular, we may now identify the possible limits.

Corollary 7.13 *(null-array limits, Feller, Khinchin) Let (ξ_{nj}) be a null array of random vectors in R^d, such that $\sum_j \xi_{nj} \xrightarrow{d} \xi$ for a random vector ξ. Then ξ is infinitely divisible.*

Proof: The random vectors $\tilde{\eta}_{nj}$ in Lemma 7.11 are infinitely divisible, and so the same thing is true for the sums $\sum_j (\tilde{\eta}_{nj} - b_{nj})$. The infinite divisibility of ξ then follows by Theorem 7.7 (i). □

To obtain explicit convergence criteria for general null arrays in R^d, it remains to combine Theorem 7.7 with Proposition 7.11. The present result extends Theorem 6.16 for Gaussian limits and Corollary 7.9 for i.i.d. arrays. For convenience, we write $\mathrm{Cov}(\xi; A)$ for the covariance matrix of the random vector $1_A \xi$.

Theorem 7.14 *(null-array convergence, Doeblin, Gnedenko) Let (ξ_{nj}) be a null array of random vectors in R^d, and let ξ be infinitely divisible with characteristics (a, b, ν). Then for fixed $h > 0$ with $\nu\{|x| = h\} = 0$, we have $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff*

(i) $\sum_j \mathcal{L}(\xi_{nj}) \xrightarrow{v} \nu$ *on* $\overline{\mathsf{R}^d} \setminus \{0\}$,
(ii) $\sum_j \mathrm{Cov}\Big(\xi_{nj};\, |\xi_{nj}| \le h\Big) \to a^h$,
(iii) $\sum_j E\Big(\xi_{nj};\, |\xi_{nj}| \le h\Big) \to b^h$.

Proof: Define for any n and j

$$\begin{aligned} a_{nj} &= \mathrm{Cov}\Big(\xi_{nj};\, |\xi_{nj}| \le h\Big), \\ b_{nj} &= E\Big(\xi_{nj};\, |\xi_{nj}| \le h\Big). \end{aligned}$$

By Theorems 7.7 and 7.11, we have $\sum_j \xi_{nj} \xrightarrow{d} \xi$ iff

(i′) $\sum_j \mathcal{L}(\eta_{nj}) \xrightarrow{v} \nu$ on $\overline{\mathsf{R}^d} \setminus \{0\}$,
(ii′) $\sum_j E\Big(\eta_{nj}\eta'_{nj};\, |\eta_{nj}| \le h\Big) \to a^h$,
(iii′) $\sum_j \Big(b_{nj} + E\Big\{\eta_{nj};\, |\eta_{nj}| \le h\Big\}\Big) \to b^h$.

Here (i) ⇔ (i′) since $\max_j |b_{nj}| \to 0$. By (i) and the facts that $\max_j |b_{nj}| \to 0$ and $\nu\{|x| = h\} = 0$, we further see that the sets $\{|\eta_{nj}| \le h\}$ in (ii′) and (iii′) may be replaced by $\{|\xi_{nj}| \le h\}$. To prove (ii) ⇔ (ii′), we note that by (i)

$$\begin{aligned} \Big\| \textstyle\sum_j \Big(a_{nj} - E\Big\{\eta_{nj}\,\eta'_{nj};\, |\xi_{nj}| \le h\Big\}\Big) \Big\| &\le \Big\| \textstyle\sum_j b_{nj}\, b'_{nj} P\Big\{|\xi_{nj}| > h\Big\} \Big\| \\ &\lesssim \max_j |b_{nj}|^2 \textstyle\sum_j P\Big\{|\xi_{nj}| > h\Big\} \to 0. \end{aligned}$$

Similarly, (iii) ⇔ (iii′) because

$$\begin{aligned} \Big| \textstyle\sum_j E\Big(\eta_{nj};\, |\xi_{nj}| \le h\Big) \Big| &= \Big| \textstyle\sum_j b_{nj} P\Big\{|\xi_{nj}| > h\Big\} \Big| \\ &\le \max_j |b_{nj}| \textstyle\sum_j P\Big\{|\xi_{nj}| > h\Big\} \to 0. \end{aligned}$$ □

When $d = 1$, the first condition in Theorem 7.14 admits some interesting probabilistic interpretations, one of which involves the row-wise extremes. For random measures η and η_n on $\mathsf{R} \setminus \{0\}$, the convergence $\eta_n \xrightarrow{vd} \eta$ on $\bar{\mathsf{R}} \setminus \{0\}$ means that $\eta_n f \xrightarrow{d} \eta f$ for all $f \in \hat{C}_+(\bar{\mathsf{R}} \setminus \{0\})$.

Theorem 7.15 *(sums and extremes) Let (ξ_{nj}) be a null array of random variables with distributions μ_{nj}, and define*

$$\eta_n = \sum_j \delta_{\xi_{nj}}, \quad \alpha_n^\pm = \max_j \big(\pm \xi_{nj}\big), \qquad n \in \mathsf{N}.$$

Fix a Lévy measure ν on $\mathsf{R} \setminus \{0\}$, let η be a Poisson process on $\mathsf{R} \setminus \{0\}$ with $E\eta = \nu$, and put

$$\alpha^\pm = \sup\big\{x \geq 0;\ \eta\{\pm x\} > 0\big\}.$$

Then these conditions are equivalent:

(i) $\sum_j \mu_{nj} \xrightarrow{v} \nu$ *on* $\bar{\mathsf{R}} \setminus \{0\}$,

(ii) $\eta_n \xrightarrow{d} \eta$ *on* $\bar{\mathsf{R}} \setminus \{0\}$,

(iii) $\alpha_n^\pm \xrightarrow{d} \alpha^\pm$.

Though (i) $\Leftrightarrow$ (ii) is immediate from Theorem 30.1 below, we include a direct, elementary proof.

Proof: Condition (i) holds iff

$$\sum_j \mu_{nj}(x, \infty) \to \nu(x, \infty),$$
$$\sum_j \mu_{nj}(-\infty, -x) \to \nu(-\infty, -x),$$

for all $x > 0$ with $\nu\{\pm x\} = 0$. By Lemma 6.8, the first condition is equivalent to

$$\begin{aligned} P\{\alpha_n^+ \leq x\} &= \prod_j \big(1 - P\{\xi_{nj} > x\}\big) \\ &\to e^{-\nu(x,\infty)} \\ &= P\{\alpha^+ \leq x\}, \end{aligned}$$

which holds for all continuity points $x > 0$ iff $\alpha_n^+ \xrightarrow{d} \alpha^+$. Similarly, the second condition holds iff $\alpha_n^- \xrightarrow{d} \alpha^-$. Thus, (i) $\Leftrightarrow$ (iii).

To show that (i) $\Rightarrow$ (ii), we may write the latter condition in the form

$$\sum_j f(\xi_{nj}) \xrightarrow{d} \eta f, \quad f \in \hat{C}_+(\bar{\mathsf{R}} \setminus \{0\}). \tag{8}$$

Here the variables $f(\xi_{nj})$ form a null array with distributions $\mu_{nj} \circ f^{-1}$, and ηf is compound Poisson with characteristic measure $\nu \circ f^{-1}$. By Theorem 7.7 (ii), (8) is then equivalent to the conditions

$$\sum_j \mu_{nj} \circ f^{-1} \xrightarrow{v} \nu \circ f^{-1} \ \text{ on } \ (0, \infty], \tag{9}$$

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \sum_j \int_{f(x) \leq \varepsilon} f(x)\, \mu_{nj}(dx) = 0. \tag{10}$$

Now (9) follows immediately from (i). To deduce (10), it suffices to note that the sum on the left is bounded by $\sum_j \mu_{nj}(f \wedge \varepsilon) \to \nu(f \wedge \varepsilon)$.

Finally, assume (ii). By a simple approximation, $\eta_n(x, \infty) \xrightarrow{d} \eta(x, \infty)$ for any $x > 0$ with $\nu\{x\} = 0$. In particular, for such an x,

$$\begin{aligned} P\{\alpha_n^+ \le x\} &= P\Big\{\eta_n(x,\infty) = 0\Big\} \\ &\to P\Big\{\eta(x,\infty) = 0\Big\} \\ &= P\{\alpha^+ \le x\}, \end{aligned}$$

and so $\alpha_n^+ \xrightarrow{d} \alpha^+$. Similarly $\alpha_n^- \xrightarrow{d} \alpha^-$, which proves (iii). □

The characteristic pair or triple clearly contains important information about an infinitely divisible distribution. In particular, we have the following classical result:

Proposition 7.16 *(diffuse distributions, Doeblin) Let ξ be an infinitely divisible random vector in R^d with characteristics (a, b, ν). Then*

$$\mathcal{L}(\xi) \text{ diffuse} \quad \Leftrightarrow \quad a \ne 0 \text{ or } \|\nu\| = \infty.$$

Proof: If $a = 0$ and $\|\nu\| < \infty$, then ξ is compound Poisson apart from a shift, and so $\mu = \mathcal{L}(\xi)$ is not diffuse. When either condition fails, it does so for at least one coordinate projection, and we may take $d = 1$. If $a > 0$, the diffuseness is obvious by Lemma 1.30. Next let ν be unbounded, say with $\nu(0,\infty) = \infty$. For every $n \in \mathsf{N}$ we may then write $\nu = \nu_n + \nu'_n$, where ν'_n is supported by $(0, n^{-1})$ and has total mass $\log 2$. For μ we get a corresponding decomposition $\mu_n * \mu'_n$, where μ'_n is compound Poisson with Lévy measure ν'_n and $\mu'_n\{0\} = \frac{1}{2}$. For any $x \in \mathsf{R}$ and $\varepsilon > 0$, we get

$$\begin{aligned} \mu\{x\} &\le \mu_n\{x\}\mu'_n\{0\} + \mu_n[x-\varepsilon, x)\,\mu'_n(0,\varepsilon] + \mu'_n(\varepsilon,\infty) \\ &\le \tfrac{1}{2}\,\mu_n[x-\varepsilon, x] + \mu'_n(\varepsilon,\infty). \end{aligned}$$

Letting $n \to \infty$ and then $\varepsilon \to 0$, and noting that $\mu'_n \xrightarrow{w} \delta_0$ and $\mu_n \xrightarrow{w} \mu$, we get $\mu\{x\} \le \frac{1}{2}\,\mu\{x\}$ by Theorem 5.25, and so $\mu\{x\} = 0$. □

Exercises

1. Prove directly from definitions that a compound Poisson distribution is infinitely divisible, and identify the corresponding characteristics. Also do the same verification using generating functions.

2. Prove directly from definitions that every Gaussian distribution in R^d is infinitely divisible. (*Hint:* Write a Gaussian vector in R^d in the form $\xi = a\,\zeta + b$ for a suitable matrix a and vector b, where ζ is standard Gaussian in R^d. Then compute the distribution of $\zeta_1 + \cdots + \zeta_n$ by a simple convolution.)

3. Prove the infinite divisibility of the Poisson and Gaussian distributions from Theorems 6.7 and 6.13. (*Hint:* Apply the mentioned theorems to suitable i.i.d. arrays.)

4. Show that a distribution μ on R with characteristic function $\hat{\mu}(u) = e^{-|u|^p}$ for a $p \in (0, 2]$ is infinitely divisible. Then find the form of the associated Lévy measure.

(*Hint:* Note that μ is symmetric p-stable, and conclude that $\nu(dx) = c_p|x|^{-p-1}dx$ for a constant $c_p > 0$.)

5. Show that the Cauchy distribution μ with density $\pi^{-1}(1+x^2)^{-1}$ is infinitely divisible, and find the associated Lévy measure. (*Hint:* Check that μ is symmetric 1-stable, and use the preceding exercise. It remains to find the constant c_1.)

6. Extend Proposition 7.10 to null arrays of spherically symmetric random vectors in R^d.

7. Derive Theorems 6.7 and 6.12 from Lemma 7.2 and Theorem 7.7.

8. Show by an example that Theorem 7.11 fails without the centering at truncated means. (*Hint:* Without the centering, condition (ii) of Theorem 7.14 becomes $\sum_j E(\xi_{nj}\xi'_{nj};\, |\xi_{nj}| \le h) \to a^h$.)

9. If ξ is infinitely divisible with characteristics (a,b,ν) and $p > 0$, show that $E|\xi|^p < \infty$ iff $\int_{|x|>1} |x|^p\, \nu(dx) < \infty$. (*Hint:* If ν has bounded support, then $E|\xi|^p < \infty$ for all p. It is then enough to consider compound Poisson distributions, for which the result is elementary.)

10. Prove directly that a Z_+-valued random variable ξ is infinitely divisible (on Z_+) iff $-\log Es^\xi = \sum_k (1-s^k)\,\nu_k$, $s \in (0,1]$, for a unique, bounded measure $\nu = (\nu_k)$ on N. (*Hint:* Assuming $\mathcal{L}(\xi) = \mu_n^{*n}$, use the inequality $1 - x \le e^{-x}$ to show that the sequence $(n\mu_n)$ is tight on N. Then $n\mu_n \xrightarrow{w} \nu$ along a sub-sequence, for a bounded measure ν on N. Finally note that $-\log(1-x) \sim x$ as $x \to 0$. As for the uniqueness, take differences and use the uniqueness theorem for power series.)

11. Prove directly that a random variable $\xi \ge 0$ is infinitely divisible iff $-\log Ee^{-u\xi} = ua + \int (1-e^{-ux})\,\nu(dx)$, $u \ge 0$, for some unique constant $a \ge 0$ and measure ν on $(0,\infty)$ with $\int(|x| \wedge 1) < \infty$. (*Hint:* If $\mathcal{L}(\xi) = \mu_n^{*n}$, note that the measures $\chi_n(dx) = n(1-e^{-x})\,\mu_n(dx)$ are tight on R_+. Then $\chi_n \xrightarrow{w} \chi$ along a sub-sequence, and we may write $\chi(dx) = a\,\delta_0(dx) + (1-e^{-x})\,\nu(dx)$. The desired representation now follows as before. As for the uniqueness, take differences and use the uniqueness theorem for Laplace transforms.)

12. Prove directly that a random variable ξ is infinitely divisible iff $\psi_u = \log Ee^{iu\xi}$ exists and is given by Corollary 7.6 (i) for some unique constants $a \ge 0$ and b and a measure ν on $\mathsf{R} \setminus \{0\}$ with $\int (x^2 \wedge 1)\,\nu(dx) < \infty$. (*Hint:* Proceed as in Lemma 7.8.)

III. Conditioning and Martingales

Modern probability theory can be said to begin with the theory of conditional expectations and distributions, along with the basic properties of martingales. This theory also involves the technical machinery of filtrations, optional and predictable times, predictable processes and compensators, all of which are indispensable for the further developments. The material of Chapter 8 is constantly used in all areas of probability theory, and should be thoroughly mastered by any serious student of the subject. The same thing can be said about at least the early parts of Chapter 9, including the basic properties of optional times and discrete-time martingales. The more advanced continuous-time theory, along with results for predictable processes and compensators in Chapter 10, might be postponed for a later study.

— — —

8. Conditioning and disintegration. Here conditional expectations are introduced by the intuitive method of Hilbert space projection, which leads easily to the basic properties. Of equal importance is the notion of conditional distributions, along with the powerful disintegration and transfer theorems, constantly employed in all areas of modern probability theory. The latter result yields in particular the fundamental Daniell–Kolmogorov theorem, ensuring the existence of processes with given finite-dimensional distributions.

9. Optional times and martingales. Here we begin with the basic properties of optional times with associated σ-fields, along with a brief discussion of random time change. Martingales are then introduced as projective sequences of random variables, and we give short proofs of their basic properties, including the classical maximum inequalities, the optional sampling theorem, and the fundamental convergence and regularization theorems. After discussing a range of ramifications and applications, we conclude with some extensions to continuous-time sub-martingales.

10. Predictability and compensation. Here we introduce the classes of predictable times and processes, and give a complete, probabilistic proof of the fundamental Doob–Meyer decomposition. The latter result leads to the equally basic notion of compensator of a random measure on a product space $\mathsf{R}_+ \times S$, and to a variety of characterizations and time-change reductions. We conclude with the more advanced theory of discounted compensators and associated predictable mapping theorems.

Chapter 8

Conditioning and Disintegration

Conditional expectation, conditional variance and covariance, local property, uniform integrability, conditional probabilities, conditional distributions and disintegration, conditional independence, chain rule, projection and orthogonality, commutation criteria, iterated conditioning, extension and transfer, stochastic equations, coupling and randomization, existence of sequences and processes, extension by conditioning, infinite product measures

Modern probability theory can be said to begin with the notions of conditioning and disintegration. In particular, conditional expectations and distributions are needed already for the *definitions* of martingales and Markov processes, the two basic dependence structures beyond independence and stationarity. In many other areas throughout probability theory, conditioning is used as a universal tool, needed to describe and analyze systems involving randomness. The notion may be thought of in terms of *averaging, projection*, and *disintegration*—alternative viewpoints that are all essential for a proper understanding.

In all but the most elementary contexts, conditioning is defined with respect to a σ-field rather than a single event. Here the resulting quantity is not a constant but a random variable, measurable with respect to the conditioning σ-field. The idea is familiar from elementary constructions of conditional expectations $E(\xi|\,\eta)$, for random vectors (ξ,η) with a nice density, where the expected value becomes a function of η. This corresponds to conditioning on the generated σ-field $\mathcal{F}=\sigma(\eta)$.

General conditional expectations are traditionally constructed by means of the Radon–Nikodym theorem. However, the simplest and most intuitive approach is arguably via *Hilbert space projection*, where $E(\xi\,|\,\mathcal{F})$ is defined for any $\xi \in L^2$ as the orthogonal projection of ξ onto the linear subspace of $\mathcal{F}$-measurable random variables. The existence and basic properties of the L^2-version extend by continuity to arbitrary $\xi \in L^1$. The orthogonality yields $E\{\xi - E(\xi\,|\,\mathcal{F})\}\zeta = 0$ for any bounded, $\mathcal{F}$-measurable random variable ζ, which leads immediately to the familiar averaging characterization of $E(\xi\,|\,\mathcal{F})$ as a version of the density $d(\xi\cdot P)/dP$ on the σ-field $\mathcal{F}$.

The conditional expectation is only defined up to a P-null set, in the sense that any two versions agree a.s. We may then look for versions of the conditional probabilities $P(A\,|\,\mathcal{F}) = E(1_A|\,\mathcal{F})$ that combine into a random probability measure on Ω. In general, such *regular* versions exist only for A restricted to a sufficiently nice sub-σ-field. The basic case is for random elements ξ in

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_9

a Borel space S, where the conditional distribution $\mathcal{L}(\xi \,|\, \mathcal{F})$ exists as an $\mathcal{F}$-measurable random measure on S. Assuming in addition that $\mathcal{F} = \sigma(\eta)$ for a random element η in a space T, we may write $P\{\xi \in B \,|\, \eta\} = \mu(\eta, B)$ for a probability kernel $\mu: T \to S$, which leads to a decomposition of the distribution of (ξ, η) according to the values of η. The result is formalized by the powerful *disintegration theorem*—an extension of Fubini's theorem of constant use in subsequent chapters, especially in combination with the strong Markov property.

Conditional distributions can be used to establish the basic *transfer theorem*, often needed to convert a distributional equivalence $\xi \overset{d}{=} f(\eta)$ into an a.s. representation $\xi = f(\tilde{\eta})$, for a suitable choice of $\tilde{\eta} \overset{d}{=} \eta$. The latter result leads in turn to the fundamental *Daniell–Kolmogorov theorem*, which guarantees the existence of a random sequence or process with specified finite-dimensional distributions. A different approach yields the more general *Ionescu Tulcea extension*, where the measure is specified by a sequence of conditional distributions.

Further topics discussed in this chapter include the notion of *conditional independence*, which is fundamental for both Markov processes and exchangeability, and also plays an important role in connection with SDEs in Chapter 32, and for the random arrays treated in Chapter 28. Especially useful for various applications is the elementary but powerful *chain rule*. We finally call attention to the local property of conditional expectations, which leads in particular to simple and transparent proofs of the optional sampling and strong Markov properties.

Returning to our construction of conditional expectations, fix a probability space $(\Omega, \mathcal{A}, P)$ and an arbitrary sub-σ-field $\mathcal{F} \subset \mathcal{A}$. Introduce in $L^2 = L^2(\mathcal{A})$ the closed linear subspace M, consisting of all random variables $\eta \in L^2$ that agree a.s. with elements of $L^2(\mathcal{F})$. By the Hilbert space projection Theorem 1.35, there exists for every $\xi \in L^2$ an a.s. unique random variable $\eta \in M$ with $\xi - \eta \perp M$, and we define $E^{\mathcal{F}}\xi = E(\xi \,|\, \mathcal{F})$ as an arbitrary $\mathcal{F}$-measurable version of η.

The L^2-projection $E^{\mathcal{F}}$ is easily extended to L^1, as follows:

Theorem 8.1 *(conditional expectation, Kolmogorov) For any σ-field $\mathcal{F} \subset \mathcal{A}$, there exists an a.s. unique linear operator $E^{\mathcal{F}}: L^1 \to L^1(\mathcal{F})$, such that*

(i) $E\Big(E^{\mathcal{F}}\xi; A\Big) = E(\xi; A),\ \xi \in L^1,\ A \in \mathcal{F}.$

The operators $E^{\mathcal{F}}$ have these further properties, whenever the corresponding expressions exist for the absolute values:

(ii) $\xi \geq 0 \;\Rightarrow\; E^{\mathcal{F}}\xi \geq 0$ *a.s.*,

(iii) $E|E^{\mathcal{F}}\xi| \leq E|\xi|$,

(iv) $0 \leq \xi_n \uparrow \xi \;\Rightarrow\; E^{\mathcal{F}}\xi_n \uparrow E^{\mathcal{F}}\xi$ *a.s.*,

(v) $E^{\mathcal{F}}\xi\eta = \xi\, E^{\mathcal{F}}\eta$ *a.s. when ξ is $\mathcal{F}$-measurable*,

(vi) $E\Big(\xi\, E^{\mathcal{F}}\eta\Big) = E\Big(\eta\, E^{\mathcal{F}}\xi\Big) = E(E^{\mathcal{F}}\xi)(E^{\mathcal{F}}\eta)$,

(vii) $E^{\mathcal{F}} E^{\mathcal{G}} \xi = E^{\mathcal{F}} \xi$ *a.s.*, $\mathcal{F} \subset \mathcal{G}$.

In particular, $E^{\mathcal{F}}\xi = \xi$ a.s. iff ξ has an $\mathcal{F}$-measurable version, and $E^{\mathcal{F}}\xi = E\,\xi$ a.s. when $\xi \perp\!\!\!\perp \mathcal{F}$. We often refer to (i) as the *averaging* property, to (ii) as the *positivity*, to (iii) as the L^1-*contractivity*, to (iv) as the *monotone convergence* property, to (v) as the *pull-out* property, to (vi) as the *self-adjointness*, and to (vii) as the *tower property*[1]. Since the operator $E^{\mathcal{F}}$ is both self-adjoint by (vi) and idempotent by (vii), it may be thought of as a generalized projection on L^1.

The first assertion is an immediate consequence of Theorem 2.10. However, the following projection approach is more elementary and intuitive, and it has the further advantage of leading easily to properties (ii)$-$(vii).

Proof of Theorem 8.1: For $\xi \in L^2$, define $E^{\mathcal{F}}\xi$ by projection as above. Then for any $A \in \mathcal{F}$ we get $\xi - E^{\mathcal{F}}\xi \perp 1_A$, and (i) follows. Taking $A = \{E^{\mathcal{F}}\xi \ge 0\}$, we get in particular

$$\begin{aligned} E\big|E^{\mathcal{F}}\xi\big| &= E\Big(E^{\mathcal{F}}\xi; A\Big) - E\Big(E^{\mathcal{F}}\xi; A^c\Big) \\ &= E(\xi; A) - E(\xi; A^c) \\ &\le E|\xi|, \end{aligned}$$

proving (iii). The mapping $E^{\mathcal{F}}$ is then uniformly L^1-continuous on L^2. Since L^2 is dense in L^1 by Lemma 1.12 and L^1 is complete by Lemma 1.33, the operator $E^{\mathcal{F}}$ extends a.s. uniquely to a linear and continuous mapping on L^1.

Properties (i) and (iii) extend by continuity to L^1, and Lemma 1.26 shows that $E^{\mathcal{F}}\xi$ is a.s. determined by (i). If $\xi \ge 0$, we may combine (i) for $A = \{E^{\mathcal{F}}\xi \le 0\}$ with Lemma 1.26 to obtain $E^{\mathcal{F}}\xi \ge 0$ a.s., which proves (ii). If $0 \le \xi_n \uparrow \xi$, then $\xi_n \to \xi$ in L^1 by dominated convergence, and so by (iii) we have $E^{\mathcal{F}}\xi_n \to E^{\mathcal{F}}\xi$ in L^1. Since $E^{\mathcal{F}}\xi_n$ is a.s. non-decreasing in n by (ii), Lemma 5.2 yields the corresponding a.s. convergence, which proves (iv).

Property (vi) is obvious when $\xi, \eta \in L^2$, and it extends L^1 by (iv). To prove (v), we see from the characterization in (i) that $E^{\mathcal{F}}\xi = \xi$ a.s. when ξ is $\mathcal{F}$-measurable. In general we need to show that

$$E(\,\xi\eta; A) = E\Big(\xi E^{\mathcal{F}}\eta; A\Big), \quad A \in \mathcal{F},$$

which follows immediately from (vi). Finally, (vii) is obvious for $\xi \in L^2$ since $L^2(\mathcal{F}) \subset L^2(\mathcal{G})$, and it extends to the general case by means of (iv). □

Using conditional expectations, we can define conditional variances and covariances in the obvious way by

$$\begin{aligned} \mathrm{Var}(\xi \,|\, \mathcal{F}) &= E^{\mathcal{F}}(\xi - E^{\mathcal{F}}\xi)^2 \\ &= E^{\mathcal{F}}\xi^2 - (E^{\mathcal{F}}\xi)^2, \\ \mathrm{Cov}(\xi, \eta \,|\, \mathcal{F}) &= E^{\mathcal{F}}(\xi - E^{\mathcal{F}}\xi)(\eta - E^{\mathcal{F}}\eta) \\ &= E^{\mathcal{F}}\xi\eta - (E^{\mathcal{F}}\xi)(E^{\mathcal{F}}\eta), \end{aligned}$$

[1] also called the *chain rule*

as long as the right-hand sides exist. Using the shorthand notation $\mathrm{Var}^{\mathcal{F}}(\xi)$ and $\mathrm{Cov}^{\mathcal{F}}(\xi,\eta)$, we note the following computational rules, which may be compared with the simple rule $E\,\xi = E(F^{\mathcal{F}}\xi)$ for expected values:

Lemma 8.2 *(conditional variance and covariance) For any random variables $\xi,\eta \in L^2$ and σ-field $\mathcal{F}$, we have*

(i) $\mathrm{Var}(\xi) = E[\,\mathrm{Var}^{\mathcal{F}}(\xi)\,] + \mathrm{Var}(E^{\mathcal{F}}\xi)$,

(ii) $\mathrm{Cov}(\xi,\eta) = E[\,\mathrm{Cov}^{\mathcal{F}}(\xi,\eta)\,] + \mathrm{Cov}(E^{\mathcal{F}}\xi, E^{\mathcal{F}}\eta)$.

Proof: We need to prove only (ii), since (i) is the special case with $\xi = \eta$. Then write

$$\begin{aligned}
\mathrm{Cov}(\xi,\eta) &= E(\xi\eta) - (E\xi)(E\eta)\\
&= EE^{\mathcal{F}}(\xi\eta) - (EE^{\mathcal{F}}\xi)(EE^{\mathcal{F}}\eta)\\
&= E\Big\{\mathrm{Cov}^{\mathcal{F}}(\xi,\eta) + (E^{\mathcal{F}}\xi)(E^{\mathcal{F}}\eta)\Big\} - (EE^{\mathcal{F}}\xi)(EE^{\mathcal{F}}\eta)\\
&= E\,\mathrm{Cov}^{\mathcal{F}}(\xi,\eta) + \Big\{E(E^{\mathcal{F}}\xi)(E^{\mathcal{F}}\eta) - (EE^{\mathcal{F}}\xi)(EE^{\mathcal{F}}\eta)\Big\}\\
&= E\,\mathrm{Cov}^{\mathcal{F}}(\xi,\eta) + \mathrm{Cov}(E^{\mathcal{F}}\xi, E^{\mathcal{F}}\eta).
\end{aligned}$$ □

Next we show that the conditional expectation $E^{\mathcal{F}}\xi$ is *local* in both ξ and $\mathcal{F}$, an observation that simplifies many proofs. Given two σ-fields $\mathcal{F}$ and $\mathcal{G}$, we say that $\mathcal{F} = \mathcal{G}$ on A if $A \in \mathcal{F}\cap\mathcal{G}$ and $A\cap\mathcal{F} = A\cap\mathcal{G}$.

Lemma 8.3 *(local property) Consider some σ-fields $\mathcal{F}$, $\mathcal{G} \subset \mathcal{A}$ and functions ξ, $\eta \in L^1$, and let $A \in \mathcal{F}\cap\mathcal{G}$. Then*

$$\mathcal{F} = \mathcal{G},\ \ \xi = \eta \ \text{ a.s. on } A \quad \Rightarrow \quad E^{\mathcal{F}}\xi = E^{\mathcal{G}}\eta \ \text{ a.s. on } A.$$

Proof: Since $1_A E^{\mathcal{F}}\xi$ and $1_A E^{\mathcal{G}}\eta$ are $\mathcal{F}\cap\mathcal{G}$-measurable, we get $B \equiv A \cap \{E^{\mathcal{F}}\xi > E^{\mathcal{G}}\eta\} \in \mathcal{F}\cap\mathcal{G}$, and the averaging property yields

$$\begin{aligned}
E\Big(E^{\mathcal{F}}\xi; B\Big) &= E(\xi;B) = E(\eta;B)\\
&= E\Big(E^{\mathcal{G}}\eta; B\Big).
\end{aligned}$$

Hence, $E^{\mathcal{F}}\xi \le E^{\mathcal{G}}\eta$ a.s. on A by Lemma 1.26. The reverse inequality follows by the symmetric argument. □

The following technical result plays an important role in Chapter 9.

Lemma 8.4 *(uniform integrability, Doob) For any $\xi \in L^1$, the conditional expectations $E(\xi\,|\mathcal{F})$, $\mathcal{F} \subset \mathcal{A}$, are uniformly integrable.*

Proof: By Jensen's inequality and the self-adjointness property,

$$\begin{aligned}
E\Big(|E^{\mathcal{F}}\xi|; A\Big) &\le E\Big(E^{\mathcal{F}}|\,\xi|; A\Big)\\
&= E\Big(|\xi|\,P^{\mathcal{F}}A\Big), \quad A \in \mathcal{A}.
\end{aligned}$$

By Lemma 5.10 we need to show that this tends to zero as $PA \to 0$, uniformly in $\mathcal{F}$. By dominated convergence along sub-sequences, it is then enough to show that $P^{\mathcal{F}_n}A_n \xrightarrow{P} 0$ for any σ-fields $\mathcal{F}_n \subset \mathcal{A}$ and sets $A_n \in \mathcal{A}$ with $PA_n \to 0$, which is clear since $EP^{\mathcal{F}_n}A_n = PA_n \to 0$. □

The *conditional probability* of an event $A \in \mathcal{A}$, given a σ-field $\mathcal{F}$, is defined as

$$P^{\mathcal{F}}A = E^{\mathcal{F}}1_A \quad \text{or} \quad P(A\,|\mathcal{F}) = E(1_A|\mathcal{F}), \quad A \in \mathcal{A}.$$

Thus, $P^{\mathcal{F}}A$ is the a.s. unique random variable in $L^1(\mathcal{F})$ satisfying

$$E\Big\{P^{\mathcal{F}}A; B\Big\} = P(A \cap B), \quad B \in \mathcal{F}.$$

Note that $P^{\mathcal{F}}A = PA$ a.s. iff $A \perp\!\!\!\perp \mathcal{F}$, and that $P^{\mathcal{F}}A = 1_A$ a.s. iff A is a.s. $\mathcal{F}$-measurable. The positivity of $E^{\mathcal{F}}$ gives $0 \le P^{\mathcal{F}}A \le 1$ a.s., and the monotone convergence property yields

$$P^{\mathcal{F}} \textstyle\bigcup_n A_n = \sum_n P^{\mathcal{F}}A_n \ \text{ a.s.}, \quad A_1, A_2, \ldots \in \mathcal{A} \ \text{ disjoint.} \tag{1}$$

Still the random set function $P^{\mathcal{F}}$ may not be a measure, in general, since the exceptional null set in (1) may depend on the sequence (A_n).

For random elements η in a measurable space $(S, \mathcal{S})$, we define η-conditioning as conditioning with respect to the induced σ-field $\sigma(\eta)$, so that

$$E^{\eta}\xi = E^{\sigma(\eta)}\xi, \qquad P^{\eta}A = P^{\sigma(\eta)}A,$$

or

$$\begin{aligned} E(\xi\,|\,\eta) &= E\{\xi\,|\,\sigma(\eta)\}, \\ P(A\,|\,\eta) &= P\{A\,|\,\sigma(\eta)\}. \end{aligned}$$

By Lemma 1.14, the η-measurable function $E^{\eta}\xi$ may be represented as $f(\eta)$ for some measurable function f on S, determined a.e. $\mathcal{L}(\eta)$ by the averaging property

$$E\Big\{f(\eta);\, \eta \in B\Big\} = E\Big(\xi;\, \eta \in B\Big), \quad B \in \mathcal{S}.$$

In particular, f depends only on the distribution of (ξ, η). The case of $P^{\eta}A$ is similar. Conditioning on a σ-field $\mathcal{F}$ is the special case where η is the identity map $(\Omega, \mathcal{A}) \to (\Omega, \mathcal{F})$.

Motivated by (1), we may hope to construct some measure-valued versions of the functions $P^{\mathcal{F}}$ and P^{η}. Then recall from Chapter 3 that, for any measurable spaces $(T, \mathcal{T})$ and $(S, \mathcal{S})$, a *kernel* $\mu\colon T \to S$ is defined as a function $\mu : T \times \mathcal{S} \to \bar{\mathbb{R}}_+$, such that $\mu(t, B)$ is $\mathcal{T}$-measurable in $t \in T$ for fixed B and a measure in $B \in \mathcal{S}$ for fixed t. In particular, μ is a probability kernel if $\mu(t, S) = 1$ for all t. *Random measures* are simply kernels on the basic probability space Ω.

Now fix a σ-field $\mathcal{F} \subset \mathcal{A}$ and a random element η in a measurable space $(T, \mathcal{T})$. By a (regular) *conditional distribution of η, given $\mathcal{F}$,* we mean an $\mathcal{F}$-measurable probability kernel $\mu = \mathcal{L}(\eta\,|\,\mathcal{F})\colon \Omega \to T$, such that

$$\mu(\omega, B) = P\{\eta \in B\,|\,\mathcal{F}\}_{\omega} \ \text{ a.s.}, \quad \omega \in \Omega, \ B \in \mathcal{T}.$$

The idea is to choose versions of the conditional probabilities on the right that combine for each ω into a probability measure in B. More generally, for any random element ξ in a measurable space $(S, \mathcal{S})$, we define a conditional distribution of η, given ξ, as a random measure of the form

$$\mu(\xi, B) = P\{\eta \in B \mid \xi\} \text{ a.s.}, \quad B \in \mathcal{T}, \tag{2}$$

for a probability kernel $\mu : S \to T$. In the extreme cases where η is $\mathcal{F}$-measurable or independent of $\mathcal{F}$, we note that $P\{\eta \in B \,|\mathcal{F}\}$ has the regular version $1\{\eta \in B\}$ or $P\{\eta \in B\}$, respectively.

To ensure the existence of conditional distributions $\mathcal{L}(\eta \,|\, \mathcal{F})$, we need to impose some regularity conditions on the space T. The following existence and disintegration property is a key result of modern probability.

Theorem 8.5 *(conditional distributions, disintegration) Let ξ, η be random elements in S, T, where T is Borel. Then $\mathcal{L}(\xi, \eta) = \mathcal{L}(\xi) \otimes \mu$ for a probability kernel $\mu\colon S \to T$, where μ is unique a.e. $\mathcal{L}(\xi)$ and satisfies*

(i) $\mathcal{L}(\eta \,|\, \xi) = \mu(\xi, \cdot)$ *a.s.,*

(ii) $E\Big\{ f(\xi, \eta) \,\Big|\, \xi \Big\} = \int \mu(\xi, dt)\, f(\xi, t)$ *a.s.,* $\quad f \geq 0$.

Proof: The stated disintegration with associated uniqueness hold by Theorem 3.4. Properties (i) and (ii) then follow by the averaging characterization of conditional expectations. □

Taking differences, we may extend (ii) to suitable real-valued functions f. When $(S, \mathcal{S}) = (\Omega, \mathcal{F})$, the kernel μ becomes an $\mathcal{F}$-measurable random measure ζ on T. Letting ξ be an $\mathcal{F}$-measurable random element in a space T, we get by (ii)

$$E\Big\{ f(\xi, \eta) \,\Big|\, \mathcal{F} \Big\} = \int \zeta(dt)\, f(\xi, t), \quad f \geq 0. \tag{3}$$

Part (ii) may also be written in integrated form as

$$E f(\xi, \eta) = E \int \mu(\xi, dt)\, f(\xi, t), \quad f \geq 0, \tag{4}$$

and similarly for (3). When $\xi \perp\!\!\!\perp \eta$ we may choose $\mu(\xi, \cdot) \equiv \mathcal{L}(\eta)$, in which case (4) reduces to the identity of Lemma 4.11.

Applying (4) to functions of the form $f(\xi)$, we may extend many properties of ordinary expectations to a conditional setting. In particular, such extensions hold for the Jensen, Hölder, and Minkowski inequalities. The first of those yields the L^p-*contractivity*

$$\Big\| E^{\mathcal{F}} \xi \Big\|_p \leq \|\xi\|_p, \quad \xi \in L^p, \ p \geq 1.$$

Considering conditional distributions of entire sequences $(\xi, \xi_1, \xi_2, \ldots)$, we may further derive conditional versions of the basic continuity properties of ordinary integrals.

We list some simple applications of conditional distributions.

Lemma 8.6 *(conditioning criteria) For any random elements ξ, η in Borel spaces S, T, where $S = [0,1]$ in* (ii), *we have*

(i) $\xi \perp\!\!\!\perp \eta \Leftrightarrow \mathcal{L}(\xi) = \mathcal{L}(\xi \mid \eta)$ *a.s.* $\Leftrightarrow \mathcal{L}(\eta) = \mathcal{L}(\eta \mid \xi)$ *a.s.*,

(ii) ξ *is a.s. η-measurable* $\Leftrightarrow \xi = E(\xi \mid \eta)$ *a.s.*

Proof: (i) Compare (4) with Lemma 4.11.

(ii) Use Theorem 8.1 (v). □

In case of independence, we have some further useful properties.

Lemma 8.7 *(independence case) For any random elements $\xi \perp\!\!\!\perp \eta$ in S, T and a measurable function $f : S \times T \to U$, where S, T, U are Borel, define $X = f(\xi, \eta)$. Then*

(i) $\mathcal{L}(X \mid \xi) = \mathcal{L}\Big\{f(x, \eta)\Big\}\Big|_{x=\xi}$ *a.s.*,

(ii) X *is a.s. ξ-measurable* $\Leftrightarrow (X, \xi) \perp\!\!\!\perp \eta$.

Proof: (i) For any measurable function $g \geq 0$ on S, we get by (4) and Lemma 8.6 (i)

$$E\Big\{g(X);\, \xi \in B\Big\} = E\, E\Big\{1_B(x)(g \circ f)(x, \eta)\Big\}\Big|_{x=\xi},$$

and so

$$E\Big\{g(X)\,\Big|\,\xi\Big\} = E\Big\{(g \circ f)(x, \eta)\Big\}\Big|_{x=\xi} \text{ a.s.}$$

(ii) If X is ξ-measurable, then so is the pair (X, ξ) by Lemma 1.9, and the relation $\xi \perp\!\!\!\perp \eta$ yields $(X, \xi) \perp\!\!\!\perp \eta$. Conversely, assuming $(X, \xi) \perp\!\!\!\perp \eta$ and using (i) and Lemma 8.6 (i), we get a.s.

$$\begin{aligned}\mathcal{L}(X, \xi) &= \mathcal{L}(X, \xi \mid \eta)\\ &= \mathcal{L}\Big\{f(\xi, y), \xi\Big\}\Big|_{y=\eta},\end{aligned}$$

so that

$$(X, \xi) \stackrel{d}{=} \Big\{f(\xi, y), \xi\Big\}, \quad y \in T \text{ a.s. } \mathcal{L}(\eta).$$

Fixing a y with equality and applying the resulting relation to the set $A = \{(z, x);\, z = f(x, y)\}$, which is measurable in $U \times S$ by the measurability of the diagonal in U^2, we obtain

$$\begin{aligned}P\Big\{X = f(\xi, y)\Big\} &= P\Big\{(X, \xi) \in A\Big\}\\ &= P\Big\{\Big(f(\xi, y), \xi\Big) \in A\Big\}\\ &= P(\Omega) = 1.\end{aligned}$$

Hence, $X = f(\xi, y)$ a.s., which shows that X is a.s. ξ-measurable. □

Next we show how distributional invariance properties are preserved under suitable conditioning. Here a mapping T on S is said to be *bi-measurable*, if it is a measurable bijection with a measurable inverse.

Lemma 8.8 *(conditional invariance) For any random elements ξ, η in a Borel space S and a bi-measurable mapping T on S, we have*

$$T(\xi,\eta) \overset{d}{=} (\xi,\eta) \quad \Rightarrow \quad \Big\{\mathcal{L}(T\xi \,|\, \eta), T\xi\Big\} \overset{d}{=} \Big\{\mathcal{L}(\xi \,|\, \eta), \xi\Big\}.$$

Proof: Write $\mathcal{L}(\xi \,|\, \eta) = \mu(\eta, \cdot)$, and note that μ depends only on $\mathcal{L}(\xi,\eta)$. Since also $\sigma(T\eta) = \sigma(\eta)$ by the bi-measurability of T, we get a.s.

$$\begin{aligned}\mathcal{L}(T\xi \,|\, \eta) &= \mathcal{L}(T\xi \,|\, T\eta)\\ &= \mu(T\eta, \cdot),\end{aligned}$$

and so

$$\begin{aligned}\Big\{\mathcal{L}(T\xi \,|\, \eta), T\xi\Big\} &= \{\mu(T\eta, \cdot), T\xi\}\\ &\overset{d}{=} \{\mu(\eta, \cdot), \xi\}\\ &= \{\mathcal{L}(\xi \,|\, \eta), \xi\}.\end{aligned}$$

□

The notion of independence between σ-fields $\mathcal{F}_t$, $t \in T$, extends immediately to the conditional setting. Thus, we say that the $\mathcal{F}_t$ are *conditionally independent* given a σ-field $\mathcal{G}$, if for any distinct indices $t_1, \dots, t_n \in T$, $n \in \mathsf{N}$, we have

$$P^{\mathcal{G}} \bigcap_{k \le n} B_k = \prod_{k \le n} P^{\mathcal{G}} B_k \;\text{ a.s.}, \quad B_k \in \mathcal{F}_{t_k},\; k = 1, \dots, n.$$

This reduces to ordinary independence when $\mathcal{G}$ is the trivial σ-field $\{\Omega, \emptyset\}$. The pairwise version of conditional independence is denoted by $\perp\!\!\!\perp_{\mathcal{G}}$. Conditional independence involving events A_t or random elements ξ_t, $t \in T$, is defined as before in terms of the induced σ-fields $\sigma(A_t)$ or $\sigma(\xi_t)$, respectively, and the notation involving $\perp\!\!\!\perp$ carries over to this case.

In particular, any $\mathcal{F}$-measurable random elements ξ_t are conditionally independent given $\mathcal{F}$. If instead the ξ_t are independent of $\mathcal{F}$, their conditional independence given $\mathcal{F}$ is equivalent to ordinary independence between the ξ_t. By Theorem 8.5, every general statement or formula involving independence between countably many random elements in Borel spaces has a conditional counterpart. For example, Lemma 4.8 shows that the σ-fields $\mathcal{F}_1, \mathcal{F}_2, \dots$ are conditionally independent, given a σ-field $\mathcal{G}$, iff

$$(\mathcal{F}_1, \dots, \mathcal{F}_n) \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{F}_{n+1}, \quad n \in \mathsf{N}.$$

More can be said in the conditional case, and we begin with a basic characterization. Here and below, $\mathcal{F}, \mathcal{G}, \dots$ with or without subscripts denote sub-σ-fields of $\mathcal{A}$, and $\mathcal{F} \vee \mathcal{G} = \sigma(\mathcal{F}, \mathcal{G})$.

Theorem 8.9 *(conditional independence, Doob) For any σ-fields $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$, we have*

$$\mathcal{F} \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{H} \quad \Leftrightarrow \quad P^{\mathcal{F} \vee \mathcal{G}} = P^{\mathcal{G}} \;\textit{ a.s. on } \mathcal{H}.$$

Proof: Assuming the second condition and using the tower and pull-out properties of conditional expectations, we get for any $F \in \mathcal{F}$ and $H \in \mathcal{H}$

$$\begin{aligned} P^{\mathcal{G}}(F \cap H) &= E^{\mathcal{G}} P^{\mathcal{F}\vee\mathcal{G}}(F \cap H) \\ &= E^{\mathcal{G}}\big(P^{\mathcal{F}\vee\mathcal{G}} H; F\big) \\ &= E^{\mathcal{G}}\big(P^{\mathcal{G}} H; F\big) \\ &= (P^{\mathcal{G}} F)\,(P^{\mathcal{G}} H), \end{aligned}$$

showing that $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$. Conversely, assuming $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{H}$ and using the tower and pull-out properties, we get for any $F \in \mathcal{F}$, $G \in \mathcal{G}$, and $H \in \mathcal{H}$

$$\begin{aligned} E\big(P^{\mathcal{G}} H;\, F \cap G\big) &= E\big\{(P^{\mathcal{G}} F)\,(P^{\mathcal{G}} H); G\big\} \\ &= E\big\{P^{\mathcal{G}}(F \cap H); G\big\} \\ &= P(F \cap G \cap H). \end{aligned}$$

By a monotone-class argument, this extends to

$$E\big(P^{\mathcal{G}} H; A\big) = P(H \cap A), \quad A \in \mathcal{F} \vee \mathcal{G},$$

and the stated condition follows by the averaging characterization of $P^{\mathcal{F}\vee\mathcal{G}} H$. □

The following simple consequence is needed in Chapters 27–28.

Corollary 8.10 *(contraction) For any random elements ξ, η, ζ, we have*

$$\left.\begin{array}{c} (\xi,\eta) \overset{d}{=} (\xi,\zeta) \\ \sigma(\eta) \subset \sigma(\zeta) \end{array}\right\} \;\Rightarrow\; \xi \underset{\eta}{\perp\!\!\!\perp} \zeta.$$

Proof: For any measurable set B, we note that the variables

$$\begin{aligned} M_1 &= P\{\xi \in B \,|\, \eta\}, \\ M_2 &= P\{\xi \in B \,|\, \zeta\}, \end{aligned}$$

form a bounded martingale with $M_1 \overset{d}{=} M_2$. Then $E(M_2 - M_1)^2 = EM_2^2 - EM_1^2 = 0$, and so $M_1 = M_2$ a.s., which implies $\xi \perp\!\!\!\perp_\eta \zeta$ by Theorem 8.9. □

The last theorem yields some further useful properties. Let $\bar{\mathcal{G}}$ denote the *completion* of $\mathcal{G}$ with respect to the basic σ-field $\mathcal{A}$, generated by $\mathcal{G}$ and the family of null sets $\mathcal{N} = \{N \subset A;\ A \in \mathcal{A},\ PA = 0\}$.

Corollary 8.11 *(extension and inclusion) For any σ-fields $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$,*

(i) $\mathcal{F} \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{H} \;\Leftrightarrow\; \mathcal{F} \underset{\mathcal{G}}{\perp\!\!\!\perp} (\mathcal{G}, \mathcal{H})$,

(ii) $\mathcal{F} \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{F} \;\Leftrightarrow\; \mathcal{F} \subset \bar{\mathcal{G}}$.

Proof: (i) By Theorem 8.9, both relations are equivalent to

$$P(F|\,\mathcal{G},\mathcal{H}) = P(F|\,\mathcal{G}) \ \text{ a.s.,} \quad F \in \mathcal{F}.$$

(ii) If $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}$, Theorem 8.9 yields

$$\begin{aligned} 1_F &= P(F|\,\mathcal{F},\mathcal{G}) \\ &= P(F|\,\mathcal{G}) \ \text{ a.s.,} \quad F \in \mathcal{F}, \end{aligned}$$

which implies $\mathcal{F} \subset \bar{\mathcal{G}}$. Conversely, the latter relation implies

$$\begin{aligned} P(F|\,\mathcal{G}) &= P(F|\,\bar{\mathcal{G}}) = 1_F \\ &= P(F|\,\mathcal{F},\mathcal{G}) \ \text{ a.s.,} \quad F \in \mathcal{F}, \end{aligned}$$

and so $\mathcal{F} \perp\!\!\!\perp_{\mathcal{G}} \mathcal{F}$ by Theorem 8.9. □

The following basic result is often applied in both directions.

Theorem 8.12 *(chain rule) For any σ-fields $\mathcal{G}$, $\mathcal{H}$, $\mathcal{F}_1, \mathcal{F}_2, \ldots$, these conditions are equivalent:*

(i) $\mathcal{H} \underset{\mathcal{G}}{\perp\!\!\!\perp} (\mathcal{F}_1, \mathcal{F}_2, \ldots)$,

(ii) $\mathcal{H} \underset{\mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n}{\perp\!\!\!\perp} \mathcal{F}_{n+1}, \quad n \geq 0.$

In particular, we often need to employ the equivalence

$$\mathcal{H} \underset{\mathcal{G}}{\perp\!\!\!\perp} (\mathcal{F}, \mathcal{F}') \quad \Leftrightarrow \quad \mathcal{H} \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{F}, \ \ \mathcal{H} \underset{\mathcal{G},\mathcal{F}}{\perp\!\!\!\perp} \mathcal{F}'.$$

Proof: Assuming (i), we get by Theorem 8.9 for any $H \in \mathcal{H}$ and $n \geq 0$

$$\begin{aligned} P\Big(H \,\Big|\, \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n\Big) &= P(H \,|\, \mathcal{G}) \\ &= P\Big(H \,\Big|\, \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_{n+1}\Big), \end{aligned}$$

and (ii) follows by another application of Theorem 8.9.

Assuming (ii) instead, we get by Theorem 8.9 for any $H \in \mathcal{H}$

$$P\Big(H \,\Big|\, \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n\Big) = P\Big(H \,\Big|\, \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_{n+1}\Big), \quad n \geq 0.$$

Summing over $n < m$ gives

$$P(H \,|\, \mathcal{G}) = P\Big(H \,\Big|\, \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_m\Big), \quad m \geq 1,$$

and so Theorem 8.9 yields $\mathcal{H} \perp\!\!\!\perp_{\mathcal{G}} (\mathcal{F}_1, \ldots, \mathcal{F}_m)$ for all $m \geq 1$, which extends to (i) by a monotone-class argument. □

The last result is also useful to establish ordinary independence. Then take $\mathcal{G} = \{\emptyset, \Omega\}$ in Theorem 8.12 to get

$$\mathcal{H} \perp\!\!\!\perp (\mathcal{F}_1, \mathcal{F}_2, \ldots) \quad \Leftrightarrow \quad \mathcal{H} \underset{\mathcal{F}_1, \ldots, \mathcal{F}_n}{\perp\!\!\!\perp} \mathcal{F}_{n+1}, \ \ n \geq 0.$$

Reasoning with conditional expectations and independence is notoriously subtle and non-intuitive. It may then be helpful to translate the probabilistic notions into their geometric counterparts in suitable Hilbert spaces. For any σ-field $\mathcal{F}$ on Ω, we introduce the *centered*[2] sub-space $L_0^2(\mathcal{F}) \subset L^2$ of $\mathcal{F}$-measurable random variables ξ with $E\xi = 0$ and $E\xi^2 < \infty$. For sub-spaces $F \subset L_0^2$, let π_F denote projection onto F, and write $F^\perp$ for the orthogonal complement of F. Define $F \perp G$ by the orthogonality $E(\xi\eta) = 0$ of all variables $\xi \in F$ and $\eta \in G$, and write $F \ominus G = F \cap G^\perp$ and $F \vee G = \mathrm{span}(F, G)$. When $F \perp G$, we may also write $F \vee G = F \oplus G$. We further introduce the *conditional orthogonality*

$$F \underset{G}{\perp} H \quad \Leftrightarrow \quad \pi_{G^\perp} F \perp \pi_{G^\perp} H.$$

Theorem 8.13 *(conditional independence and orthogonality) Let $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$ be σ-fields on Ω generating the centered sub-spaces $F, G, H \subset L_0^2$. Then*

$$\mathcal{F} \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{H} \quad \Leftrightarrow \quad F \underset{G}{\perp} H.$$

By Theorem 8.1 (vi), the latter condition may also be written as $\pi_{G^\perp} F \perp H$ or $F \perp \pi_{G^\perp} H$. For the trivial σ-field $\mathcal{G} = \{\emptyset, \Omega\}$, we get the useful equivalence

$$\mathcal{F} \perp\!\!\!\perp \mathcal{H} \quad \Leftrightarrow \quad F \perp H, \tag{5}$$

which can also be verified directly.

Proof: Using Theorem 8.9 and Lemmas A4.5 and A4.7, we get

$$\begin{aligned}
\mathcal{F} \underset{\mathcal{G}}{\perp\!\!\!\perp} \mathcal{H} \quad &\Leftrightarrow \quad (\pi_{F \vee G} - \pi_G) H = 0 \\
&\Leftrightarrow \quad \pi_{(F \vee G) \ominus G} H = 0 \\
&\Leftrightarrow \quad (F \vee G) \ominus G \perp H \\
&\Leftrightarrow \quad \pi_{G^\perp} F \perp H,
\end{aligned}$$

where the last equivalence follows from the fact that, by Lemma A4.5,

$$\begin{aligned}
F \vee G &= (\pi_G F \oplus \pi_{G^\perp} F) \vee G \\
&= \pi_{G^\perp} F \vee G \\
&= \pi_{G^\perp} F \oplus G.
\end{aligned}$$

□

For a simple illustration, we note the following useful commutativity criterion, which follows immediately from Theorem 8.13 and Lemma A4.8.

Corollary 8.14 *(commutativity) For any σ-fields $\mathcal{F}, \mathcal{G}$, we have*

$$E_{\mathcal{F}} E_{\mathcal{G}} = E_{\mathcal{G}} E_{\mathcal{F}} = E_{\mathcal{F} \cap \mathcal{G}} \quad \Leftrightarrow \quad \mathcal{F} \underset{\mathcal{F} \cap \mathcal{G}}{\perp\!\!\!\perp} \mathcal{G}.$$

[2] Using $L_0^2(\mathcal{F})$ instead of $L^2(\mathcal{F})$ has some technical advantages. In particular, it makes the trivial σ-field $\{\emptyset, \Omega\}$ correspond to the null space $0 \subset L^2$.

Conditional distributions can sometimes be constructed by suitable iteration. To explain the notation, let ξ, η, ζ be random elements in some Borel spaces S, T, U. Since $S \times U$ and $T \times U$ are again Borel, there exist some probability kernels $\mu: S \to T \times U$ and $\mu': T \to S \times U$, such that a.s.

$$\begin{aligned}\mathcal{L}(\eta, \zeta \,|\, \xi) &= \mu(\xi, \cdot),\\ \mathcal{L}(\xi, \zeta \,|\, \eta) &= \mu'(\eta, \cdot),\end{aligned}$$

corresponding to the dual disintegrations

$$\begin{aligned}\mathcal{L}(\xi, \eta, \zeta) &= \mathcal{L}(\xi) \otimes \mu\\ &\tilde{=} \mathcal{L}(\eta) \otimes \mu'.\end{aligned}$$

For fixed s or t, we may regard μ_s and μ'_t as probability measures in their own right, and repeat the conditioning, leading to disintegrations of the form

$$\begin{aligned}\mu_s &= \bar{\mu}_s \otimes \nu_s,\\ \mu'_t &= \bar{\mu}'_t \otimes \nu'_t,\end{aligned}$$

where $\bar{\mu}_s = \mu_s(\cdot \times U)$ and $\bar{\mu}'_t = \mu'_t(\cdot \times U)$, for some kernels $\nu_s : T \to U$ and $\nu'_t : S \to U$. It is suggestive to write even the latter relations in terms of conditioning, as in

$$\begin{aligned}\mu_s\Big(\tilde{\zeta} \in \cdot \,\Big|\, \tilde{\eta}\Big) &= \nu_s(\tilde{\eta}, \cdot),\\ \mu'_t\Big(\tilde{\zeta} \in \cdot \,\Big|\, \tilde{\xi}\Big) &= \nu'_t(\tilde{\xi}, \cdot),\end{aligned}$$

for the coordinate variables $\tilde{\xi}, \tilde{\eta}, \tilde{\zeta}$ in S, T, U.

Since all spaces are Borel, Corollary 3.6 allows us to choose ν_s and ν'_t as kernels $S \times T \to U$. With a slight abuse of notation, we may write the iterated conditioning in the form

$$\begin{aligned}P(\cdot \,|\, \eta \,|\, \xi)_{s,t} &= \Big\{P(\cdot \,|\, \xi)_s\Big\}(\cdot \,|\, \eta)_t,\\ P(\cdot \,|\, \xi \,|\, \eta)_{t,s} &= \Big\{P(\cdot \,|\, \eta)_t\Big\}(\cdot \,|\, \xi)_s.\end{aligned}$$

Substituting $s = \xi$ and $t = \eta$, and putting $\mathcal{F} = \sigma(\xi)$, $\mathcal{G} = \sigma(\eta)$, and $\mathcal{H} = \sigma(\zeta)$, we get some (ξ, η)-measurable random probability measures on $(\Omega, \mathcal{H})$, here written suggestively as

$$\begin{aligned}(P_\mathcal{F})_\mathcal{G} &= P(\cdot \,|\, \eta \,|\, \xi)_{\xi,\eta}\,,\\ (P_\mathcal{G})_\mathcal{F} &= P(\cdot \,|\, \xi \,|\, \eta)_{\eta,\xi}\,.\end{aligned}$$

Using this notation, we may state the basic properties of iterated conditioning in a striking form. Here we say that a σ-field $\mathcal{F}$ is *Borel generated*, if $\mathcal{F} = \sigma(\xi)$ for some random element ξ in a Borel space.

Theorem 8.15 *(iterated conditioning) For any Borel generated σ-fields $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$, we have a.s. on $\mathcal{H}$*

(i) $(P_\mathcal{F})_\mathcal{G} = (P_\mathcal{G})_\mathcal{F} = P_{\mathcal{F} \vee \mathcal{G}}$,

(ii) $P_\mathcal{F} = E_\mathcal{F}(P_\mathcal{F})_\mathcal{G}$.

Part (i) must not be confused with the elementary commutativity in Corollary 8.14. Similarly, we must not confuse (ii) with the elementary tower property in Theorem 8.1 (vii), which may be written as

$$P_{\mathcal{F}} = E_{\mathcal{F}} P_{\mathcal{F}\vee\mathcal{G}} \text{ a.s.}$$

Proof: Writing

$$\mathcal{F} = \sigma(\xi), \qquad \mathcal{G} = \sigma(\eta), \qquad \mathcal{H} = \sigma(\zeta),$$

for some random elements in Borel spaces S, T, U, and putting

$$\mu_1 = \mathcal{L}(\xi), \qquad \mu_2 = \mathcal{L}(\eta), \qquad \mu_{12} = \mathcal{L}(\xi,\eta),$$
$$\mu_{13} = \mathcal{L}(\xi,\zeta), \qquad \mu_{123} = \mathcal{L}(\xi,\eta,\zeta),$$

we get from Theorem 3.7 the relations

$$\begin{aligned} \mu_{3|2|1} &= \mu_{3|12} \\ &\cong \mu_{3|1|2} \text{ a.e. } \mu_{12}, \\ \mu_{3|1} &= \mu_{2|1}\,\mu_{3|12} \text{ a.e. } \mu_1, \end{aligned}$$

which are equivalent to (i) and (ii). We can also obtain (ii) directly from (i), by noting that

$$E_{\mathcal{F}}(P_{\mathcal{F}})_{\mathcal{G}} = E_{\mathcal{F}} P_{\mathcal{F}\vee\mathcal{G}} = P_{\mathcal{F}},$$

by the tower property in Theorem 8.1. □

Regular conditional distributions can be used to construct random elements with desired properties. This may require an extension of the basic probability space. By an *extension* of $(\Omega, \mathcal{A}, P)$ we mean a product space $(\hat{\Omega}, \hat{\mathcal{A}}) = (\Omega \times S, \mathcal{A} \otimes \mathcal{S})$, equipped with a probability measure $\hat{P}$ satisfying $\hat{P}(\cdot \times S) = P$. Any random element ξ on Ω may be regarded as a function on $\hat{\Omega}$. Thus, we simply replace ξ by the random element $\hat{\xi}(\omega, s) = \xi(\omega)$ on $\hat{\Omega}$, which has the same distribution. For extensions of this type, we retain our original notation and write P and ξ instead of $\hat{P}$ and $\hat{\xi}$.

We begin with an elementary extension suggested by Theorem 8.5. The result is needed for various constructions in Chapters 27–28.

Lemma 8.16 *(extension) For any probability kernel $\mu\colon S \to T$ and random elements ξ, ζ in S, U, there exists a random element η in T, defined on a suitable extension of the probability space, such that*

$$\mathcal{L}(\eta \mid \xi, \zeta) = \mu(\xi, \cdot) \text{ a.s.}, \qquad \eta \underset{\xi}{\perp\!\!\!\perp} \zeta.$$

Proof: Put $(\hat{\Omega}, \hat{\mathcal{A}}) = (\Omega \times T, \mathcal{A} \otimes \mathcal{T})$, where $\mathcal{T}$ denotes the σ-field in T, and define a probability measure $\hat{P}$ on $\hat{\Omega}$ by

$$\hat{P}A = E \int 1_A(\cdot, t)\,\mu(\xi, dt), \quad A \in \hat{\mathcal{A}}.$$

Then clearly $\hat{P}(\cdot \times T) = P$, and the random element $\eta(\omega, t) \equiv t$ on $\hat{\Omega}$ satisfies $\hat{\mathcal{L}}(\eta \,|\, \mathcal{A}) = \mu(\xi, \cdot)$ a.s. In particular, Proposition 8.9 yields $\eta \perp\!\!\!\perp_\xi \mathcal{A}$. $\square$

Most constructions require only a single *randomization variable*, defined as a $U(0,1)$ random variable ϑ, independent of all previously introduced random elements and σ-fields. We always assume the basic probability space to be rich enough to support any randomization variables we need. This involves no essential loss of generality, since the condition becomes fulfilled after a simple extension of the original space. Thus, we may take

$$\hat{\Omega} = \Omega \times [0,1], \quad \hat{\mathcal{A}} = \mathcal{A} \otimes \mathcal{B}_{[0,1]}, \quad \hat{P} = P \otimes \lambda,$$

where λ denotes Lebesgue measure on $[0,1]$. Then $\vartheta(\omega, t) \equiv t$ is $U(0,1)$ on $\hat{\Omega}$ and $\vartheta \perp\!\!\!\perp \mathcal{A}$. By Lemma 4.21 we may use ϑ to produce a whole sequence of independent randomization variables $\vartheta_1, \vartheta_2, \ldots$, if needed.

The following powerful and commonly used result shows how a probabilistic structure can be transferred between different contexts through a suitable randomization.

Theorem 8.17 *(transfer) Let ξ, η, ζ be random elements in S, T, U, where T is Borel. Then for any $\tilde{\xi} \overset{d}{=} \xi$, there exists a random element $\tilde{\eta}$ in T with*

(i) $(\tilde{\xi}, \tilde{\eta}) \overset{d}{=} (\xi, \eta)$, (ii) $\tilde{\eta} \underset{\tilde{\xi}}{\perp\!\!\!\perp} \zeta$.

Proof: (i) By Theorem 8.5, there exists a probability kernel $\mu : S \to T$ satisfying

$$\mu(\xi, B) = P\{\eta \in B \,|\, \xi\} \text{ a.s.}, \quad B \in \mathcal{T}.$$

Next, Lemma 4.22 yields a measurable function $f : S \times [0,1] \to T$, such that for any $U(0,1)$ random variable $\vartheta \perp\!\!\!\perp (\tilde{\xi}, \zeta)$,

$$\mathcal{L}\Big\{f(s, \vartheta)\Big\} = \mu(s, \cdot), \quad s \in S.$$

Defining $\tilde{\eta} = f(\tilde{\xi}, \vartheta)$ and using Lemmas 1.24 and 4.11 together with Theorem 8.5, we get for any measurable function $g : S \times T \to \mathbb{R}_+$

$$\begin{aligned} E\, g(\tilde{\xi}, \tilde{\eta}) &= E\, g\Big\{\tilde{\xi}, f(\tilde{\xi}, \vartheta)\Big\} \\ &= E \int g\Big\{\xi, f(\xi, u)\Big\}\, du \\ &= E \int g(\xi, t)\, \mu(\xi, dt) \\ &= E g(\xi, \eta), \end{aligned}$$

which shows that $(\tilde{\xi}, \tilde{\eta}) \overset{d}{=} (\xi, \eta)$.

(ii) By Theorem 8.12 we have $\vartheta \perp\!\!\!\perp_{\tilde{\xi}} \zeta$, and so Corollary 8.11 yields $(\tilde{\xi}, \vartheta) \perp\!\!\!\perp_{\tilde{\xi}} \zeta$, which implies $\tilde{\eta} \perp\!\!\!\perp_{\tilde{\xi}} \zeta$. $\square$

The last result can be used to transfer representations of random objects:

Corollary 8.18 *(stochastic equations) For any random elements ξ, η in Borel spaces S, T and a measurable map $f: T \to S$ with $\xi \overset{d}{=} f(\eta)$, we may choose $\tilde{\eta}$ with*

$$\tilde{\eta} \overset{d}{=} \eta, \qquad \xi = f(\tilde{\eta}) \ \text{a.s.}$$

Proof: By Theorem 8.17 there exists a random element $\tilde{\eta}$ in T with $(\xi, \tilde{\eta}) \overset{d}{=} \{f(\eta), \eta\}$. In particular, $\tilde{\eta} \overset{d}{=} \eta$ and $\{\xi, f(\tilde{\eta})\} \overset{d}{=} \{f(\eta), f(\eta)\}$. Since the diagonal in S^2 is measurable, we get

$$P\Big\{\xi = f(\tilde{\eta})\Big\} = P\Big\{f(\eta) = f(\eta)\Big\} = 1,$$

and so $\xi = f(\tilde{\eta})$ a.s. □

This leads in particular to a useful extension of Theorem 5.31.

Corollary 8.19 *(extended Skorohod coupling) Let $f, f_1, f_2, \ldots$ be measurable functions from a Borel space S to a Polish space T, and let $\xi, \xi_1, \xi_2, \ldots$ be random elements in S with $f_n(\xi_n) \xrightarrow{d} f(\xi)$. Then we may choose $\tilde{\xi}, \tilde{\xi}_1, \tilde{\xi}_2, \ldots$ with*

$$\tilde{\xi} \overset{d}{=} \xi, \quad \tilde{\xi}_n \overset{d}{=} \xi_n, \qquad f_n(\tilde{\xi}_n) \to f(\tilde{\xi}) \ \text{a.s.}$$

Proof: By Theorem 5.31 there exist some $\eta \overset{d}{=} f(\xi)$ and $\eta_n \overset{d}{=} f_n(\xi_n)$ with $\eta_n \to \eta$ a.s. By Corollary 8.18 we may further choose $\tilde{\xi} \overset{d}{=} \xi$ and $\tilde{\xi}_n \overset{d}{=} \xi_n$, such that a.s. $f(\tilde{\xi}) = \eta$ and $f_n(\tilde{\xi}_n) = \eta_n$ for all n. But then $f_n(\tilde{\xi}_n) \to f(\tilde{\xi})$ a.s. □

We proceed to clarify the relationship between conditional independence and randomizations. Important applications appear in Chapters 11, 13, 27–28, and 32.

Proposition 8.20 *(conditional independence by randomization) Let ξ, η, ζ be random elements in S, T, U, where S is Borel. Then these conditions are equivalent:*

(i) $\xi \underset{\eta}{\perp\!\!\!\perp} \zeta$,

(ii) *$\xi = f(\eta, \vartheta)$ a.s. for a measurable function $f: T \times [0,1] \to S$ and a $U(0,1)$ random variable $\vartheta \perp\!\!\!\perp (\eta, \zeta)$.*

Proof: The implication (ii) ⇒ (i) holds by the argument for Theorem 8.17 (ii). Now assume (i), and let $\vartheta \perp\!\!\!\perp (\eta, \zeta)$ be $U(0,1)$. Then Theorem 8.17 yields a measurable function $f : T \times [0,1] \to S$, such that the random element $\tilde{\xi} = f(\eta, \vartheta)$ satisfies $\tilde{\xi} \overset{d}{=} \xi$ and $(\tilde{\xi}, \eta) \overset{d}{=} (\xi, \eta)$. We further see from the sufficiency part that $\tilde{\xi} \perp\!\!\!\perp_\eta \zeta$. Hence, Proposition 8.9 yields

$$\begin{aligned} \mathcal{L}\{\tilde{\xi} \,|\, \eta, \zeta\} &= \mathcal{L}(\tilde{\xi} \,|\, \eta) \\ &= \mathcal{L}(\xi \,|\, \eta) \\ &= \mathcal{L}\{\xi \,|\, \eta, \zeta\}, \end{aligned}$$

and so $(\tilde{\xi}, \eta, \zeta) \overset{d}{=} (\xi, \eta, \zeta)$. By Theorem 8.17, we may next choose $\tilde{\vartheta} \overset{d}{=} \vartheta$ with $(\xi, \eta, \zeta, \tilde{\vartheta}) \overset{d}{=} (\tilde{\xi}, \eta, \zeta, \vartheta)$. In particular, $\tilde{\vartheta} \perp\!\!\!\perp (\eta, \zeta)$ and $\{\xi, f(\eta, \tilde{\vartheta})\} \overset{d}{=} \{\tilde{\xi}, f(\eta, \vartheta)\}$. Since $\tilde{\xi} = f(\eta, \vartheta)$, and the diagonal in S^2 is measurable, we get $\xi = f(\eta, \tilde{\vartheta})$ a.s., which is the required relation with $\tilde{\vartheta}$ in place of ϑ. □

The transfer theorem can be used to construct random sequences or processes with given finite-dimensional distributions. For any measurable spaces $S_1, S_2, \ldots$, we say that a sequence of distributions[3] μ_n on $S_1 \times \cdots \times S_n$, $n \in \mathsf{N}$, is *projective* if

$$\mu_{n+1}(\cdot \times S_{n+1}) = \mu_n, \quad n \in \mathsf{N}. \tag{6}$$

Theorem 8.21 *(existence of random sequences, Daniell) For any projective sequence of distributions μ_n on $S_1 \times \cdots \times S_n$, $n \in \mathsf{N}$, where $S_2, S_3, \ldots$ are Borel, there exist some random elements ξ_n in S_n, $n \in \mathsf{N}$, such that*

$$\mathcal{L}(\xi_1, \ldots, \xi_n) = \mu_n, \quad n \in \mathsf{N}.$$

Proof: By Lemmas 4.10 and 4.21 there exist some independent random variables $\xi_1, \vartheta_2, \vartheta_3, \ldots$, such that $\mathcal{L}(\xi_1) = \mu_1$ and the ϑ_n are i.i.d. $U(0,1)$. We proceed recursively to construct $\xi_2, \xi_3, \ldots$ with the stated properties, such that each ξ_n is a measurable function of $\xi_1, \vartheta_2, \ldots, \vartheta_n$. Once $\xi_1, \ldots, \xi_n$ are constructed, let $\mathcal{L}(\eta_1, \ldots, \eta_{n+1}) = \mu_{n+1}$. Then the projective property yields $(\xi_1, \ldots, \xi_n) \overset{d}{=} (\eta_1, \ldots, \eta_n)$, and so by Theorem 8.17 we may form ξ_{n+1} as a measurable function of $\xi_1, \ldots, \xi_n, \vartheta_{n+1}$, such that $(\xi_1, \ldots, \xi_{n+1}) \overset{d}{=} (\eta_1, \ldots, \eta_{n+1})$. This completes the recursion. □

Next we show how a given process can be extended to any unbounded index set. The result is stated in an abstract form, designed to fulfill our needs in especially Chapters 12 and 17. Let ι denote the identity map on any space.

Corollary 8.22 *(projective limits) Consider some Borel spaces $S, S_1, S_2, \ldots$ and measurable maps $\pi_n \colon S \to S_n$ and $\pi_k^n \colon S_n \to S_k$, $k \le n$, such that*

$$\pi_k^n = \pi_k^m \circ \pi_m^n, \quad k \le m \le n. \tag{7}$$

Let $\bar{S}$ be the class of sequences $(s_1, s_2, \ldots) \in S_1 \times S_2 \times \cdots$ with $\pi_k^n s_n = s_k$ for all $k \le n$, and let the map $h \colon \bar{S} \to S$ be measurable with $(\pi_1, \pi_2, \ldots) \circ h = \iota$ on $\bar{S}$. Then for any distributions μ_n on S_n with

$$\mu_n \circ (\pi_k^n)^{-1} = \mu_k, \quad k \le n \in \mathsf{N}, \tag{8}$$

there exists a distribution μ on S with $\mu \circ \pi_n^{-1} = \mu_n$ for all n.

Proof: Introduce the measures

$$\bar{\mu}_n = \mu_n \circ (\pi_1^n, \ldots, \pi_n^n)^{-1}, \quad n \in \mathsf{N}, \tag{9}$$

[3] probability measures

and conclude from (7) and (8) that

$$\begin{aligned}\bar\mu_{n+1}(\cdot \times S_{n+1}) &= \mu_{n+1} \circ \left(\pi_1^{n+1}, \dots, \pi_n^{n+1}\right)^{-1}\\ &= \mu_{n+1} \circ (\pi_n^{n+1})^{-1} \circ (\pi_1^n, \dots, \pi_n^n)^{-1}\\ &= \mu_n \circ (\pi_1^n, \dots, \pi_n^n)^{-1} = \bar\mu_n.\end{aligned}$$

By Theorem 8.21 there exists a measure $\bar\mu$ on $S_1 \times S_2 \times \cdots$ with

$$\bar\mu \circ (\bar\pi_1, \dots, \bar\pi_n)^{-1} = \bar\mu_n, \quad n \in \mathsf{N}, \tag{10}$$

where $\bar\pi_1, \bar\pi_2, \dots$ denote the coordinate projections in $S_1 \times S_2 \times \cdots$. From (7), (9), and (10) we see that $\bar\mu$ is restricted to $\bar S$, which allows us to define $\mu = \bar\mu \circ h^{-1}$. It remains to note that

$$\begin{aligned}\mu \circ \pi_n^{-1} &= \bar\mu \circ (\pi_n h)^{-1}\\ &= \bar\mu \circ \bar\pi_n^{-1}\\ &= \bar\mu_n \circ \bar\pi_n^{-1}\\ &= \mu_n \circ (\pi_n^n)^{-1} = \mu_n.\end{aligned}$$ □

We often need a version of Theorem 8.21 for processes on a general index set T. Then for any collection of measurable spaces $(S_t, \mathcal{S}_t)$, $t \in T$, we define $(S_I, \mathcal{S}_I) = \bigotimes_{t\in I}(S_t, \mathcal{S}_t)$, $I \subset T$. For any random elements ξ_t in S_t, write ξ_I for the restriction of the process (ξ_t) to the index set I.

Now let $\hat T$ and $\bar T$ be the classes of finite and countable subsets of T, respectively. A family of probability measures μ_I, $I \in \hat T$ or $\bar T$, is said to be *projective* if

$$\mu_J(\cdot \times S_{J\setminus I}) = \mu_I, \quad I \subset J \text{ in } \hat T \text{ or } \bar T. \tag{11}$$

Theorem 8.23 *(existence of processes, Kolmogorov) For any Borel spaces S_t, $t \in T$, consider a projective family of probability measures μ_I on S_I, $I \in \hat T$. Then there exist some random elements X_t in S_t, $t \in T$, such that*

$$\mathcal{L}(X_I) = \mu_I, \quad I \in \hat T.$$

Proof: The product σ-field $\mathcal{S}_T$ in S_T is generated by all coordinate projections π_t, $t \in T$, and hence consists of all countable *cylinder* sets $B \times S_{T\setminus U}$, $B \in \mathcal{S}_U$, $U \in \bar T$. For every $U \in \bar T$, Theorem 8.21 yields a probability measure μ_U on S_U with

$$\mu_U(\cdot \times S_{U\setminus I}) = \mu_I, \quad I \in \hat U,$$

and Proposition 4.2 shows that the family μ_U, $U \in \bar T$, is again projective. We may then define a function $\mu\colon \mathcal{S}_T \to [0,1]$ by

$$\mu(\cdot \times S_{T\setminus U}) = \mu_U, \quad U \in \bar T.$$

To show that μ is countably additive, consider any disjoint sets $A_1, A_2, \ldots \in \mathcal{S}_T$. For every n we have $A_n = B_n \times S_{T\setminus U_n}$ for some $U_n \in \bar T$ and $B_n \in \mathcal{S}_{U_n}$. Writing $U = \bigcup_n U_n$ and $C_n = B_n \times S_{U\setminus U_n}$, we get

$$\begin{aligned}\mu \bigcup_n A_n &= \mu_U \bigcup_n C_n \\ &= \sum_n \mu_U C_n \\ &= \sum_n \mu A_n.\end{aligned}$$

We may now define the process $X = (X_t)$ as the identity map on the probability space $(S_T, \mathcal{S}_T, \mu)$. □

If the projective sequence in Theorem 8.21 is defined recursively in terms of conditional distributions, then no regularity condition is needed on the state spaces. For a precise statement, define the composition $\mu \otimes \nu$ of two kernels μ and ν as in Chapter 3.

Theorem 8.24 *(extension by conditioning, Ionescu Tulcea) For any measurable spaces $(S_n, \mathcal{S}_n)$ and probability kernels $\mu_n : S_1 \times \cdots \times S_{n-1} \to S_n$, $n \in \mathsf{N}$, there exist some random elements ξ_n in S_n, $n \in \mathsf{N}$, such that*

$$\mathcal{L}(\xi_1, \ldots, \xi_n) = \mu_1 \otimes \cdots \otimes \mu_n, \quad n \in \mathsf{N}.$$

Proof: Put $\mathcal{F}_n = \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n$ and $T_n = S_{n+1} \times S_{n+2} \times \cdots$, and note that the class $\mathcal{C} = \bigcup_n (\mathcal{F}_n \times T_n)$ is a field in T_0 generating the σ-field $\mathcal{F}_\infty$. Define an additive function μ on $\mathcal{C}$ by

$$\mu(A \times T_n) = (\mu_1 \otimes \cdots \otimes \mu_n)A, \quad A \in \mathcal{F}_n,\ n \in \mathsf{N}, \tag{12}$$

which is clearly independent of the representation $C = A \times T_n$. We need to extend μ to a probability measure on $\mathcal{F}_\infty$. By Theorem 2.5, it is then enough to show that μ is continuous at $\emptyset$.

For any $C_1, C_2, \ldots \in \mathcal{C}$ with $C_n \downarrow \emptyset$, we need to show that $\mu C_n \to 0$. Renumbering if necessary, we may assume for every n that $C_n = A_n \times T_n$ with $A_n \in \mathcal{F}_n$. Now define

$$f_k^n = (\mu_{k+1} \otimes \cdots \otimes \mu_n) 1_{A_n}, \quad k \le n, \tag{13}$$

with the understanding that $f_n^n = 1_{A_n}$ for $k = n$. By Lemmas 3.2 and 3.3, each f_k^n is an $\mathcal{F}_k$-measurable function on $S_1 \times \cdots \times S_k$, and (13) yields

$$f_k^n = \mu_{k+1} f_{k+1}^n, \quad 0 \le k < n. \tag{14}$$

Since $C_n \downarrow \emptyset$, the functions f_k^n are non-increasing in n for fixed k, say with limits g_k. By (14) and dominated convergence,

$$g_k = \mu_{k+1} g_{k+1}, \quad k \ge 0. \tag{15}$$

Combining (12) and (13) gives $\mu C_n = f_0^n \downarrow g_0$. If $g_0 > 0$, then (15) yields an $s_1 \in S_1$ with $g_1(s_1) > 0$. Continuing recursively, we may construct a sequence $\bar{s} = (s_1, s_2, \ldots) \in T_0$, such that $g_n(s_1, \ldots, s_n) > 0$ for all n. Then

$$\begin{aligned}1_{C_n}(\bar{s}) &= 1_{A_n}(s_1,\ldots,s_n)\\ &= f_n^n(s_1,\ldots,s_n)\\ &\geq g_n(s_1,\ldots,s_n) > 0,\end{aligned}$$

and so $\bar{s} \in \bigcap_n C_n$, which contradicts the assumption $C_n \downarrow \emptyset$. Thus, $g_0 = 0$, which means that $\mu C_n \to 0$. $\square$

In particular, we may choose some independent random elements with arbitrarily prescribed distributions. The result extends the elementary Theorem 4.19.

Corollary 8.25 *(infinite product measures, Łomnicki & Ulam) For any probability spaces $(S_t, \mathcal{S}_t, \mu_t)$, $t \in T$, there exist some independent random elements ξ_t in S_t with*

$$\mathcal{L}(\xi_t) = \mu_t, \quad t \in T.$$

Proof: For countable subsets $I \subset T$, the associated product measures $\mu_I = \bigotimes_{t\in I} \mu_t$ exist by Theorem 8.24. Now proceed as in the proof of Theorem 8.23. $\square$

Exercises

1. Show that $(\xi, \eta) \overset{d}{=} (\xi', \eta)$ iff $P(\xi \in B \,|\, \eta) = P(\xi' \in B \,|\, \eta)$ a.s. for any measurable set B.

2. Show that $E^{\mathcal{F}}\xi = E^{\mathcal{G}}\xi$ a.s. for all $\xi \in L^1$ iff $\bar{\mathcal{F}} = \bar{\mathcal{G}}$.

3. Show that the averaging property implies the other properties of conditional expectations listed in Theorem 8.1.

4. State the probabilistic counterpart of Lemma A4.9, and give a direct proof.

5. Let $0 \leq \xi_n \uparrow \xi$ and $0 \leq \eta \leq \xi$, where $\xi_1, \xi_2, \ldots, \eta \in L^1$, and fix a σ-field $\mathcal{F}$. Show that $E^{\mathcal{F}}\eta \leq \sup_n E^{\mathcal{F}}\xi_n$. (*Hint:* Apply the monotone convergence property to $E^{\mathcal{F}}(\xi_n \wedge \eta)$.)

6. For any $[0,\infty]$-valued random variable ξ, define $E^{\mathcal{F}}\xi = \sup_n E^{\mathcal{F}}(\xi \wedge n)$. Show that this extension of $E^{\mathcal{F}}$ satisfies the monotone convergence property. (*Hint:* Use the preceding result.)

7. Show that the above extension of $E^{\mathcal{F}}$ remains characterized by the averaging property, and that $E^{\mathcal{F}}\xi < \infty$ a.s. iff the measure $\xi \cdot P = E(\xi; \cdot)$ is σ-finite on $\mathcal{F}$. Extend $E^{\mathcal{F}}\xi$ to random variables ξ such that the measure $|\xi| \cdot P$ is σ-finite on $\mathcal{F}$.

8. Let $\xi_1, \xi_2, \ldots$ be $[0,\infty]$-valued random variables, and fix any σ-field $\mathcal{F}$. Show that $\liminf_n E^{\mathcal{F}}\xi_n \geq E^{\mathcal{F}} \liminf_n \xi_n$ a.s.

9. For any σ-field $\mathcal{F}$, and let $\xi, \xi_1, \xi_2, \ldots$ be random variables with $\xi_n \to \xi$ and $E^{\mathcal{F}} \sup_n |\xi_n| < \infty$ a.s. Show that $E^{\mathcal{F}}\xi_n \to E^{\mathcal{F}}\xi$ a.s.

10. Let the σ-field $\mathcal{F}$ be generated by a partition $A_1, A_2, \ldots \in \mathcal{A}$ of Ω. For any $\xi \in L^1$, show that $E(\xi \,|\mathcal{F}) = E(\xi \,|A_k) = E(\xi; A_k)/PA_k$ on A_k whenever $PA_k > 0$.

11. For any σ-field $\mathcal{F}$, event A, and random variable $\xi \in L^1$, show that $E(\xi \,|\mathcal{F}, 1_A) = E(\xi; A \,|\mathcal{F})/P(A \,|\mathcal{F})$ a.s. on A.

12. Let the random variables $\xi_1, \xi_2, \ldots \geq 0$ and σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be such that $E(\xi_n|\mathcal{F}_n) \xrightarrow{P} 0$. Show that $\xi_n \xrightarrow{P} 0$. (*Hint:* Consider the random variables $\xi_n \wedge 1$.)

13. Let $(\xi, \eta) \overset{d}{=} (\tilde{\xi}, \tilde{\eta})$, where $\xi \in L^1$. Show that $E(\xi \,|\, \eta) \overset{d}{=} E(\tilde{\xi}\,|\tilde{\eta})$. (*Hint:* If $E(\xi \,|\, \eta) = f(\eta)$, then $E(\tilde{\xi}\,|\tilde{\eta}) = f(\tilde{\eta})$ a.s.)

14. Let (ξ, η) be a random vector in R^2 with probability density f, put $F(y) = \int f(x, y)\, dx$, and let $g(x, y) = f(x, y)/F(y)$. Show that $P(\xi \in B \,|\, \eta) = \int_B g(x, \eta)\, dx$ a.s.

15. Use conditional distributions to deduce the monotone and dominated convergence theorems for conditional expectations from the corresponding unconditional results.

16. Assume $E^{\mathcal{F}}\xi \overset{d}{=} \xi$ for a $\xi \in L^1$. Show that ξ is a.s. $\mathcal{F}$-measurable. (*Hint:* Choose a strictly convex function f with $Ef(\xi) < \infty$, and apply the strict Jensen inequality to the conditional distributions.)

17. Assuming $(\xi, \eta) \overset{d}{=} (\xi, \zeta)$, where η is ζ-measurable, show that $\xi \perp\!\!\!\perp_\eta \zeta$. (*Hint:* Show as above that $P(\xi \in B \,|\, \eta) \overset{d}{=} P(\xi \in B|\, \zeta)$, and deduce the corresponding a.s. equality.)

18. Let ξ be a random element in a separable metric space S. Show that $\mathcal{L}(\xi\,|\mathcal{F})$ is a.s. degenerate iff ξ is a.s. $\mathcal{F}$-measurable. (*Hint:* Reduce to the case where $\mathcal{L}(\xi\,|\mathcal{F})$ is degenerate everywhere and hence equal to δ_η for an $\mathcal{F}$-measurable random element η in S. Then show that $\xi = \eta$ a.s.)

19. Give a direct proof of the equivalence (5). Then state the probabilistic counterpart of Lemma A4.7, and give a direct proof.

20. Show that if $\mathcal{G} \perp\!\!\!\perp_{\mathcal{F}_n} \mathcal{H}$ for some increasing σ-fields $\mathcal{F}_n$, then $\mathcal{G} \perp\!\!\!\perp_{\mathcal{F}_\infty} \mathcal{H}$.

21. Assuming $\xi \perp\!\!\!\perp_\eta \zeta$ and $\gamma \perp\!\!\!\perp (\xi, \eta, \zeta)$, show that $\xi \perp\!\!\!\perp_{\eta,\gamma} \zeta$ and $\xi \perp\!\!\!\perp_\eta (\zeta, \gamma)$.

22. Extend Lemma 4.6 to the context of conditional independence. Also show that Corollary 4.7 and Lemma 4.8 remain valid for conditional independence, given a σ-field $\mathcal{H}$.

23. Consider any σ-field $\mathcal{F}$ and random element ξ in a Borel space, and define $\eta = \mathcal{L}(\xi\,|\mathcal{F})$. Show that $\xi \perp\!\!\!\perp_\eta \mathcal{F}$.

24. Let ξ, η be random elements in a Borel space S. Prove the existence of a measurable function $f\colon S \times [0, 1] \to S$ and a $U(0, 1)$ random variable $\gamma \perp\!\!\!\perp \eta$ such that $\xi = f(\eta, \gamma)$ a.s. (*Hint:* Choose f with $\{f(\eta, \vartheta), \eta\} \overset{d}{=} (\xi, \eta)$ for any $U(0, 1)$ random variable $\vartheta \perp\!\!\!\perp (\xi, \eta)$, and then let $(\gamma, \tilde{\eta}) \overset{d}{=} (\vartheta, \eta)$ with $(\xi, \eta) = \{f(\gamma, \tilde{\eta}), \tilde{\eta}\}$ a.s.)

25. Let ξ, η be random elements in Borel spaces S, T such that $\xi = f(\eta)$ a.s. for a measurable function $f\colon T \to S$. Show that $\eta = g(\xi, \vartheta)$ a.s. for a measurable function $g\colon S \times [0, 1] \to T$ and an $U(0, 1)$ random variable $\vartheta \perp\!\!\!\perp \xi$. (*Hint:* Use Theorem 8.17.)

26. Suppose that the function $f\colon T \to S$ above is injective. Show that $\eta = h(\xi)$ a.s. for a measurable function $h\colon S \to T$. (*Hint:* Show that in this case $\eta \perp\!\!\!\perp \vartheta$, and define $h(x) = \lambda g(x, \cdot)$.) Compare with Theorem A1.1.

27. Let ξ, η be random elements in a Borel space S. Show that we can choose a random element $\tilde{\eta}$ in S with $(\xi, \eta) \overset{d}{=} (\xi, \tilde{\eta})$ and $\eta \perp\!\!\!\perp_\xi \tilde{\eta}$.

28. Let the probability measures P, Q on $(\Omega, \mathcal{A})$ be related by $Q = \xi \cdot P$ for a random variable $\xi \geq 0$, and fix a σ-field $\mathcal{F} \subset \mathcal{A}$. Show that $Q = E_P(\xi\,|\mathcal{F}) \cdot P$ on $\mathcal{F}$.

29. Assume as before that $Q = \xi \cdot P$ on $\mathcal{A}$, and let $\mathcal{F} \subset \mathcal{A}$. Show that $E_Q(\eta \,|\mathcal{F}) = E_P(\xi\eta \,|\mathcal{F})/E_P(\xi \,|\mathcal{F})$ a.s. Q for any random variable $\eta \geq 0$.

Chapter 9

Optional Times and Martingales

Filtrations, strictly and weakly optional times, closure properties, optional evaluation and hitting, augmentation, random time change, sub- and super-martingales, centering and convex maps, optional sampling and stopping, martingale transforms, maximum and upcrossing inequalities, sub-martingale convergence, closed martingales, limits of conditional expectations, regularization of sub-martingales, increasing limits of super-martingales

The importance of martingales and related topics can hardly be exaggerated. Indeed, filtrations and optional times as well as a wide range of sub- and super-martingales are constantly used in all areas of modern probability. They appear frequently throughout the remainder of this book.

In discrete time, a martingale is simply a sequence of integrable random variables centered at successive conditional means, a centering that can always be achieved by the elementary *Doob decomposition*. More precisely, given any discrete *filtration* $\mathcal{F} = (\mathcal{F}_n)$, defined as an increasing sequence of σ-fields in Ω, we say that the random variables $M_0, M_1, \ldots$ form a *martingale* with respect to $\mathcal{F}$ if $E(M_n \,|\, \mathcal{F}_{n-1}) = M_{n-1}$ a.s. for all n. A special role is played by the *uniformly integrable martingales*, which can be represented in the form $M_n = E(\xi \,|\mathcal{F}_n)$ for some integrable random variables ξ.

Martingale theory owes its usefulness to a number of powerful general results, such as the *optional sampling theorem*, the *sub-martingale convergence theorem*, and a variety of *maximum inequalities*. Applications discussed in this chapter include extensions of the Borel–Cantelli lemma and Kolmogorov's $0-1$ law. Martingales can also be used to establish the existence of measurable densities and to give a short proof of the law of large numbers.

Much of the discrete-time theory extends immediately to continuous time, thanks to the fundamental *regularization theorem*, which ensures that every continuous-time martingale with respect to a right-continuous filtration has a right-continuous version with left-hand limits. The implications of this result extend far beyond martingale theory. In particular, it will enable us in Chapters 16–17 to obtain right-continuous versions of independent-increment and Feller processes.

The theory of continuous-time martingales is continued in Chapters 10, 18–20, and 35 with studies of quadratic variation, random time-change, integral representations, removal of drift, additional maximum inequalities, and various decomposition theorems. Martingales also play a basic role for especially the Skorohod embedding in Chapter 22, the stochastic integration in Chapters 18

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_10

and 20, and the theories of Feller processes, SDEs, and diffusions in Chapters 17 and 32–33.

As for the closely related notion of optional times, our present treatment is continued with a more detailed study in Chapter 10. Optional times are fundamental not only for martingale theory, but also for various models involving Markov processes. In the latter context they appear frequently throughout the remainder of the book.

To begin our systematic exposition of the theory, we fix an arbitrary index set $T \subset \bar{\mathsf{R}}$. A *filtration* on T is defined as a non-decreasing family of σ-fields $\mathcal{F}_t \subset \mathcal{A}$, $t \in T$. We say that a process X on T is *adapted* to $\mathcal{F} = (\mathcal{F}_t)$ if X_t is $\mathcal{F}_t$-measurable for every $t \in T$. The smallest filtration with this property, given by $\mathcal{F}_t = \sigma\{X_s;\ s \le t\}$, $t \in T$, is called the *induced* or *generated* filtration. Here 'smallest' is understood in the sense of set inclusion for each t.

A *random time* is defined as a random element τ in $\bar{T} = T \cup \{\sup T\}$. We say that τ is $\mathcal{F}$-*optional*[1] if $\{\tau \le t\} \in \mathcal{F}_t$ for every $t \in T$, meaning that the process $X_t = 1\{\tau \le t\}$ is adapted[2]. When T is countable, it is clearly equivalent that $\{\tau = t\} \in \mathcal{F}_t$ for every $t \in T$.

With every optional time τ we associate a σ-field

$$\mathcal{F}_\tau = \Big\{A \in \mathcal{A};\ A \cap \{\tau \le t\} \in \mathcal{F}_t,\ t \in T\Big\}.$$

We list some basic properties of optional times and corresponding σ-fields.

Lemma 9.1 *(optional times) For any optional times σ, τ,*

(i) *$\sigma \vee \tau$ and $\sigma \wedge \tau$ are again optional,*

(ii) *$\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$,*

(iii) *$\mathcal{F}_\sigma \subset \mathcal{F}_\tau$ on $\{\sigma \le \tau\}$,*

(iv) *τ is $\mathcal{F}_\tau$-measurable,*

(v) *if $\tau \equiv t \in \bar{T}$, then τ is optional with $\mathcal{F}_\tau = \mathcal{F}_t$.*

Proof: (i) For any $t \in T$,

$$\{\sigma \vee \tau \le t\} = \{\sigma \le t\} \cap \{\tau \le t\} \in \mathcal{F}_t,$$
$$\{\sigma \wedge \tau > t\} = \{\sigma > t\} \cap \{\tau > t\} \in \mathcal{F}_t.$$

(ii) For any $A \in \mathcal{F}_\sigma$ and $t \in T$,

$$\begin{aligned} &A \cap \{\sigma \le \tau\} \cap \{\tau \le t\} \\ &\quad = \big(A \cap \{\sigma \le t\}\big) \cap \{\tau \le t\} \cap \Big\{\sigma \wedge t \le \tau \wedge t\Big\}, \end{aligned}$$

which belongs to $\mathcal{F}_t$ since $\sigma \wedge t$ and $\tau \wedge t$ are both $\mathcal{F}_t$-measurable. Hence,

$$\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_\tau.$$

[1] Optional times are often called *stopping times*.

[2] We often omit the prefix '$\mathcal{F}$-' when there is no risk for confusion.

The first relation now follows as we replace τ by $\sigma \wedge \tau$. Replacing σ and τ by the pairs $(\sigma \wedge \tau, \sigma)$ and $(\sigma \wedge \tau, \tau)$ gives $\mathcal{F}_{\sigma\wedge\tau} \subset \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. To prove the reverse relation, we note that for any $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$ and $t \in T$,

$$A \cap \{\sigma \wedge \tau \leq t\} = \big(A \cap \{\sigma \leq t\}\big) \cup \big(A \cap \{\tau \leq t\}\big) \in \mathcal{F}_t,$$

whence $A \in \mathcal{F}_{\sigma\wedge\tau}$.

(iii) Clear from (ii).

(iv) Applying (ii) to the pair (τ, t) gives $\{\tau \leq t\} \in \mathcal{F}_\tau$ for all $t \in T$, which extends immediately to any $t \in \mathsf{R}$. Now use Lemma 1.4.

(v) Clearly $\mathcal{F}_\tau = \mathcal{F}_\tau \cap \{\tau \leq t\} \subset \mathcal{F}_t$. Conversely, let $A \in \mathcal{F}_t$ and $s \in T$. When $s \geq t$ we get $A \cap \{\tau \leq s\} = A \in \mathcal{F}_t \subset \mathcal{F}_s$, whereas for $s < t$ we have $A \cap \{\tau \leq s\} = \emptyset \in \mathcal{F}_s$. Thus, $A \in \mathcal{F}_\tau$, proving that $\mathcal{F}_t \subset \mathcal{F}_\tau$. □

Given a filtration $\mathcal{F}$ on R_+, we may define a new filtration $\mathcal{F}^+$ by $\mathcal{F}_t^+ = \bigcap_{u>t} \mathcal{F}_u$, $t \geq 0$, and say that $\mathcal{F}$ is *right-continuous* if $\mathcal{F}^+ = \mathcal{F}$. In particular, $\mathcal{F}^+$ is right-continuous for any filtration $\mathcal{F}$. We say that a random time τ is *weakly $\mathcal{F}$-optional* if $\{\tau < t\} \in \mathcal{F}_t$ for every $t > 0$. Then $\tau + h$ is clearly $\mathcal{F}$-optional for every $h > 0$, and we may define $\mathcal{F}_{\tau+} = \bigcap_{h>0} \mathcal{F}_{\tau+h}$. When the index set is Z_+, we take $\mathcal{F}^+ = \mathcal{F}$ and make no difference between strictly and weakly optional times.

The notions of optional and weakly optional times agree when $\mathcal{F}$ is right-continuous:

Lemma 9.2 *(weakly optional times) A random time τ is weakly $\mathcal{F}$-optional iff it is $\mathcal{F}^+$-optional, in which case*

$$\mathcal{F}_{\tau+} = \mathcal{F}_\tau^+ = \big\{A \in \mathcal{A};\ A \cap \{\tau < t\} \in \mathcal{F}_t,\ t > 0\big\}. \tag{1}$$

Proof: For any $t \geq 0$, we note that

$$\{\tau \leq t\} = \bigcap_{r>t} \{\tau < r\}, \qquad \{\tau < t\} = \bigcup_{r<t} \{\tau \leq r\}, \tag{2}$$

where r may be restricted to the rationals. If $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+}$ for all t, we get by (2) for any $t > 0$

$$A \cap \{\tau < t\} = \bigcup_{r<t} \big(A \cap \{\tau \leq r\}\big) \in \mathcal{F}_t.$$

Conversely, if $A \cap \{\tau < t\} \in \mathcal{F}_t$ for all t, then (2) yields for any $t \geq 0$ and $h > 0$

$$A \cap \{\tau \leq t\} = \bigcap_{r \in (t, t+h)} \big(A \cap \{\tau < r\}\big) \in \mathcal{F}_{t+h},$$

and so $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+}$. For $A = \Omega$ this proves the first assertion, and for general $A \in \mathcal{A}$ it proves the second relation in (1).

To prove the first relation, we note that $A \in \mathcal{F}_{\tau+}$ iff $A \in \mathcal{F}_{\tau+h}$ for each $h > 0$, that is, iff $A \cap \{\tau + h \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$ and $h > 0$. But this is equivalent to $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+h}$ for all $t \geq 0$ and $h > 0$, hence to $A \cap \{\tau \leq t\} \in \mathcal{F}_{t+}$

for every $t \geq 0$, which means that $A \in \mathcal{F}_\tau^+$. □

We have seen that the maximum and minimum of two optional times are again optional. The result extends to countable collections, as follows.

Lemma 9.3 *(closure properties) For any random times $\tau_1, \tau_2, \ldots$ and filtration $\mathcal{F}$ on R_+ or Z_+, we have*

(i) *if the τ_n are $\mathcal{F}$-optional, then so is $\tau = \sup_n \tau_n$,*

(ii) *if the τ_n are weakly $\mathcal{F}$-optional, then so is $\sigma = \inf_n \tau_n$, and*

$$\mathcal{F}_\sigma^+ = \bigcap_n \mathcal{F}_{\tau_n}^+.$$

Proof: To prove (i) and the first assertion in (ii), we write

$$\{\tau \leq t\} = \bigcap_n \{\tau_n \leq t\}, \qquad \{\sigma < t\} = \bigcup_n \{\tau_n < t\}, \tag{3}$$

where the strict inequalities may be replaced by $\leq$ for the index set $T = \mathsf{Z}_+$. To prove the second assertion in (ii), we note that $\mathcal{F}_\sigma^+ \subset \bigcap_n \mathcal{F}_{\tau_n}^+$ by Lemma 9.1. Conversely, assuming $A \in \bigcap_n \mathcal{F}_{\tau_n}^+$, we get by (3) for any $t \geq 0$

$$\begin{aligned} A \cap \{\sigma < t\} &= A \cap \bigcup_n \{\tau_n < t\} \\ &= \bigcup_n \big(A \cap \{\tau_n < t\}\big) \in \mathcal{F}_t, \end{aligned}$$

with the indicated modification for $T = \mathsf{Z}_+$. Thus, $A \in \mathcal{F}_\sigma^+$. □

In particular, part (ii) of the last result is useful for the approximation of optional times from the right.

Lemma 9.4 *(discrete approximation) For any weakly optional time τ in $\bar{\mathsf{R}}_+$, there exist some countably-valued optional times $\tau_n \downarrow \tau$.*

Proof: We may define

$$\tau_n = 2^{-n}[2^n \tau + 1], \quad n \in \mathsf{N}.$$

Then $\tau_n \in 2^{-n}\bar{\mathsf{N}}$ for all n, and $\tau_n \downarrow \tau$. The τ_n are optional, since

$$\{\tau_n \leq k\,2^{-n}\} = \{\tau < k\,2^{-n}\} \in \mathcal{F}_{k2^{-n}} \quad k, n \in \mathsf{N}.$$ □

We may now relate the optional times to random processes. Say that a process X on R_+ is *progressively measurable* or simply *progressive*, if its restriction to $\Omega \times [0, t]$ is $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$ -measurable for every $t \geq 0$. Note that any progressive process is adapted by Lemma 1.28. Conversely, we may approximate from the left or right to show that any adapted and left- or right-continuous process is progressive. A set $A \subset \Omega \times \mathsf{R}_+$ is said to be progressive if the corresponding indicator function 1_A has this property, and we note that the progressive sets form a σ-field.

Lemma 9.5 *(optional evaluation) Consider a filtration $\mathcal{F}$ on an index set T, a process X on T with values in a measurable space $(S,\mathcal{S})$, and an optional time τ in T. Then X_τ is $\mathcal{F}_\tau$-measurable under each of these conditions:*
(i) *T is countable and X is adapted,*
(ii) *$T = \mathsf{R}_+$ and X is progressive.*

Proof: In both cases, we need to show that

$$\left\{X_\tau \in B,\, \tau \le t\right\} \in \mathcal{F}_t, \quad t \ge 0,\ B \in \mathcal{S}.$$

This is clear in case (i), if we write

$$\{X_\tau \in B\} = \bigcup_{s \le t} \left\{X_s \in B,\, \tau = s\right\} \in \mathcal{F}_t, \quad B \in \mathcal{S}.$$

In case (ii) it is enough to show that $X_{\tau\wedge t}$ is $\mathcal{F}_t$-measurable for every $t \ge 0$. We may then take $\tau \le t$ and prove instead that X_τ is $\mathcal{F}_t$-measurable. Writing $X_\tau = X \circ \psi$ with $\psi(\omega) = \{\omega, \tau(\omega)\}$, we note that ψ is measurable from $\mathcal{F}_t$ to $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$, whereas X is measurable on $\Omega \times [0,t]$ from $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$ to $\mathcal{S}$. The required measurability of X_τ now follows by Lemma 1.7. □

For any process X on R_+ or Z_+ and a set B in the range space of X, we introduce the *hitting time*

$$\tau_B = \inf\left\{t > 0;\, X_t \in B\right\}.$$

We often need to certify that τ_B is optional. The following elementary result covers some common cases.

Lemma 9.6 *(hitting times) For a filtration $\mathcal{F}$ on $T = \mathsf{R}_+$ or Z_+, let X be an $\mathcal{F}$-adapted process on T with values in a measurable space $(S,\mathcal{S})$, and let $B \in \mathcal{S}$. Then τ_B is weakly optional under each of these conditions:*
(i) *$T = \mathsf{Z}_+$,*
(ii) *$T = \mathsf{R}_+$, S is a metric space, B is closed, and X is continuous,*
(iii) *$T = \mathsf{R}_+$, S is a topological space, B is open, and X is right-continuous.*

Proof: (i) Write

$$\{\tau_B \le n\} = \bigcup_{k \in [1,n]} \{X_k \in B\} \in \mathcal{F}_n, \quad n \in \mathsf{N}.$$

(ii) Letting ρ be the metric on S, we get for any $t > 0$

$$\{\tau_B \le t\} = \bigcup_{h>0} \bigcap_{n \in \mathsf{N}} \bigcup_{r \in \mathsf{Q} \cap [h,t]} \left\{\rho(X_r, B) \le n^{-1}\right\} \in \mathcal{F}_t.$$

(iii) By Lemma 9.2, it suffices to write

$$\{\tau_B < t\} = \bigcup_{r \in \mathsf{Q} \cap (0,t)} \{X_r \in B\} \in \mathcal{F}_t, \quad t > 0.$$ □

For special purposes, we need the following more general but much deeper result, known as the *debut theorem*. Here and below, a filtration $\mathcal{F}$ is said to be *complete*, if the basic σ-field $\mathcal{A}$ is complete and each $\mathcal{F}_t$ contains all P-null sets in $\mathcal{A}$.

Theorem 9.7 *(first entry, Doob, Hunt) Let the set $A \subset \mathsf{R}_+ \times \Omega$ be progressive for a right-continuous, complete filtration $\mathcal{F}$. Then the time $\tau(\omega) = \inf\{t \geq 0;\ (t,\omega) \in A\}$ is $\mathcal{F}$-optional.*

Proof: Since A is progressive, $A \cap [0,t) \in \mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$ for every $t > 0$. Noting that $\{\tau < t\}$ is the projection of $A \cap [0,t)$ onto Ω, we get $\{\tau < t\} \in \mathcal{F}_t$ by Theorem A1.2, and so τ is optional by Lemma 9.2. □

For applications of this and other results, we may need to extend a given filtration $\mathcal{F}$ on R_+, to make it both right-continuous and complete. Writing $\overline{\mathcal{A}}$ for the completion of $\mathcal{A}$, put $\mathcal{N} = \{A \in \overline{\mathcal{A}};\ PA = 0\}$ and define $\overline{\mathcal{F}}_t = \sigma\{\mathcal{F}_t, \mathcal{N}\}$. Then $\overline{\mathcal{F}} = (\overline{\mathcal{F}}_t)$ is the smallest complete extension of $\mathcal{F}$. Similarly, $\mathcal{F}^+ = (\mathcal{F}_{t+})$ is the smallest right-continuous extension of $\mathcal{F}$. We show that the two extensions commute and can be combined into a smallest right-continuous and complete version, known as the (usual) *augmentation* of $\mathcal{F}$.

Lemma 9.8 *(augmented filtration) Any filtration $\mathcal{F}$ on R_+ has a smallest right-continuous, complete extension $\mathcal{G}$, given by*

$$\mathcal{G}_t = \overline{\mathcal{F}_{t+}} = \big(\overline{\mathcal{F}}\big)_{t+}, \quad t \geq 0. \tag{4}$$

Proof: First we note that

$$\overline{\mathcal{F}_{t+}} \subset \overline{\big(\overline{\mathcal{F}}\big)_{t+}} \subset \big(\overline{\mathcal{F}}\big)_{t+}, \quad t \geq 0.$$

Conversely, let $A \in \big(\overline{\mathcal{F}}\big)_{t+}$. Then $A \in \big(\overline{\mathcal{F}}\big)_{t+h}$ for every $h > 0$, and so as in Lemma 1.27 there exist some sets $A_h \in \mathcal{F}_{t+h}$ with $P(A \triangle A_h) = 0$. Now choose $h_n \to 0$, and define $A' = \{A_{h_n} \text{ i.o.}\}$. Then $A' = \mathcal{F}_{t+}$ and $P(A \triangle A') = 0$, and so $A \in \overline{\mathcal{F}_{t+}}$. Thus, $\big(\overline{\mathcal{F}}\big)_{t+} \subset \overline{\mathcal{F}_{t+}}$, which proves the second relation in (4).

In particular, the filtration $\mathcal{G}$ in (4) contains $\mathcal{F}$ and is both right-continuous and complete. For any filtration $\mathcal{H}$ with those properties, we have

$$\begin{aligned} \mathcal{G}_t &= \overline{\mathcal{F}}_{t+} \subset \overline{\mathcal{H}}_{t+} \\ &= \mathcal{H}_{t+} = \mathcal{H}_t, \quad t \geq 0, \end{aligned}$$

which proves the required minimality of $\mathcal{G}$. □

The σ-fields $\mathcal{F}_\tau$ arise naturally in connection with a random time change:

Proposition 9.9 *(random time-change) Let $X \geq 0$ be a non-decreasing, right-continuous process, adapted to a right-continuous filtration $\mathcal{F}$, and define*

$$\tau_s = \inf\Big\{t > 0;\ X_t > s\Big\}, \quad s \geq 0.$$

Then

(i) *the τ_s form a right-continuous process of optional times, generating a right-continuous filtration $\mathcal{G}_s = \mathcal{F}_{\tau_s}$, $s \geq 0$,*

(ii) *if X is continuous and τ is $\mathcal{F}$-optional, then X_τ is $\mathcal{G}$-optional with $\mathcal{F}_\tau \subset \mathcal{G}_{X_\tau}$,*

(iii) *when X is strictly increasing, we have $\mathcal{F}_\tau = \mathcal{G}_{X_\tau}$.*

In case (iii), we have in particular $\mathcal{F}_t = \mathcal{G}_{X_t}$ for all t, so that the processes (τ_s) and (X_t) play symmetric roles.

Proof, (i)−(ii): The times τ_s are optional by Lemmas 9.2 and 9.6, and since (τ_s) is right-continuous, so is $(\mathcal{G}_s)$ by Lemma 9.3. If X is continuous, then Lemma 9.1 yields for any $\mathcal{F}$-optional time $\tau > 0$ and set $A \in \mathcal{F}_\tau$

$$\begin{aligned} A \cap \{X_\tau \le s\} &= A \cap \{\tau \le \tau_s\} \\ &\in \mathcal{F}_{\tau_s} = \mathcal{G}_s, \quad s \ge 0. \end{aligned}$$

For $A = \Omega$ it follows that X_τ is $\mathcal{G}$-optional, and for general A we get $A \in \mathcal{G}_{X_\tau}$. Thus, $\mathcal{F}_\tau \subset \mathcal{G}_{X_\tau}$. Both statements extend by Lemma 9.3 to arbitrary τ.

(iii) For any $A \in \mathcal{G}_{X_t}$ with $t > 0$, we have

$$\begin{aligned} A \cap \{t \le \tau_s\} &= A \cap \{X_t \le s\} \\ &\in \mathcal{G}_s = \mathcal{F}_{\tau_s}, \quad s \ge 0, \end{aligned}$$

and so

$$A \cap \left\{t \le \tau_s \le u\right\} \in \mathcal{F}_u, \quad s \ge 0,\ u > t.$$

Taking the union over all $s \in \mathsf{Q}_+$ gives[3] $A \in \mathcal{F}_u$, and as $u \downarrow t$ we get $A \in \mathcal{F}_{t+} = \mathcal{F}_t$. Hence, $\mathcal{F}_t = \mathcal{G}_{X_t}$, which extends as before to $t = 0$. By Lemma 9.1, we obtain for any $A \in \mathcal{G}_{X_\tau}$

$$\begin{aligned} A \cap \{\tau \le t\} &= A \cap \{X_\tau \le X_t\} \\ &\in \mathcal{G}_{X_t} = \mathcal{F}_t, \quad t \ge 0, \end{aligned}$$

and so $A \in \mathcal{F}_\tau$. Thus, $\mathcal{G}_{X_\tau} \subset \mathcal{F}_\tau$, and so the two σ-fields agree. □

To motivate the definition of martingales, fix a filtration $\mathcal{F}$ on an index set T and a random variable $\xi \in L^1$, and introduce the process

$$M_t = E(\xi \,|\, \mathcal{F}_t), \quad t \in T.$$

Then M is clearly integrable (for each t) and adapted, and the tower property of conditional expectations yields

$$M_s = E(M_t | \, \mathcal{F}_s) \ \text{ a.s.}, \quad s \le t. \tag{5}$$

Any integrable and adapted process M satisfying (5) is called a *martingale with respect to* $\mathcal{F}$, or an $\mathcal{F}$*-martingale*. When $T = \mathsf{Z}_+$, it is enough to require (5) for $t = s + 1$, so that the condition becomes

$$E\Big(\Delta M_n \,\Big|\, \mathcal{F}_{n-1}\Big) = 0 \ \text{ a.s.}, \quad n \in \mathsf{N}, \tag{6}$$

[3]Recall that Q_+ denotes the set of rational numbers $r \ge 0$.

where $\Delta M_n = M_n - M_{n-1}$. A martingale in R^d is a process $M = (M^1, \ldots, M^d)$ where $M^1, \ldots, M^d$ are one-dimensional martingales.

We also consider the cases where (5) or (6) are replaced by corresponding inequalities. Thus, we define a *sub-martingale* as an integrable and adapted process X satisfying

$$X_s \leq E(X_t | \mathcal{F}_s) \text{ a.s.}, \quad s \leq t. \tag{7}$$

Reversing the inequality sign[4] yields the notion of a *super-martingale*. In particular, the mean is non-decreasing for sub-martingales and non-increasing for super-martingales.

Given a filtration $\mathcal{F}$ on Z_+, we say that a random sequence $A = (A_n)$ with $A_0 = 0$ is predictable with respect to $\mathcal{F}$ or simply $\mathcal{F}$*-predictable*, if A_n is $\mathcal{F}_{n-1}$-measurable for every $n \in \mathsf{N}$, so that the shifted sequence $(\theta A)_n = A_{n+1}$ is adapted. The following elementary result, known as the *Doob decomposition*, is often used to derive results for sub-martingales from the corresponding martingale versions. An extension to continuous time is proved in Chapter 10.

Lemma 9.10 *(centering, Doob) An integrable, $\mathcal{F}$-adapted process X on Z_+ has an a.s. unique decomposition $M + A$, where M is an $\mathcal{F}$-martingale and A is $\mathcal{F}$-predictable with $A_0 = 0$. In particular, X is a sub-martingale iff A is a.s. non-decreasing.*

Proof: If $X = M + A$ for some processes M and A as stated, then clearly $\Delta A_n = E(\Delta X_n | \mathcal{F}_{n-1})$ a.s. for all $n \in \mathsf{N}$, and so

$$A_n = \sum_{k \leq n} E\Big(\Delta X_k \,\Big|\, \mathcal{F}_{k-1}\Big) \text{ a.s.}, \quad n \in \mathsf{Z}_+, \tag{8}$$

which proves the required uniqueness. In general, we may define a predictable process A by (8). Then $M = X - A$ is a martingale, since

$$E\Big(\Delta M_n \,\Big|\, \mathcal{F}_{n-1}\Big) = E\Big(\Delta X_n \,\Big|\, \mathcal{F}_{n-1}\Big) - \Delta A_n = 0 \text{ a.s.}, \quad n \in \mathsf{N}. \qquad \square$$

Next we show how the martingale and sub-martingale properties are preserved by suitable transformations.

Lemma 9.11 *(convex maps) Consider a random sequence $M = (M_n)$ in R^d and a convex function $f : \mathsf{R}^d \to \mathsf{R}$, such that $X = f(M)$ is integrable. Then X is a sub-martingale under each of these conditions:*

(i) *M is a martingale in R^d,*

(ii) *M is a sub-martingale in R and f is non-decreasing.*

[4]The sign conventions are suggested by analogy with sub- and super-harmonic functions. Note that sub-martingales tend to be increasing whereas super-martingales tend to be decreasing. For branching processes the conventions are the opposite.

Proof: (i) By the conditional version of Jensen's inequality, we have for $s \le t$

$$\begin{aligned} f(M_s) &= f\big\{E(M_t|\,\mathcal{F}_s)\big\} \\ &\le E\Big\{f(M_t)\Big|\,\mathcal{F}_s\Big\} \quad \text{a.s.} \end{aligned} \tag{9}$$

(ii) Here the first relation in (9) becomes $f(M_s) \le f\{E(M_t|\,\mathcal{F}_s)\}$, and the conclusion remains valid. □

The last result is often applied with $f(x) = |x|^p$ for some $p \ge 1$, or with $f(x) = x_+ = x \vee 0$ when $d = 1$. We turn to a basic version of the powerful *optional sampling theorem.* An extension to continuous-time sub-martingales appears as Theorem 9.30. Say that an optional time τ is *bounded* if $\tau \le u$ a.s. for some $u \in T$. This always holds when T has a last element.

Theorem 9.12 *(optional sampling, Doob) Let M be a martingale on a countable index set T with filtration $\mathcal{F}$, and consider some optional times σ, τ, where τ is bounded. Then M_τ is integrable, and*

$$M_{\sigma\wedge\tau} = E(M_\tau|\,\mathcal{F}_\sigma) \quad a.s.$$

Proof: By Lemmas 8.3 and 9.1, we get for any $t \le u$ in T

$$\begin{aligned} E(M_u|\,\mathcal{F}_\tau) &= E(M_u|\,\mathcal{F}_t) \\ &= M_t = M_\tau \ \text{ a.s. on } \{\tau = t\}, \end{aligned}$$

and so $E(M_u|\,\mathcal{F}_\tau) = M_\tau$ a.s. whenever $\tau \le u$ a.s. If $\sigma \le \tau \le u$, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$ by Lemma 9.1, and we get a.s.

$$\begin{aligned} E(M_\tau|\,\mathcal{F}_\sigma) &= E\Big\{E(M_u|\,\mathcal{F}_\tau)\Big|\,\mathcal{F}_\sigma\Big\} \\ &= E(M_u|\,\mathcal{F}_\sigma) = M_\sigma. \end{aligned}$$

Further note that $E(M_\tau|\,\mathcal{F}_\sigma) = M_\tau$ a.s. when $\tau \le \sigma \wedge u$. In general, using Lemmas 8.3 and 9.1, we may combine the two special cases to get a.s.

$$\begin{aligned} E(M_\tau|\,\mathcal{F}_\sigma) &= E(M_\tau|\,\mathcal{F}_{\sigma\wedge\tau}) \\ &= M_{\sigma\wedge\tau} \ \text{ on } \{\sigma \le \tau\}, \\ E(M_\tau|\mathcal{F}_\sigma) &= E(M_{\sigma\wedge\tau}|\,\mathcal{F}_\sigma) \\ &= M_{\sigma\wedge\tau} \ \text{ on } \{\sigma > \tau\}. \end{aligned}$$

□

In particular, the martingale property is preserved by a random time change:

Corollary 9.13 *(random-time change) Consider an $\mathcal{F}$-martingale M and a non-decreasing family of bounded optional times τ_s taking countably many values. Then the process N is a $\mathcal{G}$-martingale, where*

$$N_s = M_{\tau_s}, \qquad \mathcal{G}_s = \mathcal{F}_{\tau_s}.$$

This leads in turn to a useful martingale criterion.

Corollary 9.14 *(martingale criterion) Let M be an integrable, adapted process on T. Then these conditions are equivalent:*

(i) *M is a martingale,*

(ii) *$EM_\sigma = EM_\tau$ for any bounded, optional times σ, τ taking countably many values in T.*

In (ii) *it is enough to consider optional times taking at most two values.*

Proof: If $s < t$ in T and $A \in \mathcal{F}_s$, then $\tau = s1_A + t1_{A^c}$ is optional, and so

$$\begin{aligned} 0 &= EM_t - EM_\tau \\ &= EM_t - E(M_s; A) - E(M_t; A^c) \\ &= E(M_t - M_s; A). \end{aligned}$$

Since A is arbitrary, it follows that $E(M_t - M_s | \mathcal{F}_s) = 0$ a.s. □

The following predictable transformation of martingales is basic for the theory of stochastic integration.

Corollary 9.15 *(martingale transform) Let M be a martingale on an index set T with filtration $\mathcal{F}$, fix an optional time τ taking countably many values, and let η be a bounded, $\mathcal{F}_\tau$-measurable random variable. Then we may form another martingale*

$$N_t = \eta\,(M_t - M_{t\wedge\tau}), \quad t \in T.$$

Proof: The integrability holds by Theorem 9.12, and the adaptedness is clear if we replace η by $\eta\,1\{\tau \le t\}$ in the expression for N_t. Now fix any bounded, optional time σ taking countably many values. By Theorem 9.12 and the pull-out property of conditional expectations, we get a.s.

$$\begin{aligned} E(N_\sigma | \mathcal{F}_\tau) &= \eta\, E\Big(M_\sigma - M_{\sigma\wedge\tau} \Big| \mathcal{F}_\tau\Big) \\ &= \eta\,(M_{\sigma\wedge\tau} - M_{\sigma\wedge\tau}) = 0, \end{aligned}$$

and so $EN_\sigma = 0$. Thus, N is a martingale by Lemma 9.14. □

In particular, the martingale property is preserved by *optional stopping*, in the sense that the *stopped* process $M_t^\tau = M_{\tau\wedge t}$ is a martingale whenever M is a martingale and τ is an optional time taking countably many values. More generally, we may consider *predictable step processes* of the form

$$V_t = \sum_{k\le n} \eta_k 1\{t > \tau_k\}, \quad t \in T,$$

where $\tau_1 \le \cdots \le \tau_n$ are optional times and each η_k is a bounded, $\mathcal{F}_{\tau_k}$-measurable random variable. For any process X, the associated *elementary stochastic integral*

$$(V \cdot X)_t \equiv \int_0^t V_s\, dX_s = \sum_{k\le n} \eta_k (X_t - X_{t\wedge\tau_k}), \quad t \in T,$$

is a martingale by Corollary 9.15, whenever X is a martingale and each τ_k takes countably many values. In discrete time, we may take V to be a bounded, predictable sequence, in which case

$$(V \cdot X)_n = \sum_{k \le n} V_k \, \Delta X_k, \quad n \in \mathsf{Z}_+.$$

The result for martingales extends in an obvious way to sub-martingales X, provided that $V \ge 0$.

We proceed with some basic martingale inequalities, beginning with some extensions of Kolmogorov's maximum inequality from Lemma 5.15.

Theorem 9.16 *(maximum inequalities, Bernstein, Lévy) Let X be a sub-martingale on a countable index set T. Then for any $r \ge 0$ and $u \in T$,*

(i) $r\, P\Big\{\sup_{t \le u} X_t \ge r\Big\} \le E\Big(X_u; \sup_{t \le u} X_t \ge r\Big) \le EX_u^+$,

(ii) $r\, P\Big\{\sup_t |X_t| \ge r\Big\} \le 3\, \sup_t E|X_t|$.

Proof: (i) By dominated convergence it is enough to consider finite index sets, so we may take $T = \mathsf{Z}_+$. Define $\tau = u \wedge \inf\{t;\, X_t \ge r\}$ and $B = \{\max_{t \le u} X_t \ge r\}$. Then τ is an optional time bounded by u, and we have $B \in \mathcal{F}_\tau$ and $X_\tau \ge r$ on B. Hence, Lemma 9.10 and Theorem 9.12 yield

$$\begin{aligned} r\, PB &\le E(X_\tau; B) \\ &\le E(X_u; B) \le EX_u^+. \end{aligned}$$

(ii) For the Doob decomposition $X = M + A$, relation (i) applied to $-M$ yields

$$\begin{aligned} r\, P\Big\{\min_{t \le u} X_t \le -r\Big\} &\le r\, P\Big\{\min_{t \le u} M_t \le -r\Big\} \\ &\le EM_u^- = EM_u^+ - EM_u \\ &\le EX_u^+ - EX_0 \\ &\le 2 \max_{t \le u} E|X_t|. \end{aligned}$$

In remains to combine with (i). □

We turn to a powerful norm inequality. For processes X on an index set T, we define

$$X_t^* = \sup_{s \le t} |X_s|, \qquad X^* = \sup_{t \in T} |X_t|.$$

Theorem 9.17 *(L^p-inequality, Doob) Let M be a martingale on a countable index set T, and fix any $p, q > 1$ with $p^{-1} + q^{-1} = 1$. Then*

$$\|M_t^*\|_p \le q\, \|M_t\|_p, \quad t \in T.$$

Proof: By monotone convergence, we may take $T = \mathsf{Z}_+$. If $\|M_t\|_p < \infty$, then $\|M_s\|_p < \infty$ for all $s \le t$ by Jensen's inequality, and so we may assume that $0 < \|M_t^*\|_p < \infty$. Applying Theorem 9.16 to the sub-martingale $|M|$, we get

$$r\,P\{M_t^* > r\} \le E\Big(|M_t|;\, M_t^* > r\Big), \quad r > 0.$$

Using Lemma 4.4, Fubini's theorem, and Hölder's inequality, we obtain

$$\begin{aligned}
\|M_t^*\|_p^p &= p\int_0^\infty P\{M_t^* > r\}\, r^{p-1} dr \\
&\le p\int_0^\infty E\Big(|M_t|;\, M_t^* > r\Big)\, r^{p-2}\, dr \\
&= p\,E\,|M_t| \int_0^{M_t^*} r^{p-2}\, dr \\
&= q\,E\,|M_t|\, M_t^{*(p-1)} \\
&\le q\,\|M_t\|_p \,\Big\|\, M_t^{*(p-1)} \Big\|_q \\
&= q\,\|M_t\|_p\, \|M_t^*\|_p^{p-1}.
\end{aligned}$$

Now divide by the last factor on the right. □

The next inequality is needed to prove the basic convergence theorem. For any function $f : T \to \mathsf{R}$ and constants $a < b$, we define the *number of* $[a, b]$*-crossings* of f up to time t as the supremum of all $n \in \mathsf{Z}_+$, such that there exist times $s_1 < t_1 < s_2 < t_2 < \cdots < s_n < t_n \le t$ in T with $f(s_k) \le a$ and $f(t_k) \ge b$ for all k. This supremum may clearly be infinite.

Lemma 9.18 *(upcrossing inequality, Doob, Snell) Let X be a sub-martingale on a countable index set T, and let $N_a^b(t)$ be the number of $[a, b]$-crossings of X up to time t. Then*

$$E N_a^b(t) \le \frac{E(X_t - a)^+}{b - a}, \quad t \in T,\ a < b \text{ in } \mathsf{R}.$$

Proof: As before we may take $T = \mathsf{Z}_+$. Since $Y = (X - a)^+$ is again a sub-martingale by Lemma 9.11, and the $[a, b]$-crossings of X correspond to $[0, b-a]$-crossings of Y, we may take $X \ge 0$ and $a = 0$. Then define recursively the optional times $0 = \tau_0 \le \sigma_1 < \tau_1 < \sigma_2 < \cdots$ by

$$\begin{aligned}
\sigma_k &= \inf\Big\{n \ge \tau_{k-1};\, X_n = 0\Big\}, \\
\tau_k &= \inf\Big\{n \ge \sigma_k;\, X_n \ge b\Big\}, \qquad k \in \mathsf{N},
\end{aligned}$$

and introduce the predictable process

$$V_n = \sum_{k \ge 1} 1\Big\{\sigma_k < n \le \tau_k\Big\}, \quad n \in \mathsf{N}.$$

Since $(1 - V) \cdot X$ is again a sub-martingale by Corollary 9.15, we get

$$E\Big\{(1 - V) \cdot X\Big\}_t \ge E\Big\{(1 - V) \cdot X\Big\}_0 = 0, \quad t \ge 0.$$

Since also $(V \cdot X)_t \ge b\, N_0^b(t)$, we obtain

$$\begin{aligned} b\, EN_0^b(t) &\leq E(V \cdot X)_t \\ &\leq E(1 \cdot X)_t \\ &= EX_t - EX_0 \\ &\leq EX_t. \end{aligned}$$ □

We may now prove the fundamental convergence theorem for sub-martingales. Say that a process X is *L^p-bounded* if $\sup_t \|X_t\|_p < \infty$.

Theorem 9.19 *(convergence and regularity, Doob) Let X be an L^1-bounded sub-martingale on a countable index set T. Then outside a fixed P-null set, X converges along every increasing or decreasing sequence in T.*

Proof: By Theorem 9.16 we have $X^* < \infty$ a.s., and Lemma 9.18 shows that X has a.s. finitely many upcrossings of every interval $[a, b]$ with rational $a < b$. Then X has clearly the asserted property, outside the P-null set where any of these countably many conditions fails. □

We consider an interesting and useful application.

Proposition 9.20 *(one-sided bound) Let M be a martingale on Z_+ with $\Delta M \leq c$ a.s. for a constant $c < \infty$. Then a.s.*

$$\{M_n \text{ converges}\} = \Big\{\sup_n M_n < \infty\Big\}.$$

Proof: Since $M - M_0$ is again a martingale, we may take $M_0 = 0$. Consider the optional times

$$\tau_m = \inf\Big\{n;\, M_n \geq m\Big\}, \quad m \in \mathsf{N}.$$

The processes M^{τ_m} are again martingales by Corollary 9.15. Since $M^{\tau_m} \leq m+c$ a.s., we have $E|M^{\tau_m}| \leq 2(m+c) < \infty$, and so M^{τ_m} converges a.s. by Theorem 9.19. Hence, M converges a.s. on

$$\begin{aligned} \Big\{\sup_n M_n < \infty\Big\} &= \textstyle\bigcup_m \{\tau_m = \infty\} \\ &\subset \textstyle\bigcup_m \{M = M^{\tau_m}\}. \end{aligned}$$

The reverse implication is obvious, since every convergent sequence in R is bounded. □

The last result yields a useful extension of the Borel–Cantelli lemma from Theorem 4.18:

Corollary 9.21 *(extended Borel–Cantelli lemma, Lévy) For a filtration $\mathcal{F}$ on Z_+, let $A_n \in \mathcal{F}_n$, $n \in \mathsf{N}$. Then a.s.*

$$\{A_n \text{ i.o.}\} = \Big\{\textstyle\sum_n P(A_n|\,\mathcal{F}_{n-1}) = \infty\Big\}.$$

Proof: The sequence

$$M_n = \sum_{k \le n} \left\{ 1_{A_k} - P(A_k | \mathcal{F}_{k-1}) \right\}, \quad n \in \mathsf{Z}_+,$$

is a martingale with $|\Delta M_n| \le 1$, and so by Proposition 9.20 the events $\{\sup_n (\pm M_n) = \infty\}$ agree a.s., which implies

$$P\{M_n \to \infty\} = P\{M_n \to -\infty\} = 0.$$

Hence, we have a.s.

$$\begin{aligned} \{A_n \text{ i.o.}\} &= \left\{ \textstyle\sum_n 1_{A_n} = \infty \right\} \\ &= \left\{ \textstyle\sum_n P(A_n | \mathcal{F}_{n-1}) = \infty \right\}. \end{aligned}$$

□

A martingale M is said to be *closed* if $u = \sup T$ belongs to T, in which case $M_t = E(M_u | \mathcal{F}_t)$ a.s. for all $t \in T$. When $u \notin T$, we say that M is *closable* if it can be extended to a martingale on $\bar{T} = T \cup \{u\}$. If $M_t = E(\xi \,|\, \mathcal{F}_t)$ for a variable $\xi \in L^1$, we may clearly choose $M_u = \xi$. We give some general criteria for closability. An extension to continuous-time sub-martingales appears as part of Theorem 9.30.

Theorem 9.22 *(uniform integrability and closure, Doob) For a martingale M on an unbounded index set T, these conditions are equivalent:*

(i) *M is uniformly integrable,*
(ii) *M is closable at* $\sup T$,
(iii) *M is L^1-convergent at* $\sup T$.

Under those conditions, M may be closed by the limit in (iii).

Proof: Clearly (ii) ⇒ (i) by Lemma 8.4 and (i) ⇒ (iii) by Theorem 9.19 and Proposition 5.12. Now let $M_t \to \xi$ in L^1 as $t \to u \equiv \sup T$. Using the L^1-contractivity of conditional expectations, we get as $t \to u$ for fixed s

$$\begin{aligned} M_s &= E(M_t | \mathcal{F}_s) \\ &\to E(\xi \,|\mathcal{F}_s) \ \text{ in } L^1. \end{aligned}$$

Thus, $M_s = E(\xi \,|\mathcal{F}_s)$ a.s., and we may take $M_u = \xi$. This shows that (iii) ⇒ (ii). □

For comparison, we consider the case of L^p-convergence for $p > 1$.

Corollary 9.23 *(L^p-convergence) Let M be a martingale on an unbounded index set T, and fix any $p > 1$. Then these conditions are equivalent:*

(i) *M converges in L^p,*
(ii) *M is L^p-bounded.*

Proof: We may clearly assume that T is countable. If M is L^p-bounded, it converges in L^1 by Theorem 9.19. Since $|M|^p$ is uniformly integrable by Theorem 9.17, the convergence extends to L^p by Proposition 5.12. Conversely, if M converges in L^p, it is L^p-bounded by Lemma 9.11. □

We turn to the convergence of martingales of the special form $M_t = E(\xi \,|\mathcal{F}_t)$, as t increases or decreases along a sequence. Without loss of generality, we may take the index set T to be unbounded above or below, and define respectively

$$\mathcal{F}_\infty = \bigvee_{t\in T} \mathcal{F}_t, \qquad \mathcal{F}_{-\infty} = \bigcap_{t\in T} \mathcal{F}_t.$$

Theorem 9.24 *(conditioning limits, Jessen, Lévy) Let $\mathcal{F}$ be a filtration on a countable index set $T \subset \mathsf{R}$, unbounded above or below. Then for any $\xi \in L^1$, we have as $t \to \pm\infty$*

$$E(\xi \,|\, \mathcal{F}_t) \to E(\xi \,|\, \mathcal{F}_{\pm\infty}) \quad \text{a.s. and in } L^1.$$

Proof: By Theorems 9.19 and 9.22, the martingale $M_t = E(\xi \,|\, \mathcal{F}_t)$ converges a.s. and in L^1 as $t \to \pm\infty$, where the limit $M_{\pm\infty}$ may clearly be taken to be $\mathcal{F}_{\pm\infty}$-measurable. To obtain $M_{\pm\infty} = E(\xi \,|\, \mathcal{F}_{\pm\infty})$ a.s., we need to verify that

$$E(M_{\pm\infty}; A) = E(\xi; A), \quad A \in \mathcal{F}_{\pm\infty}. \tag{10}$$

Then note that, by the definition of M,

$$E(M_t; A) = E(\xi; A), \quad A \in \mathcal{F}_s, \;\; s \le t. \tag{11}$$

This clearly remains true for $s = -\infty$, and as $t \to -\infty$ we get the 'minus' version of (10). To get the 'plus' version, let $t \to \infty$ in (11) for fixed s, and extend by a monotone-class argument to arbitrary $A \in \mathcal{F}_\infty$. □

In particular, we note the following special case.

Corollary 9.25 *(Lévy) For any filtration $\mathcal{F}$ on Z_+,*

$$P(A \,|\, \mathcal{F}_n) \to 1_A \quad \text{a.s.}, \quad A \in \mathcal{F}_\infty.$$

For a simple application, we prove an extension of Kolmogorov's 0−1 law in Theorem 4.13. Here a relation $\mathcal{F}_1 \subset \mathcal{F}_2$ between two σ-fields is said to hold a.s. if $A \in \mathcal{F}_1$ implies $A \in \bar{\mathcal{F}}_2$, where $\bar{\mathcal{F}}_2$ denotes the completion of $\mathcal{F}_2$.

Corollary 9.26 *(tail σ-field) Let the σ-fields $\mathcal{F}_1, \mathcal{F}_2, \dots$ be conditionally independent, given $\mathcal{G}$. Then*

$$\bigcap_{n\ge 1} \sigma\Big\{\mathcal{F}_n, \mathcal{F}_{n+1}, \dots\Big\} \subset \mathcal{G} \quad \text{a.s.}$$

Proof: Let $\mathcal{T}$ be the σ-field on the left, and note that $\mathcal{T} \perp\!\!\!\perp_{\mathcal{G}} (\mathcal{F}_1 \vee \cdots \vee \mathcal{F}_n)$ by Theorem 8.12. Using Theorem 8.9 and Corollary 9.25, we get for any $A \in \mathcal{T}$

$$P(A \,|\, \mathcal{G}) = P\Big(A \,\Big|\, \mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n\Big) \to 1_A \;\text{ a.s.},$$

which shows that $\mathcal{T} \subset \mathcal{G}$ a.s. □

The last theorem yields a short proof of the strong law of large numbers. Here we put $S_n = \xi_1 + \ldots + \xi_n$ for some i.i.d. random variables $\xi_1, \xi_2, \ldots$ in L^1, and define $\mathcal{F}_{-n} = \sigma\{S_n, S_{n+1}, \ldots\}$. Then $\mathcal{F}_{-\infty}$ is trivial by Theorem 4.15, and $E(\xi_k | \mathcal{F}_{-n}) = E(\xi_1 | \mathcal{F}_{-n})$ a.s. for every $k \le n$, since $(\xi_k, S_n, S_{n+1}, \ldots) \overset{d}{=} (\xi_1, S_n, S_{n+1}, \ldots)$. Hence, Theorem 9.24 yields

$$\begin{aligned} n^{-1} S_n &= E\Big(n^{-1} S_n \,\Big|\, \mathcal{F}_{-n}\Big) \\ &= n^{-1} \sum_{k \le n} E(\xi_k | \mathcal{F}_{-n}) \\ &= E(\xi_1 | \mathcal{F}_{-n}) \\ &\to E(\xi_1 | \mathcal{F}_{-\infty}) = E\xi_1. \end{aligned}$$

For a further application of Theorem 9.24, we prove a kernel version of the regularization Theorem 8.5, needed in Chapter 32.

Theorem 9.27 *(kernel densities) Consider a probability kernel $\mu: S \to T \times U$, where T, U are Borel. Then the densities*

$$\nu(s,t,B) = \frac{\mu(s, dt \times B)}{\mu(s, dt \times U)}, \quad s \in S,\ t \in T,\ B \in \mathcal{U}, \tag{12}$$

have versions combining into a probability kernel $\nu: S \times T \to U$.

Proof: We may take T, U to be Borel subsets of R, in which case μ can be regarded as a probability kernel from S to R^2. Letting $\mathcal{I}_n$ be the σ-field in R generated by the intervals $I_{nk} = 2^{-n}[(k-1), k)$, $k \in \mathsf{Z}$, we define

$$M_n(s,t,B) = \sum_k \frac{\mu(s, I_{nk} \times B)}{\mu(s, I_{nk} \times U)} 1\{t \in I_{nk}\}, \quad s \in S,\ t \in T,\ B \in \mathcal{B},$$

subject to the convention $0/0 = 0$. Then $M_n(s, \cdot, B)$ is a version of the density in (12) on the σ-field $\mathcal{I}_n$, and for fixed s and B it is also a martingale with respect to $\mu(s, \cdot \times U)$. By Theorem 9.24 we get $M_n(s, \cdot, B) \to \nu(s, \cdot, B)$ a.e. $\mu(s, \cdot \times U)$. Thus, ν has the product-measurable version

$$\nu(s,t,B) = \limsup_{n \to \infty} M_n(s,t,B), \quad s \in S,\ \ t \in T,\ \ B \in \mathcal{U}.$$

To construct a kernel version of ν, we may proceed as in the proof of Theorem 3.4, noting that in each step the exceptional (s,t)-set A lies in $\mathcal{S} \otimes \mathcal{T}$, with sections $A_s = \{t \in T;\ (s,t) \in A\}$ satisfying $\mu(s, A_s \times U) = 0$ for all $s \in S$. □

To extend the previous theory to continuous time, we need to form suitably regular versions of the various processes. The following closely related regularizations may then be useful. Say that a process X on R_+ is right-continuous with left-hand limits (abbreviated[5] as *rcll*) if $X_t = X_{t+}$ for all $t \geq 0$, and the left-hand limits X_{t-} exist and are finite for all $t > 0$. For any process Y on Q_+, we define a process Y^+ by $(Y^+)_t = Y_{t+}$, $t \geq 0$, whenever the right-hand limits exist.

Theorem 9.28 *(regularization, Doob) For any $\mathcal{F}$-sub-martingale X on R_+ with restriction Y to Q_+, we have*

(i) *Y^+ exists and is rcll outside a fixed P-null set A, and $Z = 1_{A^c} Y^+$ is a sub-martingale with respect to the augmented filtration $\overline{\mathcal{F}^+}$,*

(ii) *when $\mathcal{F}$ is right-continuous, X has an rcll version iff EX is right-continuous, hence in particular when X is a martingale.*

The proof requires an extension of Theorem 9.22 to suitable sub-martingales.

Lemma 9.29 *(uniform integrability) Let X be a sub-martingale on Z_-. Then these conditions are equivalent:*

(i) *X is uniformly integrable,*

(ii) *EX is bounded.*

Proof: Let EX be bounded. Form the predictable sequence

$$\alpha_n = E\Big(\Delta X_n \,\Big|\, \mathcal{F}_{n-1}\Big) \geq 0, \quad n \leq 0,$$

and note that

$$E\sum_{n\leq 0} \alpha_n = EX_0 - \inf_{n\leq 0} EX_n < \infty.$$

Hence, $\sum_n \alpha_n < \infty$ a.s., and so we may define

$$A_n = \sum_{k\leq n} \alpha_k, \quad M_n = X_n - A_n, \qquad n \leq 0.$$

Since $EA^* < \infty$, and M is a martingale closed at 0, both A and M are uniformly integrable. $\Box$

Proof of Theorem 9.28: (i) By Lemma 9.11 the process $Y \vee 0$ is L^1-bounded on bounded intervals, and so the same thing is true for Y. Thus, by Theorem 9.19 the right- and left-hand limits $Y_{t\pm}$ exist outside a fixed P-null set A, and so $Z = 1_{A^c} Y^+$ is rcll. Also note that Z is adapted to $\overline{\mathcal{F}^+}$.

To prove that Z is a sub-martingale for $\overline{\mathcal{F}^+}$, fix any times $s < t$, and choose $s_n \downarrow s$ and $t_n \downarrow t$ in Q_+ with $s_n < t$. Then $Y_{s_m} \leq E(Y_{t_n}|\mathcal{F}_{s_m})$ a.s. for all m and n, and Theorem 9.24 yields $Z_s \leq E(Y_{t_n}|\mathcal{F}_{s+})$ a.s. as $m \to \infty$. Since $Y_{t_n} \to Z_t$ in L^1 by Lemma 9.29, it follows that

[5]The French acronym *càdlàg* is often used, even in English texts.

$$\begin{aligned} Z_s &\le E(Z_t|\,\mathcal{F}_{s+}) \\ &= E(Z_t|\,\bar{\mathcal{F}}_{s+}) \ \text{ a.s.} \end{aligned}$$

(ii) For any $t < t_n \in \mathsf{Q}_+$,

$$\begin{aligned} (EX)_{t_n} &= E(Y_{t_n}), \\ X_t &\le E(Y_{t_n}|\,\mathcal{F}_t) \ \text{ a.s.,} \end{aligned}$$

and as $t_n \downarrow t$ we get by Lemma 9.29 and the right-continuity of $\mathcal{F}$

$$\begin{aligned} (EX)_{t+} &= EZ_t, \\ X_t &\le E(Z_t|\,\mathcal{F}_t) = Z_t \ \text{ a.s.} \end{aligned} \tag{13}$$

If X has a right-continuous version, then clearly $Z_t = X_t$ a.s. Hence, (13) yields $(EX)_{t+} = EX_t$, which shows that EX is right-continuous. If instead EX is right-continuous, then (13) gives

$$E|Z_t - X_t| = EZ_t - EX_t = 0,$$

and so $Z_t = X_t$ a.s., which means that Z is a version of X. □

Justified by the last theorem, we may henceforth take all sub-martingales to be rcll, unless otherwise specified, and also let the underlying filtration be right-continuous and complete. Most of the previously quoted results for sub-martingales on a countable index set extend immediately to continuous time. In particular, this is true for the convergence Theorem 9.19 and the inequalities in Theorem 9.16 and Lemma 9.18. We proceed to show how Theorems 9.12 and 9.22 extend to sub-martingales in continuous time.

Theorem 9.30 *(optional sampling and closure, Doob) Let X be an $\mathcal{F}$-sub-martingale on R_+, where X and $\mathcal{F}$ are right-continuous, and consider some optional times σ, τ, where τ is bounded. Then X_τ is integrable, and*

$$X_{\sigma\wedge\tau} \le E(X_\tau|\,\mathcal{F}_\sigma) \ \text{ a.s.} \tag{14}$$

This extends to unbounded times τ iff X^+ is uniformly integrable.

Proof: Introduce the optional times $\sigma_n = 2^{-n}[2^n\sigma+1]$ and $\tau_n = 2^{-n}[2^n\tau+1]$, and conclude from Lemma 9.10 and Theorem 9.12 that

$$X_{\sigma_m\wedge\tau_n} \le E\Big(X_{\tau_n}\,\Big|\,\mathcal{F}_{\sigma_m}\Big) \ \text{ a.s.,} \quad m,n \in \mathsf{N}.$$

As $m \to \infty$, we get by Lemma 9.3 and Theorem 9.24

$$X_{\sigma\wedge\tau_n} \le E(X_{\tau_n}|\,\mathcal{F}_\sigma) \ \text{ a.s.,} \quad n \in \mathsf{N}. \tag{15}$$

By the result for the index sets $2^{-n}\,\mathsf{Z}_+$, the random variables $X_0; \dots, X_{\tau_2}, X_{\tau_1}$ form a sub-martingale with bounded mean, and hence are uniformly integrable by Lemma 9.29. Thus, (14) follows as we let $n \to \infty$ in (15).

If X^+ is uniformly integrable, then X is L^1-bounded and hence converges a.s. toward some $X_\infty \in L^1$. By Proposition 5.12 we get $X_t^+ \to X_\infty^+$ in L^1, and so $E(X_t^+ | \mathcal{F}_s) \to E(X_\infty^+ | \mathcal{F}_s)$ in L^1 for each s. Letting $t \to \infty$ along a sequence, we get by Fatou's lemma

$$\begin{aligned} X_s &\le \lim_{t\to\infty} E(X_t^+ | \mathcal{F}_s) - \liminf_{t\to\infty} E(X_t^- | \mathcal{F}_s) \\ &\le E(X_\infty^+ | \mathcal{F}_s) - E(X_\infty^- | \mathcal{F}_s) \\ &= E(X_\infty | \mathcal{F}_s). \end{aligned}$$

Approximating as before, we obtain (14) for arbitrary σ, τ.

Conversely, the stated condition yields the existence of an $X_\infty \in L^1$ with $X_s \le E(X_\infty | \mathcal{F}_s)$ a.s. for all $s > 0$, and so $X_s^+ \le E(X_\infty^+ | \mathcal{F}_s)$ a.s. by Lemma 9.11. Hence, X^+ is uniformly integrable by Lemma 8.4. □

For a simple application, we consider the hitting probabilities of a continuous martingale. The result will be useful in Chapters 18, 22, and 33.

Corollary 9.31 *(first hit) Let M be a continuous martingale with $M_0 = 0$ and $P\{M^* > 0\} > 0$, and define $\tau_x = \inf\{t > 0;\, M_t = x\}$. Then for $a < 0 < b$,*

$$P\Big\{\tau_a < \tau_b \,\Big|\, M^* > 0\Big\} \le \frac{b}{b-a} \le P\Big\{\tau_a \le \tau_b \,\Big|\, M^* > 0\Big\}.$$

Proof: Since $\tau = \tau_a \wedge \tau_b$ is optional by Lemma 9.6, Theorem 9.30 yields $EM_{\tau\wedge t} = 0$ for all $t > 0$, and so by dominated convergence $EM_\tau = 0$. Hence,

$$\begin{aligned} 0 &= a\,P\{\tau_a < \tau_b\} + b\,P\{\tau_b < \tau_a\} + E\Big(M_\infty;\, \tau = \infty\Big) \\ &\le a\,P\{\tau_a < \tau_b\} + b\,P\Big\{\tau_b \le \tau_a,\, M^* > 0\Big\} \\ &= b\,P\{M^* > 0\} - (b-a)\,P\{\tau_a < \tau_b\}, \end{aligned}$$

which implies the first inequality. The second one follows as we take complements. □

The next result plays a crucial role in Chapter 17.

Lemma 9.32 *(absorption) For any right-continuous super-martingale $X \ge 0$, we have $X = 0$ a.s. on $[\tau, \infty)$, where*

$$\tau = \inf\Big\{t \ge 0;\, X_t \wedge X_{t-} = 0\Big\}.$$

Proof: By Theorem 9.28 the process X remains a super-martingale with respect to the right-continuous filtration $\mathcal{F}^+$. The times $\tau_n = \inf\{t \ge 0;\, X_t < n^{-1}\}$ are $\mathcal{F}^+$-optional by Lemma 9.6, and the right-continuity of X yields $X_{\tau_n} \le n^{-1}$ on $\{\tau_n < \infty\}$. Hence, Theorem 9.30 yields

$$E\Big(X_t;\, \tau_n \le t\Big) \le E\Big(X_{\tau_n};\, \tau_n \le t\Big) \le n^{-1}, \quad t \ge 0,\ n \in \mathbb{N}.$$

Noting that $\tau_n \uparrow \tau$, we get by dominated convergence $E(X_t;\, \tau \le t) = 0$, and so $X_t = 0$ a.s. on $\{\tau \le t\}$. The assertion now follows, as we apply this result to all $t \in \mathsf{Q}_+$ and use the right-continuity of X. $\square$

Finally we show how the right-continuity of an increasing sequence of super-martingales extends to the limit. The result is needed in Chapter 34.

Theorem 9.33 *(increasing super-martingales, Meyer) Let $X^1 \le X^2 \le \cdots$ be right-continuous super-martingales with $\sup_n EX_0^n < \infty$. Then $X_t = \sup_n X_t^n$ is again an a.s. right-continuous super-martingale.*

Proof (Doob): By Theorem 9.28 we may take the filtration to be right-continuous. The super-martingale property carries over to X by monotone convergence. To prove the asserted right-continuity, we may take X^1 to be bounded below by an integrable random variable, since we can otherwise consider the processes obtained by optional stopping at the times $m \wedge \inf\{t;\, X_t^1 < -m\}$, for arbitrary $m > 0$.

For fixed $\varepsilon > 0$, let $\mathcal{T}$ be the class of optional times τ with

$$\limsup_{u \downarrow t} |X_u - X_t| \le 2\varepsilon, \quad t < \tau,$$

and put $p = \inf_{\tau \in \mathcal{T}} Ee^{-\tau}$. Choose $\sigma_1, \sigma_2, \ldots \in \mathcal{T}$ with $Ee^{-\sigma_n} \to p$, and note that $\sigma \equiv \sup_n \sigma_n \in \mathcal{T}$ with $Ee^{-\sigma} = p$. We need to show that $\sigma = \infty$ a.s. Then introduce the optional times

$$\tau_n = \inf\Big\{t > \sigma;\, |X_t^n - X_\sigma| > \varepsilon\Big\}, \quad n \in \mathsf{N},$$

and put $\tau = \limsup_n \tau_n$. Noting that

$$|X_t - X_\sigma| = \liminf_{n \to \infty} |X_t^n - X_\sigma| \le \varepsilon, \quad t \in [\sigma, \tau),$$

we obtain $\tau \in \mathcal{T}$.

By the right-continuity of X^n, we have $|X_{\tau_n}^n - X_\sigma| \ge \varepsilon$ on $\{\tau_n < \infty\}$ for every n. Furthermore, on the set $A = \{\sigma = \tau < \infty\}$,

$$\begin{aligned} \liminf_{n \to \infty} X_{\tau_n}^n &\ge \sup\nolimits_k \lim_{n \to \infty} X_{\tau_n}^k \\ &= \sup\nolimits_k X_\sigma^k = X_\sigma\,, \end{aligned}$$

and so

$$\liminf_{n \to \infty} X_{\tau_n}^n \ge X_\sigma + \varepsilon \text{ on } A.$$

Since $A \in \mathcal{F}_\sigma$ by Lemma 9.1, we get by Fatou's lemma, optional sampling, and monotone convergence

$$\begin{aligned} E(X_\sigma + \varepsilon; A) &\le E\Big(\liminf_{n \to \infty} X_{\tau_n}^n; A\Big) \\ &\le \liminf_{n \to \infty} E(X_{\tau_n}^n; A) \\ &\le \lim_{n \to \infty} E(X_\sigma^n; A) \\ &= E(X_\sigma; A). \end{aligned}$$

Thus, $PA = 0$, and so $\tau > \sigma$ a.s. on $\{\sigma < \infty\}$. Since $p > 0$ would yield the contradiction $Ee^{-\tau} < p$, we have $p = 0$, and so $\sigma = \infty$ a.s. $\square$

Exercises

1. For any optional times σ, τ, show that $\{\sigma = \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$ and $\mathcal{F}_\sigma = \mathcal{F}_\tau$ on $\{\sigma = \tau\}$. However, $\mathcal{F}_\tau$ and $\mathcal{F}_\infty$ may differ on $\{\tau = \infty\}$.

2. Show that if σ, τ are optional times on the time scale R_+ or Z_+, then so is $\sigma + \tau$.

3. Give an example of a random time that is weakly optional but not optional. (*Hint:* Let $\mathcal{F}$ be the filtration induced by the process $X_t = \vartheta\, t$ with $P\{\vartheta = \pm 1\} = \frac{1}{2}$, and take $\tau = \inf\{t;\, X_t > 0\}$.)

4. Fix a random time τ and a random variable ξ in $\mathsf{R} \setminus \{0\}$. Show that the process $X_t = \xi\, 1\{\tau \le t\}$ is adapted to a given filtration $\mathcal{F}$ iff τ is $\mathcal{F}$-optional and ξ is $\mathcal{F}_\tau$-measurable. Give corresponding conditions for the process $Y_t = \xi\, 1\{\tau < t\}$.

5. Let $\mathcal{P}$ be the class of sets $A \in \mathsf{R}_+ \times \Omega$ such that the process 1_A is progressive. Show that $\mathcal{P}$ is a σ-field and that a process X is progressive iff it is $\mathcal{P}$-measurable.

6. Let X be a progressive process with induced filtration $\mathcal{F}$, and fix any optional time $\tau < \infty$. Show that $\sigma\{\tau, X^\tau\} \subset \mathcal{F}_\tau \subset \mathcal{F}_\tau^+ \subset \sigma\{\tau, X^{\tau+h}\}$ for every $h > 0$. (*Hint:* The first relation becomes an equality when τ takes only countably many values.) Note that the result may fail when $P\{\tau = \infty\} > 0$.

7. Let M be an $\mathcal{F}$-martingale on a countable index set, and fix an optional time τ. Show that $M - M^\tau$ remains a martingale, conditionally on $\mathcal{F}_\tau$. (*Hint:* Use Theorem 9.12 and Lemma 9.14.) Extend the result to continuous time.

8. Show that any sub-martingale remains a sub-martingale with respect to the induced filtration.

9. Let $X^1, X^2, \dots$ be sub-martingales such that the process $X = \sup_n X^n$ is integrable. Show that X is again a sub-martingale. Also show that $\limsup_n X^n$ is a sub-martingale when even $\sup_n |X^n|$ is integrable.

10. Show that the Doob decomposition of an integrable random sequence $X = (X_n)$ depends on the filtration, unless X is a.s. X_0-measurable. (*Hint:* Compare the filtrations induced by X and by the sequence $Y_n = (X_0, X_{n+1})$.)

11. Fix a random time τ and a random variable $\xi \in L^1$, and define $M_t = \xi\, 1\{\tau \le t\}$. Show that M is a martingale with respect to the induced filtration $\mathcal{F}$ iff $E(\xi;\, \tau \le t \,|\, \tau > s) = 0$ for any $s < t$. (*Hint:* The set $\{\tau > s\}$ is an atom of $\mathcal{F}_s$.)

12. Let $\mathcal{F}$, $\mathcal{G}$ be filtrations on the same probability space. Show that every $\mathcal{F}$-martingale is a $\mathcal{G}$-martingale iff $\mathcal{F}_t \subset \mathcal{G}_t \perp\!\!\!\perp_{\mathcal{F}_t} \mathcal{F}_\infty$ for every $t \ge 0$. (*Hint:* For the necessity, consider $\mathcal{F}$-martingales of the form $M_s = E(\xi \,|\, \mathcal{F}_s)$ with $\xi \in L^1(\mathcal{F}_t)$.)

13. For any rcll super-martingale $X \ge 0$ and constant $r \ge 0$, show that $r\, P\{\sup_t X_t \ge r\} \le E X_0$.

14. Let M be an L^2-bounded martingale on Z_+. Mimicing the proof of Lemma 5.16, show that M_n converges a.s. and in L^2.

15. Give an example of a martingale that is L^1-bounded but not uniformly integrable. (*Hint:* Every positive martingale is L^1-bounded.)

16. For any optional times σ, τ with respect to a right-continuous filtration $\mathcal{F}$, show that $E^{\mathcal{F}_\sigma}$ and $E^{\mathcal{F}_\tau}$ commute on L^1 with product $E^{\mathcal{F}_{\sigma \wedge \tau}}$, and that $\mathcal{F}_\sigma \perp\!\!\!\perp_{\mathcal{F}_{\sigma\wedge\tau}} \mathcal{F}_\tau$. (*Hint:* The first condition holds by Theorem 9.30, and the second one since $\mathcal{F}_{\sigma\wedge\tau} =$

$\mathcal{F}_\sigma$ on $\{\sigma \le \tau\}$ and $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\tau$ on $\{\tau \le \sigma\}$ by Lemma 9.1.. The two conditions are also seen to be equivalent by Corollary 8.14.)

17. Let $\xi_n \to \xi$ in L^1. For any increasing σ-fields $\mathcal{F}_n$, show that $E(\xi_n|\mathcal{F}_n) \to E(\xi\,|\mathcal{F}_\infty)$ in L^1.

18. Let $\xi,\, \xi_1, \xi_2, \ldots \in L^1$ with $\xi_n \uparrow \xi$ a.s. For any increasing σ-fields $\mathcal{F}_n$, show that $E(\xi_n|\,\mathcal{F}_n) \to E(\xi\,|\,\mathcal{F}_\infty)$ a.s. (*Hint:* Note that $\sup_m E(\xi - \xi_n|\mathcal{F}_m) \xrightarrow{P} 0$ by Proposition 9.16, and use the monotonicity.)

19. Show that any right-continuous sub-martingale is a.s. rcll.

20. Given a super-martingale $X \ge 0$ on Z_+ and some optional times $\tau_0 \le \tau_1 \le \cdots$, show that the sequence (X_{τ_n}) is again a super-martingale. (*Hint:* Truncate the times τ_n, and use the conditional Fatou lemma.) Show by an example that the result fails for sub-martingales.

21. For any random time $\tau \ge 0$ and right-continuous filtration $\mathcal{F} = (\mathcal{F}_t)$, show that the process $X_t = P(\tau \le t\,|\,\mathcal{F}_t)$ has a right-continuous version. (*Hint:* Use Theorem 9.28 (ii).)

Chapter 10

Predictability and Compensation

Predictable times and processes, strict past, accessible and totally inaccessible times, decomposition of optional times, Doob–Meyer decomposition, natural and predictable processes, dual predictable projection, predictable martingales, compensation and decomposition of random measures, ql-continuous filtrations, induced and discounted compensators, fundamental martingale, product moments, predictable maps

The theory of martingales and optional times from the previous chapter leads naturally into a *general theory of processes*, of fundamental importance for various advanced topics of stochastic calculus, random measure theory, and many other subjects throughout the remainder of this book. Here a crucial role is played by the notions of *predictable processes* and associated *predictable times*, forming basic subclasses of adapted processes and optional times.

Among the many fundamental results proved in this chapter, we note in particular the celebrated *Doob–Meyer decomposition*, a continuous-time counterpart of the elementary Doob decomposition from Lemma 9.10, here proved directly[1] by a discrete approximation based on Dunford's characterization of weak compactness, combined with Doob's ingenious approximation of inaccessible times. The result may be regarded as a probabilistic counterpart of some basic decompositions in potential theory, as explained in Chapter 34.

Even more important for probabilists are the applications to increasing processes, leading to the notion of *compensator* of a random measure or point process on a product space $\mathsf{R}_+ \times S$, defined as a predictable random measure on the same space, providing the instantaneous rate of increase of the underlying process. In particular, we will see in Chapter 15 how the compensator can be used to reduce a fairly general point process on R_+ to Poisson, in a similar way as a continuous martingale can be time-changed into a Brownian motion. A suitable transformation leads to the equally important notion of *discounted compensator*, needed in Chapters 15 and 27 to establish some general mapping theorems.

The present material is related in many ways to material discussed elsewhere. Apart from the already mentioned connections, we note in particular the notions of tangential processes, general semi-martingales, and stochastic integration in Chapters 16, 20, and 35, the quasi-left continuity of Feller processes in Chapter 17, and the predictable mapping theorems in Chapters 15, 19, and 27.

[1] The standard proof, often omitted, is based on a subtle use of capacity theory.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_11

We take all random objects in this chapter to be defined on a fixed probability space $(\Omega, \mathcal{A}, P)$ with a right-continuous and complete filtration $\mathcal{F}$. A random time τ in $[0, \infty]$ is said to be *predictable*, if it is *announced* by some optional times $\tau_n \uparrow \tau$ with $\tau_n < \tau$ a.s. on $\{\tau > 0\}$ for all n. Note that predictable times are optional by Lemma 9.3 (i). With any optional time τ we may associate the σ-field $\mathcal{F}_{\tau-}$ generated by $\mathcal{F}_0$ and the classes $\mathcal{F}_t \cap \{\tau > t\}$ for all $t > 0$, representing the *strict past* of τ. The following properties of predictable times and associated σ-fields may be compared with similar results for optional times and σ-fields $\mathcal{F}_\tau$ in Lemmas 9.1 and 9.3.

Lemma 10.1 *(predictable times and strict past) For random times σ, τ, τ_1, τ_2, … in $[0, \infty]$, we have*

(i) $\mathcal{F}_\sigma \cap \{\sigma < \tau\} \subset \mathcal{F}_{\tau-} \subset \mathcal{F}_\tau$ *when σ, τ are optional,*

(ii) $\{\sigma < \tau\} \in \mathcal{F}_{\sigma-} \cap \mathcal{F}_{\tau-}$ *for optional σ and predictable τ,*

(iii) $\bigvee_n \mathcal{F}_{\tau_n} = \mathcal{F}_{\tau-}$ *when τ is predictable and announced by (τ_n),*

(iv) *if $\tau_1, \tau_2, \ldots$ are predictable, so is $\tau = \sup_n \tau_n$ with $\bigvee_n \mathcal{F}_{\tau_n -} = \mathcal{F}_{\tau-}$,*

(v) *if σ, τ are predictable, so is $\sigma \wedge \tau$ with $\mathcal{F}_{\sigma\wedge\tau-} \subset \mathcal{F}_{\sigma-} \cap \mathcal{F}_{\tau-}$.*

Proof: (i) For any $A \in \mathcal{F}_\sigma$, we have

$$A \cap \{\sigma < \tau\} = \bigcup_{r \in \mathbb{Q}_+} \Big(A \cap \{\sigma \le r\} \cap \{r < \tau\} \Big) \in \mathcal{F}_{\tau-},$$

since the intersections on the right are generators of $\mathcal{F}_{\tau-}$, and so $\mathcal{F}_\sigma \cap \{\sigma < \tau\} \in \mathcal{F}_{\tau-}$. The second relation holds since all generators of $\mathcal{F}_{\tau-}$ lie in $\mathcal{F}_\tau$.

(ii) If τ is announced by (τ_n), then (i) yields

$$\{\tau \le \sigma\} = \{\tau = 0\} \cup \textstyle\bigcap_n \{\tau_n < \sigma\} \in \mathcal{F}_{\sigma-}.$$

(iii) For any $A \in \mathcal{F}_{\tau_n}$, we get by (i)

$$A = \Big(A \cap \{\tau_n < \tau\} \Big) \cup \Big(A \cap \{\tau_n = \tau = 0\} \Big) \in \mathcal{F}_{\tau-},$$

and so $\bigvee_n \mathcal{F}_{\tau_n} \subset \mathcal{F}_{\tau-}$. Conversely, (i) yields for any $t \ge 0$ and $A \in \mathcal{F}_t$

$$\begin{aligned} A \cap \{\tau > t\} &= \textstyle\bigcup_n \Big(A \cap \{\tau_n > t\} \Big) \\ &\in \textstyle\bigvee_n \mathcal{F}_{\tau_n -} \\ &\subset \textstyle\bigvee_n \mathcal{F}_{\tau_n}, \end{aligned}$$

which shows that $\mathcal{F}_{\tau-} \subset \bigvee_n \mathcal{F}_{\tau_n}$.

(iv) If each τ_n is announced by $\sigma_{n1}, \sigma_{n2}, \ldots$, then τ is announced by the optional times $\sigma_n = \sigma_{1n} \vee \cdots \vee \sigma_{nn}$ and is therefore predictable. By (i) we get for all n

$$\begin{aligned} \mathcal{F}_{\sigma_n} &= \mathcal{F}_0 \cup \bigcup_{k \le n} \Big(\mathcal{F}_{\sigma_n} \cap \{\sigma_n < \tau_k\} \Big) \\ &\subset \bigvee_{k \le n} \mathcal{F}_{\tau_k -}. \end{aligned}$$

Combining with (i) and (iii) gives

$$\begin{aligned}\mathcal{F}_{\tau-} &= \bigvee_n \mathcal{F}_{\sigma_n}\\ &\subset \bigvee_n \mathcal{F}_{\tau_n-}\\ &= \bigvee_{n,k} \mathcal{F}_{\sigma_{nk}} \subset \mathcal{F}_{\tau-},\end{aligned}$$

and so equality holds throughout, proving our claim.

(v) If σ and τ are announced by $\sigma_1, \sigma_2, \ldots$ and $\tau_1, \tau_2, \ldots$, then $\sigma \wedge \tau$ is announced by the optional times $\sigma_n \wedge \tau_n$, and is therefore predictable. Combining (iii) above with Lemma 9.1 (ii), we get

$$\begin{aligned}\mathcal{F}_{\sigma\wedge\tau-} &= \bigvee_n \mathcal{F}_{\sigma_n\wedge\tau_n}\\ &= \bigvee_n (\mathcal{F}_{\sigma_n} \cap \mathcal{F}_{\tau_n})\\ &\subset \bigvee_n \mathcal{F}_{\sigma_n} \cap \bigvee_n \mathcal{F}_{\tau_n}\\ &= \mathcal{F}_{\sigma-} \cap \mathcal{F}_{\tau-}.\end{aligned}$$

□

On the product space $\Omega \times \mathsf{R}_+$ we may introduce the *predictable σ-field* $\mathcal{P}$, generated by all continuous, adapted processes on R_+. The elements of $\mathcal{P}$ are called *predictable sets*, and the $\mathcal{P}$-measurable functions on $\Omega \times \mathsf{R}_+$ are called *predictable processes*. Note that every predictable process is progressive. We provide some useful characterizations of the predictable σ-field.

Lemma 10.2 *(predictable σ-field) The predictable σ-field is generated by each of these classes of sets or processes:*

(i) $\mathcal{F}_0 \times \mathsf{R}_+$ *and all sets* $A \times (t, \infty)$ *with* $A \in \mathcal{F}_t$, $t \geq 0$,

(ii) $\mathcal{F}_0 \times \mathsf{R}_+$ *and all intervals* (τ, ∞) *with optional* τ,

(iii) *all left-continuous, adapted processes.*

Proof: Let $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$ be the σ-fields generated by the classes in (i)$-$(iii). Since continuous functions are left-continuous, we have trivially $\mathcal{P} \subset \mathcal{P}_3$. To obtain $\mathcal{P}_3 \subset \mathcal{P}_1$, we note that any left-continuous process X can be approximated by the processes

$$X_t^n = X_0 1_{[0,1]}(nt) + \sum_{k\geq 1} X_{k/n} 1_{(k,k+1]}(nt), \quad t \geq 0.$$

Next, $\mathcal{P}_1 \subset \mathcal{P}_2$ holds since the time $t_A = t \cdot 1_A + \infty \cdot 1_{A^c}$ is optional for any $t \geq 0$ and $A \in \mathcal{F}_t$. Finally, we get $\mathcal{P}_2 \subset \mathcal{P}$ since, for any optional time τ, the process $1_{(\tau,\infty)}$ can be approximated by the continuous, adapted processes $X_t^n = \{n(t-\tau)_+\} \wedge 1$, $t \geq 0$. □

Next we examine the relationship between predictable processes and the σ-fields $\mathcal{F}_{\tau-}$. Similar results for progressive processes and the σ-fields $\mathcal{F}_\tau$ were obtained in Lemma 9.5.

Lemma 10.3 *(predictable times and processes)*

(i) *For any optional time τ and predictable process X, the random variable $X_\tau 1\{\tau < \infty\}$ is $\mathcal{F}_{\tau-}$-measurable.*

(ii) *For any predictable time τ and $\mathcal{F}_{\tau-}$-measurable random variable α, the process $X_t = \alpha\, 1\{\tau \le t\}$ is predictable.*

Proof: (i) If $X = 1_{A\times(t,\infty)}$ for some $t > 0$ and $A \in \mathcal{F}_t$, then clearly

$$\Big\{X_\tau 1\{\tau < \infty\} = 1\Big\} = A \cap \{t < \tau < \infty\} \in \mathcal{F}_{\tau-}.$$

This extends by a monotone-class argument and a subsequent approximation, first to predictable indicator functions, and then to the general case.

(ii) We may clearly take α to be integrable. For any announcing sequence (τ_n) of τ, we define

$$X_t^n = E(\alpha \,|\, \mathcal{F}_{\tau_n}) \Big(1\{\tau_n \in (0,t)\} + 1\{\tau_n = 0\}\Big), \quad t \ge 0.$$

The X^n are left-continuous and adapted, hence predictable. Moreover, $X^n \to X$ on R_+ a.s. by Theorem 9.24 and Lemma 10.1 (iii). □

An optional time τ is said to be *totally inaccessible*, if $P\{\sigma = \tau < \infty\} = 0$ for every predictable time σ. An *accessible time* may then be defined as an optional time τ, such that $P\{\sigma = \tau < \infty\} = 0$ for every totally inaccessible time σ. For any random time τ, we introduce the associated *graph*

$$[\tau] = \Big\{(t,\omega) \in \mathsf{R}_+ \times \Omega;\ \tau(\omega) = t\Big\},$$

which allows us to express the previous conditions on σ and τ as $[\sigma] \cap [\tau] = \emptyset$ a.s. For any optional time τ and set $A \in \mathcal{F}_\tau$, the time $\tau_A = \tau 1_A + \infty \cdot 1_{A^c}$ is again optional and called the *restriction* of τ to A. We consider a basic decomposition of optional times. Related decompositions of random measures and martingales are given in Theorems 10.17 and 20.16.

Proposition 10.4 *(decomposition of optional time) For any optional time τ,*

(i) *there exists an a.s. unique set $A \in \mathcal{F}_\tau \cap \{\tau < \infty\}$, such that τ_A is accessible and τ_{A^c} is totally inaccessible,*

(ii) *there exist some predictable times $\tau_1, \tau_2, \ldots$ with $[\tau_A] \subset \bigcup_n [\tau_n]$ a.s.*

Proof: Define

$$p = \sup_{(\tau_n)} P \textstyle\bigcup_n \{\tau = \tau_n < \infty\}, \tag{1}$$

where the supremum extends over all sequences of predictable times τ_n. Combining sequences where the probability in (1) approaches p, we may construct a sequence (τ_n) for which the supremum is attained. For such a maximal sequence, we define A as the union in (1).

To see that τ_A is accessible, let σ be totally inaccessible. Then $[\sigma] \cap [\tau_n] = \emptyset$ a.s. for every n, and so $[\sigma] \cap [\tau_A] = \emptyset$ a.s. If τ_{A^c} is not totally inaccessible, then $P\{\tau_{A^c} = \tau_0 < \infty\} > 0$ for some predictable time τ_0, which contradicts the maximality of $\tau_1, \tau_2, \ldots$. This shows that A has the desired property.

To prove that A is a.s. unique, let B be another set with the stated properties. Then $\tau_{A\setminus B}$ and $\tau_{B\setminus A}$ are both accessible and totally inaccessible, and so $\tau_{A\setminus B} = \tau_{B\setminus A} = \infty$ a.s., which implies $A = B$ a.s. □

We turn to the celebrated *Doob–Meyer decomposition*, a cornerstone of modern probability. By an *increasing process* we mean a non-decreasing, right-continuous, and adapted process A with $A_0 = 0$. It is said to be *integrable* if $EA_\infty < \infty$. Recall that all sub-martingales are taken to be right-continuous. Local sub-martingales and locally integrable processes are defined by localization, in the usual way.

Theorem 10.5 *(Doob–Meyer decomposition, Meyer, Doléans) For an adapted process X, these conditions are equivalent:*

(i) *X is a local sub-martingale,*

(ii) *$X = M + A$ a.s. for a local martingale M and a locally integrable, non-decreasing, predictable process A with $A_0 = 0$.*

The processes M and A are then a.s. unique.

Note that the decomposition depends in a crucial way on the underlying filtration $\mathcal{F}$. We often refer to A as the *compensator* of X, and write $A = \hat{X}$. Several proofs of this result are known, most of which seem to rely on the deep section theorems. Here we give a more elementary proof, based on weak compactness in L^1 and an approximation of totally inaccessible times. For clarity, we divide the proof into several lemmas.

Let (D) be the class of measurable processes X, such that the family $\{X_\tau\}$ is uniformly integrable, where τ ranges over the set of finite optional times. We show that it is enough to consider sub-martingales of class (D).

Lemma 10.6 *(uniform integrability) A local sub-martingale X with $X_0 = 0$ is locally of class (D).*

Proof: First reduce to the case where X is a true sub-martingale. Then introduce for each n the optional time $\tau = n \wedge \inf\{t > 0; |X_t| > n\}$. Here $|X^\tau| \le n \vee |X_\tau|$, which is integrable by Theorem 9.30, and so X^τ is of class (D). □

An increasing process A is said to be *natural*, if it is integrable and such that $E\int_0^\infty (\Delta M_t)\, dA_t = 0$ for any bounded martingale M. We first establish a preliminary decomposition, where the compensator A is claimed to be natural rather than predictable.

Lemma 10.7 *(preliminary decomposition, Meyer) Any sub-martingale X of class (D) admits a decomposition $M + A$, where*

(i) *M is a uniformly integrable martingale,*

(ii) *A is a natural, increasing process.*

Proof (Rao): We may take $X_0 = 0$. Introduce the *n-dyadic* times $t^n_k = k2^{-n}$, $k \in \mathsf{Z}_+$, and define for any process Y the associated differences $\Delta^n_k Y = Y_{t^n_{k+1}} - Y_{t^n_k}$. Let

$$A^n_t = \sum_{k<2^n t} E\Big(\Delta^n_k X \Big| \mathcal{F}_{t^n_k}\Big), \quad t \geq 0, \ n \in \mathsf{N},$$

and note that $M^n = X - A^n$ is a martingale on the n-dyadic set.

Writing $\tau^n_r = \inf\{t;\, A^n_t > r\}$ for $n \in \mathsf{N}$ and $r > 0$, we get by optional sampling, for any n-dyadic time t,

$$\begin{aligned}
\tfrac{1}{2}\, E\Big(A^n_t;\, A^n_t > 2r\Big) &\leq E(A^n_t - A^n_t \wedge r) \\
&\leq E\Big(A^n_t - A^n_{\tau^n_r \wedge t}\Big) \\
&= E\Big(X_t - X_{\tau^n_r \wedge t}\Big) \\
&= E\Big(X_t - X_{\tau^n_r \wedge t};\, A^n_t > r\Big). \qquad (2)
\end{aligned}$$

By the martingale property and uniform integrability, we further obtain

$$rP\{A^n_t > r\} \leq EA^n_t = EX_t \lesssim 1,$$

and so the probability on the left tends to zero as $r \to \infty$, uniformly in t and n. Since the variables $X_t - X_{\tau^n_r \wedge t}$ are uniformly integrable by (D), the same property holds for A^n_t by (2) and Lemma 5.10. In particular, the sequence (A^n_∞) is uniformly integrable, and each M^n is a uniformly integrable martingale.

By Lemma 5.13 there exists a random variable $\alpha \in L^1(\mathcal{F}_\infty)$, such that $A^n_\infty \to \alpha$ weakly in L^1 along a sub-sequence $N' \subset \mathsf{N}$. Define

$$M_t = E\Big(X_\infty - \alpha \Big| \mathcal{F}_t\Big), \qquad A = X - M,$$

and note that $A_\infty = \alpha$ a.s. by Theorem 9.24. For dyadic t and bounded random variables ξ, the martingale and self-adjointness properties yield

$$\begin{aligned}
E\,(A^n_t - A_t)\,\xi &= E\,(M_t - M^n_t)\,\xi \\
&= E\,E\Big(M_\infty - M^n_\infty \Big| \mathcal{F}_t\Big)\,\xi \\
&= E(M_\infty - M^n_\infty)\,E(\xi \,|\, \mathcal{F}_t) \\
&= E(A^n_\infty - \alpha)\,E(\xi \,|\, \mathcal{F}_t) \to 0,
\end{aligned}$$

as $n \to \infty$ along N'. Thus, $A^n_t \to A_t$ weakly in L^1 for dyadic t. In particular, we get for any dyadic $s < t$

$$\begin{aligned}
0 \ \leq\ & E\Big(A^n_t - A^n_s;\, A_t - A_s < 0\Big) \\
\to\ & E\Big\{(A_t - A_s) \wedge 0\Big\} \leq 0.
\end{aligned}$$

Thus, the last expectation vanishes, and so $A_t \geq A_s$ a.s. By right-continuity it follows that A is a.s. non-decreasing. Also note that $A_0 = 0$ a.s., since $A^n_0 = 0$ for all n.

To see that A is natural, consider any bounded martingale N, and conclude from Fubini's theorem and the martingale properties of N and $A^n - A = M - M^n$ that

$$\begin{aligned} E(N_\infty A^n_\infty) &= \textstyle\sum_k E(N_\infty \Delta^n_k A^n) \\ &= \textstyle\sum_k E(N_{t^n_k} \Delta^n_k A^n) \\ &= \textstyle\sum_k E(N_{t^n_k} \Delta^n_k A) \\ &= E \textstyle\sum_k (N_{t^n_k} \Delta^n_k A). \end{aligned}$$

Using weak convergence on the left and dominated convergence on the right, and combining with Fubini's theorem and the martingale property of N, we get

$$\begin{aligned} E\int_0^\infty N_{t-}\, dA_t \;&=\; E(N_\infty A_\infty) \\ &= \textstyle\sum_k E(N_\infty \Delta^n_k A) \\ &= \textstyle\sum_k E(N_{t^n_{k+1}} \Delta^n_k A) \\ &= E\textstyle\sum_k (N_{t^n_{k+1}} \Delta^n_k A) \\ &\to E\int_0^\infty N_t\, dA_t. \end{aligned}$$

Hence, $E\int_0^\infty (\Delta N_t)\, dA_t = 0$, as required. □

To complete the proof of Theorem 10.5, we need to show that the process A in the last lemma is predictable. This will be inferred from the following ingenious approximation of totally inaccessible times.

Lemma 10.8 *(approximation of inaccessible times, Doob) For a totally inaccessible time τ, put $\tau_n = 2^{-n}[2^n\tau]$, and let X^n be a right-continuous version of the process $P\{\tau_n \le t \,|\, \mathcal{F}_t\}$. Then*

$$\lim_{n\to\infty} \sup_{t\ge 0} \Big| X^n_t - 1\{\tau \le t\} \Big| = 0 \quad a.s. \tag{3}$$

Proof: Since $\tau_n \uparrow \tau$, we may assume that $X^1_t \ge X^2_t \ge \cdots \ge 1\{\tau \le t\}$ for all $t \ge 0$. Then $X^n_t = 1$ for $t \in [\tau, \infty)$, and on the set $\{\tau = \infty\}$ we have $X^1_t \le P(\tau < \infty \,|\, \mathcal{F}_t) \to 0$ a.s. as $t \to \infty$ by Theorem 9.24. Thus, $\sup_n |X^n_t - 1\{\tau \le t\}| \to 0$ a.s. as $t \to \infty$. To prove (3), it is then enough to show that for every $\varepsilon > 0$, the optional times

$$\sigma_n = \inf\Big\{ t \ge 0;\ X^n_t - 1\{\tau \le t\} > \varepsilon \Big\}, \quad n \in \mathsf{N},$$

tend a.s. to infinity. The σ_n are clearly non-decreasing, say with limit σ. Note that either $\sigma_n \le \tau$ or $\sigma_n = \infty$ for all n.

By optional sampling, Theorem 8.5, and Lemma 9.1, we have

$$\begin{aligned} X^n_\sigma 1\{\sigma < \infty\} \;&=\; P\Big\{ \tau_n \le \sigma < \infty \,\Big|\, \mathcal{F}_\sigma \Big\} \\ &\to P\Big\{ \tau \le \sigma < \infty \,\Big|\, \mathcal{F}_\sigma \Big\} \\ &= 1\{\tau \le \sigma < \infty\}. \end{aligned}$$

Hence, $X^n_\sigma \to 1\{\tau \le \sigma\}$ a.s. on $\{\sigma < \infty\}$, and so by right continuity we have on the latter set $\sigma_n < \sigma$ for large enough n. Thus, σ is predictable and announced by the times $\sigma_n \wedge n$.

Applying the optional sampling and disintegration theorems to the optional times σ_n, we obtain

$$\begin{aligned}
\varepsilon P\{\sigma < \infty\} &\le \varepsilon P\{\sigma_n < \infty\} \\
&\le E\Big(X^n_{\sigma_n};\, \sigma_n < \infty\Big) \\
&= P\Big\{\tau_n \le \sigma_n < \infty\Big\} \\
&= P\Big\{\tau_n \le \sigma_n \le \tau < \infty\Big\} \\
&\to P\{\tau = \sigma < \infty\} = 0,
\end{aligned}$$

where the last equality holds since τ is totally inaccessible. Thus, $\sigma = \infty$ a.s. □

It follows easily that A has only accessible jumps:

Lemma 10.9 *(accessibility of jumps) For a natural, increasing process A and a totally inaccessible time τ, we have*

$$\Delta A_\tau = 0 \ \ \textit{a.s. on } \{\tau < \infty\}.$$

Proof: Rescaling if necessary, we may assume that A is a.s. continuous at dyadic times. Define $\tau_n = 2^{-n}[2^n\tau]$. Since A is natural, we have

$$E\int_0^\infty P\Big\{\tau_n > t \,\Big|\, \mathcal{F}_t\Big\}\, dA_t = E\int_0^\infty P\Big\{\tau_n > t \,\Big|\, \mathcal{F}_{t-}\Big\}\, dA_t,$$

and since τ is totally inaccessible, Lemma 10.8 yields

$$\begin{aligned}
E\, A_{\tau-} &= E\int_0^\infty 1\{\tau > t\}\, dA_t \\
&= E\int_0^\infty 1\{\tau \ge t\}\, dA_t = EA_\tau.
\end{aligned}$$

Hence, $E(\Delta A_\tau;\, \tau < \infty) = 0$, and so $\Delta A_\tau = 0$ a.s. on $\{\tau < \infty\}$. □

We may now show that A is predictable.

Lemma 10.10 *(Doléans) Any natural, increasing process is predictable.*

Proof: Fix a natural increasing process A. Consider a bounded martingale M and a predictable time $\tau < \infty$ announced by $\sigma_1, \sigma_2, \ldots$. Then $M^\tau - M^{\sigma_k}$ is again a bounded martingale, and since A is natural, we get by dominated convergence $E(\Delta M_\tau)(\Delta A_\tau) = 0$. In particular, we may take $M_t = P(B \,|\, \mathcal{F}_t)$ with $B \in \mathcal{F}_\tau$. By optional sampling, we have $M_\tau = 1_B$ and

$$\begin{aligned}
M_{\tau-} &\leftarrow M_{\sigma_k} \\
&= P(B \,|\, \mathcal{F}_{\sigma_k}) \\
&\to P(B \,|\, \mathcal{F}_{\tau-}).
\end{aligned}$$

Thus, $\Delta M_\tau = 1_B - P(B \mid \mathcal{F}_{\tau-})$, and so

$$\begin{aligned} E(\Delta A_\tau; B) &= E\Big\{\Delta A_\tau\, P(B \mid \mathcal{F}_{\tau-})\Big\} \\ &= E\Big\{E\Big(\Delta A_\tau \Big| \mathcal{F}_{\tau-}\Big); B\Big\}. \end{aligned}$$

Since $B \in \mathcal{F}_\tau$ was arbitrary, we get $\Delta A_\tau = E(\Delta A_\tau | \mathcal{F}_{\tau-})$ a.s., and so the process $A'_t = (\Delta A_\tau) 1\{\tau \le t\}$ is predictable by Lemma 10.3 (ii). It is also natural, since for any bounded martingale M

$$E\,(\Delta A_\tau\, \Delta M_\tau) = E\Big\{\Delta A_\tau\, E(\Delta M_\tau | \mathcal{F}_{\tau-})\Big\} = 0.$$

An elementary construction yields $\{t > 0;\, \Delta A_t > 0\} \subset \bigcup_n [\tau_n]$ a.s. for some optional times $\tau_n < \infty$, where by Proposition 10.4 and Lemma 10.9 we may choose the latter to be predictable. Taking $\tau = \tau_1$ in the previous argument, we conclude that the process $A^1_t = (\Delta A_{\tau_1}) 1\{\tau_1 \le t\}$ is both natural and predictable. Repeating the argument for the process $A - A^1$ with $\tau = \tau_2$ and proceeding by induction, we see that the jump component A^d of A is predictable. Since $A - A^d$ is continuous and hence predictable, the predictability of A follows. □

To prove the uniqueness, we need to show that every predictable martingale of integrable variation is a constant. For continuous martingales, this will be seen by elementary arguments in Proposition 18.2 below. The two versions are in fact equivalent, since every predictable martingale is continuous by Proposition 10.16 below.

Lemma 10.11 *(predictable martingale) Let M be a martingale of integrable variation. Then*

$$M \text{ is predictable} \quad \Leftrightarrow \quad M \text{ is a.s. constant.}$$

Proof: On the predictable σ-field $\mathcal{P}$, we introduce the signed measure

$$\mu B = E \int_0^\infty 1_B(t)\, dM_t, \quad B \in \mathcal{P},$$

where the inner integral is an ordinary Lebesgue–Stieltjes integral. The martingale property shows that μ vanishes for the sets $B = F \times (t, \infty)$ with $F \in \mathcal{F}_t$. By Lemma 10.2 and a monotone-class argument, it follows that $\mu = 0$ on $\mathcal{P}$. Since M is predictable, the same thing is true for the process $\Delta M_t = M_t - M_{t-}$, and then also for the sets $J_\pm = \{t > 0;\, \pm \Delta M_t > 0\}$. Thus, $\mu J_\pm = 0$, and so $\Delta M = 0$ a.s., which means that M is a.s. continuous. But then $M_t \equiv M_0$ a.s. by Proposition 18.2 below. □

Proof of Theorem 10.5: The sufficiency is obvious, and the uniqueness holds by Lemma 10.11. It remains to show that any local sub-martingale X has the stated decomposition. By Lemmas 10.6 and 10.11, we may take X to be of class (D). Then Lemma 10.7 yields $X = M + A$ for a uniformly integrable martingale M and a natural increasing process A, and the latter is predictable by Lemma 10.10. $\square$

The two properties in Lemma 10.10 are in fact equivalent.

Theorem 10.12 *(natural and predictable processes, Doléans) Let A be an integrable, increasing process. Then*

$$A \text{ is natural} \quad \Leftrightarrow \quad A \text{ is predictable.}$$

Proof: If an integrable, increasing process A is natural, it is also predictable by Lemma 10.10. Now let A be predictable. By Lemma 10.7, we have $A = M + B$ for a uniformly integrable martingale M and a natural increasing process B, and Lemma 10.10 shows that B is predictable. But then $A = B$ a.s. by Lemma 10.11, and so A is natural. $\square$

The following criterion is essentially implicit in earlier proofs.

Lemma 10.13 *(dual predictable projection) Let X, Y be locally integrable, increasing processes, where Y is predictable. Then X has compensator Y iff*

$$E\int V dX = E\int V dY, \quad V \geq 0 \text{ predictable.}$$

Proof: By localization, we may take X, Y to be integrable. Then Y is the compensator of X iff $M = Y - X$ is a martingale, which holds iff $EM_\tau = 0$ for every optional time τ. This is equivalent to the stated relation for $V = 1_{[0,\tau]}$, and the general result follows by a straightforward monotone-class argument. $\square$

We may now establish the fundamental relationship between predictable times and processes.

Theorem 10.14 *(predictable times and processes, Meyer) For any optional time τ, these conditions are equivalent:*

(i) *τ is predictable,*
(ii) *the process $1\{\tau \leq t\}$ is predictable,*
(iii) *$E\Delta M_\tau = 0$ for any bounded martingale M.*

Proof (Chung & Walsh): Since (i) ⇒ (ii) by Lemma 10.3 (ii) and (ii) ⇔ (iii) by Theorem 10.12, it remains to show that (iii) ⇒ (i). Then introduce the martingale $M_t = E(e^{-\tau} | \mathcal{F}_t)$ and super-martingale

$$\begin{aligned} X_t &= e^{-\tau\wedge t} - M_t \\ &= E\Big(e^{-\tau\wedge t} - e^{-\tau}\Big| \mathcal{F}_t\Big) \geq 0, \quad t \geq 0, \end{aligned}$$

and note that $X_\tau = 0$ a.s. by optional sampling. Letting $\sigma = \inf\{t \geq 0;\, X_{t-} \wedge X_t = 0\}$, we see from Lemma 9.32 that $\{t \geq 0;\, X_t = 0\} = [\sigma, \infty)$ a.s., and in particular $\sigma \le \tau$ a.s. Using optional sampling again, we get $E(e^{-\sigma} - e^{-\tau}) = EX_\sigma = 0$, and so $\sigma = \tau$ a.s. Hence, $X_t \wedge X_{t-} > 0$ a.s. on $[0, \tau)$. Finally, (iii) yields

$$\begin{aligned} EX_{\tau-} &= E(e^{-\tau} - M_{\tau-}) \\ &= E(e^{-\tau} - M_\tau) \\ &= EX_\tau = 0, \end{aligned}$$

and so $X_{\tau-} = 0$. It is now clear that τ is announced by the optional times $\tau_n = \inf\{t;\, X_t < n^{-1}\}$. □

To illustrate the power of the last result, we give a short proof of the following useful statement, which can also be proved directly.

Corollary 10.15 *(predictable restriction) For any predictable time τ and set $A \in \mathcal{F}_{\tau-}$, the restriction τ_A is again predictable.*

Proof: The process $1_A 1\{\tau \le t\} = 1\{\tau_A \le t\}$ is predictable by Lemma 10.3, and so the time τ_A is predictable by Theorem 10.14. □

Using the last theorem, we may show that predictable martingales are continuous.

Proposition 10.16 *(predictable martingale) For a local martingale M,*

$$M \text{ is predictable} \quad \Leftrightarrow \quad M \text{ is a.s. continuous}$$

Proof: The sufficiency is clear by definitions. To prove the necessity, we note that for any optional time τ,

$$M_t^\tau = M_t\, 1_{[0,\tau]}(t) + M_\tau\, 1_{(\tau,\infty)}(t), \quad t \ge 0.$$

Thus, the predictability is preserved by optional stopping, and so we may take M to be is a uniformly integrable martingale. Now fix any $\varepsilon > 0$, and introduce the optional time $\tau = \inf\{t > 0;\, |\Delta M_t| > \varepsilon\}$. Since the left-continuous version M_{t-} is predictable, so is the process ΔM_t, as well as the random set $A = \{t > 0;\, |\Delta M_t| > \varepsilon\}$. Hence, the interval $[\tau, \infty) = A \cup (\tau, \infty)$ has the same property, and so τ is predictable by Theorem 10.14. Choosing an announcing sequence (τ_n), we conclude that, by optional sampling, martingale convergence, and Lemmas 10.1 (iii) and 10.3 (i),

$$\begin{aligned} M_{\tau-} &\leftarrow M_{\tau_n} \\ &= E(M_\tau | \mathcal{F}_{\tau_n}) \\ &\to E(M_\tau | \mathcal{F}_{\tau-}) = M_\tau. \end{aligned}$$

Thus, $\tau = \infty$ a.s., and ε being arbitrary, it follows that M is a.s. continuous. □

We turn to a more refined decomposition of increasing processes, extending the decomposition of optional times in Proposition 10.4. For convenience here and below, any right-continuous, non-decreasing process X on R_+ starting at 0 may be identified with a random measure ξ on $(0,\infty)$, via the relationship in Theorem 2.14, so that $\xi[0,t] = X_t$ for all $t \geq 0$. We say that ξ is *adapted* or *predictable*[2] if the corresponding property holds for X. The *compensator* of ξ is defined as the predictable random measure $\hat{\xi}$ associated with the compensator $\hat{X}$ of X.

An rcll process X or filtration $\mathcal{F}$ is said to be *quasi-left continuous*[3] (*ql-continuous* for short) if $\Delta X_\tau = 0$ a.s. on $\{\tau < \infty\}$ or $\mathcal{F}_\tau = \mathcal{F}_{\tau-}$, respectively, for every predictable time τ. We further say that X has *accessible jumps*, if $\Delta X_\tau = 0$ a.s. on $\{\tau < \infty\}$ for every totally inaccessible time τ. Accordingly, a random measure ξ on $(0,\infty)$ is said to be *ql-continuous* if $\xi\{\tau\} = 0$ a.s. for every predictable time τ, and *accessible* if it is purely atomic with $\xi\{\tau\} = 0$ a.s. for every totally inaccessible time τ, where $\mu\{\infty\} = 0$ for measures μ on R_+.

Theorem 10.17 *(decomposition of random measure) Let ξ be an adapted random measure on $(0,\infty)$ with compensator $\hat{\xi}$. Then*

(i) *ξ has an a.s. unique decomposition $\xi = \xi^c + \xi^q + \xi^a$, where ξ^c is diffuse, $\xi^q + \xi^a$ is purely atomic, ξ^q is ql-continuous, and ξ^a is accessible,*

(ii) *ξ^a is a.s. supported by $\bigcup_n [\tau_n]$, for some predictable times $\tau_1, \tau_2, \ldots$ with disjoint graphs,*

(iii) *when ξ is locally integrable, $\xi^c + \xi^q$ has compensator $(\hat{\xi})^c$, and ξ is ql-continuous iff $\hat{\xi}$ is a.s. continuous.*

Proof: Subtracting the predictable component ξ^c, we may take ξ to be purely atomic. Put $\eta = \hat{\xi}$ and $A_t = \eta(0,t]$. Consider the locally integrable process $X_t = \sum_{s\leq t}(\Delta A_s \wedge 1)$ with compensator $\hat{X}$, and define

$$\begin{aligned} A_t^q &= \int_0^{t+} 1\{\Delta \hat{X}_s = 0\}\, dA_s, \\ A_t^a &= A_t - A_t^q, \qquad\qquad t \geq 0. \end{aligned}$$

For any predictable time $\tau < \infty$, the graph $[\tau]$ is again predictable by Theorem 10.14, and so by Theorem 10.13,

$$\begin{aligned} E(\Delta A_\tau^q \wedge 1) &= E\Big(\Delta X_\tau;\, \Delta \hat{X}_\tau = 0\Big) \\ &= E\Big(\Delta \hat{X}_\tau;\, \Delta \hat{X}_\tau = 0\Big) = 0, \end{aligned}$$

which shows that A^q is ql-continuous.

Now let $\tau_{n0} = 0$ for $n \in \mathsf{N}$, and define recursively the random times

$$\tau_{nk} = \inf\Big\{t > \tau_{n,k-1};\, \Delta \hat{X}_t \in 2^{-n}(1,2]\Big\}, \quad n,k \in \mathsf{N}.$$

[2]Then ξB is clearly measurable for every $B \in \mathcal{B}$, so that ξ is indeed a random measure in the sense of Chapters 15, 23, and 29–31 below.

[3]This horrible term is well established.

They are predictable by Theorem 10.14, and $\{t > 0;\, \Delta A^a_t > 0\} = \bigcup_{n,k}[\tau_{nk}]$ a.s. by the definition of A^a. Thus, for any totally inaccessible time τ, we have $\Delta A^a_\tau = 0$ a.s. on $\{\tau < \infty\}$, which shows that A^a has accessible jumps.

If $A = B^q + B^a$ is another decomposition with the stated properties, then $Y = A^q - B^q = B^a - A^a$ is ql-continuous with accessible jumps, and so Lemma 10.4 gives $\Delta Y_\tau = 0$ a.s. on $\{\tau < \infty\}$ for any optional time τ, which means that Y is a.s. continuous. Since it is also purely discontinuous, we get $Y = 0$ a.s., which proves the asserted uniqueness.

When A is locally integrable, we may write instead $A^q = 1\{\Delta\hat{A} = 0\} \cdot A$, and note that $(\hat{A})^c = 1\{\Delta\hat{A} = 0\} \cdot \hat{A}$. For any predictable process $V \geq 0$, we get by Theorem 10.13

$$\begin{aligned} E\int V dA^q &= E\int 1\{\Delta\hat{A} = 0\}\, V dA \\ &= E\int 1\{\Delta\hat{A} = 0\}\, V d\hat{A} \\ &= E\int V d(\hat{A})^c, \end{aligned}$$

which shows that A^q has compensator $(\hat{A})^c$. □

By the *compensator* of an optional time τ we mean the compensator of the associated jump process $X_t = 1\{\tau \leq t\}$. We may characterize the different categories of optional times in terms of their compensators.

Corollary 10.18 *(compensation of optional time) Let τ be an optional time with compensator A. Then*

(i) *τ is predictable iff A is a.s. constant apart from a possible unit jump,*
(ii) *τ is accessible iff A is a.s. purely discontinuous,*
(iii) *τ is totally inaccessible iff A is a.s. continuous,*
(iv) *τ has the accessible part τ_D, where $D = \{\Delta A_\tau > 0,\, \tau < \infty\}$.*

Proof: (i) If τ is predictable, so is the process $X_t = 1\{\tau \leq t\}$ by Theorem 10.14, and hence $A = X$ a.s. Conversely, if $A_t = 1\{\sigma \leq t\}$ for an optional time σ, the latter is predictable by Theorem 10.14, and Lemma 10.13 yields

$$\begin{aligned} P\{\sigma = \tau < \infty\} &= E\Big(\Delta X_\sigma;\, \sigma < \infty\Big) \\ &= E\Big(\Delta A_\sigma;\, \sigma < \infty\Big) \\ &= P\{\sigma < \infty\} \\ &= EA_\infty = EX_\infty \\ &= P\{\tau < \infty\}. \end{aligned}$$

Thus, $\tau = \sigma$ a.s., and so τ is predictable.

(ii) Clearly τ is accessible iff X has accessible jumps, which holds by Theorem 10.17 iff $A = A^d$ a.s.

(iii) The time τ is totally inaccessible iff X is ql-continuous, which holds by Theorem 10.17 iff $A = A^c$ a.s.

(iv) Use (ii) and (iii). □

Next we characterize quasi-left continuity for filtrations and martingales.

Proposition 10.19 *(ql-continuous filtration, Meyer) For a filtration $\mathcal{F}$, these conditions are equivalent:*

(i) *every accessible time is predictable,*
(ii) $\mathcal{F}_{\tau-} = \mathcal{F}_\tau$ *on* $\{\tau < \infty\}$ *for all predictable times* τ*,*
(iii) $\Delta M_\tau = 0$ *a.s. on* $\{\tau < \infty\}$ *for any martingale* M *and predictable time* τ*.*

If the basic σ-field in Ω is $\mathcal{F}_\infty$, then $\mathcal{F}_{\tau-} = \mathcal{F}_\tau$ on $\{\tau = \infty\}$ for any optional time τ, and the relation in (ii) extends to all of Ω.

Proof, (i) ⇒ (ii): Let τ be a predictable time, and fix any $B \in \mathcal{F}_\tau \cap \{\tau < \infty\}$. Then $[\tau_B] \subset [\tau]$, and so τ_B is accessible, hence by (i) even predictable. The process $X_t = 1\{\tau_B \le t\}$ is then predictable by Theorem 10.14, and since

$$X_\tau 1\{\tau < \infty\} = 1\Big\{\tau_B \le \tau < \infty\Big\} = 1_B,$$

Lemma 10.3 (i) yields $B \in \mathcal{F}_{\tau-}$.

(ii) ⇒ (iii): Fix a martingale M, and let τ be a bounded, predictable time announced by (τ_n). Using (ii) and Lemma 10.1 (iii), we get as before

$$\begin{aligned} M_{\tau-} \leftarrow M_{\tau_n} &= E(M_\tau | \mathcal{F}_{\tau_n}) \\ &\to E(M_\tau | \mathcal{F}_{\tau-}) \\ &= E(M_\tau | \mathcal{F}_\tau) = M_\tau, \end{aligned}$$

and so $M_{\tau-} = M_\tau$ a.s.

(iii) ⇒ (i): If τ is accessible, Proposition 10.4 yields some predictable times τ_n with $[\tau] \subset \bigcup_n [\tau_n]$ a.s. By (iii) we have $\Delta M_{\tau_n} = 0$ a.s. on $\{\tau_n < \infty\}$ for every martingale M and all n, and so $\Delta M_\tau = 0$ a.s. on $\{\tau < \infty\}$. Hence, τ is predictable by Theorem 10.14. □

The following inequality will be needed in Chapter 20.

Proposition 10.20 *(norm relation, Garsia, Neveu) Let A be a right- or left-continuous, predictable, increasing process, and let $\zeta \ge 0$ be a random variable such that a.s.*

$$E\Big(A_\infty - A_t \,\Big|\, \mathcal{F}_t\Big) \le E(\zeta \,|\, \mathcal{F}_t), \quad t \ge 0. \tag{4}$$

Then

$$\|A_\infty\|_p \le p\, \|\zeta\|_p, \quad p \ge 1.$$

In the left-continuous case, predictability is clearly equivalent to adaptedness. The proper interpretation of (4) is to take $E(A_t | \mathcal{F}_t) \equiv A_t$, and choose right-continuous versions of the martingales $E(A_\infty | \mathcal{F}_t)$ and $E(\zeta | \mathcal{F}_t)$. For a right-continuous A, we may clearly take $\zeta = Z^*$, where Z is the super-martingale on the left of (4). We also note that if A is the compensator of an increasing process X, then (4) holds with $\zeta = X_\infty$.

Proof: We may take A to be right-continuous, the left-continuous case being similar but simpler. We can also take A to be bounded, since we may otherwise replace A by the process $A \wedge u$ for arbitrary $u > 0$, and let $u \to \infty$ in the resulting formula. For each $r > 0$, the random time $\tau_r = \inf\{t;\ A_t \geq r\}$ is predictable by Theorem 10.14. By optional sampling and Lemma 10.1, we note that (4) remains valid with t replaced by τ_r-. Since τ_r is $\mathcal{F}_{\tau_r-}$-measurable by the same lemma, we obtain

$$\begin{aligned} E\Big(A_\infty - r;\ A_\infty > r\Big) &\leq E\Big(A_\infty - r;\ \tau_r < \infty\Big) \\ &\leq E\Big(A_\infty - A_{\tau_r-};\ \tau_r < \infty\Big) \\ &\leq E(\zeta;\ \tau_r < \infty) \\ &\leq E(\zeta;\ A_\infty \geq r). \end{aligned}$$

Writing $A_\infty = \alpha$ and letting $p^{-1}+q^{-1} = 1$, we get by Fubini's theorem, Hölder's inequality, and some calculus

$$\begin{aligned} \|\alpha\|_p^p &= p^2\, q^{-1} E \int_0^\alpha (\alpha - r)\, r^{p-2}\, dr \\ &= p^2\, q^{-1} \int_0^\infty E\Big(\alpha - r;\ \alpha > r\Big)\, r^{p-2}\, dr \\ &\leq p^2\, q^{-1} \int_0^\infty E\Big(\zeta;\ \alpha \geq r\Big)\, r^{p-2}\, dr \\ &= p^2\, q^{-1} E\, \zeta \int_0^\alpha r^{p-2}\, dr \\ &= p\, E\, \zeta \alpha^{p-1} \\ &\leq p\, \|\zeta\|_p\, \|\alpha\|_p^{p-1}. \end{aligned}$$

If $\|\alpha\|_p > 0$, we may finally divide both sides by $\|\alpha\|_p^{p-1}$. □

We can't avoid the random measure point of view when turning to processes on a product space $\mathsf{R}_+ \times S$, where S is a localized Borel space. Here a *random measure* is simply a locally finite kernel $\xi : \Omega \to (0, \infty) \times S$, and we say that ξ is *adapted*, *predictable*, or *locally integrable* if the corresponding properties hold for the real-valued processes $\xi_t B = \xi\{(0, t] \times B\}$ with $B \in \hat{\mathcal{S}}$ fixed. The *compensator* of ξ is defined as a predictable random measure $\hat{\xi}$ on $(0, \infty) \times S$, such that $\hat{\xi}_t B$ is a compensator of $\xi_t B$ for every $B \in \hat{\mathcal{S}}$.

For an alternative approach, suggested by Lemma 10.13, we say that a process V on $\mathsf{R}_+ \times S$ is *predictable* if it is $\mathcal{P} \times \mathcal{S}$-measurable, where $\mathcal{P}$ denotes the predictable σ-field on R_+. The compensator $\hat{\xi}$ may then be characterized as the a.s. unique, predictable random measure on $(0, \infty) \times S$, such that $E\, \xi V = E\, \hat{\xi} V$ for every predictable process $V \geq 0$ on $\mathsf{R}_+ \times S$.

Theorem 10.21 *(compensation of random measure, Grigelionis, Jacod) Let ξ be a locally integrable, adapted random measure on $(0, \infty) \times S$, where S*

is Borel. Then there exists an a.s. unique predictable random measure $\hat{\xi}$ *on* $(0,\infty)\times S$*, such that for any predictable process* $V\geq 0$ *on* $\mathsf{R}_+\times S$*,*

$$E\int V d\xi = E\int V d\hat{\xi}.$$

Again we emphasize that the compensator $\hat{\xi}$ depends in a crucial way on the choice of underlying filtration $\mathcal{F}$. Our proof relies on a technical lemma, which can be established by straightforward monotone-class arguments.

Lemma 10.22 *(predictable random measures)*

(i) *For any predictable random measure* ξ *and predictable process* $V\geq 0$ *on* $(0,\infty)\times S$*, the process* $V\cdot\xi$ *is again predictable.*

(ii) *For any predictable process* $V\geq 0$ *on* $(0,\infty)\times S$ *and predictable, measure-valued process* ρ *on* S*, the process* $Y_t=\int V_{t,s}\,\rho_t(ds)$ *is again predictable.*

Proof of Theorem 10.21: Since ξ is locally integrable, we may choose a predictable process $V>0$ on $\mathsf{R}_+\times S$ such that $E\int V d\xi<\infty$. If the random measure $\zeta=V\cdot\xi$ has compensator $\hat{\zeta}$, then Lemma 10.22 shows that ξ has compensator $\hat{\xi}=V^{-1}\cdot\hat{\zeta}$. Thus, we may henceforth assume that $E\xi\{(0,\infty)\times S\}=1$.

Put $\eta=\xi(\cdot\times S)$. Using the kernel composition of Chapter 3, we may introduce the probability measure $\mu=P\otimes\xi$ on $\Omega\times\mathsf{R}_+\times S$ with projection $\nu=P\otimes\eta$ onto $\Omega\times\mathsf{R}_+$. Applying Theorem 3.4 to the restrictions of μ and ν to the σ-fields $\mathcal{P}\otimes\mathcal{S}$ and $\mathcal{P}$, respectively, we obtain a probability kernel $\rho\colon(\Omega\times\mathsf{R}_+,\mathcal{P})\to(S,\mathcal{S})$ satisfying $\mu=\nu\otimes\rho$, or

$$P\otimes\xi=P\otimes\eta\otimes\rho \quad \text{on} \quad (\Omega\times\mathsf{R}_+\times S,\,\mathcal{P}\times\mathcal{S}).$$

Letting $\hat{\eta}$ be the compensator of η, we may introduce the random measure $\hat{\xi}=\hat{\eta}\otimes\rho$ on $\mathsf{R}_+\times S$.

To see that $\hat{\xi}$ is the compensator of ξ, we note that $\hat{\xi}$ is predictable by Lemma 10.22 (i). For any predictable process $V\geq 0$ on $\mathsf{R}_+\times S$, the process $Y_s=\int V_{s,t}\,\rho_t(ds)$ is again predictable by Lemma 10.22 (ii). By Theorem 8.5 and Lemma 10.13, we get

$$\begin{aligned} E\int V d\hat{\xi} &= E\int \hat{\eta}(dt)\int V_{s,t}\,\rho_t(ds)\\ &= E\int \eta(dt)\int V_{s,t}\,\rho_t(ds)\\ &= E\int V d\xi. \end{aligned}$$

Finally note that $\hat{\xi}$ is a.s. unique by Lemma 10.13. □

The theory simplifies when the filtration $\mathcal{F}$ is *induced* by ξ, in the sense of being the smallest right-continuous, complete filtration making ξ adapted, in which case we call $\hat{\xi}$ the *induced compensator*. Here we consider only the special case of a single point mass $\xi=\delta_{\tau,\chi}$, where τ is a random time with associated *mark* χ in S.

Proposition 10.23 *(induced compensator) Let (τ, ζ) be a random element in $(0, \infty] \times S$ with distribution μ, where S is Borel. Then $\xi = \delta_{\tau,\zeta}$ has the induced compensator*

$$\hat{\xi}_t B = \int_{(0,\, t\wedge\tau]} \frac{\mu(dr \times B)}{\mu([r, \infty] \times S)}, \quad t \geq 0,\;\; B \in \mathcal{S}. \tag{5}$$

Proof: The process $\eta_t B$ on the right of (5) is clearly predictable for every $B \in \mathcal{S}$. It remains to show that $M_t = \xi_t B - \eta_t B$ is a martingale, hence that $E(M_t - M_s; A) = 0$ for any $s < t$ and $A \in \mathcal{F}_s$. Since $M_t = M_s$ on $\{\tau \leq s\}$, and the set $\{\tau > s\}$ is a.s. an atom of $\mathcal{F}_s$, it suffices to show that $E(M_t - M_s) = 0$, or $EM_t \equiv 0$. Then use Fubini's theorem to get

$$\begin{aligned} E\,\eta_t B &= E \int_{(0,t\wedge\tau]} \frac{\mu(dr \times B)}{\mu([r, \infty] \times S)} \\ &= \int_{(0,\infty]} \mu(dx) \int_{(0,t\wedge x]} \frac{\mu(dr \times B)}{\mu([r, \infty] \times S)} \\ &= \int_{(0,t]} \frac{\mu(dr \times B)}{\mu([r, \infty] \times S)} \int_{[r,\infty]} \mu(dx) \\ &= \mu\{(0, t] \times B\} = E\,\xi_t B. \end{aligned}$$

□

Now return to the case of a general filtration $\mathcal{F}$:

Theorem 10.24 *(discounted compensator) For any adapted pair (τ, χ) in $(0, \infty] \times S$ with compensator η, there exists a unique, predictable random measure ζ on $(0, \tau] \times S$ with $\|\zeta\| \leq 1$, satisfying* (5) *with μ replaced by ζ. Writing $Z_t = 1 - \bar{\zeta}_t$, we have a.s.*

(i) $\zeta = Z_- \cdot \eta$, *and Z is the unique solution of $Z = 1 - Z_- \cdot \bar{\eta}$,*

(ii) $Z_t = \exp(-\bar{\eta}^c_t) \prod_{s \leq t} (1 - \Delta\bar{\eta}_s), \quad t \geq 0,$

(iii) $\Delta\bar{\eta} \begin{cases} < 1 \text{ on } [0, \tau), \\ \leq 1 \text{ on } [\tau], \end{cases}$

(iv) *Z is non-increasing and ≥ 0 with $Z_{\tau-} > 0$,*

(v) *$Y = Z^{-1}$ satisfies $d\,Y_t = Y_t\, d\,\bar{\eta}_t$ on $\{Z_t > 0\}$.*

The process in (ii) is a special case of *Doléans' exponential* in Chapter 20. Though the subsequent discussion involves some stochastic differential equations, the present versions are elementary, and there is no need for the general theory in Chapters 18–20 and 32–33.

Proof: (iii) The time $\sigma = \inf\{t \geq 0;\, \Delta\bar{\eta}_t \geq 1\}$ is optional, by an elementary approximation. Hence, the random interval (σ, ∞) is predictable, and so the same thing is true for the graph $[\sigma] = \{\Delta\bar{\eta} \geq 1\} \setminus (\sigma, \infty)$. Thus,

$$P\{\tau = \sigma\} = E\,\bar{\xi}[\sigma] = E\,\bar{\eta}[\sigma].$$

Since also

$$1\{\tau = \sigma\} \le 1\{\sigma < \infty\} \le \bar\eta[\sigma],$$

by the definition of σ, we obtain

$$1\{\tau = \sigma\} = \bar\eta[\sigma] = \Delta\bar\eta_\sigma \text{ a.s. on } \{\sigma < \infty\},$$

which implies $\Delta\bar\eta_t \le 1$ and $\tau = \sigma$ a.s. on the same set.

(i)–(ii): If (5) holds with $\mu = \zeta$ for some random measure ζ on $(0, \tau] \times S$, then the process $Z_t = 1 - \bar\zeta_t$ satisfies

$$\begin{aligned} d\,\bar\eta_t &= Z_{t-}^{-1} d\,\bar\zeta_t \\ &= -Z_{t-}^{-1} d\,Z_t, \end{aligned}$$

and so $d\,Z_t = -Z_t\,d\,\bar\eta_t$, which implies $Z_t - 1 = -(Z \cdot \bar\eta)_t$, or $Z = 1 - Z_- \cdot \bar\eta$. Conversely, any solution Z to the latter equation yields a measure solving (5) with $B = S$, and so (5) has the general solution $\zeta = Z_- \cdot \eta$. Furthermore, the equation $Z = 1 - Z_- \cdot \bar\eta$ has the unique solution (ii), by an elementary special case of Theorem 20.8 below. In particular, Z is predictable, and the predictability of ζ follows by Lemma 10.22.

(iv) Since $1 - \Delta\bar\eta \ge 0$ a.s. by (iii), (ii) shows that Z is a.s. non-increasing with $Z \ge 0$. Since also $\sum_t \Delta\bar\eta_t \le \bar\eta_\tau < \infty$ a.s., and $\sup_{t<\tau} \Delta\bar\eta_t < 1$ a.s. by (iii), we have $Z_{\tau-} > 0$.

(v) By an elementary integration by parts, we get on the set $\{t \ge 0;\, Z_t > 0\}$

$$\begin{aligned} 0 &= d(Z_t\,Y_t) \\ &= Z_{t-}\,d\,Y_t + Y_t\,d\,Z_t. \end{aligned}$$

Using the chain rule for Stieltjes integrals along with equation $Z = 1 - Z_- \cdot \bar\eta$, we obtain

$$\begin{aligned} d\,Y_t &= -Z_{t-}^{-1}\,Y_t\,d\,Z_t \\ &= Z_{t-}^{-1}\,Y_t\,Z_{t-}\,d\,\bar\eta_t \\ &= Y_t\,d\,\bar\eta_t. \end{aligned} \qquad \square$$

The power of the discounted compensator is mainly due to the following result and its sequels, which will play key roles in Chapters 15 and 27. Letting τ, χ, η, ζ, Z be such as in Theorem 10.24, we introduce a predictable process V on $\mathsf{R}_+ \times S$ with $\zeta|V| < \infty$ a.s., and form the processes[4]

$$U_{t,x} = V_{t,x} + Z_t^{-1} \int_0^{t+}\!\!\int V\,d\,\zeta, \qquad t \ge 0,\ x \in S, \tag{6}$$

$$M_t = U_{\tau,\chi} 1\{\tau \le t\} - \int_0^{t+}\!\!\int U\,d\,\eta, \quad t \ge 0, \tag{7}$$

whenever these expressions exist.

Proposition 10.25 *(fundamental martingale) For τ, χ, η, ζ, Z as in Theorem 10.24, let V be a predictable process on $\mathsf{R}_+ \times S$ with $\zeta|V| < \infty$ a.s. and $\zeta V = 0$ a.s. on $\{Z_\tau = 0\}$. Then for U, M as above, we have*

[4]Recall our convention 0/0=0.

(i) *M exists on $[0,\infty]$ and satisfies $M_\infty = V_{\tau,\chi}$ a.s.,*

(ii) *$E|U_{\tau,\chi}| < \infty$ implies*[5] *$EV_{\tau,\chi} = 0$, and M is a uniformly integrable martingale with $\|M^*\|_p \lesssim \|V_{\tau,\chi}\|_p$ for all $p > 1$.*

Proof: (i) Write $Y = Z^{-1}$. Using the conditions on V, the definition of ζ, and Theorem 10.24 (iv), we obtain

$$\begin{aligned}\eta|V| &= \zeta(Y_-|V|)\\ &\le Y_{\tau-}\,\zeta|V| < \infty.\end{aligned}$$

Next, we see from (6) and Theorem 10.24 (v) that

$$\begin{aligned}\eta|U-V| &\le \zeta|V|\,\bar{\eta}Y\\ &= (Y_\tau - 1)\,\zeta|V| < \infty,\end{aligned}$$

whenever $Z_\tau > 0$. If instead $Z_\tau = 0$, we have

$$\begin{aligned}\eta|U-V| &\le \zeta|V| \int_0^{\tau-} Y\,d\bar{\eta}\\ &= (Y_{\tau-} - 1)\,\zeta|V| < \infty.\end{aligned}$$

In either case U is η-integrable, and M is well defined.

Now let $t \ge 0$ with $Z_t > 0$. Using (6), Theorem 10.24 (v), Fubini's theorem, and the definition of ζ, we get for any $x \in S$

$$\begin{aligned}\int_0^{t+}\!\!\int (U-V)\,d\eta &= \int_0^{t+} Y_s\,d\bar{\eta}_s \int_0^{s+}\!\!\int V\,d\zeta\\ &= \int_0^{t+} dY_s \int_0^{s+}\!\!\int V\,d\zeta\\ &= \int_0^{t+}\!\!\int (Y_t - Y_{s-})\,V_{s,y}\,d\zeta_{s,y}\\ &= U_{t,x} - V_{t,x} - \int_0^{t+}\!\!\int V\,d\eta.\end{aligned}$$

Simplifying and combining with (6), we get

$$\int_0^{t+}\!\!\int U\,d\eta = U_{t,x} - V_{t,x} = Y_t \int_0^{t+}\!\!\int V\,d\zeta.$$

To extend to general t, suppose that $Z_\tau = 0$. Using the previous version of (8), the definition of ζ, and the conditions on V, we get

$$\begin{aligned}\int_0^{\tau+}\!\!\int U\,d\eta &= \int_0^{\tau-}\!\!\int U\,d\eta + \int_{[\tau]}\!\int U\,d\eta\\ &= Y_{\tau-}\int_0^{\tau-}\!\!\int V\,d\zeta + \int_{[\tau]}\!\int V\,d\eta\\ &= Y_{\tau-}\int_0^{\tau+}\!\!\int V\,d\zeta = 0\\ &= U_{\tau,x} - V_{\tau,x},\end{aligned}$$

[5]Thus, a mere *integrability* condition on U implies a moment *identity* for V.

which shows that (8) is generally true. In particular,

$$\begin{aligned} V_{\tau,\chi} &= U_{\tau,\chi} - \int_0^{\tau+}\!\!\int U\,d\eta \\ &= M_\tau = M_\infty. \end{aligned}$$

(ii) If $E\,|U_{\tau,\chi}| < \infty$, then $E\,\eta|U| < \infty$ by compensation, which shows that M is uniformly integrable. For any optional time σ, the process $1_{[0,\sigma]}\,U$ is again predictable and $E\eta$-integrable, and so by the compensation property and (7),

$$\begin{aligned} EM_\sigma &= E\Big(U_{\tau,\chi};\, \tau \le \sigma\Big) - E\int_0^{\sigma+}\!\!\int U\,d\eta \\ &= E\,\delta_{\tau,\chi}(1_{[0,\sigma]}\,U) - E\,\eta(1_{[0,\sigma]}\,U) = 0, \end{aligned}$$

which shows that M is a martingale. Thus, in view of (i),

$$EV_{\tau,\chi} = EM_\infty = EM_0 = 0.$$

Furthermore, by Doob's inequality,

$$\|M^*\|_p \lesssim \|M_\infty\|_p = \|V_{\tau,\chi}\|_p, \quad p > 1. \qquad \square$$

We also need a multi-variate version of the last result. Here the optional times $\tau_1, \dots, \tau_n$ are said to be *orthogonal*, if they are a.s. distinct, and the atomic parts of their compensators have disjoint supports.

Corollary 10.26 *(product moments) For every $j \le m$, consider an adapted pair (τ_j, χ_j) in $(0,\infty) \times S_j$ and a predictable process V_j on $\mathsf{R}_+ \times S_j$, where the τ_j are orthogonal. Define Z_j, ζ_j, U_j as in Theorem 10.24 and* (6)*, and suppose that for all $j \le m$,*

(i) $\zeta_j|V_j| < \infty$, $\zeta_j V_j = 0$ *a.s. on* $\{Z_j(\tau_j) = 0\}$,

(ii) $E\,|U_j(\tau_j, \chi_j)| < \infty$, $E\,|V_j(\tau_j, \chi_j)|^{p_j} < \infty$,

for some $p_1, \dots, p_m > 0$ with $\sum_j p_j^{-1} \le 1$. Then

(iii) $E\prod_{j \le m} V_j(\tau_j, \chi_j) = 0.$

Proof: Let the pairs (τ_j, χ_j) have compensators $\eta_1, \dots, \eta_m$, and define the martingales $M_1, \dots, M_m$ as in (7). Fix any $i \ne j$ in $\{1, \dots, m\}$, and choose some predictable times $\sigma_1, \sigma_2, \dots$ as in Theorem 10.17, such that $\{t > 0;\, \Delta\bar\eta_t > 0\} = \bigcup_k [\sigma_k]$ a.s. By compensation and orthogonality, we have for any $k \in \mathsf{N}$

$$\begin{aligned} E\,\delta_{\sigma_k}[\tau_j] &= P\{\tau_j = \sigma_k\} \\ &= E\,\delta_{\tau_j}[\sigma_k] \\ &= E\,\eta_j[\sigma_k] = 0. \end{aligned}$$

Summing over k gives $\eta_k[\tau_j] = 0$ a.s., which shows that $(\Delta M_i)(\Delta M_j) = 0$ a.s. for all $i \ne j$. Integrating repeatedly by parts, we conclude that $M = \prod_j M_j$ is a local martingale. Letting $p^{-1} = \sum_j p_j^{-1} \le 1$, we get by Hölder's inequality, Proposition 10.25 (ii), and the various hypotheses

$$\begin{aligned} \|M^*\|_1 &\le \|M^*\|_p \\ &\le \textstyle\prod_j \|M_j^*\|_{p_j} \\ &\lesssim \textstyle\prod_j \|V_{\tau_j,\chi_j}\|_{p_j} < \infty. \end{aligned}$$

Thus, M is a uniformly integrable martingale, and so by Lemma 10.25 (i),

$$\begin{aligned} E\textstyle\prod_j V_j(\tau_j,\chi_j) &= E\textstyle\prod_j M_j(\infty) \\ &= EM(\infty) \\ &= EM(0) = 0. \end{aligned}$$

□

This yields a powerful predictable mapping theorem based on the discounted compensator. Important extensions and applications appear in Chapters 15 and 27.

Theorem 10.27 *(predictable mapping) Suppose that for all $j \le m$,*

(i) (τ_j, χ_j) *is adapted in* $(0,\infty) \times K_j$ *with discounted compensator* ζ_j,

(ii) Y_j *is a predictable map of* $\mathsf{R}_+ \times K_j$ *into a probability space* (S_j, μ_j),

(iii) *the* τ_j *are orthogonal and* $\zeta_j \circ Y_j^{-1} \le \mu_j$ *a.s.*

Then the variables $\gamma_j = Y_j(\tau_j, \chi_j)$ *are independent with distributions*[6] μ_j.

Proof: For fixed $B_j \in \mathcal{S}_j$, consider the predictable processes

$$V_j(t,x) = 1_{B_j} \circ Y_j(t,x) - \mu_j B_j, \quad t \ge 0,\ x \in K_j,\ j \le m.$$

By definitions and hypotheses,

$$\begin{aligned} \int_0^{t+}\!\!\int V_j\, d\zeta_j &= \int_0^{t+}\!\!\int 1_{B_j}(Y_j)\, d\zeta_j - \mu_j B_j \big\{1 - Z_j(t)\big\} \\ &\le Z_j(t)\, \mu_j B_j. \end{aligned}$$

Since changing from B_j to B_j^c affects only the sign of V_j, the two versions combine into

$$-Z_j(t)\, \mu_j B_j^c \le \int_0^{t+}\!\!\int V_j\, d\zeta_j \le Z_j(t)\, \mu_j B_j.$$

In particular, $|\zeta_j V_j| \le Z_j(\tau_j)$ a.s., and so $\zeta_j V_j = 0$ a.s. on $\{Z_j(\tau_j) = 0\}$. Defining U_j as in (6) and using the previous estimates, we get

$$\begin{aligned} -1 &\le -1_{B_j^c} \circ Y_j \\ &= V_j - \mu_j B_j^c \le U_j \\ &\le V_j + \mu_j B_j \\ &= 1_{B_j} \circ Y_j \le 1, \end{aligned}$$

[6] Thus, the mere *inequality* in (iii) implies the *equality* $\mathcal{L}(\gamma_j) = \mu_j$.

which implies $|U_j| \leq 1$. Letting $\emptyset \neq J \subset \{1,\dots,m\}$ and applying Corollary 10.26 with $p_j = |J|$ for all $j \in J$, we obtain

$$E\prod_{j\in J}\Big\{1_{B_j}(\gamma_j) - \mu_j B_j\Big\} = E\prod_{j\in J} V_j(\tau_j,\chi_j) = 0. \tag{8}$$

We may now use induction on $|J|$ to prove that

$$P\bigcap_{j\in J}\{\gamma_j \in B_j\} = \prod_{j\in J}\mu_j B_j, \quad J \subset \{1,\dots,m\}. \tag{9}$$

For $|J| = 1$ this holds by (8). Now assume (9) for all $|J| < k$, and proceed to a J with $|J| = k$. Expanding the product in (8), and applying the induction hypothesis to each term involving at least one factor $\mu_j B_j$, we see that all terms but one reduce to $\pm\prod_{j\in J}\mu_j B_j$, whereas the remaining term equals $P\bigcap_{j\in J}\{\gamma_j \in B_j\}$. Thus, (9) remains true for J, which completes the induction. By a monotone-class argument, the formula for $J = \{1,\dots,m\}$ extends to $\mathcal{L}(\gamma_1,\dots,\gamma_m) = \bigotimes_j \mu_j$. □

Exercises

1. Show by an example that the σ-fields $\mathcal{F}_\tau$ and $\mathcal{F}_{\tau-}$ may differ. (*Hint:* Take τ to be a constant.)

2. Give examples of optional times that are predictable, accessible but not predictable, and totally inaccessible. (*Hint:* Use Corollary 10.18.)

3. Give an example that a right-continuous, adapted process that is not predictable. (*Hint:* Use Theorem 10.14.)

4. Given a Brownian motion B on $[0,1]$, let $\mathcal{F}$ be the filtration induced by $X_t = (B_t, B_1)$. Find the Doob–Meyer decomposition $B = M + A$ on $[0,1)$, and show that A has a.s. finite variation on $[0,1]$.

5. For any totally inaccessible time τ, show that $\sup_t |P(\tau \leq t+\varepsilon|\,\mathcal{F}_t) - 1\{\tau \leq t\}| \to 0$ a.s. as $\varepsilon \to 0$. Derive a corresponding result for the compensator. (*Hint:* Use Lemma 10.8.)

6. Let the process X be adapted and rcll. Show that X is predictable iff it has accessible jumps and ΔX_τ is $\mathcal{F}_{\tau-}$-measurable for every predictable time $\tau < \infty$. (*Hint:* Use Proposition 10.17 and Lemmas 10.1 and 10.3.)

7. Show that the compensator A of a ql-continuous, local sub-martingale is a.s. continuous. (*Hint:* Note that A has accessible jumps. Use optional sampling at an arbitrary predictable time $\tau < \infty$ with announcing sequence (τ_n).)

8. Show that any general inequality involving an increasing process A and its compensator $\hat{A}$ remains valid in discrete time. (*Hint:* Embed the discrete-time process and filtration into continuous time.)

9. Let ξ be a binomial process on R_+ with $E\xi = n\mu$ for a diffuse probability measure μ. Show that if ξ has points $\tau_1 < \cdots < \tau_n$, then ξ has induced compensator $\hat{\xi} = \sum_{k\leq n} 1_{[\tau_{k-1},\tau_k)}\,\mu_k$ for some non-random measures $\mu_1,\dots,\mu_n$, to be identified.

10. Let ξ be a point process on R_+ with $\|\xi\| = n$ and with induced compensator of the form $\hat{\xi} = \sum_{k\leq n} 1_{[\tau_{k-1},\tau_k)}\,\mu_k$ for some non-random measures $\mu_1,\dots,\mu_n$. Show that $\mathcal{L}(\xi)$ is uniquely determined by those measures. Under what condition is ξ a binomial process?

11. Determine the natural compensator of a renewal process ξ based on a measure μ. Under what conditions on μ has ξ accessible atoms, and when is ξ ql-continuous?

12. Let $\tau_1 < \tau_2 < \cdots$ form a renewal process based on a diffuse measure μ. Find the associated discounted compensators $\zeta_1, \zeta_2, \ldots$.

13. Let ξ be a binomial process with $\|\xi\| = n$, based on a diffuse probability measure μ on R_+. Find the natural compensator $\hat{\xi}$ of ξ. Also find the discounted compensators $\zeta_1, \ldots, \zeta_n$ of the points $\tau_1 < \cdots < \tau_n$ of ξ.

IV. Markovian and Related Structures

Here we introduce the notion of Markov processes, representing another basic dependence structure of probability theory and playing a prominent role throughout the remainder of the book. After a short discussion of the general Markov property, we establish the basic convergence and invariance properties of discrete- and continuous-time Markov chains. In the continuous-time case, we focus on the structure of pure jump-type processes, and include a brief discussion of branching processes. Chapter 12 is devoted to a detailed study of random walks and renewal processes, exploring in particular some basic recurrence properties and a two-sided version of the celebrated renewal theorem. A hurried reader might concentrate on especially the beginnings of Chapters 11 and 13, along with selected material from Chapter 12.

— — —

11. Markov properties and discrete-time chains. The beginning of this chapter, exploring the nature of the general Markov property, may be regarded as essential core material. After discussing in particular the role of transition kernels and the significance of space and time homogeneity, we prove an elementary version of the strong Markov property. We conclude with a discussion of the basic recurrence and ergodic properties of discrete-time Markov chains, including criteria for the existence of invariant distributions and convergence of transition probabilities.

12. Random walks and renewal processes. After establishing the recurrence dichotomy of random walks in R^d and giving criteria for recurrence and transience, we proceed to a study of fluctuation properties in one dimension, involving the notions of ladder times and heights and criteria for divergence to infinity. The remainder of the chapter deals with the main results of renewal theory, including the stationary version of a given renewal process, a two-sided version of the basic renewal theorem, and solutions of the renewal equation. The mentioned subjects will prepare for the related but more sophisticated theories of Lévy processes, local time, and excursion theory in later chapters.

13. Jump-type chains and branching processes. Here we begin with a detailed study of pure jump-type Markov processes, leading to a description in terms of a rate kernel, determining both the jump structure and the nature of holding times. This material is of fundamental importance and may serve as an introduction to the theory of Feller processes and their generators. After discussing the recurrence and ergodic properties of continuous-time Markov chains, we conclude with a brief introduction to branching processes, of importance for both theory and applications.

Chapter 11

Markov Properties and Discrete-Time Chains

Extended Markov property, transition kernels, finite-dimensional distributions, Chapman–Kolmogorov relation, existence, invariance and independence, random dynamical systems, iterated optional times, strong Markov property, space and time homogeneity, stationarity and invariance, strong homogeneity, occupation times, recurrence, periodicity, excursions, irreducible chains, convergence dichotomy, mean recurrence times, absorption

Markov processes are without question the most important processes of modern probability, and various aspects of their theory and applications will appear throughout the remainder of this book. They may be thought of informally as random dynamical systems, which explains their fundamental importance for the subject. They will be considered in discrete and continuous time, and in a wide variety of spaces.

To make the mentioned description precise, we may fix an arbitrary Borel space S and filtration $\mathcal{F}$. An adapted process X in S is said to be *Markov*, if for any times $s < t$ we have $X_t = f_{s,t}(X_s, \vartheta_{s,t})$ a.s. for some measurable functions $f_{s,t}$ and $U(0,1)$ random variables $\vartheta_{s,t} \perp\!\!\!\perp \mathcal{F}_s$. The stated condition is equivalent to the less transparent conditional independence $X_t \perp\!\!\!\perp_{X_s} \mathcal{F}_s$. The process is said to be *time-homogeneous* if we can choose $f_{s,t} \equiv f_{0,t-s}$, and *space-homogeneous* (in Abelian groups) if $f_{s,t}(x,\cdot) \equiv f_{s,t}(0,\cdot) + x$. The evolution is formally described in terms of a family of *transition kernels* $\mu_{s,t}(x,\cdot) = \mathcal{L}\{f_{s,t}(x,\vartheta)\}$, satisfying an a.s. version of the *Chapman–Kolmogorov relation* $\mu_{s,t}\,\mu_{t,u} = \mu_{s,u}$. In the usual axiomatic setup, we assume the latter equation to hold identically.

This chapter is devoted to the most basic parts of Markov process theory. Space homogeneity is shown to be equivalent to independence of the increments, which motivates our discussion of random walks and Lévy processes in Chapters 12 and 16. In the time-homogeneous case, we prove a primitive form of the *strong Markov property*, and show how the result simplifies when the process is also space-homogeneous. We further show how invariance of the initial distribution implies stationarity of the process, which motivates our treatment of stationary processes in Chapter 25.

The general theory of Markov processes is more advanced and will not be continued until Chapter 17, where we develop the basic theory of Feller processes. In the meantime we will consider several important sub-classes, such

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_12

as the pure jump-type processes in Chapter 13, Brownian motion and related processes in Chapters 14 and 19, and the mentioned random walks and Lévy processes in Chapters 12 and 16. A detailed study of diffusion processes appears in Chapters 32–33, and further aspects of Brownian motion are considered in Chapters 29, 34–35.

To begin our systematic study of Markov processes, consider an arbitrary time scale $T \subset \mathsf{R}$ equipped with a filtration $\mathcal{F} = (\mathcal{F}_t)$, and fix a measurable space $(S, \mathcal{S})$. An S-valued process X on T is called a *Markov process*, if it is adapted to $\mathcal{F}$ and such that

$$\mathcal{F}_t \underset{X_t}{\perp\!\!\!\perp} X_u, \quad t \le u \ \text{ in } \ T. \tag{1}$$

Just as for the martingale property, the Markov property depends on the choice of filtration, with the weakest version obtained for the filtration induced by X. The simple property (1) may be strengthened as follows.

Lemma 11.1 *(extended Markov property) The Markov property* (1) *extends to*[1]

$$\mathcal{F}_t \underset{X_t}{\perp\!\!\!\perp} \{X_u;\, u \ge t\}, \quad t \in T. \tag{2}$$

Proof: Fix any $t = t_0 \le t_1 \le \cdots$ in T. Then (1) yields $\mathcal{F}_{t_n} \perp\!\!\!\perp_{X_{t_n}} X_{t_{n+1}}$ for every $n \ge 0$, and so by Theorem 8.12

$$\mathcal{F}_t \underset{X_{t_0},\ldots,X_{t_n}}{\perp\!\!\!\perp} X_{t_{n+1}}, \quad n \ge 0.$$

By the same result, this is equivalent to

$$\mathcal{F}_t \underset{X_t}{\perp\!\!\!\perp} (X_{t_1}, X_{t_2}, \ldots),$$

and (2) follows by a monotone-class argument. $\square$

For any times $s \le t$ in T, we assume the existence of some regular conditional distributions

$$\begin{aligned} \mu_{s,t}(X_s, B) &= P(X_t \in B \mid X_s) \\ &= P(X_t \in B \mid \mathcal{F}_s) \ \text{ a.s.}, \quad B \in \mathcal{S}, \end{aligned} \tag{3}$$

where the *transition kernels* $\mu_{s,t}$ exist by Theorem 8.5 when S is Borel. We may further introduce the one-dimensional distributions $\nu_t = \mathcal{L}(X_t)$, $t \in T$. When T begins at 0, we show that the distribution of X is uniquely determined by the kernels $\mu_{s,t}$ together with the *initial distribution* ν_0.

For a precise statement, it is convenient to use the kernel operations introduced in Chapter 3. Recall that if μ, ν are kernels on S, the kernels $\mu \otimes \nu : S \to S^2$ and $\mu\nu : S \to S$ are given, for any $s \in S$ and $B \in \mathcal{S}^2$ or $\mathcal{S}$, by

[1] Using the shift operators θ_t, this becomes $\mathcal{F}_t \perp\!\!\!\perp_{X_t} \theta_t X$ for all $t \in T$. Informally, the past and future are conditionally independent, given the present.

$$\begin{aligned}(\mu \otimes \nu)(s, B) &= \int \mu(s, dt) \int \nu(t, du)\, 1_B(t, u),\\ (\mu\nu)(s, B) &= (\mu \otimes \nu)(s, S \times B)\\ &= \int \mu(s, dt)\, \nu(t, B).\end{aligned}$$

Proposition 11.2 *(finite-dimensional distributions) Let X be a Markov process on T with one-dimensional distributions ν_t and transition kernels $\mu_{s,t}$. Then for any $t_0 \leq \cdots \leq t_n$ in T,*

(i) $\mathcal{L}\Big(X_{t_0}, \ldots, X_{t_n}\Big) = \nu_{t_0} \otimes \mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}$,

(ii) $\mathcal{L}\Big(X_{t_1}, \ldots, X_{t_n}\Big| \mathcal{F}_{t_0}\Big) = \Big(\mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}\Big)(X_{t_0}, \cdot)$.

Proof: (i) This is obvious for $n = 0$. Proceeding by induction, assume the statement with n replaced by $n-1$, and fix any bounded, measurable function f on S^{n+1}. Since $X_{t_0}, \ldots, X_{t_{n-1}}$ are $\mathcal{F}_{t_{n-1}}$-measurable, we get by Theorem 8.5 and the induction hypothesis

$$\begin{aligned}Ef(X_{t_0}, \ldots, X_{t_n}) &= E\, E\Big\{ f(X_{t_0}, \ldots, X_{t_n}) \,|\, \mathcal{F}_{t_{n-1}}\Big\}\\ &= E \int f(X_{t_0}, \ldots, X_{t_{n-1}}, x)\, \mu_{t_{n-1},t_n}(X_{t_{n-1}}, dx)\\ &= \Big(\nu_{t_0} \otimes \mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}\Big) f,\end{aligned}$$

as desired. This completes the induction.

(ii) From (i) we get for any $B \in \mathcal{S}$ and $C \in \mathcal{S}^n$

$$\begin{aligned}&P\Big\{(X_{t_0}, \ldots, X_{t_n}) \in B \times C\Big\}\\ &\quad = \int_B \nu_{t_0}(dx) \Big(\mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}\Big)(x, C)\\ &\quad = E\Big\{\Big(\mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}\Big)(X_{t_0}, C);\ X_{t_0} \in B\Big\},\end{aligned}$$

and the assertion follows by Theorem 8.1 and Lemma 11.1. □

An obvious consistency requirement leads to the basic *Chapman–Kolmogorov relation* between the transition kernels. Say that two kernels μ, μ' agree a.s. if $\mu(x, \cdot) = \mu'(x, \cdot)$ for almost every x.

Corollary 11.3 *(Chapman, Smoluchovsky) For a Markov process in a Borel space S, we have*

$$\mu_{s,u} = \mu_{s,t}\, \mu_{t,u} \quad a.s.\ \nu_s, \qquad s \leq t \leq u.$$

Proof: By Proposition 11.2, we have a.s. for any $B \in \mathcal{S}$

$$\begin{aligned}\mu_{s,u}(X_s, B) &= P\Big\{X_u \in B \,\Big|\, \mathcal{F}_s\Big\}\\ &= P\Big\{(X_t, X_u) \in S \times B \,\Big|\, \mathcal{F}_s\Big\}\\ &= (\mu_{s,t} \otimes \mu_{t,u})(X_s,\ S \times B)\\ &= (\mu_{s,t}\, \mu_{t,u})(X_s, B).\end{aligned}$$

Since S is Borel, we may choose a common null set for all B. □

We will henceforth require the stated relation to hold *identically*, so that

$$\mu_{s,u} = \mu_{s,t}\, \mu_{t,u}, \quad s \le t \le u. \tag{4}$$

Thus, we define a Markov process by condition (3), in terms of some transition kernels $\mu_{s,t}$ satisfying (4). In the *discrete-time* case of $T = \mathsf{Z}_+$, the latter relation imposes no restriction, since we may then start from any versions of the kernels $\mu_n = \mu_{n-1,n}$, and define $\mu_{m,n} = \mu_{m+1} \cdots \mu_n$ for arbitrary $m < n$.

Given such a family of transition kernels $\mu_{s,t}$ and an arbitrary initial distribution ν, we show that an associated Markov process exists.

Theorem 11.4 *(existence, Kolmogorov) Consider a time scale T starting at 0, a Borel space $(S, \mathcal{S})$, a probability measure ν on S, and a family of probability kernels $\mu_{s,t}$ on S, $s \le t$ in T, satisfying (4). Then there exists an S-valued Markov process X on T with initial distribution ν and transition kernels $\mu_{s,t}$.*

Proof: Consider the probability measures

$$\nu_{t_1,\dots,t_n} = \nu \mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}, \quad 0 = t_0 \le t_1 \le \cdots \le t_n,\ n \in \mathsf{N}.$$

To see that the family $(\nu_{t_0,\dots,t_n})$ is projective, let $B \in \mathcal{S}^{n-1}$ be arbitrary, and define for any $k \in \{1, \dots, n\}$ the set

$$B_k = \Big\{(x_1, \dots, x_n) \in S^n;\ (x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \in B\Big\}.$$

Then (4) yields

$$\begin{aligned} \nu_{t_1,\dots,t_n} B_k &= \Big(\nu \mu_{t_0,t_1} \otimes \cdots \otimes \mu_{t_{k-1},t_{k+1}} \otimes \cdots \otimes \mu_{t_{n-1},t_n}\Big) B \\ &= \nu_{t_1,\dots,t_{k-1},t_{k+1},\dots,t_n} B, \end{aligned}$$

as desired. By Theorem 8.23 there exists an S-valued process X on T with

$$\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = \nu_{t_1,\dots,t_n}, \quad t_1 \le \cdots \le t_n,\ n \in \mathsf{N}, \tag{5}$$

and in particular $\mathcal{L}(X_0) = \nu_0 = \nu$.

To see that X is Markov with transition kernels $\mu_{s,t}$, fix any times $s_1 \le \cdots \le s_n = s \le t$ and sets $B \in \mathcal{S}^n$ and $C \in \mathcal{S}$, and conclude from (5) that

$$\begin{aligned} &P\Big\{(X_{s_1}, \dots, X_{s_n}, X_t) \in B \times C\Big\} \\ &\quad = \nu_{s_1,\dots,s_n,t}(B \times C) \\ &\quad = E\Big\{\mu_{s,t}(X_s, C);\ (X_{s_1}, \dots, X_{s_n}) \in B\Big\}. \end{aligned}$$

Letting $\mathcal{F}$ be the filtration induced by X, we get by a monotone-class argument

$$P\Big\{X_t \in C; A\Big\} = E\Big\{\mu_{s,t}(X_s, C); A\Big\}, \quad A \in \mathcal{F}_s,$$

and so $P\{X_t \in C \,|\, \mathcal{F}_s\} = \mu_{s,t}(X_s, C)$ a.s. $\square$

Now let $(G, \mathcal{G})$ be a measurable group with identity element ι. Recall that a kernel μ on G is said to be *invariant* if $\mu_r = \mu_\iota \circ \theta_r^{-1} \equiv \theta_r \mu_\iota$, meaning that

$$\mu(r, B) = \mu(\iota, r^{-1}B), \quad r \in G, \; B \in \mathcal{G}.$$

An G-valued Markov process is said to be *space-homogeneous* if its transition kernels $\mu_{s,t}$ are invariant. We further say that a process X in G has *independent increments*, if for any times $t_0 \leq \cdots \leq t_n$ the increments $X_{t_{k-1}}^{-1} X_{t_k}$ are mutually independent and independent of X_0. More generally, given a filtration $\mathcal{F}$ on T, we say that X has *$\mathcal{F}$-independent increments*, if X is adapted to $\mathcal{F}$ and such that $X_s^{-1} X_t \perp\!\!\!\perp \mathcal{F}_s$ for all $s \leq t$ in T. Note that the elementary notion of independence corresponds to the case where $\mathcal{F}$ is induced by X. For Abelian groups G, we may use an additive notation and write the increments in the usual way as $X_t - X_s$.

Theorem 11.5 *(space homogeneity and independence) Let X be an $\mathcal{F}$-adapted process on T with values in a measurable group G. Then these conditions are equivalent:*

(i) *X is a G-homogeneous $\mathcal{F}$-Markov process,*

(ii) *X has $\mathcal{F}$-independent increments.*

In that case, X has transition kernels

$$\mu_{s,t}(r, B) = P\Big\{X_s^{-1} X_t \in r^{-1}B\Big\}, \quad r \in G, \; B \in \mathcal{G}, \; s \leq t \text{ in } T. \tag{6}$$

Proof, (i) $\Rightarrow$ (ii): Let X be $\mathcal{F}$-Markov with transition kernels

$$\mu_{s,t}(r, B) = \mu_{s,t}(r^{-1}B), \quad r \in G, \; B \in \mathcal{G}, \; s \leq t \text{ in } T. \tag{7}$$

Then Theorem 8.5 yields for any $s \leq t$ in T and $B \in \mathcal{G}$

$$\begin{aligned} P\Big\{X_s^{-1} X_t \in B \,\Big|\, \mathcal{F}_s\Big\} &= P\Big\{X_t \in X_s B \,\Big|\, \mathcal{F}_s\Big\} \\ &= \mu_{s,t}(X_s, X_s B) \\ &= \mu_{s,t} B. \end{aligned}$$

Thus, $X_s^{-1} X_t$ is independent of $\mathcal{F}_s$ with distribution $\mu_{s,t}$, and (6) follows by means of (7).

(ii) $\Rightarrow$ (i): Let $X_s^{-1} X_t$ be independent of $\mathcal{F}_s$ with distribution $\mu_{s,t}$. Defining the associated kernel $\mu_{s,t}$ by (7), we get by Theorem 8.5 for any s, t, and B as before

$$\begin{aligned} P\Big\{X_t \in B \,\Big|\, \mathcal{F}_s\Big\} &= P\Big\{X_s^{-1} X_t \in X_s^{-1} B \,\Big|\, \mathcal{F}_s\Big\} \\ &= \mu_{s,t}(X_s^{-1} B) \\ &= \mu_{s,t}(X_s, B), \end{aligned}$$

and so X is Markov with the invariant transition kernels in (7). $\square$

We now specialize to the *time-homogeneous* case, where $T = \mathsf{R}_+$ or Z_+, and the transition kernels are of the form $\mu_{s,t} = \mu_{t-s}$, so that

$$P\Big\{X_t \in B \,\Big|\, \mathcal{F}_s\Big\} = \mu_{t-s}(X_s, B) \ \text{ a.s.}, \quad B \in \mathcal{S}, \ \ s \le t \ \text{ in } T.$$

Introducing the initial distribution $\nu = \mathcal{L}(X_0)$, we get by Proposition 11.2

$$\begin{aligned}\mathcal{L}(X_{t_0}, \ldots, X_{t_n}) &= \nu\mu_{t_0} \otimes \mu_{t_1-t_0} \otimes \cdots \otimes \mu_{t_n-t_{n-1}},\\ \mathcal{L}\Big(X_{t_1}, \ldots, X_{t_n}\Big|\, \mathcal{F}_{t_0}\Big) &= \Big(\mu_{t_1-t_0} \otimes \cdots \otimes \mu_{t_n-t_{n-1}}\Big)(X_{t_0}, \cdot).\end{aligned}$$

The Chapman–Kolmogorov relation now reduces to the *semi-group property*

$$\mu_{s+t} = \mu_s\, \mu_t, \quad s, t \in T,$$

which is again assumed to hold identically.

The following result suggests that we think of a discrete-time Markov process as a stochastic dynamical system. Let ϑ be a generic $U(0,1)$ random variable.

Proposition 11.6 *(random dynamical system) Let X be a process on Z_+ with values in a Borel space S. Then these conditions are equivalent:*

(i) *X is Markov with transition kernels μ_n on S,*

(ii) *there exist some measurable functions $f_1, f_2, \ldots \colon S \times [0,1] \to S$ and i.i.d. $U(0,1)$ random variables $\vartheta_1, \vartheta_2, \ldots \perp\!\!\!\perp X_0$, such that*

$$X_n = f_n(X_{n-1}, \vartheta_n) \ \text{ a.s.}, \quad n \in \mathsf{N}.$$

In that case $\mu_n(s, \cdot) = \mathcal{L}\{f_n(s, \vartheta)\}$ a.s. $\mathcal{L}(X_{n-1})$, and we may choose $f_n \equiv f$ iff X is time-homogeneous.

Proof, (ii)$\Rightarrow$(i): Assume (ii), and define $\mu_n(s, \cdot) = \mathcal{L}\{f_n(s, \vartheta)\}$. Writing $\mathcal{F}$ for the filtration induced by X, we get by Theorem 8.5 for any $B \in \mathcal{S}$

$$\begin{aligned}P\Big\{X_n \in B \,\Big|\, \mathcal{F}_{n-1}\Big\} &= P\Big\{f_n(X_{n-1}, \vartheta_n) \in B \,\Big|\, \mathcal{F}_{n-1}\Big\}\\ &= \lambda\Big\{t;\, f_n(X_{n-1}, t) \in B\Big\}\\ &= \mu_n(X_{n-1}, B),\end{aligned}$$

proving (i).

(i)$\Rightarrow$(ii): Assume (i). By Lemma 4.22 we may choose the associated functions f_n as above. Let $\tilde\vartheta_1, \tilde\vartheta_2, \ldots$ be i.i.d. $U(0,1)$ and independent of $\tilde X_0 \overset{d}{=} X_0$, and define recursively $\tilde X_n = f_n(\tilde X_{n-1}, \tilde\vartheta_n)$ for $n \in \mathsf{N}$. Then $\tilde X$ is Markov as before with transition kernels μ_n. Hence, $\tilde X \overset{d}{=} X$ by Proposition 11.2, and so Theorem 8.17 yields some random variables ϑ_n with $\{X, (\vartheta_n)\} \overset{d}{=} \{\tilde X, (\tilde\vartheta_n)\}$. Now (ii) follows, since the diagonal in S^2 is measurable. The last assertion is clear by construction. $\square$

Now fix a transition semi-group (μ_t) on a Borel space S. For any probability measure ν on S, Theorem 11.4 yields an associated Markov process X_ν, and Proposition 4.2 shows that the associated distribution P_ν is uniquely determined by ν. Note that P_ν is a probability measure on the path space $(S^T, \mathcal{S}^T)$. For $\nu = \delta_x$, we write P_x instead of P_{δ_x}. Integration with respect to P_ν or P_x is denoted by E_ν or E_x, respectively.

Lemma 11.7 *(mixtures) The measures P_x form a probability kernel from S to S^T, and for any initial distribution ν,*

$$P_\nu A = \int_S P_x(A)\,\nu(dx), \quad A \in \mathcal{S}^T.$$

Proof: The measurability of $P_x A$ and the stated formula are obvious for cylinder sets of the form $A = (\pi_{t_1}, \dots, \pi_{t_n})^{-1}B$. The general case follows easily by a monotone-class argument. □

Rather than considering one Markov process X_ν for every initial distribution ν, we may introduce a single *canonical* process X, defined as the identity mapping on the path space $(S^T, \mathcal{S}^T)$, equipped with the different probability measures P_ν. Then X_t agrees with the evaluation map $\pi_t \colon x \mapsto x_t$ on S^T, which is measurable by the definition of $\mathcal{S}^T$. For our present purposes, it suffices to endow the path space S^T with the *canonical filtration* $\mathcal{F}$ induced by X. On S^T we further introduce the *shift operators* $\theta_t \colon S^T \to S^T$, $t \in T$, given by

$$(\theta_t x)_s = x_{s+t}, \quad s, t \in T,\ x \in S^T,$$

which are clearly measurable with respect to $\mathcal{S}^T$. In the canonical case we note that $\theta_t X = \theta_t = X \circ \theta_t$.

Optional times with respect to a Markov process are often constructed recursively in terms of shifts on the underlying path space. Thus, for any pair of optional times σ and τ on the canonical space, we may consider the random time $\gamma = \sigma + \tau \circ \theta_\sigma$, which is understood to be infinite when $\sigma = \infty$. Under weak restrictions on space and filtration, we show that γ is again optional. Here C_S and D_S denote the spaces of continuous or rcll functions, respectively, from R_+ to S.

Proposition 11.8 *(iterated optional times) For a metric space S, let σ, τ be optional times on the canonical space S^∞, C_S, or D_S, endowed with the right-continuous, induced filtration. Then we may form an optional time*

$$\gamma = \sigma + \tau \circ \theta_\sigma.$$

Proof: Since $\sigma \wedge n + \tau \circ \theta_{\sigma \wedge n} \uparrow \gamma$, Lemma 9.3 justifies taking σ to be bounded. Let X be the canonical process with induced filtration $\mathcal{F}$. Since X is $\mathcal{F}^+$-progressive, $X_{\sigma+s} = X_s \circ \theta_\sigma$ is $\mathcal{F}^+_{\sigma+s}$-measurable for every $s \geq 0$ by

Lemma 9.5. Hence, for fixed $t \geq 0$, all sets $A = \{X_s \in B\}$ with $s \leq t$ and $B \in \mathcal{S}$ satisfy $\theta_\sigma^{-1} A \in \mathcal{F}^+_{\sigma+t}$. Since the latter sets form a σ-field, we get

$$\theta_\sigma^{-1} \mathcal{F}_t \subset \mathcal{F}^+_{\sigma+t}, \quad t \geq 0. \tag{8}$$

Now fix any $t \geq 0$, and note that

$$\{\gamma < t\} = \bigcup_{r \in \mathbb{Q} \cap (0,t)} \Big\{\sigma < r,\ \tau \circ \theta_\sigma < t - r\Big\}. \tag{9}$$

For every $r \in (0, t)$ we have $\{\tau < t - r\} \in \mathcal{F}_{t-r}$, and so $\theta_\sigma^{-1}\{\tau < t - r\} \in \mathcal{F}^+_{\sigma+t-r}$ by (8), and Lemma 9.2 gives

$$\Big\{\sigma < r,\ \tau \circ \theta_\sigma < t - r\Big\} = \Big\{\sigma + t - r < t\Big\} \cap \theta_\sigma^{-1}\Big\{\tau < t - r\Big\} \in \mathcal{F}_t.$$

Thus, (9) yields $\{\gamma < t\} \in \mathcal{F}_t$, and so γ is $\mathcal{F}^+$-optional by Lemma 9.2. □

We may now extend the Markov property to suitable optional times. The present statement is only preliminary, and stronger versions will be established, under appropriate conditions, in Theorems 13.1, 14.11, and 17.17.

Proposition 11.9 *(strong Markov property) Let X be a time-homogeneous Markov process on $T = \mathbb{R}_+$ or $\mathbb{Z}_+$, and let τ be an optional time taking countably many values. Then in the general and canonical cases, we have respectively*

(i) $P\Big\{\theta_\tau X \in A \Big| \mathcal{F}_\tau\Big\} = P_{X_\tau} A$ *a.s. on* $\{\tau < \infty\}$,

(ii) $E_\nu(\xi \circ \theta_\tau | \mathcal{F}_\tau) = E_{X_\tau}\xi$, P_ν *-a.s. on* $\{\tau < \infty\}$,

for any set $A \in \mathcal{S}^T$, distribution ν on S, and bounded or non-negative random variable ξ.

Since $\{\tau < \infty\} \in \mathcal{F}_\tau$, formulas (i)−(ii) make sense by Lemma 8.3, even though $\theta_\tau X$ and P_{X_τ} are only defined for $\tau < \infty$.

Proof: By Lemmas 8.3 and 9.1, we may take $\tau = t$ to be finite and non-random. For sets of the form

$$A = (\pi_{t_1}, \ldots, \pi_{t_n})^{-1} B, \quad t_1 \leq \cdots \leq t_n,\ B \in \mathcal{S}^n,\ n \in \mathbb{N}, \tag{10}$$

Proposition 11.2 yields

$$\begin{aligned} P\Big\{\theta_t X \in A \Big| \mathcal{F}_t\Big\} &= P\Big\{(X_{t+t_1}, \ldots, X_{t+t_n}) \in B \Big| \mathcal{F}_t\Big\} \\ &= \Big(\mu_{t_1} \otimes \mu_{t_2 - t_1} \otimes \cdots \otimes \mu_{t_n - t_{n-1}}\Big)(X_t, B) = P_{X_t} A, \end{aligned}$$

which extends to arbitrary $A \in \mathcal{S}^T$ by a monotone-class argument.

For canonical X, (i) is equivalent to (ii) with $\xi = 1_A$, since in that case $\xi \circ \theta_\tau = 1\{\theta_\tau X \in A\}$. The result extends to general ξ by linearity and monotone convergence. □

When X is both space- and time-homogeneous, we can state the strong Markov property without reference to the family (P_x). For notational clarity, we consider only processes in Abelian groups, the general case being similar.

Theorem 11.10 *(space and time homogeneity) Let X be a space- and time-homogeneous Markov process in a measurable Abelian group S. Then*

(i) $P_xA = P_0(A - x), \quad x \in S,\ A \in \mathcal{S}^T,$

(ii) *the strong Markov property holds at an optional time $\tau < \infty$, iff X_τ is a.s. $\mathcal{F}_\tau$-measurable and*

$$X - X_0 \overset{d}{=} \theta_\tau X - X_\tau \perp\!\!\!\perp \mathcal{F}_\tau.$$

Proof: (i) By Proposition 11.2, we get for sets A as in (10)

$$\begin{aligned} P_xA &= P_x \circ (\pi_{t_1}, \dots, \pi_{t_n})^{-1}B \\ &= \Big(\mu_{t_1} \otimes \mu_{t_2 - t_1} \otimes \cdots \otimes \mu_{t_n - t_{n-1}}\Big)(x, B) \\ &= \Big(\mu_{t_1} \otimes \mu_{t_2 - t_1} \otimes \cdots \otimes \mu_{t_n - t_{n-1}}\Big)(0, B - x) \\ &= P_0 \circ (\pi_{t_1}, \dots, \pi_{t_n})^{-1}(B - x) \\ &= P_0(A - x), \end{aligned}$$

which extends to general A by a monotone-class argument.

(ii) Assume the strong Markov property at τ. Letting $A = \pi_0^{-1}B$ with $B \in \mathcal{S}$, we get

$$\begin{aligned} 1_B(X_\tau) &= P_{X_\tau}\{\pi_0 \in B\} \\ &= P\Big\{X_\tau \in B \,\Big|\, \mathcal{F}_\tau\Big\} \quad \text{a.s.,} \end{aligned}$$

and so X_τ is a.s. $\mathcal{F}_\tau$-measurable. By (i) and Theorem 8.5,

$$P\Big\{\theta_\tau X - X_\tau \in A \,\Big|\, \mathcal{F}_\tau\Big\} = P_{X_\tau}(A + X_\tau) = P_0A, \quad A \in \mathcal{S}^T, \tag{11}$$

which shows that $\theta_\tau X - X_\tau \perp\!\!\!\perp \mathcal{F}_\tau$ with distribution P_0. Taking $\tau = 0$, we get in particular $\mathcal{L}(X - X_0) = P_0$, and the asserted properties follow.

Now assume the stated conditions. To deduce the strong Markov property at τ, let $A \in \mathcal{S}^T$ be arbitrary, and conclude from (i) and Theorem 8.5 that

$$\begin{aligned} P\Big\{\theta_\tau X \in A \,\Big|\, \mathcal{F}_\tau\Big\} &= P\Big\{\theta_\tau X - X_\tau \in A - X_\tau \,\Big|\, \mathcal{F}_\tau\Big\} \\ &= P_0(A - X_\tau) \\ &= P_{X_\tau}A. \end{aligned}$$

□

If a time-homogeneous Markov process X has initial distribution ν, the distribution at time $t \in T$ equals $\nu_t = \nu\mu_t$, or

$$\nu_t B = \int \nu(dx)\, \mu_t(x, B), \quad B \in \mathcal{S},\ t \in T.$$

A distribution ν is said to be *invariant* for the semi-group (μ_t) if ν_t is independent of t, so that $\nu\mu_t = \nu$ for all $t \in T$. We also say that a process X on T is *stationary*, if $\theta_t X \overset{d}{=} X$ for all $t \in T$. The two notions[2] are related as follows.

[2]Stationarity and invariance are often confused. Here the distinction is of course crucial.

Lemma 11.11 *(stationarity and invariance) Let X be a time-homogeneous Markov process on T with transition kernels μ_t and initial distribution ν. Then*

$$X \text{ is stationary} \quad \Leftrightarrow \quad \nu \text{ is invariant for } (\mu_t).$$

Proof: If ν is invariant, then Proposition 11.2 yields

$$(X_{t+t_1}, \dots, X_{t+t_n}) \overset{d}{=} (X_{t_1}, \dots, X_{t_n}), \quad t,\ t_1 \le \dots \le t_n \text{ in } T,$$

and so X is stationary by Proposition 4.2. The converse is obvious. □

The strong Markov property has many applications. We begin with a classical maximum inequality for random walks, which will be useful in Chapter 23.

Proposition 11.12 *(maximum inequality, Ottaviani) Let $\xi_1, \xi_2, \dots$ be i.i.d. random variables with $E\xi_i = 0$ and $E\xi_i^2 = 1$, and put $X_n = \sum_{i \le n} \xi_i$. Then*

$$P\Big\{X_n^* \ge 2\, r\sqrt{n}\Big\} \le \frac{P\Big\{|X_n| \ge r\sqrt{n}\Big\}}{1 - r^{-2}}, \quad r > 1,\ n \in \mathsf{N}.$$

Proof: Put $c = r\sqrt{n}$, and define $\tau = \inf\{k \in \mathsf{N};\ |X_k| \ge 2c\}$. By the strong Markov property at τ and Theorem 8.5,

$$\begin{aligned} P\{|X_n| \ge c\} &\ge P\Big\{|X_n| \ge c,\ X_n^* \ge 2c\Big\} \\ &\ge P\Big\{\tau \le n,\ |X_n - X_\tau| \le c\Big\} \\ &\ge P\Big\{X_n^* \ge 2\, c\Big\} \min_{k \le n} P\Big\{|X_k| \le c\Big\}, \end{aligned}$$

and Chebyshev's inequality yields

$$\begin{aligned} \min_{k \le n} P\Big\{|X_k| \le c\Big\} &\ge \min_{k \le n} \Big(1 - k\, c^{-2}\Big) \\ &\ge \Big(1 - n\, c^{-2}\Big) \\ &= 1 - r^{-2}. \end{aligned}$$

□

Similar ideas can be used to prove the following more sophisticated result, which will find important applications in Chapters 16 and 27.

Corollary 11.13 *(uniform convergence) For $r \in \mathsf{Q}_+$, let X^r be rcll processes in R^d, such that the increments in r are independent and satisfy $(X^r - X^s)_t^* \xrightarrow{P} 0$ as $r, s \to \infty$ for fixed $t > 0$. Then there exists an rcll process X in R^d, such that*

$$(X^r - X)_t^* \to 0 \ \ a.s., \quad t > 0.$$

Proof: Fixing a $t > 0$, we can choose a sequence $r_n \to \infty$ in Q_+, such that

$$(X^{r_m} - X^{r_n})_t^* \to 0 \ \text{ a.s.}, \quad m, n \to \infty.$$

By completeness there exists a process X in R^d with $(X^{r_n} - X)_t^* \to 0$ a.s., and we note that X is again a.s. rcll. Now define $Y^r = X^r - X$, and note that

$(Y^r)^*_t \xrightarrow{P} 0$. Fixing any $\varepsilon, r > 0$ and a finite set $A \subset [r, \infty) \cap \mathsf{Q}$, and putting $\sigma = \inf\{s \in A;\, (Y^s)^*_t > 2\varepsilon\}$, we get as in Proposition 11.12

$$\begin{aligned} P\{(Y^r)^*_t > \varepsilon\} &\geq P\Big\{(Y^r)^*_t > \varepsilon,\, \max_{s\in A}(Y^s)^*_t > 2\varepsilon\Big\} \\ &\geq P\Big\{\sigma < \infty,\, (Y^r - Y^s)^*_t \leq \varepsilon\Big\} \\ &\geq P\Big\{\max_{s\in A}(Y^s)^*_t > 2\varepsilon\Big\} \min_{s\in A} P\Big\{(Y^r - Y^s)^*_t \leq \varepsilon\Big\}, \end{aligned}$$

which extends immediately to $A = [r, \infty) \cap \mathsf{Q}$. Solving for the first factor on the right gives

$$P\Big\{\sup_{s\geq r}(Y^s)^*_t > 2\varepsilon\Big\} \leq \frac{P\{(Y^r)^*_t > \varepsilon\}}{1 - \sup_{s\geq r} P\Big\{(Y^r - Y^s)^*_t > \varepsilon\Big\}},$$

and so $\sup_{s\geq r}(Y^s)^*_t \xrightarrow{P} 0$ which implies $(X^r - X)^*_t = (Y^r)^*_t \to 0$ a.s. The assertion now follows since $t > 0$ was arbitrary. □

We proceed to clarify the nature of the strong Markov property of a process X at a finite optional time τ. The condition is clearly a combination of the properties

$$\mathcal{F}_\tau \underset{X_\tau}{\perp\!\!\!\perp} \theta_\tau X, \qquad \mathcal{L}(\,\theta_\tau X \mid X_\tau) = P_{X_\tau} \ \text{ a.s.}$$

Though the latter relation—here referred to as *strong homogeneity*—appears to be weaker than the strong Markov property, the two conditions are in fact essentially equivalent. Here we fix any right-continuous filtration $\mathcal{F}$ on R_+.

Theorem 11.14 *(strong homogeneity) Let X be an $\mathcal{F}$-adapted, rcll process in a separable metric space (S, ρ), and consider a probability kernel $(P_x)\colon S \to D_S$. Then these conditions are equivalent:*

(i) *X is strongly homogeneous at every bounded, optional time τ,*

(ii) *X satisfies the strong Markov property at every optional time $\tau < \infty$.*

Our proof is based on a 0−1 law for absorption, involving the sets

$$I = \Big\{x \in D;\, x_t \equiv x_0\Big\}, \qquad A = \Big\{s \in S;\, P_s I = 1\Big\}. \tag{12}$$

Lemma 11.15 *(absorption) For any X as in Theorem* 11.14 (i) *and optional times $\tau < \infty$, we have*

$$P_{X_\tau} I = 1_I(\theta_\tau X) = 1_A(X_\tau) \ \ a.s.$$

Proof: We may clearly take τ to be bounded, say by $n \in \mathsf{N}$. Fix any $h > 0$, and partition S into disjoint Borel sets $B_1, B_2, \ldots$ of diameter $< h$. For each $k \in \mathsf{N}$, define

$$\tau_k = n \wedge \inf\Big\{t > \tau;\, \rho(X_\tau, X_t) > h\Big\} \ \text{ on } \ \{X_\tau \in B_k\}, \tag{13}$$

and put $\tau_k = \tau$ otherwise. The times τ_k are again bounded and optional, and clearly

$$\{X_{\tau_k} \in B_k\} \subset \Big\{X_\tau \in B_k,\, \sup_{t\in[\tau,n]} \rho(X_\tau, X_t) \leq h\Big\}. \tag{14}$$

Using property (i) and (14), we get as $n \to \infty$ and $h \to 0$

$$\begin{aligned}
E\Big(P_{X_\tau} I^c;\, \theta_\tau X \in I\Big) &= \textstyle\sum_k E\Big(P_{X_\tau} I^c;\, \theta_\tau X \in I,\, X_\tau \in B_k\Big)\\
&\le \textstyle\sum_k E\Big(P_{X_{\tau_k}} I^c;\, X_{\tau_k} \in B_k\Big)\\
&= \textstyle\sum_k P\Big\{\theta_{\tau_k} X \notin I,\, X_{\tau_k} \in B_k\Big\}\\
&\le \textstyle\sum_k P\Big\{\theta_\tau X \notin I,\, X_\tau \in B_k,\, \sup_{t\in[\tau,n]} \rho(X_\tau, X_t) \le h\Big\}\\
&\to P\Big\{\theta_\tau X \notin I,\, \sup_{t\ge\tau} \rho(X_\tau, X_t) = 0\Big\} = 0,
\end{aligned}$$

and so $P_{X_\tau} I = 1$ a.s. on $\{\theta_\tau X \in I\}$. Since also $E P_{X_\tau} I = P\{\theta_\tau X \in I\}$ by (i), we obtain the first of the displayed equalities. The second one follows by the definition of A. □

Proof of Theorem 11.14: Assume (i), and define I and A as in (12). To prove (ii) on $\{X_\tau \in A\}$, fix any times $t_1 < \cdots < t_n$ and Borel sets $B_1, \ldots, B_n$, write $B = \bigcap_k B_k$, and conclude from (i) and Lemma 11.15 that

$$\begin{aligned}
P\Big(\textstyle\bigcap_k \{X_{\tau+t_k} \in B_k\} \,\Big|\, \mathcal{F}_\tau\Big) &= P\{X_\tau \in B \,|\, \mathcal{F}_\tau\}\\
&= 1\{X_\tau \in B\}\\
&= P\{X_\tau \in B \,|\, X_\tau\}\\
&= P_{X_\tau}\{x_0 \in B\}\\
&= P_{X_\tau} \textstyle\bigcap_k \{x_{t_k} \in B_k\},
\end{aligned}$$

which extends to (ii) by a monotone-class argument.

To prove (ii) on $\{X_\tau \notin A\}$, we may take $\tau \le n$ a.s., and divide A^c into disjoint Borel sets B_k of diameter $< h$. Fix any $F \in \mathcal{F}_\tau$ with $F \subset \{X_\tau \notin A\}$. For each $k \in \mathsf{N}$, define τ_k as in (13) on the set $F^c \cap \{X_\tau \in B_k\}$, and let $\tau_k = \tau$ otherwise. Note that (14) remains true on F^c. Using (i), (14), and Lemma 11.15, we get as $n \to \infty$ and $h \to 0$

$$\begin{aligned}
&\Big|\,\mathcal{L}(\theta_\tau X\,; F) - E(P_{X_\tau}; F)\Big|\\
&\quad= \Big|\textstyle\sum_k E\Big(1\{\theta_\tau X \in \cdot\} - P_{X_\tau};\, X_\tau \in B_k,\, F\Big)\Big|\\
&\quad= \Big|\textstyle\sum_k E\Big(1\{\theta_{\tau_k} X \in \cdot\} - P_{X_{\tau_k}};\, X_{\tau_k} \in B_k,\, F\Big)\Big|\\
&\quad= \Big|\textstyle\sum_k E\Big(1\{\theta_{\tau_k} X \in \cdot\} - P_{X_{\tau_k}};\, X_{\tau_k} \in B_k,\, F^c\Big)\Big|\\
&\quad\le \textstyle\sum_k P\Big\{X_{\tau_k} \in B_k;\, F^c\Big\}\\
&\quad\le \textstyle\sum_k P\Big\{X_\tau \in B_k,\, \sup_{t\in[\tau,n]} \rho(X_\tau, X_t) \le h\Big\}\\
&\quad\to P\Big\{X_\tau \notin A,\, \sup_{t\ge\tau} \rho(X_\tau, X_t) = 0\Big\} = 0,
\end{aligned}$$

which shows that the left-hand side is zero. □

We conclude with some asymptotic and invariance properties for discrete-time Markov chains. First we consider the successive visits to a fixed state

$y \in S$. Assuming the process to be canonical, we introduce the hitting time $\tau_y = \inf\{n \in \mathsf{N};\, X_n = y\}$, and define recursively

$$\tau_y^{k+1} = \tau_y^k + \tau_y \circ \theta_{\tau_y^k}, \quad k \in \mathsf{Z}_+,$$

starting with $\tau_y^0 = 0$. We further introduce the *occupation times*

$$\begin{aligned} \kappa_y &= \sup\left\{k;\, \tau_y^k < \infty\right\} \\ &= \sum_{n \ge 1} 1\{X_n = y\}, \quad y \in S, \end{aligned}$$

and show how the distribution of κ_y can be expressed in terms of the hitting probabilities

$$\begin{aligned} r_{xy} &= P_x\{\tau_y < \infty\} \\ &= P_x\{\kappa_y > 0\}, \quad x, y \in S. \end{aligned}$$

Lemma 11.16 *(occupation times) For any $x, y \in S$ and $k \in \mathsf{N}$,*

(i) $P_x\{\kappa_y \ge k\} = P_x\{\tau_y^k < \infty\} = r_{xy}\, r_{yy}^{k-1}$,

(ii) $E_x \kappa_y = \dfrac{r_{xy}}{1 - r_{yy}}$.

Proof: (i) By the strong Markov property, we have for any $k \in \mathsf{N}$

$$\begin{aligned} P_x\left\{\tau_y^{k+1} < \infty\right\} &= P_x\left\{\tau_y^k < \infty,\ \tau_y \circ \theta_{\tau_y^k} < \infty\right\} \\ &= P_x\{\tau_y^k < \infty\}\, P_y\{\tau_y < \infty\} \\ &= r_{yy}\, P_x\{\tau_y^k < \infty\}, \end{aligned}$$

and the second relation follows by induction on k. The first relation holds since $\kappa_y \ge k$ iff $\tau_y^k < \infty$.

(ii) By (i) and Lemma 4.4, we have

$$\begin{aligned} E_x \kappa_y &= \sum_{k \ge 1} P_x\{\kappa_y \ge k\} \\ &= \sum_{k \ge 1} r_{xy}\, r_{yy}^{k-1} = \frac{r_{xy}}{1 - r_{yy}}. \end{aligned}$$

□

In particular, we get for $x = y$

$$P_x\{\kappa_x \ge k\} = P_x\{\tau_x^k < \infty\} = r_{xx}^k, \quad k \in \mathsf{N}.$$

Thus, under P_x, the number of visits to x is either a.s. infinite or geometrically distributed with mean $E_x \kappa_x + 1 = (1 - r_{xx})^{-1} < \infty$. This leads to a corresponding division of S into *recurrent* and *transient* states.

Recurrence can often be deduced from the existence of an invariant distribution. Here and below we write $p_{xy}^n = \mu_n(x, \{y\})$.

Lemma 11.17 *(invariant distributions and recurrence) Let ν be an invariant distribution on S. Then for any $x \in S$,*

$$\nu\{x\} > 0 \quad \Rightarrow \quad x \text{ is recurrent.}$$

Proof: By the invariance of ν,

$$0 < \nu\{x\} = \int \nu(dy)\, p_{yx}^n, \quad n \in \mathsf{N}. \tag{15}$$

Thus, by Lemma 11.16 and Fubini's theorem,

$$\begin{aligned}\infty &= \sum_{n\geq 1} \int \nu(dy)\, p_{yx}^n \\ &= \int \nu(dy) \sum_{n\geq 1} p_{yx}^n \\ &= \int \nu(dy) \frac{r_{yx}}{1-r_{xx}} \leq \frac{1}{1-r_{xx}}.\end{aligned}$$

Hence, $r_{xx} = 1$, and so x is recurrent. □

The *period* d_x of a state x is defined as the greatest common divisor of the set $\{n \in \mathsf{N};\, p_{xx}^n > 0\}$, and we say that x is *aperiodic* if $d_x = 1$.

Lemma 11.18 *(periodicity) If a state $x \in S$ has period $d < \infty$, we have $p_{xx}^{nd} > 0$ for all but finitely many n.*

Proof: Put $C = \{n \in \mathsf{N};\, p_{xx}^{nd} > 0\}$, and conclude from the Chapman–Kolmogorov relation that C is closed under addition. Since C has greatest common divisor 1, the generated additive group equals Z. In particular, there exist some $n_1, \ldots, n_k \in C$ and $z_1, \ldots, z_k \in \mathsf{Z}$ with $\sum_j z_j\, n_j = 1$. Writing $m = n_1 \sum_j |z_j|\, n_j$, we note that any number $n \geq m$ can be represented, for suitable $h \in \mathsf{Z}_+$ and $r \in \{0, \ldots, n_1 - 1\}$, as

$$\begin{aligned}n &= m + h\, n_1 + r \\ &= h\, n_1 + \sum_{j\leq k} \Big(n_1|z_j| + r z_j\Big)\, n_j \in C.\end{aligned}$$

□

For every $x \in S$, we define the *excursions* of X from x by

$$Y_n = X^{\tau_x} \circ \theta_{\tau_x^n}, \quad n \in \mathsf{Z}_+,$$

as long as $\tau_x^n < \infty$. To allow infinite excursions, we may introduce an extraneous element $\Delta \notin S$, and define $Y_n = \bar{\Delta} \equiv (\Delta, \Delta, \ldots)$ whenever $\tau_x^n = \infty$. Conversely, X may be recovered from the Y_n through the formulas

$$\tau_n = \sum_{k<n} \inf\Big\{t > 0;\, Y_k(t) = x\Big\}, \tag{16}$$

$$X_t = Y_n(t - \tau_n), \quad t \in [\tau_n, \tau_{n+1}), \qquad n \in \mathsf{Z}_+. \tag{17}$$

The distribution $\nu_x = P_x \circ Y_0^{-1}$ is called the *excursion law* at x. When x is recurrent and $r_{yx} = 1$, Proposition 11.9 shows that $Y_1, Y_2, \ldots$ are i.i.d. ν_x under P_y. The result extends to the general case, as follows.

Lemma 11.19 *(excursions) Let X be a discrete-time Markov process in a Borel space S, and fix any $x \in S$. Then there exist some independent processes $Y_0, Y_1, \ldots$ in S, all but Y_0 with distribution ν_x, such that X is a.s. given by* (16) *and* (17)*.*

Proof: Put $\tilde{Y}_0 \overset{d}{=} Y_0$, and let $\tilde{Y}_1, \tilde{Y}_2, \ldots$ be independent of $\tilde{Y}_0$ and i.i.d. ν_x. Construct associated random times $\tilde{\tau}_0, \tilde{\tau}_1, \ldots$ as in (16), and define a process $\tilde{X}$ as in (17). By Corollary 8.18, it is enough to show that $X \overset{d}{=} \tilde{X}$. Writing

$$\kappa = \sup\Big\{n \geq 0;\, \tau_n < \infty\Big\},$$
$$\tilde{\kappa} = \sup\Big\{n \geq 0;\, \tilde{\tau}_n < \infty\Big\},$$

it is equivalent to show that

$$(Y_0, \ldots, Y_\kappa, \bar{\Delta}, \bar{\Delta}, \ldots) \overset{d}{=} \Big(\tilde{Y}_0, \ldots, \tilde{Y}_{\tilde{\kappa}}, \bar{\Delta}, \bar{\Delta}, \ldots\Big). \tag{18}$$

Using the strong Markov property on the left and the independence of the $\tilde{Y}_n$ on the right, it is easy to check that both sides are Markov processes in $S^{\mathsf{Z}_+} \cup \{\bar{\Delta}\}$ with the same initial distribution and transition kernel. Hence, (18) holds by Proposition 11.2. □

We now consider discrete-time Markov chains X on the time scale Z_+, taking values in a countable state space S. Here the Chapman–Kolmogorov relation becomes

$$p_{ik}^{m+n} = \textstyle\sum_j p_{ij}^m\, p_{jk}^n, \quad i, k \in S,\ m, n \in \mathsf{N}, \tag{19}$$

where $p_{ij}^n = \mu_n(i, \{j\})$, $i, j \in S$. This may be written in matrix form as $p^{m+n} = p^m p^n$, which shows that the matrix p^n of n-step transition probabilities is simply the n-th power of the matrix $p = p^1$, justifying our notation. Regarding the initial distribution ν as a row vector (ν_i), we may write the distribution at time n as νp^n.

Now define $r_{ij} = P_i\{\tau_j < \infty\}$ as before, where $\tau_j = \inf\{n > 0;\, X_n = j\}$. A Markov chain in S is said to be *irreducible* if $r_{ij} > 0$ for all $i, j \in S$, so that every state can be reached from any other state. For irreducible chains, all states have the same recurrence and periodicity properties:

Proposition 11.20 *(irreducible chains) For an irreducible Markov chain,*

(i) *either all states are recurrent or all are transient,*

(ii) *all states have the same period,*

(iii) *if ν is invariant, then $\nu_i > 0$ for all i.*

For the proof of (i) we need the following lemma.

Lemma 11.21 *(recurrence classes) For a recurrent $i \in S$, define $S_i = \{j \in S;\, r_{ij} > 0\}$. Then*

(i) $r_{jk} = 1$ *for all* $j, k \in S_i$,

(ii) *all states in S_i are recurrent.*

Proof: By the recurrence of i and the strong Markov property, we get for any $j \in S_i$

$$\begin{aligned} 0 &= P_i\Big\{\tau_j < \infty,\, \tau_i \circ \theta_{\tau_j} = \infty\Big\} \\ &= P_i\{\tau_j < \infty\}\, P_j\{\tau_i = \infty\} \\ &= r_{ij}(1 - r_{ji}). \end{aligned}$$

Since $r_{ij} > 0$ by hypothesis, we obtain $r_{ji} = 1$. Fixing any $m, n \in \mathsf{N}$ with $p^m_{ij},\, p^n_{ji} > 0$, we get by (19)

$$\begin{aligned} E_j\kappa_j &\geq \sum_{s>0} p^{m+n+s}_{jj} \\ &\geq \sum_{s>0} p^n_{ji}\, p^s_{ii}\, p^m_{ij} \\ &= p^n_{ji}\, p^m_{ij}\, E_i\kappa_i = \infty, \end{aligned}$$

and so j is recurrent by Lemma 11.16. Reversing the roles of i and j gives $r_{ij} = 1$. Finally, we get for any $j, k \in S_i$

$$\begin{aligned} r_{jk} &\geq P_j\Big\{\tau_i < \infty,\, \tau_k \circ \theta_{\tau_i} < \infty\Big\} \\ &= r_{ji} r_{ik} = 1. \end{aligned}$$

□

Proof of Proposition 11.20: (i) Use Lemma 11.21.

(ii) Fix any $i, j \in S$, and choose $m, n \in \mathsf{N}$ with $p^m_{ij}, p^n_{ji} > 0$. By (19),

$$p^{m+h+n}_{jj} \geq p^n_{ji}\, p^h_{ii}\, p^m_{ij}, \quad h \geq 0.$$

For $h = 0$ we get $p^{m+n}_{jj} > 0$, and so[3] $d_j | (m+n)$. Hence, in general $p^h_{ii} > 0$ implies $d_j | h$, and we get $d_j \leq d_i$. Reversing the roles of i and j yields the opposite inequality.

(iii) Fix any $i \in S$. Choosing $j \in S$ with $\nu_j > 0$, and then $n \in \mathsf{N}$ with $p^n_{ji} > 0$, we see from (15) that even $\nu_i > 0$. □

We turn to the basic limit theorem for irreducible Markov chains. Related results appear in Chapters 13, 17, and 33. Write $\xrightarrow{u}$ for convergence in the total variation norm $\|\cdot\|$, and let $\hat{\mathcal{M}}_S$ be the class of probability measures on S.

Theorem 11.22 *(convergence dichotomy, Markov, Kolmogorov, Orey) For an irreducible, aperiodic Markov chain in S, one of these cases occurs:*

(i) *There exists a unique invariant distribution ν, the latter satisfies $\nu_i > 0$ for all $i \in S$, and as $n \to \infty$,*

$$P_\mu \circ \theta_n^{-1} \xrightarrow{u} P_\nu, \quad \mu \in \hat{\mathcal{M}}_S. \tag{20}$$

(ii) *No invariant distribution exists, and as $n \to \infty$,*

$$p^n_{ij} \to 0, \quad i, j \in S. \tag{21}$$

[3] Here $m|n$ (pronounced m divides n) means that n is divisible by m.

Markov chains satisfying (i) are clearly recurrent, whereas those in (ii) may be either recurrent or transient. This suggests a further division of the irreducible, aperiodic, and recurrent Markov chains into *positive-recurrent* and *null-recurrent* ones, depending on whether (i) or (ii) occurs.

We shall prove Theorem 11.22 by a simple *coupling*[4] argument. For our present purposes, an elementary coupling by independence is sufficient.

Lemma 11.23 *(coupling) Let X, Y be independent Markov chains in S, T with transition matrices $(p_{ii'})$ and $(q_{jj'})$, respectively. Then*

(i) *(X, Y) is a Markov chain in $S \times T$ with transition matrix $r_{ij,i'j'} = p_{ii'}\, q_{jj'}$,*

(ii) *if X and Y are irreducible, aperiodic, then so is (X, Y),*

(iii) *for X, Y as in* (ii) *with invariant distributions μ, ν, the chain (X, Y) is recurrent.*

Proof: (i) Using Proposition 11.2, we may compute the finite-dimensional distributions of (X, Y) for an arbitrary initial distribution $\mu \otimes \nu$ on $S \times T$.

(ii) For X, Y as stated, fix any $i, i' \in S$ and $j, j' \in T$. Then Proposition 11.18 yields $r^n_{ij,\,i'j'} = p^n_{ii'}\, q^n_{jj'} > 0$ for all but finitely many $n \in \mathsf{N}$, and so (X, Y) has again the stated properties.

(iii) Since the product measure $\mu \otimes \nu$ is clearly invariant for (X, Y), the assertion follows by Proposition 11.17. □

The point of the construction is that, if the coupled processes eventually meet, their distributions are asymptotically the same.

Lemma 11.24 *(strong ergodicity) For an irreducible, recurrent Markov chain in S^2 with transition matrix $p_{ii'}\, p_{jj'}$, we have as $n \to \infty$*

$$\left\| P_\mu \circ \theta_n^{-1} - P_\nu \circ \theta_n^{-1} \right\| \to 0, \quad \mu, \nu \in \hat{\mathcal{M}}_S. \tag{22}$$

Proof (Doeblin): Let X, Y be independent with distributions P_μ, P_ν. By Lemma 11.23 the pair (X, Y) is again Markov with respect to the induced filtration $\mathcal{F}$, and by Proposition 11.9 it satisfies the strong Markov property at every finite optional time τ. Choosing $\tau = \inf\{n \geq 0;\ X_n = Y_n\}$, we get for any measurable set $A \subset S^\infty$

$$\begin{aligned} P\Big\{\theta_\tau X \in A \,\Big|\, \mathcal{F}_\tau\Big\} &= P_{X_\tau} A = P_{Y_\tau} A \\ &= P\Big\{\theta_\tau Y \in A \,\Big|\, \mathcal{F}_\tau\Big\}. \end{aligned}$$

[4]The idea is to study the limiting behavior of a process X by approximating with a process $\tilde{X}$ whose asymptotic behavior is known, thus replacing a technical analysis of distributional properties by a simple pathwise comparison. The method is frequently employed in subsequent chapters.

In particular, $(\tau, X^\tau, \theta_\tau X) \overset{d}{=} (\tau, X^\tau, \theta_\tau Y)$. Putting $\tilde{X}_n = X_n$ for $n \le \tau$ and $\tilde{X}_n = Y_n$ otherwise, we obtain $\tilde{X} \overset{d}{=} X$, and so for any A as above,

$$\begin{aligned}
&\left|P\{\theta_n X \in A\} - P\{\theta_n Y \in A\}\right| \\
&\qquad = \left|P\{\theta_n \tilde{X} \in A\} - P\{\theta_n Y \in A\}\right| \\
&\qquad = \left|P\Big\{\theta_n \tilde{X} \in A,\, \tau > n\Big\} - P\Big\{\theta_n Y \in A,\, \tau > n\Big\}\right| \\
&\qquad \le P\{\tau > n\} \to 0.
\end{aligned}$$

□

The next result ensures the existence of an invariant distribution. Here a coupling argument is again useful.

Lemma 11.25 *(existence) If* (21) *fails, then an invariant distribution exists.*

Proof: Suppose that (21) fails, so that $\limsup_n p^n_{i_0,j_0} > 0$ for some $i_0, j_0 \in S$. By a diagonal argument, we have $p^n_{i_0,j} \to c_j$ along a sub-sequence $N' \subset \mathsf{N}$ for some constants c_j with $c_{j_0} > 0$, where $0 < \sum_j c_j \le 1$ by Fatou's lemma.

To extend the convergence to arbitrary i, let X, Y be independent processes with the same transition matrix (p_{ij}), and conclude from Lemma 11.23 that (X, Y) is an irreducible Markov chain on S^2 with transition probabilities $q_{ij,i'j'} = p_{ii'} p_{jj'}$. If (X, Y) is transient, then Proposition 11.16 yields

$$\textstyle\sum_n (p^n_{ij})^2 = \sum_n q^n_{ii,jj} < \infty, \quad i, j \in S,$$

and (21) follows. The pair (X, Y) is then recurrent, and Lemma 11.24 gives $p^n_{ij} - p^n_{i_0,j} \to 0$ for all $i, j \in I$. Hence, $p^n_{ij} \to c_j$ along N' for all i, j.

Next conclude from the Chapman–Kolmogorov relation that

$$\begin{aligned}
p^{n+1}_{ik} &= \textstyle\sum_j p^n_{ij}\, p_{jk} \\
&= \textstyle\sum_j p_{ij}\, p^n_{jk}, \quad i, k \in S.
\end{aligned}$$

Letting $n \to \infty$ along N', and using Fatou's lemma on the left and dominated convergence on the right, we get

$$\textstyle\sum_j c_j\, p_{jk} \le \sum_j p_{ij}\, c_k = c_k, \quad k \in S. \tag{23}$$

Summing over k gives $\sum_j c_j \le 1$ on both sides, and so (23) holds with equality. Thus, (c_i) is invariant, and we may form the invariant distribution $\nu_i = c_i / \sum_j c_j$. □

Proof of Theorem 11.22: If no invariant distribution exists, then (21) holds by Lemma 11.25. Now let ν be an invariant distribution, and note that $\nu_i > 0$ for all i by Proposition 11.20. By Lemma 11.23 the coupled chain in Lemma 11.24 is irreducible and recurrent, and so (22) holds for any initial distribution μ, and (20) follows since $P_\nu \circ \theta_n^{-1} = P_\nu$ by Lemma 11.11. If even ν' is invariant, then (20) yields $P_{\nu'} = P_\nu$, and so $\nu' = \nu$. □

The limits in Theorem 11.22 may be expressed as follows in terms of the *mean recurrence times* $E_j \tau_j$.

Theorem 11.26 *(mean recurrence times, Kolmogorov) For a Markov chain in S and states $i, j \in S$ with j aperiodic, we have as $n \to \infty$*

$$p_{ij}^n \to \frac{P_i\{\tau_j < \infty\}}{E_j \tau_j}. \tag{24}$$

Proof: First let $i = j$. If j is transient, then $p_{jj}^n \to 0$ and $E_j \tau_j = \infty$, and so (24) is trivially true. If instead j is recurrent, then the restriction of X to the set $S_j = \{i;\, r_{ji} > 0\}$ is irreducible and recurrent by Lemma 11.21, and aperiodic by Proposition 11.20. Hence, p_{jj}^n converges by Theorem 11.22.

To identify the limit, define

$$\begin{aligned} L_n &= \sup\left\{k \in \mathsf{Z}_+;\, \tau_j^k \le n\right\} \\ &= \sum_{k=1}^n 1\{X_k = j\}, \quad n \in \mathsf{N}. \end{aligned}$$

The τ_j^n form a random walk under P_j, and so the law of large numbers yields

$$\frac{L(\tau_j^n)}{\tau_j^n} = \frac{n}{\tau_j^n} \to \frac{1}{E_j \tau_j} \quad \text{a.s. } P_j.$$

By the monotonicity of L_k and τ_j^n, we obtain $L_n/n \to (E_j \tau_j)^{-1}$ a.s. P_j. Since $L_n \le n$, we get by dominated convergence

$$\frac{1}{n} \sum_{k=1}^n p_{jj}^k = \frac{E_j L_n}{n} \to \frac{1}{E_j \tau_j},$$

and (24) follows.

Now let $i \ne j$. Using the strong Markov property, the disintegration theorem, and dominated convergence, we get

$$\begin{aligned} p_{ij}^n &= P_i\{X_n = j\} \\ &= P_i\left\{\tau_j \le n,\, (\theta_{\tau_j} X)_{n-\tau_j} = j\right\} \\ &= E_i\left(p_{jj}^{n-\tau_j};\, \tau_j \le n\right) \\ &\to \frac{P_i\{\tau_j < \infty\}}{E_j \tau_j}. \end{aligned}$$

□

Exercises

1. Let X be a process with $X_s \perp\!\!\!\perp_{X_t} \{X_u,\, u \ge t\}$ for all $s < t$. Show that X is Markov with respect to the induced filtration.

2. Show that the Markov property is preserved under time-reversal, but the possible space or time homogeneity is not, in general. (*Hint:* Use Lemma 11.1, and consider a random walk starting at 0.)

3. Let X be a Markov process in a space S, and fix a measurable function f on S. Show by an example that the process $Y_t = f(X_t)$ need not be Markov. (*Hint:* Let X be a simple symmetric random walk in Z, and take $f(x) = [x/2]$.)

4. Let X be a Markov process in R with transition functions μ_t satisfying $\mu_t(x, B) = \mu_t(-x, -B)$. Show that the process $Y_t = |X_t|$ is again Markov.

5. Fix any process X on R_+, and define $Y_t = X^t = \{X_{s\wedge t};\, s \geq 0\}$. Show that Y is Markov with respect to the induced filtration.

6. Consider a random element ξ in a Borel space and a filtration $\mathcal{F}$ with $\mathcal{F}_\infty \subset \sigma\{\xi\}$. Show that the measure-valued process $X_t = \mathcal{L}(\xi\,|\mathcal{F}_t)$ is Markov. (*Hint:* Note that $\xi \perp\!\!\!\perp_{X_t} \mathcal{F}_t$ for all t.)

7. Let X be a time-homogeneous Markov process in a Borel space S. Prove the existence of some measurable functions $f_h\colon S\times[0,1] \to S$, $h \geq 0$, and $U(0,1)$ random variables $\vartheta_{t,h} \perp\!\!\!\perp X^t$, $t, h \geq 0$, such that $X_{t+h} = f_h(X_t, \vartheta_{t,h})$ a.s. for all $t, h \geq 0$.

8. Let X be a time-homogeneous, rcll Markov process in a Polish space S. Prove the existence of a measurable function $f\colon S \times [0,1] \to D_{\mathsf{R}_+,S}$ and some $U(0,1)$ random variables $\vartheta_t \perp\!\!\!\perp X^t$ such that $\theta_t X = f(X_t, \vartheta_t)$ a.s. Extend the result to optional times taking countably many values.

9. Let X be a process on R_+ with state space S, and define $Y_t = (X_t, t)$, $t \geq 0$. Show that X, Y are simultanously Markov, and that Y is then time-homogeneous. Give a relation between the transition kernels for X, Y. Express the strong Markov property of Y at a random time τ in terms of the process X.

10. Let X be a discrete-time Markov process in S with invariant distribution ν. Show that $P_\nu\{X_n \in B \text{ i.o.}\} \geq \nu B$ for every measurable set $B \subset S$. Use the result to give an alternative proof of Proposition 11.17. (*Hint:* Use Fatou's lemma.)

11. Fix an irreducible Markov chain in S with period d. Show that S has a unique partition into subsets $S_1, \ldots, S_d$, such that $p_{ij} = 0$ unless $i \in S_k$ and $j \in S_{k+1}$ for some $k \in \{1, \ldots, d\}$, with addition defined modulo d.

12. Let X be an irreducible Markov chain with period d, and define $S_1, \ldots, S_d$ as above. Show that the restrictions of (X_{nd}) to $S_1, \ldots, S_d$ are irreducible, aperiodic and either all positive recurrent or all null recurrent. In the former case, show that the original chain has a unique invariant distribution ν. Further show that (20) holds iff $\mu S_k = 1/d$ for all k. (*Hint:* If (X_{nd}) has an invariant distribution ν^k on S_k, then $\nu_j^{k+1} = \sum_i \nu_i^k p_{ij}$ form an invariant distribution in S_{k+1}.)

13. Given a Markov chain X in S, define the classes C_i as in Lemma 11.21. Show that if $j \in C_i$ but $i \notin C_j$ for some $i, j \in S$, then i is transient. If instead $i \in C_j$ for every $j \in C_i$, show that C_i is irreducible (so that the restriction of X to C_i is an irreducible Markov chain). Further show that the irreducible sets are disjoint, and that every state outside the irreducible sets is transient.

14. For any Markov chain, show that (20) holds iff $\sum_j |p_{ij}^n - \nu_j| \to 0$ for all i.

15. Let X be an irreducible, aperiodic Markov chain in N. Show that X is transient iff $X_n \to \infty$ a.s. under every initial distribution, and null recurrent iff the same divergence holds in probability but not a.s.

16. For every irreducible, positive recurrent subset $S_k \subset S$, there exists a unique invariant distribution ν_k restricted to S_k, and every invariant distribution is a convex combination $\sum_k c_k\, \nu_k$.

17. Show that a Markov chain in a finite state space S has at least one irreducible set and one invariant distribution. (*Hint:* Starting from any $i_0 \in S$, choose $i_1 \in C_{i_0}$, $i_2 \in C_{i_1}$, etc. Then $\bigcap_n C_{i_n}$ is irreducible.)

18. Let $X \perp\!\!\!\perp Y$ be Markov processes with transition kernels $\mu_{s,t}$ and $\nu_{s,t}$. Show that (X, Y) is again Markov with transition kernels $\mu_{s,t}(x, \cdot) \otimes \nu_{s,t}(y, \cdot)$. (*Hint:* Compute the finite-dimensional distributions from Proposition 11.2, or use Proposition 8.12 with no computations.)

19. Let $X \perp\!\!\!\perp Y$ be irreducible Markov chains with periods d_1, d_2. Show that $Z = (X, Y)$ is irreducible iff d_1, d_2 are relatively prime, and that Z has then period $d_1 d_2$.

20. State and prove a discrete-time version of Theorem 11.14. Further simplify the continuous-time proof when S is countable.

Chapter 12

Random Walks and Renewal Processes

Recurrence dichotomy and criteria, transience for $d \geq 3$, Wald equations, first maximum and last return, duality, ladder times and heights, fluctuations, Wiener–Hopf factorization, ladder distributions, boundedness and divergence, stationary renewal process, occupation measure, two-sided renewal theorem, asymptotic delay, renewal equation

Before turning to the continuous-time Markov chains, we consider the special case of space- and time-homogeneous processes in discrete time, where the increments are independent, identically distributed (i.i.d.), and thus determined by the single one-step distribution μ. Such processes are known as *random walks*. The corresponding continuous-time processes are the Lévy processes, to be discussed in Chapter 16. For simplicity, we consider only random walks in Euclidean spaces, though the definition makes sense in an arbitrary measurable group.

Thus, a *random walk* in R^d is defined as a discrete-time random process $X = (X_n)$ evolving by i.i.d. steps $\xi_n = \Delta X_n = X_n - X_{n-1}$. For most purposes we may take $X_0 = 0$, so that $X_n = \xi_1 + \ldots + \xi_n$ for all n. Though random walks may be regarded as the simplest of all Markov chains, they exhibit many basic features of the more general discrete-time processes, and hence may serve as a good introduction to the general subject. We shall further see how random walks enter naturally into the description and analysis of certain continuous-time phenomena.

Some basic facts about random walks have already been noted in previous chapters. Thus, we proved some simple $0-1$ laws in Chapter 4, and in Chapters 5–6 we established the ultimate versions of the law of large numbers and the central limit theorem, both of which deal with the asymptotic behavior of $n^{-c}X_n$ for suitable constants $c > 0$. More sophisticated limit theorems of this kind will be derived in Chapters 22–23 and 30, often through approximation by a Brownian motion or a more general Lévy process.

Random walks in R^d are either recurrent or transient, and our first aim is to derive *recurrence criteria* in terms of the transition distribution μ. We proceed with some striking connections between maxima and return times, anticipating the arcsine laws in Chapters 14 and 22. This is followed by a detailed study of *ladder times* and *heights* for one-dimensional random walks, culminating with the *Wiener–Hopf factorization* and *Baxter's formula*. Finally, we prove a two-sided version of the *renewal theorem*, describing the asymptotic behavior of the occupation measure and its intensity for a transient random walk.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_13

In addition to the mentioned connections, we note the relevance of renewal theory for the study of continuous-time Markov chains, as considered in Chapter 13. Renewal processes are simple instances of *regenerative sets*, to be studied in full generality in Chapter 29, in connection with local time and excursion theory.

To begin our systematic discussion of random walks, assume as before that $X_n = \xi_1 + \cdots + \xi_n$ for all $n \in \mathsf{Z}_+$, where the ξ_n are i.i.d. random vectors in R^d. The distribution of X is then determined by the common distribution $\mu = \mathcal{L}(\xi_n)$ of the increments. By the *effective dimension* of X we mean the dimension of the linear subspace spanned by the support of μ. For most purposes, we may assume that the effective dimension agrees with the dimension of the underlying space, since we may otherwise restrict our attention to the generated subspace.

A random walk X in R^d is said to be *recurrent* if $\liminf_{n\to\infty} |X_n| = 0$ a.s. and *transient* if $|X_n| \to \infty$ a.s. In terms of the *occupation measure* $\xi = \sum_n \delta_{X_n}$, transience means that ξ is a.s. locally finite, and recurrence that $\xi B_0^\varepsilon = \infty$ a.s. for every $\varepsilon > 0$, where $B_x^r = \{y;\, |x-y| < r\}$. We define the *recurrence set* A of X as the set of points $x \in \mathsf{R}^d$ with $\xi B_x^\varepsilon = \infty$ a.s. for every $\varepsilon > 0$.

We now state the basic dichotomy for random walks in R^d. Here the *intensity (measure)* $E\xi$ of ξ is defined by $(E\xi)f = E(\xi f)$.

Theorem 12.1 *(recurrence dichotomy) Let X be a random walk in R^d with occupation measure ξ and recurrence set A. Then exactly one of these cases occurs:*

(i) *X is recurrent with $A = \operatorname{supp} E\xi$, which is then a closed subgroup of R^d,*

(ii) *X is transient and $E\xi$ is locally finite.*

Proof: The event $|X_n| \to \infty$ has probability 0 or 1 by Kolmogorov's 0−1 law. When $|X_n| \to \infty$ a.s., the Markov property at time m yields for any $m, n \in \mathsf{N}$ and $r > 0$

$$\begin{aligned} P\Big\{ |X_m| < r,\ \inf_{k\ge n} |X_{m+k}| \ge r \Big\} &\ge P\Big\{ |X_m| < r,\ \inf_{k\ge n} |X_{m+k} - X_m| \ge 2r \Big\} \\ &= P\{|X_m| < r\}\, P\Big\{ \inf_{k\ge n} |X_k| \ge 2r \Big\}. \end{aligned}$$

Noting that the event on the left can occur for at most n different values of m, we get

$$P\Big\{ \inf_{k\ge n} |X_k| \ge 2r \Big\} \sum_{m\ge 1} P\{|X_m| < r\} < \infty, \quad n \in \mathsf{N}.$$

Since $|X_k| \to \infty$ a.s., the probability on the left is positive for large enough n, and so for $r > 0$

$$\begin{aligned} E\,\xi B_0^r &= E \sum_{m\ge 1} 1\{|X_m| < r\} \\ &= \sum_{m\ge 1} P\{|X_m| < r\} < \infty, \end{aligned}$$

which shows that $E\xi$ is locally finite.

Next let $|X_n| \not\to \infty$ a.s., so that $P\{|X_n| < r \text{ i.o.}\} > 0$ for some $r > 0$. Covering B_0^r by finitely many balls $G_1, \dots, G_n$ of radius $\varepsilon/2$, we conclude that $P\{X_n \in G_k \text{ i.o.}\} > 0$ for at least one k. By the Hewitt–Savage 0−1 law, the latter probability is in fact 1. Thus, the optional time $\tau = \inf\{n;\, X_n \in G_k\}$ is a.s. finite, and so the strong Markov property yields

$$\begin{aligned} 1 &= P\Big\{X_n \in G_k \text{ i.o.}\Big\} \\ &\le P\Big\{|X_{\tau+n} - X_\tau| < \varepsilon \text{ i.o.}\Big\} \\ &= P\Big\{|X_n| < \varepsilon \text{ i.o.}\Big\} \\ &= P\{\xi B_0^\varepsilon = \infty\}, \end{aligned}$$

which shows that $0 \in A$, and X is recurrent.

The set $S = \operatorname{supp} E\xi$ clearly contains A. To prove the converse, fix any $x \in S$ and $\varepsilon > 0$. By the strong Markov property at $\sigma = \inf\{n \ge 0;\, |X_n - x| < \varepsilon/2\}$ and the recurrence of X, we get

$$\begin{aligned} P\{\xi B_x^\varepsilon = \infty\} &= P\Big\{|X_n - x| < \varepsilon \text{ i.o.}\Big\} \\ &\ge P\Big\{\sigma < \infty,\ |X_{\sigma+n} - X_\sigma| < \varepsilon/2 \text{ i.o.}\Big\} \\ &= P\{\sigma < \infty\}\, P\Big\{|X_n| < \varepsilon/2 \text{ i.o.}\Big\} > 0. \end{aligned}$$

By the Hewitt–Savage 0−1 law, the probability on the left then equals 1, which means that $x \in A$. Thus, $S \subset A \subset S$, and so in this case $A = S$.

The set S is clearly a closed additive semi-group in R^d. To see that in this case it is even a group, it remains to show that $x \in S$ implies $-x \in S$. Defining σ as before and using the strong Markov property, along with the fact that $x \in S \subset A$, we get

$$\begin{aligned} P\{\xi B_{-x}^\varepsilon = \infty\} &= P\Big\{|X_n + x| < \varepsilon \text{ i.o.}\Big\} \\ &= P\Big\{|X_{\sigma+n} - X_\sigma + x| < \varepsilon \text{ i.o.}\Big\} \\ &\ge P\Big\{|X_\sigma - x| < \varepsilon,\ |X_n| < \varepsilon/2 \text{ i.o.}\Big\} = 1, \end{aligned}$$

which shows that $-x \in A = S$. □

We give some simple sufficient conditions for recurrence.

Theorem 12.2 *(recurrence for $d \le 2$) A random walk X in R^d is recurrent under each of these conditions:*

- *for $d = 1$:* $n^{-1} X_n \xrightarrow{P} 0$,
- *for $d = 2$:* $E\xi_1 = 0$ *and* $E|\xi_1|^2 < \infty$.

For $d = 1$ we recognize the weak law of large numbers, which was characterized in Theorem 6.17. In particular, the condition holds when $E\xi_1 = 0$.

By contrast, $E\xi_1 \in (0,\infty]$ implies $X_n \to \infty$ a.s. by the strong law of large numbers, so in that case X is transient.

Our proof of Theorem 12.2 is based on the following scaling relation[1].

Lemma 12.3 *(scaling) For a random walk X in $\mathbb{R}^d$,*

$$\sum_{n\geq 0} P\{|X_n| \leq r\varepsilon\} \lesssim r^d \sum_{n\geq 0} P\{|X_n| \leq \varepsilon\}, \quad r \geq 1, \ \varepsilon > 0.$$

Proof: The ball $B_0^{r\varepsilon}$ may be covered by balls $G_1, \dots, G_m$ of radius $\varepsilon/2$, where $m \lesssim r^d$. Introducing the optional times $\tau_k = \inf\{n;\, X_n \in G_k\}$, $k = 1, \dots, m$, we see from the strong Markov property that

$$\begin{aligned} \textstyle\sum_n P\{|X_n| \leq r\varepsilon\} &\leq \textstyle\sum_{k,n} P\{X_n \in G_k\} \\ &\leq \textstyle\sum_{k,n} P\Big\{|X_{\tau_k+n} - X_{\tau_k}| \leq \varepsilon;\ \tau_k < \infty\Big\} \\ &= \textstyle\sum_k P\{\tau_k < \infty\} \sum_n P\{|X_n| \leq \varepsilon\} \\ &\lesssim r^d \textstyle\sum_n P\{|X_n| \leq \varepsilon\}. \end{aligned}$$

□

Proof of Theorem 12.2 *(Chung & Ornstein):* ($d = 1$) Fix any $\varepsilon > 0$ and $r \geq 1$, and conclude from Lemma 12.3 that

$$\begin{aligned} \textstyle\sum_n P\{|X_n| \leq \varepsilon\} &\gtrsim r^{-1} \textstyle\sum_n P\{|X_n| \leq r\varepsilon\} \\ &= \int_0^\infty P\Big\{|X_{[rt]}| \leq r\varepsilon\Big\}\, dt. \end{aligned}$$

Since the integrand on the right tends to 1 as $r \to \infty$, the integral tends to ∞ by Fatou's lemma, and the recurrence of X follows by Theorem 12.1.

($d = 2$) We may take X to have effective dimension 2, since the 1-dimensional case is already covered by (i). Then the central limit theorem yields $n^{-1/2}X_n \xrightarrow{d} \zeta$ for a non-degenerate normal random vector ζ. In particular, $P\{|\zeta| \leq c\} \gtrsim c^2$ for bounded $c > 0$. Fixing any $\varepsilon > 0$ and $r \geq 1$, we conclude from Lemma 12.3 that

$$\begin{aligned} \textstyle\sum_n P\{|X_n| \leq \varepsilon\} &\gtrsim r^{-2} \textstyle\sum_n P\{|X_n| \leq r\varepsilon\} \\ &= \int_0^\infty P\Big\{|X_{[r^2t]}| \leq r\varepsilon\Big\}\, dt. \end{aligned}$$

As $r \to \infty$, we get by Fatou's lemma

$$\begin{aligned} \textstyle\sum_n P\{|X_n| \leq \varepsilon\} &\gtrsim \int_0^\infty P\Big\{|\zeta| \leq \varepsilon t^{-1/2}\Big\}\, dt \\ &\gtrsim \varepsilon^2 \int_1^\infty t^{-1} dt = \infty, \end{aligned}$$

and the recurrence follows again by Theorem 12.1. □

An exact recurrence criterion can be given in terms of the characteristic function $\hat{\mu}$ of μ. Here we write $B_\varepsilon = B_0^\varepsilon$ for an open ε-ball around the origin.

[1] Here $a \lesssim b$ means that $a \leq cb$ for some constant $c > 0$.

Theorem 12.4 *(recurrence criterion, Chung & Fuchs) Let X be a random walk in R^d based on a distribution μ, and fix an $\varepsilon > 0$. Then X is recurrent iff*

$$\sup_{0<r<1} \int_{B_\varepsilon} \Re \frac{1}{1 - r\hat{\mu}_t}\, dt = \infty. \tag{1}$$

The proof is based on a classical identity.

Lemma 12.5 *(Parseval) Let μ, ν be probability measures on R^d with characteristic functions $\hat{\mu}$, $\hat{\nu}$. Then*

$$\int \hat{\mu}\, d\nu = \int \hat{\nu}\, d\mu.$$

Proof: Use Fubini's theorem. □

Proof of Theorem 12.4: The function $f(s) = (1-|s|)_+$ has Fourier transform $\hat{f}(t) = 2\,t^{-2}(1 - \cos t)$, and so the tensor product $f^{\otimes d}(s) = \prod_{k\le d} f(s_k)$ on R^d has Fourier transform $\hat{f}^{\otimes d}(t) = \prod_{k\le d} \hat{f}(t_k)$. Writing $\mu^{*n} = \mathcal{L}(X_n)$, we get by Lemma 12.5 for any $a > 0$ and $n \in \mathsf{Z}_+$

$$\int \hat{f}^{\otimes d}(x/a)\, \mu^{*n}(dx) = a^d \int f^{\otimes d}(at)\, \hat{\mu}_t^n\, dt.$$

By Fubini's theorem, it follows that for any $r \in (0,1)$

$$\int \hat{f}^{\otimes d}(x/a) \sum_{n\ge 0} r^n \mu^{*n}(dx) = a^d \int \frac{f^{\otimes d}(at)}{1 - r\hat{\mu}_t}\, dt. \tag{2}$$

Now assume that (1) fails. Putting $\delta = \varepsilon^{-1} d^{1/2}$, we get by (2)

$$\begin{aligned} \textstyle\sum_n P\{|X_n| < \delta\} &= \textstyle\sum_n \mu^{*n}(B_\delta) \\ &\lesssim \int \hat{f}^{\otimes d}(x/\delta) \textstyle\sum_n \mu^{*n}(dx) \\ &= \delta^d \sup_{r<1} \int \frac{f^{\otimes d}(\delta t)}{1 - r\hat{\mu}_t}\, dt \\ &\lesssim \varepsilon^{-d} \sup_{r<1} \int_{B_\varepsilon} \frac{dt}{1 - r\hat{\mu}_t} < \infty, \end{aligned}$$

and so X is transient by Theorem 12.1.

To prove the converse, we note that $\hat{f}^{\otimes d}$ has Fourier transform $(2\pi)^d f^{\otimes d}$. Hence, (2) remains true with f and $\hat{f}$ interchanged, apart from a factor $(2\pi)^d$ on the left. If X is transient, then for any $\varepsilon > 0$ with $\delta = \varepsilon^{-1} d^{1/2}$,

$$\begin{aligned} \sup_{r<1} \int_{B_\varepsilon} \frac{dt}{1 - r\hat{\mu}_t} &\lesssim \sup_{r<1} \int \frac{\hat{f}^{\otimes d}(t/\varepsilon)}{1 - r\hat{\mu}_t}\, dt \\ &\lesssim \varepsilon^d \int f^{\otimes d}(\varepsilon x) \textstyle\sum_n \mu^{*n}(dx) \\ &\le \varepsilon^d \textstyle\sum_n \mu^{*n}(B_\delta) < \infty. \end{aligned}$$

□

When μ is *symmetric* in the sense that $\xi_1 \overset{d}{=} -\xi_1$, then $\hat\mu$ is real valued, and the last criterion becomes

$$\int_{B_\varepsilon} \frac{dt}{1-\hat\mu_t} = \infty.$$

By a *symmetrization* of X we mean a random walk $\tilde X = X - X'$, where X' is an independent copy of X. We may relate the recurrence behavior of X to that of $\tilde X$:

Corollary 12.6 *(symmetrization) Let X be a random walk in R^d with symmetrization $\tilde X$. Then*

$$X \text{ is recurrent} \quad \Rightarrow \quad \tilde X \text{ is recurrent.}$$

Proof: Since clearly $(\Re z)(\Re z^{-1}) \le 1$ for any $z \ne 0$ in C, we get

$$\Re \frac{1}{1-r\hat\mu^2} \le \frac{1}{1-r\,\Re\,\hat\mu^2} \le \frac{1}{1-r|\hat\mu|^2}.$$

Thus, if $\tilde X$ is transient, then so is the random walk (X_{2n}) by Theorem 12.4. But then $|X_{2n}| \to \infty$ a.s. by Theorem 12.1, and so $|X_{2n+1}| \to \infty$ a.s. By combination, $|X_n| \to \infty$ a.s., which means that X is transient. □

The following sufficient conditions for recurrence or transience are often useful for applications:

Corollary 12.7 *(sufficient conditions) Let X be a random walk in R^d, and fix any $\varepsilon > 0$. Then*

(i) $\displaystyle\int_{B_\varepsilon} \Re \frac{1}{1-\hat\mu_t}\, dt = \infty \quad \Rightarrow \quad X$ *is recurrent,*

(ii) $\displaystyle\int_{B_\varepsilon} \frac{dt}{1-\Re\,\hat\mu_t} < \infty \quad \Rightarrow \quad X$ *is transient.*

Proof: (i) Under the stated condition, Fatou's lemma yields for any $r_n \uparrow 1$

$$\begin{aligned}\liminf_{n\to\infty} \int_{B_\varepsilon} \Re \frac{1}{1-r_n\hat\mu} &\ge \int_{B_\varepsilon} \lim_{n\to\infty} \Re \frac{1}{1-r_n\hat\mu} \\ &= \int_{B_\varepsilon} \Re \frac{1}{1-\hat\mu} = \infty.\end{aligned}$$

Thus, (1) holds and X is recurrent.

(ii) Let μ be such as stated. Decreasing ε if necessary, we may assume that $\Re\,\hat\mu \ge 0$ on B_ε. Then as before,

$$\int_{B_\varepsilon} \Re \frac{1}{1-r\hat\mu} \le \int_{B_\varepsilon} \frac{1}{1-r\,\Re\,\hat\mu} \le \int_{B_\varepsilon} \frac{1}{1-\Re\,\hat\mu} < \infty,$$

and so (1) fails. Thus, X is transient. □

We may now supplement Theorem 12.2 with some conclusive information for $d \ge 3$.

Theorem 12.8 *(transience for $d \geq 3$) Let X be a random walk of effective dimension d. Then*

$$d \geq 3 \quad \Rightarrow \quad X \text{ is transient.}$$

Proof: We may take the symmetrized distribution to be d-dimensional, since μ is otherwise supported by a hyper-plane outside the origin, and the transience follows by the strong law of large numbers. By Corollary 12.6 it is enough to prove that the symmetrized random walk $\tilde{X}$ is transient, and so we may assume that μ is symmetric. Considering the conditional distributions on B_r and B_r^c for large enough $r > 0$, we may write μ as a convex combination $c\,\mu_1 + (1-c)\,\mu_2$, where μ_1 is symmetric and d-dimensional with bounded support. Writing (r_{ij}) for the covariance matrix of μ_1, we get as in Lemma 6.9

$$\hat{\mu}_1(t) = 1 - \tfrac{1}{2}\textstyle\sum_{i,j} r_{ij}\, t_i t_j + o(|t|^2), \quad t \to 0.$$

Since the matrix (r_{ij}) is positive definite, it follows that $1 - \hat{\mu}_1(t) \gtrsim |t|^2$ for small enough $|t|$, say for $t \in B_\varepsilon$. A similar relation then holds for $\hat{\mu}$, and so

$$\int_{B_\varepsilon} \frac{dt}{1 - \hat{\mu}_t} \lesssim \int_{B_\varepsilon} \frac{dt}{|t|^2} \lesssim \int_0^\varepsilon r^{d-3} dr < \infty.$$

Thus, X is transient by Theorem 12.4. □

For optional times τ, some moments of X_τ are the same as if[2] $X \perp\!\!\!\perp \tau$. Given a discrete filtration $\mathcal{F}$, we say that X is an $\mathcal{F}$*-random walk*, if it is adapted to $\mathcal{F}$ and such that $\mathcal{F}_k \perp\!\!\!\perp \xi_{k+1}$ for all k.

Theorem 12.9 *(Wald equations) Let $X_n = \xi_1 + \cdots + \xi_n$ be an $\mathcal{F}$-random walk with $E\xi_1 = \mu$ and $\mathrm{Var}(\xi_1) = \sigma^2$, and let τ be an $\mathcal{F}$-optional time. Then*

(i) *for $\xi_1 \in L^1$ and $E\tau < \infty$,*

$$E X_\tau = \mu\, E\tau,$$

(ii) *for $\mu = 0$, $\sigma^2 < \infty$, and $E\tau < \infty$,*

$$E X_\tau^2 = \sigma^2 E\tau,$$

(iii) *for $X \perp\!\!\!\perp \tau$, $\sigma^2 < \infty$, and $E\tau^2 < \infty$,*

$$\mathrm{Var}(X_\tau) = \sigma^2 E\tau + \mu^2\, \mathrm{Var}(\tau).$$

Proof: (i) First let $\xi_k \geq 0$ a.s. for all k. Since $\{\tau \geq k\} \perp\!\!\!\perp \xi_k$ for all k, we get by Fubini's theorem and Lemma 4.4

[2] This illustrates the idea of *decoupling*, where the distribution of a random pair (ξ, η) is compared with that of a pair $(\tilde{\xi}, \tilde{\eta})$ with independent $\tilde{\xi} \overset{d}{=} \xi$ and $\tilde{\eta} \overset{d}{=} \eta$. Further instances appear in Chapters 15–16 and 27–28.

$$\begin{aligned} E X_\tau &= E \textstyle\sum_k \xi_k 1\{\tau \ge k\} \\ &= \textstyle\sum_k E(\xi_k;\, \tau \ge k) \\ &= \textstyle\sum_k E\xi_k\, P\{\tau \ge k\} \\ &= \mu \textstyle\sum_k P\{\tau \ge k\} = \mu\, E\tau. \end{aligned}$$

The general result follows by subtraction of the formulas for random walks with terms $\xi_k^{\pm} = (\pm\xi_k) \vee 0$.

(ii) First let τ be bounded, so that all sums below are finite. Since $\mu = 0$ and $\xi_j 1\{\tau \ge k\} \perp\!\!\!\perp \xi_k$ when $j < k$, we get by Lemma 4.4

$$\begin{aligned} E X_\tau^2 &= E\Big(\textstyle\sum_k \xi_k 1\{\tau \ge k\}\Big)^2 \\ &= E \textstyle\sum_k \xi_k^2\, 1\{\tau \ge k\} + 2\, E \displaystyle\sum_{j<k} \xi_j\, \xi_k 1\{\tau \ge k\} \\ &= \textstyle\sum_k E\Big(\xi_k^2;\, \tau \ge k\Big) + 2 \displaystyle\sum_{j<k} E\Big(\xi_j\, \xi_k;\, \tau \ge k\Big) \\ &= \textstyle\sum_k E\xi_k^2\, P\{\tau \ge k\} + 2 \displaystyle\sum_{j<k} E\Big(\xi_j;\, \tau \ge k\Big)\, E\xi_k \\ &= \sigma^2 \textstyle\sum_k P\{\tau \ge k\} = \sigma^2 E\tau. \end{aligned}$$

Similarly, we get for $m \le n$

$$\begin{aligned} E\Big(X_{\tau\wedge m} - X_{\tau\wedge n}\Big)^2 &= E\Big(\sum_{k=m+1}^{n} \xi_k 1\{\tau \ge k\}\Big)^2 \\ &= \sigma^2 \sum_{k=m+1}^{n} P\{\tau \ge k\} \\ &= \sigma^2 \Big\{E(\tau\wedge n) - E(\tau \wedge m)\Big\} \to 0. \end{aligned}$$

Hence, $X_{\tau\wedge n} \to X_\tau$ holds both a.s. and in L^2, and the relation $EX_{\tau\wedge n}^2 = \sigma^2 E(\tau \wedge n)$ extends in the limit to $EX_\tau^2 = \sigma^2 E\tau$.

(iii) Using Lemma 8.2 (i), and noting that $\mathcal{L}(\zeta, \tau \,|\, \tau) = \mathcal{L}(\zeta, t)_{t=\tau}$ when $\zeta \perp\!\!\!\perp \tau$, we get

$$\begin{aligned} \mathrm{Var}(X_\tau) &= E\ \mathrm{Var}(X_\tau |\, \tau) + \mathrm{Var}\Big\{E(X_\tau |\, \tau)\Big\} \\ &= E\,\tau\, \mathrm{Var}(\xi_1) + \mathrm{Var}(\tau\, E\xi_1) \\ &= \sigma^2 E\tau + \mu^2\, \mathrm{Var}(\tau). \end{aligned}$$

□

We proceed with a detailed study of one-dimensional random walks $X_n = \xi_1 + \cdots + \xi_n$, $n \in \mathsf{Z}_+$. Say that X is *simple* if $|\xi_1| = 1$ a.s. For a simple, symmetric random walk X we note that

$$u_n \equiv P\{X_{2n} = 0\} = 2^{-2n}\binom{2n}{n}, \quad n \in \mathsf{Z}_+. \tag{3}$$

First we give a surprising connection between the u_n and the times of last return to the origin.

Theorem 12.10 *(last return, Feller) Let X be a simple, symmetric random walk in Z, put $\sigma_n = \max\{k \le n;\, X_{2k} = 0\}$, and define u_n by* (3). *Then*

$$P\{\sigma_n = k\} = u_k\, u_{n-k}, \quad 0 \le k \le n.$$

The proof is based on a classical symmetry property, which will also appear in a continuous-time version as Lemma 14.14.

Lemma 12.11 *(reflection principle, André) For any symmetric random walk X and optional time τ, we have $\tilde{X} \stackrel{d}{=} X$, where*

$$\tilde{X}_n = X_{n\wedge\tau} - (X_n - X_{n\wedge\tau}), \quad n \ge 0.$$

Proof: We may clearly assume that $\tau < \infty$ a.s. Writing $X'_n = X_{\tau+n} - X_\tau$, $n \in \mathsf{Z}_+$, we get by the strong Markov property $X \stackrel{d}{=} X' \perp\!\!\!\perp (X^\tau, \tau)$, and by symmetry $-X \stackrel{d}{=} X$. Hence, by combination $(-X', X^\tau, \tau) \stackrel{d}{=} (X', X^\tau, \tau)$, and the assertion follows by suitable assembly. □

Proof of Theorem 12.10: The Markov property at time $2k$ yields

$$P\{\sigma_n = k\} = P\{X_{2k} = 0\}\, P\{\sigma_{n-k} = 0\}, \quad 0 \le k \le n,$$

which reduces the proof to the case of $k = 0$. Thus, it remains to show that

$$P\Big\{X_2 \ne 0, \ldots, X_{2n} \ne 0\Big\} = P\{X_{2n} = 0\}, \quad n \in \mathsf{N}.$$

By the Markov property at time 1, the left-hand side equals

$$\tfrac{1}{2}\, P\Big\{\min_{k<2n} X_k = 0\Big\} + \tfrac{1}{2}\, P\Big\{\max_{k<2n} X_k = 0\Big\} = P\{M_{2n-1} = 0\},$$

where $M_n = \max_{k\le n} X_k$. Using Lemma 12.11 with $\tau = \inf\{k;\, X_k = 1\}$, we get

$$\begin{aligned}
1 - P&\{M_{2n-1} = 0\} \\
&= P\{M_{2n-1} \ge 1\} \\
&= P\Big\{M_{2n-1} \ge 1,\, X_{2n-1} \ge 1\Big\} + P\Big\{M_{2n-1} \ge 1,\, X_{2n-1} \le 0\Big\} \\
&= P\{X_{2n-1} \ge 1\} + P\{X_{2n-1} \ge 2\} \\
&= 1 - P\{X_{2n-1} = 1\} \\
&= 1 - P\{X_{2n} = 0\}.
\end{aligned}$$

□

We continue with a striking connection between the maximum of a symmetric random walk and the last return probabilities in Theorem 12.10. Related results for Brownian motion and more general random walks will appear in Theorems 14.16 and 22.11.

Theorem 12.12 *(first maximum, Sparre-Andersen) Let X be a random walk based on a symmetric, diffuse distribution, and put*

$$M_n = \max_{k\le n} X_k, \qquad \tau_n = \min\Big\{k \ge 0;\, X_k = M_n\Big\}.$$

Define σ_n as in Proposition 12.10, in terms of a simple, symmetric random walk. Then $\tau_n \stackrel{d}{=} \sigma_n$ for every $n \ge 0$.

Here and below, we will use the relation

$$(X_1,\ldots,X_n) \stackrel{d}{=} \Big(X_n - X_{n-1},\ldots,\, X_n - X_0\Big), \quad n \in \mathsf{N}, \tag{4}$$

valid for any random walk X. The formula is obvious from the fact that $(\xi_1,\ldots,\xi_n) \stackrel{d}{=} (\xi_n,\ldots,\xi_1)$.

Proof of Theorem 12.12: By the symmetry of X together with (4), we have

$$\begin{aligned} v_k &\equiv P\{\tau_k = 0\} \\ &= P\{\tau_k = k\}, \quad k \geq 0. \end{aligned} \tag{5}$$

Hence, the Markov property at time k yields

$$\begin{aligned} P\{\tau_n = k\} &= P\{\tau_k = k\}\, P\{\tau_{n-k} = 0\} \\ &= v_k\, v_{n-k}, \qquad\qquad 0 \leq k \leq n. \end{aligned} \tag{6}$$

Clearly $\sigma_0 = \tau_0 = 0$. Proceeding by induction, let $\sigma_k \stackrel{d}{=} \tau_k$ and hence $u_k = v_k$ for all $k < n$. Comparing (6) with Theorem 12.10 gives

$$P\{\sigma_n = k\} = P\{\tau_n = k\}, \quad 0 < k < n,$$

which extends to $0 \leq k \leq n$ by (5). Thus, $\sigma_n \stackrel{d}{=} \tau_n$. □

For a one-dimensional random walk X, the *ascending ladder times* $\tau_1, \tau_2, \ldots$ are given recursively by

$$\tau_n = \inf\Big\{k > \tau_{n-1};\ X_k > X_{\tau_{n-1}}\Big\}, \quad n \in \mathsf{N}, \tag{7}$$

starting with $\tau_0 = 0$. The associated *ascending ladder heights* are defined as the random variables X_{τ_n}, $n \in \mathsf{N}$, where X_∞ may be interpreted as ∞. In a similar way, we define the *descending* ladder times τ_n^- and heights $X_{\tau_n^-}$, $n \in \mathsf{N}$. The times τ_n and τ_n^- are clearly optional. By the strong Markov property, the pairs (τ_n, X_{τ_n}) and $(\tau_n^-, X_{\tau_n^-})$ form possibly terminating random walks in $\bar{\mathsf{R}}^2$.

Replacing the relation $X_k > X_{\tau_{n-1}}$ in (7) by $X_k \geq X_{\tau_{n-1}}$, we obtain the *weak ascending ladder times* σ_n and *heights* X_{σ_n}. Similarly, we may introduce the *weak descending* ladder times σ_n^- and heights $X_{\sigma_n^-}$. The mentioned sequences are connected by a pair of simple but powerful duality[3] relations.

Lemma 12.13 *(duality) Let* η, η', ζ, ζ' *be the occupation measures of the sequences* (X_{τ_n}), (X_{σ_n}), $\{X_n;\, n < \tau_1^-\}$, *and* $\{X_n;\, n < \sigma_1^-\}$, *respectively. Then*

$$E\eta = E\zeta', \qquad E\eta' = E\zeta.$$

[3]Duality methods are often used in probability. Further notable instances appear in Chapters 21 and 31.

Proof: By (4) we have for any $B \in \mathcal{B}_{(0,\infty)}$ and $n \in \mathsf{N}$

$$\begin{aligned} P\Big\{X_1 \wedge \cdots \wedge X_{n-1} > 0,\ X_n \in B\Big\} \\ &= P\Big\{X_1 \vee \cdots \vee X_{n-1} < X_n \in B\Big\} \\ &= \textstyle\sum_k P\Big\{\tau_k = n,\ X_{\tau_k} \in B\Big\}. \end{aligned} \tag{8}$$

Summing over $n \geq 1$ gives $E\zeta' B = E\eta B$, and the first assertion follows. The proof of the second assertion is similar. □

The last result yields some remarkable information. Thus, for a simple, symmetric random walk, the expected number of visits to an arbitrary state $k \neq 0$, before the first return to 0, is constant and equal to 1. In particular, the mean recurrence time is infinite, and so X is a null-recurrent Markov chain.

The asymptotic behavior of a random walk is related to the expected values of the ladder times:

Proposition 12.14 *(fluctuations and mean ladder times) For a non-degenerate random walk X in R, exactly one of these cases occurs:*

(i) $X_n \to \infty$ *a.s. and* $E\tau_1 < \infty$,

(ii) $X_n \to -\infty$ *a.s. and* $E\tau_1^- < \infty$,

(iii) $\limsup_n(\pm X_n) = \infty$ *a.s. and* $E\sigma_1 = E\sigma_1^- = \infty$.

Proof: By Corollary 4.17 there are only the three possibilities $X_n \to \infty$ a.s., $X_n \to -\infty$ a.s., and $\limsup_n(\pm X_n) = \infty$ a.s. In the first case, $\sigma_n^- < \infty$ for finitely many n, say for $n < \kappa < \infty$. Then κ is geometrically distributed, and so $E\tau_1 = E\kappa < \infty$ by Lemma 12.13. The proof in case (ii) is similar. In case (iii), all variables τ_n and τ_n^- are finite, and Lemma 12.13 yields $E\sigma_1 = E\sigma_1^- = \infty$. □

Next we explore the relationship between the asymptotic behavior of a random walk and the expected values of X_1 and X_{τ_1}. Define $E\gamma = E\gamma^+ - E\gamma^-$ whenever $E\gamma^+ \wedge E\gamma^- < \infty$.

Proposition 12.15 *(fluctuations and mean ladder heights) Let X be a non-degenerate random walk in R. Then*

(i) $EX_1 = 0 \ \Rightarrow\ \limsup_n(\pm X_n) = \infty$ *a.s.,*

(ii) $EX_1 \in (0,\infty] \ \Rightarrow\ X_n \to \infty$ *a.s. and* $EX_{\tau_1} = E\tau_1\, EX_1$,

(iii) $EX_1^+ = EX_1^- = \infty \ \Rightarrow\ EX_{\tau_1} = -EX_{\tau_1^-} = \infty$.

Part (i) is clear from Theorem 12.2 (i). It can also be obtained more directly, as follows.

Proof: (i) By symmetry, we may assume that $\limsup_n X_n = \infty$ a.s. If $E\tau_1 < \infty$, we may apply the law of large numbers to each of the three ratios in the equation

$$\frac{X_{\tau_n}}{\tau_n}\frac{\tau_n}{n} = \frac{X_{\tau_n}}{n}, \quad n \in \mathsf{N},$$

to get $0 = EX_1 E\tau_1 = EX_{\tau_1} > 0$. The contradiction shows that $E\tau_1 = \infty$, and so $\liminf_n X_n = -\infty$ by Proposition 12.14.

(ii) Here $X_n \to \infty$ a.s. by the law of large numbers, and the formula $EX_{\tau_1} = E\tau_1 \, EX_1$ follows as before.

(iii) This is clear from the relations $X_{\tau_1} \geq X_1^+$ and $X_{\tau_1^-} \leq -X_1^-$. □

We proceed with a celebrated factorization, providing some more detailed information about the distributions of ladder times and heights. Here we write $\chi^\pm$ for the possibly defective distributions of the pairs (τ_1, X_{τ_1}) and $(\tau_1^-, X_{\tau_1^-})$, respectively, and let $\psi^\pm$ denote the corresponding distributions of (σ_1, X_{σ_1}) and $(\sigma_1^-, X_{\sigma_1^-})$. Put $\chi_n^\pm = \chi^\pm(\{n\} \times \cdot)$ and $\psi_n^\pm = \psi^\pm(\{n\} \times \cdot)$, and define a measure χ^0 on N by

$$\begin{aligned} \chi_n^0 &= P\Big\{X_1 \wedge \cdots \wedge X_{n-1} > 0 = X_n\Big\} \\ &= P\Big\{X_1 \vee \cdots \vee X_{n-1} < 0 = X_n\Big\}, \quad n \in \mathsf{N}, \end{aligned}$$

where the second equality holds by (4).

Theorem 12.16 *(Wiener–Hopf factorization) For a random walk in* R *based on a distribution* μ*, we have*

(i) $$\begin{aligned} \delta_0 - \delta_1 \otimes \mu &= (\delta_0 - \chi^+) * (\delta_0 - \psi^-) \\ &= (\delta_0 - \psi^+) * (\delta_0 - \chi^-), \end{aligned}$$

(ii) $$\delta_0 - \psi^\pm = (\delta_0 - \chi^\pm) * (\delta_0 - \chi^0).$$

Note that the convolutions in (i) are defined on the space $\mathsf{Z}_+ \times \mathsf{R}$, whereas those in (ii) can be regarded as defined on Z_+. Alternatively, we may regard χ^0 as a measure on $\mathsf{N} \times \{0\}$, and think of all convolutions as defined on $\mathsf{Z}_+ \times \mathsf{R}$.

Proof: (i) Define the measures $\rho_1, \rho_2, \ldots$ on $(0, \infty)$ by

$$\begin{aligned} \rho_n B &= P\Big\{X_1 \wedge \cdots \wedge X_{n-1} > 0,\ X_n \in B\Big\} \\ &= E \textstyle\sum_k 1\Big\{\tau_k = n,\ X_{\tau_k} \in B\Big\}, \quad n \in \mathsf{N},\ B \in \mathcal{B}_{(0,\infty)}, \end{aligned} \tag{9}$$

where the second equality holds by (8). Put $\rho_0 = \delta_0$, and regard the sequence $\rho = (\rho_n)$ as a measure on $\mathsf{Z}_+ \times (0, \infty)$. Noting that the corresponding measures on R equal $\rho_n + \psi_n^-$ and using the Markov property at time $n-1$, we get

$$\begin{aligned} \rho_n + \psi_n^- &= \rho_{n-1} * \mu \\ &= \Big\{\rho * (\delta_1 \otimes \mu)\Big\}_n, \quad n \in \mathsf{N}. \end{aligned} \tag{10}$$

Applying the strong Markov property at τ_1 to the second expression in (9), we see that also

$$\begin{aligned} \rho_n &= \sum_{k=1}^n \chi_k^+ * \rho_{n-k} \\ &= (\chi^+ * \rho)_n, \quad n \in \mathsf{N}. \end{aligned} \tag{11}$$

Recalling the values at zero, we get from (10) and (11)

$$\begin{aligned} \rho + \psi^- &= \delta_0 + \rho * (\delta_1 \otimes \mu), \\ \rho &= \delta_0 + \chi^+ * \rho. \end{aligned}$$

Eliminating ρ between the two equations yields the first relation in (i), and the second relation follows by symmetry.

(ii) Since the restriction of ψ^+ to $(0,\infty)$ equals $\psi_n^+ - \chi_n^0$, we get for any $B \in \mathcal{B}_{(0,\infty)}$

$$\left(\chi_n^+ - \psi_n^+ + \chi_n^0\right) B = P\left\{\max_{k<n} X_k = 0,\ X_n \in B\right\}.$$

Decomposing the event on the right according to the time of first return to 0, we get

$$\begin{aligned} \chi_n^+ - \psi_n^+ + \chi_n^0 &= \sum_{k=1}^{n-1} \chi_k^0\, \chi_{n-k}^+ \\ &= (\chi^0 * \chi^+)_n, \quad n \in \mathsf{N}, \end{aligned}$$

and so $\chi^+ - \psi^+ + \chi^0 = \chi^0 * \chi^+$, which is equivalent to the 'plus' version of (ii). The 'minus' version follows by symmetry. □

The preceding factorization yields an explicit formula for the joint distribution of first ladder time and height.

Theorem 12.17 *(ladder distributions, Sparre-Andersen, Baxter) Let X be a random walk in R. Then for $|s| < 1$ and $u \geq 0$, we have*

$$E\, s^{\tau_1} \exp(-\, uX_{\tau_1}) = 1 - \exp\left\{-\sum_{n\geq 1} \frac{s^n}{n} E\left(e^{-uX_n};\, X_n > 0\right)\right\}. \tag{12}$$

A similar relation holds for (σ_1, X_{σ_1}), with $X_n > 0$ replaced by $X_n \geq 0$.

Proof: Consider the mixed generating and characteristic functions

$$\begin{aligned} \hat\chi_{s,t}^+ &= E\, s^{\tau_1} \exp\left(itX_{\tau_1}\right), \\ \hat\psi_{s,t}^- &= E\, s^{\sigma_1^-} \exp\left(itX_{\sigma_1^-}\right), \end{aligned}$$

and note that the first relation in Theorem 12.16 (i) is equivalent to

$$1 - s\,\hat\mu_t = (1 - \hat\chi_{s,t}^+)(1 - \hat\psi_{s,t}^-), \quad |s| < 1,\ t \in \mathsf{R}.$$

Taking logarithms and expanding in Taylor series, we obtain

$$\textstyle\sum_n n^{-1}(s\,\hat\mu_t)^n = \sum_n n^{-1}(\hat\chi_{s,t}^+)^n + \sum_n n^{-1}(\hat\psi_{s,t}^-)^n.$$

For fixed $s \in (-1,1)$, this equation is of the form $\hat\nu = \hat\nu^+ + \hat\nu^-$, where ν and $\nu^\pm$ are bounded signed measures on R, $(0,\infty)$, and $(-\infty,0]$, respectively. By the uniqueness theorem for characteristic functions, we get $\nu = \nu^+ + \nu^-$. In particular, ν^+ equals the restriction of ν to $(0,\infty)$. Thus, the corresponding Laplace transforms agree, and (12) follows by summation of a Taylor series for the logarithm. A similar argument yields the formula for (σ_1, X_{σ_1}). □

The last result yields expressions for the probability that a random walk stays negative or non-positive, as well as criteria for its divergence to $-\infty$.

Corollary 12.18 *(boundedness and divergence) For a random walk X in R,*

(i) $\quad P\{\tau_1 = \infty\} = \dfrac{1}{E\sigma_1^-} = \exp\Big\{-\sum_{n\geq 1} n^{-1}P\{X_n > 0\}\Big\},$

(ii) $\quad P\{\sigma_1 = \infty\} = \dfrac{1}{E\tau_1^-} = \exp\Big\{-\sum_{n\geq 1} n^{-1}P\{X_n \geq 0\}\Big\},$

and the a.s. divergence $X_n \to -\infty$ is equivalent to each of the conditions:

$$\sum_{n\geq 1} n^{-1}P\{X_n > 0\} < \infty, \qquad \sum_{n\geq 1} n^{-1}P\{X_n \geq 0\} < \infty.$$

Proof: The last expression for $P\{\tau_1 = \infty\}$ follows from (12) with $u = 0$, as we let $s \to 1$. Similarly, the formula for $P\{\sigma_1 = \infty\}$ is obtained from the version of (12) for the pair (σ_1, X_{σ_1}). In particular, $P\{\tau_1 = \infty\} > 0$ iff the series in (i) converges, and similarly for the condition $P\{\sigma_1 = \infty\} > 0$ in terms of the series in (ii). Since both conditions are equivalent to $X_n \to -\infty$ a.s., the last assertion follows. Finally, the first equalities in (i) and (ii) are obtained most easily from Lemma 12.13, if we note that the number of strict or weak ladder times $\tau_n < \infty$ or $\sigma_n < \infty$ is geometrically distributed. □

We turn to a detailed study of the *occupation measure* $\xi = \sum_{n\geq 0} \delta_{X_n}$ of a transient random walk in R, based on the transition and initial distributions μ and ν. Recall from Theorem 12.1 that the associated intensity measure $E\xi = \nu * \sum_n \mu^{*n}$ is locally finite. By the strong Markov property, the sequence $(X_{\tau+n} - X_\tau)$ has the same distribution for every finite optional time τ. Thus, a similar invariance holds for the occupation measure, and the associated intensities agree. A *renewal* is then said to occur at time τ, and the whole subject is known as *renewal theory*. When μ and ν are supported by R_+, we refer to ξ as a *renewal process* based on μ and ν, and to $E\xi$ as the associated *renewal measure*. The standard choice is to take $\nu = \delta_0$, though in general we may also allow a non-degenerate *delay* distribution ν.

The occupation measure ξ is clearly a *random measure* on R, defined as a kernel from Ω to R. By Lemma 15.1 below, the distribution of a random measure on R_+ is determined by the distributions of the integrals $\xi f = \int f\, d\xi$, for all f belonging to the space $\hat{C}_+(\mathsf{R}_+)$ of continuous functions $f\colon \mathsf{R}_+ \to \mathsf{R}_+$ with bounded support. For any measure μ on R and constant $t \in \mathsf{R}$ we define $\theta_t\mu = \mu \circ \theta_t^{-1}$ with $\theta_t x = x + t$, so that $(\theta_t\mu)B = \mu(B - t)$ and $(\theta_t\mu)f = \mu(f \circ \theta_t)$ for any measurable set $B \subset \mathsf{R}$ and function $f \geq 0$. A random measure ξ is said to be *stationary* if $\theta_t\xi \overset{d}{=} \xi$ on R_+.

When ξ is a renewal process based on a transition distribution μ, the delayed process $\eta = \delta_\alpha * \xi$ is called a *stationary version* of ξ, if it is stationary on R_+ with the same spacing distribution μ. We show that such a version exists whenever μ has finite mean, in which case ν is uniquely determined by μ. Recall that λ denotes Lebesgue measure on R_+.

Theorem 12.19 *(stationary renewal process) Let ξ be a renewal process based on a distribution μ on R_+ with mean $m > 0$. Then*

(i) ξ *has a stationary version* η *on* R_+ *iff* $m < \infty$,

(ii) *the distribution of* η *is unique with* $E\eta = m^{-1}\lambda$ *and delay distribution*

$$\nu[0,t] = m^{-1}\int_0^t \mu(s,\infty)\,ds, \quad t \geq 0,$$

(iii) $\nu = \mu$ *iff* ξ *is a stationary Poisson process on* R_+.

Proof, (i)$-$(ii): For a delayed renewal process ξ based on μ and ν, Fubini's theorem yields

$$\begin{aligned} E\xi = E\sum_{n\geq 0}\delta_{X_n} &= \sum_{n\geq 0}\mathcal{L}(X_n) \\ &= \sum_{n\geq 0}\nu * \mu^{*n} \\ &= \nu + \mu * \sum_{n\geq 0}\nu * \mu^{*n} \\ &= \nu + \mu * E\xi, \end{aligned}$$

and so $\nu = (\delta_0 - \mu) * E\xi$. If ξ is stationary, then $E\xi$ is shift invariant, and Theorem 2.6 yields $E\xi = c\,\lambda$ for some constant $c > 0$. Thus, $\nu = c\,(\delta_0 - \mu) * \lambda$, and the asserted formula follows with m^{-1} replaced by c. As $t \to \infty$, we get $1 = c\,m$ by Lemma 4.4, which implies $m < \infty$ and $c = m^{-1}$.

Conversely, let $m < \infty$, and choose ν as stated. Then

$$\begin{aligned} E\xi &= \nu * \sum_{n\geq 0}\mu^{*n} \\ &= m^{-1}(\delta_0 - \mu) * \lambda * \sum_{n\geq 0}\mu^{*n} \\ &= m^{-1}\lambda * \Big(\sum_{n\geq 0}\mu^{*n} - \sum_{n\geq 1}\mu^{*n}\Big) = m^{-1}\lambda, \end{aligned}$$

which shows that $E\xi$ is invariant. By the strong Markov property, the shifted random measure $\theta_t\xi$ is again a renewal process based on μ, say with delay distribution ν_t. Since $E\,\theta_t\xi = \theta_t E\xi$, we get as before

$$\begin{aligned} \nu_t &= (\delta_0 - \mu) * (\theta_t E\xi) \\ &= (\delta_0 - \mu) * E\xi \,=\, \nu, \end{aligned}$$

which implies $\theta_t\xi \overset{d}{=} \xi$, showing that ξ is stationary.

(iii) By Theorem 13.6 below, ξ is a stationary Poisson process on R_+ with rate $c > 0$, iff it is a delayed renewal process based on the exponential distribution with mean c^{-1}, in which case $\mu = \nu$. Conversely, let ξ be a stationary renewal process based on a common distribution $\mu = \nu$ with mean $m < \infty$. Then the tail probability $f(t) = \mu(t,\infty)$ satisfies the differential equation $mf' + f = 0$ with initial condition $f(0) = 1$, and so $f(t) = e^{-t/m}$, which shows that $\mu = \nu$ is exponential with mean m. As before, ξ is then a stationary Poisson process with rate m^{-1}. □

The last result yields a similar statement for the occupation measure of a general random walk.

Corollary 12.20 *(stationary occupation measure) Let ξ be the occupation measure of a random walk X in R based on distributions μ, ν, where μ has mean $m \in (0,\infty)$, and ν is defined as in Theorem 12.19 in terms of the ladder height distribution $\tilde\mu$ and its mean $\tilde m$. Then ξ is stationary on R_+ with intensity m^{-1}.*

Proof: Since $X_n \to \infty$ a.s., Propositions 12.14 and 12.15 show that the ladder times τ_n and heights $H_n = X_{\tau_n}$ have finite mean, and by Proposition 12.19 the renewal process $\zeta = \sum_n \delta_{H_n}$ is stationary for the prescribed choice of ν. Fixing $t \ge 0$ and putting $\sigma_t = \inf\{n \in \mathsf{Z}_+;\ X_n \ge t\}$, we note in particular that $X_{\sigma_t} - t$ has distribution ν. By the strong Markov property at σ_t, the sequence $X_{\sigma_t+n} - t$, $n \in \mathsf{Z}_+$, has then the same distribution as X. Since $X_k < t$ for $k < \sigma_t$, we get $\theta_t \xi \overset{d}{=} \xi$ on R_+, which proves the asserted stationarity.

To identify the intensity, let ξ_n be the occupation measure of the sequence $X_k - H_n$, $\tau_n \le k < \tau_{n+1}$, and note that $H_n \perp\!\!\!\perp \xi_n \overset{d}{=} \xi_0$ for each n by the strong Markov property. Hence, Fubini's theorem yields

$$\begin{aligned} E\xi &= E\sum_{n\ge 0} \xi_n * \delta_{H_n} \\ &= \sum_{n\ge 0} E(\delta_{H_n} * E\,\xi_n) \\ &= E\xi_0 * E\sum_{n\ge 0} \delta_{H_n} \\ &= E\,\xi_0 * E\zeta. \end{aligned}$$

Noting that $E\zeta = \tilde m^{-1}\lambda$ by Proposition 12.19, that $E\,\xi_0(0,\infty) = 0$, and that $\tilde m = m\,E\tau_1$ by Proposition 12.15, we get on R_+

$$E\,\xi = \frac{E\,\xi_0 \mathsf{R}_-}{\tilde m}\,\lambda = \frac{E\tau_1}{\tilde m}\,\lambda = m^{-1}\lambda. \qquad \square$$

We proceed to study the asymptotic behavior of the occupation measure ξ and its intensity $E\xi$. Under weak restrictions on μ, we show that $\theta_t\xi$ approaches the stationary version $\tilde\xi$, whereas $\theta_t\,E\xi$ is asymptotically proportional to Lebesgue measure. For simplicity, we assume that the mean of μ exists in $\bar{\mathsf{R}} = [-\infty,\infty]$.

We will use the standard notation for convergence on measure spaces.[4] Thus, for locally finite measures $\nu, \nu_1, \nu_2, \dots$ on R_+, the vague convergence $\nu_n \xrightarrow{v} \nu$ means that $\nu_n f \to \nu f$ for all $f \in \hat C_+(\mathsf{R}_+)$. Similarly, for random measures $\xi, \xi_1, \xi_2, \dots$ on R_+, the distributional convergence $\xi_n \xrightarrow{vd} \xi$ is defined by the condition $\xi_n f \xrightarrow{d} \xi f$ for every $f \in \hat C_+(\mathsf{R}_+)$. A measure μ on R is said to be *non-arithmetic*, if the additive subgroup generated by $\operatorname{supp}\mu$ is dense in R.

Theorem 12.21 *(two-sided renewal theorem, Blackwell, Feller & Orey) Let ξ be the occupation measure of a random walk X in R based on distributions μ and ν, where μ is non-arithmetic with mean $m \in \bar{\mathsf{R}} \setminus \{0\}$. When $m \in (0,\infty)$, let $\tilde\xi$ be the stationary version in Proposition 12.20, and otherwise put $\tilde\xi = 0$. Then as $t \to \infty$,*

[4] For a detailed discussion of such convergence, see Chapter 23 and Appendix 5.

(i) $\theta_t \xi \xrightarrow{vd} \tilde{\xi}$,

(ii) $\theta_t E\xi \xrightarrow{v} E\tilde{\xi} = (m^{-1} \vee 0)\, \lambda$.

Our proof will be based on two further lemmas. When $m < \infty$, the crucial step is to prove convergence of the ladder height distributions of the sequences $X - t$. This will be accomplished by a coupling argument:

Lemma 12.22 *(asymptotic delay) For X as in Theorem 12.21 with $m \in (0,\infty)$, let ν_t be the distribution of the first ladder height ≥ 0 of the sequence $X - t$. Then $\nu_t \xrightarrow{w} \tilde{\nu}$ as $t \to \infty$.*

Proof: Let α and α' be independent random variables with distributions ν and $\tilde{\nu}$. Choose some i.i.d. sequences $(\xi_k) \perp\!\!\!\perp (\vartheta_k)$ independent of α and α', such that $\mathcal{L}(\xi_k) = \mu$ and $P\{\vartheta_k = \pm 1\} = \frac{1}{2}$. Then

$$M_n = \alpha' - \alpha - \sum_{k \leq n} \vartheta_k\, \xi_k, \quad n \in \mathsf{Z}_+,$$

is a random walk based on a non-arithmetic distribution with mean 0. By Theorems 12.1 and 12.2, the range $\{M_n\}$ is then a.s. dense in R, and so for any $\varepsilon > 0$ the optional time $\sigma = \inf\{n \geq 0;\, M_n \in [0,\varepsilon]\}$ is a.s. finite.

Next define $\vartheta'_k = \pm\vartheta_k$, with plus sign iff $k > \tau$, and note as in Lemma 12.11 that $\{\alpha', (\xi_k, \vartheta'_k)\} \overset{d}{=} \{\alpha', (\xi_k, \vartheta_k)\}$. Let $\kappa_1, \kappa_2, \ldots$ and $\kappa'_1, \kappa'_2, \ldots$ be the values of k where $\vartheta_k = 1$ or $\vartheta'_k = 1$, respectively. The sequences

$$\begin{aligned} X_n &= \alpha + \sum_{j \leq n} \xi_{\kappa_j}, \\ X'_n &= \alpha' + \sum_{j \leq n} \xi_{\kappa'_j}, \qquad n \in \mathsf{Z}_+, \end{aligned}$$

are again random walks based on μ with initial distributions ν and ν', respectively. Writing $\sigma_\pm = \sum_{k \leq \tau} (\pm\vartheta_k \vee 0)$, we note that

$$X'_{\sigma_- + n} - X_{\sigma_+ + n} = M_\tau \in [0,\varepsilon], \quad n \in \mathsf{Z}_+.$$

Putting $\beta = X^*_{\sigma_+} \vee X'^*_{\sigma_-}$, and considering the first entries of X and X' into the interval $[t,\infty)$, we obtain for any $r \geq \varepsilon$

$$\begin{aligned} \tilde{\nu}[\varepsilon, r] - P\{\beta \geq t\} &\leq \nu_t[0,r] \\ &\leq \tilde{\nu}[0, r+\varepsilon] + P\{\beta \geq t\}. \end{aligned}$$

Letting $t \to \infty$ and then $\varepsilon \to 0$, and noting that $\tilde{\nu}\{0\} = 0$ by stationarity, we get $\nu_t[0,r] \to \tilde{\nu}[0,r]$, which implies $\nu_t \xrightarrow{w} \tilde{\nu}$. □

To prove (i) ⇒ (ii) in the main theorem, we need the following technical result, which also plays a crucial role when $m = \infty$.

Lemma 12.23 *(uniform integrability) Let ξ be the occupation measure of a transient random walk X in R^d with arbitrary initial distribution, and fix a bounded set $B \in \mathcal{B}^d$. Then the random variables $\xi(B + x)$, $x \in \mathsf{R}^d$, are uniformly integrable.*

Proof: Fix any $x \in \mathsf{R}^d$, and put $\tau = \inf\{t \geq 0;\, X_n \in B + x\}$. Writing η for the occupation measure of an independent random walk starting at 0, we get by the strong Markov property

$$\begin{aligned} \xi(B+x) &\overset{d}{=} \eta(B + x - X_\tau)\, 1\{\tau < \infty\} \\ &\leq \eta(B - B). \end{aligned}$$

Finally, note that $E\eta(B - B) < \infty$ by Theorem 12.1, since X is transient. □

Proof of Theorem 12.21 ($m < \infty$): By Lemma 12.23 it is enough to prove (i). If $m < 0$, then $X_n \to -\infty$ a.s. by the law of large numbers, and so $\theta_t \xi = 0$ for sufficiently large t, and (i) follows. If instead $m \in (0, \infty)$, then $\nu_t \xrightarrow{w} \tilde{\nu}$ by Lemma 12.22, and we may choose some random variables α_t and α with distributions ν_t and ν, respectively, such that $\alpha_t \to \alpha$ a.s. We also introduce the occupation measure ξ_0 of an independent random walk starting at 0.

Now let $f \in \hat{C}_+(\mathsf{R}_+)$, and extend f to R by putting $f(x) = 0$ for $x < 0$. Since $\tilde{\nu} \ll \lambda$, we have $\xi_0\{-\alpha\} = 0$ a.s. Hence, by the strong Markov property and dominated convergence,

$$\begin{aligned} (\theta_t \xi) f &\overset{d}{=} \int f(\alpha_t + x)\, \xi_0(dx) \\ &\to \int f(\alpha + x)\, \xi_0(dx) \overset{d}{=} \tilde{\xi} f. \end{aligned}$$

($m = \infty$): Here it is clearly enough to prove (ii). The strong Markov property yields $E\,\xi = \nu * E\chi * E\zeta$, where χ is the occupation measure of the ladder height sequence of $(X_n - X_0)$ and ζ is the occupation measure of the same process, prior to the first ladder time. Here $E\zeta \mathsf{R}_- < \infty$ by Proposition 12.14, and so by dominated convergence it suffices to show that $\theta_t E\chi \xrightarrow{v} 0$. Since the ladder heights have again infinite mean by Proposition 12.15, we may assume instead that $\nu = \delta_0$, and let μ be an arbitrary distribution on R_+ with mean $m = \infty$.

Put $I = [0, 1]$, and note that $E\,\xi(I + t)$ is bounded by Lemma 12.23. Define $b = \limsup_t E\,\xi(I + t)$ and choose some $t_k \to \infty$ with $E\,\xi(I + t_k) \to b$. Subtracting the bounded measures μ^{*j} for $j < n$, we get $(\mu^{*n} * E\,\xi)(I + t_k) \to b$ for all $n \in \mathsf{Z}_+$. Using the reverse Fatou lemma, we obtain for any $B \in \mathcal{B}_{\mathsf{R}_+}$

$$\begin{aligned} &\liminf_{k\to\infty} E\,\xi(I - B + t_k)\, \mu^{*n} B \\ &\qquad \geq \liminf_{k\to\infty} \int_B E\,\xi(I - x + t_k)\, \mu^{*n}(dx) \\ &\qquad = b - \limsup_{k\to\infty} \int_{B^c} E\,\xi(I - x + t_k)\, \mu^{*n}(dx) \\ &\qquad \geq b - \int_{B^c} \limsup_{k\to\infty} E\,\xi(I - x + t_k)\, \mu^{*n}(dx) \\ &\qquad \geq b - \int_{B^c} b\, \mu^{*n}(dx) = b\, \mu^{*n} B. \end{aligned}$$

Since n was arbitrary, the 'lim inf' on the left is then $\geq b$ for every B with $E\,\xi B > 0$.

Now fix any $h > 0$ with $\mu(0, h] > 0$, and write $J = [0, a]$ with $a = h + 1$. Noting that $E\,\xi[r, r + h] > 0$ for all $r \ge 0$, we get

$$\liminf_{k\to\infty} E\xi(J + t_k - r) \ge b, \quad r \ge a.$$

Since $\delta_0 = (\delta_0 - \mu) * E\,\xi$, we further note that

$$\begin{aligned} 1 &= \int_0^{t_k} \mu(t_k - x, \infty)\, E\,\xi(dx) \\ &\ge \sum_{n\ge 1} \mu(na, \infty)\, E\xi(J + t_k - na). \end{aligned}$$

As $k \to \infty$, we get $1 \ge b \sum_{n\ge 1} \mu(na, \infty)$ by Fatou's lemma. Here the sum diverges since $m = \infty$, which implies $\theta_t E\,\xi \xrightarrow{v} 0$. □

The preceding theory may be used to study the *renewal equation* $F = f + F * \mu$, which often arises in applications. Here the convolution $F * \mu$ is defined by

$$(F * \mu)_t = \int_0^t F(t - s)\, \mu(ds), \quad t \ge 0,$$

whenever the integrals on the right exist. Under suitable regularity conditions, the stated equation has the unique solution $F = f * \bar\mu$, where $\bar\mu$ denotes the *renewal measure* $\sum_{n\ge 0} \mu^{*n}$. Additional conditions ensure convergence at ∞ of the solution F.

The precise statements require some further terminology. By a *regular step function* we mean a function on R_+ of the form

$$f_t = \sum_{j\ge 1} a_j 1_{[j-1,j)}(t/h), \quad t \ge 0, \tag{13}$$

where $h > 0$ and $a_1, a_2, \ldots \in \mathsf{R}$. A measurable function f on R_+ is said to be *directly Riemann integrable*, if $\lambda|f| < \infty$, and there exist some regular step functions $f_n^\pm$ with $f_n^- \le f \le f_n^+$ and $\lambda(f_n^+ - f_n^-) \to 0$.

Theorem 12.24 *(renewal equation) Consider a distribution $\mu \ne \delta_0$ on R_+ with associated renewal measure $\bar\mu$ and a locally bounded, measurable function f on R_+. Then*

(i) *equation $F = f + F * \mu$ has the unique, locally bounded solution $F = f * \bar\mu$,*

(ii) *when f is directly Riemann integrable and μ is non-arithmetic with mean m, we have*

$$F_t \to m^{-1}\lambda f, \quad t \to \infty.$$

Proof: (i) Iterating the renewal equation gives

$$F = \sum_{k<n} f * \mu^{*k} + F * \mu^{*n}, \quad n \in \mathsf{N}. \tag{14}$$

Since $\mu^{*n}[0, t] \to 0$ as $n \to \infty$ for fixed $t \ge 0$ by the weak law of large numbers, we have $F * \mu^{*n} \to 0$ for any locally bounded F. If even f is locally bounded, then by (14) and Fubini's theorem,

$$\begin{aligned} F &= \sum_{k\ge 0} f * \mu^{*k} \\ &= f * \sum_{k\ge 0} \mu^{*k} = f * \bar\mu. \end{aligned}$$

Conversely, $f + f * \bar\mu * \mu = f * \bar\mu$, which shows that $F = f * \bar\mu$ solves the given equation.

(ii) Let μ be non-arithmetic. If f is a regular step function as in (13), then by Theorem 12.21 and dominated convergence, we get as $t \to \infty$

$$\begin{aligned} F_t &= \int_0^t f(t-s)\,\bar\mu(ds) \\ &= \sum_{j\geq 1} a_j\,\bar\mu\big\{(0,h] + t - jh\big\} \\ &\to m^{-1}h \sum_{j\geq 1} a_j = m^{-1}\lambda f. \end{aligned}$$

In general, we may introduce some regular step functions $f_n^\pm$ with $f_n^- \leq f \leq f_n^+$ and $\lambda(f_n^+ - f_n^-) \to 0$, and note that

$$(f_n^- * \bar\mu)_t \leq F_t \leq (f_n^+ * \bar\mu)_t, \quad t \geq 0,\ n \in \mathsf{N}.$$

Letting $t \to \infty$ and then $n \to \infty$, we obtain $F_t \to m^{-1}\lambda f$. □

Exercises

1. Show that if X is recurrent, then so is the random walk (X_{nk}) for every $k \in \mathsf{N}$. (*Hint:* If (X_{nk}) is transient, then so is (X_{nk+j}) for every $j > 0$.)

2. For any non-degenerate random walk X in R^d, show that $|X_n| \xrightarrow{P} \infty$. (*Hint:* Use Lemma 6.1.)

3. Let X be a random walk in R, based on a symmetric, non-degenerate distribution with bounded support. Show that X is recurrent. (*Hint:* Recall that $\limsup_n(\pm X_n) = \infty$ a.s.)

4. Show that the accessible set A equals the closed semi-group generated by $\operatorname{supp}\mu$. Also show by examples that A may or may not be a group.

5. Let ν be an invariant measure on the accessible set of a recurrent random walk in R^d. Show by examples that $E\xi$ may or may not be of the form $\infty \cdot \nu$.

6. Show that a non-degenerate random walk in R^d has no invariant distribution. (*Hint:* If ν is invariant, then $\mu * \nu = \nu$.)

7. Show by examples that the conditions in Theorem 12.2 are not necessary. (*Hint:* For $d = 2$, consider mixtures of $N(0, \sigma^2)$ and use Lemma 6.19.)

8. Let X be a random walk in R based on the symmetric, p-stable distribution with characteristic function $e^{-|t|^p}$. Show that X is recurrent for $p \geq 1$ and transient for $p < 1$.

9. Let X be a random walk in R^2 based on the distribution μ^2, where μ is symmetric p-stable. Show that X is recurrent for $p = 2$ and transient for $p < 2$.

10. Let $\mu = c\,\mu_1 + (1-c)\,\mu_2$ for some symmetric distributions μ_1, μ_2 on R^d and a constant $c \in (0,1)$. Show that a random walk based on μ is recurrent iff recurrence holds for random walks based on μ_1 and μ_2.

11. Let $\mu = \mu_1 * \mu_2$, where μ_1, μ_2 are symmetric distributions on R^d. Show that if a random walk based on μ is recurrent, then so are the random walks based on

μ_1, μ_2. Also show by an example that the converse is false. (*Hint:* For the latter part, let μ_1, μ_2 be supported by orthogonal sub-spaces.)

12. For a symmetric, recurrent random walk in Z^d, show that the expected number of visits to an accessible state $k \neq 0$ before return to the origin equals 1. (*Hint:* Compute the distribution, assuming a probability p for return before visit to k.)

13. Use Proposition 12.14 to show that any non-degenerate random walk in Z^d has an infinite mean recurrence time. Compare with the preceding problem.

14. Show how part (i) of Proposition 12.15 can be strengthened by means of Theorems 6.17 and 12.2.

15. For a non-degenerate random walk in R, show that $\limsup_n X_n = \infty$ a.s. iff $\sigma_1 < \infty$ a.s., and that $X_n \to \infty$ a.s. iff $E\sigma_1 < \infty$. In both conditions, note that σ_1 can be replaced by τ_1.

16. Let ξ be a renewal process based on a non-arithmetic distribution on R_+. Show that $\sup\{t > 0;\, E\xi[t, t+\varepsilon] = 0\} < \infty$ for any $\varepsilon > 0$. (*Hint:* Mimic the proof of Proposition 11.18.)

17. Let μ be a distribution on Z_+, such that the group generated by $\operatorname{supp}\mu$ equals Z. Show that Proposition 12.19 remains valid with $\nu\{n\} = c^{-1}\mu(n, \infty)$, $n \geq 0$, and prove a corresponding version of Proposition 12.20.

18. Let ξ be the occupation measure of a random walk in Z based on a distribution μ with mean $m \in \bar{\mathsf{R}} \setminus \{0\}$, such that the group generated by $\operatorname{supp}\mu$ equals Z. Show as in Theorem 12.21 that $E\xi\{n\} \to m^{-1} \vee 0$.

19. Prove the renewal theorem for random walks in Z_+ from the ergodic theorem for discrete-time Markov chains, and conversely. (*Hint:* Given a distribution μ on N, construct a Markov chain X in Z_+ with $X_{n+1} = X_n + 1$ or 0, such that the recurrence times at 0 are i.i.d. μ. Note that X is aperiodic iff Z is the smallest group containing $\operatorname{supp}\mu$.)

20. Fix a distribution μ on R with symmetrization $\tilde{\mu}$. Note that if $\tilde{\mu}$ is non-arithmetic, then so is μ. Show by an example that the converse is false.

21. Simplify the proof of Lemma 12.22, in the case where even the symmetrization $\tilde{\mu}$ is non-arithmetic. (*Hint:* Let $\xi_1, \xi_2, \ldots$ and $\xi_1', \xi_2', \ldots$ be i.i.d. μ, and define $\tilde{X}_n = \alpha' - \alpha + \sum_{k \leq n} (\xi_k' - \xi_k)$.)

22. Show that any monotone, Lebesgue integrable function on R_+ is directly Riemann integrable.

23. State and prove the counterpart of Corollary 12.24 for arithmetic distributions.

24. Let (ξ_n), (η_n) be independent i.i.d. sequences with distributions μ, ν, put $X_n = \sum_{k \leq n}(\xi_k + \eta_k)$, and define $U = \bigcup_{n \geq 0} [X_n, X_n + \xi_{n+1})$. Show that $F_t = P\{t \in U\}$ satisfies the renewal equation $F = f + F * \mu * \nu$ with $f_t = \mu(t, \infty)$. For μ, ν with finite means, show also that F_t converges as $t \to \infty$, and identify the limit.

25. Consider a renewal process ξ based on a non-arithmetic distribution μ with mean $m < \infty$, fix any $h > 0$, and define $F_t = P\{\xi[t, t+h] = 0\}$. Show that $F = f + F * \mu$ with $f_t = \mu(t + h, \infty)$. Further show that F_t converges as $t \to \infty$, and identify the limit. (*Hint:* Consider the first point of ξ in $(0, t)$, if any.)

26. For ξ as above, let $\tau = \inf\{t \geq 0;\, \xi[t, t+h] = 0\}$, and put $F_t = P\{\tau \leq t\}$.

Show that $F_t = \mu(h,\infty) + \int_0^{h\wedge t} \mu(ds)\, F_{t-s}$, or $F = f + F * \mu_h$, where $\mu_h = 1_{[0,h]} \cdot \mu$ and $f \equiv \mu(h,\infty)$.

Chapter 13

Jump-Type Chains and Branching Processes

Strong Markov property, renewals in Poisson process, rate kernel, embedded Markov chain, existence and explosion, compound and pseudo-Poisson processes, backward equation, invariant distribution, convergence dichotomy, mean recurrence times, Bienaymé process, extinction and asymptotics, binary splitting, Brownian branching tree, genealogy and Yule tree, survival rate and distribution, diffusion approximation

After our detailed study of discrete-time Markov chains in Chapter 11, we turn our attention to the corresponding processes in continuous time, where the paths are taken to be piecewise constant apart from isolated jumps. The evolution of the process is then governed by a *rate kernel* α, determining both the rate at which transitions occur and the associated transition probabilities. For bounded α we get a *pseudo-Poisson* process, which may be described as a discrete-time Markov chain with transition times given by an independent, homogeneous Poisson process.

Of special interest is the space-homogeneous case of *compound Poisson* processes, where the underlying Markov chain is a random walk. In Chapter 17 we show how every Feller process can be approximated in a natural way by a sequence of pseudo-Poisson processes, characterized by the boundedness of their generators. A similar compound Poisson approximation of general Lévy processes plays a basic role in Chapter 16.

The chapter ends with a brief introduction to branching processes in discrete and continuous time. In particular, we study the extinction probability and asymptotic population size, and identify the ancestral structure of a critical Brownian branching tree in terms of a suitable Yule process. For a critical Bienaymé process, we further derive the asymptotic survival rate and distribution, and indicate how a suitably scaled version can be approximated by a Feller diffusion. The study of spatial branching processes will be resumed in Chapter 30.

A process X in a measurable space $(S, \mathcal{S})$ is said to be of *pure jump-type*, if its paths are a.s. right-continuous and constant apart from isolated jumps. We may then denote the jump times of X by $\tau_1, \tau_2, \ldots$, with the understanding that $\tau_n = \infty$ if there are fewer than n jumps. By a simple approximation based on Lemma 9.3, the times τ_n are optional with respect to the right-continuous filtration $\mathcal{F} = (\mathcal{F}_t)$ induced by X. For convenience we may choose X to be the identity map on the canonical path space Ω. When X is Markov, we write P_x

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_14

for the distribution with initial state x, and note that the mapping $x \mapsto P_x$ is a kernel from $(S, \mathcal{S})$ to $(\Omega, \mathcal{F}_\infty)$.

We begin our study of pure jump-type Markov processes by proving an extension of the elementary strong Markov property in Proposition 11.9. A further extension appears as Theorem 17.17.

Theorem 13.1 *(strong Markov property, Doob) A pure jump-type Markov process satisfies the strong Markov property at every optional time.*

Proof: For any optional time τ, we may choose some optional times $\sigma_n \geq \tau + 2^{-n}$ taking countably many values, such that $\sigma_n \to \tau$ a.s. When $A \in \mathcal{F}_\tau \cap \{\tau < \infty\}$ and $B \in \mathcal{F}_\infty$, we get by Proposition 11.9

$$P\Big\{\theta_{\sigma_n} X \in B; A\Big\} = E\Big(P_{X_{\sigma_n}} B; A\Big). \tag{1}$$

The right-continuity of X yields $P\{X_{\sigma_n} \neq X_\tau\} \to 0$. If B depends on finitely many coordinates, it is also clear that

$$P\Big(\{\theta_{\sigma_n} X \in B\} \triangle \{\theta_\tau X \in B\}\Big) \to 0, \quad n \to \infty.$$

Hence, (1) remains true for such sets B with σ_n replaced by τ, and the relation extends to the general case by a monotone-class argument. □

We now examine the structure of a general pure jump-type Markov process. Here the crucial step is to determine the distributions associated with the first jump, occurring at time τ_1. A random variable $\gamma \geq 0$ is said to be *exponentially distributed* if $P\{\gamma > t\} = e^{ct}$, $t \geq 0$, for a constant $c > 0$, in which case $E\gamma = \int_0^\infty e^{-ct} dt = c^{-1}$. Say that a state $x \in S$ is *absorbing* if

$$P_x\{X \equiv x\} = P_x\{\tau_1 = \infty\} = 1.$$

Lemma 13.2 *(first jump) Let X be a pure jump-type Markov process, and fix a non-absorbing state x. Then under P_x,*

(i) τ_1 *is exponentially distributed,*

(ii) $\theta_{\tau_1} X \perp\!\!\!\perp \tau_1$.

Proof: (i) Put $\tau_1 = \tau$. By the Markov property at fixed times, we get for any $s, t \geq 0$

$$\begin{aligned} P_x\Big\{\tau > s + t\Big\} &= P_x\Big\{\tau > s,\ \tau \circ \theta_s > t\Big\} \\ &= P_x\{\tau > s\}\, P_x\{\tau > t\}. \end{aligned}$$

This is a Cauchy equation in s and t, whose only non-increasing solutions are of the form $P_x\{\tau > t\} = e^{-ct}$ with $c \in [0, \infty]$. Since x is non-absorbing and $\tau > 0$ a.s., we have $c \in (0, \infty)$, and so τ is exponentially distributed with parameter c.

(ii) By the Markov property at fixed times, we get for any $B \in \mathcal{F}_\infty$

$$\begin{aligned} P_x\Big\{\tau > t,\, \theta_\tau X \in B\Big\} &= P_x\Big\{\tau > t, (\theta_\tau X)\circ\theta_t \in B\Big\} \\ &= P_x\{\tau > t\}\, P_x\Big\{\theta_\tau X \in B\Big\}, \end{aligned}$$

which shows that $\tau \perp\!\!\!\perp \theta_\tau X$. □

Writing $X_\infty = x$ when X eventually gets absorbed at x, we define the *rate function* c and *jump transition kernel* μ by

$$\begin{aligned} c(x) &= (E_x \tau_1)^{-1}, \\ \mu(x,B) &= P_x\{X_{\tau_1} \in B\}, \qquad x \in S,\ B \in \mathcal{S}, \end{aligned}$$

combined into a *rate kernel* $\alpha = c\,\mu$, or

$$\alpha(x,B) = c(x)\,\mu(x,B), \quad x \in S,\ B \in \mathcal{S}.$$

The required measurability follows from that for the kernel (P_x). Assuming in addition that $\mu(x,\cdot) = \delta_x$ when $\alpha(x,\cdot) = 0$, we may reconstruct μ from α, so that μ becomes a measurable function of α.

We proceed to show how X can be represented in terms of a discrete-time Markov chain and an independent sequence of exponential random variables. In particular, we will see how the distributions P_x are uniquely determined by the rate kernel α. In the sequel, we always assume the existence of the required randomization variables.

Theorem 13.3 *(embedded Markov chain) For a pure jump-type Markov process X with rate kernel $\alpha = c\,\mu$, there exist a Markov chain Y on Z_+ with transition kernel μ and some i.i.d., exponential random variables $\gamma_1, \gamma_2, \ldots \perp\!\!\!\perp Y$ with mean* 1*, such that a.s. for all $n \in \mathsf{Z}_+$,*

(i) $X_t = Y_n, \quad t \in [\tau_n, \tau_{n+1})$,

(ii) $\tau_n = \sum_{k=1}^{n} \dfrac{\gamma_k}{c(Y_{k-1})}$.

Proof: Let $\tau_1, \tau_2, \ldots$ be the jump times of X, and put $\tau_0 = 0$, so that (i) holds with $Y_n = X_{\tau_n}$ for $n \in \mathsf{Z}_+$. Introduce some i.i.d., exponentially distributed random variables $\gamma_1', \gamma_2', \ldots \perp\!\!\!\perp X$ with mean 1, and define for $n \in \mathsf{N}$

$$\gamma_n = (\tau_n - \tau_{n-1})\, c(Y_n)\, 1\{\tau_{n-1} < \infty\} + \gamma_n'\, 1\Big\{c(Y_n) = 0\Big\}.$$

By Lemma 13.2, we get for any $t \ge 0$, $B \in \mathcal{S}$, and $x \in S$ with $c(x) > 0$

$$\begin{aligned} P_x\Big\{\gamma_1 > t,\, Y_1 \in B\Big\} &= P_x\Big\{\tau_1 c(x) > t,\, Y_1 \in B\Big\} \\ &= e^{-t}\mu(x,B), \end{aligned}$$

which clearly remains true when $c(x) = 0$. By the strong Markov property, we obtain for any n, a.s. on $\{\tau_n < \infty\}$,

$$\begin{aligned} P_x\Big\{ \gamma_{n+1} > t,\, Y_{n+1} \in B \,\Big|\, \mathcal{F}_{\tau_n}\Big\} &= P_{Y_n}\Big\{\gamma_1 > t,\, Y_1 \in B\Big\} \\ &= e^{-t}\mu(Y_n, B). \end{aligned} \tag{2}$$

The strong Markov property also gives $\tau_{n+1} < \infty$ a.s. on the set $\{\tau_n < \infty, c(Y_n) > 0\}$. Arguing recursively, we get $\{c(Y_n) = 0\} = \{\tau_{n+1} = \infty\}$ a.s., and (ii) follows. The same relation shows that (2) remains a.s. true on $\{\tau_n = \infty\}$, and in both cases we may replace $\mathcal{F}_{\tau_n}$ by $\mathcal{G}_n = \mathcal{F}_{\tau_n} \vee \sigma\{\gamma_1', \dots, \gamma_n'\}$. Thus, the pairs (γ_n, Y_n) form a discrete-time Markov chain with the desired transition kernel. By Proposition 11.2, the latter property along with the initial distribution determine uniquely the joint distribution of Y and (γ_n). □

In applications we are often faced with the converse problem of constructing a Markov process X with a given rate kernel α. Then write $\alpha(x, B) = c(x)\,\mu(x, B)$ for some rate function $c\colon S \to \mathsf{R}_+$ and transition kernel μ on S, such that $\mu(x, \cdot) = \delta_x$ when $c(x) = 0$ and otherwise $\mu(x, \{x\}) = 0$. If X does exist, it may clearly be constructed as in Theorem 13.3. The construction fails when $\zeta \equiv \sup_n \tau_n < \infty$, in which case an *explosion* is said to occur at time ζ.

Theorem 13.4 *(synthesis) For a finite kernel $\alpha = c\,\mu$ on S with $\alpha(x, \{x\}) \equiv 0$, let Y be a Markov chain in S with transition kernel μ, and let $\gamma_1, \gamma_2, \ldots \perp\!\!\!\perp Y$ be i.i.d., exponential random variables with mean* 1. *Then these conditions are equivalent:*

(i) $\sum_n \gamma_n / c(Y_{n-1}) = \infty$ *a.s. under every initial distribution for* Y,

(ii) *there exists a pure jump-type Markov process on* R_+ *with rate kernel* α.

Proof: Assuming (i), let P_x be the joint distribution of the sequences $Y = (Y_n)$ and $\Gamma = (\gamma_n)$ when $Y_0 = x$. Regarding (Y, Γ) as the identity map on the canonical space $\Omega = (S \times \mathsf{R}_+)^\infty$, we may construct X from (Y, Γ) as in Theorem 13.3, where $X_t = x_0$ is arbitrary for $t \ge \sup_n \tau_n$, and introduce the filtrations $\mathcal{G} = (\mathcal{G}_n)$ induced by (Y, γ) and $\mathcal{F} = (\mathcal{F}_t)$ induced by X. We need to verify the Markov property $\mathcal{L}_x(\,\theta_t X \mid \mathcal{F}_t) = \mathcal{L}_{X_t}(X)$ for the distributions under P_x and P_{X_t}, since the rate kernel can then be identified via Theorem 13.3.

For any $t \ge 0$ and $n \in \mathsf{Z}_+$, put

$$\kappa = \sup\{k;\, \tau_k \le t\}, \qquad \beta = (t - \tau_n)\, c(Y_n),$$

and define

$$\begin{aligned} T^m(Y, \Gamma) &= \Big\{(Y_k, \gamma_{k+1});\, k \ge m\Big\}, \\ (Y', \Gamma') &= T^{n+1}(Y, \Gamma), \qquad \gamma' = \gamma_{n+1}. \end{aligned}$$

Since clearly

$$\mathcal{F}_t = \mathcal{G}_n \vee \sigma\{\gamma' > \beta\} \;\text{ on }\; \{\kappa = n\},$$

it suffices by Lemma 8.3 to prove that

$$\mathcal{L}_x\Big\{ (Y',\Gamma');\, \gamma'-\beta>r \,\Big|\, \mathcal{G}_n, \gamma'>\beta \Big\} = \mathcal{L}_{Y_n}\Big\{ T(Y,\Gamma);\, \gamma_1>r \Big\}.$$

Noting that $(Y',\Gamma') \perp\!\!\!\perp_{\mathcal{G}_n} (\gamma',\beta)$ since $\gamma' \perp\!\!\!\perp (\mathcal{G}_n, Y', \Gamma')$, we see that the left-hand side equals

$$\begin{aligned}\frac{\mathcal{L}_x\Big\{(Y',\Gamma');\, \gamma'-\beta>r \,\Big|\, \mathcal{G}_n\Big\}}{P_x\{\gamma'>\beta \,|\, \mathcal{G}_n\}} &= \mathcal{L}_x\Big\{(Y',\Gamma') \,\Big|\, \mathcal{G}_n\Big\} \frac{P_x\{\gamma'-\beta>r \,|\, \mathcal{G}_n\}}{P_x\{\gamma'>\beta \,|\, \mathcal{G}_n\}} \\ &= (P_{Y_n}\circ T^{-1})\, e^{-r},\end{aligned}$$

as required. □

To complete the picture, we need a convenient criterion for non-explosion.

Proposition 13.5 *(explosion) For any rate kernel α and initial state x, let (Y_n) and (τ_n) be such as in Theorem* 13.3. *Then a.s.*

$$\tau_n \to \infty \quad \Leftrightarrow \quad \sum_n \frac{1}{c(Y_n)} = \infty. \tag{3}$$

In particular, $\tau_n \to \infty$ a.s. when x is recurrent for (Y_n).

Proof: Write $\beta_n = \{c(Y_{n-1})\}^{-1}$. Noting that $Ee^{-u\gamma_n} = (1+u)^{-1}$ for all $u \ge 0$, we get by Theorem 13.3 (ii) and Fubini's theorem

$$\begin{aligned} E(\, e^{-u\zeta} | \, Y) &= \textstyle\prod_n (1+u\beta_n)^{-1} \\ &= \exp\Big\{-\textstyle\sum_n \log(1+u\beta_n)\Big\} \quad \text{a.s.} \end{aligned} \tag{4}$$

Since $\frac{1}{2}(r\wedge 1) \le \log(1+r) \le r$ for all $r>0$, the series on the right converges for every $u>0$ iff $\sum_n \beta_n < \infty$. Letting $u \to 0$ in (4), we get by dominated convergence

$$P(\,\zeta<\infty \,|\, Y) = 1\Big\{ \textstyle\sum_n \beta_n < \infty \Big\} \quad \text{a.s.},$$

which implies (3). If x is visited infinitely often, the series $\sum_n \beta_n$ has infinitely many terms $c_x^{-1} > 0$, and the last assertion follows. □

The simplest and most basic pure jump-type Markov processes are the homogeneous Poisson processes, which are space and time homogeneous processes of this kind starting at 0 and proceeding by unit jumps. Here we define a *stationary Poisson process* on R_+ with *rate* $c>0$ as a simple point process ξ with stationary, independent increments, such that $\xi[0,t)$ is Poisson distributed with mean ct. To prove the existence, we may use generating functions to see that the Poisson distributions μ_t with mean t satisfy the semi-group property $\mu_s * \mu_t = \mu_{s+t}$. The desired existence now follows by Theorem 8.23.

Theorem 13.6 *(stationary Poisson process, Bateman) Let ξ be a simple point process on R_+ with points $\tau_1 < \tau_2 < \cdots$, and put $\tau_0 = 0$. Then these conditions are equivalent:*

(i) *ξ is a stationary Poisson process,*

(ii) *ξ has stationary, independent increments,*

(iii) *the variables $\tau_n - \tau_{n-1}$, $n \in \mathsf{N}$, are i.i.d., exponentially distributed.*

In that case ξ has rate $c = (E\tau_1)^{-1}$.

Proof, (i) $\Rightarrow$ (ii): Clear by definition.

(ii) $\Rightarrow$ (iii): Under (ii), Theorem 11.5 shows that $X_t = \xi[0,t]$, $t \geq 0$, is a space- and time-homogeneous Markov process of pure jump type, and so by Lemma 13.2, τ_1 is exponential and independent of $\theta_{\tau_1}\xi$. Since $\theta_{\tau_1}\xi \stackrel{d}{=} \xi$ by Theorem 13.1, we may proceed recursively, and (iii) follows.

(iii) $\Rightarrow$ (i): Assuming (iii) with $E\tau_1 = c^{-1}$, we may choose a stationary Poisson process η with rate c and points $\sigma_1 < \sigma_2 < \cdots$, and conclude as before that $(\sigma_n) \stackrel{d}{=} (\tau_n)$. Hence,

$$\xi = \textstyle\sum_n \delta_{\tau_n} \stackrel{d}{=} \sum_n \delta_{\sigma_n} = \eta. \qquad \square$$

By a *pseudo-Poisson* process in a measurable space S we mean a process of the form $X = Y \circ N$ a.s., where Y is a discrete-time Markov process in S and $N \perp\!\!\!\perp Y$ is an homogeneous Poisson process. Letting μ be the transition kernel of Y and writing c for the constant rate of N, we may construct a kernel

$$\alpha(x,B) = c\,\mu\Big(x, B \setminus \{x\}\Big), \quad x \in S,\ B \in \mathcal{S}, \tag{5}$$

which is measurable since $\mu(x,\{x\})$ is a measurable function of x. We may characterize pseudo-Poisson processes in terms of the rate kernel:

Proposition 13.7 *(pseudo-Poisson process, Feller) For a process X in a Borel space S, these conditions are equivalent:*

(i) *$X = Y \circ N$ a.s. for a discrete-time Markov chain Y in S with transition kernel μ and a Poisson process $N \perp\!\!\!\perp Y$ with constant rate $c > 0$,*

(ii) *X is a pure jump-type Markov process with bounded rate kernel α.*

In that case, we may choose α and (μ, c) to be related by (5).

Proof, (i) $\Rightarrow$ (ii): Assuming (i), write $\tau_1, \tau_2, \ldots$ for the jump times of N, and let $\mathcal{F}$ be the filtration induced by the pair (X,N). As in Theorem 13.4, we see that X is $\mathcal{F}$-Markov. To identify the rate kernel α, fix any initial state x, and note that the first jump of X occurs at the time τ_n when Y_n first leaves x. For each transition of Y, this happens with probability $p_x = \mu(x,\{x\}^c)$. By Proposition 15.3, the time until first jump is then exponentially distributed with parameter $c\,p_x$. If $p_x > 0$, the position of X after the first jump has distribution $p_x^{-1}\mu(x,\cdot \setminus \{x\})$. Thus, α is given by (5).

(ii) $\Rightarrow$ (i): Assuming (ii), put $r_x = \alpha(x,S)$ and $c = \sup_x r_x$, and note that the kernel

$$\mu(x,\cdot) = c^{-1}\Big\{\alpha(x,\cdot) + (c - r_x)\,\delta_x\Big\}, \quad x \in S,$$

satisfies (5). Thus, if $X' = Y' \circ N'$ is a pseudo-Poisson process based on μ and c, then X' is again Markov with rate kernel α, and so $X \stackrel{d}{=} X'$. Hence, Corollary 8.18 yields $X = Y \circ N$ a.s. for a pair $(Y, N) \stackrel{d}{=} (Y', N')$. □

If the underlying Markov chain Y is a random walk in a measurable Abelian group S, then $X = Y \circ N$ is called a *compound Poisson process*. Here $X - X_0 \perp\!\!\!\perp X_0$, the jump sizes are i.i.d., and the jump times form an independent, homogeneous Poisson process. Thus, the distribution of $X - X_0$ is determined by the *characteristic measure*[1] $\nu = c\,\mu$, where c is the rate of the jump-time process and μ is the common jump distribution.

Compound Poisson processes may be characterized analytically in terms of the rate kernel, and probabilistically in terms of the increments of the process. A kernel α on S is said to be *invariant* if $\alpha_x = \alpha_0 \circ \theta_x^{-1}$ or $\alpha(x, B) = \alpha(0, B - x)$ for all x and B. We also say that a process X in S has *independent increments* if $X_t - X_s \perp\!\!\!\perp \{X_r;\, r \le s\}$ for all $s < t$.

Corollary 13.8 *(compound Poisson process) For a pure jump-type process X in a measurable Abelian group, these conditions are equivalent:*

(i) *X is Markov with invariant rate kernel,*
(ii) *X has stationary, independent increments,*
(iii) *X is compound Poisson.*

Proof: If a pure jump-type Markov process is space-homogeneous, then its rate kernel is clearly invariant, and the converse follows from the representation in Theorem 13.3. Thus, (i) ⇔ (ii) by Proposition 11.5. Next, Theorem 13.3 yields (i) ⇒ (iii), and the converse follows by Theorem 13.4. □

We now derive a combined differential and integral equation for the transition kernels μ_t. An abstract version of this result appears in Theorem 17.6. For any measurable and suitably integrable function $f\colon S \to \mathsf{R}$, we define

$$\begin{aligned} T_t f(x) &= \int f(y)\, \mu_t(x, dy) \\ &= E_x f(X_t), \quad x \in S,\ t \ge 0. \end{aligned}$$

Theorem 13.9 *(backward equation, Kolmogorov) Let α be the rate kernel of a pure jump-type Markov process in S, and fix a bounded, measurable function $f\colon S \to \mathsf{R}$. Then $T_t f(x)$ is continuously differentiable in t for fixed x, and*

$$\frac{\partial}{\partial t} T_t f(x) = \int \alpha(x, dy) \left\{ T_t f(y) - T_t f(x) \right\}, \quad t \ge 0,\ x \in S. \tag{6}$$

[1]also called *Lévy measure*, as in the general case of Chapter 16

Proof: Put $\tau = \tau_1$, and let $x \in S$ and $t \geq 0$. By the strong Markov property at $\sigma = \tau \wedge t$ and Theorem 8.5,

$$\begin{aligned} T_t f(x) &= E_x f(X_t) = E_x f\Big\{(\theta_\sigma X)_{t-\sigma}\Big\} \\ &= E_x T_{t-\sigma} f(X_\sigma) \\ &= f(x)\, P_x\{\tau > t\} + E_x\Big\{T_{t-\tau} f(X_\tau);\, \tau \leq t\Big\} \\ &= f(x)\, e^{-t\,c_x} + \int_0^t e^{-s\,c_x} ds \int \alpha(x, dy)\, T_{t-s} f(y), \end{aligned}$$

and so

$$e^{t\,c_x} T_t f(x) = f(x) + \int_0^t e^{s\,c_x} ds \int \alpha(x, dy)\, T_s f(y). \tag{7}$$

Here the use of the disintegration theorem is justified by the fact that $X(\omega, t)$ is product measurable on $\Omega \times \mathsf{R}_+$, by the right continuity of the paths.

From (7) we see that $T_t f(x)$ is continuous in t for each x, and so by dominated convergence the inner integral on the right is continuous in s. Hence, $T_t f(x)$ is continuously differentiable in t, and (6) follows by an easy computation. □

Next we show how the invariant distributions of a pure jump-type Markov process are related to those of the embedded Markov chain.

Proposition 13.10 *(invariant measures) Let the processes X, Y be related as in Theorem* 13.3, *and fix a probability measure ν on S with $\int c\, d\nu < \infty$. Then*

$$\nu \text{ is invariant for } X \quad \Leftrightarrow \quad c \cdot \nu \text{ is invariant for } Y.$$

Proof: By Theorem 13.9 and Fubini's theorem, we have for any bounded measurable function $f: S \to \mathsf{R}$

$$E_\nu f(X_t) = \int f(x)\, \nu(dx) + \int_0^t ds \int \nu(dx) \int \alpha(x, dy)\, \Big\{T_s f(y) - T_s f(x)\Big\}.$$

Thus, ν is invariant for X iff the second term on the right is identically zero. Now (6) shows that $T_t f(x)$ is continuous in t, and by dominated convergence this remains true for the integral

$$I_t = \int \nu(dx) \int \alpha(x, dy)\, \Big\{T_t f(y) - T_t f(x)\Big\}, \quad t \geq 0.$$

Thus, the condition becomes $I_t \equiv 0$. Since f is arbitrary, it is enough to take $t = 0$. Our condition then reduces to $(\nu\alpha) f \equiv \nu(cf)$ or $(c \cdot \nu)\mu = c \cdot \nu$, which means that $c \cdot \nu$ is invariant for Y. □

We turn to a study of pure jump-type Markov processes in a countable state space S, also called *continuous-time Markov chains*. Here the kernels μ_t are determined by the *transition functions* $p_{ij}^t = \mu_t(i, \{j\})$. The connectivity properties are simpler than in the discrete case, and the notion of periodicity has no continuous-time counterpart.

Lemma 13.11 *(positivity) Consider a continuous-time Markov chain in S with transition functions p_{ij}^t. Then for fixed $i, j \in S$, exactly one of these cases occurs:*

(i) $p_{ij}^t > 0,\ t > 0$,

(ii) $p_{ij}^t = 0,\ t \geq 0$.

In particular, $p_{ii}^t > 0$ for all t and i.

Proof: Let $q = (q_{ij})$ be the transition matrix of the embedded Markov chain Y in Theorem 13.3. If $q_{ij}^n = P_i\{Y_n = j\} = 0$ for all $n \geq 0$, then clearly $1\{X_t \neq j\} \equiv 1$ a.s. P_i, and so $p_{ij}^t = 0$ for all $t \geq 0$. If instead $q_{ij}^n > 0$ for some $n \geq 0$, there exist some states $i = i_0$ and $i_1, \ldots, i_n = j$ with $q_{i_{k-1}, i_k} > 0$ for $k = 1, \ldots, n$. Since the distribution of $(\gamma_1, \ldots, \gamma_{n+1})$ has the positive density $\prod_{k \leq n+1} e^{-x_k} > 0$ on R_+^{n+1}, we obtain for any $t > 0$

$$p_{ij}^t \geq P\left\{ \sum_{k=1}^{n} \frac{\gamma_k}{c_{i_{k-1}}} \leq t < \sum_{k=1}^{n+1} \frac{\gamma_k}{c_{i_{k-1}}} \right\} \prod_{k=1}^{n} q_{i_{k-1}, i_k} > 0.$$

Noting that $p_{ii}^0 = q_{ii}^0 = 1$, we get in particular $p_{ii}^t > 0$ for all $t \geq 0$. □

A continuous-time Markov chain is said to be *irreducible* if $p_{ij}^t > 0$ for all $i, j \in S$ and $t > 0$. This clearly holds iff the underlying discrete-time process Y in Theorem 13.3 is irreducible, in which case

$$\sup\Big\{t > 0;\ X_t = j\Big\} < \infty \quad \Leftrightarrow \quad \sup\Big\{n > 0;\ Y_n = j\Big\} < \infty.$$

Thus, when Y is recurrent, the sets $\{t;\ X_t = j\}$ are a.s. unbounded under P_i for all $i \in S$; otherwise they are a.s. bounded. The two possibilities are again referred to as *recurrence* and *transience*, respectively.

The basic ergodic Theorem 11.22 for discrete-time Markov chains has the following continuous-time counterpart. Further extensions are considered in Chapter 17 and 26. Recall that $\xrightarrow{u}$ means convergence in total variation, and let $\hat{\mathcal{M}}_S$ be the class of probability measures on S.

Theorem 13.12 *(convergence dichotomy, Kolmogorov) For an irreducible, continuous-time Markov chain in S, exactly one of these cases occurs:*

(i) *there exists a unique invariant distribution ν, the latter satisfies $\nu_i > 0$ for all $i \in S$, and as $t \to \infty$ we have*

$$P_\mu \circ \theta_t^{-1} \xrightarrow{u} P_\nu, \quad \mu \in \hat{\mathcal{M}}_S,$$

(ii) *no invariant distribution exists, and as $t \to \infty$,*

$$p_{ij}^t \to 0, \quad i, j \in S.$$

Proof: By Lemma 13.11, the discrete-time chain X_{nh}, $n \in \mathsf{Z}_+$, is irreducible and aperiodic. Suppose that (X_{nh}) is positive recurrent for some $h > 0$, say with invariant distribution ν. Then the chain $(X_{nh'})$ is positive recurrent for every h' of the form $2^{-m}h$, and by the uniqueness in Theorem 11.22 it has the same invariant distribution. Since the paths are right-continuous, a simple approximation shows that ν is invariant even for the original process X.

For any distribution μ on S, we have

$$\begin{aligned}\left\|P_\mu \circ \theta_t^{-1} - P_\nu\right\| &= \left\| \textstyle\sum_i \mu_i \sum_j (p_{ij}^t - \nu_j) P_j \right\| \\ &\le \textstyle\sum_i \mu_i \sum_j \left|p_{ij}^t - \nu_j\right|.\end{aligned}$$

Thus, (i) will follow by dominated convergence, if we can show that the inner sum on the right tends to zero. This is clear if we put $n = [t/h]$ and $r = t - nh$, and note that by Theorem 11.22,

$$\begin{aligned}\textstyle\sum_k \left| p_{ik}^t - \nu_k \right| &\le \textstyle\sum_{j,k} \left| p_{ij}^{nh} - \nu_j \right| p_{jk}^r \\ &= \textstyle\sum_j \left| p_{ij}^{nh} - \nu_j \right| \to 0.\end{aligned}$$

It remains to consider the case where (X_{nh}) is null-recurrent or transient for every $h > 0$. Fixing any $i, k \in S$ and writing $n = [t/h]$ and $r = t - nh$ as before, we get

$$\begin{aligned}p_{ik}^t &= \textstyle\sum_j p_{ij}^r \, p_{jk}^{nh} \\ &\le p_{ik}^{nh} + \sum_{j \ne i} p_{ij}^r \\ &= p_{ik}^{nh} + (1 - p_{ii}^r),\end{aligned}$$

which tends to zero as $t \to \infty$ and then $h \to 0$, by Theorem 11.22 and the continuity of p_{ii}^t. □

As in discrete time, condition (ii) of the last theorem holds for any transient Markov chain, whereas a recurrent chain may satisfy either condition. Recurrent chains satisfying (i) and (ii) are again referred to as *positive-recurrent* and *null-recurrent*, respectively. Note that X may be positive recurrent even if the embedded, discrete-time chain Y is null-recurrent, and vice versa. On the other hand, X clearly has the same asymptotic properties as the discrete-time processes (X_{nh}), $h > 0$.

For any $j \in S$, we now introduce the *first exit* and *recurrence times*

$$\begin{aligned}\gamma_j &= \inf\{t > 0;\ X_t \ne j\}, \\ \tau_j &= \inf\{t > \gamma_j;\ X_t = j\}.\end{aligned}$$

As for Theorem 11.26 in the discrete-time case, we may express the asymptotic transition probabilities in terms of the mean recurrence times $E_j \tau_j$. To avoid trite exceptions, we consider only non-absorbing states.

Theorem 13.13 *(mean recurrence times, Kolmogorov) For a continuous-time Markov chain in S and states $i, j \in S$ with j non-absorbing, we have as $t \to \infty$*

$$p_{ij}^t \to \frac{P_i\{\tau_j < \infty\}}{c_j \, E_j \tau_j}\,. \tag{8}$$

Proof: We may take $i = j$, since the general statement will then follow as in the proof of Theorem 11.26. If j is transient, then $1\{X_t = j\} \to 0$ a.s. P_j, and so by dominated convergence $p_{jj}^t = P_j\{X_t = j\} \to 0$. This agrees with (8), since in this case $P_j\{\tau_j = \infty\} > 0$. Turning to the recurrent case, let S_j be the class of states i accessible from j. Then S_j is clearly irreducible, and so p_{jj}^t converges by Theorem 13.12.

To identify the limit, define

$$\begin{aligned} L_t^j &= \lambda\Big\{s \le t;\, X_s = j\Big\} \\ &= \int_0^t 1\{X_s = j\}\, ds, \quad t \ge 0, \end{aligned}$$

and let τ_j^n be the time of n-th return to j. Letting $m, n \to \infty$ with $|m-n| \le 1$ and using the strong Markov property and the law of large numbers, we get a.s. P_j

$$\begin{aligned} \frac{L^j(\tau_j^m)}{\tau_j^n} &= \frac{L^j(\tau_j^m)}{m} \cdot \frac{n}{\tau_j^n} \cdot \frac{m}{n} \\ &\to \frac{E_j \gamma_j}{E_j \tau_j} = \frac{1}{c_j E_j \tau_j}. \end{aligned}$$

By the monotonicity of L^j it follows that $t^{-1} L_t^j \to (c_j E_j \tau_j)^{-1}$ a.s. Hence, by Fubini's theorem and dominated convergence,

$$\frac{1}{t}\int_0^t p_{jj}^s\, ds = \frac{E_j L_t^j}{t} \to \frac{1}{c_j E_j \tau_j},$$

and (8) follows. □

We conclude with some applications to branching processes and their diffusion approximations. By a *Bienaymé process*[2] we mean a discrete-time branching process $X = (X_n)$, where X_n represents the total number of individuals in generation n. Given that $X_n = m$, the individuals in the n-th generation give rise to independent families of progeny $Y_1, \dots, Y_m$ starting at time n, each distributed like the original process X with $X_0 = 1$. This clearly defines a discrete-time Markov chain taking values in Z_+. We say that X is *critical* when $E_1 X_1 = 1$, and *super-* or *sub-critical* when $E_1 X_1 > 1$ or < 1, respectively.[3] To avoid some trite exceptions, we assume that both possibilities $X_1 = 0$ and $X_1 > 1$ have positive probabilities.

Theorem 13.14 *(Bienaymé process) Let X be a Bienaymé process with $E_1 X_1 = \mu$ and $\mathrm{Var}_1(X_1) = \sigma^2$, and put $f(s) = E_1 s^{X_1}$. Then*

(i) *$P_1\{X_n \to 0\}$ is the smallest $s \in [0,1]$ with $f(s) = s$, and so*

$$X_n \to 0 \ \textit{a.s.} \quad \Leftrightarrow \quad \mu \le 1,$$

(ii) *when $\mu < \infty$, we have $\mu^{-n} X_n \to \gamma$ a.s. for a random variable $\gamma \ge 0$,*

[2]also called a *Galton–Watson process*

[3]The conventions of martingale theory are the opposite. Thus, a branching process is super-critical iff it is a sub-martingale, and sub-critical iff it is a super-martingale.

(iii) *when*[4] $\sigma^2 < \infty$, *we have* $E_1 \gamma = 1 \Leftrightarrow \mu > 1$, *and*

$$\{\gamma = 0\} = \{X_n \to 0\} \text{ a.s. } P_1.$$

The proof is based on some simple algebraic facts:

Lemma 13.15 *(recursion and moments) Let X be a Bienaymé process with $E_1 X_1 = \mu$ and $\mathrm{Var}_1(X_1) = \sigma^2$. Then*

(i) *the functions $f^{(n)}(s) = E_1 s^{X_n}$ on $[0,1]$ satisfy*

$$f^{(m+n)}(s) = f^{(m)} \circ f^{(n)}(s), \quad m, n \in \mathsf{N},$$

(ii) $E_1 X_n = \mu^n$, *and when $\sigma^2 < \infty$ we have*

$$\mathrm{Var}_1(X_n) = \begin{cases} n\,\sigma^2, & \mu = 1, \\ \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}, & \mu \neq 1. \end{cases}$$

Proof: (i) By the Markov and independence properties of X, we have

$$\begin{aligned} f^{(m+n)}(s) &= E_1 s^{X_{m+n}} = E_1 E_{X_m} s^{X_n} \\ &= E_1 (E_1 s^{X_n})^{X_m} \\ &= f^{(m)}(E_1 s^{X_n}) \\ &= f^{(m)} \circ f^{(n)}(s). \end{aligned}$$

(ii) Write $\mu_n = E_1 X_n$ and $\sigma_n^2 = \mathrm{Var}_1(X_n)$, whenever those moments exist. The Markov and independence properties yield

$$\begin{aligned} \mu_{m+n} &= E_1 X_{m+n} = E_1 E_{X_m} X_n \\ &= E_1 (X_m\, E_1 X_n) \\ &= (E_1 X_m)(E_1 X_n) \\ &= \mu_m\, \mu_n, \end{aligned}$$

and so by induction $\mu_n = \mu^n$. Furthermore, Theorem 12.9 (iii) gives

$$\sigma_{m+n}^2 = \sigma_m^2\, \mu_n + \mu_m^2\, \sigma_n^2.$$

The variance formula is clearly true for $n = 1$. Proceeding by induction, suppose it holds for a given n. Then for $n+1$ we get when $\mu \neq 1$

$$\begin{aligned} \sigma_{n+1}^2 &= \sigma_n^2\, \mu + \mu_n^2\, \sigma^2 \\ &= \mu\, \sigma^2\, \mu^{n-1} \frac{\mu^n - 1}{\mu - 1} + \mu^{2n}\, \sigma^2 \\ &= \frac{\sigma^2\, \mu^n}{\mu - 1} \left\{ \mu^n - 1 + \mu^n\, (\mu - 1) \right\} \\ &= \sigma^2\, \mu^n\, \frac{\mu^{n+1} - 1}{\mu - 1}, \end{aligned}$$

as required. When $\mu = 1$, the proof simplifies to

[4] This assertion remains true under the weaker condition $E_1(X_1 \log X_1) < \infty$.

$$\begin{aligned}\sigma_{n+1}^2 &= \sigma_n^2 + \sigma^2 \\ &= n\,\sigma^2 + \sigma^2 \\ &= (n+1)\,\sigma^2.\end{aligned}$$ □

Proof of Theorem 13.14: (i) Noting that

$$P_1\{X_n = 0\} = E_1 0^{X_n} = f^{(n)}(0),$$

we get by Lemma 13.15 (i)

$$P_1\{X_{n+1} = 0\} = f\big(P_1\{X_n = 0\}\big).$$

Since also $P_1\{X_0 = 0\} = 0$ and $P_1\{X_n = 0\} \uparrow P\{X_n \to 0\}$, we conclude that the extinction probability q is the smallest root in $[0,1]$ of the equation $f(s) = s$. Now f is strictly convex with $f'(1) = \mu$ and $f(0) \geq 0$, and so the equation $f(s) = s$ has the single root 1 when $\mu \leq 1$, whereas for $\mu > 1$ it has two roots $q < 1$ and 1.

(ii) Since $E_1 X_1 = \mu$, the Markov and independence properties yield

$$\begin{aligned}E(X_{n+1} \mid X_0, \ldots, X_n) &= X_n(E_1 X_1) \\ &= \mu\, X_n, \quad n \in \mathsf{N}.\end{aligned}$$

Dividing both sides by μ^{n+1}, we conclude that the sequence $M_n = \mu^{-n} X_n$ is a martingale. Since it is non-negative and hence L^1-bounded, it converges a.s. by Theorem 9.19.

(iii) If $\mu \leq 1$, then $X_n \to 0$ a.s. by (i), and so $E_1\gamma = 0$. If instead $\mu > 1$ and $\sigma^2 < \infty$, then Lemma 13.15 (ii) yields

$$\mathrm{Var}(M_n) = \sigma^2 \frac{1 - \mu^{-n}}{\mu(\mu - 1)} \to \frac{\sigma^2}{\mu(\mu - 1)} < \infty,$$

and so (M_n) is uniformly integrable, which implies $E_1\gamma = E_1 M_0 = 1$. Furthermore,

$$\begin{aligned}P_1\{\gamma = 0\} &= E_1 P_{X_1}\{\gamma = 0\} \\ &= E_1\big(P_1\{\gamma = 0\}\big)^{X_1} \\ &= f\big(P_1\{\gamma = 0\}\big),\end{aligned}$$

which shows that even $P_1\{\gamma = 0\}$ is a root of the equation $f(s) = s$. It can't be 1 since $E_1\gamma = 1$, so in fact $P_1\{\gamma = 0\} = P_1\{X_n \to 0\}$. Since also $\{X_n \to 0\} \subset \{\gamma = 0\}$, the two events agree a.s. □

The simplest case is when X proceeds by *binary splitting*, so that the offspring distribution is confined to the values 0 and 2. To obtain a continuous-time Markov chain, we may choose the life lengths until splitting or death to be independent and exponentially distributed with constant mean. This makes $X = (X_t)$ a birth & death process with rates $n\lambda$ and $n\mu$, and we say that X is *critical* of *rate* 1 if $\lambda = \mu = 1$. We also consider the *Yule process*, where $\lambda = 1$ and $\mu = 0$, so that X is non-decreasing and no deaths occur.

Theorem 13.16 *(binary splitting) Let X be a critical, unit rate binary-splitting process, and define $p_t = (1+t)^{-1}$. Then*

(i) $P_1\{X_t > 0\} = p_t, \quad t > 0,$

(ii) $P_1\big\{X_t = n \,\big|\, X_t > 0\big\} = p_t(1-p_t)^{n-1}, \quad t > 0,\ n \in \mathsf{N},$

(iii) *for fixed $t > 0$, the ancestors of X_t at times $s \le t$ form a Yule process Z_s^t, $s \in [0,t]$, with rate function $(1+t-s)^{-1}$.*

Proof, (i) The generating functions

$$F(s,t) = E_1 s^{X_t}, \quad s \in [0,1],\ t > 0,$$

satisfy Kolmogorov's backward equation

$$\frac{\partial}{\partial t} F(s,t) = \big\{1 - F(s,t)\big\}^2, \qquad F(s,0) = s,$$

which has the unique solution

$$\begin{aligned} F(s,t) &= \frac{s+t-st}{1+t-st} = q_t + \frac{p_t^2 s}{1-q_t s} \\ &= q_t + \sum_{n \ge 1} p_t^2\, q_t^{n-1} s^n, \end{aligned}$$

where $q_t = 1 - p_t$. In particular, $P_1\{X_t = 0\} = F(0,t) = q_t$.

(ii) For any $n \in \mathsf{N}$,

$$\begin{aligned} P_1\big\{X_t = n \,\big|\, X_t > 0\big\} &= \frac{P_1\{X_t = n\}}{P_1\{X_t > 0\}} \\ &= \frac{p_t^2\, q_t^{n-1}}{p_t} = p_t\, q_t^{n-1}. \end{aligned}$$

(iii) Assuming $X_0 = 1$, we get

$$\begin{aligned} P\{Z_s^t = 1\} &= EP\big\{ Z_s^t = 1 \,\big|\, X_s \big\} \\ &= E\Big(X_s\, p_{t-s}\, q_{t-s}^{X_s - 1} \Big) \\ &= \sum_{n \ge 1} n\, p_s^2\, q_s^{n-1} p_{t-s}\, q_{t-s}^{n-1} \\ &= p_s^2\, p_{t-s} \sum_{n \ge 1} n (q_s\, q_{t-s})^{n-1} \\ &= \frac{p_s^2\, p_{t-s}}{(1 - q_s\, q_{t-s})^2} = \frac{1+t-s}{(1+t)^2}\,. \end{aligned}$$

Writing $\tau_1 = \inf\{s \le t;\ Z_s^t > 1\}$, we conclude that

$$\begin{aligned} P\big\{ \tau_1 > s \,\big|\, X_t > 0 \big\} &= P\big\{ Z_s^t = 1 \,\big|\, X_t > 0 \big\} \\ &= \frac{1+t-s}{1+t}\,, \end{aligned}$$

which shows that τ_1 has the defective probability density $(1+t)^{-1}$ on $[0,t]$. By Proposition 10.23, the associated compensator has density $(1+t-s)^{-1}$. The Markov and branching properties carry over to the process Z_s^t by Theorem 8.12. Using those properties and proceeding recursively, we see that Z_s^t is a Yule process in $s \in [0,t]$ with rate function $(1+t-s)^{-1}$. □

Now let the individual particles perform independent Brownian motions[5] in R^d, so that the discrete branching process $X = (X_t)$ turns into a random branching tree $\xi = (\xi_t)$. For fixed $t > 0$, even the ancestral processes Z_s^t combine into a random branching tree $\zeta^t = (\zeta_s^t)$ on $[0,t]$.

Corollary 13.17 *(Brownian branching tree) Let the processes ξ_t form a critical, unit rate Brownian branching tree in R^d, and define $p_t = (1+t)^{-1}$. Then for $s \le s+h = t$,*

(i) *the ancestors of ξ_t at time s form a p_h-thinning ζ_s^t of ξ_s,*
(ii) *ξ_t is a sum of conditionally independent clusters of age h, rooted at the points of ζ_s^t,*
(iii) *for fixed $t > 0$, the processes ζ_s^t, $s \in [0,t]$, form a Brownian Yule tree with rate function $(1+t-s)^{-1}$.*

Similar results hold asymptotically for critical Bienaymé processes.

Theorem 13.18 *(survival rate and distribution, Kolmogorov, Yaglom) Let X be a critical Bienaymé process with $\mathrm{Var}_1(X_1) = \sigma^2 \in (0,\infty)$. Then as $n \to \infty$,*

(i) $n\,P_1\{X_n > 0\} \to 2\,\sigma^{-2}$,
(ii) $P_1\Big\{n^{-1}X_n > r \,\Big|\, X_n > 0\Big\} \to e^{-2\,r\sigma^{-2}}, \quad r \ge 0.$

Our proof is based on an elementary approximation of generating functions:

Lemma 13.19 *(generating functions) For X as above, define $f_n(s) = E_1 s^{X_n}$. Writing $\sigma^2 = 2\,c$ we have as $n \to \infty$, uniformly in $s \in [0,1)$,*

$$\frac{1}{n}\left\{\frac{1}{1-f_n(s)} - \frac{1}{1-s}\right\} \to c.$$

Proof (Spitzer): For the restrictions to Z_+ of a binary splitting process X, we get by Theorem 13.16

$$P_1\{X_1 = k\} = (1-p)^2\,p^{k-1}, \quad k \in \mathsf{N},$$

for a constant $p \in (0,1)$, and an easy calculation yields

$$\frac{1}{1-f_n(s)} - \frac{1}{1-s} = \frac{n\,p}{1-p} = n\,c.$$

[5]For the notions of Brownian motion and p-thinnings, see Chapters 14–15.

Thus, equality holds in the stated relation with p and c related by

$$p = \frac{c}{1+c} \in (0,1).$$

For general X, fix any constants $c_1, c_2 > 0$ with $c_1 < c < c_2$, and let $X^{(1)}$, $X^{(2)}$ be associated binary splitting processes. By elementary estimates there exist some constants $n_1, n_2 \in \mathsf{Z}$, such that the associated generating functions satisfy

$$f^{(1)}_{n+n_1} \le f_n \le f^{(2)}_{n+n_2}, \quad n > |n_1| \vee |n_2|.$$

Combining this with the identities for $f^{(1)}_n$, $f^{(2)}_n$, we obtain

$$c_1 \frac{(n+n_1)}{n} \le \frac{1}{n}\left\{\frac{1}{1-f_n(s)} - \frac{1}{1-s}\right\} \le c_2 \frac{(n+n_2)}{n}.$$

Here let $n \to \infty$, and then $c_1 \uparrow c$ and $c_2 \downarrow c$. □

Proof of Theorem 13.18: (i) Taking $s = 0$ in Lemma 13.19 gives $f_n(0) \to 1$.

(ii) Writing $s_n = e^{-r/n}$ and using the uniformity in Lemma 13.19, we get

$$\frac{1}{n\big\{1 - f_n(s_n)\big\}} \approx \frac{1}{n(1-s_n)} + c \to \frac{1}{r} + c,$$

where $a \approx b$ means $a - b \to 0$. Since also $n\{1 - f_n(0)\} \to c^{-1}$ by (i), we obtain

$$\begin{aligned} E_1\Big(e^{-rX_n/n}\,\Big|\,X_n > 0\Big) &= \frac{f_n(s_n) - f_n(0)}{1 - f_n(0)} \\ &= 1 - \frac{n\big\{1 - f_n(s_n)\big\}}{1 - f_n(0)} \\ &\to 1 - \frac{c}{r^{-1} + c} = \frac{1}{1 + c\,r}, \end{aligned}$$

recognized as the Laplace transform of the exponential distribution with mean c^{-1}. Now use Theorem 6.3. □

We finally state a basic approximation property of Bienaymé branching processes, anticipating some weak convergence and SDE theory from Chapters 23 and 31–32.

Theorem 13.20 *(diffusion approximation, Feller) For a critical Bienaymé process X with* $\mathrm{Var}_1(X_1) = 1$, *consider versions $X^{(n)}$ with $X^{(n)}_0 = n$, and put*

$$Y^{(n)}_t = n^{-1} X^{(n)}_{[nt]}, \quad t \ge 0.$$

Then $Y^{(n)} \xrightarrow{sd} Y$ in D_{R_+}, where Y satisfies the SDE

$$dY_t = Y_t^{1/2} dB_t, \qquad Y_0 = 1. \tag{9}$$

Note that uniqueness in law holds for (9) by Theorem 33.1 below. By Theorem 33.3 we have even pathwise uniqueness, which implies strong existence by Theorem 32.14. The solutions are known as *squared Bessel processes of order* 0.

Proof (outline): By Theorems 7.2 and 13.18 we have $Y_t^{(n)} \overset{d}{\sim} \tilde{Y}_t^{(n)}$ for fixed $t > 0$, where the variables $\tilde{Y}_t^{(n)}$ are compound Poisson with characteristic measures $\nu_t^{(n)} \xrightarrow{w} \nu_t$, and

$$\nu_{2t}(dx) = t^{-2} e^{-x/t} dx, \quad x, t > 0.$$

Hence, Lemma 7.8 yields $Y_t^{(n)} \xrightarrow{d} Y_t$, where the limit is infinitely divisible in R_+ with characteristics $(0, \nu_t)$. By suitable scaling, we obtain a similar convergence of the general transition kernels $\mathcal{L}(Y_{s+t}^{(n)} \mid Y_s^{(n)})_x$, except that the limiting law $\mu_t(x, \cdot)$ is now infinitely divisible with characteristics $(0, x\,\nu_t)$. We also note that the time-homogeneous Markov property carries over to the limit.

By Lemma 7.1 (i) we may write

$$Y_t \overset{d}{=} \xi_1^t + \cdots + \xi_{\kappa_t}^t, \quad t > 0,$$

where the ξ_k^t are independent, exponential random variables with mean $t/2$, and κ_t is independent and Poisson distributed with mean $2/t$. In particular, $EY_t = (E\kappa_t)(E\xi_1^t) = 1$, and Theorem 12.9 (iii) yields

$$\begin{aligned} \mathrm{Var}(Y_t) &= E\,\kappa_t \,\mathrm{Var}(\xi_1^t) + \mathrm{Var}(\kappa_t)\,(E\,\xi_1^t)^2 \\ &= \frac{2}{t}\Big(\frac{t}{2}\Big)^2 + \frac{2}{t}\Big(\frac{t}{2}\Big)^2 = t. \end{aligned}$$

By Theorem 23.11 below, the processes $Y^{(n)}$ are tight in D_{R_+}. Since the jumps of $Y^{(n)}$ are a.s. of order n^{-1}, every limiting process Y has continuous paths. The previous moment calculations suggest that Y is then a diffusion process with drift 0 and diffusion rate $\sigma^2(x) = x$, which leads to the stated SDE (details omitted). The asserted convergence now follows by Theorem 17.28. □

Exercises

1. Prove the implication (iii) ⇒ (i) in Theorem 13.6 by direct computation. (*Hint:* Note that $\{N_t \geq n\} = \{\tau_n \leq t\}$ for all t and n.)

2. Give a non-computational proof of the implication (i) ⇒ (iii) in Theorem 13.6, using the reverse implication. (*Hint:* Note as in Theorem 2.19 that $\tau_1, \tau_2, \ldots$ are measurable functions of ξ.)

3. Prove the implication (i) ⇒ (iii) in Theorem 13.6 by direct computation.

4. Show that the Poisson process is the only renewal process that is also a Markov process.

5. Let $X_t = \xi[0,t]$ for a binomial process ξ with $\|\xi\| = n$ based on a diffuse probability measure μ on R_+. Show that we can choose μ such that X becomes a Markov process, and determine the corresponding rate kernel.

6. For a pure jump-type Markov process in S, show that $P_x\{\tau_2 \le t\} = o(t)$ for all $x \in S$. Also note that the bound can be strengthened to $O(t^2)$ if the rate function is bounded, but not in general. (*Hint:* Use Lemma 13.2 and dominated convergence.)

7. Show that a transient, discrete-time Markov chain Y can be embedded into an exploding (resp., non-exploding) continuous-time chain X. (*Hint:* Use Propositions 11.16 and 13.5.)

8. In Corollary 13.8, use the measurability of the mapping $X = Y \circ N$ to derive the implication (iii) $\Rightarrow$ (i) from its converse. (*Hint:* Proceed as in the proof of Theorem 13.6.)

9. Consider a pure jump-type Markov process in $(S, \mathcal{S})$ with transition kernels μ_t and rate kernel α. For any $x \in S$ and $B \in \mathcal{S}$, show that $\alpha(x, B) = \dot{\mu}_0(x, B \setminus \{x\})$. (*Hint:* Apply Theorem 13.9 with $f = 1_{B\setminus\{x\}}$, and use dominated convergence.)

10. Use Theorem 13.9 to derive a system of differential equations for the transition functions $p_{ij}(t)$ of a continuous-time Markov chain. (*Hint:* Take $f(i) = \delta_{ij}$ for fixed j.)

11. Give an example of a positive recurrent, continuous-time Markov chain, such that the embedded discrete-time chain is null-recurrent, and vice versa. (*Hint:* Use Proposition 13.10.)

12. Prove Theorem 13.12 by a direct argument, mimicking the proof of Theorem 11.22.

13. Let X be a Bienaymé process as in Theorem 13.14. Determine the extinction probability and asymptotic growth rate if instead we start with m individuals.

14. Let $X = (X_t)$ be a binary-splitting process as in Theorem 13.16, and define $Y_n = X_{nh}$ for fixed $h > 0$. Show that $Y = (Y_n)$ is a critical Bienaymé process, and find the associated offspring distribution and asymptotic survival rate.

15. Show that the binary-splitting process in Theorem 13.16 is a pure jump-type Markov process, and calculate the associated rate kernel.

16. For a critical Bienaymé process with $\mathrm{Var}_1(X_1) = \sigma^2 < \infty$, find the asymptotic behavior of $P_m\{X_n > 0\}$ and $P_m\{n^{-1}X_n > r\}$ as $n \to \infty$ for fixed $m \in \mathsf{N}$.

17. How will Theorem 13.20 change if we assume instead that $\mathrm{Var}_1(X_1) = \sigma^2 > 0$?

V. Some Fundamental Processes

The Brownian motion and Poisson processes constitute the basic building blocks of probability theory, and both classes of processes enjoy a wealth of remarkable properties of constant use throughout the subject. Lévy processes are continuous-time counterparts of random walks, representable as mixtures of Brownian and Poisson processes. They may also be regarded as prototypes of Feller processes, which form a broad class of Markov processes, restricted only by some natural regularity assumptions that simplify the analysis. Most of the included material is of fundamental importance, though some later parts of especially Chapters 16–17 are more advanced and might be postponed until a later stage of study.

— — —

14. Gaussian processes and Brownian motion. After a preliminary study of general Gaussian processes, we introduce Brownian motion and derive some of its basic properties, including the path regularity, quadratic variation, reflection principle, the arcsine laws, and the laws of the iterated logarithm. The chapter concludes with a study of multiple Wiener–Itô integrals and the associated chaos expansion of Gaussian functionals. Needless to say, the basic theory of Brownian motion is of utmost importance.

15. Poisson and related processes. The basic theory of Poisson and related processes is equally important. Here we note in particular the mapping and marking theorems, the relations to binomial processes, and the roles of Cox transforms and thinnings. The independence properties lead to a Poisson representation of random measures with independent increments. We also note how a simple point process can be transformed to Poisson through a random time change derived from the compensator.

16. Independent-increment and Lévy processes. Here we note how processes with independent increments are composed of Gaussian and Poisson variables, reducing in the time-homogeneous case to mixtures of Brownian motions and Poisson processes. The resulting processes may be regarded as both continuous-time counterparts of random walks and as space-homogeneous Feller processes. After discussing some related approximation theorems, we conclude with an introduction to tangential processes.

17. Feller processes and semi-groups. The Feller processes form a general class of Markov processes, broad enough to cover most applications of interest, and yet restricted by some regularity conditions ensuring technical flexibility and convenience. Here we study the underlying semigroup theory, prove the general path regularity and strong Markov property, and establish some basic approximation and convergence theorems. Though considerably more advanced, this material is clearly of fundamental importance.

Chapter 14

Gaussian Processes and Brownian Motion

Covariance and independence, rotational symmetry, isonormal Gaussian process, independent increments, Brownian motion and bridge, scaling and inversion, Gaussian Markov processes, quadratic variation, path irregularity, strong Markov and reflection properties, Bessel processes, maximum process, arcsine and uniform laws, laws of the iterated logarithm, Wiener integral, spectral and moving-average representations, Ornstein–Uhlenbeck process, multiple Wiener–Itô integrals, chaos expansion of variables and processes

Here we initiate the study of Brownian motion, arguably the single most important object of modern probability. Indeed, we will see in Chapters 22–23 how the Gaussian limit theorems of Chapter 6 extend to pathwise approximations of broad classes of random walks and discrete-time martingales by a Brownian motion. In Chapter 19 we show how every continuous local martingale can be represented as a time-changed Brownian motion. Similarly, we will see in Chapters 32–33 how large classes of diffusion processes may be constructed from Brownian motion by various pathwise transformations. Finally, the close relationship between Brownian motion and classical potential theory is explored in Chapter 34.

The easiest construction of Brownian motion is via an isonormal Gaussian process on $L^2(\mathsf{R}_+)$, whose existence is a consequence of the characteristic spherical symmetry of multi-variate Gaussian distributions. Among the many basic properties of Brownian motion, this chapter covers the Hölder continuity and existence of quadratic variation, the strong Markov and reflection properties, the three arcsine laws, and the law of the iterated logarithm.

The values of an isonormal Gaussian process on $L^2(\mathsf{R}_+)$ can be identified with integrals of L^2-functions with respect to an associated Brownian motion. Many processes of interest have representations in terms of such integrals, and in particular we consider spectral and moving-average representations of stationary Gaussian processes. More generally, we introduce the multiple Wiener–Itô integrals $\zeta^n f$ of functions $f \in L^2(\mathsf{R}_+^n)$, and establish the fundamental chaos expansion of Brownian L^2-functionals.

The present material is related to practically every other chapter in the book. Thus, we refer to Chapter 6 for the definition of Gaussian distributions and the basic Gaussian limit theorem, to Chapter 8 for the transfer theorem, to Chapter 9 for properties of martingales and optional times, to Chapter 11

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_15

for basic facts about Markov processes, to Chapter 12 for similarities with random walks, to Chapter 27 for some basic symmetries, and to Chapter 15 for analogies with the Poisson process.

Our study of Brownian motion per se is continued in Chapter 19 with the basic recurrence or transience dichotomy, some further invariance properties, and a representation of Brownian martingales. Brownian local time and additive functionals are studied in Chapter 29. In Chapter 34 we consider some basic properties of Brownian hitting distributions, and examine the relationship between excessive functions and additive functionals of Brownian motion. A further discussion of multiple integrals and chaos expansions appears in Chapters 19 and 21.

To begin with some basic definitions, we say that a process X on a parameter space T is *Gaussian*, if the random variable $c_1X_{t_1}+\cdots+c_nX_{t_n}$ is Gaussian[1] for every choice of $n \in \mathsf{N}$, $t_1,\dots,t_n \in T$, and $c_1,\dots,c_n \in \mathsf{R}$. This holds in particular if the X_t are independent Gaussian random variables. A Gaussian process X is said to be *centered* if $EX_t = 0$ for all $t \in T$. We also say that the processes X^k on T_k, $k \in I$, are *jointly Gaussian*, if the combined process $X = \{X_t^k;\, t \in T_k,\, k \in I\}$ is Gaussian. The latter condition is certainly fulfilled when the Gaussian processes X^k are independent.

Some simple facts clarify the fundamental role of the covariance function. As usual, we take all distributions to be defined on the σ-field generated by all evaluation maps.

Lemma 14.1 *(covariance and independence) Let X and X^k, $k \in I$, be jointly Gaussian processes on an index set T. Then*

(i) *the distribution of X is uniquely determined by the functions*

$$m_t = EX_t, \quad r_{s,t} = \mathrm{Cov}(X_s, X_t), \qquad s,t \in T,$$

(ii) *the processes X^k are independent*[2] *iff*

$$\mathrm{Cov}(X_s^j, X_t^k) = 0, \qquad s \in T_j,\ t \in T_k, \quad j \neq k \text{ in } I.$$

Proof: (i) If the processes X, Y are Gaussian with the same mean and covariance functions, then the mean and variance agree for the variables $c_1X_{t_1} + \cdots + c_nX_{t_n}$ and $c_1Y_{t_1} + \cdots + c_nY_{t_n}$ for any $c_1,\dots,c_n \in \mathsf{R}$ and $t_1,\dots,t_n \in T$, $n \in \mathsf{N}$. Since the latter are Gaussian, their distributions must then agree. Hence, the Cramér–Wold theorem yields $(X_{t_1},\dots,X_{t_n}) \overset{d}{=} (Y_{t_1},\dots,Y_{t_n})$ for any $t_1,\dots,t_n \in T$, $n \in \mathsf{N}$, and so $X \overset{d}{=} Y$ by Proposition 4.2.

(ii) Assume the stated condition. To prove the asserted independence, we may take I to be finite. Introduce some independent processes Y^k, $k \in I$, with the same distributions as the X^k, and note that the combined processes

[1] also said to be *normally distributed*

[2] This (mutual) independence must not be confused with independence between the values of $X^k = (X_t^k)$ for each k.

$X = (X^k)$ and $Y = (Y^k)$ have the same means and covariances. Hence, their joint distributions agree by part (i). In particular, the independence between the processes Y^k implies the corresponding property for the processes X^k. $\square$

The multi-variate Gaussian distributions are characterized by a simple symmetry property:

Proposition 14.2 *(spherical symmetry, Maxwell) For independent random variables $\xi_1, \ldots, \xi_d$ with $d \geq 2$, these conditions are equivalent:*

(i) *$\mathcal{L}(\xi_1, \ldots, \xi_d)$ is spherically symmetric,*

(ii) *the ξ_k are i.i.d. centered Gaussian.*

Proof, (i) $\Rightarrow$ (ii): Assuming (i), let φ be the common characteristic function of $\xi_1, \ldots, \xi_d$. Noting that $-\xi_1 \overset{d}{=} \xi_1$, we see that φ is real and symmetric. Since also $s\,\xi_1 + t\,\xi_2 \overset{d}{=} \xi_1\sqrt{s^2+t^2}$, we obtain the functional equation

$$\varphi(s)\,\varphi(t) = \varphi\Big(\sqrt{s^2+t^2}\Big), \quad s, t \in \mathsf{R}.$$

This is a Cauchy equation for the function $\psi(x) = \varphi(|x|^{1/2})$, and we get $\varphi(t) = e^{-c\,t^2}$ for a constant c, which is positive since $|\varphi| \leq 1$.

(ii) $\Rightarrow$ (i): Assume (ii). By scaling we may take the ξ_k to be i.i.d. $N(0,1)$. Then $(\xi_1, \ldots, \xi_d)$ has the joint density

$$\prod_k (2\pi)^{-1/2}\, e^{-x_k^2/2} = (2\pi)^{-d/2}\, e^{-|x|^2/2}, \quad x \in \mathsf{R}^d,$$

and the spherical symmetry follows from that for $|x|^2 = x_1^2 + \cdots + x_d^2$. $\square$

In infinite dimensions, the Gaussian property is essentially a consequence of the spherical symmetry alone, without any requirement of independence.

Theorem 14.3 *(unitary invariance, Schoenberg, Freedman) For an infinite sequence of random variables $\xi_1, \xi_2, \ldots$, these conditions are equivalent:*

(i) *$\mathcal{L}(\xi_1, \ldots, \xi_n)$ is spherically symmetric for every $n \in \mathsf{N}$,*

(ii) *the ξ_k are conditionally i.i.d. $N(0, \sigma^2)$, given a random variable $\sigma^2 \geq 0$.*

Proof: Under (i) the ξ_k are exchangeable, and so by de Finetti's Theorem 27.2 below they are conditionally μ-i.i.d. for a random probability measure μ on R. By the law of large numbers,

$$\mu B = \lim_{n\to\infty} n^{-1} \sum_{k \leq n} 1\{\xi_k \in B\} \text{ a.s.}, \quad B \in \mathcal{B},$$

and in particular μ is a.s. $\{\xi_3, \xi_4, \ldots\}$-measurable. By spherical symmetry, we have for any orthogonal transformation T on R^2

$$P\Big\{(\xi_1, \xi_2) \in B \,\Big|\, \xi_3, \ldots, \xi_n\Big\} = P\Big\{T(\xi_1, \xi_2) \in B \,\Big|\, \xi_3, \ldots, \xi_n\Big\}, \quad B \in \mathcal{B}_{\mathsf{R}^2}.$$

As $n \to \infty$, we get $\mu^2 = \mu^2 \circ T^{-1}$ a.s. Applying this to a countable, dense set of mappings T, we see that the exceptional null set can be chosen to be independent of T. Thus, μ^2 is a.s. spherically symmetric, and so μ is a.s. centered Gaussian by Proposition 14.2. It remains to take $\sigma^2 = \int x^2 \mu(dx)$. □

Now fix a real, separable Hilbert space[3] H with inner product $\langle \cdot, \cdot \rangle$. By an *isonormal Gaussian process* on H we mean a centered Gaussian process ζh, $h \in H$, such that $E(\zeta h \, \zeta k) = \langle h, k \rangle$ for all $h, k \in H$. For its existence, fix any ortho-normal basis (ONB) $h_1, h_2, \ldots \in H$, and let $\xi_1, \xi_2, \ldots$ be i.i.d. $N(0,1)$. Then for any element $h = \sum_j b_j h_j$ we define $\zeta h = \sum_j b_j \xi_j$, where the series converges a.s. and in L^2, since $\sum_j b_j^2 < \infty$. Note that ζ is centered Gaussian, and also *linear*, in the sense that

$$\zeta(ah + bk) = a\,\zeta h + b\,\zeta k \ \text{ a.s.}, \quad h, k \in H, \ \ a, b \in \mathsf{R}.$$

If also $j = \sum_j c_j h_j$, we obtain

$$\begin{aligned} E(\zeta h \, \zeta k) &= \textstyle\sum_{i,j} b_i c_j \, E(\xi_i \, \xi_j) \\ &= \textstyle\sum_i b_i c_i = \langle h, k \rangle. \end{aligned}$$

By Lemma 14.1, the stated conditions determine uniquely the distribution of ζ. In particular, the symmetry in Proposition 14.2 extends to a distributional invariance of ζ under unitary transformations on H.

The Gaussian distributions arise naturally in the context of processes with independent increments. A similar Poisson criterion is given in Theorem 15.10.

Theorem 14.4 *(independent increments, Lévy) Let X be a continuous process in R^d with $X_0 = 0$. Then these conditions are equivalent:*

(i) *X has independent increments,*

(ii) *X is Gaussian, and there exist some continuous functions b in R^d and a in R^{d^2}, where a has non-negative definite increments, such that*

$$\mathcal{L}(X_t - X_s) = N(b_t - b_s, \, a_t - a_s), \quad s < t.$$

Proof, (i)⇒(ii): Fix any $s < t$ in R_+ and $u \in \mathsf{R}^d$. For every $n \in \mathsf{N}$, divide the interval $[s, t]$ into n sub-intervals of equal length, and let $\xi_{n1}, \ldots, \xi_{nn}$ be the corresponding increments of uX. Then $\max_j |\xi_{nj}| \to 0$ a.s. since X is continuous, and so the sum $u(X_t - X_s) = \sum_j \xi_{nj}$ is Gaussian by Theorem 6.16. Since the increments of X are independent, the entire process X is then Gaussian. Writing $b_t = EX_t$ and $a_t = \mathrm{Cov}(X_t)$, we get by linearity and independence

$$\begin{aligned} E(X_t - X_s) &= EX_t - EX_s \\ &= b_t - b_s, \\ \mathrm{Cov}(X_t - X_s) &= \mathrm{Cov}(X_t) - \mathrm{Cov}(X_s) \\ &= a_t - a_s, \qquad s < t. \end{aligned}$$

[3]See Appendix 3 for some basic properties of Hilbert spaces.

The continuity of X yields $X_s \xrightarrow{d} X_t$ as $s \to t$, and so $b_s \to b_t$ and $a_s \to a_t$. Thus, both functions are continuous.

(ii)⇒(i): Use Lemma 14.1. □

It is not so obvious that *continuous* processes such as in Theorem 14.4 really exist. If X has stationary, independent increments, the mean and covariance functions would clearly be linear. For $d = 1$ the simplest choice is to take $b = 0$ and $a_t = t$, so that $X_t - X_s$ is $N(0, t-s)$ for all $s < t$. We prove the existence of such a process and estimate its local modulus of continuity.

Theorem 14.5 *(Brownian motion, Thiele, Bachelier, Wiener) There exists a process B on R_+ with $B_0 = 0$, such that*

(i) *B has stationary, independent increments,*
(ii) *B_t is $N(0,t)$ for every $t \geq 0$,*
(iii) *B is a.s. locally Hölder continuous of order p, for every $p \in (0, \frac{1}{2})$.*

In Theorem 14.18 we will see that the stated order of Hölder continuity can't be improved. Related results are given in Proposition 14.10 and Lemma 22.7 below.

Proof: Let ζ be an isonormal Gaussian process on $L^2(\mathsf{R}_+, \lambda)$, and introduce a process $B_t = \zeta 1_{[0,t]}$, $t \geq 0$. Since indicator functions of disjoint intervals are orthogonal, the increments of B are uncorrelated and hence independent. Furthermore, $\|1_{(s,t]}\|^2 = t - s$ for any $s \leq t$, and so $B_t - B_s$ is $N(0, t-s)$. For any $s \leq t$ we get

$$\begin{aligned} B_t - B_s &\overset{d}{=} B_{t-s} \\ &\overset{d}{=} (t-s)^{1/2} B_1, \end{aligned} \tag{1}$$

whence

$$E|B_t - B_s|^p = (t-s)^{p/2} E|B_1|^p < \infty, \quad p > 0.$$

The asserted Hölder continuity now follows by Theorem 4.23. □

A process B as in Theorem 14.5 is called a (standard) *Brownian motion*[4]. A d-dimensional Brownian motion is a process $B = (B^1, \dots, B^d)$ in R^d, where $B^1, \dots, B^d$ are independent, one-dimensional Brownian motions. By Proposition 14.2, the distribution of B is invariant under orthogonal transformations of R^d. Furthermore, any continuous process X in R^d with stationary independent increments and $X_0 = 0$ can be written as $X_t = bt + aB_t$ for some vector b and matrix a.

Other important Gaussian processes can be constructed from a Brownian motion. For example, we define a *Brownian bridge* as a process on $[0,1]$ distributed as $X_t = B_t - tB_1$, $t \in [0,1]$. An easy computation shows that X has covariance function $r_{s,t} = s(1-t)$, $0 \leq s \leq t \leq 1$.

[4]More appropriately, it is also called a *Wiener process*.

The Brownian motion and bridge have many nice symmetry properties. For example, if B is a Brownian motion, then so is $-B$, as well as the process $r^{-1}B(r^2t)$ for any $r > 0$. The latter transformations, known as *Brownian scaling*, are especially useful. We also note that, for any $u > 0$, the processes $B_{u\pm t} - B_u$ are Brownian motions on R_+ and $[0, u]$, respectively. If B is instead a Brownian bridge, then so are the processes $-B_t$ and B_{1-t}.

We list some less obvious invariance properties. Further, possibly random mappings preserving the distribution of a Brownian motion or bridge are exhibited in Theorem 14.11, Lemma 14.14, and Proposition 27.16.

Lemma 14.6 *(scaling and inversion)*

(i) *If B is a Brownian motion, then so is the process $tB_{1/t}$, whereas these processes are Brownian bridges:*
$$(1-t)B_{t/(1-t)}, \quad tB_{(1-t)/t}, \qquad t \in (0,1),$$

(ii) *If B is a Brownian bridge, then so is the process B_{1-t}, whereas these processes are Brownian motions:*
$$(1+t)B_{t/(1+t)}, \quad (1+t)B_{1/(1+t)}, \qquad t \geq 0.$$

Proof: Since the mentioned processes are centered Gaussian, it suffices by Lemma 14.1 to verify that they have the required covariance functions. This is clear from the expressions $s \wedge t$ and $(s \wedge t)(1 - s \vee t)$ for the covariance functions of the Brownian motion and bridge. □

Combining Proposition 11.5 with Theorem 14.4, we note that any space- and time-homogeneous, continuous Markov process in R^d has the form $aB_t + b\,t + c$ for a Brownian motion B in R^d, a $d \times d$ matrix a, and some vectors $b, c \in \mathsf{R}^d$. We turn to a general characterization of Gaussian Markov processes. As before, we use the convention $0/0 = 0$.

Proposition 14.7 *(Markov property) Let X be a Gaussian process on an index set $T \subset \mathsf{R}$, and put $r_{s,t} = \mathrm{Cov}(X_s, X_t)$. Then X is Markov iff*
$$r_{s,u} = \frac{r_{s,t}\, r_{t,u}}{r_{t,t}}, \quad s \leq t \leq u \ \text{ in } T. \tag{2}$$
If X is also stationary on $T = \mathsf{R}$, then $r_{s,t} = a\, e^{-b|s-t|}$ for some constants $a \geq 0$ and $b \in [0, \infty]$.

Proof: Subtracting the means, if necessary, we may take $EX_t \equiv 0$. Fixing any $t \leq u$ in T, we may choose $a \in \mathsf{R}$ such that $X'_u \equiv X_u - aX_t \perp X_t$. Then $a = r_{t,u}/r_{t,t}$ when $r_{t,t} \neq 0$, and if $r_{t,t} = 0$ we may take $a = 0$. By Lemma 14.1 we get $X'_u \perp\!\!\!\perp X_t$.

Now let X be Markov, and fix any $s \leq t$. Then $X_s \perp\!\!\!\perp_{X_t} X_u$, and so $X_s \perp\!\!\!\perp_{X_t} X'_u$. Since also $X_t \perp\!\!\!\perp X'_u$ by the choice of a, Proposition 8.12 yields $X_s \perp\!\!\!\perp X'_u$. Hence, $r_{s,u} = a\, r_{s,t}$, and (2) follows as we insert the expression for a.

Conversely, (2) implies $X_s \perp X'_u$ for all $s \le t$, and so $\mathcal{F}_t \perp\!\!\!\perp X'_u$ by Lemma 14.1, where $\mathcal{F}_t = \sigma\{X_s;\, s \le t\}$. Then Proposition 8.12 yields $\mathcal{F}_t \perp\!\!\!\perp_{X_t} X_u$, which is the required Markov property of X at t.

If X is stationary, we have $r_{s,t} = r_{|s-t|,0} = r_{|s-t|}$, and (2) reduces to the Cauchy equation $r_0\, r_{s+t} = r_s\, r_t$, $s,t \ge 0$, allowing the only bounded solutions $r_t = a\, e^{-bt}$. □

A continuous, centered Gaussian process X on R with covariance function $r_{s,t} = e^{-|t-s|}$ is called a stationary *Ornstein–Uhlenbeck process*[5]. Such a process can be expressed as $X_t = e^{-t} B(e^{2t})$, $t \in \mathsf{R}$, in terms of a Brownian motion B. The previous result shows that the Ornstein–Uhlenbeck process is essentially the only stationary Gaussian process that is also a Markov process.

We consider some basic path properties of Brownian motion.

Lemma 14.8 *(level sets) For a Brownian motion or bridge B,*

$$\lambda\Big\{t;\, B_t = u\Big\} = 0 \quad a.s., \qquad u \in \mathsf{R}.$$

Proof: Defining $X^n_t = B_{[nt]/n}$, $t \in \mathsf{R}_+$ or $[0,1]$, $n \in \mathsf{N}$, we note that $X^n_t \to B_t$ for every t. Since the processes X^n are product-measurable on $\Omega \times \mathsf{R}_+$ or $\Omega \times [0,1]$, so is B. Hence, Fubini's theorem yields

$$E\,\lambda\Big\{t;\, B_t = u\Big\} = \int P\{B_t = u\}\, dt = 0, \qquad u \in \mathsf{R}. \qquad \square$$

Next we show that Brownian motion has locally finite quadratic variation. Extensions to general semi-martingales appear in Proposition 18.17 and Corollary 20.15. A sequence of partitions is said to be *nested*, if each of them is a refinement of the previous one.

Theorem 14.9 *(quadratic variation, Lévy) Let B be a Brownian motion, and fix any $t > 0$. Then for any partitions $0 = t_{n,0} < t_{n,1} < \cdots < t_{n,k_n} = t$ with $h_n \equiv \max_k (t_{n,k} - t_{n,k-1}) \to 0$, we have*

$$\zeta_n \equiv \textstyle\sum_k \Big(B_{t_{n,k}} - B_{t_{n,k-1}}\Big)^2 \to t \quad in \quad L^2. \qquad (3)$$

The convergence holds a.s. when the partitions are nested.

Proof (Doob): To prove (3), we see from the scaling property $B_t - B_s \overset{d}{=} |t-s|^{1/2} B_1$ that

$$\begin{aligned} E\zeta_n &= \textstyle\sum_k E\Big(B_{t_{n,k}} - B_{t_{n,k-1}}\Big)^2 \\ &= \textstyle\sum_k (t_{n,k} - t_{n,k-1})\, E B_1^2 = t, \\ \mathrm{Var}(\zeta_n) &= \textstyle\sum_k \mathrm{Var}\Big(B_{t_{n,k}} - B_{t_{n,k-1}}\Big)^2 \\ &= \textstyle\sum_k (t_{n,k} - t_{n,k-1})^2\, \mathrm{Var}(B_1^2) \\ &\le h_n\, t\, E B_1^4 \to 0. \end{aligned}$$

[5]In Chapters 21 and 32 we use a slightly different normalization.

For nested partitions, the a.s. convergence follows once we show that the sequence (ζ_n) is a reverse martingale, that is

$$E\Big(\zeta_{n-1} - \zeta_n \,\Big|\, \zeta_n, \zeta_{n+1}, \ldots\Big) = 0 \ \text{ a.s.}, \quad n \in \mathbb{N}. \tag{4}$$

Inserting intermediate partitions if necessary, we may assume that $k_n = n$ for all n. Then there exist some $t_1, t_2, \ldots \in [0, t]$, such that the n-th partition has division points $t_1, \ldots, t_n$. To verify (4) for a fixed n, we introduce an auxiliary random variable $\vartheta \perp\!\!\!\perp B$ with $P\{\vartheta = \pm 1\} = \frac{1}{2}$, and replace B by the Brownian motion

$$B'_s = B_{s\wedge t_n} + \vartheta\,(B_s - B_{s\wedge t_n}), \quad s \geq 0.$$

Since the sums $\zeta_n, \zeta_{n+1}, \ldots$ for B' agree with those for B, whereas $\zeta_n - \zeta_{n-1}$ is replaced by $\vartheta(\zeta_n - \zeta_{n-1})$, it suffices to show that $E\{\vartheta(\zeta_n - \zeta_{n-1}) \,|\, \zeta_n, \zeta_{n+1}, \ldots\} = 0$ a.s. This is clear from the choice of ϑ, if we first condition on $\zeta_{n-1}, \zeta_n, \ldots$. □

The last result shows that the linear variation of B is locally unbounded, which is why we can't define the stochastic integral $\int V dB$ as an ordinary Stieltjes integral, prompting us in Chapter 18 to use a more sophisticated approach. We give some more precise information about the irregularity of paths, supplementing the regularity property in Theorem 14.5.

Proposition 14.10 *(path irregularity, Paley, Wiener, & Zygmund) Let B be a Brownian motion in $\mathbb{R}$, and fix any $p > \frac{1}{2}$. Then a.s.*

(i) *B is nowhere*[6] *Hölder continuous of order p,*

(ii) *B is nowhere differentiable.*

Proof: (i) Fix any $p > \frac{1}{2}$ and $m \in \mathbb{N}$. If B_ω is Hölder continuous of order p at a point $t_\omega \in [0, 1]$, then for $s \in [0, 1]$ with $|s - t| \leq mh$ we have

$$\begin{aligned} |\Delta_h B_s| &\leq |B_s - B_t| + |B_{s+h} - B_t| \\ &\lesssim (mh)^p \lesssim h^p. \end{aligned}$$

Applying this to m fixed, disjoint intervals of length h between the bounds $t \pm mh$, we see that the probability of the stated Hölder continuity is bounded by expressions of the form

$$\begin{aligned} h^{-1}\Big(P\Big\{|B_h| \leq c\,h^p\Big\}\Big)^m &= h^{-1}\Big(P\Big\{|B_1| \leq c\,h^{p-1/2}\Big\}\Big)^m \\ &\lesssim h^{-1} h^{m(p-1/2)}, \end{aligned}$$

for some constants $c > 0$. Choosing $m > (p - \frac{1}{2})^{-1}$, we see that the right-hand side tends to 0 as $h \to 0$.

[6]Thus, for any ω outside a P-null set, there is no point $t \geq 0$ where the path $B(\omega, \cdot)$ has the stated regularity property. This is clearly stronger than the corresponding statement for fixed points $t \geq 0$ or intervals $I \subset \mathbb{R}_+$. Since the stated events may not be measurable, the results should be understood in the sense of inner measure or completion.

(ii) If B_ω is differentiable at a point $t_\omega \geq 0$, it is Lipschitz continuous at t, hence Hölder continuous of order 1, which contradicts (i) for $p = 1$. □

Proposition 11.5 shows that Brownian motion B is a space-homogeneous Markov process with respect to its induced filtration. If the Markov property holds for a more general filtration $\mathcal{F} = (\mathcal{F}_t)$ —i.e., if B is $\mathcal{F}$-adapted and such that the process $B'_t = B_{s+t} - B_s$ is independent of $\mathcal{F}_s$ for each $s \geq 0$ —we say that B is a Brownian motion with respect to $\mathcal{F}$ or an *$\mathcal{F}$-Brownian motion*. In particular, we may take $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{N}$, $t \geq 0$, where $\mathcal{G}$ is the filtration induced by B and $\mathcal{N} = \sigma\{N \subset A;\, A \in \mathcal{A},\, PA = 0\}$. With this choice, $\mathcal{F}$ becomes right-continuous by Corollary 9.26.

We now extend the Markov property of B to suitable optional times. A more general version of this result appears in Theorem 17.17. As in Chapter 9 we write $\mathcal{F}_t^+ = \mathcal{F}_{t+}$.

Theorem 14.11 *(strong Markov property, Hunt) For an $\mathcal{F}$-Brownian motion B in R^d and an $\mathcal{F}^+$-optional time $\tau < \infty$, we have these equivalent properties:*

(i) *B has the strong $\mathcal{F}^+$-Markov property at τ,*

(ii) *the process $B'_t = B_{\tau+t} - B_\tau$ satisfies $B \overset{d}{=} B' \perp\!\!\!\perp \mathcal{F}_\tau^+$.*

Proof: (ii) As in Lemma 9.4, there exist some optional times $\tau_n \to \tau$ with $\tau_n \geq \tau + 2^{-n}$ taking countably many values. Then $\mathcal{F}_\tau^+ \subset \bigcap_n \mathcal{F}_{\tau_n}$ by Lemmas 9.1 and 9.3, and so by Proposition 11.9 and Theorem 11.10, each process $B_t^n = B_{\tau_n+t} - B_{\tau_n}$, $t \geq 0$, is a Brownian motion independent of $\mathcal{F}_\tau^+$. The continuity of B yields $B_t^n \to B'_t$ a.s. for every t. For any $A \in \mathcal{F}_\tau^+$ and $t_1, \ldots, t_k \in \mathsf{R}_+$, $k \in \mathsf{N}$, and for bounded, continuous functions $f : \mathsf{R}^k \to \mathsf{R}$, we get by dominated convergence

$$E\Big\{f(B'_{t_1}, \ldots, B'_{t_k}); A\Big\} = Ef(B_{t_1}, \ldots, B_{t_k}) \cdot PA.$$

The general relation $\mathcal{L}(B'; A) = \mathcal{L}(B) \cdot PA$ now follows by a straightforward extension argument. □

When B is a Brownian motion in R^d, the process $X \overset{d}{=} |B|$ is called a *Bessel process* of order d. More general Bessel processes are obtained as solutions to suitable SDEs. We show that $|B|$ inherits the strong Markov property from B.

Corollary 14.12 *(Bessel processes) For an $\mathcal{F}$-Brownian motion B in R^d, $|B|$ is a strong $\mathcal{F}^+$-Markov process.*

Proof: By Theorem 14.11, it is enough to show that $|B + x| \overset{d}{=} |B + y|$ whenever $|x| = |y|$. Then choose an orthogonal transformation T on R^d with $Tx = y$, and note that

$$\begin{aligned} |B + x| &= |T(B + x)| \\ &= |TB + y| \overset{d}{=} |B + y|. \end{aligned}$$

□

The strong Markov property can be used to derive the distribution of the maximum of Brownian motion up to a fixed time. A stronger result will be obtained in Corollary 29.3.

Proposition 14.13 *(maximum process, Bachelier) Let B be a Brownian motion in R, and define $M_t = \sup_{s\le t} B_s$, $t \ge 0$. Then*

$$M_t \overset{d}{=} M_t - B_t \overset{d}{=} |B_t|, \quad t \ge 0.$$

Our proof relies on the following continuous-time counterpart of Lemma 12.11, where the given Brownian motion B is compared with a suitably reflected version $\tilde{B}$.

Lemma 14.14 *(reflection principle) For a Brownian motion B and an associated optional time τ, we have $B \overset{d}{=} \tilde{B}$ with*

$$\tilde{B}_t = B_{t\wedge\tau} - (B_t - B_{t\wedge\tau}), \quad t \ge 0.$$

Proof: It suffices to compare the distributions up to a fixed time t, and so we may assume that $\tau < \infty$. Define $B_t^\tau = B_{\tau\wedge t}$ and $B'_t = B_{\tau+t} - B_\tau$. By Theorem 14.11, the process B' is a Brownian motion independent of (τ, B^τ). Since also $-B' \overset{d}{=} B'$, we get $(\tau, B^\tau, B') \overset{d}{=} (\tau, B^\tau, -B')$, and it remains to note that

$$\begin{aligned} B_t &= B_t^\tau + B'_{(t-\tau)_+}, \\ \tilde{B}_t &= B_t^\tau - B'_{(t-\tau)_+}, \qquad t \ge 0. \end{aligned}$$ □

Proof of Proposition 14.13: By scaling we may take $t = 1$. Applying Lemma 14.14 with $\tau = \inf\{t;\, B_t = x\}$ gives

$$P\Big\{M_1 \ge x,\, B_1 \le y\Big\} = P\Big\{\tilde{B}_1 \ge 2x - y\Big\}, \quad x \ge y \vee 0.$$

By differentiation, the joint distribution of (M_1, B_1) has density $-2\,\varphi'(2x-y)$, where φ denotes the standard normal density. Changing variables, we conclude that $(M_1, M_1 - B_1)$ has density $-2\,\varphi'(x+y)$, $x, y \ge 0$. In particular, M_1 and $M_1 - B_1$ have the same marginal density $2\,\varphi(x)$ on R_+. □

To prepare for the next major result, we prove another elementary sample path property.

Lemma 14.15 *(local extremes) The local maxima and minima of a Brownian motion or bridge are a.s. distinct.*

Proof: For a Brownian motion B and any intervals $I = [a, b]$ and $J = [c, d]$ with $b < c$, we may write

$$\sup_{t\in J} B_t - \sup_{t\in I} B_t = \sup_{t\in J}\,(B_t - B_c) + (B_c - B_b) - \sup_{t\in I}\,(B_t - B_b).$$

Here the distribution of second term on the right is diffuse, which extends by independence to the whole expression. In particular, the difference on the left is a.s. non-zero. Since I and J are arbitrary, this proves the result for local maxima. The proofs for local minima and the mixed case are similar.

The result for the Brownian bridge B^o follows from that for Brownian motion, since the distributions of the two processes are mutually absolutely continuous on every interval $[0,t]$ with $t<1$. To see this, construct from B and B^o the corresponding 'bridges'

$$\begin{aligned} X_s &= B_s - s\,t^{-1}B_t \\ Y_s &= B_s^o - s\,t^{-1}B_t^o, \quad s\in[0,t], \end{aligned}$$

and check that $B_t \perp\!\!\!\perp X \overset{d}{=} Y \perp\!\!\!\perp B_t^o$. The stated equivalence now follows since $N(0,t)\sim N(0,t(1-t))$ for any $t\in[0,1)$. □

The *arcsine law* may be defined[7] as the distribution of $\xi = \sin^2\alpha$ when α is $U(0,2\pi)$. Its name derives from the fact that

$$\begin{aligned} P\{\xi\le t\} &= P\Big\{|\sin\alpha|\le\sqrt{t}\Big\} \\ &= \tfrac{2}{\pi}\arcsin\sqrt{t}, \quad t\in[0,1]. \end{aligned}$$

Note that the arcsine distribution is symmetric about $\frac{1}{2}$, since

$$\begin{aligned} \xi &= \sin^2\alpha \overset{d}{=} \cos^2\alpha \\ &= 1-\sin^2\alpha \\ &= 1-\xi. \end{aligned}$$

We consider three basic functionals of Brownian motion, which are all arcsine distributed.

Theorem 14.16 *(arcsine laws, Lévy) For a Brownian motion B on $[0,1]$ with maximum M_1, these random variables are all arcsine distributed:*

$$\begin{aligned} \tau_1 &= \lambda\{t;\, B_t>0\}, \\ \tau_2 &= \inf\Big\{t;\, B_t = M_1\Big\}, \\ \tau_3 &= \sup\{t;\, B_t = 0\}. \end{aligned}$$

The relations $\tau_1 \overset{d}{=} \tau_2 \overset{d}{=} \tau_3$ may be compared with the discrete-time analogues in Theorem 12.12 and Corollary 27.8. In Theorems 16.18 and 22.11, the arcsine laws are extended by approximation to appropriate random walks and Lévy processes.

Proof (OK): To see that $\tau_1 \overset{d}{=} \tau_2$, let $n\in\mathsf{N}$, and conclude from Corollary 27.8 below that

$$n^{-1}\sum_{k\le n} 1\{B_{k/n}>0\} \overset{d}{=} n^{-1}\min\Big\{k\ge 0;\, B_{k/n} = \max_{j\le n} B_{j/n}\Big\}.$$

[7] Equivalently, we may take α to be $U(0,\pi)$ or $U(0,\pi/2)$.

By Lemma 14.15 the right-hand side tends a.s. to τ_2 as $n \to \infty$. To see that the left-hand side converges to τ_1, we note that by Lemma 14.8

$$\lambda\Big\{t \in [0,1];\ B_t > 0\Big\} + \lambda\Big\{t \in [0,1];\ B_t < 0\Big\} = 1 \quad \text{a.s.}$$

It remains to note that, for any open set $G \subset [0,1]$,

$$\liminf_{n\to\infty} n^{-1} \sum_{k\le n} 1_G(k/n) \ge \lambda G.$$

In case of τ_2, fix any $t \in [0,1]$, let ξ and η be independent $N(0,1)$, and let α be $U(0,2\pi)$. Using Proposition 14.13 and the circular symmetry of the distribution of (ξ,η), we get

$$\begin{aligned}
P\{\tau_2 \le t\} &= P\Big\{\sup_{s\le t}(B_s - B_t) \ge \sup_{s\ge t}(B_s - B_t)\Big\} \\
&= P\Big\{|B_t| \ge |B_1 - B_t|\Big\} \\
&= P\Big\{\, t\,\xi^2 \ge (1-t)\,\eta^2\Big\} \\
&= P\left\{\frac{\eta^2}{\xi^2+\eta^2} \le t\right\} \\
&= P\Big\{\sin^2\alpha \le t\Big\}.
\end{aligned}$$

In case of τ_3, we write instead

$$\begin{aligned}
P\{\tau_3 < t\} &= P\Big\{\sup_{s\ge t} B_s < 0\Big\} + P\Big\{\inf_{s\ge t} B_s > 0\Big\} \\
&= 2\,P\Big\{\sup_{s\ge t}(B_s - B_t) < -B_t\Big\} \\
&= 2\,P\Big\{|B_1 - B_t| < B_t\Big\} \\
&= P\Big\{|B_1 - B_t| < |B_t|\Big\} \\
&= P\{\tau_2 \le t\}.
\end{aligned}$$

□

The first two arcsine laws have simple counterparts for the Brownian bridge:

Proposition 14.17 *(uniform laws) For a Brownian bridge B with maximum M_1, these random variables are $U(0,1)$:*

$$\begin{aligned}
\tau_1 &= \lambda\{t;\ B_t > 0\}, \\
\tau_2 &= \inf\Big\{t;\ B_t = M_1\Big\}.
\end{aligned}$$

Proof: The relation $\tau_1 \overset{d}{=} \tau_2$ may be proved in the same way as for Brownian motion. To see that τ_2 is $U(0,1)$, write $(x) = x - [x]$, and consider for each $u \in [0,1]$ a process $B_t^u = B_{(u+t)} - B_u$, $t \in [0,1]$. Here clearly $B^u \overset{d}{=} B$ for each u, and the maximum of B^u occurs at $(\tau_2 - u)$. Hence, Fubini's theorem yields for any $t \in [0,1]$

$$\begin{aligned} P\{\tau_2 \le t\} &= \int_0^1 P\Big\{(\tau_2 - u) \le t\Big\}\, du \\ &= E\,\lambda\Big\{u;\ (\tau_2 - u) \le t\Big\} = t. \end{aligned} \qquad \square$$

From Theorem 14.5 we see that $t^{-c}B_t \to 0$ a.s. as $t \to 0$ for fixed $c \in [0, \frac{1}{2})$. The following classical result gives the exact growth rate of Brownian motion at 0 and ∞. Extensions to random walks and renewal processes appear in Corollaries 22.8 and 22.14, and a functional version is given in Theorem 24.18.

Theorem 14.18 *(laws of the iterated logarithm, Khinchin) For a Brownian motion B in R, we have a.s.*

$$\limsup_{t\to 0} \frac{B_t}{\sqrt{2t \log\log(1/t)}} = \limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = 1.$$

Proof: Since the Brownian inversion $\tilde{B}_t = tB_{1/t}$ of Lemma 14.6 converts the two formulas into one another, it is enough to prove the result for $t \to \infty$. Then note that as $u \to \infty$

$$\begin{aligned} \int_u^\infty e^{-x^2/2}\, dx &\sim u^{-1} \int_u^\infty x\, e^{-x^2/2}\, dx \\ &= u^{-1} e^{-u^2/2}. \end{aligned}$$

Letting $M_t = \sup_{s\le t} B_s$ and using Proposition 14.13, we hence obtain, uniformly in $t > 0$,

$$\begin{aligned} P\Big\{M_t > u\, t^{1/2}\Big\} &= 2P\Big\{B_t > u\, t^{1/2}\Big\} \\ &\sim (2/\pi)^{1/2}\, u^{-1} e^{-u^2/2}. \end{aligned}$$

Writing $h_t = (2t\log\log t)^{1/2}$, we get for any $r > 1$ and $c > 0$

$$P\Big\{M(r^n) > c\, h(r^{n-1})\Big\} \lesssim n^{-c^2/r}(\log n)^{-1/2}, \quad n \in \mathsf{N}.$$

Fixing $c > 1$ and choosing $r < c^2$, we see from the Borel–Cantelli lemma that

$$P\Big\{\limsup_{t\to\infty}\,(B_t/h_t) > c\Big\} \le P\Big\{M(r^n) > c\, h(r^{n-1}) \text{ i.o.}\Big\} = 0,$$

which implies $\limsup_{t\to\infty}(B_t/h_t) \le 1$ a.s.

To prove the reverse inequality, we may write

$$P\Big\{B(r^n) - B(r^{n-1}) > c\, h(r^n)\Big\} \gtrsim n^{-c^2 r/(r-1)}(\log n)^{-1/2}, \quad n \in \mathsf{N}.$$

Taking $c = \{(r-1)/r\}^{1/2}$, we get by the Borel–Cantelli lemma

$$\begin{aligned} \limsup_{t\to\infty} \frac{B_t - B_{t/r}}{h_t} &\ge \limsup_{n\to\infty} \frac{B(r^n) - B(r^{n-1})}{h(r^n)} \\ &\ge \left(\frac{r-1}{r}\right)^{1/2} \quad \text{a.s.} \end{aligned}$$

The previous upper bound yields $\limsup_{t\to\infty}(-B_{t/r}/h_t) \le r^{-1/2}$, and so by combination

$$\limsup_{t\to\infty} \frac{B_t}{h_t} \ge (1-r^{-1})^{1/2} - r^{-1/2} \quad \text{a.s.}$$

Letting $r \to \infty$, we obtain $\limsup_{t\to\infty}(B_t/h_t) \ge 1$ a.s. □

In Theorem 14.5 we constructed a Brownian motion B from an isonormal Gaussian process ζ on $L^2(\mathsf{R}_+, \lambda)$ by setting $B_t = \zeta 1_{[0,t]}$ a.s. for all $t \ge 0$. Conversely, given a Brownian motion B on R_+, Theorem 8.17 ensures the existence of a corresponding isonormal Gaussian process ζ. Since every function $h \in L^2(\mathsf{R}_+, \lambda)$ can be approximated by simple step functions, as in the proof of Lemma 1.37, the random variables ζh are a.s. unique. They can also be constructed directly from B, as suitable *Wiener integrals* $\int h\, dB$. Since the latter may fail to exist in the pathwise Stieltjes sense, a different approach is required.

For a first step, consider the class $\mathcal{S}$ of simple step functions of the form

$$h_t = \sum_{j\le n} a_j 1_{(t_{j-1},t_j]}(t), \quad t \ge 0,$$

where $n \in \mathsf{Z}_+$, $0 = t_0 < \cdots < t_n$, and $a_1, \ldots, a_n \in \mathsf{R}$. For any $h \in \mathcal{S}$, we define the integral ζh in the obvious way as

$$\begin{aligned} \zeta h &= \int_0^\infty h_t\, dB_t = Bh \\ &= \sum_{j\le n} a_j (B_{t_j} - B_{t_{j-1}}), \end{aligned}$$

which is clearly centered Gaussian with variance

$$\begin{aligned} E(\zeta h)^2 &= \sum_{j\le n} a_j^2 (t_j - t_{j-1}) \\ &= \int_0^\infty h_t^2\, dt = \|h\|^2, \end{aligned}$$

where $\|h\|$ denotes the norm in $L^2(\mathsf{R}_+, \lambda)$. Thus, the integration $h \mapsto \zeta h = \int h\, dB$ defines a linear isometry from $\mathcal{S} \subset L^2(\mathsf{R}_+, \lambda)$ into $L^2(\Omega, P)$.

Since $\mathcal{S}$ is dense in $L^2(\mathsf{R}_+, \lambda)$, the integral extends by continuity to a linear isometry $h \mapsto \zeta h = \int h\, dB$ from $L^2(\lambda)$ to $L^2(P)$. The variable ζh is again centered Gaussian for every $h \in L^2(\lambda)$, and by linearity the whole process $h \mapsto \zeta h$ is Gaussian. By a polarization argument, the integration preserves inner products, in the sense that

$$E(\zeta h\, \zeta k) = \int_0^\infty h_t\, k_t\, dt = \langle h, k\rangle, \quad h, k \in L^2(\lambda).$$

We consider two general representations of stationary Gaussian processes in terms of Wiener integrals ζh. Here a complex notation is convenient. By a *complex, isonormal Gaussian process* on a (real) Hilbert space H, we mean a process $\zeta = \xi + i\eta$ on H such that ξ and η are independent, real, isonormal

Gaussian processes on H. For any $f = g + ih$ with $g, h \in H$, we define $\zeta f = \xi g - \eta h + i(\xi h + \eta g)$.

Now let X be a stationary, centered Gaussian process on R with covariance function $r_t = E\, X_s X_{s+t}$, $s, t \in \mathsf{R}$. Then r is non-negative definite, and it is further continuous whenever X is continuous in probability, in which case Bochner's theorem yields a unique *spectral representation*

$$r_t = \int_{-\infty}^{\infty} e^{itx} \mu(dx), \quad t \in \mathsf{R},$$

in terms of a bounded, symmetric *spectral measure* μ on R.

We may derive a similar spectral representation of the process X itself. By a different argument, the result extends to suitable non-Gaussian processes. As usual, we take the basic probability space to be rich enough to support the required randomization variables.

Proposition 14.19 *(spectral representation, Stone, Cramér) Let X be an L^2-continuous, stationary, centered Gaussian process on R with spectral measure μ. Then there exists a complex, isonormal Gaussian process ζ on $L^2(\mu)$, such that*

$$X_t = \Re \int_{-\infty}^{\infty} e^{itx} d\zeta_x \quad a.s., \quad t \in \mathsf{R}. \tag{5}$$

Proof: Writing Y for the right-hand side of (5), we obtain

$$\begin{aligned} E\, Y_s Y_t &= E \int \Big(\cos sx \; d\xi_x - \sin sx \; d\eta_x\Big) \int \Big(\cos tx \; d\xi_x - \sin tx \; d\eta_x\Big) \\ &= \int \Big(\cos sx \; \cos tx - \sin sx \; \sin tx\Big) \, \mu(dx) \\ &= \int \cos(s-t)x \; \mu(dx) \\ &= \int e^{i(s-t)x} \mu(dx) = r_{s-t}. \end{aligned}$$

Since X, Y are both centered Gaussian, Lemma 14.1 yields $Y \overset{d}{=} X$. Since X and ζ are further continuous and defined on the separable spaces $L^2(X)$ and $L^2(\mu)$, they may be regarded as random elements in suitable Borel spaces. The a.s. representation (5) then follows by Theorem 8.17. □

Our second representation requires a suitable regularity condition on the spectral measure μ. Write λ for Lebesgue measure on R.

Proposition 14.20 *(moving average representation) Let X be an L^2-continuous, stationary, centered Gaussian process on R with spectral measure $\mu \ll \lambda$. Then there exist an isonormal Gaussian process ζ on $L^2(\lambda)$ and a function $f \in L^2(\lambda)$, such that*

$$X_t = \int_{-\infty}^{\infty} f_{t-s} \, d\zeta_s \quad a.s., \quad t \in \mathsf{R}. \tag{6}$$

Proof: Choose a symmetric density $g \ge 0$ of μ, and define $h = \sqrt{g}$. Since h lies in $L^2(\lambda)$, it admits a Fourier transform in the sense of Plancherel

$$f_s = \hat{h}_s = (2\pi)^{-1/2} \lim_{a\to\infty} \int_{-a}^{a} e^{isx} h_x \, dx, \quad s \in \mathsf{R}, \tag{7}$$

which is again real and square integrable. For each $t \in \mathsf{R}$, the function $k_x = e^{-itx} h_x$ has Fourier transform $\hat{k}_s = f_{s-t}$, and so by Parseval's identity,

$$\begin{aligned} r_t &= \int_{-\infty}^{\infty} e^{itx}\, h_x^2 \, dx \\ &= \int_{-\infty}^{\infty} h_x \bar{k}_x \, dx = \int_{-\infty}^{\infty} f_s f_{s-t} \, ds. \end{aligned} \tag{8}$$

Now let ζ be an isonormal Gaussian process on $L^2(\lambda)$. For f as in (7), define a process Y on R by the right-hand side of (6). Then (8) gives $E\, Y_s Y_{s+t} = r_t$ for all $s, t \in \mathsf{R}$, and so $Y \overset{d}{=} X$ by Lemma 14.1. Again, Theorem 8.17 yields the desired a.s. representation of X. $\square$

For an example, we consider a moving average representation of the stationary Ornstein–Uhlenbeck process. Then let ζ be an isonormal Gaussian process on $L^2(\mathsf{R}, \lambda)$, and introduce the process

$$X_t = \sqrt{2} \int_{-\infty}^{t} e^{s-t} \, d\zeta_s, \quad t \ge 0,$$

which is clearly centered Gaussian. Furthermore,

$$\begin{aligned} r_{s,t} &= E\, X_s X_t \\ &= 2 \int_{-\infty}^{s\wedge t} e^{u-s}\, e^{u-t} \, du \\ &= e^{-|t-s|}, \quad s, t \in \mathsf{R}, \end{aligned}$$

as desired. The Markov property of X is clear from the fact that

$$X_t = e^{s-t} X_s + \sqrt{2} \int_s^t e^{u-t} \, d\zeta_u, \quad s \le t.$$

We turn to a discussion of *multiple Wiener–Itô integrals*[8] ζ^n with respect to an isonormal Gaussian process ζ on a separable, infinite-dimensional Hilbert space H. Without loss of generality, we may choose H to be of the form $L^2(S, \mu)$, so that the tensor product $H^{\otimes n}$ may be identified with $L^2(S^n, \mu^{\otimes n})$, where $\mu^{\otimes n}$ denotes the n-fold product measure $\mu \otimes \cdots \otimes \mu$. The tensor product $\bigotimes_{k\le n} h_k = h_1 \otimes \cdots \otimes h_n$ of $h_1, \dots, h_n \in H$ is then equivalent to the function $h_1(t_1) \cdots h_n(t_n)$ on S^n. Recall that for any ONB $h_1, h_2, \dots$ in H, the products $\bigotimes_{j\le n} h_{r_j}$ with $r_1, \dots, r_n \in \mathsf{N}$ form an ONB in $H^{\otimes n}$.

We begin with the basic characterization and existence result for the multiple integrals ζ^n.

[8] They are usually denoted by I_n. However, we often need a notation exhibiting the dependence on ζ, such as in the higher-dimensional versions of Theorem 28.19

Theorem 14.21 *(multiple stochastic integrals, Wiener, Itô) Let ζ be an isonormal Gaussian process on a separable Hilbert space H. Then for every $n \in \mathsf{N}$ there exists a unique, continuous, linear map $\zeta^n \colon H^{\otimes n} \to L^2(P)$, such that a.s.*

$$\zeta^n \bigotimes_{k \le n} h_k = \prod_{k \le n} \zeta h_k, \quad h_1, \dots, h_n \in H \ \ \textit{orthogonal.}$$

Here the *uniqueness* means that $\zeta^n h$ is a.s. unique for every h, and the *linearity* that

$$\zeta^n(af + bg) = a\,\zeta^n f + b\,\zeta^n g \ \text{ a.s.}, \quad a, b \in \mathsf{R}, \ f, g \in H^{\otimes n}.$$

In particular we have $\zeta^1 h = \zeta h$ a.s., and for consistency we define ζ^0 as the identity mapping on R.

For the proof, we may take $H = L^2(I, \lambda)$ with $I = [0,1]$. Let $\mathcal{E}_n$ be the class of *elementary* functions of the form

$$f = \sum_{j \le m} c_j \bigotimes_{k \le n} 1_{A_j^k}, \tag{9}$$

where the sets $A_j^1, \dots, A_j^n \in \mathcal{B}_I$ are disjoint for each $j \in \{1, \dots, m\}$. The indicator functions $1_{A_j^k}$ are then orthogonal for fixed j, and we need to take

$$\zeta^n f = \sum_{j \le m} c_j \prod_{k \le n} \zeta A_j^k, \tag{10}$$

where $\zeta A = \zeta 1_A$. By the linearity in each factor, $\zeta^n f$ is a.s. independent of the choice of representation (9) for f.

To extend ζ^n to the entire space $L^2(I^n, \lambda^{\otimes n})$, we need some further lemmas. For any function f on I^n, we introduce the *symmetrization*

$$\tilde{f}(t_1, \dots, t_n) = (n!)^{-1} \sum_{p \in \mathcal{P}_n} f(t_{p_1}, \dots, t_{p_n}), \quad t_1, \dots, t_n \in \mathsf{R}_+,$$

where $\mathcal{P}_n$ is the set of permutations of $\{1, \dots, n\}$. We begin with the basic L^2-structure, which will later carry over to general f.

Lemma 14.22 *(isometry) The elementary integrals $\zeta^n f$ in* (10) *are orthogonal for different n and satisfy*

$$E(\zeta^n f)^2 = n!\, \|\tilde{f}\|^2 \le n!\, \|f\|^2, \quad f \in \mathcal{E}_n. \tag{11}$$

Proof: The second relation in (11) follows from Minkowski's inequality. To prove the remaining assertions, we first reduce to the case where all sets A_j^k belong to a fixed collection of disjoint sets $B_1, B_2, \dots$. For any finite index sets $J \ne K$ in N, we get

$$E \prod_{j \in J} \zeta B_j \prod_{k \in K} \zeta B_k = \prod_{j \in J \cap K} E(\zeta B_j)^2 \prod_{j \in J \triangle K} E\zeta B_j = 0,$$

which proves the asserted orthogonality. Since clearly $\langle f, g \rangle = 0$ when f and g involve different index sets, it also reduces the proof of the isometry in (11) to the case where all terms in f involve the same sets $B_1, \dots, B_n$, though in possibly different order. Since $\zeta^n f = \zeta^n \tilde{f}$, we may further assume that $f = \bigotimes_k 1_{B_k}$. Then

$$\begin{aligned} E(\zeta^n f)^2 &= \prod_k E(\zeta B_k)^2 \\ &= \prod_k \lambda B_k \\ &= \|f\|^2 = n!\, \|\tilde{f}\|^2, \end{aligned}$$

where the last relation holds since, in the present case, the permutations of f are orthogonal. □

To extend the integral, we need to show that the set of elementary functions is dense in $L^2(\lambda^{\otimes n})$.

Lemma 14.23 *(approximation) The set $\mathcal{E}_n$ is dense in $L^2(\lambda^{\otimes n})$.*

Proof: By a standard argument based on monotone convergence and a monotone-class argument, any function $f \in L^2(\lambda^{\otimes n})$ can be approximated by linear combinations of products $\bigotimes_{k \le n} 1_{A_k}$, and so it suffices to approximate functions f of the latter type. For each m, we then divide I into 2^m intervals I_{mj} of length 2^{-m}, and define

$$f_m = f \circ \sum_{j_1,\dots,j_n} \bigotimes_{k \le n} 1_{I_{m,j_k}}, \tag{12}$$

where the summation extends over all collections of *distinct* indices $j_1, \dots, j_n \in \{1, \dots, 2^m\}$. Here $f_m \in \mathcal{E}_n$ for each m, and the sum in (12) tends to 1 a.e. $\lambda^{\otimes n}$. Thus, by dominated convergence, $f_m \to f$ in $L^2(\lambda^{\otimes n})$. □

So far we have defined the integral ζ^n as a uniformly continuous mapping on a dense subset of $L^2(\lambda^{\otimes n})$. By continuity it extends immediately to all of $L^2(\lambda^{\otimes n})$, with preservation of both linearity and the norm relations in (11). To complete the proof of Theorem 14.21, it remains to show that $\zeta^n \bigotimes_{k \le n} h_k = \prod_k \zeta h_k$ for orthogonal $h_1, \dots, h_n \in L^2(\lambda)$. This follows from the following technical result, where for any $f \in L^2(\lambda^{\otimes n})$ and $g \in L^2(\lambda)$ we write

$$(f \otimes_1 g)(t_1, \dots, t_{n-1}) = \int f(t_1, \dots, t_n)\, g(t_n)\, dt_n.$$

Lemma 14.24 *(recursion) For any $f \in L^2(\lambda^{\otimes n})$ and $g \in L^2(\lambda)$ with $n \in \mathsf{N}$, we have*

$$\zeta^{n+1}(f \otimes g) = \zeta^n f \cdot \zeta g - n\, \zeta^{n-1}\Big(\tilde{f} \otimes_1 g\Big). \tag{13}$$

Proof: By Fubini's theorem and Cauchy's inequality, we have

$$\begin{aligned} \|f \otimes g\| &= \|f\|\, \|g\|, \\ \|\tilde{f} \otimes_1 g\| &\le \|\tilde{f}\|\, \|g\| \le \|f\|\, \|g\|, \end{aligned}$$

and so both sides of (13) are continuous in probability in f and g. Hence, it suffices to prove the formula for $f \in \mathcal{E}_n$ and $g \in \mathcal{E}_1$. By the linearity of each side, we may next reduce to the case of functions $f = \bigotimes_{k \le n} 1_{A_k}$ and $g = 1_A$, where $A_1, \ldots, A_n$ are disjoint and either $A \cap \bigcup_k A_k = \emptyset$ or $A = A_1$. In the former case, we have $\tilde{f} \otimes_1 g = 0$, and (13) follows from the definitions. In the latter case, (13) becomes

$$\zeta^{n+1}\Big(A^2 \times A_2 \times \cdots \times A_n\Big) = \Big\{(\zeta A)^2 - \lambda A\Big\}\, \zeta A_2 \cdots \zeta A_n. \tag{14}$$

Approximating 1_{A^2} as in Lemma 14.23 by functions $f_m \in \mathcal{E}_2$ with support in A^2, we see that the left-hand side equals $(\zeta^2 A^2)(\zeta A_2)\cdots(\zeta A_n)$, which reduces the proof of (14) to the two-dimensional version $\zeta^2 A^2 = (\zeta A)^2 - \lambda A$. Here we may divide A for each m into 2^m subsets B_{mj} of measure $\leq 2^{-m}$, and note as in Theorem 14.9 and Lemma 14.23 that

$$\begin{aligned}(\zeta A)^2 &= \textstyle\sum_i (\zeta B_{mi})^2 + \sum\limits_{i \neq j} (\zeta B_{mi})(\zeta B_{mj}) \\ &\to \lambda A + \zeta^2 A^2 \quad \text{ in } L^2.\end{aligned}$$

□

We proceed to derive an explicit representation of the integrals ζ^n in terms of *Hermite polynomials* $p_0, p_1, \ldots$, defined as orthogonal polynomials of degrees $0, 1, \ldots$ with respect to the standard normal distribution $N(0,1)$. This determines the p_n up to normalizations, which we choose for convenience to give leading coefficients 1. The first few polynomials are then

$$p_0(x) = 1, \quad p_1(x) = x, \quad p_2(x) = x^2 - 1, \quad p_3(x) = x^3 - 3x, \quad \ldots .$$

Theorem 14.25 *(polynomial representation, Itô) Let ζ be an isonormal Gaussian process on a separable Hilbert space H with associated multiple Wiener–Itô integrals $\zeta^1, \zeta^2, \ldots$. Then for any ortho-normal elements $h_1, \ldots, h_m \in H$ and integers $n_1, \ldots, n_m \geq 1$ with sum n, we have*

$$\zeta^n \bigotimes_{j \leq m} h_j^{\otimes n_j} = \prod_{j \leq m} p_{n_j}(\zeta h_j).$$

Using the linearity of ζ^n and writing $\hat{h} = h/\|h\|$, we see that the stated formula is equivalent to the factorization

$$\zeta^n \bigotimes_{j \leq m} h_j^{\otimes n_j} = \prod_{j \leq m} \zeta^{n_j} h_j^{\otimes n_j}, \quad h_1, \ldots, h_k \in H \text{ orthogonal}, \tag{15}$$

together with the representation of the individual factors

$$\zeta^n h^{\otimes n} = \|h\|^n p_n(\zeta \hat{h}), \quad h \in H \setminus \{0\}. \tag{16}$$

Proof: We prove (15) by induction on n. Assuming equality for all integrals of order $\leq n$, we fix any ortho-normal elements $h, h_1, \ldots, h_m \in H$ and integers $k, n_1, \ldots, n_m \in \mathbb{N}$ with sum $n+1$, and write $f = \bigotimes_{j \leq m} h_j^{\otimes n_j}$. By Lemma 14.24 and the induction hypothesis, we get

$$\begin{aligned}\zeta^{n+1}\Big(f \otimes h^{\otimes k}\Big) &= \zeta^n\Big(f \otimes h^{\otimes (k-1)}\Big) \cdot \zeta h - (k-1)\, \zeta^{n-1}\Big(f \otimes h^{\otimes (k-2)}\Big) \\ &= (\zeta^{n-k+1} f)\Big\{\zeta^{k-1} h^{\otimes (k-1)} \cdot \zeta h - (k-1)\, \zeta^{k-2} h^{\otimes (k-2)}\Big\} \\ &= (\zeta^{n-k+1} f)(\zeta^k h^{\otimes k}).\end{aligned}$$

Using the induction hypothesis again, we obtain the desired extension to ζ^{n+1}.

It remains to prove (16) for an arbitrary element $h \in H$ with $\|h\| = 1$. Then conclude from Lemma 14.24 that

$$\zeta^{n+1}h^{\otimes(n+1)} = (\zeta^n h^{\otimes n})(\zeta h) - n\,\zeta^{n-1}h^{\otimes(n-1)}, \quad n \in \mathsf{N}.$$

Since $\zeta^0 1 = 1$ and $\zeta^1 h = \zeta h$, we see by induction that $\zeta^n h^{\otimes n}$ is a polynomial in ζh of degree n with leading coefficient 1. By the definition of Hermite polynomials, it remains to show that the integrals $\zeta^n h^{\otimes n}$ with different n are orthogonal, which holds by Lemma 14.22. □

Given an isonormal Gaussian process ζ on a separable Hilbert space H, we introduce the space $L^2(\zeta) = L^2(\Omega, \sigma\{\zeta\}, P)$ of ζ-measurable random variables ξ with $E\xi^2 < \infty$. The n-th *polynomial chaos* $\mathcal{P}_n$ is defined as the closed linear sub-space generated by all polynomials of degree $\leq n$ in the random variables ζh, $h \in H$. For every $n \in \mathsf{Z}_+$, we also introduce the n-th *homogeneous chaos* $\mathcal{H}_n$, consisting of all integrals $\zeta^n f$, $f \in H^{\otimes n}$.

The relationship between the mentioned spaces is clarified by the following result, where we write $\oplus$ for direct sum and $\ominus$ for orthogonal complement.

Theorem 14.26 *(chaos decomposition, Wiener) Let ζ be an isonormal Gaussian process on a separable Hilbert space H, with associated polynomial and homogeneous chaoses $\mathcal{P}_n$ and $\mathcal{H}_n$, respectively. Then*

(i) *the $\mathcal{H}_n$ are orthogonal, closed, linear sub-spaces of $L^2(\zeta)$, satisfying*

$$\mathcal{P}_n = \bigoplus_{k=0}^{n} \mathcal{H}_k, \quad n \in \mathsf{Z}_+; \qquad L^2(\zeta) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n,$$

(ii) *for any $\xi \in L^2(\zeta)$, there exist some unique symmetric elements $f_n \in H^{\otimes n}$, $n \geq 0$, such that*

$$\xi = \sum_{n\geq 0} \zeta^n f_n \ \ a.s.$$

In particular, we note that $\mathcal{H}_0 = \mathcal{P}_0 = \mathsf{R}$ and

$$\mathcal{H}_n = \mathcal{P}_n \ominus \mathcal{P}_{n-1}, \quad n \in \mathsf{N}.$$

Proof: (i) The properties in Lemma 14.22 extend to arbitrary integrands, and so the spaces $\mathcal{H}_n$ are mutually orthogonal, closed, linear sub-spaces of $L^2(\zeta)$. From Lemma 14.23 or Theorem 14.25 we see that also $\mathcal{H}_n \subset \mathcal{P}_n$. Conversely, let ξ be a polynomial of degree n in the variables ζh. We may then choose some ortho-normal elements $h_1, \ldots, h_m \in H$, such that ξ is a polynomial in $\zeta h_1, \ldots, \zeta h_m$ of degree n. Since any power $(\zeta h_j)^k$ is a linear combination of the variables $p_0(\zeta h_j), \ldots, p_k(\zeta h_j)$, Theorem 14.25 shows that ξ is a linear combination of multiple integrals $\zeta^k f$ with $k \leq n$, which means that $\xi \in \bigoplus_{k\leq n}\mathcal{H}_k$. This proves the first relation.

To prove the second relation, let $\xi \in L^2(\zeta) \ominus \bigoplus_n \mathcal{H}_n$. In particular, $\xi \perp (\zeta h)^n$ for every $h \in H$ and $n \in \mathsf{Z}_+$. Since $\sum_n |\zeta h|^n/n! = e^{|\zeta h|} \in L^2$, the series $e^{i\zeta h} = \sum_n (i\zeta h)^n/n!$ converges in L^2, and we get $\xi \perp e^{i\zeta h}$ for every $h \in H$. By linearity of the integral ζh, we hence obtain for any $h_1, \ldots, h_n \in H$, $n \in \mathsf{N}$,

$$E\Big(\xi \,\exp \sum_{k\le n} iu_k \zeta h_k\Big) = 0, \quad u_1,\dots,u_n \in \mathsf{R}.$$

Applying Theorem 6.3 to the distributions of $(\zeta h_1,\dots,\zeta h_n)$ under the bounded measures $\mu^\pm = E(\xi^\pm;\cdot)$, we conclude that

$$E\Big\{\xi;\, (\zeta h_1,\dots,\zeta h_n) \in B\Big\} = 0, \quad B \in \mathcal{B}^n,$$

which extends by a monotone-class argument to $E(\xi; A) = 0$ for arbitrary $A \in \sigma\{\zeta\}$. Since ξ is ζ-measurable, it follows that $\xi = E(\xi|\zeta) = 0$ a.s.

(ii) By (i) any element $\xi \in L^2(\zeta)$ has an orthogonal expansion

$$\xi = \sum_{n\ge 0} \zeta^n f_n = \sum_{n\ge 0} \zeta^n \tilde{f}_n,$$

for some elements $f_n \in H^{\otimes n}$ with symmetrizations $\tilde{f}_n$, $n \in \mathsf{Z}_+$. Now assume that also $\xi = \sum_n \zeta^n g_n$. Projecting onto $\mathcal{H}_n$ and using the linearity of ζ^n, we get $\zeta^n(g_n - f_n) = 0$. By the isometry in (11) it follows that $\|\tilde{g}_n - \tilde{f}_n\| = 0$, and so $\tilde{g}_n = \tilde{f}_n$. □

We may often write the decomposition in (ii) as $\xi = \sum_n J_n\xi$, where $J_n\xi = \zeta^n f_n$. We proceed to derive a Wiener chaos expansion in $L^2(\zeta, H)$, stated in terms of the WI-integrals $\zeta^n f_n(\cdot, t)$ for suitable functions f_n on T^{n+1}.

Corollary 14.27 *(chaos expansion of processes) Let ζ be an isonormal Gaussian process on $H = L^2(T,\mu)$, and let $X \in L^2(\zeta, H)$. Then there exist some functions $f_n \in L^2(\mu^{n+1})$, $n \ge 0$, each symmetric in the first n arguments, such that*

(i) $X_t = \sum_n \zeta^n f_n(\cdot, t)$ *in* $L^2(\zeta, H)$,

(ii) $\|X\|^2 = \sum_n n!\,\|f_n\|^2$.

Proof: We may regard X as an element in the space $L^2(\zeta) \otimes H$, admitting an ONB of the form $(\alpha_i \otimes h_j)$, for suitable ONB's $\alpha_1, \alpha_2, \ldots \in L^2(\zeta)$ and $h_1, h_2, \ldots \in H$. This yields an orthogonal expansion $X = \sum_j \xi_j \otimes h_j$ in terms of some elements $\xi_j \in L^2(\zeta)$. Applying the classical Wiener chaos expansion to each ξ_j, we get an orthogonal decomposition

$$\begin{aligned} X &= \textstyle\sum_{n,j} J_n \xi_j \otimes h_j \\ &= \textstyle\sum_{n,j} \zeta^n g_{n,j} \otimes h_j, \end{aligned} \tag{17}$$

for some symmetric functions $g_{n,j} \in L^2(\mu^n)$ satisfying

$$\begin{aligned} \|X\|^2 &= \textstyle\sum_{n,j} \|\zeta^n g_{n,j}\|^2 \\ &= \textstyle\sum_{n,j} n!\,\|g_{n,j}\|^2. \end{aligned} \tag{18}$$

Regarding the h_j as functions in $L^2(\mu)$, we may now introduce the functions

$$f_n(s,t) = \textstyle\sum_j g_{n,j}(s)\, h_j(t) \ \text{ in } L^2(\mu^{n+1}), \quad s \in T^n,\ t \in T,\ n \ge 0,$$

which may clearly be chosen to be product measurable on T^{n+1} and symmetric in $s \in T^n$. The relations (i) and (ii) are obvious from (17) and (18). □

Next we will see how the chaos expansion is affected by conditioning.

Corollary 14.28 *(conditioning) Let $\xi \in L^2(\zeta)$ with chaos expansion $\xi = \sum_n \zeta^n f_n$. Then*

$$E(\xi \mid 1_A\zeta) = \sum_{n\geq 0} \zeta^n(1_A^{\otimes n} f_n) \ \text{a.s.}, \quad A \in \mathcal{T}.$$

Proof: By Lemma 14.23 we may choose $\xi = \zeta^n f$ with $f \in \mathcal{E}_n$, and by linearity we may take $f = \bigotimes_j 1_{B_j}$ for some disjoint sets $B_1, \ldots, B_n \in \mathcal{T}$. Writing

$$\begin{aligned}
\xi &= \prod_{j\leq n} \zeta B_j \\
&= \prod_{j\leq n} \left\{\zeta(B_j \cap A) + \zeta(B_j \cap A^c)\right\} \\
&= \sum_{J\in 2^n} \prod_{j\in J} \zeta(B_j \cap A) \prod_{k\in J^c} \zeta(B_k \cap A^c),
\end{aligned}$$

and noting that the conditional expectation of the last product vanishes by independence unless $J^c = \emptyset$, we get

$$\begin{aligned}
E(\xi \mid 1_A\zeta) &= \prod_{j\leq n} \zeta(B_j \cap A) \\
&= \zeta^n \bigotimes_{j\leq n} 1_{B_j\cap A} \\
&= \zeta^n(1_A^{\otimes n} f).
\end{aligned}$$

□

We also need a related property for processes adapted to a Brownian motion.

Corollary 14.29 *(adapted processes) Let B be a Brownian motion generated by an isonormal Gaussian process ζ on $L^2(\mathsf{R}_+, \lambda)$. Then a process $X \in L^2(\zeta, \lambda)$ is B-adapted iff it can be represented as in Corollary 14.27, with $f_n(\cdot, t)$ supported by $[0,t]^n$ for every $t \geq 0$ and $n \geq 0$.*

Proof: The sufficiency is obvious. Now let X be adapted. We may then apply Corollary 14.28 for every fixed $t \geq 0$. Alternatively, we may apply Corollary 14.27 with ζ replaced by $1_{[0,t]}\zeta$, and use the a.e. uniqueness of the representation. In either case, it remains to check that the resulting functions $1_{[0,t]}^{\otimes n} f_n(\cdot, t)$ remain product measurable. □

Exercises

1. Let $\xi_1, \ldots, \xi_n$ be i.i.d. $N(m, \sigma^2)$. Show that the random variables $\bar{\xi} = n^{-1}\sum_k \xi_k$ and $s^2 = (n-1)^{-1}\sum_k (\xi_k - \bar{\xi})^2$ are independent, and that $(n-1)s^2 \stackrel{d}{=} \sum_{k<n}(\xi_k - m)^2$. (*Hint:* To avoid calculations, use the symmetry in Proposition 14.2.)

2. For a Brownian motion B, put $t_{nk} = k\,2^{-n}$, and define $\xi_{0,k} = B_k - B_{k-1}$ and $\xi_{nk} = B_{t_{n,2k-1}} - \frac{1}{2}(B_{t_{n-1,k-1}} + B_{t_{n-1,k}})$, $k, n \geq 1$. Show that the ξ_{nk} are independent Gaussian. Use this fact to construct a Brownian motion from a sequence of i.i.d. $N(0,1)$ random variables.

3. Let B be a Brownian motion on $[0,1]$, and define $X_t = B_t - tB_1$. Show that $X \perp\!\!\!\perp B_1$. Use this fact to express the conditional distribution of B, given B_1, in terms of a Brownian bridge.

4. Combine the transformations in Lemma 14.6 with the Brownian scaling $B'_t = c^{-1}B(c^2t)$ to construct a class of transformations preserving the distribution of a Brownian bridge.

5. Show that the Brownian bridge is an inhomogeneous Markov process. (*Hint:* Use the transformations in Lemma 14.6 or verify the condition in Proposition 14.7.)

6. Let $B = (B^1, B^2)$ be a Brownian motion in R^2, and choose the t_{nk} as in Theorem 14.9. Show that $\sum_k (B^1_{t_{n,k}} - B^1_{t_{n,k-1}})(B^2_{t_{n,k}} - B^2_{t_{n,k-1}}) \to 0$ in L^2 or a.s., respectively. (*Hint:* Reduce to the case of quadratic variation.)

7. Use Theorem 9.28 to construct an rcll version B of Brownian motion. Then show as in Theorem 14.9 that B has quadratic variation $[B]_t \equiv t$, and conclude that B is a.s. continuous.

8. For a Brownian motion B, show that $\inf\{t > 0;\, B_t > 0\} = 0$ a.s. (*Hint:* By Kolmogorov's 0−1 law, the stated event has probability 0 or 1. Alternatively, use Theorem 14.18.)

9. For a Brownian motion B, define $\tau_a = \inf\{t > 0;\, B_t = a\}$. Compute the density of the distribution of τ_a for $a \neq 0$, and show that $E\tau_a = \infty$. (*Hint:* Use Proposition 14.13.)

10. For a Brownian motion B, show that $Z_t = \exp(cB_t - \frac{1}{2}c^2t)$ is a martingale for every c. Use optional sampling to compute the Laplace transform of τ_a above, and compare with the preceding result.

13. Show that the local maxima of a Brownian motion are a.s. dense in R, and the corresponding times are a.s. dense in R_+. (*Hint:* Use the preceding result.)

14. Show by a direct argument that $\limsup_t t^{-1/2}B_t = \infty$ a.s. as $t \to 0$ and ∞, where B is a Brownian motion. (*Hint:* Use Kolmogorov's 0−1 law.)

15. Show that the local law of the iterated logarithm for Brownian motion remains valid for the Brownian bridge.

16. For a Brownian motion B in R^d, show that the process $|B|$ satisfies the law of the iterated logarithm at 0 and ∞.

17. Let $\xi_1, \xi_2, \ldots$ be i.i.d. $N(0,1)$. Show that $\limsup_n (2\log n)^{-1/2}\xi_n = 1$ a.s.

18. For a Brownian motion B, show that $M_t = t^{-1}B_t$ is a reverse martingale, and conclude that $t^{-1}B_t \to 0$ a.s. and in L^p, $p > 0$, as $t \to \infty$. (*Hint:* The limit is degenerate by Kolmogorov's 0−1 law.) Deduce the same result from Theorem 25.9.

19. For a Brownian bridge B, show that $M_t = (1-t)^{-1}B_t$ is a martingale on $[0,1)$. Check that M is not L^1-bounded.

20. Let ζ^n be the n-fold Wiener–Itô integral w.r.t. Brownian motion B on R_+. Show that the process $M_t = \zeta^n(1_{[0,t]^n})$ is a martingale. Express M in terms of B, and compute the expression for $n = 1, 2, 3$. (*Hint:* Use Theorem 14.25.)

21. Let $\zeta_1, \ldots, \zeta_n$ be independent, isonormal Gaussian processes on a separable Hilbert space H. Show that there exists a unique continuous linear mapping $\bigotimes_k \zeta_k$ from $H^{\otimes n}$ to $L^2(P)$ such that $\bigotimes_k \zeta_k \bigotimes_k h_k = \prod_k \zeta_k h_k$ a.s. for all $h_1, \ldots, h_n \in H$. Also show that $\bigotimes_k \zeta_k$ is an isometry.

Chapter 15

Poisson and Related Processes

Random measures, Laplace transforms, binomial processes, Poisson and Cox processes, transforms and thinnings, mapping and marking, mixed Poisson and binomial processes, existence and randomization, Cox and thinning uniqueness, one-dimensional uniqueness criteria, independent increments, symmetry, Poisson reduction, random time change, Poisson convergence, positive, symmetric, and centered Poisson integrals, double Poisson integrals, decoupling

Brownian motion and the Poisson processes are universally recognized as the basic building blocks of modern probability. After studying the former process in the previous chapter, we turn to a detailed study of Poisson and related processes. In particular, we construct Poisson processes on bounded sets as mixed binomial processes, and derive a variety of Poisson characterizations in terms of independence, symmetry, and renewal properties. A randomization of the underlying intensity measure leads to the richer class of Cox processes. We also consider related randomizations of general point processes, obtainable through independent mappings of the individual points. In particular, we will see how such transformations preserve the Poisson property.

For most purposes, it is convenient to regard Poisson and other point processes as integer-valued random measures. The relevant parts of this chapter then serve as an introduction to general random measure theory. Here we will see how Cox processes and thinnings can be used to derive some general uniqueness criteria for simple point processes and diffuse random measures. The notions and results of this chapter form a basis for the corresponding convergence theory in Chapters 23 and 30, where typically Poisson and Cox processes arise as distributional limits.

We also stress the present significance of the compensators and predictable processes from Chapter 10. In particular, we may use a predictable transformation to reduce a ql-continuous simple point process on R_+ to Poisson, in a similar way as a continuous local martingale can be time-changed into a Brownian motion. This gives a basic connection to the stochastic calculus of Chapters 18–19. Using discounted compensators, we can even go beyond the ql-continuous case.

We finally recall the fundamental role of compound Poisson processes for the theory of Lévy processes and their distributions in Chapters 7 and 16, and for the excursion theory in Chapter 29. In such connections, we often encounter Poisson integrals of various kind, whose existence is studied toward the end of this chapter. We also characterize the existence of double Poisson integrals of

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_16

the form $\xi^2 f$ or $\xi\eta f$, which provides a connection to the multiple Wiener–Itô integrals in Chapters 14, 19, and 21.

Given a localized Borel space S, we define a *random measure* on S as a locally finite kernel from the basic probability space $(\Omega, \mathcal{A}, P)$ into S. Thus, $\xi(\omega, B)$ is a locally finite measure in $B \in \mathcal{S}$ for fixed ω and a random variable in $\omega \in \Omega$ for fixed B. Equivalently, we may regard ξ as a random element in the space $\mathcal{M}_S$ of locally finite measures on S, endowed with the σ-field generated by all evaluation maps $\pi_B : \mu \mapsto \mu B$ with $B \in \mathcal{S}$. The *intensity* $E\xi$ of ξ is the measure on S defined by $(E\xi)B = E(\xi B)$ on $\mathcal{S}$.

A *point process* on S is defined as an integer-valued random measure ξ. Alternatively, it may be regarded as a random element in the space $\mathcal{N}_S \subset \mathcal{M}_S$ of all locally finite, integer-valued measures on S. By Theorem 2.18, it may be written as $\xi = \sum_{k \le \kappa} \delta_{\sigma_k}$ for some random elements $\sigma_1, \sigma_2, \ldots$ in S and κ in $\bar{\mathsf{Z}}_+$, and we say that ξ is *simple* if the σ_k are all distinct. For a general ξ, we may eliminate the possible multiplicities to form a simple point process ξ^*, equal to the counting measure on the support of ξ. Its measurability as a function of ξ is ensured by Theorem 2.19.

We begin with some basic uniqueness criteria for random measures. Stronger results for simple point processes and diffuse random measures are given in Theorem 15.8, and related convergence criteria appear in Theorem 23.16. Recall that $\hat{\mathcal{S}}_+$ denotes the class of measurable functions $f \ge 0$ with bounded support. For countable classes $\mathcal{I} \subset \mathcal{S}$, we write $\hat{\mathcal{I}}_+$ for the family of simple, $\mathcal{I}$-measurable functions $f \ge 0$ on S. Dissecting systems were defined in Chapter 1.

Lemma 15.1 *(uniqueness) Let ξ, η be random measures on S, and fix any dissecting semi-ring $\mathcal{I} \subset \hat{\mathcal{S}}$. Then $\xi \overset{d}{=} \eta$ under each of the conditions*

(i) $\xi f \overset{d}{=} \eta f, \quad f \in \hat{\mathcal{I}}_+,$

(ii) $E e^{-\xi f} = E e^{-\eta f}, \quad f \in \hat{\mathcal{S}}_+.$

When S is Polish, we may take $f \in \hat{C}_S$ or $f \le 1$ in (i)–(ii).

Proof: (i) By the Cramér–Wold Theorem 6.5, the stated condition implies

$$(\xi I_1, \ldots, \xi I_n) \overset{d}{=} (\eta I_1, \ldots, \eta I_n), \quad I_1, \ldots, I_n \in \mathcal{I}, \; n \in \mathsf{N}.$$

By a monotone-class argument in $\mathcal{M}_S$, we get $\xi \overset{d}{=} \eta$ on the σ-field $\Sigma_{\mathcal{I}}$ in $\mathcal{M}_S$ generated by all projections $\pi_I : \mu \to \mu I$, $I \in \mathcal{I}$. By a monotone-class argument in S, π_B remains $\Sigma_{\mathcal{I}}$-measurable for every $B \in \hat{\mathcal{S}}$. Hence, $\Sigma_{\mathcal{I}}$ contains the σ-field generated by the latter projections, and so it agrees with the basic σ-field on $\mathcal{M}_S$, which shows that $\xi \overset{d}{=} \eta$. The statement for Polish spaces follows by a simple approximation.

(ii) Use the uniqueness Theorem 6.3 for Laplace transforms. □

A random measure ξ on a Borel space S is said to have *independent increments*, if the random variables $\xi B_1, \ldots, \xi B_n$ are independent for disjoint sets

$B_1, \ldots, B_n \in \hat{\mathcal{S}}$. By a *Poisson process* on S with intensity measure $\mu \in \mathcal{M}_S$ we mean a point process ξ on S with independent increments, such that ξB is Poisson distributed with mean μB for every $B \in \hat{\mathcal{S}}$. By Lemma 15.1 the stated conditions define a distribution for ξ, determined by the intensity measure μ. More generally, for any random measure η on S, we say that a point process ξ is a *Cox process*[1] *directed by* η, if it is conditionally Poisson given η with $E(\xi|\,\eta) = \eta$ a.s. Choosing $\eta = \alpha\,\mu$ for some measure $\mu \in \mathcal{M}_S$ and random variable $\alpha \geq 0$, we get a *mixed Poisson process* based on μ and α.

Given a probability kernel $\nu\colon S \to T$ and a point measure $\mu = \sum_k \delta_{s_k}$, we define a ν*-transform* of μ as a point process $\eta = \sum_k \delta_{\tau_k}$ on T, where the τ_k are independent random elements in T with distributions ν_{s_k}, assumed to be such that η becomes a.s. locally finite. The distribution of η is clearly a kernel P_μ on the appropriate subset of $\mathcal{N}_T$. Replacing μ by a general point process ξ on S, we may define a ν-transform η of ξ by $\mathcal{L}(\eta\,|\,\xi) = P_\xi$, as long as this makes η a.s. locally finite.

By a ρ*-randomization* of ξ we mean a transform with respect to the kernel $\nu_s = \delta_s \otimes \rho_s$ from S to $S \times T$, where ρ is a probability kernel from S to T. In particular, we may form a *uniform randomization* by choosing ρ to be Lebesgue measure λ on $[0,1]$. We may also think of a randomization as a ρ*-marking* of ξ, where independent random marks of distributions ρ_s are attached to the points of ξ with locations s. We further consider p*-thinnings* of ξ, obtained by independently deleting the points of ξ with probability $1-p$.

A *binomial process*[2] on S is defined as a point process of the form $\xi = \sum_{i \leq n} \delta_{\sigma_i}$, where $\sigma_1, \ldots, \sigma_n$ are independent random elements in S with a common distribution μ. The name is suggested by the fact that ξB is binomially distributed for every $B \in \mathcal{S}$. Replacing n by a random variable $\kappa \perp\!\!\!\perp (\sigma_i)$ in Z_+ yields a *mixed binomial process*.

To examine the relationship between the various processes, it is often helpful to use the *Laplace transforms* of random measures ξ on S, defined for any measurable function $f \geq 0$ on the state space S by $\psi_\xi(f) = Ee^{-\xi f}$. Note that ψ_ξ determines the distribution $\mathcal{L}(\xi)$ by Lemma 15.1. Here we list some basic formulas, where we recall that a kernel ν between measurable spaces S and T may be regarded as an operator between the associated function spaces, given by $\nu f(s) = \int \nu(s, dt) f(t)$.

Lemma 15.2 *(Laplace transforms) Let $f \in \hat{\mathcal{S}}_+$ with S Borel. Then a.s.*

(i) *for a mixed binomial process ξ based on μ and κ,*

$$E(e^{-\xi f}|\,\kappa) = (\mu e^{-f})^\kappa,$$

(ii) *for a Cox process ξ directed by η,*

$$E(e^{-\xi f}|\,\eta) = \exp\Big\{-\eta(1 - e^{-f})\Big\},$$

[1] also called a *doubly stochastic Poisson process*

[2] Binomial processes have also been called *sample* or *Bernoulli* processes, or, when $S \subset \mathsf{R}$, processes with the *order statistics property*.

(iii) *for a ν-transform ξ of η,*

$$E(e^{-\xi f}\,|\,\eta) = \exp\Big(\eta \log \nu e^{-f}\Big),$$

(iv) *for a p-thinning ξ of η,*

$$E(e^{-\xi f}\,|\,\eta) = \exp\Big\{-\eta \log\Big(1 - p(1-e^{-f})\Big)\Big\}.$$

Throughout, we use the compact notation explained in previous chapters. For example, in a more explicit notation the formula in part (iii) becomes[3]

$$E\Big\{\exp\Big(-\int f(s)\,\xi(ds)\Big)\Big|\,\eta\Big\} = \exp\Big\{\int \log\Big(\int e^{-f(t)}\,\nu_s(dt)\Big)\,\eta(ds)\Big\}.$$

Proof: (i) Let $\xi = \sum_{k\le n}\delta_{\sigma_k}$, where $\sigma_1,\dots,\sigma_n$ are i.i.d. μ and $n \in \mathsf{Z}_+$. Then

$$\begin{aligned} Ee^{-\xi f} &= E\exp\textstyle\sum_k\Big\{-f(\sigma_k)\Big\} \\ &= \textstyle\prod_k Ee^{-f(\sigma_k)} \\ &= (\mu e^{-f})^n. \end{aligned}$$

The general result follows by conditioning on κ.

(ii) For a Poisson random variable κ with mean c, we have

$$\begin{aligned} Es^\kappa &= e^{-c}\sum_{k\ge 0}\frac{c^k}{k!}\,s^k \\ &= e^{-c}e^{cs} = e^{-c(1-s)}, \quad s \in (0,1]. \end{aligned} \tag{1}$$

Hence, if ξ is a Poisson process with intensity μ, we get for any function $f = \sum_{k\le n} c_k 1_{B_k}$ with disjoint $B_1,\dots,B_n \in \hat{\mathcal{S}}$ and constant $c_1,\dots,c_n \ge 0$

$$\begin{aligned} Ee^{-\xi f} &= E\exp\textstyle\sum_k\Big(-c_k\,\xi B_k\Big) \\ &= \textstyle\prod_k E\,(e^{-c_k})^{\xi B_k} \\ &= \textstyle\prod_k \exp\Big\{-\mu B_k(1-e^{-c_k})\Big\} \\ &= \exp\textstyle\sum_k\Big\{-\mu B_k(1-e^{-c_k})\Big\} \\ &= \exp\Big\{-\mu(1-e^{-f})\Big\}, \end{aligned}$$

which extends by monotone and dominated convergence to arbitrary $f \in \mathcal{S}_+$. Here the general result follows by conditioning on η.

(iii) First let $\eta = \sum_k \delta_{s_k}$ be non-random in $\mathcal{N}_S$. Choosing some independent random elements τ_k in T with distributions ν_{s_k}, we get

$$\begin{aligned} Ee^{-\xi f} &= E\exp\textstyle\sum_k\{-f(\tau_k)\} \\ &= \textstyle\prod_k Ee^{-f(\tau_k)} \\ &= \textstyle\prod_k \nu_{s_k}e^{-f} \\ &= \exp\textstyle\sum_k \log\nu_{s_k}e^{-f} \\ &= \exp\Big(\eta\log\nu e^{-f}\Big). \end{aligned}$$

[3]Such transliterations are henceforth left to the reader, who may soon see the advantage of our compact notation.

Again, the general result follows by conditioning on η.

(iv) This is easily deduced from (iii). For a direct proof, let $\eta = \sum_k \delta_{s_k}$ as before, and put $\xi = \sum_k \vartheta_k \delta_{s_k}$ for some independent random variables ϑ_k in $\{0,1\}$ with $E\vartheta_k = p(s_k)$. Then

$$\begin{aligned}
Ee^{-\xi f} &= E \exp \textstyle\sum_k \big\{-\vartheta_k f(s_k)\big\} \\
&= \textstyle\prod_k Ee^{-\vartheta_k f(s_k)} \\
&= \textstyle\prod_k \Big\{1 - p(s_k)\Big(1 - e^{-f(s_k)}\Big)\Big\} \\
&= \exp \textstyle\sum_k \log\Big\{1 - p(s_k)\Big(1 - e^{-f(s_k)}\Big)\Big\} \\
&= \exp\Big\{\eta \log\Big(1 - p\,(1 - e^{-f})\Big)\Big\},
\end{aligned}$$

which extends by conditioning to general η. □

We proceed with some basic closure properties for Cox processes and point process transforms. In particular, we note that the Poisson property is preserved by randomizations.

Theorem 15.3 *(mapping and marking, Bartlett, Doob, Prékopa, OK) Let $\mu : S \to T$ and $\nu : T \to U$ be probability kernels between the Borel spaces S, T, U.*

(i) *For a Cox process ξ directed by a random measure η on S, consider a locally finite μ-transform $\zeta \perp\!\!\!\perp_\xi \eta$ of ξ. Then ζ is a Cox process on U directed by $\eta\mu$.*

(ii) *For a μ-transform ξ of a point process η on S, consider a locally finite ν-transform $\zeta \perp\!\!\!\perp_\xi \eta$ of ξ. Then ζ is a $\mu\nu$-transform of η.*

Recall that the kernel $\mu\nu \colon S \to U$ is given by

$$\begin{aligned}
(\mu\nu)_s f &= \int (\mu\nu)_s(du)\, f(u) \\
&= \int \mu_s(dt) \int \nu_t(du)\, f(u).
\end{aligned}$$

Proof: (i) By Proposition 8.9 and Lemma 15.2 (ii) and (iii), we have

$$\begin{aligned}
E(e^{-\zeta f}|\,\eta) &= E\Big\{E\Big(e^{-\zeta f}\Big|\,\xi, \eta\Big)\Big|\,\eta\Big\} \\
&= E\Big\{E(e^{-\zeta f}|\,\xi)\,\Big|\,\eta\Big\} \\
&= E\Big\{\exp\Big(\xi \log \mu e^{-f}\Big)\Big|\,\eta\Big\} \\
&= \exp\Big\{-\eta(1 - \mu e^{-f})\Big\} \\
&= \exp\Big\{-\eta\mu(1 - e^{-f})\Big\}.
\end{aligned}$$

Now use Lemmas 15.1 (ii) and 15.2 (ii).

(ii) Using Lemma 15.2 (iii) twice, we get

$$\begin{aligned} E(e^{-\zeta f} | \, \eta) &= E\Big\{ E(e^{-\zeta f} | \, \xi) \, \Big| \, \eta \Big\} \\ &= E\Big\{ \exp\Big(\xi \log \nu e^{-f} \Big) \Big| \, \eta \Big\} \\ &= \exp\Big(\eta \log \mu \nu e^{-f} \Big). \end{aligned}$$

Now apply Lemmas 15.1 (ii) and 15.2 (ii). □

We turn to a basic relationship between mixed Poisson and binomial processes. The equivalence (i) ⇔ (ii) is elementary for bounded S, and leads in that case to an easy construction of the general Poisson process. The significance of mixed Poisson and binomial processes is further clarified by Theorem 15.14 below. Write $1_B \mu = 1_B \cdot \mu$ for the restriction of μ to B.

Theorem 15.4 *(mixed Poisson and binomial processes, Wiener, Moyal, OK) For a point process ξ on S, these conditions are equivalent:*

(i) *ξ is a mixed Poisson or binomial process,*

(ii) *$1_B \xi$ is a mixed binomial process for every $B \in \hat{\mathcal{S}}$.*

In that case,

(iii) $\quad 1_B \xi \underset{\xi B}{\perp\!\!\!\perp} 1_{B^c} \xi, \quad B \in \hat{\mathcal{S}}.$

Proof, (i) ⇒ (ii): First let ξ be a mixed Poisson process based on μ and ρ. Since $1_B \xi$ is again mixed Poisson and based on $1_B \mu$ and ρ, we may take $B = S$ and $\|\mu\| = 1$. Now consider a mixed binomial process η based on μ and κ, where κ is conditionally Poisson given ρ with $E(\kappa | \, \rho) = \rho$. Using (1) and Lemma 15.2 (i)−(ii), we get for any $f \in \mathcal{S}_+$

$$\begin{aligned} Ee^{-\eta f} &= E(\mu e^{-f})^\kappa \\ &= EE\Big\{ (\mu e^{-f})^\kappa \Big| \, \rho \Big\} \\ &= E \exp\Big\{ -\rho(1 - \mu e^{-f}) \Big\} \\ &= E \exp\Big\{ -\rho\mu(1 - e^{-f}) \Big\} = Ee^{-\xi f}, \end{aligned}$$

and so $\xi \overset{d}{=} \eta$ by Lemma 15.1.

Next, let ξ be a mixed binomial process based on μ and κ. Fix any $B \in \mathcal{S}$ with $\mu B > 0$, and consider a mixed binomial process η based on $\hat{\mu}_B = 1_B \mu / \mu B$ and β, where β is conditionally binomially distributed given κ, with parameters κ and μB. Letting $f \in \mathcal{S}_+$ be supported by B and using Lemma 15.2 (i) twice, we get

$$\begin{aligned} Ee^{-\eta f} &= E\Big(\hat{\mu}_B e^{-f} \Big)^\beta \\ &= EE\Big\{ \Big(\hat{\mu}_B e^{-f} \Big)^\beta \Big| \, \kappa \Big\} \\ &= E\Big\{ 1 - \mu B \Big(1 - \hat{\mu}_B e^{-f} \Big) \Big\}^\kappa \\ &= E(\mu e^{-f})^\kappa = Ee^{-\xi f}, \end{aligned}$$

and so $1_B\xi \stackrel{d}{=} \eta$ by Lemma 15.1.

(ii) $\Rightarrow$ (i): Let $1_B\xi$ be mixed binomial for every $B \in \hat{\mathcal{S}}$. Fix a localizing sequence $B_n \uparrow S$ in $\hat{\mathcal{S}}$, so that the restrictions $1_{B_n}\xi$ are mixed binomial processes based on some probability measures μ_n. Since the μ_n are unique, we see as above that $\mu_m = 1_{B_m}\mu_n/\mu_n B_m$ whenever $m \le n$, and so there exists a measure $\mu \in \mathcal{M}_S$ with $\mu_n = 1_{B_n}\mu/\mu B_n$ for all $n \in \mathsf{N}$.

When $\|\mu\| < \infty$, we may first normalize to $\|\mu\| = 1$. Since $s_k^{n_k} \to s^n$ as $s_k \to s \in (0,1)$ and $n_k \to n \in \bar{\mathsf{N}}$, we get for any $f \in \hat{\mathcal{S}}_+$ with $\mu f > 0$, by Lemma 15.2 (i) and monotone and dominated convergence,

$$\begin{aligned} Ee^{-\xi f} &\leftarrow Ee^{-\xi_n f} \\ &= E(\mu_n e^{-f})^{\xi B_n} \\ &\to E(\mu e^{-f})^{\xi S}. \end{aligned}$$

Choosing $f = \varepsilon 1_{B_n}$ and letting $\varepsilon \downarrow 0$, we get $\xi S < \infty$ a.s., and so the previous calculation applies to arbitrary $f \in \hat{\mathcal{S}}_+$, which shows that ξ is a mixed binomial process based on μ.

Next let $\|\mu\| = \infty$. If $f \in \hat{\mathcal{S}}_+$ is supported by B_m for a fixed $m \in \mathsf{N}$, then for any $n \ge m$,

$$\begin{aligned} Ee^{-\xi f} &= E(\mu_n e^{-f})^{\xi B_n} \\ &= E\left\{1 - \frac{\mu(1-e^{-f})}{\mu B_n}\right\}^{\mu B_n \alpha_n}, \end{aligned}$$

where $\alpha_n = \xi B_n/\mu B_n$. By Theorem 6.20, we have convergence $\alpha_n \xrightarrow{d} \alpha$ in $[0,\infty]$ along a sub-sequence. Noting that $(1-m_n^{-1})^{m_n x_n} \to e^{-x}$ as $m_n \to \infty$ and $0 \le x_n \to x \in [0,\infty]$, we get by Theorem 5.31

$$Ee^{-\xi f} = E\exp\left\{-\alpha\mu(1-e^{-f})\right\}.$$

As before, we get $\alpha < \infty$ a.s., and so by monotone and dominated convergence, we may extend the displayed formula to arbitrary $f \in \hat{\mathcal{S}}_+$. By Lemmas 15.1 and 15.2 (ii), ξ is then distributed as a mixed Poisson process based on μ and α.

(iii) We may take $\|\mu\| < \infty$, since the case of $\|\mu\| = \infty$ will then follow by a martingale or monotone-class argument. Then consider any mixed binomial process $\xi = \sum_{k\le\kappa}\delta_{\sigma_k}$, where the σ_k are i.i.d. μ and independent of κ. For any $B \in \mathcal{S}$, we note that $1_B\xi$ and $1_{B^c}\xi$ are conditionally independent binomial processes based on $1_B\mu$ and $1_{B^c}\mu$, respectively, given the variables κ and $\vartheta_k = 1\{k \le \kappa,\ \sigma_k \in B\}$, $k \in \mathsf{N}$. Thus, $1_B\xi$ is conditionally a binomial process based on μ_B and ξB, given $1_{B^c}\xi$, ξB, and $\vartheta_1, \vartheta_2, \ldots$. Since the conditional distribution depends only on ξB, we obtain $1_B\xi \perp\!\!\!\perp_{\xi B} (1_{B^c}\xi, \vartheta_1, \vartheta_2, \ldots)$, and the assertion follows. $\square$

Using the easy necessity part of Theorem 15.4, we may prove the existence of Cox processes and transforms. The result also covers the cases of binomial processes and randomizations.

Theorem 15.5 *(Cox and transform existence)*

(i) *For any random measure η on S, there exists a Cox process ξ directed by η.*

(ii) *For any point process ξ on S and probability kernel $\nu : S \to T$, there exists a ν-transform ζ of ξ.*

Proof: (i) First let $\eta = \mu$ be non-random with $\mu S \in (0, \infty)$. By Corollary 8.25 we may choose a Poisson distributed random variable κ with mean μS and some independent i.i.d. random elements $\sigma_1, \sigma_2, \ldots$ in S with distribution $\mu/\mu S$. By Theorem 15.4, the random measure $\xi = \sum_{j \le \kappa} \delta_{\sigma_j}$ is then Poisson with intensity μ.

For general $\mu \in \mathcal{M}_\mu$, we may choose a partition of S into disjoint subsets $B_1, B_2, \ldots \in \hat{\mathcal{S}}$, such that $\mu B_k \in (0, \infty)$ for all k. As before, there exists for every k a Poisson process ξ_k on S with intensity $\mu_k = 1_{B_k}\mu$, and by Corollary 8.25 we may choose the ξ_k to be independent. Writing $\xi = \sum_k \xi_k$ and using Lemma 15.2 (i), we get for any measurable function $f \ge 0$ on S

$$\begin{aligned} Ee^{-\xi f} &= \textstyle\prod_k Ee^{-\xi_k f} \\ &= \textstyle\prod_k \exp\left\{-\mu_k(1-e^{-f})\right\} \\ &= \exp\left\{-\textstyle\sum_k \mu_k(1-e^{-f})\right\} \\ &= \exp\left\{-\mu(1-e^{-f})\right\}. \end{aligned}$$

Using Lemmas 15.1 (ii) and 15.2 (i), we conclude that ξ is a Poisson process with intensity μ.

Now let ξ_μ be a Poisson process with intensity μ. Then for any $m_1, \ldots, m_n \in \mathsf{Z}_+$ and disjoint $B_1, \ldots, B_n \in \hat{\mathcal{S}}$,

$$P\bigcap_{k \le n}\left\{\xi_\mu B_k = m_k\right\} = \prod_{k \le n} e^{-\mu B_k}(\mu B_k)^{m_k}/m_k!,$$

which is a measurable function of μ. The measurability extends to arbitrary sets $B_k \in \hat{\mathcal{S}}$, since the general probability on the left is a finite sum of such products. Now the sets on the left form a π-system generating the σ-field in $\mathcal{N}_S$, and so Theorem 1.1 shows that $P_\mu = \mathcal{L}(\xi_\mu)$ is a probability kernel from $\mathcal{M}_S$ to $\mathcal{N}_S$. Then Lemma 8.16 ensures the existence, for any random measure η on S, of a Cox process ξ directed by η.

(ii) First let $\mu = \sum_k \delta_{s_k}$ be non-random in $\mathcal{N}_S$. Then Corollary 8.25 yields some independent random elements τ_k in T with distributions ν_{s_k}, making $\zeta_\mu = \sum_k \delta_{\tau_k}$ a ν-transform of μ. Letting $B_1, \ldots, B_n \in \mathcal{T}$ and $s_1, \ldots, s_n \in (0,1)$, we get by Lemma 15.2 (iii)

$$\begin{aligned} E\textstyle\prod_k s_k^{\zeta_\mu B_k} &= E\exp \zeta_\mu \textstyle\sum_k 1_{B_k} \log s_k \\ &= \exp \mu \log \nu \exp \textstyle\sum_k 1_{B_k} \log s_k \\ &= \exp \mu \log \nu \textstyle\prod_k (s_k)^{1_{B_k}}. \end{aligned}$$

Using Lemma 3.2 (i) twice, we see that $\nu \prod_k (s_k)^{1_{B_k}}$ is a measurable function on S for fixed $s_1, \ldots, s_n$, and so the right-hand side is a measurable function of μ. Differentiating m_k times with respect to s_k for each k and setting $s_1 = \cdots = s_n = 1$, we conclude that the probability $P \bigcap_k \{\zeta_\mu B_k = m_k\}$ is a measurable function of μ for any $m_1, \ldots, m_n \in \mathsf{Z}_+$. As before, $P_\mu = \mathcal{L}(\zeta_\mu)$ is then a probability kernel from $\mathcal{N}_S$ to $\mathcal{N}_T$, and the general result follows by Lemma 8.16. □

We proceed with some basic properties of Poisson and Cox processes.

Lemma 15.6 *(Cox simplicity and integrability) Let ξ be a Cox process directed by a random measure η on S. Then*

(i) *ξ is a.s. simple $\Leftrightarrow$ η is a.s. diffuse,*

(ii) $\{\xi f < \infty\} = \big\{\eta(f \wedge 1) < \infty\big\}$ *a.s. for all $f \in \mathcal{S}_+$.*

Proof: (i) First reduce by conditioning to the case where ξ is Poisson with intensity μ. Since both properties are local, we may also assume that $\|\mu\| \in (0, \infty)$. By a further conditioning based on Theorem 15.4, we may then take ξ to be a binomial process based on μ, say $\xi = \sum_{k \le n} \delta_{\sigma_k}$, where the σ_k are i.i.d. μ. Then Fubini's theorem yields

$$\begin{aligned} P\{\sigma_i = \sigma_j\} &= \int \mu\{s\}\, \mu(ds) \\ &= \textstyle\sum_s (\mu\{s\})^2, \quad i \ne j, \end{aligned}$$

which shows that the σ_k are a.s. distinct iff μ is diffuse.

(ii) Conditioning on η and using Theorem 8.5, we may again reduce to the case where ξ is Poisson with intensity μ. For any $f \in \mathcal{S}_+$, we get as $0 < r \to 0$

$$\begin{aligned} \exp\Big\{-\mu(1 - e^{-rf})\Big\} &= E e^{-r\xi f} \\ &\to P\{\xi f < \infty\}. \end{aligned}$$

Since $1 - e^{-f} \asymp f \wedge 1$, we have $\mu(1 - e^{-rf}) \equiv \infty$ when $\mu(f \wedge 1) = \infty$, whereas $\mu(1 - e^{-rf}) \to 0$ by dominated convergence when $\mu(f \wedge 1) < \infty$. □

The following uniqueness property will play an important role below.

Lemma 15.7 *(Cox and thinning uniqueness, Krickeberg) Let ξ, $\tilde{\xi}$ be Cox processes on S directed by some random measures η, $\tilde{\eta}$, or p-thinnings of some point processes η, $\tilde{\eta}$, for a fixed $p \in (0, 1]$. Then*

$$\xi \overset{d}{=} \tilde{\xi} \quad \Leftrightarrow \quad \eta \overset{d}{=} \tilde{\eta}.$$

Proof: The implications to the left are obvious. To prove the implication to the right for the Cox transform, we may invert Lemma 15.2 (ii) to get for $f \in \hat{\mathcal{S}}_+$

$$E e^{-\eta f} = E \exp\Big\{\xi \log(1 - f)\Big\}, \quad f < 1.$$

In case of p-thinnings, we may use Lemma 15.2 (iv) instead to get

$$Ee^{-\eta f} = E\exp\Big\{\xi\log\Big(1-p^{-1}(1-e^{-f})\Big)\Big\}, \quad f < -\log(1-p).$$

In both cases, the assertion follows by Lemma 15.1 (ii). □

Using Cox and thinning transforms, we may partially strengthen the general uniqueness criteria of Theorem 15.1. Related convergence criteria are given in Chapter 23 and 30.

Theorem 15.8 *(one-dimensional uniqueness criteria, Mönch, Grandell, OK) Let S be a Borel space with a dissecting ring $\mathcal{U} \subset \hat{\mathcal{S}}$. Then*

(i) *for any point processes ξ, η on S, we have $\xi^* \stackrel{d}{=} \eta^*$ iff*

$$P\{\xi U = 0\} = P\{\eta U = 0\}, \quad U \in \mathcal{U},$$

(ii) *for any simple point processes or diffuse random measures ξ, η on S, we have $\xi \stackrel{d}{=} \eta$ iff for a fixed $c > 0$,*

$$E\,e^{-c\xi U} = E\,e^{-c\eta U}, \quad U \in \mathcal{U},$$

(iii) *for a simple point process or diffuse random measure ξ and an arbitrary random measure η on S, we have $\xi \stackrel{d}{=} \eta$ iff*

$$\xi U \stackrel{d}{=} \eta U, \quad U \in \mathcal{U}.$$

Proof: (i) Let $\mathcal{C}$ be the class of sets $\{\mu;\, \mu U = 0\}$ in $\mathcal{N}_S$ with $U \in \mathcal{U}$, and note that $\mathcal{C}$ is a π-system, since for any $B, C \in \mathcal{U}$,

$$\{\mu B = 0\} \cap \{\mu C = 0\} = \Big\{\mu(B \cup C) = 0\Big\} \in \mathcal{C},$$

by the ring property of $\mathcal{U}$. Furthermore, the sets $M \in \mathcal{B}_{\mathcal{N}_S}$ with $P\{\xi \in M\} = P\{\eta \in M\}$ form a λ-system $\mathcal{D}$, which contains $\mathcal{C}$ by hypothesis. Hence, Theorem 1.1 (i) yields $\mathcal{D} \supset \sigma(\mathcal{C})$, which means that $\xi \stackrel{d}{=} \eta$ on $\sigma(\mathcal{C})$.

Since the ring $\mathcal{U}$ is dissecting, it contains a dissecting system (I_{nj}), and for any $\mu \in \mathcal{N}_S$ we have

$$\textstyle\sum_j \Big\{\mu(U \cap I_{nj}) \wedge 1\Big\} \to \mu^* U, \quad U \in \mathcal{U},$$

since the atoms of μ are ultimately separated by the sets I_{nj}. Since the sum on the left is $\sigma(\mathcal{C})$-measurable, so is the limit $\mu^* U$ for every $U \in \mathcal{U}$. By the proof of Lemma 15.1, the measurability extends to the map $\mu \mapsto \mu^*$, and we conclude that $\xi^* \stackrel{d}{=} \eta^*$.

(ii) First let ξ, η be diffuse. Letting $\tilde{\xi}, \tilde{\eta}$ be Cox processes directed by $c\xi$, $c\eta$, respectively, we get by conditioning and hypothesis

$$\begin{aligned} P\{\tilde{\xi} U = 0\} &= E\,e^{-c\xi U} = E\,e^{-c\eta U} \\ &= P\{\tilde{\eta} U = 0\}, \quad U \in \mathcal{U}. \end{aligned}$$

Since $\tilde{\xi}$, $\tilde{\eta}$ are a.s. simple by Lemma 15.6 (i), part (i) yields $\tilde{\xi} \stackrel{d}{=} \tilde{\eta}$, and so $\xi \stackrel{d}{=} \eta$ by Lemma 15.7 (i).

Next let ξ, η be simple point processes. For $p = 1 - e^{-c}$, let $\tilde{\xi}$, $\tilde{\eta}$ be p-thinnings of ξ, η, respectively. Since Lemma 15.2 (iv) remains true when $0 \le f \le \infty$, we may take $f = \infty \cdot 1_U$ for any $U \in \mathcal{U}$ to get

$$\begin{aligned} P\{\tilde{\xi}U = 0\} &= E\exp\Big\{\xi \log(1 - p\,1_U)\Big\} \\ &= E\exp\Big\{\xi U \log(1-p)\Big\} = Ee^{-c\xi}, \end{aligned}$$

and similarly for $\tilde{\eta}$ in terms of η. Hence, the simple point processes $\tilde{\xi}$, $\tilde{\eta}$ satisfy the condition in (i), and so $\tilde{\xi} \stackrel{d}{=} \tilde{\eta}$, which implies $\xi \stackrel{d}{=} \eta$.

(iii) First let ξ be a simple point process. The stated condition yields $\eta U \in \mathsf{Z}_+$ a.s. for every $U \in \mathcal{U}$, and so η is a point process by Lemma 2.18. By (i) we get $\xi \stackrel{d}{=} \eta^*$, and in particular $\eta U \stackrel{d}{=} \xi U \stackrel{d}{=} \eta^* U$ for all $U \in \mathcal{U}$, which implies $\eta = \eta^*$ since the ring $\mathcal{U}$ is covering.

Next let η be a.s. diffuse. Letting $\tilde{\xi}$, $\tilde{\eta}$ be Cox processes directed by ξ, η, respectively, we get $\tilde{\xi}U \stackrel{d}{=} \tilde{\eta}U$ for any $U \in \mathcal{U}$ by Lemma 15.2 (ii). Since $\tilde{\xi}$ is simple by Lemma 15.6 (i), we conclude as before that $\tilde{\xi} \stackrel{d}{=} \tilde{\eta}$, and so $\xi \stackrel{d}{=} \eta$ by Lemma 15.7 (i). □

The last result yields a nice characterization of Poisson processes:

Corollary 15.9 *(Poisson criterion, Rényi) Let ξ be a random measure on S with $E\xi\{s\} \equiv 0$ a.s., and fix a dissecting ring $\mathcal{U} \subset \hat{\mathcal{S}}$. Then these conditions are equivalent:*

(i) *ξ is a Poisson process,*

(ii) *ξU is Poisson distributed for every $U \in \mathcal{U}$.*

Proof: Assume (ii). Since $\lambda = E\xi \in \mathcal{M}_S$, Theorem 15.5 yields a Poisson process η on S with $E\eta = \lambda$. Here $\xi U \stackrel{d}{=} \eta U$ for all $U \in \mathcal{U}$, and since η is simple by Corollary 15.6 (i), we get $\xi \stackrel{d}{=} \eta$ by Theorem 15.8 (iii), proving (i). □

Much of the previous theory extends to processes with marks. Given some Borel spaces S and T, we define a *T-marked point process on S* as a point process ξ on $S \times T$ with $\xi(\{s\} \times T) \le 1$ for all $s \in S$. Say that ξ has *independent increments*, if the point processes $\xi(B_1 \times \cdot), \dots, \xi(B_n \times \cdot)$ on T are independent for any disjoint sets $B_1, \dots, B_n \in \hat{\mathcal{S}}$. We also say that ξ is a *Poisson process*, if ξ is Poisson in the usual sense on the product space $S \times T$. We may now characterize Poisson processes in terms of the independence property, extending Theorem 13.6. The result plays a crucial role in Chapters 16 and 29. A related characterization in terms of compensators is given in Corollary 15.17.

For future purposes, we consider a slightly more general case. For any S-marked point process ξ on T, we introduce the discontinuity set $D = \{t \in T;$

$E\xi(\{t\} \times S) > 0\}$, and say that ξ is an *extended Poisson process*, if it is an ordinary Poisson process on D^c, and the contributions ξ_t to the points $t \in D$ are independent point measures on S with mass ≤ 1. Note that $\mathcal{L}(\xi)$ is then uniquely determined by $E\xi$.

Theorem 15.10 *(independent-increment point processes, Bateman, Copeland & Regan, Itô, Kingman) For an S-marked point process ξ on T, these conditions are equivalent:*

(i) *ξ has independent T-increments,*

(ii) *ξ is an extended Poisson process on $T \times S$.*

Proof (OK): Subtracting the fixed discontinuities, we may reduce to the case where $E\xi(\{t\} \times S) \equiv 0$. Here we first consider a simple point process ξ on $T = [0,1]$ with $E\xi\{t\} \equiv 0$. Define a set function

$$\rho B = -\log E e^{-\xi B}, \quad B \in \hat{\mathcal{T}},$$

and note that $\rho B \ge 0$ for all B since $\xi B \ge 0$ a.s. If ξ has independent increments, then for disjoint $B, C \in \hat{\mathcal{T}}$,

$$\begin{aligned}\rho(B \cup C) &= -\log E e^{-\xi B - \xi C} \\ &= -\log\Big(E e^{-\xi B} E e^{-\xi C}\Big) \\ &= \rho B + \rho C,\end{aligned}$$

which shows that ρ is finitely additive. It is even countably additive, since $B_n \uparrow B$ in $\hat{\mathcal{T}}$ implies $\rho B_n \uparrow \rho B$ by monotone and dominated convergence. It is further diffuse, since $\rho\{t\} = -\log E e^{-\xi\{t\}} = 0$ for all $t \in T$.

Now Theorem 15.5 yields a Poisson process η on T, such that $E\eta = c^{-1}\rho$ with $c = 1 - e^{-1}$. For any $B \in \hat{\mathcal{S}}$, we get by Lemma 15.2 (ii)

$$\begin{aligned}E e^{-\eta B} &= \exp\Big\{-c^{-1}\rho(1 - e^{-1_B})\Big\} \\ &= e^{-\rho B} = E e^{-\xi B}.\end{aligned}$$

Since ξ, η are simple by hypothesis and Lemma 15.6 (i), we get $\xi \overset{d}{=} \eta$ by Theorem 15.8 (ii).

In the marked case, fix any $B \in \mathcal{B}_{T\times S}$, and define a simple point process on T by $\eta = 1_B\xi(\cdot \times S)$. Then the previous case shows that $\xi B = \|\eta\|$ is Poisson distributed, and so ξ is Poisson by Corollary 15.9. $\square$

The last result extends to a representation of random measures with independent increments[4]. A extension to general processes on R_+ will be given in Theorem 16.3.

[4]also said to be *completely random*

Corollary 15.11 *(independent-increment random measures, Kingman) A random measure ξ on S has independent increments iff a.s.*

$$\xi B = \alpha B + \int_0^\infty x\,\eta(B \times dx), \quad B \in \mathcal{S}, \tag{2}$$

for a fixed measure $\alpha \in \mathcal{M}_S$ and an extended Poisson process η on $S \times (0, \infty)$ with

$$\int_0^\infty (x \wedge 1)\, E\,\eta(B \times dx) < \infty, \quad B \in \hat{\mathcal{S}}. \tag{3}$$

Proof (OK): On $S \times (0, \infty)$ we define a point process $\eta = \sum_s \delta_{s, \xi\{s\}}$, where the required measurability follows by a simple approximation. Since η has independent S-increments and

$$\eta\Big(\{s\} \times (0, \infty)\Big) = 1\Big\{\xi\{s\} > 0\Big\} \le 1, \quad s \in S,$$

Theorem 15.10 shows that η is a Poisson process. Subtracting the atomic part from ξ, we obtain a diffuse random measure α satisfying (2), which has again independent increments and hence is a.s. non-random by Theorem 6.12. Next, Lemma 15.2 (i) yields for any $B \in \mathcal{S}$ and $r > 0$

$$-\log E \exp\Big\{-r \int_0^\infty x\,\eta(B \times dx)\Big\} = \int_0^\infty (1 - e^{-rx})\, E\eta(B \times dx).$$

As $r \to 0$, we see by dominated convergence that $\int_0^\infty x\,\eta(B \times dx) < \infty$ a.s. iff (3) holds. □

We summarize the various Poisson criteria, to highlight some remarkable equivalences. The last condition anticipates Corollary 15.17 below.

Corollary 15.12 *(Poisson criteria) Let ξ be a simple point process on S with $E\,\xi\{s\} = 0$ for all $s \in S$. Then these conditions are equivalent*[5]*:*

(i) *ξ is a Poisson process,*
(ii) *ξ has independent increments,*
(iii) *ξB is Poisson distributed for every $B \in \hat{\mathcal{S}}$.*

When ξ is stationary on $S = \mathsf{R}_+$, it is further equivalent that

(iv) *ξ is a renewal process with exponential holding times,*
(v) *ξ has induced compensator $c\lambda$ for a constant $c \ge 0$.*

Proof: Here (i) ⇔ (ii) by Theorem 15.10 and (i) ⇔ (iii) by Corollary 15.9. For stationary ξ, the equivalence of (i)–(ii), (iv) was proved in Theorem 13.6. □

In the multi-variate case, we have a further striking equivalence:

[5]The equivalence (ii) ⇔ (iii) is remarkable, since *both* conditions are usually required to define a Poisson process.

Corollary 15.13 *(multi-variate Poisson criteria) Let ξ be a point process on $S \times T$ with a simple, locally finite S-projection $\bar{\xi} = \xi(\cdot \times T)$, such that $E\bar{\xi}\{s\} = 0$ for all $s \in S$. Then these conditions are equivalent:*

(i) *ξ is a Poisson process on $S \times T$,*

(ii) *ξ has independent S-increments,*

(iii) *ξ is a ν-randomization of a Poisson process $\bar{\xi}$ on S, for a probability kernel $\nu\colon S \to T$.*

Proof: Since (iii) $\Rightarrow$ (ii) is obvious and (ii) $\Rightarrow$ (i) by Theorem 15.10, it remains to prove that (i) $\Rightarrow$ (iii). Then let ξ be Poisson with intensity $\lambda = E\xi$, and note that $\bar{\xi}$ is again Poisson with intensity $\bar{\lambda} = \lambda(\cdot \times T)$. By Theorem 3.4 there exists a probability kernel $\nu\colon S \to T$ such that $\lambda = \bar{\lambda} \otimes \nu$. Letting η be a ν-randomization of $\bar{\xi}$, we see from Theorem 15.3 (i) that η is again Poisson with intensity λ. Hence, $(\xi, \bar{\xi}) \overset{d}{=} (\eta, \bar{\xi})$, and so $\mathcal{L}(\xi|\,\bar{\xi}) = \mathcal{L}(\eta|\,\bar{\xi})$, which shows that even ξ is a ν-randomization of $\bar{\xi}$. $\square$

We may next characterize the mixed Poisson and binomial processes by a basic symmetry property. Related results for more general processes appear in Theorems 27.9 and 27.10. Given a random measure ξ and a diffuse measure λ on S, we say that ξ is λ-*symmetric* if $\xi \circ f^{-1} \overset{d}{=} \xi$ for every λ-preserving map f on S.

Theorem 15.14 *(symmetry, Davidson, Matthes et al., OK) Let S be a Borel space with a diffuse measure $\lambda \in \mathcal{M}_S$. Then*

(i) *a simple point process ξ on S is λ-symmetric iff it is a mixed Poisson or binomial process based on λ,*

(ii) *a diffuse random measure ξ on S is λ-symmetric iff $\xi = \alpha\lambda$ a.s. for a random variable $\alpha \geq 0$.*

In both cases, this holds[6] *iff $\mathcal{L}(\xi B)$ depends only on λB.*

Proof: (i) If $S = [0,1]$, then $\|\xi\|$ is invariant under λ-preserving transformations, and so ξ remains conditionally λ-symmetric, given $\|\xi\|$. We may then reduce by conditioning to the case of a constant $\|\xi\| = n$, so that $\xi = \sum_{k\leq n} \delta_{\sigma_k}$ for some random variables $\sigma_1 < \cdots < \sigma_n$ in $[0,1]$. Now consider any subintervals $I_1 < \cdots < I_n$ of $[0,1]$, where $I < J$ means that $s < t$ for all $s \in I$ and $t \in J$. By contractability, the probability

$$P \textstyle\bigcap_k \{\sigma_k \in I_k\} = P \bigcap_k \{\xi I_k > 0\}$$

is invariant under individual shifts of $I_1, \ldots, I_n$, subject to the mentioned order restriction. Thus, $\mathcal{L}(\sigma_1, \ldots, \sigma_n)$ is shift invariant on the set $\Delta_n = \{s_1 < \cdots <$

[6]In particular, the notions of exchangeability and contractability are equivalent for simple or diffuse random measures on a finite or infinite interval. Note that the corresponding statements for finite sequences or processes on $[0,1]$ are false. See Chapters 27–28.

$s_n\} \subset [0,1]^n$, and hence equals $n!\,\lambda^n$ on Δ_n, which means that ξ is a binomial process on $[0,1]$ based on λ and n. If instead $S = \mathsf{R}_+$, the restrictions $1_{[0,t]}\xi$ are mixed binomial processes based on $1_{[0,t]}\lambda$, and so ξ is mixed Poisson on R_+ by Theorem 15.4.

(ii) We may take $S = [0,1]$. As in (i), we may further reduce by conditioning to the case where $\|\xi\| = 1$ a.s. Then $E\xi = \lambda$ by contractability, and for any intervals $I_1 < I_2$ in I, the moment $E\xi^2(I_1 \times I_2)$ is invariant under individual shifts of I_1 and I_2, so that $E\xi^2$ is proportional to λ^2 on Δ_2. Since ξ is diffuse, we have $E\xi^2 = 0$ on the diagonal $\{s_1 = s_2\}$ in $[0,1]^2$. Hence, by symmetry and normalization, $E\xi^2 = \lambda^2$ on $[0,1]^2$, and so

$$\begin{aligned} \operatorname{Var}(\xi B) &= E(\xi B)^2 - (E\xi B)^2 \\ &= \lambda^2 B^2 - (\lambda B)^2 = 0, \quad B \in \mathcal{B}_{[0,1]}. \end{aligned}$$

Then $\xi B = \lambda B$ a.s. for all B, and so $\xi = \lambda$ a.s. by a monotone-class argument.

Now suppose that $\mathcal{L}(\xi B)$ depends only on λB, and define

$$Ee^{-r\xi B} = \varphi_r(\lambda B), \quad B \in \mathcal{S},\ r \geq 0.$$

For any λ-preserving map f on S, we get

$$\begin{aligned} Ee^{-r\xi\circ f^{-1}B} &= \varphi_r \circ \lambda(f^{-1}B) \\ &= \varphi_r(\lambda B) \\ &= Ee^{-r\xi B}, \end{aligned}$$

and so $\xi(f^{-1}B) \stackrel{d}{=} \xi B$ by Theorem 6.3. Then Theorem 15.8 (iii) yields $\xi \circ f^{-1} \stackrel{d}{=} \xi$, which shows that ξ is λ-symmetric. $\Box$

A Poisson process ξ on R_+ is said to be *stationary* with *rate* $c \geq 0$ if $E\xi = c\lambda$. Then Proposition 11.5 shows that $N_t = \xi[0,t]$, $t \geq 0$, is a space- and time-homogeneous Markov process. We show how a point process on R_+, subject to a suitable regularity condition, can be transformed to Poisson by a predictable map. This leads to various time-change formulas for point processes, similar to those for continuous local martingales in Chapter 19.

When ξ is an *S-marked point process* on $(0,\infty)$, it is automatically locally integrable, which ensures the existence of the associated compensator $\hat{\xi}$. Recall that ξ is said to be *ql-continuous* if $\xi([\tau] \times S) = 0$ a.s. for every predictable time τ, so that the process $\hat{\xi}_t B$ is a.s. continuous for every $B \in \hat{\mathcal{S}}$.

Theorem 15.15 *(predictable mapping to Poisson) For any Borel spaces S, T and measure $\mu \in \mathcal{M}_T$, let ξ be a ql-continuous, S-marked point process on $(0,\infty)$ with compensator $\hat{\xi}$, and consider a predictable map $Y\colon \Omega \times \mathsf{R}_+ \times S \to T$. Then*

$$\hat{\xi} \circ Y^{-1} = \mu \ \text{ a.s.} \quad \Rightarrow \quad \eta = \xi \circ Y^{-1} \ \text{ is Poisson } (\mu).$$

In particular, any simple, ql-continuous point process ξ on R_+ is unit rate Poisson, on the time scale given by the compensator $\hat{\xi}$. Note the analogy with the time-change reduction of continuous local martingales to Brownian motion in Theorem 19.4.

Proof: For disjoint sets $B_1, \ldots, B_n \in \mathcal{B}_T$ with finite μ-measure, we need to show that $\eta B_1, \ldots, \eta B_n$ are independent Poisson variables with means $\mu B_1, \ldots, \mu B_n$. Then define for each $k \le n$ the processes

$$J_t^k = \int_S \int_0^{t+} 1_{B_k}(Y_{s,x})\, \xi(ds\, dx),$$
$$\hat{J}_t^k = \int_S \int_0^{t} 1_{B_k}(Y_{s,x})\, \hat{\xi}(ds\, dx).$$

Here $\hat{J}_\infty^k = \mu B_k < \infty$ a.s. by hypothesis, and so the J^k are simple, integrable point processes on R_+ with compensators $\hat{J}^k$. For fixed $u_1, \ldots, u_n \ge 0$, we put

$$X_t = \sum_{k \le n} \left\{ u_k J_t^k - (1 - e^{-u_k}) \hat{J}_t^k \right\}, \quad t \ge 0.$$

The process $M_t = e^{-X_t}$ has bounded variation and finitely many jumps, and so by an elementary change of variables,

$$\begin{aligned} M_t - 1 &= \sum_{s \le t} \Delta e^{-X_s} - \int_0^t e^{-X_s} dX_s^c \\ &= \sum_{k \le n} \int_0^{t+} e^{-X_{s-}} (1 - e^{-u_k})\, d\Big(\hat{J}_s^k - J_s^k\Big). \end{aligned}$$

Since the integrands on the right are bounded and predictable, M is a uniformly integrable martingale, and so $EM_\infty = 1$. Thus,

$$E \exp\Big(-\sum\nolimits_k u_k\, \eta B_k\Big) = \exp\Big\{-\sum\nolimits_k (1 - e^{-u_k})\, \mu B_k\Big\},$$

and the assertion follows by Theorem 6.3. □

We also have a multi-variate time-change result, similar to Proposition 19.8 for continuous local martingales. Say that the simple point processes $\xi_1, \ldots, \xi_n$ are *orthogonal* if $\sum_k \xi_k\{s\} \le 1$ for all s.

Corollary 15.16 *(time-change reduction to Poisson, Papangelou, Meyer) Let $\xi_1, \ldots, \xi_n$ be orthogonal, ql-continuous point processes on $(0, \infty)$ with a.s. unbounded compensators $\hat{\xi}_1, \ldots, \hat{\xi}_n$, and define*

$$Y_k(t) = \inf\Big\{s > 0;\ \hat{\xi}_k[0, s] > t\Big\}, \quad t \ge 0,\ \ k \le n.$$

Then the processes $\eta_k = \xi_k \circ Y_k^{-1}$ on R_+ are independent, unit-rate Poisson.

Proof: The processes η_k are Poisson by Theorem 15.15, since each Y_k is predictable and

$$\hat{\xi}_k[0, Y_k(t)] = \hat{\xi}_k\Big\{s \ge 0;\ \hat{\xi}_k[0, s] \le t\Big\} = t \ \text{ a.s.}, \quad t \ge 0,$$

by the continuity of $\hat{\xi}_k$. The multi-variate statement follows from the same theorem, if we apply the predictable map $Y = (Y_1, \ldots, Y_n)$ to the marked point process $\xi = (\xi_1, \ldots, \xi_n)$ on $(0, \infty) \times \{1, \ldots, n\}$ with compensator $\hat{\xi} = (\hat{\xi}_1, \ldots, \hat{\xi}_n)$. $\square$

From Theorem 15.15 we see that a ql-continuous, marked point process ξ is Poisson iff its compensator is a.s. non-random, similarly to the characterization of Brownian motion in Theorem 19.3. The following more general statement may be regarded as an extension of Theorem 15.10.

Theorem 15.17 *(extended Poisson criterion, Watanabe, Jacod) Let ξ be an S-marked, $\mathcal{F}$-adapted point process on $(0, \infty)$ with compensator $\hat{\xi}$, where S is Borel. Then these conditions are equivalent:*

(i) *$\hat{\xi} = \mu$ is a.s. non-random,*

(ii) *ξ is extended Poisson with $E\xi = \mu$.*

Proof (OK): It is enough to prove that (i) $\Rightarrow$ (ii). Then write

$$\mu = \nu + \sum\nolimits_k (\delta_{t_k} \otimes \mu_k), \qquad \xi = \eta + \sum\nolimits_k (\delta_{t_k} \otimes \xi_k),$$

where $t_1, t_2, \ldots > 0$ are the times with $\mu(\{t_k\} \times S) > 0$. Here ν is the restriction of μ to the remaining set $D^c \times S$ with $D = \bigcup_k \{t_k\}$, η is the corresponding restriction of ξ with $E\eta = \nu$, and the ξ_k are point processes on S with $\|\xi_k\| \le 1$ and $E\xi_k = \mu_k$. We need to show that η, $\xi_1, \xi_2, \ldots$ are independent, and that η is a Poisson process.

Then for any measurable $A \subset D^c \times S$ with $\nu A < \infty$, we define

$$X_t = \int_0^{t+} \!\! \int 1_A(r, s)\, \xi(dr\, ds), \qquad N_t = e^{-\hat{X}_t} 1\{X_t = 0\}.$$

As in Theorem 15.15, we may check that N is a uniformly integrable martingale with $N_0 = 1$ and $N_\infty = e^{\nu A} 1\{\eta A = 0\}$. For any $B_1, B_2, \ldots \in \mathcal{S}$, we further introduce the bounded martingales

$$M_t^k = (\xi_k - \mu_k) B_k 1\{t \ge t_k\}, \quad t \ge 0, \;\; k \in \mathsf{N}.$$

Since N and $M^1, M^2, \ldots$ are mutually orthogonal, the process

$$M_t^n = N_t \prod_{k \le n} M_t^k, \quad t \ge 0,$$

is again a uniformly integrable martingale, and so $EM_\infty^0 = 1$ and $EM_0^n = 0$ for all $n > 0$. Proceeding by induction, as in the proof of Theorem 10.27, we conclude that

$$E\Big(\prod_{k \le n} \xi_k B_k;\, \eta A = 0 \Big) = e^{-\nu A} \prod_{k \le n} \mu_k B_k, \quad n \in \mathsf{N}.$$

The desired joint distribution now follows by Corollary 15.9. $\square$

We may use discounted compensators to extend the previous time-change results beyond the ql-continuous case. To avoid distracting technicalities, we consider only unbounded point processes.

Theorem 15.18 *(time-change reduction to Poisson) Let $\xi = \sum_j \delta_{\tau_j}$ be a simple, unbounded, $\mathcal{F}$-adapted point process on $(0,\infty)$ with compensator η, let $\vartheta_1, \vartheta_2, \ldots \perp\!\!\!\perp \mathcal{F}$ be i.i.d. $U(0,1)$, and define*

$$\begin{aligned} \rho_t &= 1 - \textstyle\sum_j \vartheta_j \, 1\{t = \tau_j\}, \\ Y_t &= \eta_t^c - \sum_{s \le t} \log\Big(1 - \rho_s \, \Delta\eta_s\Big), \quad t \ge 0. \end{aligned}$$

Then $\tilde{\xi} = \xi \circ Y^{-1}$ is a unit rate Poisson process on R_+.

Proof: Write $\sigma_0 = 0$, and let $\sigma_j = Y_{\tau_j}$ for $j > 0$. We claim that the differences $\sigma_j - \sigma_{j-1}$ are i.i.d. and exponentially distributed with mean 1. Since the τ_j are orthogonal, it is enough to consider σ_1. Letting $\mathcal{G}$ be the right-continuous filtration induced by $\mathcal{F}$ and the pairs (σ_j, ϑ_j), we note that (τ_1, ϑ_1) has $\mathcal{G}$-compensator $\tilde{\eta} = \eta \otimes \lambda$ on $[0, \tau_1] \times [0,1]$. Write ζ for the associated discounted version with projection $\bar{\zeta}$ onto R_+, and put $Z_t = 1 - \bar{\zeta}(0,t]$. Define the $\mathcal{G}$-predictable processes U and $V = e^{-U}$ on $\mathsf{R}_+ \times [0,1]$ by

$$\begin{aligned} U_{t,r} &= \eta_t^c - \sum_{s<t} \log(1 - \Delta\eta_s) - \log(1 - r\,\Delta\eta_t), \\ V_{t,r} &= \exp(-\eta_t^c) \prod_{s<t} (1 - \Delta\eta_s)\,(1 - r\,\Delta\eta_t) \\ &= Z_{t-}\,(1 - r\Delta\eta_t), \end{aligned}$$

where the last equality holds by Theorem 10.24 (ii). For any random variable γ with distribution function F, we have

$$\begin{aligned} F_{t-} &= P\{F_\gamma < F_{t-}\} \\ &\le P\{F_\gamma \le F_t\} = F_t, \quad t \in \mathsf{R}. \end{aligned}$$

Thus, $\zeta \circ V^{-1} \le \lambda$ a.s. on $[0,1]$, and so $V(\tau_1, 1 - \vartheta_1) = e^{-\sigma_1}$ is $U(0,1)$ by Theorem 10.27, which yields the required distribution for σ_1. □

The last result yields an asymptotic version of the Poisson characterization in Corollary 15.17.

Corollary 15.19 *(Poisson convergence, Brown) Let $\xi_1, \xi_2, \ldots$ be simple point processes on $(0,\infty)$ with compensators $\hat{\xi}_n$, and let ξ be a unit rate Poisson process on R_+. Then*[7]

$$\hat{\xi}_n[0,t] \xrightarrow{P} t, \;\; t > 0 \quad \Rightarrow \quad \xi_n \xrightarrow{vd} \xi.$$

Proof (OK): For any sub-sequence $N' \subset \mathsf{N}$, we have a.s. $\hat{\xi}_n[0,t] \to t$ for all $t \ge 0$ along a further sub-sequence N''. In particular, $\sup_{t \le c} \hat{\xi}_n\{t\} \to 0$ a.s. for all $c > 0$. Writing $\xi_n = \sum_j \delta_{\tau_{nj}}$ and putting $\sigma_{nj} = Y_n(\tau_{nj})$ with Y_n as before, we get a.s. for every $j \in \mathsf{N}$

$$|\,\tau_{nj} - \sigma_{nj}| \le \Big|\,\tau_{nj} - \eta_n[0,\tau_{nj}]\,\Big| + \Big|\,\eta_n[0,\tau_{nj}] - \sigma_{nj}\Big| \to 0.$$

[7] For the convergence $\xrightarrow{vd}$, see Chapter 23.

Letting $n \to \infty$ along N'', we get

$$\xi_n = \sum_j \delta_{\tau_{nj}} \xrightarrow{vd} \sum_j \delta_{\sigma_{nj}} \stackrel{d}{=} \xi,$$

which yields the required convergence. □

Poisson integrals of various kind occur frequently in applications. Here we give criteria for the existence of the integrals ξf, $(\xi - \xi')f$, and $(\xi - \mu)f$, where ξ and ξ' are independent Poisson processes with a common intensity μ. In each case, the integral may be defined as a limit in probability of elementary integrals ξf_n, $(\xi - \xi')f_n$, or $(\xi - \mu)f_n$, respectively, where the f_n are bounded with compact support and such that $|f_n| \le |f|$ and $f_n \to f$. We say that the integral of f *exists*, if the appropriate limit exists and is independent of the choice of approximating functions f_n.

Proposition 15.20 *(Poisson integrals) Let $\xi \perp\!\!\!\perp \xi'$ be Poisson processes on S with a common diffuse intensity μ. Then for measurable functions f on S, we have*

(i) ξf *exists* $\Leftrightarrow$ $\mu(|f| \wedge 1) < \infty$,

(ii) $(\xi - \xi')f$ *exists* $\Leftrightarrow$ $\mu(f^2 \wedge 1) < \infty$,

(iii) $(\xi - \mu)f$ *exists* $\Leftrightarrow$ $\mu(f^2 \wedge |f|) < \infty$.

In each case, the existence is equivalent to tightness of the corresponding set of approximating elementary integrals.

Proof: (i) This is a special case of Lemma 15.6 (ii).

(ii) For counting measures $\nu = \sum_k \delta_{s_k}$ we form a symmetrization $\tilde{\nu} = \sum_k \vartheta_k \delta_{s_k}$, where $\vartheta_1, \vartheta_2, \ldots$ are i.i.d. random variables with $P\{\vartheta_k = \pm 1\} = \frac{1}{2}$. By Theorem 5.17, the series $\tilde{\nu} f$ converges a.s. iff $\nu f^2 < \infty$, and otherwise $|\tilde{\nu} f_n| \xrightarrow{P} \infty$ for any bounded approximations $f_n = 1_{B_n} f$ with $B_n \in \hat{\mathcal{S}}$. The result extends by conditioning to any point process ν with symmetric randomization $\tilde{\nu}$. Now Proposition 15.3 exhibits $\xi - \xi'$ as such a randomization of the Poisson process $\xi + \xi'$, and (i) yields $(\xi + \xi')f^2 < \infty$ a.s. iff $\mu(f^2 \wedge 1) < \infty$.

(iii) Write $f = g + h$, where $g = f1\{|f| \le 1\}$ and $h = f1\{|f| > 1\}$. First let $\mu g^2 + \mu|h| = \mu(f^2 \wedge |f|) < \infty$. Since clearly $E(\xi f - \mu f)^2 = \mu f^2$, the integral $(\xi - \mu)g$ exists. Since also ξh exists by (i), even $(\xi - \mu)f = (\xi - \mu)g + \xi h - \mu h$ exists.

Conversely, if $(\xi - \mu)f$ exists, then so does $(\xi - \xi')f$, and (ii) yields $\mu g^2 + \mu\{h \ne 0\} = \mu(f^2 \wedge 1) < \infty$. The existence of $(\xi - \mu)g$ now follows by the direct assertion, and trivially even ξh exists. The existence of $\mu h = (\xi - \mu)g + \xi h - (\xi - \mu)f$ then follows, and so $\mu|h| < \infty$. □

The multi-variate case is more difficult, and so we consider only positive double integrals $\xi^2 f$ and $\xi\eta f$, where $\xi^2 = \xi \otimes \xi$ and $\xi\eta = \xi \otimes \eta$. Write $f_1 = \mu_2 \hat{f}$ and $f_2 = \mu_1 \hat{f}$, where $\hat{f} = f \wedge 1$ and μ_i denotes μ-integration in the i-th coordinate. Say that a function f on S^2 is *non-diagonal* if $f(x, x) \equiv 0$.

Theorem 15.21 *(double Poisson integrals, OK & Szulga) Let $\xi \perp\!\!\!\perp \eta$ be Poisson processes on S with a common intensity μ, and consider a measurable, non-diagonal function $f \geq 0$ on S^2. Then $\xi\eta f < \infty$ a.s. iff $\xi^2 f < \infty$ a.s., which holds iff*

(i) $f_1 \vee f_2 < \infty$ *a.e.* μ,

(ii) $\mu\{f_1 \vee f_2 > 1\} < \infty$,

(iii) $\mu^2\big(\hat{f};\, f_1 \vee f_2 \leq 1\big) < \infty$.

Note that this result covers even the case where $E\xi \neq E\eta$. For the proof, we may clearly take μ to be diffuse, and whenever convenient we may even choose $S = \mathsf{R}_+$ and $\mu = \lambda$. Define $\psi(x) = 1 - e^{-x}$.

Lemma 15.22 *(Poisson moments) Let ξ be Poisson with $E\xi = \mu$. Then for any measurable set $B \subset S$ and function $f \geq 0$ on S or S^2, we have*

(i) $E\,\psi(\xi f) = \psi\big\{\mu(\psi \circ f)\big\}$,

(ii) $P\{\xi B > 0\} = \psi(\mu B)$,

(iii) $E\,(\xi f)^2 = \mu f^2 + (\mu f)^2$,
$E\,\xi\eta f = \mu^2 f$,

(iv) $E\,(\xi\eta f)^2 = \mu^2 f^2 + \mu(\mu_1 f)^2 + \mu(\mu_2 f)^2 + (\mu^2 f)^2$.

Proof: (i)$-$(ii): Use Lemma 15.2 (ii) with $\eta = \mu$.

(iii) When $\mu f < \infty$, the first relation becomes $\mathrm{Var}(\xi f) = \mu f^2$, which holds by linearity, independence, and monotone convergence. For the second relation, use Fubini's theorem.

(iv) Use (iii) and Fubini's theorem to get

$$\begin{aligned} E(\xi\eta f)^2 &= E\{\xi(\eta f)\}^2 \\ &= E\mu_1(\eta f)^2 + E(\mu_1 \eta f)^2 \\ &= \mu_1 E(\eta f)^2 + E\{\eta(\mu_1 f)\}^2 \\ &= \mu^2 f^2 + \mu_1(\mu_2 f)^2 + \mu_2(\mu_1 f)^2 + (\mu^2 f)^2. \end{aligned}$$

□

Lemma 15.23 *(tail estimate) For measurable $f\colon S^2 \to [0,1]$ with $\mu^2 f < \infty$, we have*

$$P\Big\{\xi\eta f > \tfrac{1}{2}\,\mu^2 f\Big\} \gtrsim \psi\bigg(\frac{\mu^2 f}{1 + f_1^* \vee f_2^*}\bigg).$$

Proof: We may clearly take $\mu^2 f > 0$. By Lemma 15.22 (iii)$-$(iv), we have $E\,\xi\eta f = \mu^2 f$ and

$$\begin{aligned} E(\xi\eta f)^2 &\leq (\mu^2 f)^2 + \mu^2 f + f_1^* \mu f_1 + f_2^* \mu f_2 \\ &= (\mu^2 f)^2 + (1 + f_1^* + f_2^*)\,\mu^2 f, \end{aligned}$$

and so by Lemma 5.1,

$$\begin{aligned} P\Big\{\xi\eta f > \tfrac{1}{2}\,\mu^2 f\Big\} &\geq \frac{(1 - \frac{1}{2})^2\,(\mu^2 f)^2}{(\mu^2 f)^2 + (1 + f_1^* + f_2^*)\,\mu^2 f} \\ &\gtrsim \bigg(1 + \frac{1 + f_1^* \vee f_2^*}{\mu^2 f}\bigg)^{-1} \\ &\gtrsim \psi\bigg(\frac{\mu^2 f}{1 + f_1^* \vee f_2^*}\bigg). \end{aligned}$$

□

Lemma 15.24 *(decoupling) For non-diagonal, measurable functions $f \geq 0$ on S^2, we have*

$$E\,\psi(\xi^2 f) \asymp E\,\psi(\xi\eta f).$$

Proof: Here it is helpful to take $\mu = \lambda$ on R_+. We may further assume that f is supported by the wedge $\{s < t\}$ in R_+^2. It is then equivalent to show that

$$E\Big\{(V\cdot\xi)_\infty \wedge 1\Big\} \asymp E\Big\{(V\cdot\eta)_\infty \wedge 1\Big\},$$

where $V_t = \xi f(\cdot,t)\wedge 1$. Since V is predictable with respect to the filtration induced by ξ, η, the random time

$$\tau = \inf\Big\{t \geq 0;\ (V\cdot\eta)_t > 1\Big\}$$

is optional, and so the process $1\{\tau > t\}$ is again predictable. Noting that ξ and η are both compensated by λ, we get

$$\begin{aligned} E\Big\{(V\cdot\xi)_\infty;\ \tau = \infty\Big\} &\leq E(V\cdot\xi)_\tau \\ &= E(V\cdot\lambda)_\tau \\ &= E(V\cdot\eta)_\tau \\ &\leq E\Big\{(V\cdot\eta)_\infty \wedge 1\Big\} + 2\,P\{\tau < \infty\}, \end{aligned}$$

and so

$$\begin{aligned} E\Big\{(V\cdot\xi)_\infty \wedge 1\Big\} &\leq E\Big\{(V\cdot\xi)_\infty;\ \tau = \infty\Big\} + P\{\tau < \infty\} \\ &\leq E\Big\{(V\cdot\eta)_\infty \wedge 1\Big\} + 3\,P\{\tau < \infty\} \\ &\leq 4\,E\Big\{(V\cdot\eta)_\infty \wedge 1\Big\}. \end{aligned}$$

The same argument applies with the roles of ξ and η interchanged. □

Proof of Theorem 15.21: Since $\xi\otimes\eta$ is a.s. simple, we have $\xi\eta f < \infty$ iff $\xi\eta\hat{f} < \infty$ a.s., which allows us to take $f \leq 1$. First assume (i)–(iii). Since by (i)

$$E\,\xi\eta\Big(f;\ f_1 \vee f_2 = \infty\Big) \leq \textstyle\sum_i \mu^2(f;\ f_i = \infty) = 0,$$

we may assume that $f_1, f_2 < \infty$. Then Lemma 15.20 (i) gives $\eta f(s,\cdot) < \infty$ and $\xi f(\cdot,s) < \infty$ a.s. for all $s \geq 0$. Furthermore, (ii) implies $\xi\{f_1 > 1\} < \infty$ and $\eta\{f_2 > 1\} < \infty$ a.s. By Fubini's theorem, we get a.s.

$$\xi\eta\Big(f;\ f_1 \vee f_2 > 1\Big) \leq \xi\Big(\eta f;\ f_1 > 1\Big) + \eta\Big(\xi f;\ f_2 > 1\Big) < \infty,$$

which allows us to assume that even $f_1, f_2 \leq 1$. Then (iii) yields $E\,\xi\eta f = \mu^2 f < \infty$, which implies $\xi\eta f < \infty$ a.s.

Conversely, suppose that $\xi\eta f < \infty$ a.s. for a function $f\colon S^2 \to [0,1]$. By Lemma 15.22 (i) and Fubini's theorem,

$$E\,\psi\Big\{\mu\,\psi(t\,\eta f)\Big\} = E\,\psi(t\,\xi\eta f) \to 0, \quad t \downarrow 0,$$

which implies $\mu\,\psi(t\,\eta f) \to 0$ a.s. and hence $\eta f < \infty$ a.e. $\mu \otimes P$. By Lemma 15.20 and Fubini's theorem we get $f_1 = \mu_2 f < \infty$ a.e. μ, and the symmetric argument yields $f_2 = \mu_1 f < \infty$ a.e. This proves (i).

Next, Lemma 15.22 (i) yields on the set $\{f_1 > 1\}$

$$\begin{aligned} E\,\psi(\eta f) &= \psi\Big\{\mu_2(\psi \circ f)\Big\} \\ &\geq \psi\Big\{(1 - e^{-1}) f_1\Big\} \\ &\geq \psi(1 - e^{-1}) \equiv c > 0. \end{aligned}$$

Hence, for any measurable set $B \subset \{f_1 > 1\}$,

$$E\,\mu\Big\{1 - \psi(\eta f);\, B\Big\} \leq (1 - c)\,\mu B,$$

and so by Chebyshev's inequality,

$$\begin{aligned} P\Big\{\mu\,\psi(\eta f) < \tfrac{1}{2}\,c\,\mu B\Big\} &\leq P\Big\{\mu\{1 - \psi(\eta f);\, B\} > (1 - \tfrac{1}{2}\,c)\,\mu B\Big\} \\ &\leq \frac{E\,\mu\Big\{1 - \psi(\eta f);\, B\Big\}}{(1 - \frac{1}{2}\,c)\,\mu B} \leq \frac{1 - c}{1 - \frac{1}{2}\,c}\,. \end{aligned}$$

Since B was arbitrary, we conclude that

$$P\Big\{\mu\,\psi(\eta f) \geq \tfrac{1}{2}\,c\,\mu\{f_1 > 1\}\Big\} \geq 1 - \frac{1 - c}{1 - \frac{1}{2}\,c} = \frac{c}{2 - c} > 0.$$

Noting that $\mu\,\psi(\eta f) < \infty$ a.s. by Lemma 15.20 and Fubini's theorem, we obtain $\mu\{f_1 > 1\} < \infty$. Combining with a similar result for f_2, we obtain (ii).

Applying Lemma 15.23 to the function $f\,1\{f_1 \vee f_2 \leq 1\}$ gives

$$P\Big\{\xi\eta f > \tfrac{1}{2}\,\mu^2(f;\, f_1 \vee f_2 \leq 1)\Big\} \gtrsim \psi\Big\{\tfrac{1}{2}\,\mu^2(f;\, f_1 \vee f_2 \leq 1)\Big\},$$

which implies (iii), since the opposite statement would yield the contradiction $P\{\xi\eta f = \infty\} > 0$.

To extend the result to the integral $\xi^2 f$, we see from Lemma 15.24 that

$$E\,\psi(t\,\xi^2 f) \asymp E\,\psi(t\,\xi\eta f), \quad t > 0.$$

Letting $t \to 0$ gives

$$P\{\xi^2 f = \infty\} \asymp P\{\xi\eta f = \infty\},$$

which implies $\xi^2 f < \infty$ a.s. iff $\xi\eta f < \infty$ a.s. □

Exercises

1. Let ξ be a point process on a Borel space S. Show that $\xi = \sum_k \delta_{\tau_k}$ for some random elements τ_k in $S \cup \{\Delta\}$, where $\Delta \notin S$ is arbitrary. Extend the result to general random measures. (*Hint:* We may assume that $S = \mathsf{R}_+$.)

2. Show that two random measures ξ, η are independent iff $Ee^{-\xi f-\eta g} = Ee^{-\xi f} \times Ee^{-\eta g}$ for all measurable $f, g \geq 0$. When ξ, η are simple point processes, prove the equivalence of $P\{\xi B + \eta C = 0\} = P\{\xi B = 0\}P\{\eta C = 0\}$ for any $B, C \in \mathcal{S}$. (*Hint:* Regard (ξ, η) as a random measure on $2\,S$.)

3. Let $\xi_1, \xi_2, \ldots$ be independent Poisson processes with intensity measures $\mu_1, \mu_2, \ldots$, such that the measure $\mu = \sum_k \mu_k$ is σ-finite. Show that $\xi = \sum_k \xi_k$ is again Poisson with intensity measure μ.

4. Show that the classes of mixed Poisson and binomial processes are preserved by randomization.

5. Let ξ be a Cox process on S directed by a random measure η, and let f be a measurable mapping into a space T, such that $\eta \circ f^{-1}$ is a.s. σ-finite. Prove directly from definitions that $\xi \circ f^{-1}$ is a Cox process on T directed by $\eta \circ f^{-1}$. Derive a corresponding result for p-thinnings. Also show how the result follows from Proposition 15.3.

6. Use Proposition 15.3 to show that (iii) $\Rightarrow$ (ii) in Corollary 13.8, and use Theorem 15.10 to prove the converse.

7. Consider a p-thinning η of ξ and a q-thinning ζ of η with $\zeta \perp\!\!\!\perp_\eta \xi$. Show that ζ is a $p\,q$-thinning of ξ.

8. Let ξ be a Cox process directed by η or a p-thinning of η with $p \in (0,1)$, and fix two disjoint sets $B, C \in \mathcal{S}$. Show that $\xi B \perp\!\!\!\perp \xi C$ iff $\eta B \perp\!\!\!\perp \eta C$. (*Hint:* Compute the Laplace transforms. The 'if' assertions can also be obtained from Proposition 8.12.)

9. Use Lemma 15.2 to derive expressions for $P\{\xi B = 0\}$ when ξ is a Cox process directed by η, a μ-randomization of η, or a p-thinning of η. (*Hint:* Note that $Ee^{-t\,\xi B} \to P\{\xi B = 0\}$ as $t \to \infty$.)

10. Let ξ be a p-thinning of η, where $p \in (0,1)$. Show that ξ, η are simultaneously Cox. (*Hint:* Use Lemma 15.7.)

11. (*Fichtner*) For a fixed $p \in (0,1)$, let η be a p-thinning of a point process ξ on S. Show that ξ is Poisson iff $\eta \perp\!\!\!\perp \xi - \eta$. (*Hint:* Extend by iteration to arbitrary p. Then a uniform randomization of ξ on $S \times [0,1]$ has independent increments in the second variable, and the result follows by Theorem 19.3.)

12. Use Theorem 15.8 to give a simplified proof of Theorem 15.4, in the case of simple ξ.

13. Derive Theorem 15.4 from Theorem 15.14. (*Hint:* Note that ξ is symmetric on S iff it is symmetric on B_n for every n. If ξ is simple, the assertion follows immediately from Theorem 15.14. Otherwise, apply the same result to a uniform randomization on $S \times [0,1]$.)

14. For ξ as in Theorem 15.14, show that $P\{\xi B = 0\} = \varphi(\mu B)$ for a completely monotone function φ. Conclude from the Hausdorff–Bernstein characterization and Theorem 15.8 that ξ is a mixed Poisson or binomial process based on μ.

15. Show that the distribution of a simple point process ξ on R may *not* be determined by the distributions of ξI for all intervals I. (*Hint:* If ξ is restricted to $\{1, \ldots, n\}$, then the distributions of all ξI give $\sum_{k \leq n} k(n-k+1) \leq n^3$ linear relations between the $2^n - 1$ parameters.)

16. Show that the distribution of a point process may *not* be determined by the one-dimensional distributions. (*Hint:* If ξ is restricted to $\{0,1\}$ with $\xi\{0\} \vee \xi\{1\} \leq n$, then the one-dimensional distributions give $4n$ linear relations between the $n(n+2)$ parameters.)

17. Show that Lemma 15.1 remains valid with $B_1, \dots, B_n$ restricted to an arbitrary pre-separating class $\mathcal{C}$, as defined in Chapter 23 or Appendix 6. Also show that Theorem 15.8 holds with B restricted to a separating class. (*Hint:* Extend to the case where $\mathcal{C} = \{B \in \mathcal{S};\, (\xi + \eta)\partial B = 0 \text{ a.s.}\}$. Then use monotone-class arguments for sets in S and $\mathcal{M}_S$.)

18. Show that Theorem 15.10 may fail without the condition $E\xi(\{s\} \times K) \equiv 0$.

19. Give an example of a non-Poisson point process ξ on S such that ξB is Poisson for every $B \in \mathcal{S}$. (*Hint:* Take $S = \{0, 1\}$.)

20. Describe the marked point processes ξ with independent increments, in the presence of possible fixed atom sites. (*Hint:* Note that there are at most countably many fixed discontinuities, and that the associated component of ξ is independent of the remaining part. Now use Proposition 5.14.)

21. Describe the random measures ξ with independent increments, in the presence of possible fixed atom sites. Write the general representation as in Corollary 15.11, for a suitably generalized Poisson process η.

22. Show by an example that Theorem 15.15 may fail without the assumption of ql-continuity.

23. Use Theorem 15.15 to show that any ql-continuous simple point process ξ on $(0, \infty)$ with a.s. unbounded compensator $\hat{\xi}$ can be time-changed into a stationary Poisson process η w.r.t. the correspondingly time-changed filtration. Also extend the result to possibly bounded compensators, and describe the reverse map $\eta \mapsto \xi$.

24. Extend Corollary 15.16 to possibly bounded compensators. Show that the result fails in general, when the compensators are not continuous.

25. For $\xi_1, \dots, \xi_n$ as in Corollary 15.16 with compensators $\hat{\xi}_1, \dots, \hat{\xi}_n$, let $Y_1, \dots, Y_n$ be predictable maps into T such that the measures $\hat{\xi}_k \circ Y_k^{-1} = \mu_k$ are non-random in $\mathcal{M}_T$. Show that the processes $\xi_k \circ Y_k^{-1}$ are independent Poisson $\mu_1, \dots, \mu_n$.

26. Prove Theorem 15.20 (i), (iii) by means of characteristic functions.

27. Let $\xi \perp\!\!\!\perp \eta$ be Poisson processes on S with $E\xi = E\eta = \mu$, and let $f_1, f_2, \dots : S \to \mathsf{R}$ be measurable with $\infty > \mu(f_n^2 \wedge 1) \to \infty$. Show that $|(\xi - \eta) f_n| \xrightarrow{P} \infty$. (*Hint:* Consider the symmetrization $\tilde{\nu}$ of a fixed measure $\nu \in \mathcal{N}_S$ with $\nu f_n^2 \to \infty$, and argue along sub-sequences, as in the proof of Theorem 5.17.)

28. Give conditions for $\xi^2 f < \infty$ a.s. when ξ is Poisson with $E\xi = \lambda$ and $f \geq 0$ on R_+^2. (*Hint:* Combine Theorems 15.20 and 15.21.)

Chapter 16

Independent-Increment and Lévy Processes

Infinitely divisible distributions, independent-increment processes, Lévy processes and subordinators, orthogonality and independence, stable Lévy processes, first-passage times, coupling of Lévy processes, approximation of random walks, arcsine laws, time change of stable integrals, tangential existence and comparison

Just as random walks are the most basic Markov processes in discrete time, so also the Lévy processes constitute the most fundamental among continuous-time Markov processes, as they are precisely the processes of this kind that are both space- and time-homogeneous. They may also be characterized as processes with stationary, independent increments, which suggests that they be thought of simply as random walks in continuous time. Basic special cases are given by Brownian motion and homogeneous Poisson processes.

Dropping the stationarity assumption leads to the broader class of processes with independent increments, and since many of the basic properties are the same, it is natural to extend our study to this larger class. Under weak regularity conditions, their paths are right-continuous with left-hand limits (rcll), and the one-dimensional distributions agree with the infinitely divisible laws studied in Chapter 7. As the latter are formed by a combination of Gaussian and compound Poisson distributions, the general independent-increment processes are composed of Gaussian processes and suitable Poisson integrals.

The classical limit theorems of Chapters 6–7 suggest similar approximations of random walks by Lévy processes, which may serve as an introduction to the general weak convergence theory in Chapters 22–23. In this context, we will see how results for random walks or Brownian motion may be extended to a more general continuous-time context. In particular, two of the arcsine laws derived for Brownian motion in Chapter 14 remain valid for general symmetric Lévy processes.

Independent-increment processes may also be regarded as the simplest special cases of semi-martingales, to be studied in full generality in Chapter 20. We conclude the chapter with a discussion of tangential processes, where the idea is to approximate a general semi-martingale X by a process $\tilde{X}$ with conditionally independent increments, such that X and $\tilde{X}$ have similar asymptotic properties. This is useful, since the asymptotic behavior of the latter processes can often be determined by elementary methods.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_17

To resume our discussion of general processes with independent increments, we say that an R^d-valued process X is *continuous in probability*[1] if $X_s \xrightarrow{P} X_t$ whenever $s \to t$. This clearly prevents the existence of fixed discontinuities. To examine the structure of processes with independent increments, we may first choose regular versions of the paths. Recall that X is said to be *rcll*[2], if it has finite limits $X_{t\pm}$ from the right and left at every point $t \geq 0$, and the former equals X_t. Then the only possible discontinuities are jumps, and we say that X has a *fixed jump* at a time $t > 0$ if $P\{X_t \neq X_{t-}\} > 0$.

Lemma 16.1 *(infinitely divisible distributions, de Finetti) For a random vector ξ in R^d, these conditions are equivalent:*

(i) *ξ is infinitely divisible,*

(ii) *$\xi \overset{d}{=} X_1$ for a process X in R^d with stationary, independent increments and $X_0 = 0$.*

We may then choose X to be continuous in probability, in which case

(iii) *$\mathcal{L}(X)$ is uniquely determined by $\mathcal{L}(\xi)$,*

(iv) *$\xi \in \mathsf{R}^d_+$ a.s. $\Leftrightarrow$ $X_t \in \mathsf{R}^d_+$ a.s., $t \geq 0$.*

Proof: To prove (i) ⇔ (ii), let ξ be infinitely divisible. Using the representation in Theorem 7.5, we can construct the finite-dimensional distributions of an associated process X in R^d with stationary, independent increments, and then invoke Theorem 8.23 for the existence of an associated process. Assertions (iii)−(iv) also follow from the same theorem. □

We first describe the processes with stationary, independent increments:

Theorem 16.2 *(Lévy processes and subordinators, Lévy, Itô) Let the process X in R^d be continuous in probability with $X_0 = 0$. Then X has stationary, independent increments iff it has an rcll version, given for $t \geq 0$ by*

$$X_t = b\,t + \sigma B_t + \int_0^t \int_{|x|\leq 1} x\,(\eta - E\eta)(ds\,dx) + \int_0^t \int_{|x|>1} x\,\eta(ds\,dx),$$

where $b \in \mathsf{R}^d$, σ is a $d \times d$ matrix, B is a Brownian motion in R^d, and η is an independent Poisson process on $\mathsf{R}_+ \times (\mathsf{R}^d \setminus \{0\})$ with $E\eta = \lambda \otimes \nu$, where $\int(|x|^2 \wedge 1)\,\nu(dx) < \infty$. When X is R^d_+-valued, the representation simplifies to

$$X_t = a\,t + \int_0^t \int x\,\eta(ds\,dx) \;\; a.s., \quad t \geq 0,$$

where $a \in \mathsf{R}^d_+$, and η is a Poisson process on $\mathsf{R}_+ \times (\mathsf{R}^d_+ \setminus \{0\})$ with $E\eta = \lambda \otimes \nu$, where $\int(|x| \wedge 1)\,\nu(dx) < \infty$. The triple (b, a, ν) with $a = \sigma'\sigma$ or pair (a, ν) is then unique, and any choice with the stated properties may occur.

[1] also said to be *stochastically continuous*

[2] right-continuous with left-hand limits; the French acronym *càdlàg* is also common, even in English texts

A process X in R^d with stationary, independent increments is called a *Lévy processes*, and when X is R_+-valued it is also called a *subordinator*[3]. Furthermore, the measure ν is called the *Lévy measure* of X. The result follows easily from the theory of infinitely divisible distributions:

Proof: Given Theorem 7.5, it is enough to show that X has an rcll version. This being obvious for the first two terms as well as for the final integral term, it remains to prove that the first double integral M_t has an rcll version. By Theorem 9.17 we have $M_t^\varepsilon \to M_t$ uniformly in L^2 as $\varepsilon \to 0$, where M_t^ε is the sum of compensated jumps of size $|x| \in (\varepsilon, 1]$, hence a martingale. Proceeding along an a.s. convergent sub-sequence, we see that the rcll property extends to the limit. □

The last result can also be proved by probabilistic methods. We will develop the probabilistic approach in a more general setting, where we drop the stationarity assumption and even allow some fixed discontinuities. Given an rcll process X in R^d, we define the associated *jump point process* η on $(0,\infty) \times (\mathsf{R}^d \setminus \{0\}$ by

$$\eta = \sum_{t>0} \delta_{t,\Delta X_t} = \sum_{t>0} 1\Big\{(t, \Delta X_t) \in \cdot\Big\}, \tag{1}$$

where the summations extend over all times $t > 0$ with $\Delta X_t = X_t - X_{t-} \neq 0$.

Theorem 16.3 *(independent-increment processes, Lévy, Itô, Jacod, OK) Let the process X in R^d be rcll in probability with independent increments and $X_0 = 0$. Then*

(i) *X has an rcll version with jump point process η, given for $t \geq 0$ by*

$$X_t = b_t + G_t + \int_0^t \int_{|x|\leq 1} x\,(\eta - E\eta)(ds\,dx) + \int_0^t \int_{|x|>1} x\,\eta(ds\,dx), \tag{2}$$

where b is an rcll function in R^d with $b_0 = 0$, G is a continuous, centered Gaussian process in R^d with independent increments and $G_0 = 0$, and η is an independent, extended Poisson process on $\mathsf{R}'_+ \times (\mathsf{R}^d \setminus \{0\})$ with[4]

$$\int_0^t \int (|x|^2 \wedge 1)\, E\eta'(ds\,dx) < \infty, \quad t > 0, \tag{3}$$

(ii) *for non-decreasing X, the representation simplifies to*

$$X_t = a_t + \int_0^t \int x\,\eta(ds\,dx) \ \ a.s., \quad t \geq 0,$$

where a is a non-decreasing, rcll function in R_+^d with $a_0 = 0$, and η is an extended Poisson process on $\mathsf{R}'_+ \times (\mathsf{R}_+^d \setminus \{0\})$ with

$$\int_0^t \int (|x| \wedge 1)\, E\eta(ds\,dx) < \infty, \quad t > 0.$$

[3]This curious term comes from the fact that the composition of a Markov process X with a subordinator T yields a new Markov process $Y = X \circ T$.

[4]Here we need to replace η by a slightly modified version η', as explained below.

Both representations are a.s. unique, and any processes η and G with the stated properties may occur.

First we need to clarify the meaning of the jump component in (i) and the associated integrability condition. Here we consider separately the continuous Poisson component and the process of fixed discontinuities. For the former, let η be a stochastically continuous Poisson process on $\mathsf{R}_+ \times (\mathsf{R}^d \setminus \{0\}$, and introduce the elementary processes

$$Y_t^r = \int_0^t \int_{r<|x|\le 1} x\,(\eta - E\eta)(ds\,dx) + \int_0^t \int_{|x|>1} x\,\eta(ds\,dx), \quad t \ge 0,$$

with associated symmetrizations $\tilde{Y}_t^r$.

Proposition 16.4 *(Poisson component) For any stochastically continuous Poisson process η on $\mathsf{R}_+ \times (\mathsf{R}^d \setminus \{0\})$, these conditions are equivalent:*

(i) *the family $\{\tilde{Y}_t^r;\, r > 0\}$ is tight for every $t \ge 0$,*

(ii)
$$\int_0^t \int (|x|^2 \wedge 1)\, E\eta(ds\,dx) < \infty, \quad t > 0,$$

(iii) *there exists an rcll process Y in R^d with*

$$(Y^r - Y)_t^* \to 0 \ \textit{a.s.}, \quad t \ge 0.$$

Proof: Since clearly (iii) $\Rightarrow$ (i), it is enough to prove (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii).

(i) $\Rightarrow$ (ii): Assume (i), and fix any $t > 0$. By Lemma 6.2 we can choose $a > 0$, such that $Ee^{iu\tilde{Y}_t^r} \ge \frac{1}{2}$ for all $r < 1$ and $u \le a$. Then by Lemma 15.2 and Fubini's theorem,

$$\begin{aligned} -\tfrac{1}{2}\int_0^a \log E\, e^{iu\tilde{Y}_t^r}\, du &= \int_0^a du \int_{|x|>r} (1 - \cos ux)\, E\eta_t(dx) \\ &= a \int_{|x|>r} \left(1 - \frac{\sin ax}{ax}\right) E\eta_t(dx) \\ &\approx a \int_{|x|>r} (|ax|^2 \wedge 1)\, E\eta_t(dx). \end{aligned}$$

Now take the lim sup as $r \to 0$.

(ii) $\Rightarrow$ (iii): By Lemma 15.22, we have $\mathrm{Var}|\eta f| = E\eta|f|^2$ for suitably bounded functions f. Hence as $r, r' \to 0$,

$$\begin{aligned} E\,|Y_t^r - Y_t^{r'}|^2 &= \mathrm{Var} \int_0^t \int_{r<|x|<r'} x\,\eta(ds\,dx) \\ &= \int_0^t \int_{r<|x|<r'} |x|^2 E\eta(ds\,dx) \to 0. \end{aligned}$$

Now the differences $Y^r - Y^{r'}$ are martingales, and so by Doob's inequality

$$(Y^r - Y^{r'})_t^* \xrightarrow{P} 0, \quad t > 0, \ r, r' \to 0.$$

Since the r-increments of Y^r are further independent, Lemma 11.13 yields an rcll process Y with

$$(Y^r - Y)_t^* \to 0 \text{ a.s.}, \quad t > 0, \ r \to 0 \text{ along } \mathsf{Q}_+.$$

We may finally extend the convergence to R_+, since Y^r is a.s. right continuous in $r > 0$, uniformly on $[0, t]$. □

As for the fixed discontinuities, let $D \subset (0, \infty)$ be the countable supporting set and write $\eta = \sum_{t \in D} \delta_{t, \xi_t}$, where the ξ_t are independent random vectors in R^d, and we interpret $\delta_{t,0}$ as 0. Fixing any finite sets $D^n \uparrow D$ and putting $D_t^n = D^n \cap [0, t]$, we introduce the elementary pure jump-type processes

$$Y_t^n = \sum_{s \in D_t^n} \Big(\xi_s - E\{\xi_s;\, |\xi_s| \le 1\} \Big), \quad t \ge 0, \; n \in \mathsf{N},$$

with associated symmetrizations $\tilde{Y}_t^n$. Put $D_t = [0, t] \cap D$.

Proposition 16.5 *(fixed discontinuities) Let the ξ_t be independent random vectors in R^d indexed by a countable set $D \subset (0, \infty)$, such that $\sum_{s \le t} 1\{|\xi_s| > 1\} < \infty$ a.s. for all $t > 0$. Then these conditions are equivalent:*

(i) *the family $\{\tilde{Y}_t^n;\, n > 0\}$ is tight for every $t \ge 0$,*

(ii) $$\sum_{s \in D_t} \Big(\mathrm{Var}\{\xi_s;\, |\xi_s| \le 1\} + P\{|\xi_s| > 1\} \Big) < \infty, \quad t > 0,$$

(iii) *there exists an rcll process Y in R^d with*

$$Y_t^n \to Y_t \;\; a.s., \quad t \ge 0.$$

Proof: Here again (iii) ⇒ (i), and so it remains to prove (i) ⇒ (ii) ⇒ (iii).

(i) ⇒ (ii): Fix any $t > 0$. By Theorem 4.18 we may assume that $|\xi_t| \le 1$ for all t. Then Theorem 5.17 yields

$$\sum_{s \in D_t} \mathrm{Var}(\xi_s) = \tfrac{1}{2} \sum_{s \in D_t} \mathrm{Var}(\tilde{\xi}_s) < \infty.$$

(ii) ⇒ (iii): Under (i), we have convergence $Y_t^n \to Y_t$ a.s. for every $t \ge 0$ by Theorem 5.18, and we also note that the limiting process Y is rcll in probability. The existence of a version with rcll paths then follows by Proposition 16.6 below. □

We proceed to establish the pathwise regularity.

Proposition 16.6 *(regularization) For a process X in R^d with independent increments, these conditions are equivalent:*

(i) *X is rcll in probability,*

(ii) *X has a version with rcll paths.*

This requires a simple analytic result.

Lemma 16.7 *(complex exponentials) For any constants $x_1, x_2, \ldots \in \mathsf{R}^d$, these conditions are equivalent:*

(i) *x_n converges in R^d,*

(ii) *e^{iux_n} converges in C for almost every $u \in \mathsf{R}^d$.*

Proof: Assume (ii). Fix a standard Gaussian random vector ζ in R^d, and note that $\exp\{it\,\zeta(a_m - a_n)\} \to 1$ a.s. as $m, n \to \infty$ for fixed $t \in \mathsf{R}$. Then by dominated convergence $E\exp\{it\,\zeta(a_m - a_n)\} \to 1$, and so $\zeta(a_m - a_n) \xrightarrow{P} 0$ by Theorem 6.3, which implies $a_m - a_n \to 0$. Thus, the sequence (a_n) is Cauchy, and (i) follows. □

Proof of Proposition 16.6: Assume (i). In particular, $X_t \xrightarrow{P} 0$ as $t \to 0$, and so $\varphi_t(u) \to 1$ for every $u \in \mathsf{R}^d$, where

$$\varphi_t(u) = Ee^{iuX_t}, \quad t \geq 0, \ u \in \mathsf{R}^d.$$

For fixed $u \in \mathsf{R}^d$ we may then choose a $t_u > 0$, such that $\varphi_t(u) \neq 0$ for all $t \in [0, t_u)$. Then the process

$$M_t^u = \frac{e^{iuX_t}}{\varphi_t(u)}, \quad t < t_u, \ u \in \mathsf{R}^d,$$

exists and is a martingale in t, and so by Theorem 9.28 it has a version with rcll paths. Since $\varphi_t(u)$ has the same property, we conclude that even e^{iuX_t} has an rcll version. By a similar argument, every $s > 0$ has a neighborhood I_s, such that e^{iuX_t} has an rcll version on I_s, except for a possible jump at s. The rcll property then holds on all of I_s, and by compactness we conclude that e^{iuX_t} has an rcll version on the entire domain R_+.

Now let A be the set of pairs $(\omega, u) \in \Omega \times \mathsf{R}^d$, such that e^{iuX_t} has right and left limits along Q at every point. Expressing this in terms of upcrossings, we see that A is product measurable, and the previous argument shows that all u-sections A_u have probability 1. Hence, by Fubini's theorem, P-almost every ω-section A^ω has full λ^d-measure. This means that, on a set Ω' of probability 1, e^{iuX_t} has the stated continuity properties for almost every u. Then by Lemma 16.7, X_t itself has the same continuity properties on Ω'. Now define $\tilde{X}_t \equiv X_{t+}$ on Ω', and put $\tilde{X}_t \equiv 0$ otherwise. Then $\tilde{X}$ is clearly rcll, and by (i) we have $\tilde{X}_t = X_t$ a.s. for all t. □

To separate the jump component from the remainder of X, we need the following independence criterion:

Lemma 16.8 *(orthogonality and independence) Let (X, Y) be an rcll processes in $(\mathsf{R}^d)^2$ with independent increments and $X_0 = Y_0 = 0$, where Y is an elementary step process. Then*

$$(\Delta X)(\Delta Y) = 0 \ \text{ a.s.} \quad \Rightarrow \quad X \perp\!\!\!\perp Y.$$

Proof: By a transformation of the jump sizes, we may assume that Y has locally integrable variation. Now introduce as before the characteristic functions

$$\varphi_t(u) = Ee^{iuX_t}, \quad \psi_t(v) = Ee^{ivY_t}, \qquad t \geq 0, \ u, v \in \mathsf{R}^d,$$

and note that both are $\neq 0$ on some interval $[0, t_{u,v})$, where $t_{u,v} > 0$ for all $u, v \in \mathsf{R}^d$. On the same interval, we can then introduce the martingales

$$M_t^u = \frac{e^{iuX_t}}{\varphi_t(u)}, \quad N_t^v = \frac{e^{ivY_t}}{\psi_t(v)}, \qquad t < t_{u,v}, \;\; u, v \in \mathsf{R}^d,$$

where N^v has again integrable variation on every compact subinterval. Fixing any $t \in (0, t_{u,v})$ and writing $h = t/n$ with $n \in \mathsf{N}$, we get by the martingale property and dominated convergence

$$\begin{aligned} E M_t^u N_t^v - 1 &= E \sum_{k \le n} \Big(M_{kh}^u - M_{(k-1)h}^u \Big) \Big(N_{kh}^v - N_{(k-1)h}^v \Big) \\ &= E \int_0^t \Big(M_{[sn+1-]h}^u - M_{[sn-]h}^u \Big)\, dN_s^v \\ &\to E \int_0^t (\Delta M_s^u)\, dN_s^v \\ &= E \sum_{s \le t} (\Delta M_s^u)(\Delta N_s^v) = 0. \end{aligned}$$

This shows that $E M_t^u N_t^v = 1$, and so

$$E e^{iuX_t + ivY_t} = E e^{iuX_t} E e^{ivY_t}, \quad t < t_{u,v}, \;\; u, v \in \mathsf{R}^d.$$

By a similar argument, along with the independence

$$(\Delta X_t,\, \Delta Y_t) \perp\!\!\!\perp (X - \Delta X_t,\, Y - \Delta Y_t), \quad t > 0,$$

every point $r > 0$ has a neighborhood $I_{u,v}$, such that

$$E e^{iuX_s^t + ivY_s^t} = E e^{iuX_s^t} E e^{ivY_s^t}, \quad s, t \in I_{u,v}, \;\; u, v \in \mathsf{R}^d,$$

where $X_s^t = X_t - X_s$ and $Y_s^t = Y_t - Y_s$. By a compactness argument, the same relation holds for all $s, t \ge 0$, and since u, v were arbitrary, Theorem 6.3 yields $X_s^t \perp\!\!\!\perp Y_s^t$ for all $s, t \ge 0$. Since (X, Y) has independent increments, the independence extends to any finite collection of increments, and we get $X \perp\!\!\!\perp Y$. □

The last lemma allows us to eliminate the compensated random jumps, which leaves us with a process with only non-random jumps. It remains to show that the latter process is Gaussian.

Proposition 16.9 *(Gaussian processes with independent increments) Let X be an rcll process in R^d with independent increments and $X_0 = 0$. Then these conditions are equivalent:*

(i) *X has only non-random jumps,*

(ii) *X is Gaussian of the form*

$$X_t = b_t + G_t \;\; a.s., \quad t \ge 0,$$

where G is a continuous, centered, Gaussian process with independent increments and $G_0 = 0$, and b is an rcll function in R^d with $b_0 = 0$.

Proof: We need to prove only (i) ⇒ (ii), so assume (i). Let $\tilde{X}$ be a symmetrization of X, and note that $\tilde{X}$ has again independent increments. It is also is a.s. continuous, since the only discontinuities of X are constant jumps, which are canceled by the symmetrization. Hence, $\tilde{X}$ is symmetric Gaussian by Theorem 14.4.

Now fix any $t > 0$, and write ξ_{nj} and $\tilde{\xi}_{nj}$ for the increments of X, $\tilde{X}$ over the n sub-intervals of length t/n. Since $\max_j |\tilde{\xi}_{nj}| \to 0$ a.s. by compactness, the $\tilde{\xi}_{nj}$ form a null array in R^d, and so by Theorem 6.12,

$$\textstyle\sum_j P\{|\tilde{\xi}_{nj}| > \varepsilon\} \to 0, \quad \varepsilon > 0.$$

Letting m_{nj} be component-wise medians of ξ_{nj}, we see from Lemma 5.19 that the differences $\xi'_{nj} = \xi_{nj} - m_{nj}$ again form a null array satisfying

$$\textstyle\sum_j P\{|\xi'_{nj}| > \varepsilon\} \to 0, \quad \varepsilon > 0. \tag{4}$$

Putting $m_n = \sum_j m_{nj}$, we may write

$$X_t = \textstyle\sum_j \xi'_{nj} + m_n = \sum_j \xi'_{nj} + (n|m_n|)\dfrac{m_n}{n|m_n|}, \quad n \in \mathsf{N},$$

where the $n(1 + |m_n|)$ terms on the right again form a null array in R^d, satisfying a condition like (4). Then X_t is Gaussian by Theorem 6.16. A similar argument shows that all increments $X_t - X_s$ are Gaussian, and so the entire process X has the same property, and the difference $G = X - EX$ is centered Gaussian. It is further continuous, since X has only constant jumps, which all go into the mean $b_t = EX_t = X_t - G_t$. The latter is clearly non-random and rcll, and the independence of the increments carries over to G. □

Proof of Theorem 16.3 (OK): By Proposition 16.6 we may assume that X has rcll paths. Then introduce the associated jump point process η on $\mathsf{R}_+ \times (\mathsf{R}^d \setminus \{0\})$, which clearly inherits the property of independent increments. By Theorem 15.10 the process η is then extended Poisson, and hence can be decomposed into a stochastically continuous Poisson process η_1 and a process $\eta_2 \perp\!\!\!\perp \eta_1$ with fixed discontinuities, supported by a countable set $D \subset \mathsf{R}_+$.

Let Y_1^r, Y_2^n be the associated approximating processes in Propositions 16.4 and 16.5, and note that by Lemma 16.8 the pairs (Y_1^r, Y_2^n) are independent from the remaining parts of X. Factoring the characteristic functions, we see from Lemma 6.2 that the corresponding families of symmetrized processes $\tilde{Y}_1^r$, $\tilde{Y}_2^n$ are tight at fixed times $t > 0$. Hence, Propositions 16.4 and 16.5 yield the stated integrability conditions for $E\eta$ adding up to (3), and there exist some limiting processes Y_1, Y_2, adding up to the integral terms in (2). The previous independence property clearly extends to $Y \perp\!\!\!\perp (X - Y)$, where $Y = Y_1 + Y_2$.

Since the jumps of X are already contained in Y and the continuous centering in Y_1 gives no new jumps, the only remaining jumps of $X - Y$ are the fixed jumps arising from the centering in Y_2. Thus, Proposition 19.2 yields $X - Y = G + b$, where G is continuous Gaussian with independent increments and b is non-random, rcll. □

Of special interest is the case where X is a semi-martingale:

Corollary 16.10 *(semi-martingales with independent increments, Jacod, OK) Let X be an rcll process with independent increments. Then*

(i) *$X = Y + b$, where Y is a semi-martingale and b is non-random rcll,*

(ii) *X is a semi-martingale iff b has locally finite variation,*

(iii) *when X is a semi-martingale,* (3) *holds with $E\eta'$ replaced by $E\eta$.*

Proof: (i) We need to show that all terms making up $X - b$ are semi-martingales. Then note that G and the first integral term are martingales, whereas the second integral term represents a process of isolated jumps.

(ii) Part (i) shows that X is a semi-martingale iff b has the same property, which holds by Corollary 20.21 below iff b has locally finite variation.

(iii) By the definition of η, the only discontinuities of b are the fixed jumps $E(\xi_t;\, |\xi_t| \le 1)$. Since b is a semi-martingale, its quadratic variation is locally finite, and we get

$$\begin{aligned}\int_{D_t}\int_{|x|\le 1} &|x|^2 \eta(ds\, dx)\\ &= \sum_{s\,\in D_t} E\Big\{|\xi_s|^2;\, |\xi_s| \le 1\Big\}\\ &= \sum_{s\,\in D_t} \Big(\mathrm{Var}\{\xi_t;\, |\xi_t|\le 1\} + |E\{\xi_t;\, |\xi_t| \le 1\}|^2\Big) < \infty. \qquad \Box\end{aligned}$$

By Proposition 11.5, a Lévy process X is Markov for the induced filtration $\mathcal{G} = (\mathcal{G}_t)$, with the translation-invariant transition kernels

$$\begin{aligned}\mu_t(x, B) &= \mu_t(0, B - x)\\ &= P\{X_t \in B - x\}, \quad t \ge 0.\end{aligned}$$

More generally, given any filtration $\mathcal{F}$, we say that a process X is Lévy with respect to $\mathcal{F}$ or simply *$\mathcal{F}$-Lévy*, if it is adapted to $\mathcal{F}$ and such that $(X_t - X_s) \perp\!\!\!\perp \mathcal{F}_s$ for all $s < t$. In particular, we may choose the filtration

$$\begin{aligned}\mathcal{F}_t &= \mathcal{G}_t \vee \mathcal{N}, \quad t \ge 0,\\ \mathcal{N} &= \sigma\{N \subset A;\, A \in \mathcal{A},\, PA = 0\},\end{aligned}$$

which is right-continuous by Corollary 9.26. Just as for Brownian motion in Theorem 14.11, we see that when X is $\mathcal{F}$-Lévy for a right-continuous, complete filtration $\mathcal{F}$, it is a *strong Markov process*, in the sense that the process $X' = \theta_\tau X - X_\tau$ satisfies $X \overset{d}{=} X' \perp\!\!\!\perp \mathcal{F}_\tau$ for every optional time $\tau < \infty$.

Turning to some symmetry conditions, we say that a process X on R_+ is *self-similar*, if for every $r > 0$ the scaled process X_{rt}, $t \ge 0$, has the same distribution as sX for a constant $s = h(r) > 0$. Excluding the trivial case of $E|X_t| \equiv 0$, we see that h satisfies the Cauchy equation $h(rs) = h(r)\, h(s)$. If

X is right-continuous, then h is continuous, and all solutions are of the form $h(r) = r^\alpha$ for some $\alpha \in \mathsf{R}$.

A Lévy process X in R is said to be *strictly stable* if it is self-similar, and *weakly stable* if it is self-similar apart from a centering, so that for every $r > 0$ the process X_{rt} has the same distribution as $sX_t + bt$ for some s and b. In either case, the symmetrized version $X - X'$ is strictly stable with the same parameter α, and so s is again of the form r^α. Since clearly $\alpha > 0$, we may introduce the *stability index* $p = \alpha^{-1}$, and say that X is strictly or weakly p*-stable*. The terminology carries over to random variables or vectors with distribution $\mathcal{L}(X_1)$.

Corollary 16.11 *(stable Lévy processes) Let X be a non-degenerate Lévy process in R with characteristics (a, b, ν). Then X is weakly p-stable for a $p > 0$ iff one of these conditions holds:*

(i) *$p = 2$ and $\nu = 0$,*

(ii) *$p \in (0, 2)$, $a = 0$, and $\nu(dx) = c_\pm |x|^{-p-1} dx$ on $\mathsf{R}_\pm$ for some $c_\pm \geq 0$.*

For subordinators, it is further equivalent that

(iii) *$p \in (0, 1)$ and $\nu(dx) = c\, x^{-p-1} dx$ on $(0, \infty)$ for a $c > 0$.*

In (i) we recognize a scaled Brownian motion with possible drift. An important case of (ii) is the symmetric, 1-stable Lévy process, known under suitable normalization as a *Cauchy process*.

Proof: Writing $S_r : x \mapsto rx$ for $r > 0$, we note that the processes $X(r^p t)$ and rX have characteristics $r^p(a, b, \nu)$ and $(r^2 a, rb, \nu \circ S_r^{-1})$, respectively. Since the latter are determined by the distributions, it follows that X is weakly p-stable iff $r^p a = r^2 a$ and $r^p \nu = \nu \circ S_r^{-1}$ for all $r > 0$. In particular, $a = 0$ when $p \neq 2$. Writing $F(x) = \nu[x, \infty)$ or $\nu(-\infty, -x]$, we also note that $r^p F(rx) = F(x)$ for all $r, x > 0$, so that $F(x) = x^{-p} F(1)$, which yields a density of the stated form. The condition $\int (x^2 \wedge 1)\, \nu(dx) < \infty$ implies $p \in (0, 2)$ when $\nu \neq 0$. When $X \geq 0$ we have the stronger requirement $\int (x \wedge 1)\, \nu(dx) < \infty$, so in this case $p < 1$. $\square$

If X is weakly p-stable with $p \neq 1$, it becomes strictly p-stable through a suitable centering. In particular, a weakly p-stable subordinator is strictly stable iff its drift component vanishes, in which case X is simply said to be *stable*. We show how stable subordinators may arise naturally even in the context of continuous processes. Given a Brownian motion B in R, we introduce the maximum process $M_t = \sup_{s \leq t} B_s$ with right-continuous inverse

$$\begin{aligned} T_r &= \inf\Big\{t \geq 0;\ M_t > r\Big\} \\ &= \inf\Big\{t \geq 0;\ B_t > r\Big\}, \quad r \geq 0. \end{aligned}$$

Theorem 16.12 *(first-passage times, Lévy) For a Brownian motion B, the process T is a $\frac{1}{2}$-stable subordinator with Lévy measure*

$$\nu(dx) = (2\pi)^{-1/2} x^{-3/2} dx, \quad x > 0.$$

Proof: By Lemma 9.6, the random times T_r are optional with respect to the right-continuous filtration $\mathcal{F}$ induced by B. By the strong Markov property of B, the process $\theta_r T - T_r$ is then independent of $\mathcal{F}_{T_r}$ and distributed as T. Since T is further adapted to the filtration $(\mathcal{F}_{T_r})$, it follows that T has stationary, independent increments and hence is a subordinator.

To see that T is $\frac{1}{2}$-stable, fix any $c > 0$, put $\tilde{B}_t = c^{-1}B(c^2\,t)$, and define $\tilde{T}_r = \inf\big\{t \geq 0;\, \tilde{B}_t > r\big\}$. Then

$$\begin{aligned} T_{cr} &= \inf\big\{t \geq 0;\, B_t > c\,r\big\} \\ &= c^2 \inf\big\{t \geq 0;\, \tilde{B}_t > r\big\} = c^2\,\tilde{T}_r. \end{aligned}$$

By Proposition 16.11, the Lévy measure of T has a density of the form $a\,x^{-3/2}$, $x > 0$, and it remains to identify a. Then note that the process

$$X_t = \exp\Big(uB_t - \tfrac{1}{2}\,u^2 t\Big), \quad t \geq 0,$$

is a martingale for any $u \in \mathsf{R}$. In particular, $EX_{\tau_r \wedge t} = 1$ for any $r, t \geq 0$, and since clearly $B_{\tau_r} = r$, we get by dominated convergence

$$E\exp\Big(-\tfrac{1}{2}\,u^2 T_r\Big) = e^{-ur}, \quad u, r \geq 0.$$

Taking $u = \sqrt{2}$ and comparing with Corollary 7.6, we obtain

$$\begin{aligned} \frac{\sqrt{2}}{a} &= \int_0^\infty (1 - e^{-x})\,x^{-3/2}dx \\ &= 2\int_0^\infty e^{-x}x^{-1/2}dx = 2\,\sqrt{\pi}, \end{aligned}$$

which shows that $a = (2\pi)^{-1/2}$. □

If we add a negative drift to a Brownian motion, the associated maximum process M becomes bounded, so that $T = M^{-1}$ terminates by a jump to infinity. Here it becomes useful to allow subordinators with infinite jumps. By an *extended subordinator* we mean a process of the form $X_t \equiv Y_t + \infty \cdot 1\{t \geq \zeta\}$ a.s., where Y is an ordinary subordinator and ζ is an independent, exponentially distributed random variable, so that X is obtained from Y by *exponential killing*. The representation in Theorem 16.3 remains valid in the extended case, except for a positive mass of ν at ∞.

The following characterization is needed in Chapter 29.

Lemma 16.13 *(extended subordinators) Let X be a non-decreasing, right-continuous process in $[0, \infty]$ with $X_0 = 0$ and induced filtration $\mathcal{F}$. Then X is an extended subordinator iff*

$$\mathcal{L}\Big(X_{s+t} - X_s\Big|\,\mathcal{F}_s\Big) = \mathcal{L}(X_t) \;\; \textit{a.s. on } \{X_s < \infty\}, \quad s, t > 0. \tag{5}$$

Proof: Writing $\zeta = \inf\{t;\, X_t = \infty\}$, we get from (5) the Cauchy equation

$$P\{\zeta > s + t\} = P\{\zeta > s\}P\{\zeta > t\}, \quad s, t \geq 0, \tag{6}$$

which implies that ζ is exponentially distributed with mean $m \in (0,\infty]$. Next define $\mu_t = \mathcal{L}(X_t \,|\, X_t < \infty)$, $t \geq 0$, and conclude from (5) and (6) that the μ_t form a semigroup under convolution. By Theorem 11.4 there exists a corresponding process Y with stationary, independent increments. From the right-continuity of X, it follows that Y is continuous in probability. Hence, Y has a subordinator version. Now choose $\tilde{\zeta} \overset{d}{=} \zeta$ with $\tilde{\zeta} \perp\!\!\!\perp Y$, and let $\tilde{X}$ denote the process Y killed at $\tilde{\zeta}$. Comparing with (5), we note that $\tilde{X} \overset{d}{=} X$. By Theorem 8.17 we may assume that even $X = \tilde{X}$ a.s., which means that X is an extended subordinator. The converse assertion is obvious. □

The weak convergence of infinitely divisible laws extends to a pathwise approximation of the corresponding Lévy processes.

Theorem 16.14 *(coupling of Lévy processes, Skorohod) For any Lévy processes X, $X^1, X^2, \ldots$ in R^d with $X_1^n \xrightarrow{d} X_1$, there exist some processes $\tilde{X}^n \overset{d}{=} X^n$ with*

$$\sup_{s\leq t} \left| \tilde{X}_s^n - X_s \right| \xrightarrow{P} 0, \quad t \geq 0.$$

For the proof, we first consider two special cases.

Lemma 16.15 *(compound Poisson case) Theorem 16.14 holds for compound Poisson processes X, $X^1, X^2, \ldots$ with characteristic measures ν, $\nu_1, \nu_2, \ldots$ satisfying $\nu_n \xrightarrow{w} \nu$.*

Proof: Adding positive masses at the origin, we may assume that ν and the ν_n have the same total mass, which may then be reduced to 1 through a suitable scaling. If $\xi_1, \xi_2, \ldots$ and $\xi_1^n, \xi_2^n, \ldots$ are associated i.i.d. sequences, we get $(\xi_1^n, \xi_2^n, \ldots) \xrightarrow{d} (\xi_1, \xi_2, \ldots)$ by Theorem 5.30, and by Theorem 5.31 we may strengthen this to a.s. convergence. Letting N be an independent, unit-rate Poisson process and defining $X_t = \sum_{j\leq N_t} \xi_j$ and $X_t^n = \sum_{j\leq N_t} \xi_j^n$, we obtain $(X^n - X)_t^* \to 0$ a.s. for all $t \geq 0$. □

Lemma 16.16 *(case of small jumps) Theorem 16.14 holds for Lévy processes X^n satisfying*

$$EX^n \equiv 0, \qquad 1 \geq (\Delta X^n)_1^* \xrightarrow{P} 0.$$

Proof: Since $(\Delta X^n)_1^* \xrightarrow{P} 0$, we may choose some constants $h_n \to 0$ with $m_n = h_n^{-1} \in \mathsf{N}$, such that $w(X^n, 1, h_n) \xrightarrow{P} 0$. By the stationarity of the increments, we get $w(X^n, t, h_n) \xrightarrow{P} 0$ for all $t \geq 0$. Further note that X is centered Gaussian by Theorem 7.7. As in Theorem 22.20 below, we may then form some processes $Y^n \overset{d}{=} (X^n_{[m_n t]h_n})$, such that $(Y^n - X)_t^* \xrightarrow{P} 0$ for all $t \geq 0$. By Corollary 8.18, we may further choose some processes $\tilde{X}^n \overset{d}{=} X^n$ with $Y^n \equiv \tilde{X}^n_{[m_n t]h_n}$ a.s. Letting $n \to \infty$ for fixed $t \geq 0$, we obtain

$$E\Big\{(\tilde{X}^n - X)_t^* \wedge 1\Big\}$$
$$\le E\Big\{(Y^n - X)_t^* \wedge 1\Big\} + E\Big\{w(X^n, t, h_n) \wedge 1\Big\} \to 0. \qquad \Box$$

Proof of Theorem 16.14: The asserted convergence is clearly equivalent to $\rho(\tilde{X}^n, X) \to 0$, where ρ denotes the metric

$$\rho(X, Y) = \int_0^\infty e^{-t} E\Big\{(X - Y)_t^* \wedge 1\Big\}\, dt.$$

For any $h > 0$, we may write

$$X = L^h + M^h + J^h,$$
$$X^n = L^{n,h} + M^{n,h} + J^{n,h},$$

with $L_t^h \equiv b^h t$ and $L_t^{n,h} \equiv b_n^h t$, where M^h and $M^{n,h}$ are martingales containing the Gaussian components and all centered jumps of size $\le h$, whereas the processes J^h and $J^{n,h}$ contain all the remaining jumps. Write G for the Gaussian component of X, and note that $\rho(M^h, G) \to 0$ as $h \to 0$ by Proposition 9.17.

For any $h > 0$ with $\nu\{|x| = h\} = 0$, Theorem 7.7 yields $b_n^h \to b^h$ and $\nu_n^h \xrightarrow{w} \nu^h$, where ν^h and ν_n^h denote the restrictions of ν and ν_n, respectively, to the set $\{|x| > h\}$. The same theorem gives $a_n^h \to a$ as $n \to \infty$ and then $h \to 0$, so under those conditions $M_1^{n,h} \xrightarrow{d} G_1$.

Now fix any $\varepsilon > 0$. By Lemma 16.16, we may choose some constants $h, r > 0$ and processes $\tilde{M}^{n,h} \overset{d}{=} M^{n,h}$, such that $\rho(M^h, G) \le \varepsilon$ and $\rho(\tilde{M}^{n,h}, G) \le \varepsilon$ for all $n > r$. Under the additional assumption $\nu\{|x| = h\} = 0$, Lemma 16.15 yields an $r' \ge r$ and some processes $\tilde{J}^{n,h} \overset{d}{=} J^{n,h}$ independent of $\tilde{M}^{n,h}$, such that $\rho(\tilde{J}^h, \tilde{J}^{n,h}) \le \varepsilon$ for all $n > r'$. We may finally choose $r'' \ge r'$ so large that $\rho(L^h, L^{n,h}) \le \varepsilon$ for all $n > r''$. Then the processes

$$\tilde{X}^n \equiv L^{n,h} + \tilde{M}^{n,h} + \tilde{J}^{n,h} \overset{d}{=} X^n, \quad n \in \mathsf{N},$$

satisfy $\rho(X, \tilde{X}^n) \le 4\,\varepsilon$ for all $n > r''$. $\Box$

Combining Theorem 16.14 with Corollary 7.9, we get a similar approximation of random walks, extending the result for Gaussian limits in Theorem 22.20. A slightly weaker result is obtained by different methods in Theorem 23.14.

Corollary 16.17 *(Lévy approximation of random walks) Consider a Lévy process X and some random walks $S^1, S^2, \ldots$ in R^d, such that $S_{k_n}^n \xrightarrow{d} X_1$ for some constants $k_n \to \infty$, and let N be an independent, unit-rate Poisson process. Then there exist some processes X^n with*

$$X^n \overset{d}{=} (S^n \circ N_{k_n t}), \qquad \sup_{s \le t} |X_s^n - X_s| \xrightarrow{P} 0, \quad t \ge 0.$$

Using the last result, we may extend the first two arcsine laws of Theorem 14.16 to symmetric Lévy processes, where symmetry means $-X \overset{d}{=} X$.

Theorem 16.18 *(arcsine laws) Let X be a symmetric Lévy process in R with $X_1 \neq 0$ a.s. Then these random variables are arcsine distributed:*

$$\tau_1 = \lambda\Big\{t \leq 1;\, X_t > 0\Big\},$$
$$\tau_2 = \inf\Big\{t \geq 0;\, X_t \vee X_{t-} = \sup_{s\leq 1} X_s\Big\}. \tag{7}$$

Proof: Introduce the random walk $S_k^n = X_{k/n}$, let N be an independent, unit-rate Poisson process, and define $X_t^n = S^n \circ N_{nt}$. Then Corollary 16.17 yields some processes $\tilde{X}^n \stackrel{d}{=} X^n$ with $(\tilde{X}^n - X)_1^* \xrightarrow{P} 0$. Define τ_1^n, τ_2^n as in (7) in terms of X^n, and conclude from Lemmas 22.12 and 7.16 that $\tau_i^n \xrightarrow{d} \tau_i$ for $i = 1, 2$.

Now define

$$\sigma_1^n = N_n^{-1} \sum_{k\leq N_n} 1\{S_k^n > 0\},$$
$$\sigma_2^n = N_n^{-1} \min\Big\{k;\, S_k^n = \max_{j\leq N_n} S_j^n\Big\}.$$

Since $t^{-1}N_t \to 1$ a.s. by the law of large numbers, we have $\sup_{t\leq 1} |n^{-1}N_{nt} - t| \to 0$ a.s., and so $\sigma_2^n - \tau_2^n \to 0$ a.s. Applying the same law to the sequence of holding times in N, we further note that $\sigma_1^n - \tau_1^n \xrightarrow{P} 0$. Hence, $\sigma_i^n \xrightarrow{d} \tau_i$ for $i = 1, 2$. Now $\sigma_1^n \stackrel{d}{=} \sigma_2^n$ by Corollary 27.8, and Theorem 22.11 yields $\sigma_2^n \xrightarrow{d} \sin^2\alpha$ where α is $U(0, 2\pi)$. Hence, $\tau_1 \stackrel{d}{=} \tau_2 \stackrel{d}{=} \sin^2\alpha$. □

In Theorem 15.15, we saw how any ql-continuous, simple point process on R_+ can be time-changed into a Poisson process. Similarly, we will see in Theorem 19.4 below how a continuous local martingale can be time-changed into a Brownian motion. Here we consider a third case of time-change reduction, involving integrals of p-stable Lévy processes. Since the general theory relies on some advanced stochastic calculus from Chapter 20, we consider only the elementary case of $p < 1$.

Proposition 16.19 *(time change of stable integrals, Rosiński & Woyczyński, OK) Let X be a strictly p-stable Lévy process with $p \in (0, 1)$, and let the process $V \geq 0$ be predictable and such that $A = V^p \cdot \lambda$ is a.s. finite but unbounded. Then*

$$(V \cdot X) \circ \tau \stackrel{d}{=} X, \qquad \tau_s = \inf\Big\{t;\, A_t > s\Big\}, \quad s \geq 0.$$

Proof: Define a point process ξ on $\mathsf{R}_+ \times (\mathsf{R} \setminus \{0\})$ by $\xi B = \sum_s 1_B(s, \Delta X_s)$, and recall from Corollary 16.2 and Proposition 16.11 that ξ is Poisson with intensity $\lambda \otimes \nu$, for some measures $\nu(dx) = c_\pm |x|^{-p-1} dx$ on $\mathsf{R}_\pm$. In particular, ξ has compensator $\hat{\xi} = \lambda \otimes \nu$. Now introduce on $\mathsf{R}_+ \times \mathsf{R}$ the predictable map $T_{s,x} = (A_s, xV_s)$. Since A is continuous, we have $\{A_s \leq t\} = \{s \leq \tau_t\}$ and $A_{\tau_t} = t$. Hence, Fubini's theorem yields for any $t, u > 0$

$$\begin{aligned}(\lambda \otimes \nu) \circ T^{-1}\Big([0,t] \times (u, \infty)\Big) \\ = (\lambda \otimes \nu)\Big\{(s,x);\, A_s \leq t,\, x\,V_s > u\Big\}\end{aligned}$$

$$\begin{aligned}
&= \int_0^{\tau_t} \nu\Big\{x;\, x\, V_s > u\Big\}\, ds \\
&= \nu(u,\infty) \int_0^{\tau_t} V_s^p\, ds \\
&= t\, \nu(u,\infty),
\end{aligned}$$

and similarly for the sets $[0,t] \times (-\infty, -u)$. Thus, $\hat{\xi} \circ T^{-1} = \hat{\xi} = \lambda \otimes \nu$ a.s., and so Theorem 15.15 gives $\xi \circ T^{-1} \overset{d}{=} \xi$. Finally, note that

$$\begin{aligned}
(V \cdot X)_{\tau_t} &= \int_0^{\tau_t +} \int x\, V_s\, \xi(ds\, dx) \\
&= \int_0^{\infty} \int x\, V_s 1\{A_s \le t\}\, \xi(ds\, dx) \\
&= \int_0^{t+} \int y\, (\xi \circ T^{-1})(dr\, dy),
\end{aligned}$$

where the process on the right has the same distribution as X. □

Anticipating our more detailed treatment in Chapter 20, we define a *semi-martingale* in R^d as an adapted process X with rcll paths, admitting a centering into a local martingale. Thus, subtracting all suitably centered jumps yields a continuous semi-martingale, in the sense of Chapter 18. Here the centering of small jumps amounts to a subtraction of the compensator $\hat{\eta}$ rather than the intensity $E\eta$ of the jump point process η, to ensure that the resulting process will be a local martingale. Note that such a centering depends on the choice of underlying filtration $\mathcal{F}$.

The set of *local characteristics* of X is given by the triple $(\hat{\eta}, [M], A)$, where M is a continuous local martingale and A is a continuous, predictable process of locally finite variation, such that X is the sum of the centered jumps and the continuous semi-martingale $M + A$. We say that X is *locally symmetric* if $\hat{\eta}$ is symmetric and $A = 0$. Two semi-martingales X, Y are said to be *$\mathcal{F}$-tangential*[5] if they have the same local characteristics with respect to $\mathcal{F}$. For filtrations $\mathcal{F}, \mathcal{G}$ on a common probability space, we say that $\mathcal{G}$ is a *standard extension* of $\mathcal{F}$ if

$$\mathcal{F}_t \subset \mathcal{G}_t \underset{\mathcal{F}_t}{\perp\!\!\!\perp} \mathcal{F}, \quad t \ge 0.$$

These are the minimal conditions ensuring that $\mathcal{G}$ will preserve all conditioning and adaptedness properties of $\mathcal{F}$.

Theorem 16.20 *(tangential existence) Let X be an $\mathcal{F}$-semi-martingale with local characteristics Y. Then there exists an $\tilde{\mathcal{F}}$-semi-martingale $\tilde{X} \perp\!\!\!\perp_Y \mathcal{F}$ with Y-conditionally independent increments, for a standard extension $\tilde{\mathcal{F}} \supset \mathcal{F}$, such that X and $\tilde{X}$ are $\tilde{\mathcal{F}}$-tangential.*

We will prove this only in the ql-continuous case where Y is continuous, in which case we may rely on some elementary constructions:

[5]This should not to be confused with the more elementary notion of *tangent processes*. For two processes to be tangential, they must be semi-martingales with respect to the *same filtration*, which typically requires us to extend of the original filtration without affecting the local characteristics, a highly non-trivial step.

Lemma 16.21 *(tangential constructions) Let M be a continuous local $\mathcal{F}$-martingale, and let ξ be an S-marked, $\mathcal{F}$-adapted point process with continuous $\mathcal{F}$-compensator η. Form a Cox process $\tilde{\xi} \perp\!\!\!\perp_\eta \mathcal{F}$ directed by η and a Brownian motion $B \perp\!\!\!\perp (\tilde{\xi}, \mathcal{F})$, put $\tilde{M} = B \circ [M]$, and let $\tilde{\mathcal{F}}$ be the filtration generated by $(\mathcal{F}, \tilde{\xi}, \tilde{M})$. Then*[6]

(i) *$\tilde{\mathcal{F}}$ is a standard extension of $\mathcal{F}$,*

(ii) *$\tilde{M}$ is a continuous local $\tilde{\mathcal{F}}$-martingale with $[\tilde{M}] = [M]$ a.s.,*

(iii) *ξ, $\tilde{\xi}$ have the same $\tilde{\mathcal{F}}$-compensator η,*

(iv) *η is both a (ξ, η)-compensator of ξ and a $(\tilde{\xi}, \eta)$-compensator of $\tilde{\xi}$.*

Proof: (i) Since $\tilde{\xi} \perp\!\!\!\perp_\eta \mathcal{F}$ and $B \perp\!\!\!\perp_{\tilde{\xi},\eta} \mathcal{F}$, Theorem 8.12 yields $(\tilde{\xi}, B) \perp\!\!\!\perp_\eta \mathcal{F}$, and so

$$(\tilde{\xi}^t, \tilde{M}^t) \underset{\eta,\,[M]}{\perp\!\!\!\perp} \mathcal{F}, \quad t \geq 0.$$

Using Theorem 8.9 and the definitions of $\tilde{\xi}$ and $\tilde{M}$, we further note that

$$(\tilde{\xi}^t, \tilde{M}^t) \underset{\eta^t,\,[M]^t}{\perp\!\!\!\perp} (\eta, [M]), \quad t \geq 0.$$

Combining those relations and using Theorem 8.12, we obtain

$$(\tilde{\xi}^t, \tilde{M}^t) \underset{\eta^t,\,[M]^t}{\perp\!\!\!\perp} \mathcal{F}, \quad t \geq 0,$$

which implies $\tilde{\mathcal{F}}_t \perp\!\!\!\perp_{\mathcal{F}_t} \mathcal{F}$ for all $t \geq 0$.

(ii) Since $B \perp\!\!\!\perp (\tilde{\xi}, \mathcal{F})$, we get

$$B \underset{[M],\,\tilde{M}^s}{\perp\!\!\!\perp} (\mathcal{F}, \tilde{\xi}, \tilde{M}^s), \quad s \geq 0,$$

and so

$$\theta_s \tilde{M} \underset{[M],\,\tilde{M}^s}{\perp\!\!\!\perp} \tilde{\mathcal{F}}_s, \quad s \geq 0.$$

Combining with the relation $\theta_s \tilde{M} \perp\!\!\!\perp_{[M]} \tilde{M}^s$ and using Theorem 8.12, we obtain

$$\theta_s \tilde{M} \underset{[M]}{\perp\!\!\!\perp} \tilde{\mathcal{F}}_s, \quad s \geq 0.$$

Localizing if necessary to ensure integrability, we get for any $s \leq t$ the desired martingale property

$$\begin{aligned} E\Big(\tilde{M}_t - \tilde{M}_s \,\Big|\, \tilde{\mathcal{F}}_s\Big) &= E\Big\{E\Big(\tilde{M}_t - \tilde{M}_s \,\Big|\, \tilde{\mathcal{F}}_s, [M]\Big) \,\Big|\, \tilde{\mathcal{F}}_s\Big\} \\ &= E\Big\{E\Big(\tilde{M}_t - \tilde{M}_s \,\Big|\, [M]\Big) \,\Big|\, \tilde{\mathcal{F}}_s\Big\} = 0, \end{aligned}$$

and the associated rate property

$$\begin{aligned} E\Big(\tilde{M}_t^2 - \tilde{M}_s^2 \,\Big|\, \tilde{\mathcal{F}}_s\Big) &= E\Big\{E\Big(\tilde{M}_t^2 - \tilde{M}_s^2 \,\Big|\, \tilde{\mathcal{F}}_s, [M]\Big) \Big|\, \tilde{\mathcal{F}}_s\Big\} \\ &= E\Big\{E\Big(\tilde{M}_t^2 - \tilde{M}_s^2 \,\Big|\, [M]\Big) \Big|\, \tilde{\mathcal{F}}_s\Big\} \\ &= E\Big\{E\Big([M]_t - [M]_s \,\Big|\, [M]\Big) \,\Big|\, \tilde{\mathcal{F}}_s\Big\}. \end{aligned}$$

[6] We may think of $\tilde{\xi}$ as a *Coxification* of ξ and of $\tilde{M}$ as a *Brownification* of M.

(iii) Property (i) shows that η remains an $\tilde{F}$-compensator of ξ. Next, the relation $\tilde{\xi} \perp\!\!\!\perp_\eta \mathcal{F}$ implies $\theta_t \tilde{\xi} \perp\!\!\!\perp_{\eta, \tilde{\xi}^t} \mathcal{F}_t$. Combining with the Cox property and using Theorem 8.12, we get $\theta_t \tilde{\xi} \perp\!\!\!\perp_\eta (\tilde{\xi}^t, \mathcal{F}_t)$. Invoking the tower property of conditional expectations and the Cox property of $\tilde{\xi}$, we obtain

$$\begin{aligned}
E(\theta_r \tilde{\xi} \mid \tilde{\mathcal{F}}_t) &= E\Big(\theta_t \tilde{\xi} \,\Big|\, \tilde{\xi}^t, \mathcal{F}_t\Big) \\
&= E\Big\{ E\Big(\theta_t \tilde{\xi} \,\Big|\, \tilde{\xi}^t, \eta, \mathcal{F}_t\Big) \Big|\, \tilde{\xi}^t, \mathcal{F}_t \Big\} \\
&= E\Big\{ E(\theta_t \tilde{\xi} \mid \eta) \,\Big|\, \tilde{\xi}^t, \mathcal{F}_t \Big\} \\
&= E\Big(\theta_t \eta \mid \tilde{\xi}^t, \mathcal{F}_t\Big) \\
&= E(\theta_t \eta \mid \tilde{\mathcal{F}}_t).
\end{aligned}$$

Since η remains $\tilde{\mathcal{F}}$-predictable, it is then an $\tilde{\mathcal{F}}$-compensator of $\tilde{\xi}$.

(iv) The martingale properties in (ii) extend to the filtrations generated by (ξ, η) and $\tilde{\xi}, \eta)$, respectively, by the tower property of conditional expectations. Since η is continuous and adapted to both filtrations, it is both (ξ, η)- and $(\tilde{\xi}, \eta)$-predictable. The assertions follow by combination of those properties. □

Proof of Theorem 16.20 for continuous η: Let ξ be the jump point process of X, and let η denote the $\mathcal{F}$-compensator of ξ. Further, let M be the continuous martingale component of X, and let A be the predictable drift component of X, for a suitable truncation function. Define $\tilde{\xi}$, $\tilde{M}$, $\tilde{\mathcal{F}}$ as in Lemma 16.21, and construct an associated semi-martingale $\tilde{X}$ by compensating the jumps given by $\tilde{\xi}$. Then $\tilde{X}$ has the same local characteristics $Y = ([M], \eta, A)$ as X, and so the two processes are $\tilde{\mathcal{F}}$-tangential. Since $\tilde{\xi} \perp\!\!\!\perp_\eta \mathcal{F}$ and $B \perp\!\!\!\perp (\tilde{\xi}, \mathcal{F})$, we have $\tilde{X} \perp\!\!\!\perp_Y \mathcal{F}$, and the independence properties of B show that $\tilde{X}$ has conditionally independent increments. □

Tangential processes have similar asymptotic properties at ∞. Here we state only the most basic result of this type. Say that the function $\varphi\colon \mathsf{R}_+ \to \mathsf{R}_+$ has *moderate growth*, if it is non-decreasing and continuous with $\varphi(0) = 0$, and there exists a constant $c > 0$ such that $\varphi(2x) \le c\,\varphi(x)$ for all $x > 0$. In that case, there exists a function $h > 0$ on $(0, \infty)$, such that $\varphi(cx) \le h(c)\,\varphi(x)$ for all $c, x > 0$. Basic examples include the power functions $\varphi(x) = |x|^p$ with $p > 0$, and the functions $\varphi(x) = x \wedge 1$ and $\varphi(x) = 1 - e^{-x}$.

Theorem 16.22 *(tangential comparison, Zinn, Hitchenko, OK) Let the processes X, Y be tangential and either increasing or locally symmetric. Then for any function $\varphi\colon \mathsf{R}_+ \to \mathsf{R}_+$ of moderate growth, we have*[7]

$$E\,\varphi(X^*) \asymp E\,\varphi(Y^*).$$

Our proof is based on some tail estimates:

[7]The domination constants are understood to depend only on φ.

Lemma 16.23 *(tail comparison) Let the processes X, Y be tangential and either increasing or locally symmetric. Then for any $c, x > 0$,*

(i) $P\Big\{(\Delta X)^* > x\Big\} \le 2\,P\Big\{(\Delta Y)^* > x\Big\}$,

(ii) $P\{X^* > x\} \le 3\,P\{Y^* > c\,x\} + 4\,c$.

Proof: (i) Let ξ, η be the jump point processes of X, Y. Fix any $x > 0$, and introduce the optional time

$$\tau = \inf\Big\{t > 0;\ |\Delta Y_t| > x\Big\}.$$

Since the set $(0, \tau]$ is predictable by Lemma 10.2, we get

$$\begin{aligned} P\Big\{(\Delta X)^* > x\Big\} &\le P\{\tau < \infty\} + E\,\xi\Big\{(0,\tau] \times [-x, x]^c\Big\} \\ &= P\{\tau < \infty\} + E\,\eta\Big\{(0,\tau] \times [-x, x]^c\Big\} \\ &= 2\,P\{\tau < \infty\} \\ &= 2\,P\Big\{(\Delta Y)^* > x\Big\}. \end{aligned}$$

(ii) Fix any $c, x > 0$. For increasing X, Y, form $\hat{X}$, $\hat{Y}$ by omitting all jumps greater than cx, which clearly preserves the tangential relation. If X, Y are instead locally symmetric, we form $\hat{X}$, $\hat{Y}$ by omitting all jumps of modulus $> 2cx$. By Lemma 20.5 below and its proof, $\hat{X}$, $\hat{Y}$ are then local L^2-martingales with jumps a.s. bounded by $4cx$. They also remain tangential, and the tangential relation carries over to the quadratic variation processes $[\hat{X}]$, $[\hat{Y}]$.

Now introduce the optional time

$$\tau = \inf\Big\{t > 0;\ |\hat{Y}_t| > c\,x\Big\}.$$

For increasing X, Y, we have

$$\begin{aligned} x\,P\{\hat{X}_\tau > x\} &\le E\hat{X}_\tau = E\hat{Y}_\tau \\ &= E\hat{Y}_{\tau-} + E\Delta\hat{Y}_\tau \\ &\le 3\,c\,x. \end{aligned}$$

If X, Y are instead locally symmetric, then using the Jensen and Bernstein–Lévy inequalities in 4.5 and 9.16, the integration by parts formula in Theorem 20.6 below, and the tangential properties and bounds, we get

$$\begin{aligned} \Big(x\,P\{\hat{X}^*_\tau > x\}\Big)^2 &\le (E|\hat{X}_\tau|)^2 \\ &\le E\hat{X}^2_\tau = E[\hat{X}]_\tau \\ &= E[\hat{Y}]_\tau = E\hat{Y}^2_\tau \\ &\le \Big(|\hat{Y}_{\tau-}| + |\Delta\hat{Y}_\tau|\Big)^2 \\ &\le (4\,c\,x)^2. \end{aligned}$$

Thus, in both cases $P\{\hat{X}^*_\tau > x\} \le 4c$.

Since $Y^* \ge \frac{1}{2}\,(\Delta Y)^*$, we have

$$\Big\{(\Delta X)^* \le 2\,c\,x\Big\} \subset \{X = \hat X\},$$
$$\{\tau < \infty\} \subset \{\hat Y^* > c\,x\} \subset \{Y^* > c\,x\}.$$

Combining with the previous tail estimate and (i), we obtain

$$\begin{aligned} P\{X^* > x\} &\le P\Big\{(\Delta X)^* > 2\,c\,x\Big\} + P\{X^* > x\} \\ &\le 2\,P\Big\{(\Delta Y)^* > 2\,c\,x\Big\} + P\{\tau < \infty\} + P\{\hat X^*_\tau > x\} \\ &\le 3\,P\{Y^* > c\,x\} + 4\,c. \end{aligned}$$

□

Proof of Theorem 16.22: For any $c, x > 0$, consider the optional times

$$\begin{aligned} \tau &= \inf\Big\{t > 0;\ |X_t| > x\Big\}, \\ \sigma &= \inf\Big\{t > 0;\ P\{(\theta_t Y)^* > c\,x \,|\, \mathcal{F}_t\} > c\Big\}. \end{aligned}$$

Since $\theta_\tau X$, $\theta_\tau Y$ remain conditionally tangential given $\mathcal{F}_\tau$, Lemma 16.23 (ii) yields a.s. on $\{\tau < \sigma\}$

$$P\Big\{(\theta_\tau X)^* > x \,\Big|\, \mathcal{F}_\tau\Big\} \le 3\,P\Big\{(\theta_\tau Y)^* > c\,x \,\Big|\, \mathcal{F}_\tau\Big\} + 4c \le 7c,$$

and since $\{\tau < \sigma\} \in \mathcal{F}_\tau$ by Lemma 9.1, we get

$$\begin{aligned} &P\Big\{X^* > 3\,x,\ (\Delta X)^* \le x,\ \sigma = \infty\Big\} \\ &\qquad \le P\Big\{(\theta_\tau X)^* > x,\ \tau < \sigma\Big\} \\ &\qquad = E\Big\{P\{(\theta_\tau X)^* > x \,|\, \mathcal{F}_\tau\};\ \tau < \sigma\Big\} \\ &\qquad \le 7c\,P\{\tau < \infty\} \\ &\qquad = 7c\,P\{X^* > x\}. \end{aligned}$$

Furthermore, Lemma 16.23 (i) yields

$$\begin{aligned} P\Big\{(\Delta X)^* > x\Big\} &\le 2\,P\Big\{(\Delta Y)^* > x\Big\} \\ &\le 2\,P\Big\{Y^* > \tfrac{1}{2}\,x\Big\}, \end{aligned}$$

and Theorem 9.16 gives

$$\begin{aligned} P\{\sigma < \infty\} &= P\Big\{\sup_t P\Big\{(\theta_t Y)^* > c\,x \,\Big|\, \mathcal{F}_t\Big\} > c\Big\} \\ &\le P\Big\{\sup_t P\Big(Y^* > \tfrac{1}{2}\,c\,x \,\Big|\, \mathcal{F}_t\Big) > c\Big\} \\ &\le c^{-1} P\Big\{Y^* > \tfrac{1}{2}\,c\,x\Big\}. \end{aligned}$$

Combining the last three estimates, we obtain

$$\begin{aligned} P\{X^* > 3\,x\} &\le P\Big\{X^* > 3\,x,\ (\Delta X)^* \le x,\ \sigma = \infty\Big\} \\ &\qquad + P\Big\{(\Delta X)^* > x\Big\} + P\{\sigma < \infty\} \\ &\le 7c\,P\{X^* > x\} + 2\,P\Big\{Y^* > \tfrac{1}{2}\,x\Big\} \\ &\qquad + c^{-1} P\Big\{Y^* > \tfrac{1}{2}\,c\,x\Big\}. \end{aligned}$$

Since φ is non-decreasing of moderate growth, so that $\varphi(rx) \le h(r)\,\varphi(x)$ for some function $h > 0$, we obtain

$$\begin{aligned}(h_3^{-1} - 7c)E\,\varphi(X^*) &\le E\,\varphi(X^*/3) - 7c\,E\,\varphi(X^*)\\ &\le 2\,E\,\varphi(2\,Y^*) + c^{-1}E\,\varphi(2\,Y^*/c)\\ &\le \big(2h_2 + c^{-1}h_{2/c}\big)E\,\varphi(Y^*).\end{aligned}$$

Now choose $c < (7h_3)^{-1}$ to get $E\,\varphi(X^*) \lesssim E\,\varphi(Y^*)$. □

Combining the last two theorems, we may simplify the tangential comparison. Say that X, Y are *weakly tangential* if their local characteristics have the same distribution. Here there is no mention of associated filtrations, and the two processes may even be defined on different probability spaces.

Corollary 16.24 *(extended comparison) Theorem 16.22 remains true when X and Y are weakly tangential.*

Proof: Let X, Y be semi-martingales on some probability spaces Ω, Ω' with filtrations $\mathcal{F}$, $\mathcal{G}$, and write A, B for the associated local characteristics. By Theorem 16.20 we may choose some processes $\tilde{X}$, $\tilde{Y}$ with conditionally independent increments, on suitable extensions of Ω, Ω', such that X, $\tilde{X}$ are $\tilde{\mathcal{F}}$-tangential while Y, $\tilde{Y}$ are $\tilde{\mathcal{G}}$-tangential, for some standard extensions $\tilde{\mathcal{F}}$ of $\mathcal{F}$ and $\tilde{\mathcal{G}}$ of $\mathcal{G}$. In particular, $\tilde{X}$, $\tilde{Y}$ have the same local characteristics A, B.

Now suppose that $A \stackrel{d}{=} B$, and conclude that also $\tilde{X} \stackrel{d}{=} \tilde{Y}$ since $\tilde{X}$, $\tilde{Y}$ have conditionally independent increments. Assuming X, Y to be either increasing or locally symmetric and letting φ have moderate growth, we get by Theorem 16.22

$$\begin{aligned}E\,\varphi(X^*) &\asymp E\,\varphi(\tilde{X}^*)\\ &= E\,\varphi(\tilde{Y}^*)\\ &\asymp E\,\varphi(Y^*).\end{aligned}$$

□

Exercises

1. Give an example of a process, whose increments are independent but not infinitely divisible. (*Hint:* Allow fixed discontinuities.)

2. Show that a non-decreasing process with independent increments has at most countably many fixed discontinuities, and that the associated component is independent of the remaining part. Then write the representation as in Theorem 16.3 for a suitably generalized Poisson process η.

3. Given a convolution semi-group of distributions μ_t on R^d, construct a Lévy process X with $\mathcal{L}(X_t) = \mu_t$ for all $t \ge 0$, starting from a suitable Poisson process and an independent Brownian motion. (*Hint:* Use Lemma 4.24 and Theorems 8.17 and 8.23.)

4. Let X be a process with stationary, independent increments and $X_0 = 0$. Show that X has a version with rcll paths. (*Hint:* Use the previous exercise.)

5. For a Lévy process X of effective dimension $d \geq 3$, show that $|X_t| \to \infty$ a.s. as $t \to \infty$. (*Hint:* Define $\tau = \inf\{t;\, |X_t| > 1\}$, and iterate to form a random walk (S_n). Show that the latter has the same effective dimension as X, and use Theorem 12.8.)

6. Let X be a real Lévy process, and fix any $p \in (0,2)$. Show that $t^{-1/p}X_t$ converges a.s. iff $E|X_1|^p < \infty$ and either $p \leq 1$ or $EX_1 = 0$. (*Hint:* Define a random walk (S_n) as before, show that S_1 satisfies the same moment condition as X_1, and apply Theorem 5.23.)

7. Show that a real Lévy process X is a subordinator iff $X_1 \geq 0$ a.s.

8. Let X be a real process as in Theorem 16.3. Show that if $X \geq 0$ a.s., then $G = 0$ a.s. and ν is restricted to $(0, \infty)$.

9. Let X be a weakly p-stable Lévy process. Show that when $p \neq 1$, we may choose $c \in \mathsf{R}$ such that the process $X_t - ct$ becomes strictly p-stable. Explain why such a centering fails for $p = 1$.

10. Show that a Lévy process is symmetric p-stable for a $p \in (0,2]$ iff $E\,e^{iuX_t} = e^{-ct|u|^p}$ for a constant $c \geq 0$. Similarly, show that a subordinator Y is strictly p-stable for a $p \in (0,1)$ iff $E\,e^{-uY_t} = e^{-ctu^p}$ for a constant $c \geq 0$. In each case, find the corresponding characteristics. (*Hint:* Derive a scaling property for the characteristic exponent.)

11. Let X be a symmetric, p-stable Lévy process and let T be a strictly q-stable subordinator, for some $p \in (0,2]$ and $q \in (0,1)$. Show that $Y = X \circ T$ is a symmetric pq-stable Lévy process. (*Hint:* Check that Y is again a Lévy process, and calculate $E\,e^{iuY_t}$.)

12. Give an example of two tangential ql-continuous, simple point processes ξ, η, such that η is Cox while ξ is not.

13. Give an example two tangential continuous martingales M, N, such that N has conditionally independent increments while M has not.

14. Let M, N be tangential continuous martingales. Show that $\|M^*\|_p \asymp \|N^*\|_p$ for all $p > 0$. (*Hint:* Use Theorem 18.7.)

15. Let M, N be tangential martingales. Show that $\|M^*\|_p \asymp \|N^*\|_p$ for all $p \geq 1$. (*Hint:* Note that $[M]$, $[N]$ are again tangential, and use Theorems 16.22 and 20.12.)

Chapter 17

Feller Processes and Semi-groups

Transition semi-groups, pseudo-Poisson processes, Feller properties, resolvents and generator, forward and backward equations, Yosida approximation, closure and cores, Lévy processes, Hille–Yosida theorem, positive-maximum principle, compactification, existence and regularization, strong Markov property, discontinuity sets, Dynkin's formula, characteristic operator, diffusions and elliptic operators, convergence of Feller processes, approximation of Markov chains, quasi-left continuity

As stressed before, Markov processes are among the most basic processes of modern probability. After studying several special cases in previous chapters, we now turn to a detailed study of the broad class of Feller processes. Those are Markov processes general enough to cover most applications of interest, yet restricted by some regularity conditions that allow for a reasonable flexibility, leading to a rich arsenal of basic tools and powerful properties.

The crucial new idea is to regard the transition kernels as operators T_t on an appropriate function space. The Chapman–Kolmogorov relation then turns into the semi-group property $T_s T_t = T_{s+t}$, which suggests a formal representation $T_t = e^{tA}$ in terms of a generator A. Under suitable regularity conditions—the so-called *Feller properties*—it is indeed possible to define a generator A describing the infinitesimal evolution of the underlying process X. Under further conditions, X will be shown to have continuous paths iff A extends an elliptic differential operator. In general, the powerful Hille–Yosida theorem provides precise conditions for the existence of a Feller process corresponding to a given operator A.

Using the basic regularity theorem for sub-martingales from Chapter 9, we show that every Feller process has a right-continuous version with left-hand limits. Given this fundamental result, it is straightforward to extend the strong Markov property to arbitrary Feller processes. We also explore some profound connections with martingale theory. Finally, we establish a general continuity theorem for Feller processes, and deduce a corresponding approximation of discrete-time Markov chains by diffusions and other continuous-time Markov processes, anticipating some weak convergence results from Chapter 23.

To clarify the connection between transition kernels and operators, let μ be an arbitrary probability kernel on a measurable space $(S, \mathcal{S})$. The associated *transition operator* T is given by

$$\begin{aligned} Tf(x) &= (Tf)(x) \\ &= \int \mu(x, dy)\, f(y), \quad x \in S, \end{aligned} \tag{1}$$

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_18

where $f: S \to \mathsf{R}$ is assumed to be measurable and either bounded or nonnegative. Approximating f by simple functions and using monotone convergence, we see that Tf is again a measurable function on S. We also note that T is a *positive contraction operator*, in the sense that $0 \le f \le 1$ implies $0 \le Tf \le 1$. A special role is played by the identity operator I, corresponding to the kernel $\mu(x,\cdot) \equiv \delta_x$. The importance of transition operators for the study of Markov processes is due to the following simple fact.

Lemma 17.1 *(semi-group property) For each $t \ge 0$, let μ_t be a probability kernel on S with associated transition operator T_t. Then for any $s, t \ge 0$,*

$$\mu_{s+t} = \mu_s \, \mu_t \quad \Leftrightarrow \quad T_{s+t} = T_s \, T_t.$$

Proof: For any $B \in \mathcal{S}$, we have $T_{s+t} 1_B(x) = \mu_{s+t}(x, B)$ and

$$\begin{aligned}(T_s \, T_t) 1_B(x) &= T_s (T_t 1_B)(x) \\ &= \int \mu_s(x, dy) \, (T_t 1_B)(y) \\ &= \int \mu_s(x, dy) \, \mu_t(y, B) \\ &= (\mu_s \, \mu_t)(x, B).\end{aligned}$$

Thus, the Chapman–Kolmogorov relation on the left is equivalent to $T_{s+t} 1_B = (T_s T_t) 1_B$ for any $B \in \mathcal{S}$, which extends to the semi-group property $T_{s+t} = T_s \, T_t$, by linearity and monotone convergence. □

By analogy with the situation for the Cauchy equation, we might hope to represent the semi-group in the form $T_t = e^{tA}$, $t \ge 0$, for a suitable *generator* A. For this formula to make sense, the operator A must be suitably bounded, so that the exponential function can be defined through a Taylor expansion. We consider a simple case where such a representation exists.

Proposition 17.2 *(pseudo-Poisson processes) Let (T_t) be the transition semi-group of a pure jump-type Markov process in S with bounded rate kernel α. Then $T_t = e^{tA}$, $t \ge 0$, where for bounded measurable functions $f: S \to \mathsf{R}$,*

$$Af(x) = \int \Big\{ f(y) - f(x) \Big\} \, \alpha(x, dy), \quad x \in S.$$

Proof: Since α is bounded, we may write $\alpha(x, B) \equiv c \, \mu(x, B \setminus \{x\})$ for a probability kernel μ and a constant $c \ge 0$. By Proposition 13.7, X is then a pseudo-Poisson process of the form $X = Y \circ N$, where Y is a discrete-time Markov chain with transition kernel μ, and N is an independent Poisson process with fixed rate c. Letting T be the transition operator associated with μ, we get for any $t \ge 0$ and f as stated

$$\begin{aligned}
T_t f(x) &= E_x f(X_t) \\
&= \sum_{n\geq 0} E_x\Big\{f(Y_n);\, N_t = n\Big\} \\
&= \sum_{n\geq 0} P\{N_t = n\}\, E_x f(Y_n) \\
&= \sum_{n\geq 0} e^{-ct}\frac{(ct)^n}{n!}\, T^n f(x) \\
&= e^{ct\,(T-I)} f(x).
\end{aligned}$$

Hence, $T_t = e^{tA}$ for all $t \geq 0$, where

$$\begin{aligned}
Af(x) &= c\,(T - I)f(x) \\
&= c\int\Big\{f(y) - f(x)\Big\}\,\mu(x, dy) \\
&= \int\Big\{f(y) - f(x)\Big\}\,\alpha(x, dy). \qquad \square
\end{aligned}$$

For the further analysis, we take S to be a locally compact, separable metric space, and write $C_0 = C_0(S)$ for the class of continuous functions $f : S \to \mathsf{R}$ with $f(x) \to 0$ as $x \to \infty$. We can make C_0 into a Banach space by introducing the norm $\|f\| = \sup_x |f(x)|$. A semi-group of positive contraction operators T_t on C_0 is called a *Feller semi-group*, if it has the additional regularity properties

(F$_1$) $T_t C_0 \subset C_0, \quad t \geq 0,$

(F$_2$) $T_t f(x) \to f(x)$ as $t \to 0, \quad f \in C_0,\ x \in S.$

In Theorem 17.6 we show that (F$_1$)−(F$_2$), together with the semi-group property, imply the *strong continuity*

(F$_3$) $T_t f \to f$ as $t \to 0, \quad f \in C_0.$

To clarify the probabilistic significance of those conditions, we assume for simplicity that S is compact, and also that (T_t) is *conservative*, in the sense that $T_t 1 = 1$ for all t. For every initial state x, we may introduce an associated Markov process X_t^x, $t \geq 0$, with transition operators T_t.

Lemma 17.3 *(Feller properties) Let (T_t) be a conservative transition semi-group on a compact metric space (S, ρ). Then the Feller properties* (F$_1$)−(F$_3$) *are equivalent to, respectively,*

(i) $X_t^x \xrightarrow{d} X_t^y$ *as* $x \to y, \quad t \geq 0,$

(ii) $X_t^x \xrightarrow{P} x$ *as* $t \to 0, \quad x \in S,$

(iii) $\sup_x E_x\Big\{\rho(X_s, X_t) \wedge 1\Big\} \to 0$ *as* $|s - t| \to 0.$

Proof: The first two equivalences are obvious, so we prove only the third one. Then let the sequence $f_1, f_2, \ldots$ be dense in $C = C_S$. By the compactness of S, we note that $x_n \to x$ in S iff $f_k(x_n) \to f_k(x)$ for each k. Thus, ρ is topologically equivalent to the metric

$$\rho'(x, y) = \sum_{k\geq 1} 2^{-k}\Big\{|f_k(x) - f_k(y)| \wedge 1\Big\}, \quad x, y \in S.$$

Since S is compact, the identity mapping on S is uniformly continuous with respect to ρ and ρ', and so we may assume that $\rho = \rho'$.

Next we note that, for any $f \in C,\ x \in S,$ and $t, h \geq 0$,

$$\begin{aligned} E_x\big\{f(X_t) - f(X_{t+h})\big\}^2 &= E_x\big(f^2 - 2f\,T_h f - T_h f^2\big)(X_t) \\ &\leq \big\|f^2 - 2f\,T_h f + T_h f^2\big\| \\ &\leq 2\|f\|\,\big\|f - T_h f\big\| + \big\|f^2 - T_h f^2\big\|. \end{aligned}$$

Assuming (F_3), we get $\sup_x E_x|f_k(X_s) - f_k(X_t)| \to 0$ as $s - t \to 0$ for fixed k, and so by dominated convergence $\sup_x E_x \rho(X_s, X_t) \to 0$. Conversely, the latter condition yields $T_h f_k \to f_k$ for each k, which implies (F_3). □

Our first aim is to construct the generator of a Feller semi-group (T_t) on C_0. Since in general there is no bounded linear operator A satisfying $T_t = e^{tA}$, we need to look for a suitable substitute. For motivation, note that if p is a real-valued function on R_+ with representation $p_t = e^{\,at}$, we can recover a from p by either a differentiation or an integration:

$$t^{-1}(p_t - 1) \to a \ \text{ as } \ t \to 0,$$
$$\int_0^\infty e^{-\lambda t} p_t\, dt = (\lambda - a)^{-1}, \quad \lambda > 0.$$

The latter formula suggests that we introduce, for every $\lambda > 0$, the *resolvent* or *potential* R_λ, defined as the Laplace transform

$$R_\lambda f = \int_0^\infty e^{-\lambda t}(T_t f)\, dt, \quad f \in C_0.$$

Note that the integral exists, since $T_t f(x)$ is bounded and right-continuous in $t \geq 0$ for fixed $x \in S$.

Theorem 17.4 *(resolvents and generator) Let (T_t) be a Feller semi-group on C_0 with resolvents R_λ, $\lambda > 0$. Then*

(i) *the λR_λ are injective contraction operators on C_0, and $\lambda R_\lambda \to I$ holds strongly as $\lambda \to \infty$,*

(ii) *the range $\mathcal{D} = R_\lambda C_0$ is independent of λ and dense in C_0,*

(iii) *there exists an operator A on C_0 with domain $\mathcal{D}$, such that $R_\lambda^{-1} = \lambda - A$ on $\mathcal{D}$ for every $\lambda > 0$,*

(iv) *A and T_t commute on $\mathcal{D}$ for every $t \geq 0$.*

Proof, (i)–(iii): If $f \in C_0$, then (F_1) yields $T_t f \in C_0$ for every t, and so by dominated convergence we have even $R_\lambda f \in C_0$. To prove the stated contraction property, we may write for any $f \in C_0$

$$\begin{aligned} \big\|\lambda R_\lambda f\big\| &\leq \lambda \int_0^\infty e^{-\lambda t}\, \|T_t f\|\, dt \\ &\leq \lambda\, \|f\| \int_0^\infty e^{-\lambda t} dt = \|f\|. \end{aligned}$$

A simple computation yields the *resolvent equation*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu, \quad \lambda, \mu > 0, \tag{2}$$

which shows that the operators R_λ commute and have a common range $\mathcal{D}$. If $f = R_1 g$ for some $g \in C_0$, we get by (2) as $\lambda \to \infty$

$$\begin{aligned} \left\| \lambda R_\lambda f - f \right\| &= \left\| (\lambda R_\lambda - I) R_1 g \right\| \\ &= \left\| (R_1 - I) R_\lambda g \right\| \\ &\le \lambda^{-1} \left\| R_1 - I \right\| \|g\| \to 0. \end{aligned}$$

The convergence extends by a simple approximation to the closure of $\mathcal{D}$.

Now introduce the one-point compactification $\hat{S} = S \cup \{\Delta\}$ of S, and extend any $f \in C_0$ to $\hat{C} = C(\hat{S})$ by putting $f(\Delta) = 0$. If $\bar{\mathcal{D}} \neq C_0$, then the Hahn–Banach theorem yields a bounded linear functional $\varphi \not\equiv 0$ on $\hat{C}$, such that $\varphi R_1 f = 0$ for all $f \in C_0$. By Riesz' representation Theorem 2.25, φ extends to a bounded, signed measure on $\hat{S}$. Letting $f \in C_0$ and using (F_2), we get by dominated convergence as $\lambda \to \infty$

$$\begin{aligned} 0 = \lambda \varphi R_\lambda f &= \int \varphi(dx) \int_0^\infty \lambda e^{-\lambda t}\, T_t f(x)\, dt \\ &= \int \varphi(dx) \int_0^\infty e^{-s}\, T_{s/\lambda} f(x)\, ds \to \varphi f, \end{aligned}$$

and so $\varphi \equiv 0$. The contradiction shows that $\mathcal{D}$ is dense in C_0.

To see that the operators R_λ are injective, let $f \in C_0$ with $R_{\lambda_0} f = 0$ for some $\lambda_0 > 0$. Then (2) yields $R_\lambda f = 0$ for every $\lambda > 0$, and since $\lambda R_\lambda f \to f$ as $\lambda \to \infty$, we get $f = 0$. Hence, the inverses R_λ^{-1} exist on $\mathcal{D}$. Multiplying (2) by R_λ^{-1} from the left and by R_μ^{-1} from the right, we get on $\mathcal{D}$ the relation $R_\mu^{-1} - R_\lambda^{-1} = \mu - \lambda$. Thus, the operator $A = \lambda - R_\lambda^{-1}$ on $\mathcal{D}$ is independent of λ.

(iv) Note that T_t and R_λ commute for any $t, \lambda > 0$, and write

$$\begin{aligned} T_t(\lambda - A) R_\lambda = T_t &= (\lambda - A) R_\lambda T_t \\ &= (\lambda - A)\, T_t R_\lambda. \end{aligned}$$

□

The operator A in Theorem 17.4 is called the *generator* of the semi-group (T_t). To emphasize the role of the domain $\mathcal{D}$, we often say that (T_t) has generator $(A, \mathcal{D})$. The term is justified by the following lemma.

Lemma 17.5 *(uniqueness) A Feller semi-group is uniquely determined by its generator.*

Proof: The operator A determines $R_\lambda = (\lambda - A)^{-1}$ for all $\lambda > 0$. By the uniqueness theorem for Laplace transforms, it then determines the measure $\mu_f(dt) = T_t f(x) dt$ on R_+ for all $f \in C_0$ and $x \in S$. Since the density $T_t f(x)$ is right-continuous in t for fixed x, the assertion follows. □

We now show that any Feller semi-group is strongly continuous, and derive general versions of Kolmogorov's forward and backward equations.

Theorem 17.6 *(strong continuity, forward and backward equations) Let (T_t) be a Feller semi-group with generator $(A, \mathcal{D})$. Then*

(i) *(T_t) is strongly continuous and satisfies*

$$T_t f - f = \int_0^t T_s A f\, ds, \quad f \in \mathcal{D},\ t \geq 0,$$

(ii) *$T_t f$ is differentiable at 0 iff $f \in \mathcal{D}$, in which case*

$$\frac{d}{dt}(T_t f) = T_t A f = A T_t f, \quad t \geq 0.$$

For the proof we use the *Yosida approximation*

$$A^\lambda = \lambda A R_\lambda = \lambda(\lambda R_\lambda - I), \quad \lambda > 0, \tag{3}$$

with associated semi-group $T_t^\lambda = e^{tA^\lambda}$, $t \geq 0$. Note that this is the transition semi-group of a pseudo-Poisson process with rate λ, based on the transition operator λR_λ.

Lemma 17.7 *(Yosida approximation) For any $f \in \mathcal{D}$,*

(i) $\left\| T_t f - T_t^\lambda f \right\| \leq t \left\| Af - A^\lambda f \right\|, \quad t, \lambda > 0,$

(ii) $A^\lambda f \to Af$ *as* $\lambda \to \infty$,

(iii) $T_t^\lambda f \to T_t f$ *as* $\lambda \to \infty$ *for each* $f \in C_0$, *uniformly for bounded* $t \geq 0$.

Proof: By Theorem 17.4, we have $A^\lambda f = \lambda R_\lambda A f \to Af$ for any $f \in \mathcal{D}$. For fixed $\lambda > 0$, we further note that $h^{-1}(T_h^\lambda - I) \to A^\lambda$ in the norm topology as $h \to 0$. Now for any commuting contraction operators B and C,

$$\begin{aligned} \left\| B^n f - C^n f \right\| &\leq \left\| B^{n-1} + B^{n-2} C + \cdots + C^{n-1} \right\| \left\| Bf - Cf \right\| \\ &\leq n \left\| Bf - Cf \right\|. \end{aligned}$$

Fixing any $f \in C_0$ and t, λ, $\mu > 0$, we hence obtain as $h = t/n \to 0$

$$\begin{aligned} \left\| T_t^\lambda f - T_t^\mu f \right\| &\leq n \left\| T_h^\lambda f - T_h^\mu f \right\| \\ &= t \left\| \frac{T_h^\lambda f - f}{h} - \frac{T_h^\mu f - f}{h} \right\| \\ &\to t \left\| A^\lambda f - A^\mu f \right\|. \end{aligned}$$

For $f \in \mathcal{D}$ it follows that $T_t^\lambda f$ is Cauchy convergent as $\lambda \to \infty$ for fixed t. Since $\mathcal{D}$ is dense in C_0, the same property holds for arbitrary $f \in C_0$. Denoting the limit by $\tilde{T}_t f$, we get in particular

$$\left\| T_t^\lambda f - \tilde{T}_t f \right\| \leq t \left\| A^\lambda f - Af \right\|, \quad f \in \mathcal{D},\ t \geq 0. \tag{4}$$

Thus, for each $f \in \mathcal{D}$ we have $T_t^\lambda f \to \tilde{T}_t f$ as $\lambda \to \infty$, uniformly for bounded t, which again extends to all $f \in C_0$.

To identify $\tilde{T}_t$, we may use the resolvent equation (2) to obtain, for any $f \in C_0$ and $\lambda, \mu > 0$,

$$\begin{aligned} \int_0^\infty e^{-\lambda t}\, T_t^\mu \mu R_\mu f\, dt &= (\lambda - A^\mu)^{-1} \mu R_\mu f \\ &= \frac{\mu}{\lambda + \mu} R_\nu f, \end{aligned} \tag{5}$$

where $\nu = \lambda\mu(\lambda + \mu)^{-1}$. As $\mu \to \infty$ we have $\nu \to \lambda$, and so $R_\nu f \to R_\lambda$. Furthermore,

$$\left\| T_t^\mu \mu R_\mu f - \tilde{T}_t f \right\| \le \left\| \mu R_\mu f - f \right\| + \left\| T_t^\mu f - \tilde{T}_t f \right\| \to 0,$$

and so by dominated convergence, (5) yields $\int e^{-\lambda t}\, \tilde{T}_t f dt = R_\lambda f$. Hence, the semi-groups (T_t) and $(\tilde{T}_t)$ have the same resolvent operators R_λ, and so they agree by Lemma 17.5. In particular, (i) then follows from (4). □

Proof of Theorem 17.6: The semi-group (T_t^λ) is clearly norm continuous in t for each $\lambda > 0$, and the strong continuity of (T_t) follows by Lemma 17.7 as $\lambda \to \infty$. We further have $h^{-1}(T_h^\lambda - I) \to A^\lambda$ as $h \downarrow 0$. Using the semi-group relation and continuity, we obtain more generally

$$\frac{d}{dt} T_t^\lambda = A^\lambda T_t^\lambda = T_t^\lambda A^\lambda, \quad t \ge 0,$$

which implies

$$T_t^\lambda f - f = \int_0^t T_s^\lambda A^\lambda f\, ds, \quad f \in C_0,\ t \ge 0. \tag{6}$$

If $f \in \mathcal{D}$, Lemma 17.7 yields as $\lambda \to \infty$

$$\left\| T_s^\lambda A^\lambda f - T_s A f \right\| \le \left\| A^\lambda f - A f \right\| + \left\| T_s^\lambda A f - T_s A f \right\| \to 0,$$

uniformly for bounded s, and so (i) follows from (6) as $\lambda \to \infty$. By the strong continuity of T_t, we may differentiate (i) to get the first relation in (ii). The second relation holds by Theorem 17.4.

Conversely, let $h^{-1}(T_h f - f) \to g$ for some functions $f, g \in C_0$. As $h \to 0$, we get

$$A R_\lambda f \leftarrow \frac{T_h - I}{h} R_\lambda f = R_\lambda \frac{T_h f - f}{h} \to R_\lambda g,$$

and so

$$\begin{aligned} f &= (\lambda - A) R_\lambda f \\ &= \lambda R_\lambda f - A R_\lambda f \\ &= R_\lambda(\lambda f - g) \in \mathcal{D}. \end{aligned}$$

□

For a given generator A, it is often hard to identify the domain $\mathcal{D}$, or the latter may simply be too large for convenient calculations. It is then useful to restrict A to a suitable sub-domain. An operator A with domain $\mathcal{D}$ on a Banach space B is said to be *closed*, if its *graph* $G = \{(f, Af);\ f \in \mathcal{D}\}$ is a closed subset of B^2. In general, we say that A is *closable*, if the closure $\bar{G}$ is the graph of a single-valued operator $\bar{A}$, called the *closure* of A. Note that A is closable iff the conditions $\mathcal{D} \ni f_n \to 0$ and $Af_n \to g$ imply $g = 0$.

When A is closed, a *core* for A is defined as a linear sub-space $D \subset \mathcal{D}$, such that the restriction $A|_D$ has closure A. Note that A is then uniquely determined by $A|_D$. We give conditions ensuring $D \subset \mathcal{D}$ to be a core, when A is the generator of a Feller semi-group (T_t) on C_0.

Lemma 17.8 *(closure and cores) Let $(A, \mathcal{D})$ be the generator of a Feller semi-group, and fix any $\lambda > 0$ and a sub-space $D \subset \mathcal{D}$. Then*

(i) *$(A, \mathcal{D})$ is closed,*

(ii) *D is a core for A $\Leftrightarrow$ $(\lambda - A)D$ is dense in C_0.*

Proof: (i) Let $f_1, f_2, \ldots \in \mathcal{D}$ with $f_n \to f$ and $Af_n \to g$. Then $(I - A)f_n \to f - g$. Since R_1 is bounded, we get $f_n \to R_1(f - g)$. Hence, $f = R_1(f - g) \in \mathcal{D}$, and $(I - A)f = f - g$ or $g = Af$. Thus, A is closed.

(ii) If D is a core for A, then for any $g \in C_0$ and $\lambda > 0$ there exist some $f_1, f_2, \ldots \in D$ with $f_n \to R_\lambda g$ and $Af_n \to AR_\lambda g$, and we get $(\lambda - A)f_n \to (\lambda - A)R_\lambda g = g$. Thus, $(\lambda - A)D$ is dense in C_0.

Conversely, let $(\lambda - A)D$ be dense in C_0. To see that D is a core, fix any $f \in \mathcal{D}$. By hypothesis, we may choose some $f_1, f_2, \ldots \in D$ with

$$\begin{aligned} g_n &\equiv (\lambda - A)f_n \\ &\to (\lambda - A)f \equiv g. \end{aligned}$$

Since R_λ is bounded, we obtain $f_n = R_\lambda g_n \to R_\lambda g = f$, and thus

$$\begin{aligned} Af_n &= \lambda f_n - g_n \\ &\to \lambda f - g = Af. \end{aligned}$$

□

A sub-space $D \subset C_0$ is said to be *invariant* under (T_t) if $T_t D \subset D$ for all $t \geq 0$. Note that for any subset $B \subset C_0$, the linear span of $\bigcup_t T_t B$ is an invariant sub-space of C_0.

Proposition 17.9 *(invariance and cores, Watanabe) Let $(A, \mathcal{D})$ be the generator of a Feller semi-group. Then every dense, invariant sub-space $D \subset \mathcal{D}$ is a core for A.*

Proof: By the strong continuity of (T_t), the operator R_1 can be approximated in the strong topology by some finite linear combinations $L_1, L_2, \ldots$ of the operators T_t. Now fix any $f \in D$, and define $g_n = L_n f$. Noting that A and L_n commute on D by Theorem 17.4, we get

$$\begin{aligned} (I - A)g_n &= (I - A)L_n f \\ &= L_n(I - A)f \\ &\to R_1(I - A)f = f. \end{aligned}$$

Since $g_n \in D$ which is dense in C_0, it follows that $(I - A)D$ is dense in C_0. Hence, D is a core by Lemma 17.8. $\Box$

The Lévy processes in R^d are archetypes of Feller processes, and we proceed to identify their generators[1]. Let C_0^∞ denote the class of all infinitely differentiable functions f on R^d, such that f and all its derivatives belong to $C_0 = C_0(\mathsf{R}^d)$.

Theorem 17.10 *(Lévy processes) Let T_t, $t \geq 0$, be transition operators of a Lévy process in R^d with characteristics (a, b, ν). Then*

(i) *(T_t) is a Feller semi-group,*

(ii) *C_0^∞ is a core for the generator A of (T_t),*

(iii) *for any $f \in C_0^\infty$ and $x \in \mathsf{R}^d$, we have*

$$\begin{aligned} Af(x) = {} & \tfrac{1}{2}\, a^{ij}\, \partial^2_{ij} f(x) + b^i\, \partial_i f(x) \\ & + \int \Big\{ f(x+y) - f(x) - y^i\, \partial_i f(x)\, 1\{|y| \leq 1\} \Big\}\, \nu(dy). \end{aligned}$$

In particular, a standard Brownian motion in R^d has generator $\frac{1}{2}\Delta$, and the uniform motion with velocity $b \in \mathsf{R}^d$ has generator $b\,\nabla$, both on the core C_0^∞, where Δ and ∇ denote the Laplace and gradient operators, respectively. The generator of the jump component has the same form as for the pseudo-Poisson processes in Proposition 17.2, apart from a compensation for small jumps by a linear drift term.

Proof of Theorem 17.10 (i), (iii): As $t \to 0$ we have $\mu_t^{*[t^{-1}]} \xrightarrow{w} \mu_1$, and so Corollary 7.9 yields $\mu_t/t \xrightarrow{v} \nu$ on $\overline{\mathsf{R}^d} \setminus \{0\}$ and

$$\begin{aligned} a_{t,h} &\equiv t^{-1} \int_{|x| \leq h} xx'\, \mu_t(dx) \to a_h, \\ b_{t,h} &\equiv t^{-1} \int_{|x| \leq h} x\, \mu_t(dx) \to b_h, \end{aligned} \tag{7}$$

for any $h > 0$ with $\nu\{|x| = h\} = 0$. Now fix any $f \in C_0^\infty$, and write

$$\begin{aligned} & t^{-1}\Big\{ T_t f(x) - f(x) \Big\} \\ & \quad = t^{-1} \int \Big\{ f(x+y) - f(x) \Big\}\, \mu_t(dy) \\ & \quad = t^{-1} \int_{|y| \leq h} \Big\{ f(x+y) - f(x) - y^i\, \partial_i f(x) - \tfrac{1}{2}\, y^i y^j\, \partial^2_{ij} f(x) \Big\}\, \mu_t(dy) \\ & \qquad + t^{-1} \int_{|y| > h} \Big\{ f(x+y) - f(x) \Big\}\, \mu_t(dy) + b^i_{t,h}\, \partial_i f(x) + \tfrac{1}{2}\, a^{ij}_{t,h}\, \partial^2_{ij} f(x). \end{aligned}$$

As $t \to 0$, the last three terms approach the expression in (iii), though with a_{ij} replaced by a^h_{ij} and with integration over $\{|x| > h\}$. To establish the required convergence, it remains to show that the first term on the right tends to zero

[1]Here and below, summation over repeated indices is understood.

as $h \to 0$, uniformly for small $t > 0$. Now this is clear from (7), since the integrand is of order $h|y|^2$ by Taylor's formula. By the uniform boundedness of the derivatives of f, the convergence is uniform in x. Thus, $C_0^\infty \subset \mathcal{D}$ by Theorem 17.6, and (iii) holds on C_0^∞.

(ii) Since C_0^∞ is dense in C_0, it suffices by Proposition 17.9 to show that it is also invariant under (T_t). Then note that, by dominated convergence, the differentiation operators commute with each T_t, and use condition (F_1). □

We proceed to characterize the linear operators A on C_0, whose closures $\bar{A}$ are generators of Feller semi-groups.

Theorem 17.11 *(generator criteria, Hille, Yosida) Let A be a linear operator on C_0 with domain $\mathcal{D}$. Then A is closable and its closure $\bar{A}$ is the generator of a Feller semi-group on C_0, iff these conditions hold:*

(i) *$\mathcal{D}$ is dense in C_0,*
(ii) *the range of $\lambda_0 - A$ is dense in C_0 for some $\lambda_0 > 0$,*
(iii) *$f \vee 0 \le f(x) \;\Rightarrow\; Af(x) \le 0$, for any $f \in \mathcal{D}$ and $x \in S$.*

Condition (iii) is known as the *positive-maximum principle.*

Proof: First assume that $\bar{A}$ is the generator of a Feller semi-group (T_t). Then (i) and (ii) hold by Theorem 17.4. To prove (iii), let $f \in \mathcal{D}$ and $x \in S$ with $f^+ = f \vee 0 \le f(x)$. Then

$$\begin{aligned} T_t f(x) &\le T_t f^+(x) \\ &\le \|T_t f^+\| \\ &\le \|f^+\| = f(x), \qquad t \ge 0, \end{aligned}$$

and so $h^{-1}(T_h f - f)(x) \le 0$. As $h \to 0$, we get $Af(x) \le 0$.

Conversely, let A satisfy (i)$-$(iii). For any $f \in \mathcal{D}$, choose an $x \in S$ with $|f(x)| = \|f\|$, and put $g = f \operatorname{sgn} f(x)$. Then $g \in \mathcal{D}$ with $g^+ \le g(x)$, and so (iii) yields $Ag(x) \le 0$. Thus, we get for any $\lambda > 0$

$$\begin{aligned} \big\|(\lambda - A)f\big\| &\ge \lambda g(x) - Ag(x) \\ &\ge \lambda g(x) = \lambda \|f\|. \end{aligned} \tag{8}$$

To show that A is closable, let $f_1, f_2, \ldots \in \mathcal{D}$ with $f_n \to 0$ and $Af_n \to g$. By (i) we may choose $g_1, g_2, \ldots \in \mathcal{D}$ with $g_n \to g$, and (8) yields

$$\big\|(\lambda - A)(g_m + \lambda f_n)\big\| \ge \lambda \big\|g_m + \lambda f_n\big\|, \quad m, n \in \mathsf{N}, \;\; \lambda > 0.$$

As $n \to \infty$, we get $\|(\lambda - A)g_m - \lambda g\| \ge \lambda \|g_m\|$. Dividing by λ and letting $\lambda \to \infty$, we obtain $\|g_m - g\| \ge \|g_m\|$, which gives $\|g\| = 0$ as $m \to \infty$. Thus, A is closable, and by (8) the closure $\bar{A}$ satisfies

$$\big\|(\lambda - \bar{A})f\big\| \ge \lambda \|f\|, \quad \lambda > 0, \;\; f \in \operatorname{dom}(\bar{A}). \tag{9}$$

Now let $\lambda_n \to \lambda > 0$ and $(\lambda_n - \bar{A})f_n \to g$ for some $f_1, f_2, \ldots \in \mathrm{dom}(\bar{A})$. By (9) the sequence (f_n) is then Cauchy, say with limit $f \in C_0$. The definition of $\bar{A}$ yields $(\lambda - \bar{A})f = g$, and so g belongs to the range of $\lambda - \bar{A}$. Letting Λ be the set of constants $\lambda > 0$ such that $\lambda - \bar{A}$ has range C_0, it follows in particular that Λ is closed. If it can be shown to be open as well, then (ii) gives $\Lambda = (0, \infty)$.

Then fix any $\lambda \in \Lambda$, and conclude from (9) that $\lambda - \bar{A}$ has a bounded inverse R_λ with norm $\|R_\lambda\| \le \lambda^{-1}$. For any $\mu > 0$ with $|\lambda - \mu|\|R_\lambda\| < 1$, we may form a bounded linear operator

$$\tilde{R}_\mu = \sum_{n \ge 0} (\lambda - \mu)^n R_\lambda^{n+1},$$

and note that

$$(\mu - \bar{A})\tilde{R}_\mu = (\lambda - \bar{A})\tilde{R}_\mu - (\lambda - \mu)\tilde{R}_\mu = I.$$

In particular $\mu \in \Lambda$, which shows that $\lambda \in \Lambda^o$.

To prove the resolvent equation (2), we start from the identity $(\lambda - \bar{A})R_\lambda = (\mu - \bar{A})R_\mu = I$. A simple rearrangement yields

$$(\lambda - \bar{A})(R_\lambda - R_\mu) = (\mu - \lambda)R_\mu,$$

and (2) follows as we multiply from the left by R_λ. In particular, (2) shows that R_λ and R_μ commute for any $\lambda, \mu > 0$.

Since $R_\lambda(\lambda - \bar{A}) = I$ on $\mathrm{dom}(\bar{A})$ and $\|R_\lambda\| \le \lambda^{-1}$, we have for any $f \in \mathrm{dom}(\bar{A})$ as $\lambda \to \infty$

$$\begin{aligned} \left\|\lambda R_\lambda f - f\right\| &= \left\|R_\lambda \bar{A} f\right\| \\ &\le \lambda^{-1}\|\bar{A}f\| \to 0. \end{aligned}$$

From (i) and the contractivity of λR_λ, it follows easily that $\lambda R_\lambda \to I$ in the strong topology. Now define A^λ as in (3), and let $T_t^\lambda = e^{tA^\lambda}$. Arguing as in the proof of Lemma 17.7, we get $T_t^\lambda f \to T_t f$ for each $f \in C_0$, uniformly for bounded t, where the T_t form a strongly continuous family of contraction operators on C_0, such that $\int e^{-\lambda t}\, T_t\, dt = R_\lambda$ for all $\lambda > 0$. To deduce the semi-group property, fix any $f \in C_0$ and $s, t \ge 0$, and note that as $\lambda \to \infty$,

$$\begin{aligned} \Big(T_{s+t} - T_s T_t\Big) f &= \Big(T_{s+t} - T_{s+t}^\lambda\Big) f + T_s^\lambda \Big(T_t^\lambda - T_t\Big) f \\ &\quad + \Big(T_s^\lambda - T_s\Big) T_t f \to 0. \end{aligned}$$

The positivity of the operators T_t will follow immediately, if we can show that R_λ is positive for each $\lambda > 0$. Then fix any function $g \ge 0$ in C_0, and put $f = R_\lambda g$, so that $g = (\lambda - \bar{A})f$. By the definition of $\bar{A}$, there exist some $f_1, f_2, \ldots \in \mathcal{D}$ with $f_n \to f$ and $Af_n \to \bar{A}f$. If $\inf_x f(x) < 0$, we have $\inf_x f_n(x) < 0$ for all sufficiently large n, and we may choose some $x_n \in S$ with $f_n(x_n) \le f_n \wedge 0$. By (iii) we have $Af_n(x_n) \ge 0$, and so

$$\begin{aligned} \inf\nolimits_x (\lambda - A) f_n(x) &\le (\lambda - A) f_n(x_n) \\ &\le \lambda f_n(x_n) \\ &= \lambda \inf\nolimits_x f_n(x). \end{aligned}$$

As $n \to \infty$, we get the contradiction

$$\begin{aligned} 0 &\le \inf_x g(x) \\ &= \inf_x (\lambda - \bar{A}) f(x) \\ &\le \lambda \inf_x f(x) < 0. \end{aligned}$$

To see that $\bar{A}$ is the generator of the semi-group (T_t), we note that the operators $\lambda - \bar{A}$ are inverses of the resolvent operators R_λ. □

The previous proof shows that, if an operator A on C_0 satisfies the positive maximum principle in (iii), then it is *dissipative*, in the sense that $\|(\lambda - A) f\| \ge \lambda \|f\|$ for all $f \in \mathrm{dom}(A)$ and $\lambda > 0$. This yields the following simple observation, needed later.

Lemma 17.12 *(maximality) Let $(A, \mathcal{D})$ be the generator of a Feller semi-group on C_0, and consider an extension of A to a linear operator $(A', \mathcal{D}')$ satisfying the positive-maximum principle. Then $(A, \mathcal{D}) = (A', \mathcal{D}')$.*

Proof: Fix any $f \in \mathcal{D}'$, and put $g = (I - A') f$. Since A' is dissipative and $(I - A) R_1 = I$ on C_0, we get

$$\begin{aligned} \big\| f - R_1 g \big\| &\le \big\| (I - A')(f - R_1 g) \big\| \\ &= \big\| g - (I - A) R_1 g \big\| = 0, \end{aligned}$$

and so $f = R_1 g \in \mathcal{D}$. □

We proceed to show how a nice Markov process can be associated with every Feller semi-group (T_t). For the corresponding transition kernels μ_t to have total mass 1, we need the operators T_t to be *conservative*, in the sense that $\sup_{f \le 1} T_t f(x) = 1$ for all $x \in S$. This can be achieved by a suitable extension.

Then introduce an auxiliary state[2] $\Delta \notin S$, and form the compactified space $\hat{S} = S \cup \{\Delta\}$, where Δ is chosen as the *point at infinity* when S is non-compact, and otherwise is taken to be isolated from S. Note that any function $f \in C_0$ has a continuous extension to $\hat{S}$, achieved by putting $f(\Delta) = 0$. The original semi-group on C_0 extends as follows to a conservative semi-group on $\hat{C} = C_{\hat{S}}$.

Lemma 17.13 *(compactification) Every Feller semi-group (T_t) on C_0 extends to a conservative Feller semi-group $(\hat{T}_t)$ on $\hat{C}$, given by*

$$\hat{T}_t f = f(\Delta) + T_t \big\{ f - f(\Delta) \big\}, \quad t \ge 0, \;\; f \in \hat{C}.$$

Proof: It is straightforward to verify that $(\hat{T}_t)$ is a strongly continuous semi-group on $\hat{C}$. To show that the operators $\hat{T}_t$ are positive, fix any $f \in \hat{C}$ with $f \ge 0$, and note that $g \equiv f(\Delta) - f \in C_0$ with $g \le f(\Delta)$. Hence,

[2] often called the *coffin state* or *cemetery*; this morbid terminology is well established

$$\begin{aligned} T_t g &\le T_t g^+ \le \|T_t g^+\| \\ &\le \|g^+\| \le f(\Delta), \end{aligned}$$

and so $\hat{T}_t f = f(\Delta) - T_t g \ge 0$. The contraction and conservation properties now follow from the fact that $\hat{T}_t 1 = 1$. □

Next we construct an associated semi-group of Markov transition kernels μ_t on $\hat{S}$, satisfying

$$T_t f(x) = \int f(y)\,\mu_t(x,dy), \quad f \in C_0. \tag{10}$$

Say that a state $x \in \hat{S}$ is *absorbing* for (μ_t), if $\mu_t(x,\{x\}) = 1$ for all $t \ge 0$.

Proposition 17.14 *(existence) For any Feller semi-group (T_t) on C_0, there exists a unique semi-group of Markov transition kernels μ_t on $\hat{S}$, satisfying* (10) *and such that Δ is absorbing for (μ_t).*

Proof: For fixed $x \in S$ and $t \ge 0$, the mapping $f \mapsto \hat{T}_t f(x)$ is a positive linear functional on $\hat{C}$ with norm 1. Hence, Riesz' representation Theorem 2.25 yields some probability measures $\mu_t(x,\cdot)$ on $\hat{S}$ satisfying

$$\hat{T}_t f(x) = \int f(y)\,\mu_t(x,dy), \quad f \in \hat{C},\ x \in \hat{S},\ t \ge 0. \tag{11}$$

The measurability on the right is clear by continuity. By a standard approximation followed by a monotone-class argument, we obtain the desired measurability of $\mu_t(x,B)$, for any $t \ge 0$ and Borel set $B \subset \hat{S}$. The Chapman–Kolmogorov relation holds on $\hat{S}$ by Lemma 17.1. Formula (10) is a special case of (11), which also yields

$$\begin{aligned} \int f(y)\,\mu_t(\Delta,dy) &= \hat{T}_t f(\Delta) \\ &= f(\Delta) = 0, \quad f \in C_0, \end{aligned}$$

showing that Δ is absorbing. The uniqueness of (μ_t) is clear from the last two properties. □

For any probability measure ν on $\hat{S}$, Theorem 11.4 yields a Markov process X^ν in $\hat{S}$ with initial distribution ν and transition kernels μ_t. Let P_ν be the distribution of X^ν with associated integration operator E_ν. When $\nu = \delta_x$, we may write P_x and E_x instead, for simplicity. We now extend Theorem 16.3 to a basic regularization theorem for Feller processes. For an rcll process X, we say that Δ is *absorbing* for $X^\pm$ if $X_t = \Delta$ or $X_{t-} = \Delta$ implies $X_u = \Delta$ for all $u \ge t$.

Theorem 17.15 *(regularization, Kinney) Let X be a Feller process in $\hat{S}$ with initial distribution ν. Then*

(i) *X has an rcll version $\tilde{X}$ in $\hat{S}$, such that Δ is absorbing for $\tilde{X}^\pm$,*

(ii) *if (T_t) is conservative and ν is restricted to S, we can choose $\tilde{X}$ to be rcll even in S.*

For the proof, we introduce an associated class of super-martingales, to which we may apply the regularity theorems in Chapter 9. Let C_0^+ be the class of non-negative functions in C_0.

Lemma 17.16 *(resolvents and super-martingales) For any $f \in C_0^+$, the process*

$$Y_t = e^{-t} R_1 f(X_t), \quad t \geq 0,$$

is a super-martingale under P_ν for every ν.

Proof: Letting $(\mathcal{G}_t)$ be the filtration induced by X, we get for any $t, h \geq 0$

$$\begin{aligned} E(Y_{t+h} | \mathcal{G}_t) &= E\Big\{ e^{-t-h} R_1 f(X_{t+h}) \Big| \mathcal{G}_t \Big\} \\ &= e^{-t-h} T_h R_1 f(X_t) \\ &= e^{-t-h} \int_0^\infty e^{-s} T_{s+h} f(X_t)\, ds \\ &= e^{-t} \int_h^\infty e^{-s} T_s f(X_t)\, ds \leq Y_t. \end{aligned}$$

□

Proof of Theorem 17.15: By Lemma 17.16 and Theorem 9.28, the process $f(X_t)$ has a.s. right- and left-hand limits along Q_+ for any $f \in \mathcal{D} \equiv \mathrm{dom}(A)$. Since $\mathcal{D}$ is dense in C_0, the same property holds for every $f \in C_0$. By the separability of C_0, we can choose the exceptional null set N to be independent of f. If $x_1, x_2, \ldots \in \hat{S}$ are such that $f(x_n)$ converges for every $f \in C_0$, then by compactness of $\hat{S}$ the sequence x_n converges in the topology of $\hat{S}$. Thus, on N^c, the process X has right- and left-hand limits $X_{t\pm}$ along Q_+, and on N we may redefine X to be 0. Then clearly $\tilde{X}_t = X_{t+}$ is rcll. It remains to show that $\tilde{X}$ is a version of X, so that $X_{t+} = X_t$ a.s. for each $t \geq 0$. This holds since $X_{t+h} \stackrel{P}{\to} X_t$ as $h \downarrow 0$, by Lemma 17.3 and dominated convergence.

For any $f \in C_0$ with $f > 0$ on S, the strong continuity of (T_t) yields even $R_1 f > 0$ on S. Applying Lemma 9.32 to the super-martingale $Y_t = e^{-t} R_1 f(\tilde{X}_t)$, we conclude that $X \equiv \Delta$ a.s. on the interval $[\zeta, \infty)$, where $\zeta = \inf\big\{t \geq 0;\, \Delta \in \{\tilde{X}_t, \tilde{X}_{t-}\}\big\}$. Discarding the exceptional null set, we can make this hold identically. If (T_t) is conservative and ν is restricted to S, then $\tilde{X}_t \in S$ a.s. for every $t \geq 0$. Thus, $\zeta > t$ a.s. for all t, and hence $\zeta = \infty$ a.s., which may again be taken to hold identically. Then $\tilde{X}_t$ and $\tilde{X}_{t-}$ take values in S, and the stated regularity properties remain valid in S. □

By the last theorem, we may choose Ω as the space of rcll functions in $\hat{S}$ such that Δ is absorbing, and let X be the canonical process on Ω. Processes with different initial distributions ν are then distinguished by their distributions P_ν on Ω. The process X is clearly Markov under each P_ν, with initial distribution ν and transition kernels μ_t, and it has the regularity properties of Theorem 17.15. In particular, $X \equiv \Delta$ on the interval $[\zeta, \infty)$, where ζ denotes the *terminal time*

$$\zeta = \inf\Big\{t \geq 0;\, X_t = \Delta \text{ or } X_{t-} = \Delta\Big\}.$$

Let $(\mathcal{F}_t)$ be the right-continuous filtration generated by X, and put $\mathcal{A} = \mathcal{F}_\infty = \bigvee_t \mathcal{F}_t$. As before, the *shift operators* θ_t on Ω are given by

$$(\theta_t\,\omega)_s = \omega_{s+t}, \quad s,t \geq 0.$$

The process X with associated distributions P_ν, filtration $\mathcal{F} = (\mathcal{F}_t)$, and shift operators θ_t is called the *canonical Feller process* with semi-group (T_t).

We now state a general version of the strong Markov property. The result extends the special versions obtained in Proposition 11.9 and Theorems 13.1 and 14.11. A further instant of this property is given in Theorem 32.11.

Theorem 17.17 *(strong Markov property, Dynkin & Yushkevich, Blumenthal) For any canonical Feller process X, initial distribution ν, optional time τ, and random variable $\xi \geq 0$, we have*

$$E_\nu\Big(\xi \circ \theta_\tau \Big| \mathcal{F}_\tau\Big) = E_{X_\tau}\xi \quad \text{a.s. } P_\nu \text{ on } \{\tau < \infty\}.$$

Proof: By Lemmas 8.3 and 9.1, we may take $\tau < \infty$. Let $\mathcal{G}$ be the filtration induced by X. Then the times $\tau_n = 2^{-n}[2^n\tau + 1]$ are $\mathcal{G}$-optional by Lemma 9.4, and Lemma 9.3 gives $\mathcal{F}_\tau \subset \mathcal{G}_{\tau_n}$ for all n. Thus, Proposition 11.9 yields

$$E_\nu\Big(\xi \circ \theta_{\tau_n}; A\Big) = E_\nu\Big(E_{X_{\tau_n}}\xi; A\Big), \quad A \in \mathcal{F}_\tau,\ n \in \mathbb{N}. \tag{12}$$

To extend this to τ, we first take $\xi = \prod_{k\leq m} f_k(X_{t_k})$ for some $f_1, \dots, f_m \in C_0$ and $t_1 < \cdots < t_m$. Then $\xi \circ \theta_{\tau_n} \to \xi \circ \theta_\tau$, by the right-continuity of X and continuity of $f_1, \dots, f_m$. Writing $h_k = t_k - t_{k-1}$ with $t_0 = 0$, and using Feller property (F_1) and the right-continuity of X, we obtain

$$\begin{aligned} E_{X_{\tau_n}}\xi &= T_{h_1}\Big\{f_1 T_{h_2} \cdots (f_{m-1} T_{h_m} f_m) \cdots\Big\}(X_{\tau_n}) \\ &\to T_{h_1}\Big\{f_1 T_{h_2} \cdots (f_{m-1} T_{h_m} f_m) \cdots\Big\}(X_\tau) = E_{X_\tau}\xi. \end{aligned}$$

Thus, (12) extends to τ by dominated convergence on both sides. Using standard approximation and monotone-class arguments, we may finally extend the result to arbitrary ξ. $\square$

For a simple application, we get a useful 0−1 law:

Corollary 17.18 *(Blumenthal's 0−1 law) For a canonical Feller process, we have*

$$P_x A = 0 \text{ or } 1, \quad x \in S,\ A \in \mathcal{F}_0.$$

Proof: Taking $\tau = 0$ in Theorem 17.17, we get for any $x \in S$ and $A \in \mathcal{F}_0$

$$\begin{aligned} 1_A &= P_x(A|\mathcal{F}_0) \\ &= P_{X_0}A = P_x A \quad \text{a.s. } P_x. \end{aligned}$$ $\square$

To appreciate the last result, recall that $\mathcal{F}_0 = \mathcal{F}_{0+}$. In particular, $P_x\{\tau = 0\} = 0$ or 1 for any state $x \in S$ and $\mathcal{F}$-optional time τ.

The strong Markov property is often used in the following extended form.

Corollary 17.19 *(optional projection) For any canonical Feller process X, non-decreasing, adapted process Y, and random variable $\xi \ge 0$, we have*

$$E_x \int_0^\infty (E_{X_t}\xi)\, dY_t = E_x \int_0^\infty (\xi \circ \theta_t)\, dY_t, \quad x \in S.$$

Proof: We may take $Y_0 = 0$. Consider the right-continuous inverse

$$\tau_s = \inf\Big\{t \ge 0;\, Y_t > s\Big\}, \quad s \ge 0,$$

and note that the τ_s are optional by Lemma 9.6. By Theorem 17.17, we have

$$\begin{aligned} E_x\Big(E_{X_{\tau_s}}\xi;\, \tau_s < \infty\Big) &= E_x\Big\{E_x(\xi \circ \theta_{\tau_s} | \mathcal{F}_{\tau_s});\, \tau_s < \infty\Big\} \\ &= E_x\Big(\xi \circ \theta_{\tau_s};\, \tau_s < \infty\Big). \end{aligned}$$

Since $\tau_s < \infty$ iff $s < Y_\infty$, we get by integration

$$E_x \int_0^{Y_\infty} (E_{X_{\tau_s}}\xi)\, ds = E_x \int_0^{Y_\infty} (\xi \circ \theta_{\tau_s})\, ds,$$

and the asserted formula follows by Lemma 1.24. □

Next we show that any martingale on the canonical space of a Feller process X is a.s. continuous outside the discontinuity set of X. For Brownian motion, this also follows from the integral representation in Theorem 19.11.

Theorem 17.20 *(discontinuity sets) Let X be a canonical Feller process with initial distribution ν, and let M be a local P_ν-martingale. Then*

$$\Big\{t > 0;\, \Delta M_t \ne 0\Big\} \subset \Big\{t > 0;\, X_{t-} \ne X_t\Big\} \quad a.s. \tag{13}$$

Proof (Chung & Walsh): By localization, we may take M to be uniformly integrable and hence of the form $M_t = E(\xi|\mathcal{F}_t)$ for some $\xi \in L^1$. Let $\mathcal{C}$ be the class of random variables $\xi \in L^1$, such that the the generated process M satisfies (13). Then $\mathcal{C}$ is a linear sub-space of L^1. It is further closed, since if $M_t^n = E(\xi_n|\mathcal{F}_t)$ with $\|\xi_n\|_1 \to 0$, then

$$P\Big\{\sup_t |M_t^n| > \varepsilon\Big\} \le \varepsilon^{-1} E|\xi_n| \to 0, \quad \varepsilon > 0,$$

and so $\sup_t |M_t^n| \xrightarrow{P} 0$.

Now let $\xi = \prod_{k \le n} f_k(X_{t_k})$ for some $f_1, \ldots, f_n \in C_0$ and $t_1 < \cdots < t_n$. Writing $h_k = t_k - t_{k-1}$, we note that

$$M_t = \prod_{k \le m} f_k(X_{t_k})\, T_{t_{m+1}-t}\, g_{m+1}(X_t), \quad t \in [t_m, t_{m+1}], \tag{14}$$

where

$$g_k = f_k\, T_{h_{k+1}}\Big\{f_{k+1} T_{h_{k+2}}(\cdots T_{h_n} f_n)\cdots\Big\}, \quad k = 1, \ldots, n,$$

subject to obvious conventions when $t < t_1$ and $t > t_n$. Since $T_t g(x)$ is jointly continuous in (t, x) for each $g \in C_0$, equation (14) defines a right-continuous version of M satisfying (13), and so $\xi \in \mathcal{C}$. By a simple approximation, $\mathcal{C}$ then contains all indicator functions of sets $\bigcap_{k\le n}\{X_{t_k} \in G_k\}$ with $G_1, \dots, G_n$ open. The result extends by a monotone-class argument to any X-measurable indicator function ξ, and a routine argument yields the final extension to L_1. $\square$

A basic role in the theory is played by the processes

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Af(X_s)\, ds, \quad t \ge 0, \;\; f \in \mathcal{D}.$$

Lemma 17.21 *(Dynkin's formula) For a Feller process X,*

(i) *the processes M^f are martingales under any initial distributions ν for X,*

(ii) *for any bounded optional time τ,*

$$E_x f(X_\tau) = f(x) + E_x \int_0^\tau Af(X_s)\, ds, \quad x \in S, \;\; f \in \mathcal{D}.$$

Proof: For any $t, h \ge 0$,

$$\begin{aligned} M_{t+h}^f - M_t^f &= f(X_{t+h}) - f(X_t) - \int_t^{t+h} Af(X_s)\, ds \\ &= M_h^f \circ \theta_t, \end{aligned}$$

and so by the Markov property at t and Theorem 17.6,

$$\begin{aligned} E_\nu\Big(M_{t+h}^f \,\Big|\, \mathcal{F}_t \Big) - M_t^f &= E_\nu\Big(M_h^f \circ \theta_t \,\Big|\, \mathcal{F}_t \Big) \\ &= E_{X_t} M_h^f = 0. \end{aligned}$$

Thus, M^f is a martingale, and (ii) follows by optional sampling. $\square$

To prepare for the next major result, we introduce the optional times

$$\tau_h = \inf\Big\{ t \ge 0;\; \rho(X_t, X_0) > h \Big\}, \quad h > 0,$$

where ρ denotes the metric in S. Note that a state x is *absorbing* iff $\tau_h = \infty$ a.s. P_x for every $h > 0$.

Lemma 17.22 *(escape times) When $x \in S$ is non-absorbing, we have $E_x \tau_h < \infty$ for $h > 0$ small enough.*

Proof: If x is not absorbing, then $\mu_t(x, B_x^\varepsilon) < p < 1$ for some $t, \varepsilon > 0$, where $B_x^\varepsilon = \{y;\, \rho(x,y) \le \varepsilon\}$. By Lemma 17.3 and Theorem 5.25, we may choose $h \in (0, \varepsilon]$ so small that

$$\mu_t(y, B_x^h) \le \mu_t(y, B_x^\varepsilon) \le p, \quad y \in B_x^h.$$

Then Proposition 11.2 yields

$$P_x\{\tau_h \ge nt\} \le P_x \bigcap_{k\le n} \Big\{ X_{kt} \in B_x^h \Big\} \le p^n, \quad n \in \mathsf{Z}_+,$$

and so by Lemma 4.4,

$$\begin{aligned} E_x \tau_h &= \int_0^\infty P\{\tau_h \ge s\} ds \\ &\le t \sum_{n\ge 0} P\{\tau_h \ge nt\} \\ &= t \sum_{n \ge 0} p^n = \frac{t}{1-p} < \infty. \end{aligned}$$

□

We turn to a probabilistic description of the generator and its domain. Say that A is *maximal* within a class of linear operators, if it extends every member of the class.

Theorem 17.23 *(characteristic operator, Dynkin) Let $(A, \mathcal{D})$ be the generator of a Feller process. Then*

(i) *for $f \in \mathcal{D}$ we have $Af(x) = 0$ when x is absorbing, and otherwise*

$$Af(x) = \lim_{h\to 0} \frac{E_x f(X_{\tau_h}) - f(x)}{E_x \tau_h}, \tag{15}$$

(ii) *A is the maximal operator on C_0 with those properties.*

Proof: (i) Fix any $f \in \mathcal{D}$. If x is absorbing, then $T_t f(x) = f(x)$ for all $t \ge 0$, and so $Af(x) = 0$. For non-absorbing x, Lemma 17.21 yields instead

$$E_x f(X_{\tau_h \wedge t}) - f(x) = E_x \int_0^{\tau_h \wedge t} Af(X_s)\, ds, \quad t, h > 0. \tag{16}$$

By Lemma 17.22, we have $E\tau_h < \infty$ for sufficiently small $h > 0$, and so (16) extends to $t = \infty$ by dominated convergence. Relation (15) now follows from the continuity of Af, along with the fact that $\rho(X_s, x) \le h$ for all $s < \tau_h$.

(ii) Since the positive-maximum principle holds for any extension of A with the stated properties, the assertion follows by Lemma 17.12. □

In the special case of $S = \mathsf{R}^d$, let $\hat{C}^\infty$ be the class of infinitely differentiable functions on R^d with bounded support. An operator $(A, \mathcal{D})$ with $\mathcal{D} \supset \hat{C}^\infty$ is said to be *local* on $\hat{C}^\infty$, if $Af(x) = 0$ when f vanishes in a neighborhood of x. For a local generator A on $\hat{C}^\infty$, the positive-maximum principle yields a *local positive-maximum principle*, asserting that if $f \in \hat{C}^\infty$ has a local maximum ≥ 0 at a point x, then $Af(x) \le 0$.

We give a basic relationship between diffusion processes and elliptic differential operators. This connection is further explored in Chapters 21, 32, and 34–35.

Theorem 17.24 *(Feller diffusions and elliptic operators, Dynkin) Let $(A, \mathcal{D})$ be the generator of a Feller process X in R^d with $\hat{C}^\infty \subset \mathcal{D}$. Then these conditions are equivalent:*

(i) *X is continuous on $[0, \zeta)$, a.s. P_ν for every ν,*

(ii) *A is local on $\hat{C}^\infty$.*

In this case, there exist some functions a^{ij}, b^i, $c \in C_{\mathsf{R}^d}$ with $c \geq 0$ and (a^{ij}) symmetric, non-negative definite, such that for any $f \in \hat{C}^\infty$ and $x \in \mathsf{R}_+$,

$$Af(x) = \tfrac{1}{2}\, a^{ij}(x)\, \partial^2_{ij} f(x) + b^i(x)\, \partial_i f(x) - c(x) f(x). \tag{17}$$

Here the matrix (a_{ij}) gives the diffusion rates of X, the vector (b^i) represents the drift rate of X, and c is the rate of *killing*[3]. For semi-groups of this type, we may choose Ω as the set of paths that are continuous on $[0, \zeta)$. The Markov process X is then referred to as a *canonical Feller diffusion.*

Proof: (i) $\Rightarrow$ (ii): Use Theorem 17.23.

(ii) $\Rightarrow$ (i): Assume (ii). Fix any $x \in \mathsf{R}^d$ and $0 < h < m$, and choose an $f \in \hat{C}^\infty$ with $f \geq 0$ and support $\{y;\, h \leq |y - x| \leq m\}$. Then $Af(y) = 0$ for all $y \in B^h_x$, and so Lemma 17.21 shows that $f(X_{t \wedge \tau_h})$ is a martingale under P_x. By dominated convergence, we have $E_x f(X_{\tau_h}) = 0$, and since m was arbitrary,

$$P_x\Big\{|X_{\tau_h} - x| \leq h \;\text{ or }\; X_{\tau_h} = \Delta\Big\} = 1, \quad x \in \mathsf{R}^d,\;\; h > 0.$$

By the Markov property at fixed times, we get for any initial distribution ν

$$P_\nu \bigcap_{t \in \mathsf{Q}_+} \theta_t^{-1} \Big\{|X_{\tau_h} - X_0| \leq h \;\text{ or }\; X_{\tau_h} = \Delta\Big\} = 1, \quad h > 0,$$

which implies

$$P_\nu\Big\{ \sup_{t < \zeta} |\Delta X_t| \leq h \Big\} = 1, \quad h > 0,$$

proving (i).

Now fix any $x \in \mathsf{R}^d$, and choose $f_0^x, f_i^x, f_{ij}^x \in \hat{C}^\infty$, such that for any y in a neighborhood of x,

$$f_0^x(y) = 1, \quad f_i^x(y) = y_i - x_i,$$
$$f_{ij}^x(y) = (y_i - x_i)(y_j - x_j).$$

Putting

$$c(x) = -Af_0^x(x), \qquad b_i(x) = Af_i^x(x), \qquad a_{ij}(x) = Af_{ij}^x(x),$$

we see that (17) holds locally when $f \in \hat{C}^\infty$ agrees near x with a second-degree polynomial. In particular, choosing $f_0(y) = 1$, $f_i(y) = y_i$, and $f_{ij}(y) = y_i\, y_j$ for y near x, we obtain

$$\begin{aligned} Af_0(x) &= -c(x), \\ Af_i(x) &= b_i(x) - x_i\, c(x), \\ Af_{ij}(x) &= a_{ij}(x) + x_i\, b_j(x) + x_j\, b_i(x) - x_i x_j\, c(x). \end{aligned}$$

This shows that c, b_i, and $a_{ij} = a_{ji}$ are continuous.

Applying the local positive-maximum principle to f_0^x, we get $c(x) \geq 0$. By the same principle applied to the function

$$f = -\left(u^i f_i^x\right)^2 = -u^i u^j f_{ij}^x,$$

[3]Instantaneous transfer to the 'cemetery' Δ; this morbid terminology is again standard.

we obtain $u^i u^j a_{ij}(x) \geq 0$, which shows that (a_{ij}) is non-negative definite. Finally, let $f \in \hat{C}^\infty$ be arbitrary with a second-order Taylor expansion $\tilde{f}$ around x. Then each of the functions

$$g_\pm^\varepsilon(y) = \pm\left\{f(y) - \tilde{f}(y)\right\} - \varepsilon|x-y|^2, \quad \varepsilon > 0,$$

has a local maximum 0 at x, and so

$$Ag_\pm^\varepsilon(x) = \pm\left\{Af(x) - A\tilde{f}(x)\right\} - \varepsilon \textstyle\sum_i a_{ii}(x) \leq 0, \quad \varepsilon > 0.$$

As $\varepsilon \to 0$ we get $Af(x) = A\tilde{f}(x)$, showing that (17) is generally true. □

We turn to a basic limit theorem for Feller processes, extending some results for Lévy processes in Theorems 7.7 and 16.14.

Theorem 17.25 *(convergence, Trotter, Sova, Kurtz, Mackevičius) Let X, X^1, $X^2, \ldots$ be Feller processes in S with semi-groups (T_t), $(T_{n,t})$ and generators $(A, \mathcal{D})$, $(A_n, \mathcal{D}_n)$, respectively, and fix a core D for A. Then these conditions are equivalent:*

(i) *for any $f \in D$, there exist some $f_n \in \mathcal{D}_n$ with $f_n \to f$ and $A_n f_n \to Af$,*
(ii) *$T_{n,t} \to T_t$ strongly for each $t > 0$,*
(iii) *$T_{n,t} f \to T_t f$ for every $f \in C_0$, uniformly for bounded $t > 0$,*
(iv) *$X_0^n \xrightarrow{d} X_0$ in S $\Rightarrow$ $X^n \xrightarrow{sd} X$ in $D_{\mathbb{R}_+, \hat{S}}$.*

Our proof is based on two lemmas, the former extending Lemma 17.7.

Lemma 17.26 *(norm inequality) Let (T_t), (T'_t) be Feller semi-groups with generators $(A, \mathcal{D})$, $(A', \mathcal{D}')$, respectively, where A' is bounded. Then*

$$\left\|T_t f - T'_t f\right\| \leq \int_0^t \left\|(A - A')\, T_s f\right\| ds, \quad f \in \mathcal{D}, \ t \geq 0. \tag{18}$$

Proof: Fix any $f \in \mathcal{D}$ and $t > 0$. Since (T'_s) is norm continuous, Theorem 17.6 yields

$$\frac{\partial}{\partial s}(T'_{t-s} T_s f) = T'_{t-s}(A - A')T_s f, \quad 0 \leq s \leq t.$$

Here the right-hand side is continuous in s, by the strong continuity of (T_s), the boundedness of A', the commutativity of A and T_s, and the norm continuity of (T'_s). Hence,

$$\begin{aligned} T_t f - T'_t f &= \int_0^t \frac{\partial}{\partial s}(T'_{t-s} T_s f)\, ds \\ &= \int_0^t T'_{t-s}(A - A')T_s f\, ds, \end{aligned}$$

and (18) follows by the contractivity of T'_{t-s}. □

We may next establish a continuity property for the Yosida approximations A^λ, A_n^λ of the generators A, A_n, respectively.

Lemma 17.27 *(continuity of Yosida approximation) Let* $(A,\mathcal{D})$, $(A_n,\mathcal{D}_n)$ *be generators of some Feller semi-groups, satisfying* (i) *of Theorem* 17.25*. Then* $A_n^\lambda \to A^\lambda$ *strongly for every* $\lambda > 0$.

Proof: By Lemma 17.8, it is enough to show that $A_n^\lambda f \to A^\lambda f$ for every $f \in (\lambda - A)D$. Then define $g \equiv R^\lambda f \in D$. By (i) we may choose some $g_n \in \mathcal{D}_n$ with $g_n \to g$ and $A_n g_n \to Ag$. Then

$$\begin{aligned} f_n &\equiv (\lambda - A_n)g_n \\ &\to (\lambda - A)g = f, \end{aligned}$$

and so

$$\begin{aligned} \left\|A_n^\lambda f - A^\lambda f\right\| &= \lambda^2 \left\|R_n^\lambda f - R^\lambda f\right\| \\ &\le \lambda^2 \left\|R_n^\lambda (f - f_n)\right\| + \lambda^2 \left\|R_n^\lambda f_n - R^\lambda f\right\| \\ &\le \lambda \|f - f_n\| + \lambda^2 \|g_n - g\| \to 0. \end{aligned}$$

□

Proof of Theorem 17.25, (i) ⇒ (iii): Assume (i). Since D is dense in C_0, we may take $f \in D$. Then choose some functions f_n as in (i), and conclude from Lemmas 17.7 and 17.26 that, for any $n \in \mathsf{N}$ and $t, \lambda > 0$,

$$\begin{aligned} \left\|T_{n,t} f - T_t f\right\| &\le \left\|T_{n,t}(f - f_n)\right\| + \left\|(T_{n,t} - T_{n,t}^\lambda) f_n\right\| + \left\|T_{n,t}^\lambda (f_n - f)\right\| \\ &\quad + \left\|(T_{n,t}^\lambda - T_t^\lambda) f\right\| + \left\|(T_t^\lambda - T_t) f\right\| \\ &\le 2\,\|f_n - f\| + t\left\|(A^\lambda - A) f\right\| + t\left\|(A_n - A_n^\lambda) f_n\right\| \\ &\quad + \int_0^t \left\|(A_n^\lambda - A^\lambda) T_s^\lambda f\right\| ds. \end{aligned} \tag{19}$$

By Lemma 17.27 and dominated convergence, the last term tends to zero as $n \to \infty$. For the third term on the right, we get

$$\begin{aligned} \left\|(A_n - A_n^\lambda) f_n\right\| &\le \left\|A_n f_n - Af\right\| + \left\|(A - A^\lambda) f\right\| \\ &\quad + \left\|(A^\lambda - A_n^\lambda) f\right\| + \left\|A_n^\lambda (f - f_n)\right\|, \end{aligned}$$

which tends to $\|(A - A^\lambda) f\|$ by the same lemma. Hence, by (19),

$$\limsup_{n\to\infty} \sup_{t\le u} \left\|T_{n,t} f - T_t f\right\| \le 2u \left\|(A^\lambda - A) f\right\|, \quad u, \lambda > 0,$$

and the desired convergence follows by Lemma 17.7, as we let $\lambda \to \infty$.

(iii) ⇒ (ii): Obvious.

(ii) ⇒ (i): Assume (ii), fix any $f \in D$ and $\lambda > 0$, and define $g = (\lambda - A) f$ and $f_n = R_n^\lambda g$. By dominated convergence, $f_n \to R^\lambda g = f$. Since $(\lambda - A_n) f_n = g = (\lambda - A) f$, we also note that $A_n f_n \to Af$, proving (i).

(iv) ⇒ (ii): Assume (iv). We may take S to be compact and the semi-groups (T_t) and $(T_{n,t})$ to be conservative. We need to show that, for any $f \in C$ and

$t > 0$, we have $T_t^n f(x_n) \to T_t f(x)$ whenever $x_n \to x$ in S. Then let $X_0 = x$ and $X_0^n = x_n$. By Lemma 17.3, the process X is a.s. continuous at t. Thus, (iv) yields $X_t^n \xrightarrow{d} X_t$, and the desired convergence follows.

(i) – (iii) ⇒ (iv): Assume (i) – (iii), and let $X_0^n \xrightarrow{d} X_0$. To obtain $X^n \xrightarrow{fd} X$, it is enough to show that as $n \to \infty$, for any $f_0, \ldots, f_m \in C$ and $0 = t_0 < t_1 \cdots t_m$,

$$E \prod_{k \le m} f_k(X_{t_k}^n) \to E \prod_{k \le m} f_k(X_{t_k}). \tag{20}$$

This holds by hypothesis when $m = 0$. Proceeding by induction, we may use the Markov property to rewrite (20) in the form

$$E \prod_{k < m} f_k(X_{t_k}^n) \cdot T_{h_m}^n f_m(X_{t_{m-1}}^n) \to E \prod_{k < m} f_k(X_{t_k}) \cdot T_{h_m} f_m(X_{t_{m-1}}), \tag{21}$$

where $h_m = t_m - t_{m-1}$. Since (ii) implies $T_{h_m}^n f_m \to T_{h_m} f_m$, it is equivalent to prove (21) with $T_{h_m}^n$ replaced by T_{h_m}. The resulting condition is of the form (20) with m replaced by $m - 1$. This completes the induction and shows that $X^n \xrightarrow{fd} X$.

To strengthen the convergence to $X^n \xrightarrow{d} X$, it suffices by Theorems 23.9 and 23.11 to show that $\rho(X_{\tau_n}^n, X_{\tau_n+h_n}^n) \xrightarrow{P} 0$ for any finite optional times τ_n and positive constants $h_n \to 0$. By the strong Markov property, we may prove instead that $\rho(X_0^n, X_{h_n}^n) \xrightarrow{P} 0$ under any initial distributions ν_n. By the compactness of S and Theorem 23.2, we may then assume that $\nu_n \xrightarrow{w} \nu$ for some probability measure ν. Fixing any $f, g \in C$ and noting that $T_{h_n}^n g \to g$ by (iii), we get

$$\begin{aligned} E\, f(X_0^n)\, g(X_{h_n}^n) \;&=\; E\, f T_{h_n}^n\, g(X_0^n) \\ &\to E\, f g(X_0), \end{aligned}$$

where $\mathcal{L}(X_0) = \nu$. Then $(X_0^n, X_{h_n}^n) \xrightarrow{d} (X_0, X_0)$ as before, and in particular $\rho(X_0^n, X_{h_n}^n) \xrightarrow{d} \rho(X_0, X_0) = 0$, proving (iv). □

From the last theorem and its proof, we may derive a similar approximation of discrete-time Markov chains. This extends the approximations of random walks in Corollary 16.17 and Theorem 23.14.

Theorem 17.28 *(approximation of Markov chains) Let $Y^1, Y^2, \ldots$ be discrete-time Markov chains in S with transition operators $U_1, U_2, \ldots$, and let X be a Feller process in S with semi-group (T_t) and generator A. Fix a core D for A, and let $0 < h_n \to 0$. Then conditions* (i)–(iv) *of Theorem* 17.25 *remain equivalent for the operators and processes*

$$A_n = h_n^{-1}(U_n - I), \qquad T_{n,t} = U_n^{[t/h_n]}, \qquad X_t^n = Y_{[t/h_n]}^n.$$

Proof, (i) ⇔ (iv): Let N be an independent, unit-rate Poisson process, and note that the processes $\tilde{X}_t^n = Y^n \circ N_{t/h_n}$ are pseudo-Poisson with generators A_n. Then Theorem 17.25 yields (i) ⇔ (iv) with X^n replaced by $\tilde{X}^n$. By the strong law of large numbers for N, together with Theorem 5.29, we further note that (iv) holds simultaneously for the processes X^n and $\tilde{X}^n$.

(iv) $\Rightarrow$ (iii): Since X is a.s. continuous at fixed times, (iv) yields $X^n_{t_n} \xrightarrow{d} X_t$ whenever $t_n \to t$, and the processes X^n and X start at fixed points $x_n \to x$ in $\hat{S}$. Hence, $T_{n,t_n} f(x_n) \to T_t f(x)$ for any $f \in \hat{C}$, and (iii) follows.

(iii) $\Rightarrow$ (ii): Trivial.

(ii) $\Rightarrow$ (i): As in the previous proof, we need to show that $\tilde{R}^\lambda_n g \to R^\lambda g$ for any $\lambda > 0$ and $g \in C_0$, where $\tilde{R}^\lambda_n = (\lambda - A_n)^{-1}$. Since (ii) yields $R^\lambda_n g \to R^\lambda g$ with $R^\lambda_n = \int e^{-\lambda t} T_{n,t}\, dt$, it suffices to prove that $(R^\lambda_n - \tilde{R}^\lambda_n) g \to 0$. Then note that

$$\lambda R^\lambda_n g - \lambda \tilde{R}^\lambda_n g = Eg(Y^n_{\kappa_n - 1}) - Eg(Y^n_{\tilde{\kappa}_n - 1}),$$

where the random variables κ_n and $\tilde{\kappa}_n$ are independent of Y^n and geometrically distributed with parameters $p_n = 1 - e^{-\lambda h_n}$ and $\tilde{p}_n = \lambda h_n (1 + \lambda h_n)^{-1}$, respectively. Since $p_n \sim \tilde{p}_n$, we have $\|\mathcal{L}(\kappa_n) - \mathcal{L}(\tilde{\kappa}_n)\| \to 0$, and the desired convergence follows by Fubini's theorem. □

We finally show that canonical Feller processes and their induced filtrations are quasi-left continuous.

Proposition 17.29 *(ql-continuity of Feller processes, Blumenthal, Meyer) Let X be a canonical Feller process with arbitrary initial distribution, and fix an optional time τ. Then these conditions are equivalent:*

(i) *τ is predictable,*
(ii) *τ is accessible,*
(iii) *$X_{\tau-} = X_\tau$ a.s. on $\{\tau < \infty\}$.*

In particular we see that, when X is a.s. continuous, every optional time is predictable.

Proof, (ii) $\Rightarrow$ (iii): By Proposition 10.4, we may take τ to be finite and predictable. Now fix an announcing sequence (τ_n) and a function $f \in C_0$. By the strong Markov property, we get for any $h > 0$

$$\begin{aligned} E\Big\{f(X_{\tau_n}) - f(X_{\tau_n + h})\Big\}^2 &= E\Big(f^2 - 2f\, T_h f + T_h f^2\Big)(X_{\tau_n}) \\ &\le \Big\| f^2 - 2f\, T_h f + T_h f^2 \Big\| \\ &\le 2\|f\| \, \Big\| f - T_h f \Big\| + \Big\| f^2 - T_h f^2 \Big\|. \end{aligned}$$

Letting $n \to \infty$ and then $h \downarrow 0$, and using dominated convergence on the left and strong continuity on the right, we see that $E\{f(X_{\tau-}) - f(X_\tau)\}^2 = 0$, which implies $f(X_{\tau-}) = f(X_\tau)$ a.s. Applying this to a separating sequence of functions $f_1, f_2, \ldots \in C_0$, we obtain $X_{\tau-} = X_\tau$ a.s.

(iii) $\Rightarrow$ (i): By (iii) and Theorem 17.20, we have $\Delta M_\tau = 0$ a.s. on $\{\tau < \infty\}$ for every martingale M, and so τ is predictable by Theorem 10.14.

(i) ⇒ (ii): Obvious. □

Exercises

1. Show how the proofs of Theorems 17.4 and 17.6 can be simplified, if we assume (F_3) instead of the weaker condition (F_2).

2. Consider a pseudo-Poisson process X on S with rate kernel α. Give conditions ensuring X to be Feller.

3. Verify the resolvent equation (2), and conclude that the range of R_λ is independent of λ.

4. Show that a Feller semi-group (T_t) is uniquely determined by the resolvent operator R_λ for a fixed $\lambda > 0$. Interpret the result probabilistically in terms of an independent, exponentially distributed random variable with mean λ^{-1}. (*Hint:* Use Theorem 17.4 and Lemma 17.5.)

5. Consider a discrete-time Markov process in S with transition operator T, and let τ be an independent random variable with a fixed geometric distribution. Show that T is uniquely determined by $E_x f(X_\tau)$ for arbitrary $x \in S$ and $f \geq 0$. (*Hint:* Apply the preceding result to the associated pseudo-Poisson process.)

6. Give a probabilistic description of the Yosida approximation T_t^λ in terms the original process X and *two* independent Poisson processes with rate λ.

7. Given a Feller diffusion semi-group, write the differential equation in Theorem 17.6 (ii), for suitable f, as a PDE for the function $T_t f(x)$ on $\mathsf{R}_+ \times \mathsf{R}^d$. Also show that the backward equation of Theorem 13.9 is a special case of the same equation.

8. Consider a Feller process X and an independent subordinator T. Show that $Y = X \circ T$ is again Markov, and that Y is Lévy whenever this is true for X. If both T and X are stable, then so is Y. Find the relationship between the transition semi-groups, respectively between the indices of stability.

9. Consider a Feller process X and an independent renewal process $\tau_0, \tau_1, \ldots$. Show that $Y_n = X_{\tau_n}$ is a discrete-time Markov process, and express its transition kernel in terms of the transition semi-group of X. Also show that $Y_t = X \circ \tau_{[t]}$ may fail to be Markov, even when (τ_n) is Poisson.

10. Let X, Y be independent Feller processes in S, T with generators A, B. Show that (X, Y) is a Feller process in $S \times T$ with generator extending $\tilde{A} + \tilde{B}$, where $\tilde{A}$, $\tilde{B}$ denote the natural extensions of A, B to $C_0(S \times T)$.

11. Consider in S a Feller process with generator A and a pseudo-Poisson process with generator B. Construct a Markov process with generator $A + B$.

12. Use Theorem 17.23 to show that the generator of Brownian motion in R extends $A = \frac{1}{2}\Delta$, on the set D of functions $f \in C_0^2$ with $Af \in C_0$.

13. Let R_λ be the λ-resolvent of Brownian motion in R. For any $f \in C_0$, put $h = R_\lambda f$, and show by direct computation that $\lambda h - \frac{1}{2}h'' = f$. Conclude by Theorem 17.4 that $\frac{1}{2}\Delta$ with domain D, defined as above, extends the generator A. Thus, $A = \frac{1}{2}\Delta$ by the preceding exercise or by Lemma 17.12.

14. Show that if A is a bounded generator on C_0, then the associated Markov process is pseudo-Poisson. (*Hint:* Note as in Theorem 17.11 that A satisfies the positive-maximum principle. Next use Riesz' representation theorem to express A in terms of bounded kernels, and show that A has the form of Proposition 17.2.)

15. Let X, X^n be processes as in Theorem 23.14. Show that if $X_t^n \xrightarrow{d} X_t$ for all $t > 0$, then $X^n \xrightarrow{d} X$ in $D_{\mathsf{R}_+,\mathsf{R}^d}$, and compare with the stated theorem. Also prove a corresponding result for a sequence of Lévy processes X^n. (*Hint:* Use Theorems 17.28 and 17.25, respectively.)

VI. Stochastic Calculus and Applications

Stochastic calculus is rightfully recognized as one of the central areas of modern probability. In Chapter 18 we develop the classical theory of Itô integration with respect to continuous semi-martingales, based on a detailed study of the quadratic variation process. Chapter 19 deals with some basic applications to Brownian motion and related processes, including various martingale characterizations and transformations of martingales based on a random time change or a change of probability measure. In Chapter 20, the integration theory is extended to the more subtle case of semi-martingales with jump discontinuities, and in the final Chapter 21 we develop some basic aspects of the Malliavin calculus for functionals on Wiener space. The material of the first two chapters is absolutely fundamental, whereas the remaining chapters are more advanced and could be deferred to a later stage of studies.

— — —

18. Itô integration and quadratic variation. After a detailed study of the quadratic variation and covariation processes for continuous local martingales, we introduce the stochastic integral via an ingenious martingale characterization. This leads easily to some basic continuity and iteration properties, and to the fundamental Itô formula, showing how continuous semi-martingales are transformed by smooth mappings.

19. Continuous martingales and Brownian motion. Using stochastic calculus, we derive some basic properties of Brownian motion and continuous local martingales. Thus, we show how the latter can be reduced to Brownian motions by a random time change, and how the drift of a continuous semi-martingale can be removed by a suitable change of measure. We further derive some integral representations of Brownian functionals and continuous martingales.

20. Semi-martingales and stochastic integration. Here we extend the stochastic calculus for continuous integrators to the context of semi-martingales with jump discontinuities. Important applications include some general decompositions of semi-martingales, the Burkholder-type inequalities, the Bichteler–Dellacherie characterization of stochastic integrators, and a representation of the Doléans exponential.

21. Malliavin calculus. Here we study the Malliavin derivative D and its dual D^*, along with the Ornstein–Uhlenbeck generator L. Those three operators are closely related and admit simple representations in terms of multiple Wiener–Itô integrals. We further explain in what sense D^* extends the Itô integral, derive a differentiation property for the latter, and establish the Clark–Ocone representation of Brownian functionals. We also indicate how the theory can be used to prove the existence and smoothness of densities.

Chapter 18

Itô Integration and Quadratic Variation

Local and finite-variation martingales, completeness, covariation, continuity, norm comparison, uniform integrability, Cauchy-type inequalities, martingale integral, semi-martingales, continuity, dominated convergence, chain rule, optional stopping, integration by parts, approximation of covariation, Itô's formula, local integral, conformal mapping, Fisk–Stratonovich integral, continuity and approximation, random time change, dependence on parameter, functional representation

Here we initiate the study of stochastic calculus, arguably the most powerful tool of modern probability, with applications to a broad variety of subjects throughout the subject. For the moment, we may only mention the time-change reduction and integral representation of continuous local martingales in Chapter 19, the Girsanov theory for removal of drift in the same chapter, the predictable transformations in Chapter 27, the construction of local time in Chapter 29, and the stochastic differential equations in Chapters 32–33.

In this chapter we consider only stochastic integration with respect to continuous semi-martingales, whereas the more subtle case of integrators with possible jump discontinuities is postponed until Chapter 20, and an extension to non-anticipating integrators appears in Chapter 21. The theory of stochastic integration is inextricably linked to the notions of quadratic variation and covariation, already encountered in Chapter 14 in the special case of Brownian motion, and together the two notions are developed into a theory of amazing power and beauty.

We begin with a construction of the *covariation process* $[M, N]$ of a pair of continuous local martingales M and N, which requires an elementary approximation and completeness argument. The processes M^* and $[M] = [M, M]$ will be related by some useful continuity and norm relations, including the powerful *BDG inequalities*.

Given the quadratic variation $[M]$, we may next construct the stochastic integral $\int V dM$ for suitable progressive processes V, using a simple Hilbert space argument. Combining with the ordinary Stieltjes integral $\int V dA$ for processes A of locally finite variation, we may finally extend the integral to arbitrary continuous semi-martingales $X = M + A$. The continuity properties of quadratic variation carry over to the stochastic integral, and in conjunction with the obvious linearity they characterize the integration.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_19

A key result for applications is *Itô's formula*, which shows how semi-martingales are transformed under smooth mappings. Though the present substitution rule differs from the elementary result for Stieltjes integrals, the two formulas can be brought into agreement by a simple modification of the integral. We conclude the chapter with some special topics of importance for applications, such as the transformation of stochastic integrals under a random time-change, and the integration of processes depending on a parameter.

The present material may be thought of as a continuation of the martingale theory of Chapters 9–10. Though no results for Brownian motion are used explicitly in this chapter, the existence of the Brownian quadratic variation in Chapter 14 may serve as a motivation. We also need the representation and measurability of limits obtained in Chapter 5.

Throughout the chapter we take $\mathcal{F} = (\mathcal{F}_t)$ to be a right-continuous and complete filtration on R_+. A process M is said to be a *local martingale*, if it is adapted to $\mathcal{F}$ and such that the stopped and centered processes $M^{\tau_n} - M_0$ are true martingales for some optional times $\tau_n \uparrow \infty$. By a similar *localization* we may define local L^2-martingales, locally bounded martingales, locally integrable processes, etc. The required optional times τ_n are said to form a *localizing sequence*.

Any continuous local martingale may clearly be reduced by localization to a sequence of bounded, continuous martingales. Conversely, we see by dominated convergence that every bounded local martingale is a true martingale. The following useful result may be less obvious.

Lemma 18.1 *(local martingales) For any process M and optional times $\tau_n \uparrow \infty$, these conditions are equivalent:*

(i) *M is a local martingale,*

(ii) *M^{τ_n} is a local martingale for every n.*

Proof, (i) ⇒ (ii): If M is a local martingale with localizing sequence (σ_n) and τ is an arbitrary optional time, then the processes $(M^\tau)^{\sigma_n} = (M^{\sigma_n})^\tau$ are true martingales. Thus, M^τ is again a local martingale with localizing sequence (σ_n).

(ii) ⇒ (i): Suppose that each process M^{τ_n} is a local martingale with localizing sequence (σ_k^n). Since $\sigma_k^n \to \infty$ a.s. for each n, we may choose some indices k_n with

$$P\Big\{\sigma_{k_n}^n < \tau_n \wedge n\Big\} \le 2^{-n}, \quad n \in \mathsf{N}.$$

Writing $\tau_n' = \tau_n \wedge \sigma_{k_n}^n$, we get $\tau_n' \to \infty$ a.s. by the Borel–Cantelli lemma, and so the optional times $\tau_n'' = \inf_{m \ge n} \tau_m'$ satisfy $\tau_n'' \uparrow \infty$ a.s. It remains to note that the processes $M^{\tau_n''} = (M^{\tau_n'})^{\tau_n''}$ are true martingales. □

Next we show that every continuous martingale of finite variation is a.s. constant. An extension appears as Lemma 10.11.

Proposition 18.2 *(finite-variation martingales) Let M be a continuous local martingale. Then*

$$M \text{ has locally finite variation} \quad \Leftrightarrow \quad M \text{ is a.s. constant.}$$

Proof: By localization we may reduce to the case where $M_0 = 0$ and M has bounded variation. In fact, let V_t denote the total variation of M on the interval $[0,t]$, and note that V is continuous and adapted. For each $n \in \mathbb{N}$, we may then introduce the optional time $\tau_n = \inf\{t \geq 0;\, V_t = n\}$, and note that $M^{\tau_n} - M_0$ is a continuous martingale with total variation bounded by n. We further note that $\tau_n \to \infty$, and that if $M^{\tau_n} = M_0$ a.s. for each n, then even $M = M_0$ a.s. In the reduced case, fix any $t > 0$, write $t_{n,k} = kt/n$, and conclude from the continuity of M that a.s.

$$\begin{aligned}\zeta_n &\equiv \sum_{k\leq n}\Big(M_{t_{n,k}} - M_{t_{n,k-1}}\Big)^2\\ &\leq V_t \max_{k\leq n}\Big|M_{t_{n,k}} - M_{t_{n,k-1}}\Big| \to 0.\end{aligned}$$

Since $\zeta_n \leq V_t^2$, which is bounded by a constant, it follows by the martingale property and dominated convergence that $EM_t^2 = E\zeta_n \to 0$, and so $M_t = 0$ a.s. for each $t > 0$. □

Our construction of stochastic integrals depends on the quadratic variation and covariation processes, which therefore need to be constructed first. Here we use a direct approach, which has the further advantage of giving some insight into the nature of the basic integration-by-parts formula in Proposition 18.16. An alternative but less elementary approach would be to use the Doob–Meyer decomposition in Chapter 10.

The construction utilizes *predictable step processes* of the form

$$\begin{aligned}V_t &= \textstyle\sum_k \xi_k 1\{t > \tau_k\}\\ &= \textstyle\sum_k \eta_k 1_{(\tau_k,\tau_{k+1}]}(t), \quad t \geq 0,\end{aligned} \tag{1}$$

where the τ_n are optional times with $\tau_n \uparrow \infty$ a.s., and the ξ_k and η_k are $\mathcal{F}_{\tau_k}$-measurable random variables for all $k \in \mathbb{N}$. For any process X we consider the elementary integral process $V \cdot X$, given as in Chapter 9 by

$$\begin{aligned}(V\cdot X)_t &\equiv \int_0^t V\,dX = \textstyle\sum_k \xi_k\Big(X_t - X_t^{\tau_k}\Big)\\ &= \textstyle\sum_k \eta_k\Big(X_{\tau_{k+1}}^t - X_{\tau_k}^t\Big),\end{aligned} \tag{2}$$

where the series converge since they have only finitely many non-zero terms. Note that $(V\cdot X)_0 = 0$, and that $V\cdot X$ inherits the possible continuity properties of X. It is further useful to note that $V \cdot X = V \cdot (X - X_0)$. The following simple estimate will be needed later.

Lemma 18.3 *(martingale preservation and L^2-bound) For any continuous L^2-martingale M with $M_0 = 0$ and predictable step process V with $|V| \leq 1$, the process $V \cdot M$ is again an L^2-martingale satisfying*

$$E(V\cdot M)_t^2 \leq EM_t^2, \quad t \geq 0.$$

Proof: First suppose that the sum in (1) has only finitely many non-zero terms. Then $V \cdot M$ is a martingale by Corollary 9.15, and the L^2-bound follows by the computation

$$\begin{aligned} E(V \cdot M)_t^2 &= E \sum\nolimits_k \eta_k^2 \Big(M^t_{\tau_{k+1}} - M^t_{\tau_k} \Big)^2 \\ &\le E \sum\nolimits_k \Big(M^t_{\tau_{k+1}} - M^t_{\tau_k} \Big)^2 \\ &= E\, M_t^2. \end{aligned}$$

The estimate extends to the general case by Fatou's lemma, and the martingale property then extends by uniform integrability. □

Now consider the space $\mathcal{M}^2$ of all L^2-bounded, continuous martingales M with $M_0 = 0$, equipped with the norm $\|M\| = \|M_\infty\|_2$. Recall that $\|M^*\|_2 \le 2\,\|M\|$ by Proposition 9.17.

Lemma 18.4 *(completeness) Let $\mathcal{M}^2$ be the space of L^2-bounded, continuous martingales with $M_0 = 0$ and norm $\|M\| = \|M_\infty\|_2$. Then $\mathcal{M}^2$ is complete and hence a Hilbert space.*

Proof: For any Cauchy sequence $M^1, M^2, \dots$ in $\mathcal{M}^2$, the sequence (M^n_∞) is Cauchy in L^2 and thus converges toward some $\xi \in L^2$. Introduce the L^2-martingale $M_t = E(\xi \,|\, \mathcal{F}_t)$, $t \ge 0$, and note that $M_\infty = \xi$ a.s. since ξ is $\mathcal{F}_\infty$-measurable. Hence,

$$\begin{aligned} \Big\| (M^n - M)^* \Big\|_2 &\le 2\, \|M^n - M\| \\ &= 2\, \Big\| M^n_\infty - M_\infty \Big\|_2 \to 0, \end{aligned}$$

and so $\|M^n - M\| \to 0$. Moreover, $(M^n - M)^* \to 0$ a.s. along a sub-sequence, which implies that M is a.s. continuous with $M_0 = 0$. □

We may now establish the existence and basic properties of the *quadratic variation* and *covariation* processes $[M]$ and $[M, N]$. Extensions to possibly discontinuous processes are considered in Chapter 20.

Theorem 18.5 *(covariation) For any continuous local martingales M, N, there exists a continuous process $[M, N]$ with locally finite variation and $[M, N]_0 = 0$, such that a.s.*

(i) $MN - [M, N]$ *is a local martingale,*

(ii) $[M, N] = [N, M]$,

(iii) $[a\,M_1 + b\,M_2, N] = a\,[M_1, N] + b\,[M_2, N]$,

(iv) $[M, N] = [M - M_0, N]$,

(v) $[M] = [M, M]$ *is non-decreasing,*

(vi) $[M^\tau, N] = [M^\tau, N^\tau] = [M, N]^\tau$ *for any optional time* τ.

The process $[M, N]$ is determined a.s. uniquely by (i).

Proof: The a.s. uniqueness of $[M,N]$ follows from Proposition 18.2, and (ii)−(iii) are immediate consequences. If $[M,N]$ exists with the stated properties and τ is an optional time, then Lemma 18.1 shows that $M^\tau N^\tau - [M,N]^\tau$ is a local martingale, as is the process $M^\tau(N - N^\tau)$ by Corollary 9.15. Hence, even $M^\tau N - [M,N]^\tau$ is a local martingale, and (vi) follows. Furthermore, $MN - (M - M_0)N = M_0 N$ is a local martingale, which yields (iv) whenever either side exists. If both $[M+N]$ and $[M-N]$ exist, then

$$\begin{aligned}4MN &- \Big([M+N] - [M-N]\Big)\\ &= \Big\{(M+N)^2 - [M+N]\Big\} - \Big\{(M-N)^2 - [M-N]\Big\}\end{aligned}$$

is a local martingale, and so we may take $[M,N] = ([M+N] - [M-N])/4$. It is then enough to prove the existence of $[M]$ when $M_0 = 0$.

First let M be bounded. For each $n \in \mathsf{N}$, let $\tau_0^n = 0$, and define recursively

$$\tau_{k+1}^n = \inf\Big\{t > \tau_k^n;\ |M_t - M_{\tau_k^n}| = 2^{-n}\Big\}, \quad k \ge 0.$$

Note that $\tau_k^n \to \infty$ as $k \to \infty$ for fixed n. Introduce the processes

$$\begin{aligned} V_t^n &= \textstyle\sum_k M_{\tau_k^n} 1\Big\{t \in (\tau_k^n, \tau_{k+1}^n]\Big\},\\ Q_t^n &= \textstyle\sum_k \Big(M_{t\wedge\tau_k^n} - M_{t\wedge\tau_{k-1}^n}\Big)^2. \end{aligned}$$

The V^n are bounded, predictable step processes, and clearly

$$M_t^2 = 2\,(V^n \cdot M)_t + Q_t^n, \quad t \ge 0. \tag{3}$$

By Lemma 18.3, the integrals $V^n \cdot M$ are continuous L^2-martingales, and since $|V^n - M| \le 2^{-n}$ for each n, we have for $m \le n$

$$\begin{aligned} \Big\| V^m \cdot M - V^n \cdot M \Big\| &= \Big\| (V^m - V^n) \cdot M \Big\|\\ &\le 2^{-m+1} \|M\|. \end{aligned}$$

Hence, Lemma 18.4 yields a continuous martingale N with $(V^n \cdot M - N)^* \xrightarrow{P} 0$. The process $[M] = M^2 - 2N$ is again continuous, and by (3) we have

$$\Big(Q^n - [M]\Big)^* = 2\,\Big(N - V^n \cdot M\Big)^* \xrightarrow{P} 0.$$

In particular, $[M]$ is a.s. non-decreasing on the random time set $T = \{\tau_k^n;\ n,k \in \mathsf{N}\}$, which extends by continuity to the closure $\bar{T}$. Also note that $[M]$ is constant on each interval in $\bar{T}^c$, since this is true for M and hence also for every Q^n. Thus, (v) follows.

In the unbounded case, define

$$\tau_n = \inf\{t > 0;\ |M_t| = n\}, \quad n \in \mathsf{N}.$$

The processes $[M^{\tau_n}]$ exist as before, and clearly $[M^{\tau_m}]^{\tau_m} = [M^{\tau_n}]^{\tau_m}$ a.s. for all $m < n$. Hence, $[M^{\tau_m}] = [M^{\tau_n}]$ a.s. on $[0, \tau_m]$, and since $\tau_n \to \infty$, there exists a non-decreasing, continuous, and adapted process $[M]$, such that $[M] = [M^{\tau_n}]$

a.s. on $[0, \tau_n]$ for every n. Here $(M^{\tau_n})^2 - [M]^{\tau_n}$ is a local martingale for every n, and so $M^2 - [M]$ is a local martingale by Lemma 18.1. □

Next we establish a basic continuity property.

Proposition 18.6 *(continuity) For any continuous local martingales M_n starting at 0, we have*

$$M_n^* \xrightarrow{P} 0 \quad \Leftrightarrow \quad [M_n]_\infty \xrightarrow{P} 0.$$

Proof: First let $M_n^* \xrightarrow{P} 0$. Fix any $\varepsilon > 0$, and define $\tau_n = \inf\{t \geq 0;\ |M_n(t)| > \varepsilon\}$, $n \in \mathsf{N}$. Write $N_n = M_n^2 - [M_n]$, and note that $N_n^{\tau_n}$ is a true martingale on $\bar{\mathsf{R}}_+$. In particular, $E[M_n]_{\tau_n} \leq \varepsilon^2$, and so by Chebyshev's inequality

$$\begin{aligned} P\Big\{[M_n]_\infty > \varepsilon\Big\} &\leq P\{\tau_n < \infty\} + \varepsilon^{-1} E[M_n]_{\tau_n} \\ &\leq P\{M_n^* > \varepsilon\} + \varepsilon. \end{aligned}$$

Here the right-hand side tends to zero as $n \to \infty$ and then $\varepsilon \to 0$, which shows that $[M_n]_\infty \xrightarrow{P} 0$.

Conversely, let $[M_n]_\infty \xrightarrow{P} 0$. By a localization argument and Fatou's lemma, we see that M is L^2-bounded. Now proceed as before to get $M_n^* \xrightarrow{P} 0$. □

Next we prove a pair of basic norm inequalities[1] involving the quadratic variation, known as the *BDG inequalities*. Partial extensions to discontinuous martingales appear in Theorem 20.12.

Theorem 18.7 *(norm comparison, Burkholder, Millar, Gundy, Novikov) For a continuous local martingale M with $M_0 = 0$, we have*[2]

$$EM^{*p} \asymp E[M]_\infty^{p/2}, \quad p > 0.$$

Proof: By optional stopping, we may take M and $[M]$ to be bounded. Write $M' = M - M^\tau$ with $\tau = \inf\{t;\ M_t^2 = r\}$, and define $N = (M')^2 - [M']$. By Corollary 9.31, we have for any $r > 0$ and $c \in (0, 2^{-p})$

$$\begin{aligned} P\Big\{M^{*2} \geq 4r\Big\} - P\Big\{[M]_\infty \geq c\,r\Big\} &\leq\ P\Big\{M^{*2} \geq 4r,\ [M]_\infty < c\,r\Big\} \\ &\leq\ P\Big\{N > -c\,r,\ \sup\nolimits_t N_t > r - c\,r\Big\} \\ &\leq\ c\,P\{N^* > 0\} \\ &\leq\ c\,P\{M^{*2} \geq r\}. \end{aligned}$$

Multiplying by $(p/2)\, r^{p/2-1}$ and integrating over R_+, we get by Lemma 4.4

$$2^{-p} EM^{*p} - c^{-p/2} E[M]_\infty^{p/2} \leq c\, EM^{*p},$$

which gives the bound $\lesssim$ with domination constant $c_p = c^{-p/2}/(2^{-p} - c)$.

---

[1] Recall that $f \asymp g$ means $f \leq cg$ and $g \leq cf$ for some constant $c > 0$.

[2] The domination constants are understood to depend only on p.

Defining N as before with $\tau = \inf\{t;\, [M]_t = r\}$, we get for any $r > 0$ and $c \in (0, 2^{-p/2-2})$

$$\begin{aligned} P\Big\{[M]_\infty \geq 2\,r\Big\} - P\Big\{M^{*2} \geq c\,r\Big\} &\leq P\Big\{[M]_\infty \geq 2\,r,\, M^{*2} < c\,r\Big\} \\ &\leq P\Big\{N < 4\,c\,r,\, \inf_t N_t < 4\,c\,r - r\Big\} \\ &\leq 4c\,P\Big\{[M]_\infty \geq r\Big\}. \end{aligned}$$

Integrating as before yields

$$2^{-p/2} E[M]_\infty^{p/2} - c^{-p/2} E M^{*p} \leq 4c\, E[M]_\infty^{p/2},$$

and the bound $\gtrsim$ follows with domination constant $c_p = c^{-p/2}/(2^{-p/2} - 4\,c)$. □

We often need to certify that a given local martingale is a true martingale. The last theorem yields a useful criterion.

Corollary 18.8 *(uniform integrability) Let M be a continuous local martingale. Then M is a uniformly integrable martingale, whenever*

$$E\Big(|M_0| + [M]_\infty^{1/2}\Big) < \infty.$$

Proof: By Theorem 18.7 we have $EM^* < \infty$, and the martingale property follows by dominated convergence. □

The basic properties of $[M, N]$ suggest that we think of the covariation process as a kind of inner product. A further justification is given by the following Cauchy-type inequalities.

Proposition 18.9 *(Cauchy-type inequalities, Courrège) For any continuous local martingales M, N and product-measurable processes U, V, we have a.s.*

(i) $\Big|[M,N]\Big| \leq \int \Big|d\,[M,N]\Big| \leq \Big([M]\,[N]\Big)^{1/2}$,

(ii) $\int_0^t \Big|UV\,d\,[M,N]\Big| \leq \Big\{\Big(U^2 \cdot [M]\Big)_t \Big(V^2 \cdot [N]\Big)_t\Big\}^{1/2}, \quad t \geq 0.$

Proof: (i) By Theorem 18.5 (iii) and (v), we have a.s. for any $a, b \in \mathsf{R}$ and $t > 0$

$$\begin{aligned} 0 &\leq [aM + bN]_t \\ &= a^2\,[M]_t + 2ab\,[M,N]_t + b^2\,[N]_t. \end{aligned}$$

By continuity we may choose a common exceptional null set for all a, b, and so $[M,N]_t^2 \leq [M]_t [N]_t$ a.s. Applying this inequality to the processes $M - M^s$ and $N - N^s$ for any $s < t$, we obtain a.s.

$$\Big|\,[M,N]_t - [M,N]_s\Big| \leq \Big\{\Big([M]_t - [M]_s\Big)\Big([N]_t - [N]_s\Big)\Big\}^{1/2}, \tag{4}$$

and by continuity we may again choose a common null set. Now let $0 = t_0 < t_1 < \cdots < t_n = t$ be arbitrary, and conclude from (4) and the classical Cauchy inequality that

$$
\begin{aligned}
|\,[M,N]_t| &\le \textstyle\sum_k \Big|\,[M,N]_{t_k} - [M,N]_{t_{k-1}}\Big| \\
&\le \Big([M]_t\,[N]_t\Big)^{1/2}.
\end{aligned}
$$

It remains to take the supremum over all partitions of $[0,t]$.

(ii) Writing $d\mu = d[M]$, $d\nu = d[N]$, and $d\rho = |d[M,N]|$, we conclude from (i) that $(\rho I)^2 \le (\mu I)(\nu I)$ a.s. for every interval I. By continuity, we may choose the exceptional null set A to be independent of I. Letting $G \subset \mathsf{R}_+$ be open with connected components I_k and using Cauchy's inequality, we get on A^c

$$
\begin{aligned}
\rho G &= \textstyle\sum_k \rho I_k \\
&\le \textstyle\sum_k \Big(\mu I_k\,\nu I_k\Big)^{1/2} \\
&\le \Big(\textstyle\sum_j \mu I_j \sum_k \nu I_k\Big)^{1/2} \\
&= \Big(\mu G\cdot \nu G\Big)^{1/2}.
\end{aligned}
$$

By Lemma 1.36, the last relation extends to any $B \in \mathcal{B}_{\mathsf{R}_+}$.

Now fix any simple, measurable functions $f = \sum_k a_k 1_{B_k}$ and $g = \sum_k b_k 1_{B_k}$. Using Cauchy's inequality again, we obtain on A^c

$$
\begin{aligned}
\rho|fg| &\le \textstyle\sum_k |a_k b_k|\,\rho B_k \\
&\le \textstyle\sum_k |a_k b_k|\,\Big(\mu B_k\,\nu B_k\Big)^{1/2} \\
&\le \Big(\textstyle\sum_j a_j^2\,\mu B_j \sum_k b_k^2\,\nu B_k\Big)^{1/2} \\
&\le \Big(\mu f^2\,\nu g^2\Big)^{1/2},
\end{aligned}
$$

which extends by monotone convergence to any measurable functions f, g on R_+. By Lemma 1.35, we may choose $f(t) = U_t(\omega)$ and $g(t) = V_t(\omega)$ for any fixed $\omega \in A^c$. □

Let $\mathcal{E}$ be the class of bounded, predictable step processes with jumps at finitely many fixed times. To motivate the construction of general stochastic integrals, and for subsequent needs, we derive a basic identity for elementary integrals.

Lemma 18.10 *(covariation identity) For any continuous local martingales M, N and processes $U, V \in \mathcal{E}$, the integrals $U\cdot M$ and $V\cdot N$ are again continuous local martingales, and we have*

$$
\Big[U\cdot M,\; V\cdot N\Big] = (\,UV)\cdot [M,N] \quad a.s. \tag{5}
$$

Proof: We may clearly take $M_0 = N_0 = 0$. The first assertion follows by localization from Lemma 18.3. To prove (5), let $U_t = \sum_{k\le n} \xi_k 1_{(t_k,t_{k+1}]}(t)$, where ξ_k is bounded and $\mathcal{F}_{t_k}$-measurable for each k. By localization we may take M, N, $[M,N]$ to be bounded, so that M, N, and $MN - [M,N]$ are martingales on $\bar{\mathsf{R}}_+$. Then

$$
\begin{aligned}
E\,(U\cdot M)_\infty N_\infty &= E\,\textstyle\sum_j \xi_j\Big(M_{t_{j+1}} - M_{t_j}\Big)\sum_k\Big(N_{t_{k+1}} - N_{t_k}\Big) \\
&= E\,\textstyle\sum_k \xi_k\Big(M_{t_{k+1}}N_{t_{k+1}} - M_{t_k}N_{t_k}\Big) \\
&= E\,\textstyle\sum_k \xi_k\Big([M,N]_{t_{k+1}} - [M,N]_{t_k}\Big) \\
&= E\Big(U\cdot[M,N]\Big)_\infty.
\end{aligned}
$$

Replacing M, N by M^τ, N^τ for an arbitrary optional time τ, we get

$$\begin{aligned} E\,(U\cdot M)_\tau N_\tau &= E(\,U\cdot M^\tau)_\infty\, N^\tau_\infty \\ &= E\Big(\,U\cdot [M^\tau, N^\tau]\Big)_\infty \\ &= E\Big(\,U\cdot [M,N]\Big)_\tau. \end{aligned}$$

By Lemma 9.14, the process $(\,U\cdot M)N - U\cdot [M,N]$ is then a martingale, and so $[\,U\cdot M, N] = U\cdot [M,N]$ a.s. The general formula follows by iteration. □

To extend the stochastic integral $V\cdot M$ to more general processes V, we take (5) to be the characteristic property. For any continuous local martingale M, let $L(M)$ be the class of all progressive processes V, such that $(V^2\cdot[M])_t < \infty$ a.s. for every $t > 0$.

Theorem 18.11 *(martingale integral, Itô, Kunita & Watanabe) For any continuous local martingale M and process $V \in L(M)$, there exists an a.s. unique continuous local martingale $V\cdot M$ with $(V\cdot M)_0 = 0$, such that for any continuous local martingale N,*

$$[V\cdot M,\ N] = V\cdot [M,N] \quad \textit{a.s.}$$

Proof: To prove the uniqueness, let M', M'' be continuous local martingales with $M'_0 = M''_0 = 0$, such that

$$\begin{aligned} [M', N] &= [M'', N] \\ &= V\cdot [M,N] \ \text{ a.s.} \end{aligned}$$

for all continuous local martingales N. Then by linearity $[M' - M'', N] = 0$ a.s. Taking $N = M' - M''$ gives $[M' - M''] = 0$ a.s., and so $M' = M''$ a.s. by Proposition 18.6.

To prove the existence, we first assume $\|V\|_M^2 = E\Big(V^2\cdot [M]\Big)_\infty < \infty$. Since V is measurable, we get by Proposition 18.9 and Cauchy's inequality

$$\Big|E\Big(V\cdot [M,N]\Big)_\infty\Big| \le \|V\|_M \|N\|, \quad N \in \mathcal{M}^2.$$

The mapping $N \mapsto E\Big(V\cdot [M,N]\Big)_\infty$ is then a continuous linear functional on $\mathcal{M}^2$, and so Lemma 18.4 yields an element $V\cdot M \in \mathcal{M}^2$ with

$$E\Big(V\cdot [M,N]\Big)_\infty = E\,(V\cdot M)_\infty N_\infty, \quad N \in \mathcal{M}^2.$$

Replacing N by N^τ for an arbitrary optional time τ and using Theorem 18.5 and optional sampling, we get

$$\begin{aligned} E\Big(V\cdot [M,N]\Big)_\tau &= E\Big(V\cdot [M,N]^\tau\Big)_\infty \\ &= E\Big(V\cdot [M,N^\tau]\Big)_\infty \\ &= E(V\cdot M)_\infty N_\tau \\ &= E(V\cdot M)_\tau N_\tau. \end{aligned}$$

Since V is progressive, Lemma 9.14 shows that $V \cdot [M,N] - (V \cdot M)N$ is a martingale, which implies $[V \cdot M, N] = V \cdot [M,N]$ a.s. The last relation extends by localization to any continuous local martingale N.

For general V, define

$$\tau_n = \inf\Big\{t > 0; \big(V^2 \cdot [M]\big)_t = n\Big\}, \quad n \in \mathsf{N}.$$

By the previous argument, there exist some continuous local martingales $V \cdot M^{\tau_n}$ such that, for any continuous local martingale N,

$$\Big[V \cdot M^{\tau_n}, N\Big] = V \cdot [M^{\tau_n}, N] \ \text{a.s.}, \quad n \in \mathsf{N}. \tag{6}$$

For $m < n$ it follows that $(V \cdot M^{\tau_n})^{\tau_m}$ satisfies the corresponding relation with $[M^{\tau_m}, N]$, and so $(V \cdot M^{\tau_n})^{\tau_m} = V \cdot M^{\tau_m}$ a.s. Hence, there exists a continuous process $V \cdot M$ with $(V \cdot M)^{\tau_n} = V \cdot M^{\tau_n}$ a.s. for all n, and Lemma 18.1 shows that $V \cdot M$ is again a local martingale. Finally, (6) yields $[V \cdot M, N] = V \cdot [M,N]$ a.s. on $[0, \tau_n]$ for each n, and so the same relation holds on R_+. □

By Lemma 18.10, the stochastic integral $V \cdot M$ of Theorem 18.11 extends the previously defined elementary integral. It is also clear that $V \cdot M$ is a.s. bilinear in the pair (V, M), with the following basic continuity property.

Lemma 18.12 *(continuity) For any continuous local martingales M_n and processes $V_n \in L(M_n)$, we have*

$$(V_n \cdot M_n)^* \xrightarrow{P} 0 \quad \Leftrightarrow \quad \big(V_n^2 \cdot [M_n]\big)_\infty \xrightarrow{P} 0.$$

Proof: Note that $[V_n \cdot M_n] = V_n^2 \cdot [M_n]$, and use Proposition 18.6. □

Before continuing our study of basic properties, we extend the stochastic integral to a larger class of integrators. By a *continuous semi-martingale* we mean a process $X = M + A$, where M is a continuous local martingale and A is a continuous, adapted process with locally finite variation and $A_0 = 0$. The decomposition $X = M + A$ is then a.s. unique by Proposition 18.2, and it is often referred to as the *canonical decomposition* of X. A continuous semi-martingale in R^d is defined as a process $X = (X^1, \dots, X^d)$, where $X^1, \dots, X^d$ are continuous semi-martingales in R.

Let $L(A)$ be the class of progressive processes V, such that the processes $(V \cdot A)_t = \int_0^t V dA$ exist as elementary Stieltjes integrals. For any continuous semi-martingale $X = M + A$, we write $L(X) = L(M) \cap L(A)$, and define the X-integral of a process $V \in L(X)$ as the sum $V \cdot X = V \cdot M + V \cdot A$, which makes $V \cdot X$ a continuous semi-martingale with canonical decomposition $V \cdot M + V \cdot A$. For progressive processes V, it is further clear that

$$V \in L(X) \quad \Leftrightarrow \quad V^2 \in L([M]), \ \ V \in L(A).$$

Lemma 18.12 yields the following stochastic version of the dominated convergence theorem.

Corollary 18.13 *(dominated convergence) For any continuous semi-martingale X and processes $U, V, V_1, V_2, \ldots \in L(X)$, we have*

$$\left.\begin{array}{r} V_n \to V \\ |V_n| \le U \end{array}\right\} \Rightarrow \left(V_n \cdot X - V \cdot X\right)_t^* \xrightarrow{P} 0, \quad t \ge 0.$$

Proof: Let $X = M + A$. Since $U \in L(X)$, we have $U^2 \in L([M])$ and $U \in L(A)$. By dominated convergence for ordinary Stieltjes integrals, we obtain a.s.

$$\left\{(V_n - V)^2 \cdot [M]\right\}_t + \left(V_n \cdot A - V \cdot A\right)_t^* \to 0.$$

Here the former convergence implies $(V_n \cdot M - V \cdot M)_t^* \xrightarrow{P} 0$ by Lemma 18.12, and the assertion follows. □

We further extend the elementary chain rule of Lemma 1.25 to stochastic integrals.

Lemma 18.14 *(chain rule) For any continuous semi-martingale X and progressive processes U, V with $V \in L(X)$, we have*

(i) *$U \in L(V \cdot X) \Leftrightarrow UV \in L(X)$, and then*

(ii) *$U \cdot (V \cdot X) = (UV) \cdot X$ a.s.*

Proof: (i) Letting $X = M + A$, we have

$$\begin{aligned} U \in L(V \cdot X) \;&\Leftrightarrow\; U^2 \in L([V \cdot M]), \quad U \in L(V \cdot A), \\ UV \in L(X) \;&\Leftrightarrow\; (UV)^2 \in L([M]), \quad UV \in L(A). \end{aligned}$$

Since $[V \cdot M] = V^2 \cdot [M]$, the two pairs of conditions are equivalent.

(ii) The relation $U \cdot (V \cdot A) = (UV) \cdot A$ is elementary. To see that even $U \cdot (V \cdot M) = (UV) \cdot M$ a.s., consider any continuous local martingale N, and note that

$$\begin{aligned} \left[(UV) \cdot M, N\right] &= (UV) \cdot [M, N] \\ &= U \cdot \left(V \cdot [M, N]\right) \\ &= U \cdot [V \cdot M, N] \\ &= \left[U \cdot (V \cdot M), N\right]. \end{aligned}$$

□

Next we examine the behavior under optional stopping.

Lemma 18.15 *(optional stopping) For any continuous semi-martingale X, process $V \in L(X)$, and optional time τ, we have a.s.*

$$(V \cdot X)^\tau = V \cdot X^\tau = (V 1_{[0,\tau]}) \cdot X.$$

Proof: The relations being obvious for ordinary Stieltjes integrals, we may take $X = M$ to be a continuous local martingale. Then $(V \cdot M)^\tau$ is a continuous local martingale starting at 0, and we have

$$\begin{aligned}\big[(V \cdot M)^\tau, N\big] &= \big[V \cdot M, N^\tau\big]\\ &= V \cdot [M, N^\tau]\\ &= V \cdot [M, N]^\tau\\ &= (V 1_{[0,\tau]}) \cdot [M, N].\end{aligned}$$

Thus, $(V \cdot M)^\tau$ satisfies the conditions characterizing the integrals $V \cdot M^\tau$ and $(V 1_{[0,\tau]}) \cdot M$. □

We may extend the definitions of quadratic variation and covariation to any continuous semi-martingales $X = M + A$ and $Y = N + B$ by putting $[X] = [M]$ and $[X, Y] = [M, N]$. As a crucial step toward a general substitution rule, we show how the covariation process can be expressed in terms of stochastic integrals. For martingales X and Y, the result is implicit in the proof of Theorem 18.5.

Theorem 18.16 *(integration by parts) For any continuous semi-martingales X, Y, we have a.s.*

$$XY = X_0 Y_0 + X \cdot Y + Y \cdot X + [X, Y]. \tag{7}$$

Proof: We may take $X = Y$, since the general result will then follow by polarization. First let $X = M \in \mathcal{M}^2$, and define V^n and Q^n as in the proof of Theorem 18.5. Then $V^n \to M$ and $|V_t^n| \le M_t^* < \infty$, and so Corollary 18.13 yields $(V^n \cdot M)_t \xrightarrow{P} (M \cdot M)_t$ for each $t \ge 0$. Now (7) follows as we let $n \to \infty$ in the relation $M^2 = 2\, V^n \cdot M + Q^n$, and it extends by localization to general continuous local martingales M with $M_0 = 0$. If instead $X = A$, then (7) reduces to $A^2 = 2\, A \cdot A$, which holds by Fubini's theorem.

For general X we may take $X_0 = 0$, since the formula for general X_0 will then follow by an easy computation from the result for $X - X_0$. In this case, (7) reduces to $X^2 = 2\, X \cdot X + [M]$. Subtracting the formulas for M^2 and A^2, it remains to show that $AM = A \cdot M + M \cdot A$ a.s. Then fix any $t > 0$, put $t_k^n = (k/n)t$, and introduce for $s \in (t_{k-1}^n, t_k^n]$, $k, n \in \mathsf{N}$, the processes

$$A_s^n = A_{t_{k-1}^n}, \qquad M_s^n = M_{t_k^n}.$$

Note that

$$A_t\, M_t = (A^n \cdot M)_t + (M^n \cdot A)_t, \quad n \in \mathsf{N}.$$

Here $(A^n \cdot M)_t \xrightarrow{P} (A \cdot M)_t$ by Corollary 18.13, and $(M^n \cdot A)_t \to (M \cdot A)_t$ by dominated convergence for ordinary Stieltjes integrals. □

Our terminology is justified by the following result, which extends Theorem 14.9 for Brownian motion. It also shows that $[X, Y]$ is a.s. measurably determined[3] by X and Y.

[3]This is remarkable, since it is defined by martingale properties that depend on both probability measure and filtration.

Proposition 18.17 *(approximation of covariation, Fisk) For any continuous semi-martingales* X, Y *on* $[0,t]$ *and partitions* $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$, $n \in \mathsf{N}$, *with* $\max_k(t_k^n - t_{k-1}^n) \to 0$, *we have*

$$\zeta_n \equiv \sum\nolimits_k \Big(X_{t_k^n} - X_{t_{k-1}^n}\Big)\Big(Y_{t_k^n} - Y_{t_{k-1}^n}\Big) \xrightarrow{P} [X,Y]_t. \tag{8}$$

Proof: We may clearly take $X_0 = Y_0 = 0$. Introduce for $s \in (t_{k-1}^n, t_k^n]$, $k, n \in \mathsf{N}$, the predictable step processes

$$X_s^n = X_{t_{k-1}^n}, \qquad Y_s^n = Y_{t_{k-1}^n},$$

and note that

$$X_tY_t = (X^n \cdot Y)_t + (Y^n \cdot X)_t + \zeta_n, \quad n \in \mathsf{N}.$$

Since $X^n \to X$ and $Y^n \to Y$, and also $(X^n)_t^* \le X_t^* < \infty$ and $(Y^n)_t^* \le X_t^* < \infty$, we get by Corollary 18.13 and Theorem 18.16

$$\begin{aligned}\zeta_n &\xrightarrow{P} X_tY_t - (X \cdot Y)_t - (Y \cdot X)_t \\ &= [X,Y]_t.\end{aligned}$$

□

We turn to a version of *Itô's formula*, arguably the most important formula of modern probability.[4] It shows that the class of continuous semi-martingales is closed under smooth maps, and exhibits the canonical decomposition of the image process in terms of the components of the original process. Extended versions appear in Corollaries 18.19 and 18.20, as well as in Theorems 20.7 and 29.5.

Write $C^k = C^k(\mathsf{R}^d)$ for the class of k times continuously differentiable functions on R^d. When $f \in C^2$, we write $\partial_i f$ and $\partial_{ij}^2 f$ for the first and second order partial derivatives of f. Here and below, summation over repeated indices is understood.

Theorem 18.18 *(substitution rule, Itô) For any continuous semi-martingale* X *in* R^d *and function* $f \in C^2(\mathsf{R}^d)$, *we have a.s.*

$$f(X) = f(X_0) + \partial_i f(X) \cdot X^i + \tfrac{1}{2}\,\partial_{ij}^2 f(X) \cdot [X^i, X^j]. \tag{9}$$

This may also be written in differential form as

$$df(X) = \partial_i f(X)\, dX^i + \tfrac{1}{2}\,\partial_{ij}^2 f(X)\, d\,[X^i, X^j]. \tag{10}$$

It is suggestive to write $d\,[X^i, X^j] = dX^i\, dX^j$, and think of (10) as a second order Taylor expansion.

If X has canonical decomposition $M + A$, we get the corresponding decomposition of $f(X)$ by substituting $M^i + A^i$ for X^i on the right side of (9). When $M = 0$, the last term vanishes, and (9) reduces to the familiar substitution rule for ordinary Stieltjes integrals. In general, the appearance of this *Itô*

[4] Possible contenders might include the representation of infinitely divisible distributions, the polynomial representation of multiple WI-integrals, and the formula for the generator of a continuous Feller process.

correction term shows that the rules of ordinary calculus fail for the Itô integral.

Proof of Theorem 18.18: For notational convenience, we may take $d = 1$, the general case being similar. Then fix a continuous semi-martingale X in R, and let $\mathcal{C}$ be the class of functions $f \in C^2$ satisfying (9), now written in the form

$$f(X) = f(X_0) + f'(X) \cdot X + \tfrac{1}{2}\, f''(X) \cdot [X]. \tag{11}$$

The class $\mathcal{C}$ is clearly a linear subspace of C^2 containing the functions $f(x) \equiv 1$ and $f(x) \equiv x$. We shall prove that $\mathcal{C}$ is closed under multiplication, and hence contains all polynomials.

Then suppose that (11) holds for both f and g, so that $F = f(X)$ and $G = g(X)$ are continuous semi-martingales. Using the definition of the integral, along with Proposition 18.14 and Theorem 18.16, we get

$$\begin{aligned}
&(fg)(X) - (fg)(X_0)\\
&\quad = FG - F_0G_0\\
&\quad = F \cdot G + G \cdot F + [F, G]\\
&\quad = F \cdot \Big\{g'(X) \cdot X + \tfrac{1}{2}\, g''(X) \cdot [X]\Big\}\\
&\qquad\quad + G \cdot \Big\{f'(X) \cdot X + \tfrac{1}{2}\, f''(X) \cdot [X]\Big\} + \Big[f'(X) \cdot X,\ g'(X) \cdot X\Big]\\
&\quad = \Big(fg' + f'g\Big)(X) \cdot X + \tfrac{1}{2}\Big(fg'' + 2f'g' + f''g\Big)(X) \cdot [X]\\
&\quad = (fg)'(X) \cdot X + \tfrac{1}{2}\,(fg)''(X) \cdot [X].
\end{aligned}$$

Now let $f \in C^2$ be arbitrary. By Weierstrass' approximation theorem[5], we may choose some polynomials $p_1, p_2, \ldots,$ such that $\sup_{|x|\le c} |p_n(x) - f''(x)| \to 0$ for every $c > 0$. Integrating the p_n twice yields some polynomials f_n satisfying

$$\sup_{|x|\le c} \Big\{\Big|f_n(x) - f(x)\Big| \vee \Big|f_n'(x) - f'(x)\Big| \vee \Big|f_n''(x) - f''(x)\Big|\Big\} \to 0, \quad c > 0.$$

In particular, $f_n(X_t) \to f(X_t)$ for each $t > 0$. Letting X have canonical decomposition $M + A$ and using dominated convergence for ordinary Stieltjes integrals, we get for any $t \ge 0$

$$\Big\{f_n'(X) \cdot A + \tfrac{1}{2}\, f_n''(X) \cdot [X]\Big\}_t \to \Big\{f'(X) \cdot A + \tfrac{1}{2}\, f''(X) \cdot [X]\Big\}_t.$$

Similarly, $\{f_n'(X) - f'(X)\}^2 \cdot [M])_t \to 0$ for all t, and so by Lemma 18.12

$$\Big\{f_n'(X) \cdot M\Big\}_t \xrightarrow{P} \Big\{f'(X) \cdot M\Big\}_t, \quad t \ge 0.$$

Thus, formula (11) for the polynomials f_n extends in the limit to the same relation for f. □

We also need a local version of the last theorem, involving stochastic integrals up to the first time ζ_D when X exits a given domain $D \subset \mathsf{R}^d$. For

[5] Any continuous function on $[0, 1]$ admits a uniform approximation by polynomials.

continuous and adapted X, the time ζ_D is clearly *predictable*, in the sense of being *announced* by some optional times $\tau_n \uparrow \zeta_D$ with $\tau_n < \zeta_D$ a.s. on $\{\zeta_D > 0\}$ for all n. Indeed, writing ρ for the Euclidean metric in R^d, we may choose

$$\tau_n = \inf\Big\{t \in [0,n];\ \rho(X_t, D^c) \le n^{-1}\Big\}, \quad n \in \mathsf{N}. \tag{12}$$

We say that X is a semi-martingale on $[0, \zeta_D)$, if the stopped process X^{τ_n} is a semi-martingale in the usual sense for every $n \in \mathsf{N}$. To define the co-variation processes $[X^i, X^j]$ on the interval $[0, \zeta_D)$, we require $[X^i, X^j]^{\tau_n} = [(X^i)^{\tau_n}, (X^j)^{\tau_n}]$ a.s. for every n. Stochastic integrals with respect to $X^1, \dots, X^d$ are defined on $[0, \zeta_D)$ in a similar way.

Corollary 18.19 *(local Itô formula) For any domain $D \subset \mathsf{R}^d$, let X be a continuous semi-martingale on $[0, \zeta_D)$. Then (9) holds a.s. on $[0, \zeta_D)$ for every $f \in C^2(D)$.*

Proof: Choose some $f_n \in C^2(\mathsf{R}^d)$ with $f_n(x) = f(x)$ when $\rho(x, D^c) \ge n^{-1}$. Applying Theorem 18.18 to $f_n(X^{\tau_n})$ with τ_n as in (12), we get (9) on $[0, \tau_n]$. Since n was arbitrary, the result extends to $[0, \zeta_D)$. □

By a *complex, continuous semi-martingale* we mean a process of the form $Z = X + iY$, where X, Y are real, continuous semi-martingales. The bilinearity of the covariation process suggests that we define the quadratic variation of Z as

$$\begin{aligned}[Z] = [Z, Z] &= \Big[X + iY, X + iY\Big] \\ &= [X] + 2i\,[X, Y] - [Y].\end{aligned}$$

Let $L(Z)$ be the class of processes $W = U + iV$ with $U, V \in L(X) \cap L(Y)$. For such a process W, we define the integral by

$$\begin{aligned}W \cdot Z &= (U + iV) \cdot (X + iY) \\ &= U \cdot X - V \cdot Y + i\Big(U \cdot Y + V \cdot X\Big).\end{aligned}$$

Corollary 18.20 *(conformal mapping) Let f be an analytic function on a domain $D \subset \mathsf{C}$. Then (9) holds for every continuous semi-martingale Z in D.*

Proof: Writing $f(x + iy) = g(x, y) + ih(x, y)$ for any $x + iy \in D$, we get

$$g_1' + ih_1' = f', \qquad g_2' + ih_2' = if',$$

and so by iteration

$$g_{11}'' + ih_{11}'' = f'', \qquad g_{12}'' + ih_{12}'' = if'', \qquad g_{22}'' + ih_{22}'' = -f''.$$

Now (9) follows for $Z = X + iY$, as we apply Corollary 18.19 to the semi-martingale (X, Y) and the functions g, h. □

Under suitable regularity conditions, we may modify the Itô integral so that it will obey the rules of ordinary calculus. Then for any continuous semi-martingales X, Y, we define the *Fisk–Stratonovich integral* by

$$\int_0^t X \circ dY = (X \cdot Y)_t + \tfrac{1}{2}\,[X,Y]_t, \quad t \geq 0, \tag{13}$$

or, in differential form $X \circ dY = XdY + \frac{1}{2}\,d[X,Y]$, where the first term on the right is an ordinary Itô integral. The point of this modification is that the substitution rule simplifies to $df(X) = \partial_i f(X) \circ dX^i$, conforming with the chain rule of ordinary calculus. We may also prove a version for FS-integrals of the chain rule in Proposition 18.14.

Theorem 18.21 *(modified Itô integral, Fisk, Stratonovich) The integral in* (13) *satisfies the computational rules of elementary calculus. Thus, a.s.,*

(i) *for any continuous semi-martingale X in R^d and function $f \in C^3(\mathsf{R}^d)$,*

$$f(X_t) = f(X_0) + \int_0^t \partial_i f(X) \circ dX^i, \quad t \geq 0,$$

(ii) *for any real, continuous semi-martingales X, Y, Z,*

$$X \circ (Y \circ Z) = (XY) \circ Z.$$

Proof: (i) By Itô's formula,

$$\partial_i f(X) = \partial_i f(X_0) + \partial^2_{ij} f(X) \cdot X^j + \tfrac{1}{2}\,\partial^3_{ijk} f(X) \cdot [X^j, X^k].$$

Using Itô's formula again together with (5) and (13), we get

$$\begin{aligned}\int_0 \partial_i f(X) \circ dX^i &= \partial_i f(X) \cdot X^i + \tfrac{1}{2}\,[\partial_i f(X), X^i] \\ &= \partial_i f(X) \cdot X^i + \tfrac{1}{2}\,\partial^2_{ij} f(X) \cdot [X^j, X^i] \\ &= f(X) - f(X_0).\end{aligned}$$

(ii) By Lemma 18.14 and integration by parts,

$$\begin{aligned}X \circ (Y \circ Z) &= X \cdot (Y \circ Z) + \tfrac{1}{2}\,[X, Y \circ Z] \\ &= X \cdot (Y \cdot Z) + \tfrac{1}{2}\,X \cdot [Y,Z] + \tfrac{1}{2}\,[X, Y \cdot Z] + \tfrac{1}{4}\Big[X, [Y,Z]\Big] \\ &= XY \cdot Z + \tfrac{1}{2}\,[XY, Z] \\ &= XY \circ Z.\end{aligned}$$

□

The more convenient substitution rule of Corollary 18.21 comes at a high price: The FS-integral does not preserve the martingale property, and it requires even the integrand to be a continuous semi-martingale, which forces us to impose the stronger regularity constraint on the function f in the substitution rule.

Our next task is to establish a basic uniqueness property, justifying our reference to the process $V \cdot M$ in Theorem 18.11 as an integral.

Theorem 18.22 *(continuity) The integral $V \cdot M$ in Theorem* 18.11 *is the a.s. unique linear extension of the elementary stochastic integral, such that for any $t > 0$,*

$$\left(V_n^2 \cdot [M]\right)_t \xrightarrow{P} 0 \quad \Rightarrow \quad (V_n \cdot M)_t^* \xrightarrow{P} 0.$$

This follows immediately from Lemmas 18.10 and 18.12, together with the following approximation of progressive processes by predictable step processes.

Lemma 18.23 *(approximation) For any continuous semi-martingale $X = M + A$ and process $V \in L(X)$, there exist some processes $V_1, V_2, \ldots \in \mathcal{E}$, such that a.s., simultaneously for all $t > 0$,*

$$\left\{(V_n - V)^2 \cdot [M]\right\}_t + \left\{(V_n - V) \cdot A\right\}_t^* \to 0. \tag{14}$$

Proof: We may take $t = 1$, since we can then combine the processes V_n for disjoint, finite intervals to construct an approximating sequence on $\mathbb{R}_+$. It is further enough to consider approximations in the sense of convergence in probability, since the a.s. versions will then follow for a suitable sub-sequence. This allows us to perform the construction in steps, first approximating V by bounded and progressive processes V', next approximating each V' by continuous and adapted processes V'', and finally approximating each V'' by predictable step processes V'''.

The first and last steps being elementary, we may focus on the second step. Then let V be bounded. We need to construct some continuous, adapted processes V_n, such that (14) holds a.s. for $t = 1$. Since the V_n can be chosen to be uniformly bounded, we may replace the first term by $(|V_n - V| \cdot [M])_1$. Thus, it is enough to establish the approximation $(|V_n - V| \cdot A)_1 \to 0$ when the process A is non-decreasing, continuous, and adapted with $A_0 = 0$. Replacing A_t by $A_t + t$ if necessary, we may even take A to be strictly increasing.

To form the required approximations, we may introduce the inverse process $T_s = \sup\{t \geq 0;\, A_t \leq s\}$, and define for all $t, h > 0$

$$\begin{aligned} V_t^h &= h^{-1} \int_{T(A_t - h)}^t V\, dA \\ &= h^{-1} \int_{(A_t - h)_+}^{A_t} V(T_s)\, ds. \end{aligned}$$

Then Theorem 2.15 yields $V^h \circ T \to V \circ T$ as $h \to 0$, a.e. on $[0, A_1]$, and so by dominated convergence,

$$\int_0^1 \left|V^h - V\right| dA = \int_0^{A_1} \left|V^h(T_s) - V(T_s)\right| ds \to 0.$$

The processes V^h are clearly continuous. To prove their adaptedness, we note that the process $T(A_t - h)$ is adapted for every $h > 0$, by the definition of T. Since V is progressive, we further note that $V \cdot A$ is adapted and hence progressive. The adaptedness of $(V \cdot A)_{T(A_t - h)}$ now follows by composition. □

Though the class $L(X)$ of stochastic integrands is sufficient for most purposes, it is sometimes useful to allow integration of slightly more general processes. Given any continuous semi-martingale $X = M + A$, let $\hat{L}(X)$ denote the class of product-measurable processes V, such that $(V - \tilde{V}) \cdot [M] = 0$ and $(V - \tilde{V}) \cdot A = 0$ a.s. for some process $\tilde{V} \in L(X)$. For $V \in \hat{L}(X)$ we define $V \cdot X = \tilde{V} \cdot X$ a.s. The extension clearly enjoys all the previously established properties of stochastic integration.

We often need to see how semi-martingales, covariation processes, and stochastic integrals are transformed by a random time change. Then consider a non-decreasing, right-continuous family of finite optional times τ_s, $s \geq 0$, here referred to as a *finite random time change* τ. If even $\mathcal{F}$ is right-continuous, so is the *induced filtration* $\mathcal{G}_s = \mathcal{F}_{\tau_s}$, $s \geq 0$, by Lemma 9.3. A process X is said to be τ*-continuous*, if it is a.s. continuous on R_+ and constant on every interval $[\tau_{s-}, \tau_s]$, $s \geq 0$, where $\tau_{0-} = X_{0-} = 0$ by convention.

Theorem 18.24 *(random time change, Kazamaki) Let τ be a finite random time change with induced filtration $\mathcal{G}$, and let $X = M + A$ be a τ-continuous $\mathcal{F}$-semi-martingale. Then*

(i) *$X \circ \tau$ is a continuous $\mathcal{G}$-semi-martingale with canonical decomposition $M \circ \tau + A \circ \tau$, such that $[X \circ \tau] = [X] \circ \tau$ a.s.,*

(ii) *$V \in L(X)$ implies $V \circ \tau \in \hat{L}(X \circ \tau)$ and*

$$(V \circ \tau) \cdot (X \circ \tau) = (V \cdot X) \circ \tau \quad a.s. \tag{15}$$

Proof: (i) It is easy to check that the time change $X \mapsto X \circ \tau$ preserves continuity, adaptedness, monotonicity, and the local martingale property. In particular, $X \circ \tau$ is then a continuous $\mathcal{G}$-semi-martingale with canonical decomposition $M \circ \tau + A \circ \tau$. Since $M^2 - [M]$ is a continuous local martingale, so is the time-changed process $M^2 \circ \tau - [M] \circ \tau$, and we get

$$\begin{aligned}[X \circ \tau] &= [M \circ \tau] \\ &= [M] \circ \tau \\ &= [X] \circ \tau \quad \text{a.s.}\end{aligned}$$

If $V \in L(X)$, we also note that $V \circ \tau$ is product-measurable, since this is true for both V and τ.

(ii) Fixing any $t \geq 0$ and using the τ-continuity of X, we get

$$\begin{aligned}(1_{[0,t]} \circ \tau) \cdot (X \circ \tau) &= 1_{[0,\tau_t^{-1}]} \cdot (X \circ \tau) \\ &= (X \circ \tau)^{\tau_t^{-1}} \\ &= (1_{[0,t]} \cdot X) \circ \tau,\end{aligned}$$

which proves (15) when $V = 1_{[0,t]}$. If X has locally finite variation, the result extends by a monotone-class argument and monotone convergence to arbitrary $V \in L(X)$. In general, Lemma 18.23 yields some continuous, adapted processes $V_1, V_2, \ldots$, such that a.s.

$$\int (V_n - V)^2 d[M] + \int \big|(V_n - V)\, dA\big| \to 0.$$

By (15) the corresponding properties hold for the time-changed processes, and since the processes $V_n \circ \tau$ are right-continuous and adapted, hence progressive, we obtain $V \circ \tau \in \hat{L}(X \circ \tau)$.

Now assume instead that the approximating processes $V_1, V_2, \ldots$ are predictable step processes. Then (15) holds as before for each V_n, and the relation extends to V by Lemma 18.12. □

Next we consider stochastic integrals of processes depending on a parameter. Given a measurable space $(S, \mathcal{S})$, we say that a process V on $S \times \mathsf{R}_+$ is *progressive*[6], if its restriction to $S \times [0, t]$ is $\mathcal{S} \otimes \mathcal{B}_t \otimes \mathcal{F}_t$-measurable for every $t \geq 0$, where $\mathcal{B}_t = \mathcal{B}_{[0,t]}$. A simple version of the following result will be useful in Chapter 19.

Theorem 18.25 *(dependence on parameter, Doléans, Stricker & Yor) For measurable S, consider a continuous semi-martingale X and a progressive process $V_s(t)$ on $S \times \mathsf{R}_+$ such that $V_s \in L(X)$ for all $s \in S$. Then the process $Y_s(t) = (V_s \cdot X)_t$ has a version that is progressive and a.s. continuous at each $s \in S$.*

Proof: Let X have canonical decomposition of $M + A$, and let the processes V_s^n on $S \times \mathsf{R}_+$ be progressive and such that for any $t \geq 0$ and $s \in S$,

$$\Big\{(V_s^n - V_s)^2 \cdot [M]\Big\}_t + \Big\{(V_s^n - V_s) \cdot A\Big\}_t^* \xrightarrow{P} 0.$$

Then Lemma 18.12 yields $(V_s^n \cdot X - V_s \cdot X)_t^* \xrightarrow{P} 0$ for every s and t. Proceeding as in the proof of Proposition 5.32, we may choose a sub-sequence $\{n_k(s)\} \subset \mathsf{N}$ depending measurably on s, such that the same convergence holds a.s. along $\{n_k(s)\}$ for any s and t. Define $Y_{s,t} = \limsup_k (V_s^{n_k} \cdot X)_t$ whenever this is finite, and put $Y_{s,t} = 0$ otherwise. If the processes $(V_s^n \cdot X)_t$ have progressive versions on $S \times \mathsf{R}_+$ that are a.s. continuous for each s, then $Y_{s,t}$ is clearly a version of the process $(V_s \cdot X)_t$ with the same properties. We apply this argument in three steps.

First we reduce to the case of bounded, progressive integrands by taking $V^n = V 1\{|V| \leq n\}$. Next, we apply the transformation in the proof of Lemma 18.23, to reduce to the case of continuous and progressive integrands. Finally, we approximate any continuous, progressive process V by the predictable step processes $V_s^n(t) = V_s(2^{-n}[2^n t])$. The integrals $V_s^n \cdot X$ are then elementary, and the desired continuity and measurability are obvious by inspection. □

We turn to the related topic of functional representations. For motivation, we note that the construction of a stochastic integral $V \cdot X$ depends in a subtle way on the underlying probability measure P and filtration $\mathcal{F}$. Thus, we cannot expect a general representation $F(V, X)$ of the integral process $V \cdot X$. In view

[6]short for *progressively measurable*

of Proposition 5.32, we might still hope for a modified representation of the form $F(\mu, V, X)$, where $\mu = \mathcal{L}(V, X)$. Even this may be too optimistic, since the canonical decomposition of X also depends on $\mathcal{F}$.

Here we consider only a special case, sufficient for our needs in Chapter 32. Fixing any progressive functions σ^i_j and b^i of suitable dimension, defined on the path space $C_{\mathsf{R}_+,\mathsf{R}^d}$, we consider an adapted process X satisfying the stochastic differential equation

$$dX^i_t = \sigma^i_j(t, X)\, dB^j_t + b^i(t, X)\, dt, \tag{16}$$

where B is a Brownian motion in R^r. A detailed discussion of such equations appears in Chapter 32. Here we need only the simple fact from Lemma 32.1 that the coefficients $\sigma^i_j(t, X)$ and $b^i(t, X)$ are again progressive. Write $a^{ij} = \sigma^i_k \sigma^j_k$.

Proposition 18.26 *(functional representation) For any progressive functions σ, b, f of suitable dimension, there exists a measurable mapping*

$$F : \hat{\mathcal{M}}\Big(C_{\mathsf{R}_+,\,\mathsf{R}^d}\Big) \times C_{\mathsf{R}_+,\,\mathsf{R}^d} \to C_{\mathsf{R}_+,\,\mathsf{R}}, \tag{17}$$

such that whenever X solves (16) *with $\mathcal{L}(X) = \mu$ and $f^i(X) \in L(X^i)$ for all i, we have*

$$f^i(X) \cdot X^i = F(\mu, X) \;\; a.s.$$

Proof: From (16) we note that X is a semi-martingale with covariation processes $[X^i, X^j] = a^{ij}(X) \cdot \lambda$ and drift components $b^i(X) \cdot \lambda$. Hence, $f^i(X) \in L(X^i)$ for all i iff the processes $(f^i)^2 a^{ii}(X)$ and $f^i b^i(X)$ are a.s. Lebesgue integrable, which holds in particular when f is bounded. Letting $f_1, f_2, \ldots$ be progressive with

$$(f^i_n - f^i)^2 a^{ii}(X) \cdot \lambda + \Big|(f^i_n - f^i)\, b^i(X)\Big| \cdot \lambda \to 0, \tag{18}$$

we get by Lemma 18.12

$$\Big\{f^i_n(X) \cdot X^i - f^i(X) \cdot X^i\Big\}^*_t \xrightarrow{P} 0, \quad t \ge 0.$$

Thus, if $f^i_n(X) \cdot X^i = F_n(\mu, X)$ a.s. for some measurable mappings F_n as in (17), Proposition 5.32 yields a similar representation for the limit $f^i(X) \cdot X^i$.

As in the previous proof, we may apply this argument in steps, reducing first to the case where f is bounded, next to the case of continuous f, and finally to the case where f is a predictable step function. Here the first and last steps are again elementary. For the second step, we may now use the simpler approximation

$$f_n(t, x) = n \int_{(t-n^{-1})_+}^t f(s, x) ds, \qquad t \ge 0, \;\; n \in \mathsf{N}, \;\; x \in C_{\mathsf{R}_+,\mathsf{R}^d}.$$

By Theorem 2.15 we have $f_n(t, x) \to f(t, x)$ a.e. in t for each $x \in C_{\mathsf{R}_+,\mathsf{R}^d}$, and (18) follows by dominated convergence. □

Exercises

1. Show that if M is a local martingale and ξ is an $\mathcal{F}_0$-measurable random variable, then the process $N_t = \xi M_t$ is again a local martingale.

2. Show that every local martingale $M \geq 0$ with $EM_0 < \infty$ is a super-martingale. Also show by an example that M may fail to be a martingale. (*Hint:* Use Fatou's lemma. Then take $M_t = X_{t/(1-t)_+}$, where X is a Brownian motion starting at 1, stopped when it reaches 0.)

3. For a continuous local martingale M, show that M and $[M]$ have a.s. the same intervals of constancy. (*Hint:* For any $r \in \mathsf{Q}_+$, put $\tau = \inf\{t > r;\ [M]_t > [M]_r\}$. Then M^τ is a continuous local martingale on $[r, \infty)$ with quadratic variation 0, and so M^τ is a.s. constant on $[s, \tau]$. Use a similar argument in the other direction.)

4. For any continuous local martingales M_n starting at 0 and associated optional times τ_n, show that $(M_n)^*_{\tau_n} \xrightarrow{P} 0$ iff $[M_n]_{\tau_n} \xrightarrow{P} 0$. State the corresponding result for stochastic integrals.

5. Give examples of continuous semi-martingales $X_1, X_2, \ldots$, such that $X^*_n \xrightarrow{P} 0$, and yet $[X_n]_t \overset{P}{\not\to} 0$ for all $t > 0$. (*Hint:* Let B be a Brownian motion stopped at time 1, put $A_{k2^{-n}} = B_{(k-1)_+ 2^{-n}}$, and interpolate linearly. Define $X^n = B - A^n$.)

6. For a Brownian motion B and an optional time τ, show that $EB_\tau = 0$ when $E\tau^{1/2} < \infty$ and $EB^2_\tau = E\tau$ when $E\tau < \infty$. (*Hint:* Use optional sampling and Theorem 18.7.)

7. Deduce the first inequality of Proposition 18.9 from Proposition 18.17 and the classical Cauchy inequality.

8. For any continuous semi-martingales X, Y, show that $[X + Y]^{1/2} \leq [X]^{1/2} + [Y]^{1/2}$ a.s.

9. (Kunita & Watanabe) Let M, N be continuous local martingales, and fix any $p, q, r > 0$ with $p^{-1} + q^{-1} = r^{-1}$. Show that $\|[M, N]_t\|^2_{2r} \leq \|[M]_t\|_p \, \|[N]_t\|_q$ for all $t > 0$.

10. Let M, N be continuous local martingales with $M_0 = N_0 = 0$. Show that $M \perp\!\!\!\perp N$ implies $[M, N] \equiv 0$ a.s. Also show by an example that the converse is false. (*Hint:* Let $M = U \cdot B$ and $N = V \cdot B$ for a Brownian motion B and suitable $U, V \in L(B)$.)

11. Fix a continuous semi-martingale X, and let $U, V \in L(X)$ with $U = V$ a.s. on a set $A \in \mathcal{F}_0$. Show that $U \cdot X = V \cdot X$ a.s. on A. (*Hint:* Use Proposition 18.15.)

12. For a continuous local martingale M, let $U, U_1, U_2, \ldots$ and $V, V_1, V_2, \ldots \in L(M)$ with $|U_n| \leq V_n$, $U_n \to U$, $V_n \to V$, and $\{(V_n - V) \cdot M\}^*_t \xrightarrow{P} 0$ for all $t > 0$. Show that $(U_n \cdot M)_t \xrightarrow{P} (U \cdot M)_t$ for all t. (*Hint:* Write $(U_n - U)^2 \leq 2\,(V_n - V)^2 + 8\,V^2$, and use Theorem 1.23 and Lemmas 5.2 and 18.12.)

13. Let B be a Brownian bridge. Show that $X_t = B_{t\wedge 1}$ is a semi-martingale on R_+ with the induced filtration. (*Hint:* Note that $M_t = (1-t)^{-1} B_t$ is a martingale on $[0, 1)$, integrate by parts, and check that the compensator has finite variation.)

14. Show by an example that the canonical decomposition of a continuous semi-martingale may depend on the filtration. (*Hint:* Let B be a Brownian motion with induced filtration $\mathcal{F}$, put $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(B_1)$, and use the preceding result.)

15. Show by stochastic calculus that $t^{-p}B_t \to 0$ a.s. as $t \to \infty$, where B is a Brownian motion and $p > \frac{1}{2}$. (*Hint:* Integrate by parts to find the canonical decomposition. Compare with the L^1-limit.)

16. Extend Theorem 18.16 to a product of n semi-martingales.

17. Consider a Brownian bridge X and a bounded, progressive process V with $\int_0^1 V_t dt = 0$ a.s. Show that $E\int_0^1 V dX = 0$. (*Hint:* Integrate by parts to get $\int_0^1 V dX = \int_0^1 (V - U)dB$, where B is a Brownian motion and $U_t = (1-t)^{-1}\int_t^1 V_s ds$.)

18. Show that Proposition 18.17 remains valid for any finite optional times t and t_{nk} satisfying $\max_k(t_{nk} - t_{n,k-1}) \xrightarrow{P} 0$.

19. Let M be a continuous local martingale. Find the canonical decomposition of $|M|^p$ when $p \geq 2$. For such a p, deduce the second relation in Theorem 18.7. (*Hint:* Use Theorem 18.18. For the last part, use Hölder's inequality.)

20. Let M be a continuous local martingale with $M_0 = 0$ and $[M]_\infty \leq 1$. Show that $P\{\sup_t M_t \geq r\} \leq e^{-r^2/2}$ for all $r \geq 0$. (*Hint:* Consider the super-martingale $Z = \exp(cM - c^2[M]/2)$ for a suitable $c > 0$.)

21. Let X, Y be continuous semi-martingales. Fix a $t > 0$ and some partitions (t_{nk}) of $[0, t]$ with $\max_k(t_{nk} - t_{n,k-1}) \to 0$. Show that $\frac{1}{2}\sum_k (Y_{t_{nk}} + Y_{t_{n,k-1}})(X_{t_{nk}} - X_{t_{n,k-1}}) \xrightarrow{P} (Y \circ X)_t$. (*Hint:* Use Corollary 18.13 and Proposition 18.17.)

23. Say that a process is *predictable* if it is measurable with respect to the σ-field in $\mathsf{R}_+ \times \Omega$ induced by all predictable step processes. Show that every predictable process is progressive. Conversely, given any progressive process X and constant $h > 0$, show that the process $Y_t = X_{(t-h)_+}$ is predictable.

24. For a progressive process V and a non-decreasing, continuous, adapted process A, prove the existence of a predictable process $\tilde{V}$ with $|V - \tilde{V}| \cdot A = 0$ a.s. (*Hint:* Use Lemma 18.23.)

25. Use the preceding statement to deduce Lemma 18.23. (*Hint:* Begin with predictable V, using a monotone-class argument.)

26. Construct the stochastic integral $V \cdot M$ by approximation from elementary integrals, using Lemmas 18.10 and 18.23. Show that the resulting integral satisfies the relation in Theorem 18.11. (*Hint:* First let $M \in \mathcal{M}^2$ and $E(V^2 \cdot [M])_\infty < \infty$, and extend by localization.)

27. Let $(V, B) \stackrel{d}{=} (\tilde{V}, \tilde{B})$ for some Brownian motions B, $\tilde{B}$ on possibly different filtered probability spaces and some $V \in L(B)$, $\tilde{V} \in L(\tilde{B})$. Show that $(V,\ B,\ V \cdot B) \stackrel{d}{=} (\tilde{V},\ \tilde{B},\ \tilde{V} \cdot \tilde{B})$. (*Hint:* Argue as in the proof of Proposition 18.26.)

28. Let X be a continuous $\mathcal{F}$-semi-martingale. Show that X remains conditionally a semi-martingale given $\mathcal{F}_0$, and that the conditional quadratic variation agrees with $[X]$. Also show that if $V \in L(X)$, where $V = \sigma(Y)$ for some continuous process Y and measurable function σ, then V remains conditionally X-integrable, and the conditional integral agrees with $V \cdot X$. (*Hint:* Conditioning on $\mathcal{F}_0$ preserves martingales.)

Chapter 19

Continuous Martingales and Brownian Motion

Real and complex exponential martingales, characterization of Brownian motion, time-change reduction, harmonic and analytic maps, point polarity and transience, skew-product representation, orthogonality and independence, Brownian invariance, Brownian functionals and martingales, Gauss and Poisson reduction, iterated and multiple integrals, chaos and integral representation, change of measure, transformation of drift, preservation properties, Cameron–Martin theorem, uniform integrability, Wald's identity, removal of drift

Here we deal with a wide range of applications of stochastic calculus, the principal tools of which were introduced in the previous chapter. A recurrent theme is the notion of *exponential martingales*, which appear in both a real and a complex variety. Exploring the latter yields an easy access to Lévy's celebrated martingale characterization of Brownian motion, and to the basic time-change reduction of *isotropic* continuous local martingales to a Brownian motion. Applying the latter result to suitable compositions of a Brownian motion with harmonic or analytic maps, we obtain important information about Brownian motion in R^d. Similar methods can be used to analyze a variety of other transformations leading to Gaussian processes.

As a further application of the exponential martingales, we derive stochastic integral representations of Brownian functionals and martingales, and examine their relationship to the chaos expansions obtained by different methods in Chapter 14. In this context, we also note how the previously introduced multiple Wiener–Itô integrals can be expressed as iterated single Itô integrals. A related problem, of crucial importance for Chapter 32, is to represent a continuous local martingale with absolutely continuous covariation process in terms of stochastic integrals with respect to a suitable Brownian motion.

As a final major topic, we examine the transformations induced by absolutely continuous changes of the probability measure. Here the density process turns out to be a real exponential martingale, and any continuous local martingale remains a martingale under the new measure, apart from an additional drift term. This observation is useful for applications, where it is often employed to remove the drift of a given semi-martingale. The appropriate change of measure then depends on the given process, and it becomes important to find effective criteria for a proposed exponential process to be a true martingale.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_20

The present exposition may be regarded as a continuation of our discussion of martingales and Brownian motion in Chapters 9 and 14, respectively. Changes of time and measure are both important for the theory of stochastic differential equations, developed in Chapters 32–33. The time-change reductions of continuous martingales have counterparts for the point processes explored in Chapter 15, where the present Gaussian processes are replaced by suitable Poisson processes. The results about changes of measure are extended in Chapter 20 to the context of possibly discontinuous semi-martingales.

To begin our technical discussion of the mentioned ideas, we consider first the complex exponential martingales, which are closely related to the real versions appearing in Lemma 19.22.

Lemma 19.1 *(complex exponential martingales) Let M be a real, continuous local martingale with $M_0 = 0$. Then*

$$Z_t = \exp\Big(iM_t + \tfrac{1}{2}\,[M]_t\Big), \quad t \geq 0,$$

is a complex local martingale, satisfying

$$Z_t = 1 + i(Z \cdot M)_t \ \text{ a.s.}, \quad t \geq 0.$$

Proof: Applying Corollary 18.20 to the complex-valued semi-martingale $X_t = iM_t + \frac{1}{2}\,[M]_t$ and the entire function $f(z) = e^z$, we get

$$\begin{aligned} dZ_t &= Z_t\Big(dX_t + \tfrac{1}{2}\,d[X]_t\Big) \\ &= Z_t\Big(idM_t + \tfrac{1}{2}\,d[M]_t - \tfrac{1}{2}\,d[M]_t\Big) \\ &= iZ_t\,dM_t. \end{aligned}$$

□

We now explore the connection between continuous martingales and Gaussian processes. For a subset K of a Hilbert space H, let $\bar{K}$ denote the closed linear subspace generated by K.

Lemma 19.2 *(isonormal martingales) For a Hilbert space H, let $K \subset H$ with $\bar{K} = H$. Let M^h, $h \in K$, be continuous local martingales with $M_0^h = 0$, such that*

$$[M^h, M^k]_\infty = \langle h, k\rangle \ \text{ a.s.}, \quad h, k \in K. \tag{1}$$

Then there exists an isonormal Gaussian process $\zeta \perp\!\!\!\perp \mathcal{F}_0$ on H, such that

$$M_\infty^h = \zeta h \ \text{ a.s.}, \quad h \in K.$$

Proof: For any linear combination $h = u_1 h_1 + \cdots + u_n h_n$, let

$$N_t = u_1 M_t^{h_1} + \cdots + u_n M_t^{h_n}, \quad t \geq 0.$$

Then (1) yields

$$\begin{aligned} [N]_\infty &= \textstyle\sum_{j,k} u_j u_k \Big[M^{h_j}, M^{h_k}\Big]_\infty \\ &= \textstyle\sum_{j,k} u_j u_k \,\langle h_j, h_k\rangle = \|h\|^2. \end{aligned}$$

The process $Z = \exp(iN + \frac{1}{2}[N])$ is a.s. bounded, and so by Lemma 19.1 it is a uniformly integrable martingale. Writing $\xi = N_\infty$, we hence obtain for any $A \in \mathcal{F}_0$

$$\begin{aligned} PA &= E(Z_\infty; A) \\ &= E\Big\{\exp\Big(iN_\infty + \tfrac{1}{2}\,[N]_\infty\Big); A\Big\} \\ &= E(e^{i\xi}; A)\,\exp\Big(\tfrac{1}{2}\,\|h\|^2\Big). \end{aligned}$$

Since $u_1, \ldots, u_n$ were arbitrary, we conclude from Theorem 6.3 that the random vector $(M_\infty^{h_1}, \ldots, M_\infty^{h_n})$ is independent of $\mathcal{F}_0$ and centered Gaussian with covariances $\langle h_j, h_k \rangle$.

For any element $h = \sum_i a_i h_i$, define $\zeta h = \sum_i a_i M_\infty^{h_i}$. To see that ζh is a.s. independent of the representation of h, suppose that also $h = \sum_i b_i h_i$. Writing $c_i = a_i - b_i$, we get

$$\begin{aligned} E\Big(\textstyle\sum_i a_i M_\infty^{h_i} - \sum_i b_i M_\infty^{h_i}\Big)^2 &= E\Big(\textstyle\sum_i c_i M_\infty^{h_i}\Big)^2 \\ &= \textstyle\sum_{i,j} c_i c_j \,\langle h_i, h_j \rangle \\ &= \Big\|\textstyle\sum_i c_i h_i\Big\|^2 \\ &= \Big\|\textstyle\sum_i a_i h_i - \sum_i b_i h_i\Big\|^2 = 0, \end{aligned}$$

as desired. Note that ζ is centered Gaussian with $E(\zeta h\, \zeta k) = \langle h, k \rangle$. We may finally extend ζ by continuity to an isonormal Gaussian process on H. □

For a first application, we may establish a celebrated martingale characterization of Brownian motion.

Theorem 19.3 *(Brownian criterion, Lévy) For a continuous process $B = (B^1, \ldots, B^d)$ in R^d with $B_0 = 0$, these conditions are equivalent:*

(i) *B is an $\mathcal{F}$-Brownian motion,*

(ii) *B is a local $\mathcal{F}$-martingale with $[B^i, B^j]_t \equiv \delta_{ij} t$ a.s.*

Proof (OK): Assuming (ii), we introduce for fixed $s < t$ the continuous local martingales

$$M_r^i = B_{r\wedge t}^i - B_{r\wedge s}^i, \quad r \ge s, \ \ i \le d,$$

and conclude from Lemma 19.2 that the differences $B_t^i - B_s^i$ are i.i.d. $N(0, t-s)$ and independent of $\mathcal{F}_s$, which proves (i). □

The last result suggests that we might transform a general continuous local martingale M into a Brownian motion through a suitable random time change. In higher dimensions, this requires M to be *isotropic*, in the sense that a.s.

$$\begin{aligned} [M^i] &= [M^j], \\ [M^i, M^j] &= 0, \qquad i \ne j, \end{aligned}$$

which holds in particular when M is a Brownian motion in R^d. For continuous local martingales M in C, the condition is a.s. equivalent to $[M] = 0$, or

$$[\,\Re M] = [\Im M],$$
$$[\,\Re M, \Im M] = 0.$$

In the isotropic case, we refer to $[M^1] = \cdots = [M^d]$ or $[\,\Re M] = [\Im M]$ as the *rate process* of M.

Though the proof is straightforward when $[M]_\infty = \infty$ a.s., the general case requires a rather subtle extension of the filtered probability space. For two filtrations $\mathcal{F}$, $\mathcal{G}$ on Ω, we say that $\mathcal{G}$ is a *standard extension* of $\mathcal{F}$ if

$$\mathcal{F}_t \subset \mathcal{G}_t \underset{\mathcal{F}_t}{\perp\!\!\!\perp} \mathcal{F}, \quad t \geq 0.$$

This is precisely the condition needed to ensure that all adaptedness and conditioning properties will be preserved.

Theorem 19.4 *(time-change reduction, Doeblin, Dambis, Dubins & Schwarz) Let M be an isotropic, continuous local $\mathcal{F}$-martingale in R^d with $M_0 = 0$, and define*

$$\tau_s = \inf\Bigl\{t \geq 0;\ [M^1]_t > s\Bigr\}, \quad \mathcal{G}_s = \mathcal{F}_{\tau_s}, \qquad s \geq 0.$$

Then there exists a Brownian motion B in R^d w.r.t. a standard extension $\hat{\mathcal{G}}$ of $\mathcal{G}$, such that a.s.

$$B = M \circ \tau \ \text{ on } \ [0, [M^1]_\infty),$$
$$M = B \circ [M^1].$$

Proof (OK): We may take $d = 1$, the proof for $d > 1$ being similar. Introduce a Brownian motion $X \perp\!\!\!\perp \mathcal{F}$ with induced filtration $\mathcal{X}$, and put $\hat{\mathcal{G}}_t = \mathcal{G}_t \vee \mathcal{X}_t$. Since $\mathcal{G} \perp\!\!\!\perp \mathcal{X}$, it is clear that $\hat{\mathcal{G}}$ is a standard extension of both $\mathcal{G}$ and $\mathcal{X}$. In particular, X remains a Brownian motion under $\hat{\mathcal{G}}$. Now define

$$B_s = M_{\tau_s} + \int_0^s 1\{\tau_r = \infty\}\, dX_r, \quad s \geq 0. \tag{2}$$

Since M is τ-continuous by Proposition 18.6, Theorem 18.24 shows that the first term $M \circ \tau$ is a continuous $\mathcal{G}$-martingale, hence also a $\hat{\mathcal{G}}$-martingale, with quadratic variation

$$[M \circ \tau]_s \,=\, [M]_{\tau_s} \,=\, s \wedge [M]_\infty, \quad s \geq 0.$$

The second term in (2) has quadratic variation $s - s \wedge [M]_\infty$, and the covariation vanishes since $M \circ \tau \perp\!\!\!\perp X$. Thus, $[B]_s = s$ a.s., and so Theorem 19.3 shows that B is a $\hat{\mathcal{G}}$-Brownian motion. Finally, $B_s = M_{\tau_s}$ for $s < [M]_\infty$, which implies $M = B \circ [M]$ a.s. by the τ-continuity of M. $\square$

In two dimensions, isotropic martingales arise naturally by composition of a complex Brownian motion B with a possibly multi-valued analytic function f. For any continuous process X, we may clearly choose a continuous evolution of $f(X)$, as long as X avoids all singularities of f. Similar results hold for harmonic functions, which is especially useful in dimensions $d \geq 3$, where no analytic functions exist.

Theorem 19.5 *(harmonic and analytic maps, Lévy)*

(i) *For an harmonic function f on R^d, consider an isotropic, continuous local martingale M in R^d that a.s. avoids all singularities of f. Then $f(M)$ is a local martingale satisfying*

$$[f(M)] = |\nabla f(M)|^2 \cdot [M^1] \ \text{ a.s.}$$

(ii) *For an analytic function f, consider a complex, isotropic, continuous local martingale M that a.s. avoids all singularities of f. Then $f(M)$ is again a complex, isotropic local martingale satisfying*

$$[\,\Re f(M)\,] = |f'(M)|^2 \cdot [\,\Re M\,] \ \text{ a.s.}$$

If B is a Brownian motion and $f' \not\equiv 0$, then $[\,\Re f(B)]$ is a.s. unbounded and strictly increasing.

Proof: (i) Since M is isotropic, we get by Corollary 18.19

$$f(M) = f(M_0) + f_i' \cdot M^i + \tfrac{1}{2}\,\Delta f(M) \cdot [M^1].$$

Here the last term vanishes since f is harmonic, and so $f(M)$ is a local martingale. The isotropy of M also yields

$$\begin{aligned} [f(M)] &= \textstyle\sum_i \Big[f_i'(M) \cdot M^i\Big] \\ &= \textstyle\sum_i \Big\{f_i'(M)\Big\}^2 \cdot [M^1] \\ &= \Big|\nabla f(M)\Big|^2 \cdot [M^1]. \end{aligned}$$

(ii) Since f is analytic, Corollary 18.20 yields

$$f(M) = f(M_0) + f'(M) \cdot M + \tfrac{1}{2}\, f''(M) \cdot [M].$$

Since M is isotropic, the last term vanishes, and we get

$$\begin{aligned} [f(M)] &= \Big[f'(M) \cdot M\Big] \\ &= \Big\{f'(M)\Big\}^2 \cdot [M] = 0, \end{aligned}$$

which shows that $f(M)$ is again isotropic. Writing $M = X + iY$ and $f'(M) = U + iV$, we get

$$\begin{aligned} [\,\Re f(M)\,] &= [\,U \cdot X - V \cdot Y\,] \\ &= (\,U^2 + V^2) \cdot [X] \\ &= |f'(M)|^2 \cdot [\,\Re M\,]. \end{aligned}$$

Noting that f' has at most countably many zeros, unless it vanishes identically, we obtain by Fubini's theorem

$$E\lambda\Big\{t \ge 0;\ f'(B_t) = 0\Big\} = \int_0^\infty P\Big\{f'(B_t) = 0\Big\}\, dt = 0,$$

and so $[\Re f(B)] = |f'(B)|^2 \cdot \lambda$ is a.s. strictly increasing. To see that it is also a.s. unbounded, we note that $f(B)$ converges a.s. on the set $\{[\Re f(B)] < \infty\}$. However, $f(B)$ diverges a.s., since f is non-constant and the random walk $B_0, B_1, \ldots$ is recurrent by Theorem 12.2. $\square$

Combining the last two results, we may prove the polarity of single points when $d \geq 2$, and the transience of the process when $d \geq 3$. Note that the latter property is a continuous-time counterpart of Theorem 12.8 for random walks. Both results are important for the potential theory developed in Chapter 34. Define $\tau_a = \inf\{t > 0;\ B_t = a\}$.

Theorem 19.6 *(point polarity and transience, Lévy, Kakutani) For a Brownian motion B in R^d, we have*

(i) *for $d \geq 2$, $\tau_a = \infty$ a.s. for all $a \in \mathsf{R}^d$,*

(ii) *for $d \geq 3$, $|B_t| \to \infty$ a.s. as $t \to \infty$.*

Proof: (i) We may take $d = 2$, and hence choose B to be a complex Brownian motion. Using Theorem 19.5 (ii) with $f(z) = e^z$, we see that $M = e^B$ is a conformal local martingale with unbounded rate $[\Re M]$. By Theorem 19.4 we have $M - 1 = X \circ [\Re M]$ a.s. for a Brownian motion X, and since $M \neq 0$, it follows that X a.s. avoids -1. Hence, $\tau_{-1} = \infty$ a.s., and so by scaling and rotational symmetry we have $\tau_a = \infty$ a.s. for every $a \neq 0$. To extend this to $a = 0$, we use the Markov property at $h > 0$ to obtain

$$P_0\left\{\tau_0 \circ \theta_h < \infty\right\} = E_0 P_{B_h}\{\tau_0 < \infty\} = 0, \quad h > 0.$$

As $h \to 0$, we get $P_0\{\tau_0 < \infty\} = 0$, and so $\tau_0 = \infty$ a.s.

(ii) Here we may take $d = 3$. For any $a \neq 0$ we have $\tau_a = \infty$ a.s. by (i), and so by Theorem 19.5 (i) the process $M = |B - a|^{-1}$ is a continuous local martingale. By Fatou's lemma, M is then an L^1-bounded super-martingale, and so by Theorem 9.19 it converges a.s. toward some random variable ξ. Since clearly $M_t \xrightarrow{d} 0$ we have $\xi = 0$ a.s. $\square$

Combining (i) above with Theorem 17.11, we see that a complex, isotropic, continuous local martingale M with $M_0 = 0$ avoids every fixed point $a \neq 0$. Thus, Theorem 19.5 (ii) applies to any analytic function f with only isolated singularities. Since f is allowed to be multi-valued, the result applies even to functions with essential singularities, such as to $f(z) = \log(1+z)$. For a simple application, we consider the windings of planar Brownian motion about a fixed point.

Corollary 19.7 *(skew-product representation, Galmarino) Let B be a complex Brownian motion with $B_0 = 1$, and choose a continuous version of $V = \arg B$ with $V_0 = 0$. Then there exists a real Brownian motion $Y \perp\!\!\!\perp |B|$, such that a.s.*

$$V_t \equiv Y \circ \left(|B|^{-2} \cdot \lambda\right)_t, \quad t \geq 0.$$

Proof: Using Theorem 19.5 (ii) with $f(z) = \log(1+z)$, we see that $M_t = \log|B_t| + iV_t$ is an isotropic martingale with rate $[\,\Re M\,] = |B|^{-2}\cdot\lambda$. Hence, Theorem 19.4 yields a complex Brownian motion $Z = X + iY$ with $M = Z \circ [\,\Re M\,]$ a.s., and the assertion follows. □

For a non-isotropic, continuous local martingale M in R^d, there is no single random time change that reduces the process to a Brownian motion. However, we may transform the components M^i individually as in Theorem 19.4, to obtain a set of one-dimensional Brownian motions $B^1, \dots, B^d$. If the latter processes are independent, they may be combined into a d-dimensional Brownian motion $B = (B^1, \dots, B^d)$. We show that the required independence arises automatically, whenever the original components M^i are *strongly orthogonal*, in the sense that $[M^i, M^j] = 0$ a.s. for all $i \neq j$.

Proposition 19.8 *(multi-variate time change, Knight) Let $M^1, M^2, \dots$ be strongly orthogonal, continuous local martingales starting at 0. Then there exist some independent Brownian motions $B^1, B^2, \dots$ with*

$$M^k = B^k \circ [M^k] \ \ a.s., \quad k \in \mathsf{N}.$$

Proof: When $[M^k]_\infty = \infty$ a.s. for all k, the result is an easy consequence of Lemma 19.2. In general, choose some independent Brownian motions $X^1, X^2, \dots \perp\!\!\!\perp \mathcal{F}$ with induced filtration $\mathcal{X}$, and define

$$B^k_s = M^k(\tau^k_s) + X^k\big\{(s - [M^k]_\infty)_+\big\}, \quad s \geq 0,\ k \in \mathsf{N}.$$

Write $\psi_t = -\log(1-t)_+$, and put $\mathcal{G}_t = \mathcal{F}_{\psi_t} + \mathcal{X}_{(t-1)_+}$, $t \geq 0$. To check that $B^1, B^2, \dots$ have the desired joint distribution, we may take the $[M^k]$ to be bounded. Then the processes

$$N^k_t = M^k_{\psi_t} + X^k_{(t-1)_+}, \quad t \geq 0,$$

are strongly orthogonal, continuous $\mathcal{G}$-martingales with quadratic variations

$$[N^k]_t = [M^k]_{\psi_t} + (t-1)_+, \quad t \geq 0,$$

and we note that

$$B^k_s = N^k_{\sigma^k_s},$$
$$\sigma^k_s = \inf\big\{t \geq 0;\ [N^k]_t > s\big\}.$$

The assertion now follows from the result for $[M^k]_\infty = \infty$ a.s. □

We turn to some general invariance properties of a Brownian motion or bridge. Then for any processes U, V on $I = \mathsf{R}_+$ or $[0,1]$, we define[1]

$$\lambda V = \int_I V_t\, dt, \qquad \langle U, V \rangle = \int_I U_t\, V_t\, dt, \qquad \|V\|_2 = \langle V, V \rangle^{1/2}.$$

[1]The inner product $\langle \cdot, \cdot \rangle$ must not be confused with the predictable covariation introduced in Chapter 20.

A process B on $[0,1]$ is said to be a Brownian bridge w.r.t. a filtration $\mathcal{F}$, if it is $\mathcal{F}$-adapted and such that, conditionally on $\mathcal{F}_t$ for every $t \in [0,1]$, it is a Brownian bridge from (t, B_t) to $(1,0)$. We anticipate the fact that an $\mathcal{F}$-Brownian bridge in R^d is a continuous semi-martingale on $[0,1]$. The associated martingale component is clearly a standard Brownian motion.

Theorem 19.9 *(Brownian invariance)*

(i) *Consider an $\mathcal{F}$-Brownian motion X and some $\mathcal{F}$-predictable processes V^t on R_+, $t \geq 0$, such that*

$$\langle V^s, V^t \rangle = s \wedge t \ \ a.s., \quad s,t \geq 0.$$

Then $Y_t = (V^t \cdot X)_\infty$, $t \geq 0$, is a version of a Brownian motion.

(ii) *Consider an $\mathcal{F}$-Brownian bridge X and some $\mathcal{F}$-predictable processes V^t on $[0,1]$, $t \in [0,1]$, such that*

$$\lambda V^t = t, \quad \langle V^s, V^t \rangle = s \wedge t \ \ a.s., \qquad s,t \in [0,1].$$

Then $Y_t = (V^t \cdot X)_1$, $t \in [0,1]$, is a version of a Brownian bridge.

The existence of the integrals $V^t \cdot X$ should be regarded as part of the assertions. In case of (ii) we need to convert the relevant Brownian-bridge integrals into continuous martingales. Then for any Lebesgue integrable process V on $[0,1]$, we define

$$\bar{V}_t = (1-t)^{-1} \int_t^1 V_s\, ds, \quad t \in [0,1].$$

Lemma 19.10 *(Brownian bridge integral)*

(i) *Let X be an $\mathcal{F}$-Brownian bridge with martingale component B, and consider an $\mathcal{F}$-predictable process V on $[0,1]$, such that $E^{\mathcal{F}_0} \int_0^1 V^2 < \infty$ a.s. and λV is $\mathcal{F}_0$-measurable. Then*

$$\int_0^1 V_t\, dX_t = \int_0^1 (V_t - \bar{V}_t)\, dB_t \ \ a.s.$$

(ii) *For any processes $U, V \in L^2[0,1]$, we have $\bar{U}, \bar{V} \in L^2[0,1]$ and*

$$\int_0^1 (U_t - \bar{U}_t)(V_t - \bar{V}_t)\, dt = \int_0^1 U_t V_t\, dt - \lambda U \cdot \lambda V.$$

Proof: (ii) We may take $U = V$, since the general case will then follow by polarization. Writing $R_t = \int_t^1 V_s\, ds = (1-t)\bar{V}_t$ and integrating by parts, we get on $[0,1)$

$$\begin{aligned} \int \bar{V}_t^2\, dt &= \int (1-t)^{-2} R_t^2\, dt \\ &= (1-t)^{-1} R_t^2 + 2 \int (1-t)^{-1} R_t V_t\, dt \\ &= (1-t)\, \bar{V}_t^2 + 2 \int \bar{V}_t V_t\, dt. \end{aligned}$$

For bounded V, we conclude that

$$\int_0^1 (V_t - \bar{V}_t)^2\, dt = \int_0^1 V_t^2\, dt - (\lambda V)^2$$
$$= \int_0^1 (V_t - \lambda V)^2\, dt, \qquad (3)$$

and so by Minkowski's inequality

$$\begin{aligned} \|\bar{V}\|_2 &\le \|V\|_2 + \|V - \bar{V}\|_2 \\ &= \|V\|_2 + \|V - \lambda V\|_2 \\ &\le 2\,\|V\|_2, \end{aligned}$$

which extends to the general case by monotone convergence. This gives $\bar{V} \in L^2$, and the asserted relation follows from (3) by dominated convergence.

(i) The process $M_t = X_t/(1-t)$ is clearly a continuous martingale on $[0,1)$, and so $X_t = (1-t)M_t$ is a continuous semi-martingale on the same interval. Integrating by parts, we get for any $t \in [0,1)$

$$\begin{aligned} dX_t &= (1-t)\, dM_t - M_t\, dt \\ &= dB_t - M_t\, dt, \end{aligned}$$

and also

$$\begin{aligned} \int_0^t V_s M_s\, ds &= M_t \int_0^t V_s\, ds - \int_0^t dM_s \int_0^s V_r\, dr \\ &= \int_0^t dM_s \int_s^1 V_r\, dr - M_t \int_t^1 V_s\, ds \\ &= \int_0^t \bar{V}_s\, dB_s - X_t\, \bar{V}_t. \end{aligned}$$

Hence by combination,

$$\int_0^t V_s\, dX_s = \int_0^t (V_s - \bar{V}_s)\, dB_s + X_t\, \bar{V}_t, \quad t \in [0,1). \qquad (4)$$

By Cauchy's inequality and dominated convergence, we get as $t \to 1$

$$\begin{aligned} \left(E^{\mathcal{F}_0}|X_t\, \bar{V}_t|\right)^2 &\le E^{\mathcal{F}_0}(M_t^2)\, E^{\mathcal{F}_0}\Big(\int_t^1 |V_s|\, ds\Big)^2 \\ &\le (1-t)\, E^{\mathcal{F}_0}(M_t^2)\, E^{\mathcal{F}_0}\int_t^1 V_s^2\, ds \\ &= t\, E^{\mathcal{F}_0}\int_t^1 V_s^2\, ds \to 0, \end{aligned}$$

and so by dominated convergence $X_t\, \bar{V}_t \xrightarrow{P} 0$. By (4) and Corollary 18.13, it remains to note that $\bar{V}$ is B-integrable, which holds since $\bar{V} \in L^2$ by (ii). $\square$

Proof of Theorem 19.9: (i) Use Lemma 19.2.

(ii) By Lemma 19.10 (i), we have a.s.

$$Y_t = \int_0^1 (V_r^t - \bar{V}_r^t)\, dB_r, \quad t \in [0,1],$$

for the Brownian motion $B = X - \hat{X}$. Furthermore, Lemma 19.10 (ii) yields

$$\begin{aligned}\int_0^1 \left(V_r^s - \bar{V}_r^s\right)\left(V_r^t - \bar{V}_r^t\right) dr &= \int_0^1 V_r^s\, V_r^t\, dr - \lambda V^s \cdot \lambda V^t \\ &= s \wedge t - st.\end{aligned}$$

The assertion now follows by Lemma 19.2. □

Next we consider a basic stochastic integral representation of martingales with respect to a Brownian filtration.

Theorem 19.11 *(Brownian martingales) Let $B = (B^1, \dots, B^d)$ be a Brownian motion in R^d with complete, induced filtration $\mathcal{F}$. Then every local $\mathcal{F}$-martingale M is a.s. continuous, and there exist some $(P \times \lambda)$-a.e. unique processes $V^1, \dots, V^d \in L(B^1)$, such that*

$$M = M_0 + \sum_{k \le d} V^k \cdot B^k \quad a.s. \tag{5}$$

The proof is based on a similar representation of Brownian functionals:

Lemma 19.12 *(Brownian functionals, Itô) Let $B = (B^1, \dots, B^d)$ be a Brownian motion in R^d, and let ξ be a B-measurable random variable with $E\xi = 0$ and $E\xi^2 < \infty$. Then there exist some $(P \times \lambda)$-a.e. unique processes $V^1, \dots, V^d \in L(B^1)$, such that*[2] *a.s.*

$$E\int_0^\infty |V_t|^2\, dt < \infty, \qquad \xi = \sum_{k \le d} \int_0^\infty V^k\, dB^k.$$

Proof (Dellacherie): Let H be the Hilbert space of B-measurable random variables $\xi \in L^2$ with $E\xi = 0$, and write K for the subspace of elements ξ admitting the desired representation $\sum_k (V^k \cdot B^k)_\infty$. For such a ξ we get $E\xi^2 = E\sum_k \{(V^k)^2 \cdot \lambda\}_\infty$, which implies the asserted uniqueness. By the obvious completeness of $L(B^1)$, the same formula shows that K is closed. To obtain $K = H$, we need to show that any $\xi \in H \ominus K$ vanishes a.s.

Then fix any non-random functions $u^1, \dots, u^d \in L^2(\mathsf{R})$. Put $M = \sum_k u^k \cdot B^k$, and define the process Z as in Lemma 19.1. Then Proposition 18.14 yields

$$\begin{aligned}Z - 1 &= i\, Z \cdot M \\ &= i \sum_{k \le d} (Z u^k) \cdot B^k,\end{aligned}$$

and so $\xi \perp (Z_\infty - 1)$, or

$$E\,\xi \exp\Big\{i \textstyle\sum_k (u^k \cdot B^k)_\infty\Big\} = 0.$$

Specializing to step functions u^k and using Theorem 6.3, we get

$$E\Big\{\xi;\, (B_{t_1}, \dots, B_{t_n}) \in C\Big\} = 0, \quad t_1, \dots, t_n \in \mathsf{R}_+,\;\; C \in \mathcal{B}^n,\;\; n \in \mathsf{N},$$

[2]The moment condition is essential, since the existence would otherwise be trivial and the uniqueness would fail. This is similar to the situation for the Skorohod embedding in Theorem 22.1.

which extends by a monotone-class argument to $E(\xi; A) = 0$ for any $A \in \mathcal{F}_\infty$. Thus, $\xi = E(\xi \,|\, \mathcal{F}_\infty) = 0$ a.s. □

Proof of Theorem 19.11: We may clearly take $M_0 = 0$, and by suitable localization we may assume that M is uniformly integrable. Then M_∞ exists in $L^1(\mathcal{F}_\infty)$ and may be approximated in L^1 by some random variables $\xi_1, \xi_2, \ldots \in L^2(\mathcal{F}_\infty)$ with $E\xi_n = 0$. The martingales $M_t^n = E(\xi_n | \mathcal{F}_t)$ are a.s. continuous by Lemma 19.12, and Proposition 9.16 yields for any $\varepsilon > 0$

$$\begin{aligned} P\Big\{(\Delta M)^* > 2\varepsilon\Big\} &\le P\Big\{(M^n - M)^* > \varepsilon\Big\} \\ &\le \varepsilon^{-1} E|\,\xi_n - M_\infty| \to 0. \end{aligned}$$

Hence, $(\Delta M)^* = 0$ a.s., and so M is a.s. continuous. The remaining assertions now follow by localization from Lemma 19.12. □

We turn to the converse problem of finding a Brownian motion B satisfying (5) for given processes V^k. This result plays a crucial role in Chapter 32.

Theorem 19.13 *(integral representation, Doob) Let M be a continuous local $\mathcal{F}$-martingale in R^d with $M_0 = 0$, such that*

$$[M^i, M^j] = \sum_{k \le n} V_k^i V_k^j \cdot \lambda \;\; a.s., \quad i, j \le d,$$

for some $\mathcal{F}$-progressive processes V_k^i, $i \le d$, $k \le n$. Then there exists a Brownian motion $B = (B^1, \ldots, B^d)$ in R^d w.r.t. a standard extension $\mathcal{G}$ of $\mathcal{F}$, such that

$$M^i = \sum_{k \le n} V_k^i \cdot B^k \;\; a.s., \quad i \le d.$$

Proof: For any $t \ge 0$, let N_t and R_t be the null and range spaces of the matrix V_t, and write $N_t^\perp$ and $R_t^\perp$ for their orthogonal complements. Denote the corresponding orthogonal projections by π_{N_t}, π_{R_t}, $\pi_{N_t^\perp}$, and $\pi_{R_t^\perp}$, respectively. Note that V_t is a bijection from $N_t^\perp$ to R_t, and write V_t^{-1} for the inverse mapping from R_t to $N_t^\perp$. All these mappings are clearly Borel-measurable functions of V_t, and hence again progressive.

Now introduce a Brownian motion $X \perp\!\!\!\perp \mathcal{F}$ in R^n with induced filtration $\mathcal{X}$, and note that $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{X}_t$, $t \ge 0$, is a standard extension of both $\mathcal{F}$ and $\mathcal{X}$. Thus, V remains $\mathcal{G}$-progressive, and the martingale properties of M and X remain valid for $\mathcal{G}$. In R^n we introduce the local $\mathcal{G}$-martingale

$$B = V^{-1}\pi_R \cdot M + \pi_N \cdot X.$$

The covariation matrix of B has density

$$\begin{aligned} (V^{-1}\pi_R)\, VV'\, (V^{-1}\pi_R)' + \pi_N \pi_N' &= \pi_{N^\perp}\pi_{N^\perp}' + \pi_N\pi_N' \\ &= \pi_{N^\perp} + \pi_N = I, \end{aligned}$$

and so B is a Brownian motion by Theorem 19.3. Furthermore, the process $\pi_{R^\perp} \cdot M$ vanishes a.s., since its covariation matrix has density $\pi_{R^\perp} V V' \pi'_{R^\perp} = 0$. Hence, Proposition 18.14 yields

$$\begin{aligned} V \cdot B &= V V^{-1} \pi_R \cdot M + V \pi_N \cdot Y \\ &= \pi_R \cdot M \\ &= (\pi_R + \pi_{R^\perp}) \cdot M = M. \end{aligned}$$ □

The following joint extension of Theorems 15.15 and 19.2 will be needed in Chapter 27.

Proposition 19.14 *(Gauss and Poisson reduction) Consider a continuous local martingale M in R^d, a ql-continuous, K-marked point process ξ on $(0, \infty)$ with compensator $\hat{\xi}$, a predictable process $V\colon \mathsf{R}_+ \times K \to \hat{S}$, and a progressive process $U\colon T \times \mathsf{R}_+ \to \mathsf{R}^d$, where K, S are Borel and $\hat{S} = S \cup \{\Delta\}$ with $\Delta \notin S$. Let the random measure μ on S and process ρ on T^2 be $\mathcal{F}_0$-measurable with*

$$\mu = \hat{\xi} \circ V^{-1}, \qquad \rho_{s,t} = \textstyle\sum_{i,j} \displaystyle\int_0^\infty U^i_{s,r} U^j_{t,r}\, d[M^i, M^j]_r, \quad s, t \in T,$$

and define the point process η on S and process X on T by

$$\eta = \xi \circ V^{-1}, \qquad X_t = \textstyle\sum_i \displaystyle\int_0^\infty U^i_{t,r}\, dM^i_r, \quad t \in T.$$

Then conditionally on $\mathcal{F}_0$, we have

(i) *η and X are independent,*

(ii) *η is Poisson on S with intensity measure μ,*

(iii) *X is centered Gaussian with covariance function ρ.*

Proof: For any constants $c_1, \dots, c_m \in \mathsf{R}$, times $t_1, \dots, t_m \in T$, and disjoint sets $B_1, \dots, B_n \in \mathcal{S}$ with $\mu B_j < \infty$, consider the processes

$$\begin{aligned} N_t &= \textstyle\sum_k c_k \sum_i \displaystyle\int_0^t U^i_{t_k,r}\, dM^i_r, \\ Y^k_t &= \int_S \int_0^{t+} 1_{B_k}(V_{s,x})\, \xi(ds\, dx), \quad t \ge 0, \ k \le n. \end{aligned}$$

Next define for any $u_1, \dots, u_n \ge 0$ the exponential local martingales

$$\begin{aligned} Z^0_t &= \exp\Big(iN_t + \tfrac{1}{2}[N, N]_t \Big), \\ Z^k_t &= \exp\Big\{ -u_k Y^k_t + (1 - e^{-u_k}) \hat{Y}^k_t \Big\}, \quad t \ge 0, \ k \le n, \end{aligned}$$

where the local martingale property holds for Z^0 by Lemma 19.1 and for $Z^1, \dots, Z^n$ by Theorem 20.8, applied to the processes $A^k_t = (1 - e^{-u_k})(\hat{Y}^k_t - Y^k_t)$. The same property holds for the product $Z_t = \prod_k Z^k_t$, since the Z^k are strongly orthogonal by Theorems 20.4 and 20.6. Furthermore,

$$\begin{aligned} N_\infty &= \textstyle\sum_k c_k X_{t_k}, \qquad Y^k_\infty = \eta B_k, \\ [N, N]_\infty &= \textstyle\sum_{h,k} c_h c_k \sum_{i,j} \displaystyle\int_0^\infty U^i_{t_h,r} U^j_{t_k,r}\, d[M^i, M^j]_r \\ &= \textstyle\sum_{h,k} c_h c_k\, \rho_{t_h,t_k}, \\ \hat{Y}^k_\infty &= \int_S \int_K 1_{B_k}(V_{s,x})\, \hat{\xi}(ds\, dx) = \mu B_k, \quad k \le n. \end{aligned}$$

The product Z remains a local martingale with respect to the conditional probability measure $P_A = P(\cdot\,|A)$, for any $A \in \mathcal{F}_0$ with $PA > 0$. Choosing A such that ρ_{t_h,t_k} and μB_k are bounded on A, we see that even Z becomes bounded on A, and hence is a uniformly integrable P_A-martingale. In particular, $E(Z_\infty|A) = 1$ or $E(Z_\infty; A) = PA$, which extends immediately to arbitrary $A \in \mathcal{F}_0$. Hence, $E(Z_\infty|\,\mathcal{F}_0) = 1$, and we get

$$E\Big\{\exp\Big(i\sum_k c_k X_{t_k} - \sum_k u_k\,\eta B_k\Big)\Big|\,\mathcal{F}_0\Big\}$$
$$= \exp\Big\{-\tfrac{1}{2}\sum_{h,k} c_h c_k\,\rho_{t_h,t_k} - \sum_k (1-e^{-u_k})\,\mu B_k\Big\}.$$

By Theorem 6.3, the variables $X_{t_1},\dots,X_{t_m}$ and $\eta B_1,\dots,\eta B_n$ have then the required joint conditional distribution, and the assertion follows by a monotone-class argument. □

For a similar purpose in Chapter 27, we also need a joint extension of Theorems 10.27 and 19.2, where the ordinary compensator is replaced by its discounted version.

Proposition 19.15 *(discounted reduction maps) For $j \in \mathbb{N}$, let (τ_j,χ_j) be orthogonal, adapted pairs as in Theorem 10.27 with discounted compensators ζ_j, and let $Y_j\colon \mathbb{R}_+ \times K_j \to S_j$ be predictable with $\zeta_j \circ Y_j^{-1} \le \mu_j$ a.s. for some $\mathcal{F}_0$-measurable random distributions μ_j on S_j. Put $\gamma_j = Y_j(\tau_j,\chi_j)$, and define X as in Lemma 19.14 for a continuous local martingale M in $\mathbb{R}^d$ and some predictable processes U_t, $t \in S$, such that the process ρ on S^2 is $\mathcal{F}_0$-measurable. Then conditionally on $\mathcal{F}_0$, we have*

- (i) *$X, \gamma_1, \gamma_2, \dots$ are independent,*
- (ii) *$\mathcal{L}(\gamma_j) = \mu_j$ for all j,*
- (iii) *X is centered Gaussian with covariance function ρ.*

Proof: For $V_1,\dots,V_m$ as before, we define the associated martingales M_1, $\dots, M_m$ as in Lemma 10.25. Further introduce the exponential local martingale $Z = \exp(iN + \frac{1}{2}[N])$, where N is such as in the proof of Lemma 19.14. Then $(Z-1)\prod_j M_j$ is conditionally a uniformly integrable martingale, and we get as in Theorem 10.27

$$E^{\mathcal{F}_0}(Z_\infty - 1)\prod_j \Big\{1_{B_j}(\gamma_j) - \mu_j B_j\Big\} = 0.$$

Combining with the same relation without the factor $Z_\infty - 1$ and proceeding recursively, we get as before

$$E^{\mathcal{F}_0}\exp(iN_\infty)\prod_j 1_{B_j}(\gamma_j) = \exp\Big(-\tfrac{1}{2}[N]_\infty\Big)\prod_j \mu_j B_j.$$

Applying Theorem 6.3 to the bounded measure $\tilde{P} = \prod_j 1_{B_j}(\gamma_j)\cdot P^{\mathcal{F}_0}$, we conclude that

$$P^{\mathcal{F}_0}\Big\{X \in A;\ \gamma_j \in B_j,\ j \le m\Big\} = \nu_\rho(A)\prod_j \mu_j B_j,$$

where ν_ρ denotes the centered Gaussian distribution on R^T with covariance function ρ. $\square$

Next we show how the multiple Wiener–Itô integrals of Chapter 14 can be expressed as iterated single Itô integrals. Then introduce the *tetrahedral* sets

$$\Delta_n = \Big\{(t_1, \dots, t_n) \in \mathsf{R}_+^n;\ t_1 < \cdots < t_n\Big\}, \quad n \in \mathsf{N}.$$

Given a function $f \in L^2(\mathsf{R}_+^n, \lambda^n)$, we write $\hat f = n!\,\tilde f 1_{\Delta_n}$, where $\tilde f$ is the symmetrization of f defined in Chapter 14. For any isonormal Gaussian process ζ on $L^2(\lambda)$, we may introduce an associated Brownian motion B on R_+ satisfying $B_t = \zeta 1_{[0,t]}$ for all $t \ge 0$.

Theorem 19.16 *(multiple and iterated integrals) Let B be a Brownian motion on R_+ generated by an isonormal Gaussian process ζ on $L^2(\lambda)$. Then for any $f \in L^2(\lambda^n)$, we have*

$$\zeta^n f = \int dB_{t_n} \int dB_{t_{n-1}} \cdots \int \hat f(t_1, \dots, t_n)\, dB_{t_1} \quad a.s. \tag{6}$$

Though a formal verification is easy, the existence of the integrals on the right depends in a subtle way on the choice of suitable versions in each step. The existence of such versions is regarded as part of the assertion.

Proof: We prove by induction that the iterated integral

$$V^k_{t_{k+1},\dots,t_n} = \int dB_{t_k} \int dB_{t_{k-1}} \cdots \int \hat f(t_1, \dots, t_n)\, dB_{t_1}$$

exists for almost all $t_{k+1}, \dots, t_n$, and that V^k has a version supported by Δ_{n-k}, which is progressive in the variable t_{k+1} for fixed $t_{k+2}, \dots, t_n$. We further need to establish the isometry

$$E\Big(V^k_{t_{k+1},\dots,t_n}\Big)^2 = \int \cdots \int \Big\{\hat f(t_1, \dots, t_n)\Big\}^2 dt_1 \cdots dt_k, \tag{7}$$

which allows us in the next step to define $V^{k+1}_{t_{k+2},\dots,t_n}$ for almost all $t_{k+2}, \dots, t_n$.

The integral $V^0 = \hat f$ has clearly the stated properties. Now suppose that a version of the integral $V^{k-1}_{t_k,\dots,t_n}$ has been constructed with the desired properties. For any $t_{k+1}, \dots, t_n$ such that (7) is finite, Theorem 18.25 shows that the process

$$X^k_{t,t_{k+1},\dots,t_n} = \int_0^t V^{k-1}_{t_k,\dots,t_n}\, dB_{t_k}, \quad t \ge 0,$$

has a progressive version that is a.s. continuous in t for fixed $t_{k+1}, \dots, t_n$. By Proposition 18.15 we obtain

$$V^k_{t_{k+1},\dots,t_n} = X^k_{t_{k+1},t_{k+1},\dots,t_n} \ \text{ a.s.}, \quad t_{k+1}, \dots, t_n \ge 0,$$

and the progressivity carries over to V^k, regarded as a process in t_{k+1} for fixed $t_{k+2}, \dots, t_n$. Since V^{k-1} is supported by Δ_{n-k+1}, we may choose X^k to be supported by $\mathsf{R}_+ \times \Delta_{n-k}$, which ensures that V^k will be supported by Δ_{n-k}. Finally, formula (7) for V^{k-1} yields

$$\begin{aligned} E\Big(V^k_{t_{k+1},\dots,t_n}\Big)^2 &= E\int \Big(V^{k-1}_{t_k,\dots,t_n}\Big)^2 dt_k \\ &= \int\cdots\int \Big\{\hat f(t_1,\dots,t_n)\Big\}^2 dt_1\cdots dt_k. \end{aligned}$$

To prove (6), we note that the right-hand side is linear and L^2-continuous in f. Furthermore, the two sides agree for indicator functions of rectangles in Δ_n. The equality extends by a monotone-class argument to arbitrary indicator functions in Δ_n, and the further extension to $L^2(\Delta_n)$ is immediate. It remains to note that $\zeta^n f = \zeta^n \tilde f = \zeta^n \hat f$ for any $f \in L^2(\lambda^n)$. □

For a Brownian functional with mean 0 and finite variance, we have established both a chaos expansion in Theorem 14.26 and a stochastic integral representation in Lemma 19.12. We proceed to examine how the two formulas are related. For any $f \in L^2(\lambda^n)$, we define $f_t(t_1,\dots,t_{n-1}) = f(t_1,\dots,t_{n-1},t)$, and write $\zeta^{n-1}f(t) = \zeta^{n-1}f_t$ when $\|f_t\| < \infty$.

Proposition 19.17 *(chaos and integral representations) Let ζ be an isonormal Gaussian process on $L^2(\lambda)$, generating a Brownian motion B, and let $\xi \in L^2(\zeta)$ with chaos expansion $\sum_{n>0}\zeta^n f_n$. Then a.s.*

$$\xi = \int_0^\infty V_t\,dB_t, \qquad V_t = \sum_{n>0}\zeta^{n-1}\hat f_n(t), \quad t \ge 0.$$

Proof: Proceeding as in the previous proof, we get for any $m \in \mathsf{N}$

$$\begin{aligned} \int \sum_{n\ge m} E\Big\{\zeta^{n-1}\hat f_n(t)\Big\}^2 dt &= \sum_{n\ge m}\|\hat f_n\|^2 \\ &= \sum_{n\ge m} E(\zeta^n f_n)^2 < \infty. \end{aligned} \tag{8}$$

Since the integrals $\zeta^n f$ are orthogonal for different n, the series for V_t converges in L^2 for almost every $t \ge 0$. On the exceptional null set, we may redefine $V_t = 0$. Choosing progressive versions of the integrals $\zeta^{n-1}\hat f_n(t)$, as before, we see from the proof of Corollary 5.33 that even the sum V can be chosen to be progressive. Applying (8) with $m = 1$ gives $V \in L(B)$.

Using Theorem 19.16, we get by a formal calculation

$$\begin{aligned} \xi &= \sum_{n\ge 1}\zeta^n f_n \\ &= \sum_{n\ge 1}\int \zeta^{n-1}\hat f_n(t)\,dB_t \\ &= \int dB_t \sum_{n\ge 1}\zeta^{n-1}\hat f_n(t) \\ &= \int V_t\,dB_t. \end{aligned}$$

To justify the interchange of integration and summation, we may use (8) and let $m \to \infty$ to obtain

$$E\Big\{\int dB_t \sum_{n\ge m} \zeta^{n-1}\hat f_n(t)\Big\}^2 = \int \sum_{n\ge m} E\Big\{\zeta^{n-1}\hat f_n(t)\Big\}^2 dt$$
$$= \sum_{n\ge m} E(\zeta^n f_n)^2 \to 0. \qquad \square$$

Now let P, Q be probability measures on a common measurable space $(\Omega, \mathcal{A})$, endowed with a right-continuous and P-complete filtration $(\mathcal{F}_t)$. When $Q \ll P$ on $\mathcal{F}_t$, we write Z_t for the corresponding density, so that $Q = Z_t \cdot P$ on $\mathcal{F}_t$. Since the martingale property depends on the choice of probability measure, we need to distinguish between P-martingales and Q-martingales. Let E_P denote integration with respect to P, and write $\mathcal{F}_\infty = \bigvee_t \mathcal{F}_t$. We summarize the basic properties of such measure transformations. Recall from Theorem 9.28 that any $\mathcal{F}$-martingale has an rcll version.

Lemma 19.18 *(change of measure) Let P, Q be probability measures on a measurable space Ω with a right-continuous, P-complete filtration $\mathcal{F}$, such that*

$$Q = Z_t \cdot P \ \text{ on } \mathcal{F}_t, \quad t \ge 0,$$

for an adapted process Z. Then for any adapted process X,

(i) *Z is a P-martingale,*

(ii) *Z is uniformly integrable $\Leftrightarrow$ $Q \ll P$ on $\mathcal{F}_\infty$,*

(iii) *X is a Q-martingale $\Leftrightarrow$ XZ is a P-martingale.*

When Z is right-continuous and X is rcll, we have also

(iv) *$Q = Z_\tau \cdot P$ on $\mathcal{F}_\tau \cap \{\tau < \infty\}$ for any optional time τ,*

(v) *X is a local Q-martingale $\Leftrightarrow$ XZ is a local P-martingale,*

(vi) $\inf_{s\le t} Z_s > 0$ *a.s. Q,* $\quad t > 0$.

Proof: (iii) For any adapted process X, we note that X_t is Q-integrable iff $X_t Z_t$ is P-integrable. When this holds for all t, the Q-martingale property of X becomes
$$\int_A X_s\, dQ = \int_A X_t\, dQ, \quad A \in \mathcal{F}_s,\ s < t.$$
By the definition of Z, this is equivalent to

$$E_P\Big(X_s Z_s; A\Big) = E_P\Big(X_t Z_t; A\Big), \quad A \in \mathcal{F}_s,\ s < t,$$

which means that XZ is a P-martingale.

(i)$-$(ii): Choosing $X_t \equiv 1$ in (iii), we see that Z is a P-martingale. Now let Z be uniformly P-integrable, say with L^1-limit Z_∞. For any $t < u$ and $A \in \mathcal{F}_t$, we have $QA = E_P(Z_u; A)$. As $u \to \infty$ we get $QA = E_P(Z_\infty; A)$, which extends to arbitrary $A \in \mathcal{F}_\infty$ by a monotone-class argument. Thus, $Q = Z_\infty \cdot P$ on $\mathcal{F}_\infty$. Conversely, if $Q = \xi \cdot P$ on $\mathcal{F}_\infty$, then $E_P \xi = 1$, and the P-martingale $M_t = E_P(\xi \,|\mathcal{F}_t)$ satisfies $Q = M_t \cdot P$ on $\mathcal{F}_t$ for every t. But then $Z_t = M_t$ a.s. for all t, and Z is uniformly P-integrable with limit ξ.

(iv) By optional sampling,

$$QA = E_P\big(Z_{\tau\wedge t}; A\big), \quad A \in \mathcal{F}_{\tau\wedge t},\ t \geq 0,$$

and so

$$Q\big(A; \tau \leq t\big) = E_P\big(Z_\tau; A \cap \{\tau \leq t\}\big), \quad A \in \mathcal{F}_\tau,\ t \geq 0.$$

The assertion now follows by monotone convergence as $t \to \infty$.

(v) For any optional time τ, we need to show that X^τ is a Q-martingale iff $(XZ)^\tau$ is a P-martingale. This may be seen as in (iv), once we note that $Q = Z_t^\tau \cdot P$ on $\mathcal{F}_{\tau\wedge t}$ for all t.

(vi) By Lemma 9.32 it is enough to show that $Z_t > 0$ a.s. Q for every $t > 0$. This is clear from the fact that $Q\{Z_t = 0\} = E_P(Z_t;\, Z_t = 0) = 0$. □

The measure Q is usually not given at the outset, but needs to be constructed from the martingale Z. This requires some regularity conditions on the underlying probability space.

Lemma 19.19 *(existence) Let $Z \geq 0$ be a martingale with $Z_0 = 1$, on the canonical probability space (Ω, P) with induced filtration $\mathcal{F}$, where $\Omega = D_{\mathsf{R}_+,S}$ for a Polish space S. Then there exists a probability measure Q on Ω with*

$$Q = Z_t \cdot P \text{ on } \mathcal{F}_t, \quad t \geq 0.$$

Proof: For every $t \geq 0$ we may introduce the probability measure $Q_t = Z_t \cdot P$ on $\mathcal{F}_t$, which may be regarded as defined on $D_{[0,t],S}$. Since the latter spaces are again Polish for the Skorohod topology, Corollary 8.22 yields a probability measure Q on $D_{\mathsf{R}_+,S}$ with projections Q_t. The measure Q has clearly the required properties. □

We show how the drift of a continuous semi-martingale is transformed under a change of measure with a continuous density Z. An extension appears in Theorem 20.9.

Theorem 19.20 *(transformation of drift, Girsanov, van Schuppen & Wong) Consider some probability measures P, Q and a continuous process Z, such that*

$$Q = Z_t \cdot P \text{ on } \mathcal{F}_t, \quad t \geq 0.$$

Then for any continuous local P-martingale M, we have the local Q-martingale

$$\tilde{M} = M - Z^{-1} \cdot [M, Z].$$

Proof: First let Z^{-1} be bounded on the support of $[M]$. Then $\tilde{M}$ is a continuous P-semi-martingale, and so by Proposition 18.14 and an integration by parts,

$$\begin{aligned} \tilde{M}Z - (\tilde{M}Z)_0 &= \tilde{M} \cdot Z + Z \cdot \tilde{M} + [\tilde{M}, Z] \\ &= \tilde{M} \cdot Z + Z \cdot M - [M, Z] + [\tilde{M}, Z] \\ &= \tilde{M} \cdot Z + Z \cdot M, \end{aligned}$$

which shows that $\tilde{M}Z$ is a local P-martingale. Hence, by Lemma 19.18, $\tilde{M}$ is a local Q-martingale.

For general M, define $\tau_n = \inf\{t \ge 0;\, Z_t < 1/n\}$, and note as before that $\tilde{M}^{\tau_n}$ is a local Q-martingale for every $n \in \mathsf{N}$. Since $\tau_n \to \infty$ a.s. Q by Lemma 19.18, $\tilde{M}$ is a local Q-martingale by Lemma 18.1. □

Next we show how the basic notions of stochastic calculus are preserved by a change of measure. Then let $[X]_P$ be the quadratic variation of X under the probability measure P, write $L_P(X)$ for the class of processes V that are X-integrable under P, and let $(V\cdot X)_P$ be the corresponding stochastic integral.

Proposition 19.21 *(preservation properties) Let $Q = Z_t \cdot P$ on $\mathcal{F}_t$ for all $t \ge 0$, where Z is continuous. Then*

(i) *a continuous P-semi-martingale X is also a Q-semi-martingale, and*

$$[X]_P = [X]_Q \ \text{ a.s. } Q,$$

(ii) *$L_P(X) \subset L_Q(X)$, and for any $V \in L_P(X)$,*

$$(V\cdot X)_P = (V\cdot X)_Q \ \text{ a.s. } Q,$$

(iii) *for any continuous local P-martingale M and process $V \in L_P(M)$,*

$$(V\cdot M)^{\sim} = V\cdot \tilde{M} \ \text{ a.s. } Q,$$

whenever either side exists.

Proof: (i) Consider a continuous P-semi-martingale $X = M + A$, where M is a continuous local P-martingale and A is a process of locally finite variation. Under Q we may write

$$X = \tilde{M} + Z^{-1}\cdot[M,Z] + A,$$

where $\tilde{M}$ is the continuous local Q-martingale of Theorem 19.20, and we note that $Z^{-1}\cdot[M,Z]$ has locally finite variation, since $Z > 0$ a.s. Q by Lemma 19.18. Thus, X is also a semi-martingale under Q. The statement for $[X]$ is now clear from Proposition 18.17.

(ii) If $V \in L_P(X)$, then $V^2 \in L_P([X])$ and $V \in L_P(A)$, and so the same relations hold under Q, and we get $V \in L_Q(\tilde{M} + A)$. To get $V \in L_Q(X)$, it remains to show that $V \in L_Q(Z^{-1}\cdot[M,Z])$. Since $Z > 0$ under Q, it is equivalent to show that $V \in L_Q([M,Z])$, which is clear by Proposition 18.9, since $[M,Z]_Q = [\tilde{M},Z]_Q$ and $V \in L_Q(\tilde{M})$.

(iii) Note that $L_Q(M) = L_Q(\tilde{M})$ as before. If V belongs to either class, then under Q, Proposition 18.14 yields the a.s. relations

$$\begin{aligned}(V\cdot M)^{\sim} &= V\cdot M - Z^{-1}\cdot[V\cdot M, Z]\\ &= V\cdot M - VZ^{-1}\cdot[M,Z]\\ &= V\cdot\tilde{M}.\end{aligned}$$

□

In particular, Theorem 19.3 shows that if B is a P-Brownian motion in R^d, then $\tilde{B}$ is a Brownian motion under Q, since both processes are continuous martingales with the same covariation process.

The preceding theory simplifies when P and Q are equivalent on each $\mathcal{F}_t$, since in that case $Z > 0$ a.s. P by Lemma 19.18. If Z is also continuous, it can be expressed as an exponential martingale. More general processes of this kind are considered in Theorem 20.8.

Lemma 19.22 *(real exponential martingales) A continuous process $Z > 0$ is a local martingale iff a.s.*

$$Z_t = \mathcal{E}(M)_t \equiv \exp\Big(M_t - \tfrac{1}{2}[M]_t\Big), \quad t \geq 0, \tag{9}$$

for a continuous local martingale M. Here M is a.s. unique, and for any continuous local martingale N,

$$[M, N] = Z^{-1} \cdot [Z, N] \ \ a.s.$$

Proof: If M is a continuous local martingale, so is $\mathcal{E}(M)$ by Itô's formula. Conversely, let $Z > 0$ be a continuous local martingale. Then Corollary 18.19 yields

$$\begin{aligned} \log Z - \log Z_0 &= Z^{-1} \cdot Z - \tfrac{1}{2}\, Z^{-2} \cdot [Z] \\ &= Z^{-1} \cdot Z - \tfrac{1}{2}\, [Z^{-1} \cdot Z], \end{aligned}$$

and (9) follows with $M = \log Z_0 + Z^{-1} \cdot Z$. The last assertion is clear from this expression, and the uniqueness of M follows from Proposition 18.2. □

The drift of a continuous semi-martingale can sometimes be eliminated by a suitable change of measure. Beginning with the case of a Brownian motion B with deterministic drift, we need to show that $\mathcal{E}(B)$ is a true martingale, which follows easily by a direct computation. By $P \sim Q$ we mean that $P \ll Q$ and $Q \ll P$. Let L^2_{loc} be the class of measurable functions $f : \mathsf{R}_+ \to \mathsf{R}^d$, such that $|f|^2$ is locally Lebesgue integrable. For any $f \in L^2_{\mathrm{loc}}$ we define $f \cdot \lambda = (f^1 \cdot \lambda, \ldots, f^d \cdot \lambda)$, where the components on the right are ordinary Lebesgue integrals.

Theorem 19.23 *(shifted Brownian motion, Cameron & Martin) Let B be a canonical Brownian motion in R^d with induced, complete filtration $\mathcal{F}$, fix a continuous function $h\colon \mathsf{R}_+ \to \mathsf{R}^d$ with $h_0 = 0$, and put $P_h = \mathcal{L}(B + h)$. Then these conditions are equivalent:*

(i) $P_h \sim P_0$ *on* $\mathcal{F}_t$ *for all* $t \geq 0$,

(ii) $h = f \cdot \lambda$ *for an* $f \in L^2_{\mathrm{loc}}$.

Under those conditions,

(iii) $P_h = \mathcal{E}(f \cdot B)_t \cdot P_0$.

Proof, (i) $\Rightarrow$ (ii)–(iii): Assuming (i), Lemma 19.18 yields a P_0-martingale $Z > 0$, such that $P_h = Z_t \cdot P_0$ on $\mathcal{F}_t$ for all $t \geq 0$. Here Z is a.s. continuous by Theorem 19.11, and by Lemma 19.22 it equals $\mathcal{E}(M)$ for a continuous local P_0-martingale M. Using Theorem 19.11 again, we note that $M = V \cdot B = \sum_i V^i \cdot B^i$ a.s. for some processes $V^i \in L(B^1)$, and in particular $V \in L^2_{\mathrm{loc}}$ a.s.

By Theorem 19.20, the process

$$\begin{aligned}\tilde{B} &= B - [B, M] \\ &= B - V \cdot \lambda\end{aligned}$$

is a P_h-Brownian motion, and so under P_h the canonical process B has the semi-martingale decompositions

$$\begin{aligned}B &= \tilde{B} + V \cdot \lambda \\ &= (B - h) + h.\end{aligned}$$

Since the decomposition is a.s. unique by Proposition 18.2, we get $V \cdot \lambda = h$ a.s. Then also $h = f \cdot \lambda$ for some measurable function f on R_+, and since clearly $\lambda\{t \geq 0;\, V_t \neq f_t\} = 0$ a.s., we have even $f \in L^2_{\mathrm{loc}}$, proving (ii). Furthermore, $M = V \cdot B = f \cdot B$ a.s., and (iii) follows.

(ii) $\Rightarrow$ (i): Assume (ii). Since $M = f \cdot B$ is a time-changed Brownian motion under P_0, the process $Z = \mathcal{E}(M)$ is a P_0-martingale, and Lemma 19.19 yields a probability measure Q on $C_{\mathsf{R}_+, \mathsf{R}^d}$, such that $Q = Z_t \cdot P_0$ on $\mathcal{F}_t$ for all $t \geq 0$. Moreover, Theorem 19.20 shows that the process

$$\begin{aligned}\tilde{B} &= B - [B, M] \\ &= B - h\end{aligned}$$

is a Q-Brownian motion, and so $Q = P_h$. In particular, this proves (i). $\square$

For more general semi-martingales, Theorem 19.20 and Lemma 19.22 suggest that we might remove the drift by a change of measure of the form $Q = \mathcal{E}(M)_t \cdot P$ on $\mathcal{F}_t$ for all $t \geq 0$, where M is a continuous local martingale with $M_0 = 0$. Then by Lemma 19.18 we need $Z = \mathcal{E}(M)$ to be a true martingale, which can often be verified through the following criterion.

Theorem 19.24 *(uniform integrability, Novikov) Let M be a real, continuous local martingale with $M_0 = 0$. Then $\mathcal{E}(M)$ is a uniformly integrable martingale, whenever*

$$E \exp\left(\tfrac{1}{2} [M]_\infty\right) < \infty.$$

This will first be proved in a special case.

Lemma 19.25 *(Wald's identity) For any real Brownian motion B and optional time τ, we have*

$$E e^{\tau/2} < \infty \quad \Rightarrow \quad E \exp\left(B_\tau - \tfrac{1}{2} \tau\right) = 1.$$

Proof: First consider the hitting times

$$\tau_b = \inf\Big\{t \geq 0;\ B_t = t - b\Big\}, \quad b > 0.$$

Since the τ_b remain optional with respect to the right-continuous, induced filtration, we may take B to be a canonical Brownian motion with associated distribution $P = P_0$. Defining $h_t \equiv t$ and $Z = \mathcal{E}(B)$, we see from Theorem 19.23 that $P_h = Z_t \cdot P$ on $\mathcal{F}_t$ for all $t \geq 0$. Since $\tau_b < \infty$ a.s. under both P and P_h, Lemma 19.18 yields

$$\begin{aligned} E\exp\Big(B_{\tau_b} - \tfrac{1}{2}\,\tau_b\Big) &= EZ_{\tau_b} \\ &= E\Big(Z_{\tau_b};\ \tau_b < \infty\Big) \\ &= P_h\{\tau_b < \infty\} = 1. \end{aligned}$$

The stopped process $M_t \equiv Z_{t\wedge\tau_b}$ is a positive local martingale, and by Fatou's lemma it is also a super-martingale on $[0,\infty]$. Since clearly $EM_\infty = EZ_{\tau_b} = 1 = EM_0$, we see from the Doob decomposition that M is a true martingale on $[0,\infty]$. For τ as stated, we get by optional sampling

$$\begin{aligned} 1 &= EM_\tau = EZ_{\tau\wedge\tau_b} \\ &= E\Big(Z_\tau;\ \tau \leq \tau_b\Big) + E\Big(Z_{\tau_b};\ \tau > \tau_b\Big). \end{aligned} \tag{10}$$

Using the definition of τ_b and the condition on τ, we get as $b \to \infty$

$$\begin{aligned} E\Big(Z_{\tau_b};\ \tau > \tau_b\Big) &= e^{-b}E\Big(e^{\tau_b/2};\ \tau > \tau_b\Big) \\ &\leq e^{-b}Ee^{\tau/2} \to 0, \end{aligned}$$

and so the last term in (10) tends to zero. Since also $\tau_b \to \infty$, the first term on the right tends to EZ_τ by monotone convergence, and we get $EZ_\tau = 1$. □

Proof of Theorem 19.24: Since $\mathcal{E}(M)$ is a super-martingale on $[0,\infty]$, it suffices to show that, under the stated condition, $E\,\mathcal{E}(M)_\infty = 1$. Using Theorem 19.4 and Proposition 9.9, we may then reduce to the statement of Lemma 19.25. □

This yields in particular a classical result for Brownian motion:

Corollary 19.26 *(drift removal, Girsanov) Consider a Brownian motion B and a progressive process V in R^d, such that*

$$E\exp\Big\{\tfrac{1}{2}\Big(|V|^2 \cdot \lambda\Big)_\infty\Big\} < \infty.$$

Then

(i) $Q = \mathcal{E}(V \cdot B)_\infty \cdot P$ *is a probability measure,*

(ii) $\tilde{B} = B - V \cdot \lambda$ *is a Q-Brownian motion.*

Proof: Combine Theorems 19.20 and 19.24. □

Exercises

1. Let B be a real Brownian motion and let $V \in L(B)$. Show that $X = V \cdot B$ is a time-changed Brownian motion, express the required time change τ in terms of V, and verify that X is τ-continuous.

2. Consider a real Brownian motion B and some progressive processes $V^t \in L(B)$, $t \geq 0$. Give conditions on the V^t for the process $X_t = (V^t \cdot B)_\infty$ to be Gaussian.

3. Let M be a continuous local martingale in R^d, and let $V^1, \dots, V^d$ be predictable processes on $S \times \mathsf{R}_+$, such that $\sum_{ij} \int_0^\infty V^i_{s,r} V^j_{t,r}\, d[M^i, M^j]_r = \rho_{s,t}$ is a.s. non-random for all $s, t \in S$. Show that the process $X_s = \sum_i \int_0^\infty V^i_{s,r}\, dM^i_r$ is Gaussian on S with covariance function ρ. (*Hint:* Use Lemma 19.2.)

4. For M in Theorem 19.4, let $[M]_\infty = \infty$ a.s. Show that M is τ-continuous in the sense of Theorem 18.24, and conclude from Theorem 19.3 that $B = M \circ \tau$ is a Brownian motion. Also show for any $V \in L(M)$ that $(V \circ \tau) \cdot B = (V \cdot M) \circ \tau$ a.s.

5. Deduce Theorem 19.3 for $d = 1$ from Theorem 22.17. (*Hint:* Proceed as above to construct a discrete-time martingale with jump sizes h. Let $h \to 0$, and use a version of Proposition 18.17.)

6. Let M be a real, continuous local martingale. Show that M converges a.s. on the set $\{\sup_t M_t < \infty\}$. (*Hint:* Use Theorem 19.4.)

7. Let $\mathcal{F}$, $\mathcal{G}$ be right-continuous filtrations, where $\mathcal{G}$ is a standard extension of $\mathcal{F}$, and let τ be an $\mathcal{F}$-optional time. Show that $\mathcal{F}_\tau \subset \mathcal{G}_\tau \perp\!\!\!\perp_{\mathcal{F}_\tau} \mathcal{F}$. (*Hint:* Apply optional sampling to the uniformly integrable martingale $M_t = E(\xi|\mathcal{F}_t)$, for any $\xi \in L^1(\mathcal{F}_\infty)$.)

8. Let $\mathcal{F}$, $\mathcal{G}$ be filtrations on a common probability space $(\Omega, \mathcal{A}, P)$. Show that $\mathcal{G}$ is a standard extension of $\mathcal{F}$ iff every $\mathcal{F}$-martingale is also a $\mathcal{G}$-martingale. (*Hint:* Consider martingales of the form $M_t = E(\xi|\, \mathcal{F}_t)$, where $\xi \in L^1(\mathcal{F}_\infty)$. Here M_t is $\mathcal{G}_t$-measurable for all ξ iff $\mathcal{F}_t \subset \mathcal{G}_t$, and then $M_t = E(\xi|\, \mathcal{G}_t)$ a.s. for all ξ iff $\mathcal{F} \perp\!\!\!\perp_{\mathcal{F}_t} \mathcal{G}_t$ by Proposition 8.9.)

9. Let M be a non-trivial, isotropic, continuous local martingale in R^d, and fix an affine transformation f on R^d. Show that even $f(M)$ is isotropic iff f is conformal (i.e., the composition of a rigid motion with a change of scale).

10. Show that for any $d \geq 2$, the path of a Brownian motion in R^d has Lebesgue measure 0. (*Hint:* Use Theorem 19.6 and Fubini's theorem.)

11. Deduce Theorem 19.6 (ii) from Theorem 12.8. (*Hint:* Define $\tau = \inf\{t;\, |B_t| = 1\}$, and iterate the construction to form a random walk in R^d with steps of size 1.)

12. Give an example of a bounded, continuous martingale $M \geq 0$ with $M_0 = 1$, such that $P\{M_t = 0\} > 0$ for large $t > 0$. (*Hint:* We may take M to be a Brownian motion stopped at the levels 0 and 2.)

13. For any Brownian motion B and optional time $\tau < \infty$, show that $E \exp(B_\tau - \frac{1}{2}\tau) \leq 1$, where the inequality may be strict. (*Hint:* Truncate and use Fatou's lemma. Note that $t - 2B_t \to \infty$ by the law of large numbers.)

14. In Theorem 19.23, give necessary and sufficient conditions for $P_h \sim P_0$ on $\mathcal{F}_\infty$, and find the associated density.

15. Show by an example that Lemma 19.25 may fail without a moment condition on τ. (*Hint:* Choose a $\tau > 0$ with $B_\tau = 0$.)

Chapter 20
Semi-Martingales and Stochastic Integration

Predictable covariation, L^2-martingale and semi-martingale integrals, quadratic variation and covariation, substitution rule, Doléans exponential, change of measure, BDG inequalities, martingale integral, purely discontinuous semi-martingales, semi-martingale and martingale decompositions, exponential super-martingales and inequalities, quasi-martingales, stochastic integrators

In the last two chapters, we have seen how the Itô integral with associated transformation rules leads to a powerful stochastic calculus for continuous semi-martingales, defined as sums of continuous local martingales and continuous processes of locally finite variation. This entire theory will now be extended to semi-martingales with possible jump discontinuities. In particular, the semi-martingales turn out to be precisely the processes that can serve as stochastic integrators with reasonable adaptedness and continuity properties.

The present theory is equally important but much more subtle, as it depends crucially on the Doob–Meyer decomposition from Chapter 10 and the powerful BDG inequalities for general local martingales. Our construction of the general stochastic integral proceeds in three steps. First we imitate the definition of the L^2-integral $V \cdot M$ from Chapter 18, using a predictable version $\langle M, N\rangle$ of the covariation process. A suitable truncation then allows us to extend the integral to arbitrary semi-martingales X and bounded, predictable processes V. The ordinary covariation $[X, Y]$ can now be defined by the integration-by-parts formula, and we may use the general BDG inequalities to extend the martingale integral $V \cdot M$ to more general integrands V.

Once the stochastic integral is defined, we may develop a stochastic calculus for general semi-martingales. In particular, we prove an extension of Itô's formula, solve a basic stochastic differential equation, and establish a general Girsanov-type theorem for absolutely continuous changes of the probability measure. The latter material extends the relevant parts of Chapters 18–19.

The stochastic integral and covariation process, along with the Doob–Meyer decomposition from Chapter 10, provide the basic tools for a more detailed study of general semi-martingales. In particular, we may now establish two general decompositions, similar to the decompositions of optional times and increasing processes in Chapter 10. We further derive some exponential inequalities for martingales with bounded jumps, characterize local quasi-martingales

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_21

as special semi-martingales, and show that no continuous extension of the predictable integral exists beyond the context of semi-martingales.

Throughout this chapter, let $\mathcal{M}^2$ be the class of uniformly square-integrable martingales, and note as in Lemma 18.4 that $\mathcal{M}^2$ is a Hilbert space for the norm $\|M\| = (EM_\infty^2)^{1/2}$. Let $\mathcal{M}_0^2$ be the closed linear subspace of martingales $M \in \mathcal{M}^2$ with $M_0 = 0$. The corresponding classes $\mathcal{M}^2_{\rm loc}$ and $\mathcal{M}^2_{0,{\rm loc}}$ are defined as the sets of processes M, such that the stopped versions M^{τ_n} belong to $\mathcal{M}^2$ or $\mathcal{M}_0^2$, respectively, for suitable optional times $\tau_n \to \infty$.

For every $M \in \mathcal{M}^2_{\rm loc}$, we note that M^2 is a local sub-martingale. The corresponding compensator $\langle M \rangle$ is called the *predictable quadratic variation* of M. More generally, we define[1] the *predictable covariation* $\langle M, N \rangle$ of two processes $M, N \in \mathcal{M}^2_{\rm loc}$ as the compensator of MN, also obtainable through the *polarization formula*

$$4\langle M, N \rangle = \langle M + N \rangle - \langle M - N \rangle.$$

Note that $\langle M, M \rangle = \langle M \rangle$. If M and N are continuous, then clearly $\langle M, N \rangle = [M, N]$ a.s. Here we list some further useful properties.

Proposition 20.1 *(predictable covariation) For any $M, N, M^n \in \mathcal{M}^2_{\rm loc}$, we have a.s.*

(i) $\langle M, N \rangle$ *is symmetric, bilinear with* $\langle M, N \rangle = \langle M - M_0, N \rangle$,

(ii) $\langle M \rangle = \langle M, M \rangle$ *is non-decreasing,*

(iii) $|\langle M, N \rangle| \le \int |d\langle M, N \rangle| \le \langle M \rangle^{1/2} \langle N \rangle^{1/2}$,

(iv) $\langle M, N \rangle^\tau = \langle M^\tau, N \rangle = \langle M^\tau, N^\tau \rangle$ *for any optional time* τ,

(v) $\langle M^n \rangle_\infty \xrightarrow{P} 0 \;\Rightarrow\; (M^n - M_0^n)^* \xrightarrow{P} 0$.

Proof, (i)$-$(iv): Lemma 10.11 shows that $\langle M, N \rangle$ is the a.s. unique predictable process of locally integrable variation starting at 0, such that $MN - \langle M, N \rangle$ is a local martingale. The symmetry and bilinearity follow immediately, as does last property in (i), since MN_0, $M_0 N$, and $M_0 N_0$ are all local martingales. Moreover, we get (iii) as in Proposition 18.9, and (iv) as in Theorem 18.5.

(v) Here we may take $M_0^n = 0$ for all n. Let $\langle M^n \rangle_\infty \xrightarrow{P} 0$, fix any $\varepsilon > 0$, and define $\tau_n = \inf\{t;\, \langle M^n \rangle_t \ge \varepsilon\}$. Since $\langle M^n \rangle$ is predictable, even τ_n is predictable by Theorem 10.14, say with announcing sequence $\tau_{nk} \uparrow \tau_n$. The latter may be chosen such that M^n becomes an L^2-martingale and $(M^n)^2 - \langle M^n \rangle$ a uniformly integrable martingale on every interval $[0, \tau_{nk}]$. Then Proposition 9.17 yields

$$\begin{aligned} E(M^n)^{*2}_{\tau_{nk}} &\lesssim E(M^n)^2_{\tau_{nk}} \\ &= E\langle M^n \rangle_{\tau_{nk}} \le \varepsilon, \end{aligned}$$

[1] The predictable covariation $\langle \cdot, \cdot \rangle$ must not be confused with the inner product appearing in Chapters 14, 21, and 35.

and as $k \to \infty$ we get $E(M^n)^{*2}_{\tau_n-} \lesssim \varepsilon$. Now fix any $\delta > 0$, and write

$$\begin{aligned} P\Big\{(M^n)^{*2} > \delta\Big\} &\le P\{\tau_n < \infty\} + \delta^{-1}E(M^n)^{*2}_{\tau_n-} \\ &\lesssim P\Big\{\langle M^n\rangle_\infty \ge \varepsilon\Big\} + \delta^{-1}\varepsilon. \end{aligned}$$

Here the right-hand side tends to zero as $n \to \infty$ and then $\varepsilon \to 0$. □

The predictable quadratic variation can be used to extend the Itô integral from Chapter 18. Then let $\mathcal{E}$ be the class of bounded, predictable step processes V with jumps at finitely many fixed times. The corresponding integral $V \cdot X$ is referred to as an *elementary predictable integral*.

Given any $M \in \mathcal{M}^2_{\mathrm{loc}}$, let $L^2(M)$ be the class of predictable processes V, such that $(V^2 \cdot \langle M\rangle)_t < \infty$ a.s. for all $t > 0$. Assuming $M \in \mathcal{M}^2_{\mathrm{loc}}$ and $V \in L^2(M)$, we may form an integral process $V \cdot M$ belonging to $\mathcal{M}^2_{0,\mathrm{loc}}$. In the following statement, we assume that $M, N, M_n \in \mathcal{M}^2_{\mathrm{loc}}$, and let U, V, V_n be predictable processes such that the appropriate integrals exist.

Theorem 20.2 *(L^2-martingale integral, Courrège, Kunita & Watanabe) The elementary predictable integral extends a.s. uniquely to a bilinear map of any $M \in \mathcal{M}^2_{\mathrm{loc}}$ and $V \in L^2(M)$ into a process $V \cdot M \in \mathcal{M}^2_{0,\mathrm{loc}}$, such that a.s.*

(i) $\Big(V_n^2 \cdot \langle M_n\rangle\Big)_t \xrightarrow{P} 0 \;\Rightarrow\; \Big(V_n \cdot M_n\Big)^*_t \xrightarrow{P} 0, \quad t > 0,$

(ii) $\langle V \cdot M, N\rangle = V \cdot \langle M, N\rangle, \; N \in \mathcal{M}^2_{\mathrm{loc}},$

(iii) $U \cdot (V \cdot M) = (UV) \cdot M,$

(iv) $\Delta(V \cdot M) = V(\Delta M),$

(v) $(V \cdot M)^\tau = V \cdot M^\tau = (V 1_{[0,\tau]}) \cdot M$ *for any optional time* τ.

The integral $V \cdot M$ is a.s. uniquely determined by (ii).

The proof depends on an elementary approximation of predictable processes, similar to Lemma 18.23 in the continuous case.

Lemma 20.3 *(approximation) Let V be a predictable process with $|V|^p \in L(A)$, where A is increasing and $p \ge 1$. Then there exist some $V_1, V_2, \ldots \in \mathcal{E}$ with*

$$\Big(|V_n - V|^p \cdot A\Big)_t \to 0 \;\; a.s., \quad t > 0.$$

Proof: It is enough to establish the approximation $(|V_n - V|^p \cdot A)_t \xrightarrow{P} 0$. By Minkowski's inequality we may then approximate in steps, and by dominated convergence we may first take V to be simple. Then each term can be approximated separately, and so we may next choose $V = 1_B$ for a predictable set B. Approximating separately on disjoint intervals, we may finally assume that $B \subset \Omega \times [0, t]$ for some $t > 0$. The desired approximation then follows from Lemma 10.2 by a monotone-class argument. □

Proof of Theorem 20.2: As in Theorem 18.11, we may construct the integral $V\cdot M$ as the a.s. unique process in $\mathcal{M}^2_{0,\mathrm{loc}}$ satisfying (ii). The mapping $(V,M)\mapsto V\cdot M$ is clearly bilinear and extends the elementary predictable integral, just as in Lemma 18.10. Properties (iii) and (v) may be proved as in Propositions 18.14 and 18.15. The stated continuity follows immediately from (ii) and Proposition 20.1 (v). To prove the asserted uniqueness, it is then enough to apply Lemma 20.3 with $A=\langle M\rangle$ and $p=2$.

To prove (iv), we may apply Lemma 20.3 with

$$A_t=\langle M\rangle_t+\sum_{s\le t}(\Delta M_s)^2,\quad t\ge 0,$$

to ensure the existence of some processes $V_n\in\mathcal{E}$, such that a.s.

$$\begin{aligned} V_n(\Delta M)&\to V(\Delta M),\\ \big(V_n\cdot M-V\cdot M\big)^*&\to 0. \end{aligned}$$

In particular, $\Delta(V_n\cdot M)\to\Delta(V\cdot M)$ a.s., and (iv) follows from the corresponding relation for the elementary integrals $V_n\cdot M$, once we have shown that $\sum_{s\le t}(\Delta M_s)^2<\infty$ a.s. For the latter condition, we may take $M\in\mathcal{M}^2_0$ and define $t_{n,k}=kt\,2^{-n}$ for $k\le 2^n$. Then Fatou's lemma yields

$$\begin{aligned} E\sum_{s\le t}(\Delta M_s)^2 &\le E\liminf_{n\to\infty}\sum\nolimits_k\Big(M_{t_{n,k}}-M_{t_{n,k-1}}\Big)^2\\ &\le \liminf_{n\to\infty}E\sum\nolimits_k\Big(M_{t_{n,k}}-M_{t_{n,k-1}}\Big)^2\\ &= EM_t^2<\infty. \end{aligned}$$

□

A *semi-martingale* is defined as a right-continuous, adapted process X, admitting a decomposition $M+A$ into a local martingale M and a process A of locally finite variation starting at 0. If the variation of A is even locally integrable, we can write $X=(M+A-\hat{A})+\hat{A}$, where $\hat{A}$ denotes the compensator of A. We may then choose A to be predictable, in which case the decomposition is a.s. unique by Propositions 18.2 and 10.16, and X is called a *special semi-martingale* with *canonical decomposition* $M+A$.

The Lévy processes constitute basic examples of semi-martingales, and we note that a Lévy process is a special semi-martingale iff its Lévy measure ν satisfies $\int(x^2\wedge|x|)\,\nu(dx)<\infty$. From Theorem 10.5 we further see that any local sub-martingale is a special semi-martingale.

We may extend the stochastic integration to general semi-martingales. For the moment we consider only locally bounded integrands, which covers most applications of interest.

Theorem 20.4 *(semi-martingale integral, Doléans-Dade & Meyer) The L^2-integral of Theorem* 20.2 *and the ordinary Lebesgue–Stieltjes integral extend a.s. uniquely to a bilinear map of any semi-martingale X and locally bounded, predictable process V into a semi-martingale $V\cdot X$. The integration map also satisfies*

(i) *properties* (iii)–(v) *of Theorem* 20.2*,*

(ii) $V \geq |V_n| \to 0 \;\Rightarrow\; (V_n \cdot X)^*_t \xrightarrow{P} 0,\; t > 0,$

(iii) *when X is a local martingale, so is $V \cdot X$.*

Our proof relies on a basic truncation:

Lemma 20.5 *(truncation, Doléans-Dade, Jacod & Mémin, Yan) Any local martingale M has a decomposition $M' + M''$ into local martingales, such that*

(i) *M' has locally integrable variation,*

(ii) $|\Delta M''| \leq 1$ *a.s.*

Proof: Define

$$A_t = \sum_{s \leq t} (\Delta M_s)\, 1\Big\{|\Delta M_s| > \tfrac{1}{2}\Big\}, \quad t \geq 0.$$

By optional sampling, A has locally integrable variation. Let $\hat{A}$ be the compensator of A, and put $M' = A - \hat{A}$ and $M'' = M - M'$. Then M' and M'' are again local martingales, and M' has locally integrable variation. Since also

$$\begin{aligned} |\Delta M''| &\leq \Big|\Delta M - \Delta A\Big| + |\Delta \hat{A}| \\ &\leq \tfrac{1}{2} + |\Delta \hat{A}|, \end{aligned}$$

it remains to show that $|\Delta \hat{A}| \leq \frac{1}{2}$. Since the constructions of A and $\hat{A}$ commute with optional stopping, we may take M and M' to be uniformly integrable. Since $\hat{A}$ is predictable, the times $\tau = n \wedge \inf\{t;\, |\Delta \hat{A}| > \frac{1}{2}\}$ are predictable by Theorem 10.14, and it suffices to show that $|\Delta \hat{A}_\tau| \leq \frac{1}{2}$ a.s. Since clearly $E(\Delta M_\tau | \mathcal{F}_{\tau-}) = E(\Delta M'_\tau | \mathcal{F}_{\tau-}) = 0$ a.s., Lemma 10.3 yields

$$\begin{aligned} |\Delta \hat{A}_\tau| &= \Big|E\Big(\Delta A_\tau \Big| \mathcal{F}_{\tau-}\Big)\Big| \\ &= \Big|E\Big(\Delta M_\tau;\, |\Delta M_\tau| > \tfrac{1}{2} \Big| \mathcal{F}_{\tau-}\Big)\Big| \\ &= \Big|E\Big(\Delta M_\tau;\, |\Delta M_\tau| \leq \tfrac{1}{2} \Big| \mathcal{F}_{\tau-}\Big)\Big| \leq \tfrac{1}{2}. \end{aligned}$$

□

Proof of Theorem 20.4: By Lemma 20.5 we may write $X = M + A$, where M is a local martingale with bounded jumps, hence a local L^2-martingale, and A has locally finite variation. For any locally bounded, predictable process V, we may then define $V \cdot X = V \cdot M + V \cdot A$, where the first term is the integral in Theorem 20.2 and the second term is an ordinary Lebesgue–Stieltjes integral. If $V \geq |V_n| \to 0$, we get by dominated convergence

$$\Big(V_n^2 \cdot \langle M \rangle\Big)_t + \Big(V_n \cdot A\Big)^*_t \to 0,$$

and so by Theorem 20.2 we have $(V_n \cdot X)^*_t \xrightarrow{P} 0$ for all $t > 0$.

For the uniqueness it suffices to prove that, if $M = A$ is a local L^2-martingale of locally finite variation, then $V \cdot M = V \cdot A$ a.s. for every locally bounded, predictable process V, where $V \cdot M$ is the integral in Theorem 20.2

and $V \cdot A$ is an ordinary Stieltjes integral. The two integrals clearly agree when $V \in \mathcal{E}$. For general V, we may approximate as in Lemma 20.3 by processes $V_n \in \mathcal{E}$, such that a.s.

$$\Big\{(V_n - V)^2 \cdot \langle M \rangle\Big\}^* + \Big(|V_n - V| \cdot A\Big)^* \to 0.$$

Then for any $t > 0$,

$$\begin{aligned}(V_n \cdot M)_t &\xrightarrow{P} (V \cdot M)_t,\\ (V_n \cdot A)_t &\to (V \cdot A)_t,\end{aligned}$$

and the desired equality follows.

(iii) By Lemma 20.5 and a suitable localization, we may assume that V is bounded and X has integrable variation A. By Lemma 20.3 we may next choose some uniformly bounded processes $V_1, V_2, \ldots \in \mathcal{E}$, such that $(|V_n - V| \cdot A)_t \to 0$ a.s. for every $t \geq 0$. Then $(V_n \cdot X)_t \to (V \cdot X)_t$ a.s. for all t, which extends to L^1 by dominated convergence. Thus, the martingale property of $V_n \cdot X$ carries over to $V \cdot X$. □

For any semi-martingales X, Y, the left-continuous versions $(X_-)_t = X_{t-}$ and $(Y_-)_t = Y_{t-}$ are locally bounded and predictable, and hence may serve as integrands in general stochastic integrals. We may then define[2] the *quadratic variation* $[X]$ and *covariation* $[X, Y]$ by the *integration-by-parts* formulas

$$\begin{aligned}[X] &= X^2 - X_0^2 - 2\,X_- \cdot X,\\ [X, Y] &= X\,Y - X_0\,Y_0 - X_- \cdot Y - Y_- \cdot X\\ &= \tfrac{1}{4}\Big(\,[X+Y] - [X-Y]\,\Big),\end{aligned} \tag{1}$$

so that in particular $[X] = [X, X]$. We list some further properties of the covariation.

Theorem 20.6 *(covariation) For any semi-martingales X, Y, we have a.s.*

(i) $[X, Y]$ *is symmetric, bilinear with* $[X, Y] = [X - X_0, Y]$,

(ii) $[X] = [X, X]$ *is non-decreasing,*

(iii) $|\,[X, Y]\,| \leq \int |d\,[X, Y]| \leq [X]^{1/2}\,[Y]^{1/2}$,

(iv) $\Delta[X] = (\Delta X)^2$, $\Delta[X, Y] = (\Delta X)(\Delta Y)$,

(v) $[V \cdot X, Y] = V \cdot [X, Y]$ *for locally bounded, predictable* V,

(vi) $[X^\tau, Y] = [X^\tau, Y^\tau] = [X, Y]^\tau$ *for any optional time* τ,

(vii) $M, N \in \mathcal{M}^2_{\mathrm{loc}} \;\Rightarrow\; [M, N]$ *has compensator* $\langle M, N \rangle$,

(viii) $[X, A]_t = \sum_{s \leq t} (\Delta X_s)(\Delta A_s)$ *when A has locally finite variation.*

[2]Both covariation processes $[X, Y]$ and $\langle X, Y \rangle$ remain important, as is evident from, e.g., Theorems 20.9, 20.16, and 20.17.

Proof: (i) The symmetry and bilinearity of $[X, Y]$ are obvious from (1), and the last relation holds since $[X, Y_0] = 0$.

(ii) Proposition 18.17 extends with the same proof to general semi-martingales. In particular, $[X]_s \le [X]_t$ a.s. for all $s \le t$. By right-continuity, we can choose the exceptional null set to be independent of s and t, which means that $[X]$ is a.s. non-decreasing.

(iii) Proceed as in case of Proposition 18.9.

(iv) By (1) and Theorem 20.2 (iv),

$$\begin{aligned}\Delta[X, Y]_t &= \Delta(X\,Y)_t - \Delta(X_- \cdot Y)_t - \Delta(Y_- \cdot X)_t \\ &= X_t\, Y_t - X_{t-}\, Y_{t-} - X_{t-}(\Delta Y_t) - Y_{t-}(\Delta X_t) \\ &= (\Delta X_t)(\Delta Y_t).\end{aligned}$$

(v) For $V \in \mathcal{E}$ the relation follows most easily from the extended version in Proposition 18.17. Also note that both sides are a.s. linear in V. Now let $V, V_1, V_2, \ldots$ be locally bounded and predictable with $V \ge |V_n| \to 0$. Then $V_n \cdot [X, Y] \to 0$ by dominated convergence, and Theorem 20.4 yields

$$[V_n \cdot X,\, Y] = (V_n \cdot X)\,Y - (V_n \cdot X)_- \cdot Y - (V_n\, Y_-) \cdot X \xrightarrow{P} 0.$$

A monotone-class argument yields an extension to arbitrary V.

(vi) Use (v) with $V = 1_{[0,\tau]}$.

(vii) Since $M_- \cdot N$ and $N_- \cdot M$ are local martingales, the assertion follows from (1) and the definition of $\langle M, N\rangle$.

(viii) For step processes A, the stated relation follows from the extended version in Proposition 18.17. If instead $\Delta A \le \varepsilon$, we conclude from the same result and property (iii), along with the ordinary Cauchy inequality, that

$$\begin{aligned}[X, A]_t^2 \vee \Big|\sum_{s\le t}(\Delta X_s)(\Delta A_s)\Big|^2 &\le [X]_t\,[A]_t \\ &\le \varepsilon [X]_t \int_0^t |dA_s|.\end{aligned}$$

The assertion now follows by a simple approximation. □

We may now extend Itô's formula in Theorem 18.18 to a substitution rule for general semi-martingales. By a semi-martingale in R^d we mean a process $X = (X^1, \ldots, X^d)$ composed by one-dimensional semi-martingales X^i. Let $[X^i, X^j]^c$ denote the continuous components of the finite-variation processes $[X^i, X^j]$, and write $\partial_i f$ and $\partial_{ij}^2 f$ for the first and second order partial derivatives of f, respectively.

Theorem 20.7 *(substitution rule, Kunita & Watanabe) For any semi-martingale $X = (X^1, \dots, X^d)$ in R^d and function $f \in C^2(\mathsf{R}^d)$, we have*[3]

$$f(X_t) = f(X_0) + \int_0^t \partial_i f(X_{s-})\, dX_s^i + \tfrac{1}{2} \int_0^t \partial_{ij}^2 f(X_{s-})\, d\,[X^i, X^j]_s^c$$
$$+ \sum_{s \le t} \Big\{ \Delta f(X_s) - \partial_i f(X_{s-})(\Delta X_s^i) \Big\}. \tag{2}$$

Proof: Let $f \in C^2(\mathsf{R}^d)$ satisfy (2), where the last term has locally finite variation. We show that (2) remains true for the functions $g_k(x) = x_k f(x)$, $k = 1, \dots, n$. Then note that by (1),

$$g_k(X) = g_k(X_0) + X_-^k \cdot f(X) + f(X_-) \cdot X^k + [X^k, f(X)]. \tag{3}$$

Writing $\hat f(x, y) = f(x) - f(y) - \partial_i f(y)(x_i - y_i)$, we get by (2) and Theorem 20.2 (iii)

$$X_-^k \cdot f(X) = X_-^k\, \partial_i f(X_-) \cdot X^i + \tfrac{1}{2} X_-^k\, \partial_{ij}^2 f(X_-) \cdot [X^i, X^j]^c$$
$$+ \textstyle\sum_s X_{s-}^k\, \hat f(X_s, X_{s-}). \tag{4}$$

Next we see from (ii), (iv), (v), and (viii) of Theorem 20.6 that

$$\Big[X^k, f(X)\Big] = \partial_i f(X_-) \cdot [X^k, X^i] + \textstyle\sum_s (\Delta X_s^k)\, \hat f(X_s, X_{s-})$$
$$= \partial_i f(X_-) \cdot [X^k, X^i]^c + \textstyle\sum_s (\Delta X_s^k) \{\Delta f(X_s)\}. \tag{5}$$

Inserting (4)−(5) into (3) and using the elementary formulas

$$\partial_i\, g_k(x) = \delta_{ik} f(x) + x_k\, \partial_i f(x),$$
$$\partial_{ij}^2\, g_k(x) = \delta_{ik}\, \partial_j f(x) + \delta_{jk}\, \partial_i f(x) + x_k\, \partial_{ij}^2 f(x),$$
$$\hat g_k(x, y) = (x_k - y_k) \Big\{ f(x) - f(y) \Big\} + y_k \hat f(x, y),$$

followed by some simplification, we obtain the desired expression for $g_k(X)$.

Formula (2) is trivially true for constant functions, and it extends by induction and linearity to arbitrary polynomials. By Weierstrass' theorem, a general function $f \in C^2(\mathsf{R}^d)$ can be approximated by polynomials, such that all derivatives up to second order tend to those of f, uniformly on every compact set. To prove (2) for f, it is then enough to show that the right-hand side tends to zero in probability, as f and its first and second order derivatives tend to zero, uniformly on compact sets.

For the integrals in (2), this holds by the dominated convergence property in Theorem 20.4, and it remains to consider the last term. Writing $B_t = \{x \in \mathsf{R}^d;\ |x| \le X_t^*\}$ and $\|g\|_B = \sup_B |g|$, we get by Taylor's formula in R^d

$$\sum_{s \le t} \Big| \hat f(X_s, X_{s-}) \Big| \ \lesssim\ \sum_{i,j} \Big\| \partial_{ij}^2 f \Big\|_{B_t} \sum_{s \le t} |\Delta X_s|^2$$
$$\le\ \textstyle\sum_{i,j} \Big\| \partial_{ij}^2 f \Big\|_{B_t} \sum_i [X^i]_t \to 0.$$

[3]As always, summation over repeated indices is understood.

The same estimate shows that the last term has locally finite variation. □

To illustrate the use of the general substitution rule, we consider a partial extension of Lemma 19.22 and Proposition 32.2 to general semi-martingales.

Theorem 20.8 *(Doléans exponential) For a semi-martingale X with $X_0 = 0$, the equation $Z = 1 + Z_- \cdot X$ has the a.s. unique solution*

$$Z_t = \mathcal{E}(X) \equiv \exp\Big(X_t - \tfrac{1}{2}[X]_t^c\Big) \prod_{s \le t} (1 + \Delta X_s)\, e^{-\Delta X_s}, \quad t \ge 0. \tag{6}$$

The infinite product in (6) is a.s. absolutely convergent, since $\sum_{s \le t} (\Delta X_s)^2 \le [X]_t < \infty$. However, we may have $\Delta X_s = -1$ for some $s > 0$, in which case $Z = 0$ for $t \ge s$. The process $\mathcal{E}(X)$ in (6) is called the *Doléans exponential* of X. For continuous X we get $\mathcal{E}(X) = \exp(X - \frac{1}{2}[X])$, conforming with the notation of Lemma 19.22. For processes A of locally finite variation, formula (6) simplifies to

$$\mathcal{E}(A) = \exp(A_t^c) \prod_{s \le t} (1 + \Delta A_s), \quad t \ge 0.$$

Proof of Theorem 20.8: To check that (6) is a solution, we may write $Z = e^Y V$, where

$$\begin{aligned} Y_t &= X_t - \tfrac{1}{2}[X]_t^c, \\ V_t &= \prod_{s \le t} (1 + \Delta X_s)\, e^{-\Delta X_s}. \end{aligned}$$

Then Theorem 20.7 yields

$$\begin{aligned} Z_t - 1 = {} & (Z_- \cdot Y)_t + (e^{Y_-} \cdot V)_t + \tfrac{1}{2}(Z_- \cdot [X]^c)_t \\ & + \sum_{s \le t} \Big(\Delta Z_s - Z_{s-}\Delta X_s - e^{Y_{s-}} \Delta V_s\Big). \end{aligned} \tag{7}$$

Now $e^{Y_-} \cdot V = \sum e^{Y_-} \Delta V$ since V is of pure jump type, and furthermore $\Delta Z = Z_- \Delta X$. Hence, the right-hand side of (7) simplifies to $Z_- \cdot X$, as desired.

To prove the uniqueness, let Z be an arbitrary solution, and put $V = Ze^{-Y}$, where $Y = X - \frac{1}{2}[X]^c$ as before. Then Theorem 20.7 gives

$$\begin{aligned} V_t - 1 &= (e^{-Y_-} \cdot Z)_t - (V_- \cdot Y)_t + \tfrac{1}{2}(V_- \cdot [X]^c)_t - (e^{-Y_-} \cdot [X, Z]^c)_t \\ &\quad + \sum_{s \le t} \Big(\Delta V_s + V_{s-}\Delta Y_s - e^{-Y_{s-}} \Delta Z_s\Big) \\ &= (V_- \cdot X)_t - (V_- \cdot X)_t + \tfrac{1}{2}(V_- \cdot [X]^c)_t + \tfrac{1}{2}(V_- \cdot [X]^c)_t - (V_- \cdot [X]^c)_t \\ &\quad + \sum_{s \le t} \Big(\Delta V_s + V_{s-}\Delta X_s - V_{s-}\Delta X_s\Big) \\ &= \sum_{s \le t} \Delta V_s. \end{aligned}$$

Thus, V is purely discontinuous of locally finite variation. We may further compute

$$\begin{aligned} \Delta V_t &= Z_t\, e^{-Y_t} - Z_{t-}\, e^{-Y_{t-}} \\ &= (Z_{t-} + \Delta Z_t)\, e^{-Y_{t-} - \Delta Y_t} - Z_{t-}\, e^{-Y_{t-}} \\ &= V_{t-} \Big\{(1 + \Delta X_t)\, e^{-\Delta X_t} - 1\Big\}, \end{aligned}$$

which shows that

$$V_t = 1 + (V_- \cdot A)_t,$$
$$A_t = \sum_{s \le t} \Big\{ (1 + \Delta X_t)\, e^{-\Delta X_t} - 1 \Big\}.$$

It remains to show that the homogeneous equation $V = V_- \cdot A$ has the unique solution $V = 0$. Then define $R_t = \int_{(0,t]} |dA|$, and conclude from Theorem 20.7 and the convexity of the function $x \mapsto x^n$ that

$$\begin{aligned} R_t^n &= n\,(R_-^{n-1} \cdot R)_t + \sum_{s \le t} \Big(\Delta R_s^n - n\, R_{s-}^{n-1} \Delta R_s \Big) \\ &\ge n\,(R_-^{n-1} \cdot R)_t. \end{aligned} \tag{8}$$

We prove by induction that

$$V_t^* \le \frac{V_t^* R_t^n}{n!}, \quad t \ge 0, \ \ n \in \mathsf{Z}_+. \tag{9}$$

This is obvious for $n = 0$, and if it holds for $n - 1$, then by (8)

$$\begin{aligned} V_t^* &= (V_- \cdot A)_t^* \\ &\le \frac{V_t^* (R_-^{n-1} \cdot R)_t}{(n-1)!} \le \frac{V_t^* R_t^n}{n!}, \end{aligned}$$

as required. Since $R_t^n / n! \to 0$ as $n \to \infty$, relation (9) yields $V_t^* = 0$ for all $t > 0$. □

The equation $Z = 1 + Z_- \cdot X$ arises naturally in connection with changes of probability measure. Here we extend Theorem 19.20 to general local martingales.

Theorem 20.9 *(change of measure, van Schuppen & Wong) Consider some probability measures P, Q and a local martingale $Z \ge 0$, such that $Q = Z_t \cdot P$ on $\mathcal{F}_t$ for all $t \ge 0$, and let M be a local P-martingale such that $[M, Z]$ has locally integrable variation and P-compensator $\langle M, Z \rangle$. Then we have the local Q-martingale*

$$\tilde{M} = M - Z_-^{-1} \cdot \langle M, Z \rangle.$$

A lemma is needed for the proof.

Lemma 20.10 *(integration by parts) For any semi-martingale X with $X_0 = 0$ and predictable process A of locally finite variation, we have*

$$A X = A \cdot X + X_- \cdot A \quad a.s.$$

Proof: We need to show that $\Delta A \cdot X = [A, X]$ a.s., which is equivalent to

$$\int_{(0,t]} (\Delta A_s)\, dX_s = \sum_{s \le t} (\Delta A_s)(\Delta X_s), \quad t \ge 0,$$

by Theorem 20.6 (viii). Since the series on the right is absolutely convergent by Cauchy's inequality, we may reduce, by dominated convergence on each side, to the case where A is constant apart from finitely many jumps. By Lemma 10.3 and Theorem 10.14, we may proceed to the case where A has at most one jump occurring at a predictable time τ. Introducing an announcing sequence (τ_n) and writing $Y = \Delta A \cdot X$, we get by Theorem 20.2 (v)

$$Y_{\tau_n \wedge t} = 0 = Y_t - Y_{t \wedge \tau} \;\; \text{a.s.}, \quad t \geq 0, \;\; n \in \mathsf{N}.$$

Thus, even Y is constant, apart from a possible jump at τ. Finally, Theorem 20.2 (iv) yields $\Delta Y_\tau = (\Delta A_\tau)(\Delta X_\tau)$ a.s. on $\{\tau < \infty\}$. □

Proof of Theorem 20.9: Let $\tau_n = \inf\{t; Z_t < 1/n\}$ for $n \in \mathsf{N}$, and note that $\tau_n \to \infty$ a.s. Q by Lemma 19.18. Hence, $\tilde{M}$ is well defined under Q, and it suffices as in Lemma 19.18 to show that $(\tilde{M}Z)^{\tau_n}$ is a local P-martingale for every n. Writing $\stackrel{m}{=}$ for equality up to a local P-martingale, we see from Lemma 20.10 with $X = Z$ and $A = Z_-^{-1} \cdot \langle M, Z \rangle$ that, on every interval $[0, \tau_n]$,

$$\begin{aligned} MZ &\stackrel{m}{=} [M, Z] \stackrel{m}{=} \langle M, Z \rangle \\ &= Z_- \cdot A \stackrel{m}{=} AZ. \end{aligned}$$

Thus, $\tilde{M}Z = (M - A)\, Z \stackrel{m}{=} 0$, as required. □

Using the last theorem, we may show that the class of semi-martingales is closed under absolutely continuous changes of probability measure. A special case of this result was previously obtained as part of Proposition 19.21.

Corollary 20.11 *(preservation law, Jacod) If $Q \ll P$ on $\mathcal{F}_t$ for all $t > 0$, then every P-semi-martingale is also a Q-semi-martingale.*

Proof: Let $Q = Z_t \cdot P$ on $\mathcal{F}_t$ for all $t \geq 0$. We need to show that every local P-martingale M is a semi-martingale under Q. By Lemma 20.5, we may then take ΔM to be bounded, so that $[M]$ is locally bounded. By Theorem 20.9 it suffices to show that $[M, Z]$ has locally integrable variation, and by Theorem 20.6 (iii) it is then enough to prove that $[Z]^{1/2}$ is locally integrable. By Theorem 20.6 (iv),

$$\begin{aligned} [Z]_t^{1/2} &\leq [Z]_{t-}^{1/2} + |\Delta Z_t| \\ &\leq [Z]_{t-}^{1/2} + Z_{t-}^* + |Z_t|, \quad t \geq 0, \end{aligned}$$

and so the desired integrability follows by optional sampling. □

Next we extend the *BDG inequalities* of Theorem 18.7 to general local martingales. Such extensions are possible only for exponents $p \geq 1$.

Theorem 20.12 *(norm comparison, Burkholder, Davis, Gundy) For any local martingale M with $M_0 = 0$, we have*[4]

$$E M^{*p} \asymp E[M]_\infty^{p/2}, \quad p \geq 1. \tag{10}$$

[4]Here $M^* = \sup_t |M_t|$, and $f \asymp g$ means that $f \leq cg$ and $g \leq cf$ for some constant $c > 0$. The domination constants are understood to depend only on p.

In particular, we conclude as in Corollary 18.8 that M is a uniformly integrable martingale whenever $E[M]_\infty^{1/2} < \infty$.

Proof for $p = 1$ (Davis): Exploiting the symmetry of the argument, we write $M^\flat$ and $M^\sharp$ for the processes M^* and $[M]^{1/2}$, taken in either order. Putting $J = \Delta M$, we define

$$A_t = \sum_{s \le t} J_s 1\Big\{|J_s| > 2J^*_{s-}\Big\}, \quad t \ge 0.$$

Since $|\Delta A| \le 2\Delta J^*$, we have

$$\begin{aligned}\int_0^\infty |dA_s| &= \textstyle\sum_s |\Delta A_s| \\ &\le 2J^* \le 4M^\sharp_\infty.\end{aligned}$$

Writing $\hat{A}$ for the compensator of A and putting $D = A - \hat{A}$, we get

$$\begin{aligned} ED^\flat_\infty \vee ED^\sharp_\infty &\le E\int_0^\infty |dD_s| \\ &\lesssim E\int_0^\infty |dA_s| \lesssim EM^\sharp_\infty. \end{aligned} \tag{11}$$

To get a similar estimate for $N = M - D$, we introduce the optional times

$$\tau_r = \inf\Big\{t;\, N^\sharp_t \vee J^*_t > r\Big\}, \quad r > 0,$$

and note that

$$\begin{aligned} P\{N^\flat_\infty > r\} &\le P\{\tau_r < \infty\} + P\Big\{\tau_r = \infty,\, N^\flat_\infty > r\Big\} \\ &\le P\{N^\sharp_\infty > r\} + P\{J^* > r\} + P\{N^\flat_{\tau_r} > r\}. \end{aligned} \tag{12}$$

Arguing as in the proof of Lemma 20.5, we get $|\Delta N| \le 4J^*_-$, and so

$$\begin{aligned} N^\sharp_{\tau_r} &\le N^\sharp_\infty \wedge \Big(N^\sharp_{\tau_r-} + 4J^*_{\tau_r-}\Big) \\ &\le N^\sharp_\infty \wedge 5r. \end{aligned}$$

Since $N^2 - [N]$ is a local martingale, we get by Chebyshev's inequality or Proposition 9.16, respectively,

$$\begin{aligned} r^2 P\{N^\flat_{\tau_r} > r\} &\lesssim EN^{\sharp 2}_{\tau_r} \\ &\lesssim E(N^\sharp_\infty \wedge r)^2. \end{aligned}$$

Hence, by Fubini's theorem and calculus,

$$\int_0^\infty P\{N^\flat_{\tau_r} > r\}dr \lesssim \int_0^\infty E(N^\sharp_\infty \wedge r)^2\, r^{-2}dr \lesssim EN^\sharp_\infty.$$

Combining this with (11)−(12) and using Lemma 4.4, we get

$$\begin{aligned} EN^\flat_\infty &= \int_0^\infty P\{N^\flat_\infty > r\}dr \\ &\le \int_0^\infty \Big(P\{N^\sharp_\infty > r\} + P\{J^* > r\} + P\{N^\flat_{\tau_r} > r\}\Big)\, dr \\ &\lesssim EN^\sharp_\infty + EJ^* \lesssim EM^\sharp_\infty. \end{aligned}$$

It remains to note that $EM^\flat_\infty \le ED^\flat_\infty + EN^\flat_\infty$. □

Extension to $p > 1$ (Garsia): For any $t \ge 0$ and $B \in \mathcal{F}_t$, we may apply (10) with $p = 1$ to the local martingale $1_B(M - M^t)$ to get a.s.

$$\begin{aligned} c_1^{-1} E\Big([M - M^t]^{1/2}_\infty \,\Big|\, \mathcal{F}_t \Big) &\le E\Big\{ (M - M^t)^*_\infty \Big| \mathcal{F}_t \Big\} \\ &\le c_1 E\Big([M - M^t]^{1/2}_\infty \,\Big|\, \mathcal{F}_t \Big). \end{aligned}$$

Since

$$\begin{aligned} [M]^{1/2}_\infty - [M]^{1/2}_t &\le [M - M^t]^{1/2}_\infty \\ &\le [M]^{1/2}_\infty, \\ M^*_\infty - M^*_t &\le (M - M^t)^*_\infty \\ &\le 2M^*_\infty, \end{aligned}$$

Proposition 10.20 yields $E(A_\infty - A_t \,|\, \mathcal{F}_t) \lesssim E(\zeta \,|\, \mathcal{F}_t)$ with $A_t = [M]^{1/2}_t$ and $\zeta = M^*$, and also with $A_t = M^*_t$ and $\zeta = [M]^{1/2}_\infty$. Since

$$\begin{aligned} \Delta M^*_t &\le \Delta [M]^{1/2}_t \\ &= |\Delta M_t| \\ &\le [M]^{1/2}_t \wedge 2M^*_t, \end{aligned}$$

we have in both cases $\Delta A_\tau \lesssim E(\zeta \,|\, \mathcal{F}_\tau)$ a.s. for any optional time τ, and so the cited condition remains fulfilled for the left-continuous version A_-. Hence, Proposition 10.20 yields $\|A_\infty\|_p \lesssim \|\zeta\|_p$ for every $p \ge 1$, and (10) follows. □

The last theorem allows us to extend the stochastic integral to a larger class of integrands. Then write $\mathcal{M}$ for the space of local martingales, and let $\mathcal{M}_0$ be the subclass of processes M with $M_0 = 0$. For any $M \in \mathcal{M}$, let $L(M)$ be the class of predictable processes V such that $(V^2 \cdot [M])^{1/2}$ is locally integrable.

Theorem 20.13 *(martingale integral, Meyer) The elementary predictable integral extends a.s. uniquely to a bilinear map of any $M \in \mathcal{M}$ and $V \in L(M)$ into a process $V \cdot M \in \mathcal{M}_0$, satisfying*

(i) $|V_n| \le V$ *in* $L(M)$ *with* $(V_n^2 \cdot [M])_t \xrightarrow{P} 0 \;\Rightarrow\; (V_n \cdot M)^*_t \xrightarrow{P} 0$,

(ii) *properties* (iii)−(v) *of Theorem 20.2,*

(iii) $[V \cdot M, N] = V \cdot [M, N]$ *a.s.,* $N \in \mathcal{M}$.

The integral $V \cdot M$ is a.s. uniquely determined by (iii).

Proof: To construct the integral, we may reduce by localization to the case where $E(M - M_0)^* < \infty$ and $E(V^2 \cdot [M])^{1/2}_\infty < \infty$. For each $n \in \mathsf{N}$, define $V_n = V 1\{|V| \le n\}$. Then $V_n \cdot M \in \mathcal{M}_0$ by Theorem 20.4, and by Theorem 20.12 we have $E(V_n \cdot M)^* < \infty$. Using Theorems 20.6 (v) and 20.12, Minkowski's inequality, and dominated convergence, we obtain

$$\begin{aligned} E\Big(V_m \cdot M - V_n \cdot M\Big)^* &\lesssim E\Big[(V_m - V_n) \cdot M\Big]^{1/2}_\infty \\ &= E\Big\{(V_m - V_n)^2 \cdot [M]\Big\}^{1/2}_\infty \to 0. \end{aligned}$$

Hence, there exists a process $V \cdot M$ with $E(V_n \cdot M - V \cdot M)^* \to 0$, and we note that $V \cdot M \in \mathcal{M}_0$ and $E(V \cdot M)^* < \infty$.

To prove (iii), we note that the relation holds for each V_n by Theorem 20.6 (v). Since $E[V_n \cdot M - V \cdot M]_\infty^{1/2} \to 0$ by Theorem 20.12, we get by Theorem 20.6 (iii) for any $N \in \mathcal{M}$ and $t \geq 0$

$$\Big| [V_n \cdot M, N]_t - [V \cdot M, N]_t \Big|^2 \leq \Big[V_n \cdot M - V \cdot M\Big]_t \, [N]_t \xrightarrow{P} 0. \tag{13}$$

Next we see from Theorem 20.6 (iii) and (v) that

$$\begin{aligned}\int_0^t \Big| V_n \, d[M,N] \Big| &= \int_0^t \Big| d[V_n \cdot M, N] \Big| \\ &\leq [V_n \cdot M]_t^{1/2} \, [N]_t^{1/2}.\end{aligned}$$

As $n \to \infty$, we get by monotone convergence on the left and Minkowski's inequality on the right

$$\int_0^t \Big| V d[M,N] \Big| \leq [V \cdot M]_t^{1/2} \, [N]_t^{1/2} < \infty.$$

Hence, $V_n \cdot [M,N] \to V \cdot [M,N]$ by dominated convergence, and (iii) follows by combination with (13).

To see that $V \cdot M$ is determined by (iii), it remains to note that, if $[M] = 0$ a.s. for some $M \in \mathcal{M}_0$, then $M^* = 0$ a.s. by Theorem 20.12. To prove the stated continuity property, we may reduce by localization to the case where $E(V^2 \cdot [M])_\infty^{1/2} < \infty$. But then $E(V_n^2 \cdot [M])_\infty^{1/2} \to 0$ by dominated convergence, and Theorem 20.12 yields $E(V_n \cdot M)^* \to 0$. To prove the uniqueness of the integral, it is enough to consider bounded integrands V. We may then approximate as in Lemma 20.3 by uniformly bounded processes $V_n \in \mathcal{E}$ with $\{(V_n - V)^2 \cdot [M]\}_\infty \xrightarrow{P} 0$, and conclude that $(V_n \cdot M - V \cdot M)^* \xrightarrow{P} 0$.

Of the remaining properties in Theorem 20.2, assertion (iii) may be proved as before by means of (iii) above, whereas properties (iv)−(v) follow easily by truncation from the corresponding statements in Theorem 20.4. □

A semi-martingale $X = M + A$ is said to be *purely discontinuous*, if there exist some local martingales $M^1, M^2, \ldots$ of locally finite variation, such that $E(M - M^n)_t^{*2} \to 0$ for every $t > 0$. The property is clearly independent of the choice of decomposition $X = M + A$. To motivate the terminology, note that any martingale M of locally finite variation can be written as $M = M_0 + A - \hat{A}$, where $A_t = \sum_{s \leq t} \Delta M_s$ and $\hat{A}$ is the compensator of A. Thus, $M - M_0$ is then a compensated sum of jumps.

We proceed to establish a basic decomposition[5] of a general semi-martingale X into continuous and purely discontinuous components, corresponding to the elementary decomposition of the quadratic variation $[X]$ into a continuous part and a jump part.

[5] Not to be confused with the decomposition of functions of bounded variation in Theorem 2.23 (iii). The distinction is crucial for part (ii) below, which involves both notions.

Theorem 20.14 *(semi-martingale decomposition, Yoeurp, Meyer) For any real semi-martingale X, we have*

(i) $X = X_0 + X^c + X^d$ *a.s. uniquely, for a continuous local martingale X^c with $X_0^c = 0$ and a purely discontinuous semi-martingale X^d,*

(ii) $[X^c] = [X]^c$ *and* $[X^d] = [X]^d$ *a.s.*

Proof: (i) It is enough to consider the martingale component in an arbitrary decomposition $X = X_0 + M + A$, and by Lemma 20.5 we may take $M \in \mathcal{M}^2_{0,\mathrm{loc}}$. Then choose some optional times $\tau_n \uparrow \infty$ with $\tau_0 = 0$, such that $M^{\tau_n} \in \mathcal{M}_0^2$ for all n. It is enough to construct the desired decomposition for each process $M^{\tau_n} - M^{\tau_{n-1}}$, which reduces the discussion to the case of $M \in \mathcal{M}_0^2$. Now let $\mathcal{C}$ and $\mathcal{D}$ be the classes of continuous and purely discontinuous processes in $\mathcal{M}_0^2$, and note that both are closed linear subspaces of the Hilbert space $\mathcal{M}_0^2$. The desired decomposition will follow from Theorem 1.35, if we can show that $\mathcal{D}^\perp \subset \mathcal{C}$.

First let $M \in \mathcal{D}^\perp$. To see that M is continuous, fix any $\varepsilon > 0$, and put $\tau = \inf\{t;\, \Delta M_t > \varepsilon\}$. Define $A_t = 1\{\tau \le t\}$, let $\hat{A}$ be the compensator of A, and put $N = A - \hat{A}$. Integrating by parts and using Lemma 10.13 gives

$$\begin{aligned} \tfrac{1}{2}\, E\hat{A}_\tau^2 &\le E\int \hat{A}\, d\hat{A} \\ &= E\int \hat{A}\, dA = E\hat{A}_\tau \\ &= EA_\tau \le 1. \end{aligned}$$

Thus, N is L^2-bounded and hence lies in $\mathcal{D}$. For any bounded martingale M', we get

$$\begin{aligned} E(M'_\infty N_\infty) &= E\int M' dN \\ &= E\int (\Delta M')\, dN \\ &= E\int (\Delta M')\, dA \\ &= E\Big(\Delta M'_\tau;\, \tau < \infty\Big), \end{aligned}$$

where the first equality is obtained as in the proof of Lemma 10.7, the second holds by the predictability of M'_-, and the third holds since $\hat{A}$ is predictable and hence natural. Letting $M' \to M$ in $\mathcal{M}^2$, we obtain

$$\begin{aligned} 0 &= E(M_\infty N_\infty) \\ &= E\Big(\Delta M_\tau;\, \tau < \infty\Big) \\ &\ge \varepsilon\, P\{\tau < \infty\}. \end{aligned}$$

Thus, $\Delta M \le \varepsilon$ a.s., and so $\Delta M \le 0$ a.s. since ε is arbitrary. Similarly, $\Delta M \ge 0$ a.s., and the desired continuity follows.

Next let $M \in \mathcal{D}$ and $N \in \mathcal{C}$, and choose martingales $M^n \to M$ of locally finite variation. By Theorem 20.6 (vi) and (vii) and optional sampling, we get for any optional time τ

$$\begin{aligned} 0 &= E[M^n, N]_\tau \\ &= E(M^n_\tau N_\tau) \\ &\to E(M_\tau N_\tau) \\ &= E[M, N]_\tau, \end{aligned}$$

and so $[M, N]$ is a martingale by Lemma 9.14. Since it is also continuous by (14), Proposition 18.2 yields $[M, N] = 0$ a.s. In particular, $E(M_\infty N_\infty) = 0$, which shows that $\mathcal{C} \perp \mathcal{D}$. The asserted uniqueness now follows easily.

(ii) By Theorem 20.6 (iv), we have for any $M \in \mathcal{M}^2$

$$[M]_t = [M]^c_t + \sum_{s \le t} (\Delta M_s)^2, \quad t \ge 0. \tag{14}$$

If $M \in \mathcal{D}$, we may choose some martingales of locally finite variation $M^n \to M$. By Theorem 20.6 (vii) and (viii), we get $[M^n]^c = 0$ and $E[M^n - M]_\infty \to 0$. For any $t \ge 0$, we get by Minkowski's inequality and (14)

$$\begin{aligned} \left| \Big\{ \sum_{s \le t} (\Delta M^n_s)^2 \Big\}^{1/2} - \Big\{ \sum_{s \le t} (\Delta M_s)^2 \Big\}^{1/2} \right| &\le \Big\{ \sum_{s \le t} \Big(\Delta M^n_s - \Delta M_s \Big)^2 \Big\}^{1/2} \\ &\le [M^n - M]^{1/2}_t \xrightarrow{P} 0, \\ \left| [M^n]^{1/2}_t - [M]^{1/2}_t \right| &\le [M^n - M]^{1/2}_t \xrightarrow{P} 0. \end{aligned}$$

Taking limits in (14) for the martingales M^n, we get the same formula for M without the term $[M]^c_t$, which shows that $[M] = [M]^d$.

Now consider any $M \in \mathcal{M}^2$. Using the strong orthogonality $[M^c, M^d] = 0$, we get a.s.

$$\begin{aligned} [M]^c + [M]^d = [M] &= [M^c + M^d] \\ &= [M^c] + [M^d], \end{aligned}$$

which shows that even $[M^c] = [M]^c$ a.s. By the same argument combined with Theorem 20.6 (viii), we obtain $[X^d] = [X]^d$ a.s. for any semi-martingale X. □

The last result yields an explicit formula for the covariation of two semi-martingales.

Corollary 20.15 *(decomposition of covariation) For any semi-martingales X, Y, we have*

(i) *X^c is the a.s. unique continuous local martingale M with $M_0 = 0$, such that $[X - M]$ is a.s. purely discontinuous,*

(ii) $[X, Y]_t = [X^c, Y^c]_t + \sum_{s \le t} (\Delta X_s)(\Delta Y_s)$ *a.s.,* $\quad t \ge 0$.

In particular, $(V \cdot X)^c = V \cdot X^c$ a.s. for any semi-martingale X and locally bounded, predictable process V, and $[X^c, Y^c] = [X, Y]^c$ a.s.

Proof: If M has the stated properties, then $[(X - M)^c] = [X - M]^c = 0$ a.s., and so $(X - M)^c = 0$ a.s. Thus, $X - M$ is purely discontinuous. Formula (ii) holds by Theorems 20.6 (iv) and 20.14 when $X = Y$, and the general result follows by polarization. □

The purely discontinuous component of a local martingale has a further decomposition, similar to the decompositions of optional times and increasing processes in Propositions 10.4 and 10.17.

Corollary 20.16 *(martingale decomposition, Yoeurp) For any purely discontinuous, local martingale M, we have*

(i) $M = M_0 + M^q + M^a$ *a.s. uniquely, where* $M^q, M^a \in \mathcal{M}_0$ *are purely discontinuous,* M^q *is ql-continuous, and* M^a *has accessible jumps,*

(ii) $\left\{t;\, \Delta M_t^a \neq 0\right\} \subset \bigcup_n [\tau_n]$ *a.s. for some predictable times* $\tau_1, \tau_2, \ldots$ *with disjoint graphs,*

(iii) $[M^q] = [M]^q$ *and* $[M^a] = [M]^a$ *a.s.,*

(iv) $\langle M^q \rangle = \langle M \rangle^c$ *and* $\langle M^a \rangle = \langle M \rangle^d$ *a.s. when* $M \in \mathcal{M}^2_{\mathrm{loc}}$.

Proof: Form a locally integrable process

$$A_t = \sum_{s \le t} \left\{(\Delta M_s)^2 \wedge 1\right\}, \quad t \ge 0,$$

with compensator $\hat{A}$, and define

$$\begin{aligned} M^q &= M - M_0 - M^a \\ &= 1\left\{\Delta \hat{A}_t = 0\right\} \cdot M. \end{aligned}$$

Then Theorem 20.4 yields $M^q, M^a \in \mathcal{M}_0$ and $\Delta M^q = 1\{\Delta \hat{A} = 0\}\Delta M$ a.s., and M^q and M^a are purely discontinuous by Corollary 20.15. The proof may now be completed as in case of Proposition 10.17. □

We illustrate the use of the previous decompositions by proving two exponential inequalities for martingales with bounded jumps.

Theorem 20.17 *(exponential inequalities, OK & Sztencel) Let M be a local martingale with $M_0 = 0$, such that $|\Delta M| \le c$ for a constant $c \le 1$. Then*

(i) *if* $[M]_\infty \le 1$ *a.s., we have*

$$P\{M^* \ge r\} \lesssim \exp\left\{-\frac{r^2}{2(1+rc)}\right\}, \quad r \ge 0,$$

(ii) *if* $\langle M \rangle_\infty \le 1$ *a.s., we have*

$$P\{M^* \ge r\} \lesssim \exp\left\{-r\,\frac{\log(1+rc)}{2\,c}\right\}, \quad r \ge 0.$$

For continuous martingales M, both bounds reduce to $e^{-r^2/2}$, which can also be seen directly by more elementary methods. For the proof of Theorem 20.17 we need two lemmas, beginning with a characterization of certain pure jump-type martingales.

Lemma 20.18 *(accessible jump-type martingales) Let N be a pure jump-type process with integrable variation and accessible jumps. Then N is a martingale iff*

$$E\Big(\Delta N_\tau \Big| \mathcal{F}_{\tau-}\Big) = 0 \;\; a.s., \quad \tau < \infty \;\; predictable.$$

Proof: Proposition 10.17 yields some predictable times $\tau_1, \tau_2, \ldots$ with disjoint graphs such that $\{t > 0;\, \Delta N_t \neq 0\} \subset \bigcup_n [\tau_n]$. Under the stated condition, we get by Fubini's theorem and Lemma 10.1, for any bounded optional time τ,

$$\begin{aligned} EN_\tau &= \textstyle\sum_n E\Big(\Delta N_{\tau_n};\, \tau_n \le \tau\Big) \\ &= \textstyle\sum_n E\Big\{E\Big(\Delta N_{\tau_n}\Big|\, \mathcal{F}_{\tau_n-}\Big);\, \tau_n \le \tau\Big\} = 0, \end{aligned}$$

and so N is a martingale by Lemma 9.14. Conversely, given any uniformly integrable martingale N and finite predictable time τ, we have a.s. $E(N_\tau | \mathcal{F}_{\tau-}) = N_{\tau-}$ and hence $E(\Delta N_\tau | \mathcal{F}_{\tau-}) = 0$. □

Though for general martingales M the process $Z = e^{M-[M]/2}$ of Lemma 19.22 may fail to be a martingale, it can often be replaced by a similar super-martingale.

Lemma 20.19 *(exponential super-martingales) Let M be a local martingale with $M_0 = 0$ and $|\Delta M| \le c < \infty$ a.s., and define $a = f(c)$ and $b = g(c)$, where*

$$f(x) = -\,\frac{x + \log(1-x)_+}{x^2}, \qquad g(x) = \frac{e^x - 1 - x}{x^2}.$$

Then we have the super-martingales

$$\begin{aligned} X_t &= \exp\Big(M_t - a\,[M]_t\Big), \\ Y_t &= \exp\Big(M_t - b\,\langle M\rangle_t\Big), \quad t \ge 0. \end{aligned}$$

Proof: In case of X, we may clearly assume that $c < 1$. By Theorem 20.7 we get, in an obvious shorthand notation,

$$(X_-^{-1} \cdot X)_t = M_t - (a - \tfrac{1}{2})\,[M]_t^c + \sum_{s \le t} \Big\{ e^{\Delta M_s - a(\Delta M_s)^2} - 1 - \Delta M_s \Big\}.$$

Here the first term on the right is a local martingale, and the second term is non-increasing since $a \ge \frac{1}{2}$. To see that even the sum is non-increasing, we need to show that $\exp(x - ax^2) \le 1 + x$ or $f(-x) \le f(c)$ whenever $|x| \le c$, which is clear by a Taylor expansion of each side. Thus, $(X_-^{-1}) \cdot X$ is a local super-martingale, which remains true for $X_- \cdot (X_-^{-1} \cdot X) = X$ since $X > 0$. By Fatou's lemma, X is then a true super-martingale.

In case of Y, Theorem 20.14 and Proposition 20.16 yield a decomposition $M = M^c + M^q + M^a$, and Theorem 20.7 gives

$$\begin{aligned} (Y_-^{-1} \cdot Y)_t &= M_t - b\,\langle M\rangle_t^c + \tfrac{1}{2}\,[M]_t^c + \sum_{s \le t} \Big(e^{\Delta M_s - b\,\Delta\langle M\rangle_s} - 1 - \Delta M_s \Big) \\ &= M_t + b\,\Big([M^q]_t - \langle M^q\rangle_t\Big) - (b - \tfrac{1}{2})\,[M]_t^c \end{aligned}$$

$$+\sum_{s\le t}\left\{e^{\Delta M_s - b\,\Delta\langle M\rangle_s} - \frac{1+\Delta M_s + b\,(\Delta M_s)^2}{1+b\,\Delta\langle M\rangle_s}\right\}$$
$$+\sum_{s\le t}\left\{\frac{1+\Delta M_s^a + b\,(\Delta M_s^a)^2}{1+b\,\Delta\langle M^a\rangle_s} - 1 - \Delta M_s^a\right\}.$$

Here the first two terms on the right are martingales, and the third term is non-increasing since $b \ge \frac{1}{2}$. Even the first sum of jumps is non-increasing, since $e^x - 1 - x \le b\,x^2$ for $|x| \le c$ and $e^y \le 1 + y$ for $y \ge 0$.

The last sum clearly defines a purely discontinuous process N with locally finite variation and accessible jumps. Fixing any finite predictable time τ, and writing $\xi = \Delta M_\tau$ and $\eta = \Delta\langle M\rangle_\tau$, we note that

$$\begin{aligned} E\left|\frac{1+\xi+b\,\xi^2}{1+b\,\eta} - 1 - \xi\right| &\le E\Big|\,1+\xi+b\,\xi^2 - (1+\xi)(1+b\,\eta)\,\Big| \\ &= b\,E\Big|\,\xi^2 - (1+\xi)\,\eta\,\Big| \\ &\le b\,(2+c)\,E\xi^2. \end{aligned}$$

Since

$$\begin{aligned} E\,\textstyle\sum_t (\Delta M_t)^2 &\le E[M]_\infty \\ &= E\langle M\rangle_\infty \le 1, \end{aligned}$$

we conclude that N has integrable total variation. Furthermore, Lemmas 10.3 and 20.18 show that a.s. $E(\xi\,|\mathcal{F}_{\tau-}) = 0$ and

$$\begin{aligned} E(\,\xi^2\,|\,\mathcal{F}_{\tau-}) &= E\Big(\Delta[M]_\tau\Big|\,\mathcal{F}_{\tau-}\Big) \\ &= E(\,\eta\,|\mathcal{F}_{\tau-}) = \eta. \end{aligned}$$

Thus,

$$E\left\{\frac{1+\xi+b\,\xi^2}{1+b\,\eta} - 1 - \xi\,\middle|\,\mathcal{F}_{\tau-}\right\} = 0,$$

and Lemma 20.18 shows that N is a martingale. The proof may now be completed as before. □

Proof of Theorem 20.17: (i) Fix any $u > 0$, and conclude from Lemma 20.19 that the process

$$X_t^u = \exp\Big\{uM_t - u^2 f(uc)\,[M]_t\Big\}, \quad t \ge 0,$$

is a positive super-martingale. Since $[M] \le 1$ and $X_0^u = 1$, we get for any $r > 0$

$$\begin{aligned} P\Big\{\sup\nolimits_t M_t > r\Big\} &\le P\Big\{\sup\nolimits_t X_t^u > \exp\Big(ur - u^2 f(uc)\Big)\Big\} \\ &\le \exp\Big\{-ur + u^2 f(uc)\Big\}. \end{aligned} \tag{15}$$

The function $F(x) = 2xf(x)$ is clearly continuous and strictly increasing from $[0,1)$ onto R_+. Since also $F(x) \le x/(1-x)$, we have $F^{-1}(y) \ge y/(1+y)$. Choosing $u = F^{-1}(rc)/c$ in (15) gives

$$\begin{aligned} P\Big\{\sup\nolimits_t M_t > r\Big\} &\le \exp\left\{-\frac{rF^{-1}(rc)}{2\,c}\right\} \\ &\le \exp\left\{-\frac{r^2}{2\,(1+rc)}\right\}. \end{aligned}$$

It remains to combine with the same inequality for $-M$.

(ii) The function $G(x) = 2x\, g(x)$ is clearly continuous and strictly increasing onto R_+. Since also $G(x) \le e^x - 1$, we have $G^{-1}(y) \ge \log(1+y)$. Then as before,

$$P\Big\{\sup\nolimits_t M_t > r\Big\} \le \exp\Big\{-\frac{r\, G^{-1}(rc)}{2\, c}\Big\} \\ \le \exp\Big\{-\frac{r\, \log(1+rc)}{2\, c}\Big\},$$

and the result follows. □

By a *quasi-martingale* we mean an integrable, adapted, and right-continuous process X, such that

$$\sup_{\pi} \sum_{k\le n} E\left|X_{t_k} - E(X_{t_{k+1}}|\mathcal{F}_{t_k})\right| < \infty, \tag{16}$$

where π is the set of finite partitions of R_+ of the form $0 = t_0 < t_1 < \cdots < t_n < \infty$, and we take $t_{n+1} = \infty$ and $X_\infty = 0$ in the last term. In particular, (16) holds when X is the sum of an L^1-bounded martingale and a process of integrable variation starting at 0. We show that this is close to the general situation. Here localization is defined in the usual way in terms of some optional times $\tau_n \uparrow \infty$.

Theorem 20.20 *(quasi-martingales, Rao) A quasi-martingale is the difference of two non-negative super-martingales. Thus, for a process X with $X_0 = 0$, these conditions are equivalent:*

(i) *X is a local quasi-martingale,*
(ii) *X is a special semi-martingale.*

Proof: For any $t \ge 0$, let $\mathcal{P}_t$ be the class of partitions π of the interval $[t, \infty)$ of the form $t = t_0 < t_1 < \cdots < t_n$, and define

$$\eta_\pi^\pm = \sum_{k\le n} E\Big(\Big\{X_{t_k} - E(X_{t_{k+1}}|\mathcal{F}_{t_k})\Big\}_\pm\Big|\,\mathcal{F}_t\Big), \quad \pi \in \mathcal{P}_t,$$

where $t_{n+1} = \infty$ and $X_\infty = 0$, as before. We show that η_π^+ and η_π^- are a.s. non-decreasing under refinements of $\pi \in \mathcal{P}_t$. It is then enough to add one more division point u to π, say in the interval (t_k, t_{k+1}). Put $\alpha = X_{t_k} - X_u$ and $\beta = X_u - X_{t_{k+1}}$. By sub-additivity and Jensen's inequality, we get the desired relation

$$E\Big\{E\Big(\alpha + \beta\,\Big|\,\mathcal{F}_{t_k}\Big)_\pm\Big|\,\mathcal{F}_t\Big\} \le E\Big\{E(\alpha\,|\,\mathcal{F}_{t_k})_\pm + E(\beta\,|\,\mathcal{F}_{t_k})_\pm\Big|\,\mathcal{F}_t\Big\} \\ \le E\Big\{E(\alpha\,|\,\mathcal{F}_{t_k})_\pm + E(\beta\,|\,\mathcal{F}_u)_\pm\Big|\,\mathcal{F}_t\Big\}.$$

Now fix any $t \ge 0$, and conclude from (16) that $m_t^\pm \equiv \sup_{\pi\in\mathcal{P}_t} E\eta_\pi^\pm < \infty$. For every $n \in \mathsf{N}$, we may then choose $\pi_n \in \mathcal{P}_t$ with $E\eta_{\pi_n}^\pm > m_t^\pm - n^{-1}$. The sequences $(\eta_{\pi_n}^\pm)$ being Cauchy in L^1, they converge in L^1 toward some limits

$Y_t^\pm$. Further note that $E|\eta_\pi^\pm - Y_t^\pm| < n^{-1}$ whenever π is a refinement of π_n. Thus, $\eta_\pi^\pm \to Y_t^\pm$ in L^1 along the directed set $\mathcal{P}_t$.

Next fix any $s < t$, let $\pi \in \mathcal{P}_t$ be arbitrary, and define $\pi' \in \mathcal{P}_s$ by adding the point s to π. Then

$$\begin{aligned} Y_s^\pm \geq \eta_{\pi'}^\pm &= \Big\{X_s - E(X_t|\mathcal{F}_s)\Big\}_\pm + E(\eta_\pi^\pm \,|\, \mathcal{F}_s) \\ &\geq E(\eta_\pi^\pm \,|\, \mathcal{F}_s). \end{aligned}$$

Taking limits along $\mathcal{P}_t$ on the right, we get $Y_s^\pm \geq E(Y_t^\pm|\mathcal{F}_s)$ a.s., which means that the processes $Y^\pm$ are super-martingales. By Theorem 9.28, the right-hand limits along the rationals $Z_t^\pm = Y_{t+}^\pm$ then exist outside a fixed null set, and the processes $Z^\pm$ are right-continuous super-martingales. For $\pi \in \mathcal{P}_t$, we have $X_t = \eta_\pi^+ - \eta_\pi^- \to Y_t^+ - Y_t^-$, and so $Z_t^+ - Z_t^- = X_{t+} = X_t$ a.s. □

The following consequence is easy to believe but surprisingly hard to prove.

Corollary 20.21 *(non-random semi-martingales) Let X be a non-random process in $\mathbb{R}^d$. Then these conditions are equivalent:*

(i) *X is a semi-martingale,*

(ii) *X is rcll of locally finite variation.*

Proof (OK): Clearly (ii) ⇒ (i). Now assume (i) with $d = 1$. Then X is trivially a quasi-martingale, and so by Theorem 20.20 it is a special semi-martingale, hence $X = M + A$ for a local martingale M and a predictable process A of locally finite variation. Since $M = X - A$ is again predictable, it is continuous by Proposition 10.16. Then X and A have the same jumps, and so by subtraction we may take all processes to be continuous. Then $[M] = [X - A] = [X]$, which is non-random by Proposition 18.17, and we see as in Theorem 19.3 that M is a Brownian motion on the deterministic time scale $[X]$, hence has independent increments. Then $A = X - M$ has the same property, and since it is also continuous of locally finite variation, it is a.s. non-random by Proposition 14.4. Then even $M = X - A$ is a.s. non-random, hence 0 a.s., and so $X = A$, which has locally finite variation. □

Now we show that semi-martingales are the most general integrators for which a stochastic integral with reasonable continuity properties can be defined. As before, let $\mathcal{E}$ be the class of bounded, predictable step processes with jumps at finitely many fixed times.

Theorem 20.22 *(stochastic integrators, Bichteler, Dellacherie) A right-continuous, adapted process X is a semi-martingale iff, for any $V_1, V_2, \ldots \in \mathcal{E}$,*

$$\|V_n^*\|_\infty \to 0 \quad \Rightarrow \quad (V_n \cdot X)_t \xrightarrow{P} 0, \quad t > 0.$$

Our proof is based on three lemmas, beginning with the crucial functional-analytic part of the argument.

Lemma 20.23 *(convexity and tightness) For any convex, tight set $\mathcal{K} \subset L^1(P)$, there exists a bounded random variable $\rho > 0$ with*

$$\sup_{\xi \in \mathcal{K}} E(\rho\,\xi) < \infty.$$

Proof (Yan): Let $\mathcal{B}$ be the class of bounded, non-negative random variables, and define

$$\mathcal{C} = \Big\{ \gamma \in \mathcal{B};\, \sup_{\xi \in \mathcal{K}} E(\gamma\,\xi) < \infty \Big\}.$$

We claim that, for any $\gamma_1, \gamma_2, \ldots \in \mathcal{C}$, there exists a $\gamma \in \mathcal{C}$ with $\{\gamma > 0\} = \bigcup_n \{\gamma_n > 0\}$. Indeed, we may assume that $\gamma_n \le 1$ and $\sup_{\xi \in \mathcal{K}} E(\gamma_n\,\xi) \le 1$, in which case we may choose $\gamma = \sum_n 2^{-n} \gamma_n$. It is then easy to construct a $\rho \in \mathcal{C}$, such that $P\{\rho > 0\} = \sup_{\gamma \in \mathcal{C}} P\{\gamma > 0\}$. Clearly,

$$\{\gamma > 0\} \subset \{\rho > 0\} \ \text{ a.s.}, \quad \gamma \in \mathcal{C}, \tag{17}$$

since we could otherwise choose a $\rho' \in \mathcal{C}$ with $P\{\rho' > 0\} > P\{\rho > 0\}$.

To see that $\rho > 0$ a.s., suppose that instead $P\{\rho = 0\} > \varepsilon > 0$. Since $\mathcal{K}$ is tight, we may choose $r > 0$ so large that $P\{\xi > r\} \le \varepsilon$ for all $\xi \in \mathcal{K}$. Then $P\{\xi - \beta > r\} \le \varepsilon$ for all $\xi \in \mathcal{K}$ and $\beta \in \mathcal{B}$. By Fatou's lemma, we get $P\{\zeta > r\} \le \varepsilon$ for all ζ in the L^1-closure $\mathcal{Z} = \overline{\mathcal{K} - \mathcal{B}}$. In particular, the random variable $\zeta_0 = 2r1\{\rho = 0\}$ lies outside $\mathcal{Z}$. Now $\mathcal{Z}$ is convex and closed, and so, by a version of the Hahn–Banach theorem, there exists a $\gamma \in (L^1)^* = L^\infty$ satisfying

$$\begin{aligned} \sup_{\xi \in \mathcal{K}} E(\gamma\,\xi) - \inf_{\beta \in \mathcal{B}} E(\gamma\,\beta) &\le \sup_{\zeta \in \mathcal{Z}} E(\gamma\,\zeta) \\ &< E(\gamma\,\zeta_0) \\ &= 2rE(\gamma;\, \rho = 0). \end{aligned} \tag{18}$$

Here $\gamma \ge 0$, since we would otherwise get a contradiction by choosing $\beta = b\,1\{\gamma < 0\}$ for large enough $b > 0$. Hence, (18) reduces to

$$\sup_{\xi \in \mathcal{K}} E(\gamma\,\xi) < 2rE(\gamma;\, \rho = 0),$$

which implies $\gamma \in \mathcal{C}$ and $E(\gamma;\, \rho = 0) > 0$. But this contradicts (17), proving that indeed $\rho > 0$ a.s. □

Two further lemmas are needed for the proof of Theorem 20.22.

Lemma 20.24 *(tightness and boundedness) Let X be a right-continuous, adapted process, and let $\mathcal{T}$ be the class of optional times $\tau < \infty$ taking finitely many values. Then*

$$\{X_\tau;\, \tau \in \mathcal{T}\} \text{ is tight} \quad \Rightarrow \quad X^* < \infty \text{ a.s.}$$

Proof: By Lemma 9.4, any bounded optional time τ can be approximated from the right by optional times $\tau_n \in \mathcal{T}$, so that $X_{\tau_n} \to X_\tau$ by right continuity. Hence, Fatou's lemma yields

$$P\Big\{|X_\tau| > r\Big\} \le \liminf_{n \to \infty} P\Big\{|X_{\tau_n}| > r\Big\},$$

and so the hypothesis remains true with $\mathcal{T}$ replaced by the class $\hat{\mathcal{T}}$ of bounded optional times. By Lemma 9.6, the times

$$\tau_{t,n} = t \wedge \inf\Big\{s;\ |X_s| > n\Big\}, \quad t > 0,\ n \in \mathsf{N},$$

belong to $\hat{\mathcal{T}}$, and as $n \to \infty$ we get

$$\begin{aligned} P\{X^* > n\} &= \sup_{t>0} P\{X_t^* > n\} \\ &\le \sup_{\tau \in \hat{\mathcal{T}}} P\Big\{|X_\tau| > n\Big\} \to 0. \end{aligned}$$

□

Lemma 20.25 *(scaling) For any finite random variable ξ, there exists a bounded random variable $\rho > 0$ with $E|\rho\xi| < \infty$.*

Proof: We may take $\rho = (|\xi| \vee 1)^{-1}$. □

Proof of Theorem 20.22: The necessity is clear from Theorem 20.4. Now assume the stated condition. By Lemma 5.9, it is equivalent to assume for each $t > 0$ that the family $\mathcal{K}_t = \{(V \cdot X)_t;\ V \in \mathcal{E}_1\}$ is tight, where $\mathcal{E}_1 = \{V \in \mathcal{E};\ |V| \le 1\}$. The latter family is clearly convex, and by the linearity of the integral the convexity carries over to $\mathcal{K}_t$.

By Lemma 20.24 we have $X^* < \infty$ a.s., and so Lemma 20.25 yields a probability measure $Q \sim P$, such that $E_Q X_t^* = \int X_t^*\, dQ < \infty$. In particular, $\mathcal{K}_t \subset L^1(Q)$, and we note that $\mathcal{K}_t$ remains tight with respect to Q. Hence, Lemma 20.23 yields a probability measure $R \sim Q$ with bounded density $\rho > 0$, such that $\mathcal{K}_t$ is bounded in $L^1(R)$.

For any partition $0 = t_0 < t_1 < \cdots < t_n = t$, we note that

$$\sum_{k \le n} E_R\Big|X_{t_k} - E_R(X_{t_{k+1}}|\mathcal{F}_{t_k})\Big| = E_R(V \cdot X)_t + E_R|X_t|, \tag{19}$$

where

$$V_s = \sum_{k<n} \operatorname{sgn}\Big\{E_R(X_{t_{k+1}}|\mathcal{F}_{t_k}) - X_{t_k}\Big\}\, 1_{(t_k, t_{k+1}]}(s), \quad s \ge 0.$$

Since ρ is bounded and $V \in \mathcal{E}_1$, the right-hand side of (19) is bounded by a constant. Hence, the stopped process X^t is a quasi-martingale under R. By Theorem 20.20 it is then a semi-martingale for R, and since $P \sim R$, Corollary 20.11 shows that X^t remains a semi-martingale for P. Since t is arbitrary, we conclude that X itself is a semi-martingale under P. □

Exercises

1. Construct the quadratic variation $[M]$ of a local L^2-martingale M directly as in Theorem 18.5, and prove a corresponding version of the integration-by-parts formula. Use $[M]$ to define the L^2-integral of Theorem 20.2.

2. Show that the approximation in Proposition 18.17 remains valid for general semi-martingales.

3. For a local martingale M starting at 0 and an optional time τ, use Theorem 20.12 to give conditions for the relations $EM_\tau = 0$ and $EM_\tau^2 = [M]_\tau$.

4. Give an example of some L^2-bounded martingales M_n, such that $M_n^* \xrightarrow{P} 0$ and yet $\langle M_n \rangle_\infty \xrightarrow{P} \infty$. (*Hint:* Consider compensated Poisson processes with large jumps.)

5. Give an example of some martingales M_n, such that $[M_n]_\infty \xrightarrow{P} 0$ and yet $M_n^* \xrightarrow{P} \infty$. (*Hint:* See the preceding problem.)

6. Show that $\langle M_n \rangle_\infty \xrightarrow{P} 0$ implies $[M_n]_\infty \xrightarrow{P} 0$.

7. Give an example of a martingale M of bounded variation and a bounded, progressive process V, such that $V^2 \cdot \langle M \rangle = 0$ and yet $V \cdot M \neq 0$. Conclude that the L^2-integral in Theorem 20.2 has no continuous extension to progressive integrands.

8. Show that any general martingale inequality involving the processes M, $[M]$, and $\langle M \rangle$ remains valid in discrete time. (*Hint:* Embed M and the associated discrete filtration into a martingale and filtration on R_+.)

9. Show that the a.s. convergence in Theorem 5.23 remains valid in L^p. (*Hint:* Use Theorem 20.12 to reduce to the case where $p < 1$. Then truncate.)

10. Let $\mathcal{G}$ be an extension of the filtration $\mathcal{F}$. Show that any $\mathcal{F}$-adapted semi-martingale for $\mathcal{G}$ is also a semi-martingale for $\mathcal{F}$. Also show by an example that the converse implication fails in general. (*Hint:* Use Theorem 20.22.)

11. Show that if X is a Lévy process in R, then $[X]$ is a subordinator. Express the characteristics of $[X]$ in terms of those for X.

12. For a Lévy process X, show that if X is p-stable, then $[X]$ is strictly $p/2$-stable. Also prove the converse, in the case where X has positive or symmetric jumps. (*Hint:* Use Proposition 16.11.)

13. Extend Theorem 20.17 to the case where $[M]_\infty \leq a$ or $\langle M \rangle \leq a$ a.s. for some $a \geq 1$. (*Hint:* Apply the original result to a suitably scaled process.)

14. For a Lévy process X with Lévy measure ν, show that $X \in \mathcal{M}_{\mathrm{loc}}^2$ iff $X \in \mathcal{M}^2$, and also iff $\int x^2 \nu(dx) < \infty$, in which case $\langle X \rangle_t = t\, EX_1^2$. (*Hint:* Use Corollary 15.17.)

15. For a purely discontinuous local martingale M with positive jumps, show that $M - M_0$ is a.s. determined by $[M]$. (*Hint:* For any such processes M, N with $[M] = [N]$, apply Theorem 20.14 to $M - N$.)

16. Show that a semi-martingale X is ql-continuous or has accessible jumps iff $[X]$ has the same property. (*Hint:* Use Theorem 20.6 (iv).)

17. Show that a semi-martingale X with $|\Delta X| \leq c < \infty$ a.s. is a special semi-martingale with canonical decomposition $M + A$ satisfying $|\Delta A| \leq c$ a.s. In particular, X is a continuous semi-martingale iff it has a decomposition $M + A$ with M and A continuous. (*Hint:* Use Lemma 20.5, and note that $|\Delta A| \leq c$ a.s. implies $|\Delta \hat{A}| \leq c$ a.s.)

18. Show that a semi-martingale X is ql-continuous or has accessible jumps, iff it has a decomposition $M + A$ where M and A have the same property. Also show that, for special semi-martingales, we may choose $M + A$ to be the canonical decomposition of X. (*Hint:* Use Proposition 10.17 and Corollary 20.16, and refer to the preceding exercise.)

19. Show that a semi-martingale X is predictable, iff it is a special semi-martingale with canonical decomposition $M + A$ for a continuous M. (*Hint:* Use Proposition 10.16.)

Chapter 21

Malliavin Calculus

Malliavin derivative, integration by parts, closure, chain rule, divergence operator, chaos representations, norms and domains, conditioning and adaptation, factorization, commutation, covariance, local properties, Ornstein–Uhlenbeck semi-group and generator, elliptic operators, second order chain rule, extension and differentiation of Itô integrals, Brownian functionals, existence of smooth density, absolute continuity

The stochastic calculus of variations, known as Malliavin calculus, is an infinite-dimensional calculus on Wiener space of amazing power and beauty. Here we develop only some basic results for the differentiation and related operators, along with their representations and existence criteria in terms of multiple Wiener-Itô integrals. Already when developed that far, the theory is powerful enough to yield important new insights into various topics treated in earlier chapters. Unfortunately, we had to omit the more advanced aspects of the theory, including some fundamental Sobolev-type inequalities, with applications to solutions of SDEs, due to our obvious space limitations.

In this chapter we give a fairly complete treatment of the three basic operators of the theory—the *Malliavin derivative* D, the adjoint *divergence operator* D^*, even known as the *Skorohod integral*, and the *Ornstein–Uhlenbeck generator* L. Highlights of the theory include the explicit representations of those operators and their domains in terms of multiple Wiener–Itô integrals, the remarkable identity $L = -D^*D$, an expression of L as a second order differential operator, and some surprising connections to ordinary Itô integrals.

Most of the basic theory covered here relies on some standard functional analysis, combined with the theory of Wiener chaos expansions from Chapter 14. To ease the access for probabilists, I have replaced the sometimes peculiar traditional notation[1] by a more standard usage in probability theory, writing $\xi, \eta, \ldots$ for random variables and $X, Y, \ldots$ for random processes, and similarly for some basic notions of functional analysis. I have also tried to clarify the customary confusion[2] between the OU-semi-group and generator in the sense of Feller processes, and the corresponding notions of Malliavin calculus.

To begin our technical exposition of the theory, let ζ be an isonormal Gaussian process on a separable Hilbert space H, as defined in Chapter 14. Our first aim is to define a linear *differentiation operator* D on a suitable class

[1] where $F, G, \ldots$ are often used for random variables, $u, v, \ldots$ for random processes, δ with domain $D(\delta)$ for the adjoint of the operator D, etc.

[2] the former acting on deterministic functions, the latter on random variables

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_22

of ζ-measurable random variables. Our basic requirements are the relations $D(\zeta h) = h$ for all $h \in H$, along with a suitable chain rule for differentiation.

For motivation, we consider first the class $\mathcal{S}$ of *smooth* random variables of the form

$$\xi = f(\zeta\bar{h}) = f(\zeta h_1, \ldots, \zeta h_n), \quad h_1, \ldots, h_n \in H, \ n \in \mathsf{N}, \tag{1}$$

for functions $f \in C^1(\mathsf{R}^n)$ with bounded, first order partial derivatives $\partial_i f$, where we write $\bar{h} = (h_1, \ldots, h_n)$. The condition $D(\zeta h_i) = h_i$, along with obvious demands of linearity and continuity, suggest that we define

$$D\xi = \sum_{i\le n} \partial_i f(\zeta\bar{h})\, h_i, \quad \xi \in \mathcal{S}. \tag{2}$$

Then

$$\langle D\xi, h\rangle = \lim_{t\to 0} t^{-1}\Big\{ f\Big(\zeta\bar{h} + t\langle\bar{h}, h\rangle\Big) - f(\zeta\bar{h})\Big\}, \quad h \in H,$$

which identifies $\langle D\xi, h\rangle$ as the derivative of ξ in the direction of h. We may then think of $D\xi$ as a generalized directional derivative of ξ.

Note that the differentiation operator D maps random variables ξ into random elements $D\xi$ of H. In applications we often take H to be a function space $L^2(T, \mathcal{T}, \mu)$ whose elements are L^2-functions $h\colon T \to \mathsf{R}$, so that random elements in H may be regarded as real-valued processes on T, and $D\xi$ becomes a process $D_t\xi$ on T. In particular, we may take ζ to be an isonormal Gaussian process on $L^2([0,1], \lambda)$ generating a Brownian motion B on $[0,1]$, so that $D\xi$ becomes a process on $[0,1]$ with[3] $\|D\xi\|^2 = \int_0^1 (D_t\xi)^2 dt < \infty$ a.s.

To state the basic existence theorem, recall that a linear operator A between two Banach spaces S, T is said to be *closed*, if its graph is closed in $S \times T$, so that $x_n \to x$ in S and $Ax_n \to y$ in T imply $x \in \operatorname{dom} A$ and $y = Ax$. An operator $A : S \to T$ is *closable* and hence admits a closed extension, if $x_n \to 0$ in S and $Ax_n \to y$ in T imply $y = 0$. Let $L^p(\zeta, H)$ be the space of ζ-measurable random elements η in H with norm $E(\|\eta\|^p)^{1/p} < \infty$, and write $\hat{C}^1(\mathsf{R}^n)$ for the class of differentiable functions on R^n with bounded first derivatives. For convenience, we write $\xi = (\xi_1, \ldots, \xi_n)$.

Theorem 21.1 *(Malliavin derivative) Let ζ be an isonormal Gaussian process on a separable Hilbert space H. Then for any $p \ge 1$ there exists a unique closed, linear operator $D_p : L^p(\zeta) \to L^p(\zeta, H)$ with domain $\mathcal{D}_p$ containing ζh for all $h \in H$, such that for any $\xi_1, \ldots, \xi_n \in \mathcal{D}_p$,*

(i) $D_p(\zeta h) = h, \ \ h \in H,$

(ii) $D_p(f \circ \xi) = \sum_{i\le n} \partial_i f(\xi)\, D_p\xi_i, \quad f \in \hat{C}^1(\mathsf{R}^n).$

Note that for $\xi \in \mathcal{S}$, the *chain rule* in (ii) reduces to (2). In general it is understood that $f(\xi) \in \mathcal{D}_p$. For the proof, we define D on $\mathcal{S}$ by means of (2). The resulting operator from $L^p(\zeta)$ to $L^p(\zeta, H)$ turns out to be closable, and we may define D_p as the associated closure. We need to show that D is well-defined by (2) on the space $\mathcal{S}$.

[3] In this chapter, the inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$ are always taken in the underlying Hilbert space H, so that $\|D\xi\|$ becomes a random variable, not a constant.

Lemma 21.2 *(consistency) For any $\xi \in \mathcal{S}$, the variable $D\xi$ in* (2) *is a.s. independent of the representation of ξ in* (1).

Proof: We need to show that, if also $\xi = g(\zeta\bar{k}) = g(\zeta k_1, \dots, \zeta k_m)$, then

$$\sum_{i\le n} \partial_i f(\zeta\bar{h})\, h_i = \sum_{j\le m} \partial_i\, g(\zeta\bar{k})\, k_j.$$

Here we may choose the k_j to be ortho-normal with a linear span containing all h_i. Assuming $h_i = \sum_j T_{ij} k_j$ for a matrix T, we obtain

$$f \circ T(\zeta\bar{k}) = g(\zeta\bar{k}) \text{ a.s.}$$

Since the variables ζk_i are i.i.d. $N(0,1)$ and the functions $f\circ T$ and g are continuous, it follows that $f \circ T = g$. By the elementary chain rule for differentiation, we get

$$\begin{aligned} \textstyle\sum_j \partial_j\, g(\zeta\bar{k})\, k_j &= \textstyle\sum_j \partial_j (f\circ T)(\zeta\bar{k})\, k_j \\ &= \textstyle\sum_{i,j} \partial_i (f\circ T)(\zeta\bar{k})\, T_{ij} k_j \\ &= \textstyle\sum_i \partial_i f(\zeta\bar{h})\, h_i. \end{aligned} \qquad \Box$$

We further need some elementary identities for the operator D on $\mathcal{S}$.

Lemma 21.3 *(integration by parts) Let ξ, η be smooth functionals of ζ, and let $h \in H$. Then*

(i) $E\langle D\xi, h\rangle = E\xi(\zeta h)$,

(ii) $E\,\xi\eta(\zeta h) = E\,\xi\langle D\eta, h\rangle + E\,\eta\langle D\xi, h\rangle$.

Proof: (i) We may choose ξ as in (1) with ortho-normal $h_1, \dots, h_n \in H$, and take $h = h_1$. Writing φ for the standard normal density on R, and noting that $\varphi'(x) = -x\varphi(x)$, we get by (2) and an elementary integration by parts

$$\begin{aligned} E\langle D\xi, h\rangle &= \int \partial_1 f(x)\, \varphi^{\otimes n}(x)\, dx \\ &= \int f(x)\, \varphi^{\otimes n}(x)\, x_1\, dx \\ &= E\,\xi(\zeta h). \end{aligned}$$

(ii) Here we may represent ξ, η as in (1) in terms of a common set of elements $h_1, \dots, h_n \in H$. Using (2) and the elementary product rule for differentiation, we obtain $D(\xi\eta) = \xi D\eta + \eta D\xi$ a.s., and so by (i)

$$\begin{aligned} E\,\xi\eta(\zeta h) &= E\Big\langle D(\xi\eta), h\Big\rangle \\ &= E\Big\langle \xi D\eta + \eta D\xi, h\Big\rangle \\ &= E\,\xi\langle D\eta, h\rangle + E\,\eta\langle D\xi, h\rangle. \end{aligned} \qquad \Box$$

Next, we show that the operator D in (2) is closable.

Lemma 21.4 *(closure) For any $p \geq 1$, the operator D on $\mathcal{S}$ is closable from $L^p(\zeta)$ to $L^p(\zeta, H)$, and hence extends to a closed operator $(D_p, \mathcal{D}_p)$ between those spaces.*

Proof: Let $\xi_1, \xi_2, \ldots \in \mathcal{S}$ with $\xi_n \to 0$ in L^p, and such that $D\xi_n \to \eta$ in $L^p(\zeta, H)$. To show that $\eta = 0$ a.s., fix any $h \in H$, and let $\beta \in \mathcal{S}$ be such that $\beta(\zeta h)$ is bounded. Using Lemma 21.3 (ii), we get by dominated convergence

$$\begin{aligned} E\,\beta\langle \eta, h\rangle &\leftarrow E\,\beta\langle D\,\xi_n, h\rangle \\ &= E\Big\{\beta\,\xi_n(\zeta h) - \xi_n\langle D\beta, h\rangle\Big\} \to 0, \end{aligned}$$

since $\beta(\zeta h)$ and $\langle D\beta, h\rangle$ are bounded. Hence, $E\,\beta\langle \eta, h\rangle = 0$, and so by approximation $\langle \eta, h\rangle = 0$ a.s. Since h was arbitrary, we obtain $\eta = 0$ a.s. □

This extends D to a closed operator $D_p\colon L^p(\zeta) \to L^p(\zeta, H)$, and we denote the associated domain by $\mathcal{D}_{1,p}$ or simply $\mathcal{D}_p$. Since clearly $D_p = D_q$ on $\mathcal{D}_p \cap \mathcal{D}_q$ for any $p, q \geq 1$, we may henceforth write simply $D_p = D$. It remains to show that the extended version of D satisfies the required chain rule.

Lemma 21.5 *(chain rule) Let $\xi_1, \ldots, \xi_n \in \mathcal{D}_p$ with $p \geq 1$, and let $f \in C^1(\mathsf{R}^n)$ with bounded first derivatives. Then even $f(\xi) \in \mathcal{D}_p$, and*

$$D(f \circ \xi) = \sum_{i \leq n} \partial_i f(\xi)\, D\,\xi_i, \quad f \in \hat{C}^1(\mathsf{R}^n). \tag{3}$$

Proof: First consider some smooth variables $\xi_i = g_i(\zeta \bar{h})$, represented in terms of a common set of elements $h_1, \ldots, h_n \in H$. Then $f(\xi) = (f \circ \bar{g})(\zeta \bar{h})$ is again smooth, where $\bar{g} = (g_1, \ldots, g_m)$, and we get by (2) and the elementary chain rule

$$\begin{aligned} D(f \circ \xi) &= \textstyle\sum_j \partial_j (f \circ \bar{g})(\zeta \bar{h})\, h_j \\ &= \textstyle\sum_{i,j} \partial_i f(\xi)\, \partial_j\, g_i(\xi)\, h_j \\ &= \textstyle\sum_i \partial_i f(\xi)\, D\,\xi_i. \end{aligned}$$

In the general case, we may choose some smooth variables $\xi_i^{(n)} \to \xi_i$ in L^p, such that also $D\xi_i^{(n)} \to D\xi_i$ in $L^p(\zeta, H)$ for all i. We may also choose some smooth functions $f_n \to f$, such that $\partial_i f_n \to \partial_i f$ uniformly on compacts. Note in particular that the f_n have uniform linear growth. By dominated convergence it follows that $f_n(\xi^{(n)}) \to f(\xi)$ in L^p, and further that in $L^p(\zeta, H)$

$$\begin{aligned} D\Big(f_n \circ \xi^{(n)}\Big) &= \textstyle\sum_i \partial_i f_n(\xi^{(n)})\, D\,\xi_i^{(n)} \\ &\to \textstyle\sum_i \partial_i f(\xi)\, D\,\xi_i. \end{aligned}$$

Since D is closed, we get $f(\xi) \in \mathcal{D}_{1,p}$, and the last limit equals $D(f \circ \xi)$, proving (3). □

For any $p \geq 1$, we define a norm on $\mathcal{D}_{1,p}$ by

$$\|\xi\|_{1,p} = \left(E|\xi|^p + E\|D\xi\|^p\right)^{1/p}.$$

In particular, $\mathcal{D}_{1,2}$ is a Hilbert space with inner product

$$\langle \xi, \eta \rangle_{1,2} = E\,\xi\eta + E\langle D\xi, D\eta\rangle.$$

Note that the space $\mathcal{D}_{1,p}$ is reflexive for every $p > 1$, since it is isometrically isomorphic to the reflexive space $L^p(\zeta) \times L^p(\zeta, H)$. This yields a useful closure property:

Lemma 21.6 *(weak closure) Let $\xi_1, \xi_2, \ldots \in \mathcal{D}_{1,2}$ with $\xi_n \to \xi$ in $L^p(\zeta)$ for a $p > 1$. Then these conditions are equivalent:*

(i) *$\xi \in \mathcal{D}_{1,p}$, and $D\xi_n \to D\xi$ weakly in $L^p(\zeta, H)$,*

(ii) $\sup_n E\|D\xi_n\|^p < \infty$.

Proof, (i) $\Rightarrow$ (ii): Note that weak convergence in L^p implies L^p-boundedness, by the Banach-Steinhaus theorem.

(ii) $\Rightarrow$ (i): By (ii) the sequence (ξ_n) is bounded in the reflexive space $\mathcal{D}_{1,p}$. Thus, (ξ_n) is weakly relatively compact in $\mathcal{D}_{1,p}$, and hence converges weakly in $\mathcal{D}_{1,p}$ along a subsequence, say to some $\eta \in \mathcal{D}_{1,p}$. Since also $\xi_n \to \xi$ in $L^p(\zeta)$, we have $\eta = \xi$ a.s., and so $\xi \in \mathcal{D}_{1,p}$, and $\xi_n \to \xi$ weakly in $\mathcal{D}_{1,p}$ along the entire sequence. In particular, $D\xi_n \to D\xi$ weakly in $L^p(\zeta, H)$. $\square$

So far we have regarded the derivative D as an operator from $L^p(\zeta)$ to $L^p(\zeta, H)$. It is straightforward to extend D to an operator from $L^p(\zeta, K)$ to $L^p(\zeta, H \otimes K)$, for a possibly different separable Hilbert space K. Assuming as before that $H = L^2(\mu)$ and $K = L^2(\nu)$ for some σ-finite measures μ, ν on suitable spaces T, T', we may identify $H \otimes K$ with $L^2(\mu \otimes \nu)$. For any random variables $\xi_1, \ldots, \xi_m \in D_{1,p}$ and elements $k_1, \ldots, k_m \in K$, we define

$$D \sum_{i \leq m} \xi_i k_i = \sum_{i \leq m} D\xi_i \otimes k_i.$$

In particular, we get for smooth variables $\xi_i = f_i(\zeta \bar{h})$ with $\bar{h} = (h_1, \ldots, h_n)$

$$D \sum_{i \leq m} f_i(\zeta h)\, k_i = \sum_{i \leq m} \sum_{j \leq n} \partial_j f_i(\zeta \bar{h})\, (h_j \otimes k_i).$$

As before, D extends to a closed operator from $L^p(P \otimes \nu)$ to $L^p(P \otimes \mu \otimes \nu)$.

Iterating the construction, we may define the higher order derivatives D^m on $L^p(\zeta)$ by setting $D^{m+1}\xi = D(D^m \xi)$ with $D^1 = D$. The associated domains $\mathcal{D}_{m,p}$ are defined recursively, and in particular $\mathcal{D}_{m,2}$ is a Hilbert space with inner product

$$\langle \xi, \eta \rangle_{m,2} = E(\xi\,\eta) + \sum_{i=1}^{m} E\left\langle D^i \xi, D^i \eta \right\rangle_{H^{\otimes i}}.$$

For first order derivatives, the chaos decomposition in Theorem 14.26 yields explicit expressions of the L^2-derivative $D\xi$ and its domain $\mathcal{D}_{1,2}$, in terms of multiple Wiener-Itô integrals. For elements in $L^2(\zeta, H)$, we write $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ for the norm and inner product in H, so that the resulting quantities are random variables, not constants. This applies in particular to the norm $\|D\xi\|$.

Theorem 21.7 *(derivative of WI-integral) For an isonormal Gaussian process ζ on $H = L^2(T, \mu)$, let $\xi = \zeta^n f$ with $f \in L^2(\mu^{\otimes n})$ symmetric. Then*

(i) $\xi \in \mathcal{D}_{1,2}$ *with* $D_t\xi = n\,\zeta^{n-1} f(\cdot, t),\ t \in T$,

(ii) $E\|D\xi\|^2 = n\|\xi\|^2 = n(n!)\,\|f\|^2$.

Proof: For any ONB $h_1, h_2, \ldots \in H$, Theorem 14.25 yields an orthogonal expansion of $\zeta^n f$ in terms of products

$$\zeta^n \bigotimes_{j\le m} h_j^{\otimes n_j} = \prod_{j\le m} p_{n_j}(\zeta h_j), \tag{4}$$

where $\sum_{j\le m} n_j = n$. Using the simple chain rule (2), for smooth functions of polynomial growth, together with the elementary fact that $p_n' = n\,p_{n-1}$, we get

$$D\,\zeta^n \bigotimes_{j\le m} h_j^{\otimes n_j} = \sum_{i\le m} n_i\,p_{n_i-1}(\zeta h_i)\,h_i \prod_{j\ne i} p_{n_j}(\zeta h_j),$$

which is equivalent to $D_t\zeta^n f = n\,\zeta^{n-1}\tilde{f}(t)$ for the symmetrization $\tilde{f}$ of the elementary integrand f in (4). This extends by linearity to any linear combination of such functions.

Using Lemma 14.22, along with the orthogonality of the various terms and factors, we obtain

$$\begin{aligned} E\Big\| D\,\zeta^n \bigotimes_{j\le m} h_j^{\otimes n_j} \Big\|^2 &= \sum_{i\le m} n_i^2\,\|p_{n_i-1}(\zeta h_i)\|^2 \prod_{j\ne i} \big\|p_{n_j}(\zeta h_j)\big\|^2 \\ &= \sum_{i\le m} n_i^2 (n_i-1)! \prod_{j\ne i} (n_j)! \\ &= \sum_{i\le m} n_i \prod_{j\le m} (n_j)! \\ &= n \prod_{j\le m} \big\|p_{n_j}(\zeta h_j)\big\|^2 \\ &= n\|\zeta^n f\|^2, \end{aligned}$$

which extends by orthogonality to general linear combinations. This proves (i) and (ii) for such variables ξ, and the general result follows by the closure property of D. □

Corollary 21.8 *(chaos representation of D) For an isonormal Gaussian process ζ on a separable Hilbert space H, let $\xi \in L^2(\zeta)$ with chaos decomposition $\xi = \sum_{n\ge 0} J_n\xi$. Then*

(i) $\xi \in \mathcal{D}_{1,2} \iff \sum_{n>0} n\|J_n\xi\|^2 < \infty$,

in which case

(ii) $D\xi = \sum_{n>0} D(J_n\xi)$ *in* $L^2(\zeta, H)$,

(iii) $E\|D\xi\|^2 = \sum_{n>0} E\left\|D(J_n\xi)\right\|^2 = \sum_{n>0} n\|J_n\xi\|^2.$

Proof: Use Theorems 14.26 and 21.7, along with the closure property of D. □

Corollary 21.9 *(vanishing derivative) For any $\xi \in \mathcal{D}_{1,2}$,*

$$D\xi = 0 \text{ a.s.} \quad \Leftrightarrow \quad \xi = E\xi \text{ a.s.}$$

Proof: If $\xi = E\xi$, then $D\xi = 0$ by (2). Conversely, $D\xi = 0$ yields $J_n\xi = 0$ for all $n > 0$ by Theorem 21.8 (iii), and so $\xi = J_0\xi = E\xi$. □

Corollary 21.10 *(indicator variables) For any $A \in \sigma(\zeta)$,*

$$1_A \in \mathcal{D}_{1,2} \quad \Leftrightarrow \quad PA \in \{0,1\}.$$

Proof: If $PA \in \{0,1\}$, then $\xi = 1_A$ is a.s. a constant, and so $\xi \in \mathcal{S} \subset \mathcal{D}_{1,2}$. Conversely, let $1_A \in \mathcal{D}_{1,2}$. Applying the chain rule to the function $f(x) = x^2$ on $[0,1]$ yields $D1_A = D1_A^2 = 2 1_A D1_A$, and so $D1_A = 0$. Hence, Corollary 21.9 shows that 1_A is a.s. a constant, which implies that either $1_A = 0$ a.s. or $1_A = 1$ a.s., meaning that $PA \in \{0,1\}$. □

Corollary 21.11 *(conditioning) Let $\xi \in \mathcal{D}_{1,2}$, and fix any $A \in \mathcal{T}$. Then $E(\xi \,|\, 1_A\zeta) \in \mathcal{D}_{1,2}$ and*

$$D_t E\Big(\xi \,\Big|\, 1_A\zeta\Big) = E\Big(D_t\xi \,\Big|\, 1_A\zeta\Big)\, 1_A(t) \ \text{ a.e.}$$

Proof: Letting $\xi = \sum_n \zeta^n f_n$ and using Corollary 14.28 and Theorem 21.7 twice, we get

$$\begin{aligned} D_t\, E\Big(\xi \,\Big|\, 1_A\zeta\Big) &= D_t \textstyle\sum_n \zeta^n (f_n\, 1_A^{\otimes n}) \\ &= \textstyle\sum_n n\, \zeta^{n-1}\Big\{ f_n(\cdot,t)\, 1_A^{\otimes n}(\cdot,t)\Big\} \\ &= \textstyle\sum_n n\, \zeta^{n-1}\Big\{ f_n(\cdot,t)\, 1_A^{\otimes(n-1)}\Big\}\, 1_A(t) \\ &= E\Big(D_t\xi \,\Big|\, 1_A\zeta\Big)\, 1_A(t). \end{aligned}$$

In particular, the last series converges, so that $E(\xi \,|\, 1_A\zeta) \in \mathcal{D}_{1,2}$. □

The last result yields immediately the following:

Corollary 21.12 *(null criterion) Let $\xi \in \mathcal{D}_{1,2}$ be $1_A\zeta$-measurable for an $A \in \mathcal{T}$. Then*

$$D_t\xi = 0 \ \text{ a.e. on } A^c \times \Omega.$$

Corollary 21.13 *(adapted processes) Let B be a Brownian motion generated by an isonormal Gaussian process ζ on $L^2(\mathbb{R}_+, \lambda)$, and consider a B-adapted process $X \in \mathcal{D}_{1,2}(H)$. Then $X_s \in \mathcal{D}_{1,2}$ for almost every $s \geq 0$, and the process $(s,t) \mapsto D_t X_s$ has a version such that, for fixed $t \geq 0$,*

(i) $D_t X_s = 0$ *for all* $s \le t$,

(ii) $D_t X_s$ *is* B*-adapted in* $s \ge t$.

Proof: (i) Use Corollary 21.12.

(ii) Apply Theorems 21.7 and 21.8 to the chaos representation in Corollary 14.29. □

Next we introduce the *divergence operator*[4], defined as the adjoint $D^* : L^2(\zeta, H) \to L^2(\zeta)$ of the Malliavin derivative D on $\mathcal{D}_{1,2}$, in the sense of the *duality relation* $E\langle D\xi, Y\rangle = \langle \xi, D^* Y\rangle$, valid for all $\xi \in \mathcal{D}_{1,2}$ and suitable $Y \in L^2(\zeta, H)$, where the inner products are taken in $L^2(\zeta, H)$ on the left and in $L^2(\zeta)$ on the right. More precisely, the operator D^* and its domain $\mathcal{D}^*$ are given by:

Lemma 21.14 *(existence of D^*) For an isonormal Gaussian process ζ on H and element $Y \in L^2(\zeta, H)$, these conditions are equivalent:*

(i) *there exists an a.s. unique element* $D^*Y \in L^2(\zeta)$ *with*

$$E\langle D\xi, Y\rangle = E(\xi D^* Y), \quad \xi \in \mathcal{D}_{1,2},$$

(ii) *there exists a constant* $c > 0$ *such that*

$$\Big| E\langle D\xi, Y\rangle \Big|^2 \le c\, E|\xi|^2, \quad \xi \in \mathcal{D}_{1,2}.$$

Proof: By (i) and Cauchy's inequality, we get (ii) with $c^2 = E|D^*Y|^2$. Now assume (ii). Then the map $\varphi(\xi) = E\langle D\xi, Y\rangle$ is a bounded linear functional on $\mathcal{D}_{1,2}$, and since the latter is a dense subset of $L^2(\zeta)$, φ extends by continuity to all of $L^2(\zeta)$. Hence, the Riesz–Fischer theorem yields $\varphi(\xi) = E(\xi\gamma)$ for an a.s. unique $\gamma \in L^2(\zeta)$, and (i) follows with $D^*Y = \gamma$. □

For suitable elements $X \in L^2(\zeta, H)$, with chaos expansions as in Corollary 14.27 in terms of functions f_n on T^{n+1}, symmetric in the first n arguments, we may now express the divergence D^*X in terms of the full symmetrizations $\tilde{f}_n$ on T^{n+1}. Since $f_n(s,t)$ is already symmetric in $s \in T^n$, we note that for all $n \ge 0$,

$$(n+1)\,\tilde{f}_n(s,t) = f_n(s,t) + \sum_{i\le n} f_n\Big(s_1, \dots, s_{i-1}, t, s_{i+1}, \dots, s_n, s_i\Big).$$

Theorem 21.15 *(chaos representation of D^*) For an isonormal Gaussian process ζ on $H = L^2(\mu)$, let $X \in L^2(\zeta, H)$ with $X_t = \sum_{n\ge 0} \zeta^n f_n(\cdot, t)$, where the $f_n \in L^2(\mu^{n+1})$ are symmetric in the first n arguments. Writing $\tilde{f}_n$ for the full symmetrizations of f_n, we have*

(i) $X \in \mathcal{D}^* \quad \Leftrightarrow \quad \sum_{n\ge 0} (n+1)!\, \|\tilde{f}_n\|^2 < \infty,$

in which case

[4] also called the *Skorohod integral* and often denoted by δ or $\int$

(ii) $D^*X = \sum_{n\geq 0} D^*(J_nX) = \sum_{n\geq 0} \zeta^{n+1}\tilde{f}_n$ *in* $L^2(\zeta)$,

(iii) $\|D^*X\|^2 = \sum_{n\geq 0} \|D^*(J_nX)\|^2 = \sum_{n\geq 0} (n+1)!\, \|\tilde{f}_n\|^2$.

Proof: For an $n \geq 1$, let $\eta = \zeta^n g$ with symmetric $g \in L^2(\mu^n)$. Using Theorem 21.7 and the extension of Lemma 14.22, we get

$$\begin{aligned}
E\langle X, D\eta\rangle &= E\int X_t (D\eta)_t\, \mu(dt)\\
&= E\int X_t\, n\, \zeta^{n-1} g(\cdot,t)\, \mu(dt)\\
&= \textstyle\sum_m \int nE\Big\{\zeta^m f_m(\cdot,t)\, \zeta^{n-1} g(\cdot,t)\Big\}\, \mu(dt)\\
&= n\int E\Big\{\zeta^{n-1} f_{n-1}(\cdot,t)\, \zeta^{n-1} g(\cdot,t)\Big\}\, \mu(dt)\\
&= n!\int \Big\langle f_{n-1}(\cdot,t),\, g(\cdot,t)\Big\rangle\, \mu(dt)\\
&= n!\, \langle f_{n-1}, g\rangle = n!\, \langle \tilde{f}_{n-1}, g\rangle\\
&= E\Big(\zeta^n \tilde{f}_{n-1}\, \zeta^n g\Big) = E\,\eta(\zeta^n \tilde{f}_{n-1}).
\end{aligned}$$

Now let $X \in \mathcal{D}^*$. Comparing the above computation with Theorem 21.14 (i), we get

$$E\,\eta\Big(\zeta^n \tilde{f}_{n-1}\Big) = E\,\eta(D^*X), \quad \eta \in \mathcal{H}_n,\ n \geq 1,$$

and so

$$J_n(D^*X) = \zeta^n \tilde{f}_{n-1}, \quad n \geq 1,$$

which remains true for $n = 0$, since $E\,D^*X = E\langle D1, X\rangle = 0$ by Theorem 21.14 (i) and Corollary 21.9. Hence,

$$\begin{aligned}
D^*X &= \sum_{n>0} J_n(D^*X)\\
&= \sum_{n>0} \zeta^n \tilde{f}_{n-1} = \sum_{n\geq 0} \zeta^{n+1}\tilde{f}_n,\\
\|D^*X\|^2 &= \sum_{n>0} \Big\| J_n(D^*X)\Big\|^2\\
&= \sum_{n>0} \Big\| \zeta^n \tilde{f}_{n-1}\Big\|^2 = \sum_{n>0} n!\, \|\tilde{f}_{n-1}\|^2,
\end{aligned}$$

which proves (ii)−(iii), along with the convergence of the series in (i).

Conversely, suppose that the series in (i) converges, so that $\sum_n \zeta^n \tilde{f}_{n-1}$ converges in L^2. By the previous calculation, we get for any $\eta \in L^2(\zeta)$

$$\begin{aligned}
E\Big\langle X, D(J_n\eta)\Big\rangle &= E(J_n\eta)\Big(\zeta^n \tilde{f}_{n-1}\Big)\\
&= E\,\eta\Big(\zeta^n \tilde{f}_{n-1}\Big).
\end{aligned}$$

If $\eta \in \mathcal{D}_{1,2}$, we have $\sum_n D(J_n\eta) = D\eta$ in $L^2(\zeta, H)$ by Theorem 21.8, and so

$$E\langle X, D\eta\rangle = E\,\eta \sum_{n>0} \zeta^n \tilde{f}_{n-1}.$$

Hence, by Cauchy's inequality,

$$\Big| E\langle X, D\eta\rangle\Big| \leq \|\eta\|\, \Big\| \sum_{n>0} \zeta^n \tilde{f}_{n-1}\Big\| \lesssim \|\eta\|, \quad \eta \in \mathcal{D}_{1,2},$$

and $X \in \mathcal{D}^*$ follows by Theorem 21.14. □

Many results for D^* require a slightly stronger condition. Let $\hat{\mathcal{D}}_{1,2}$ be the class of processes $X \in L^2(P \otimes \mu)$, such that $X_t \in \mathcal{D}_{1,2}$ for almost every t, and the process $D_s X_t$ on T^2 has a product-measurable version in $L^2(\mu^2)$. Note that $\hat{\mathcal{D}}_{1,2}$ is a Hilbert space with norm

$$\|X\|_{1,2}^2 = \|X\|_{P\otimes\mu}^2 + \|DX\|_{P\otimes\mu^2}^2, \quad X \in \hat{\mathcal{D}}_{1,2},$$

and hence is isomorphic to $\mathcal{D}_{1,2}(H)$.

Corollary 21.16 *(sub-domain of D^*) For an isonormal Gaussian process ζ on $L^2(\mu)$, the class $\hat{\mathcal{D}}_{1,2}$ is a proper subset of $\mathcal{D}^*$.*

Partial proof: For the moment, we prove only the stated inclusion. Then let $X \in L^2(T,\mu)$ with a representation $X_t = \sum_n \zeta^n f_n(\cdot, t)$, as in Lemma 14.27, where the functions $f_n \in L^2(\mu^{n+1})$ are symmetric in the first n arguments. If $X \in \hat{\mathcal{D}}_{1,2}$, then $DX \in L^2$, and we get by Lemma 14.22, along with the vector-valued versions of Theorem 21.7 and Corollary 21.8,

$$\begin{aligned} E\|DX\|^2 &= \textstyle\sum_n E\Big\|D(J_n X)\Big\|^2 \\ &= \textstyle\sum_n n(n!)\,\|f_n\|^2 \\ &\gtrsim \textstyle\sum_n (n+1)!\,\|f_n\|^2 \\ &\geq \textstyle\sum_n (n+1)!\,\|\tilde{f}_n\|^2, \end{aligned}$$

with summations over $n \geq 1$. Hence, the series on the right converges, and so $X \in \mathcal{D}^*$ by Theorem 21.15. □

Now let $\mathcal{S}_H$ be the class of *smooth* random elements in $L^2(\zeta, H)$ of the form $\eta = \sum_{i\leq n} \xi_i h_i$, where $\xi_1, \ldots, \xi_n \in \mathcal{S}$ and $h_1, \ldots, h_n \in H$.

Lemma 21.17 *(divergence of smooth elements) The class $\mathcal{S}_H$ is contained in $\mathcal{D}^*$, and for any $\beta = \sum_i \xi_i h_i$ in $\mathcal{S}_H$,*

$$D^*\beta = \textstyle\sum_i \xi_i(\zeta h_i) - \sum_i \langle D\xi_i, h_i\rangle.$$

Proof: By linearity, we may take $\beta = \xi\, h$ for some $\xi \in \mathcal{S}$ and $h \in H$, which reduces the asserted formula to

$$D^*(\xi h) = \xi(\zeta h) - \langle D\xi, h\rangle.$$

By the definition of D^*, we need to check that

$$E\,\xi\langle D\eta, h\rangle = E\,\eta\Big\{\xi(\zeta h) - \langle D\xi, h\rangle\Big\}, \quad \eta \in \mathcal{D}_{1,2}.$$

This holds by Lemma 21.3 (ii) for any $\eta \in \mathcal{S}$, and the general result follows by approximation. □

The next result allows us to factor out scalar variables.

Lemma 21.18 *(scalar factors) Let $\xi \in \mathcal{D}_{1,2}$ and $Y \in \mathcal{D}^*$ with*

(i) $\xi Y \in L^2(\zeta, H)$,

(ii) $\xi D^* Y - \langle D\xi, Y\rangle \in L^2(\zeta)$.

Then $\xi Y \in \mathcal{D}^*$, and

$$D^*(\xi Y) = \xi D^* Y - \langle D\xi, Y\rangle.$$

Proof: Using the elementary product rule for differentiation and the D/D^*-duality, we get for any $\xi, \eta \in \mathcal{S}$

$$\begin{aligned} E\langle D\eta,\, \xi Y\rangle &= E\langle Y,\, \xi D\eta\rangle \\ &= E\Big\langle Y,\, D(\xi\eta) - \eta\, D\xi\Big\rangle \\ &= E\Big\{(D^*Y)\,\xi\eta - \eta\,\langle Y, D\xi\rangle\Big\} \\ &= \Big\langle \eta,\, \xi D^*Y - \langle D\xi, Y\rangle\Big\rangle, \end{aligned}$$

which extends to general $\xi, \eta \in \mathcal{D}_{1,2}$ as stated, by (i)−(ii) and the closure of D. Hence, Theorem 21.14 yields $\xi Y \in \mathcal{D}^*$, and the asserted formula follows. □

This yields a simple factorization property:

Corollary 21.19 *(factorization) Let $\xi \in L^2(\zeta)$ be $1_{A^c}\zeta$-measurable for an $A \in \mathcal{T}$. Then $1_A\xi \in \mathcal{D}^*$, and*

$$D^*(1_A\xi) = \xi(\zeta A) \ \textit{a.s.}$$

Proof: First let $\xi \in \mathcal{D}_{1,2}$. Using Lemmas 21.17 and 21.18, along with Corollary 21.12, we get

$$\begin{aligned} D^*(1_A\xi) &= \xi\, D^*(1_A) - \langle D\xi, 1_A\rangle \\ &= \xi(\zeta A) - 0 = \xi(\zeta A). \end{aligned}$$

The general result follows since $\mathcal{D}_{1,2}$ is dense in $L^2(\zeta)$ and D^* is closed, being the adjoint of a closed operator. □

End of proof of Corollary 21.16: It remains to show that the stated inclusion is strict. Then let $A, B \in \mathcal{T}$ be disjoint with $\mu A \wedge \mu B > 0$, and define $Y = 1_A\xi$ with $\xi = 1\{\zeta B > 0\}$, so that $Y \in \mathcal{D}^*$ by Corollary 21.19. On the other hand, Corollary 21.10 yields $\xi \notin \mathcal{D}_{1,2}$ since $P\{\zeta B > 0\} = \frac{1}{2}$. Noting that $1_A\xi \in \hat{\mathcal{D}}_{1,2}$ iff $\xi \in \mathcal{D}_{1,2}$ when $\mu A > 0$, we conclude that even $Y \notin \hat{\mathcal{D}}_{1,2}$. □

We turn to a key relationship between D and D^*.

Theorem 21.20 *(commutation) Let $X \in \hat{\mathcal{D}}_{1,2}$ be such that, for $t \in T$ a.e. μ,*

(i) *the process $s \mapsto D_t X_s$ belongs to $\mathcal{D}^*$,*

(ii) *the process $t \mapsto D^*(D_t X_s)$ has a version in $L^2(P \otimes \mu)$.*

Then $D^*X \in \mathcal{D}_{1,2}$, and

$$D_t(D^*X) = X_t + D^*(D_tX), \quad t \in T.$$

Proof: By Corollary 14.27 we may write $X_t = \sum_n \zeta^n f_n(\cdot, t)$, where each f_n is symmetric in the first n arguments. Since $\hat{\mathcal{D}}_{1,2} \subset \mathcal{D}^*$ by Corollary 21.16, Theorem 21.15 yields

$$D^*X = \sum\nolimits_n \zeta^{n+1} \tilde{f}_n, \qquad \sum\nolimits_n (n+1)!\, \|\tilde{f}_n\|^2 < \infty.$$

In particular,

$$J_n(D^*X) = \zeta^n \tilde{f}_{n-1},$$
$$\Big\| J_n(D^*X) \Big\|^2 = n!\, \|\tilde{f}_{n-1}\|^2,$$

and so $\sum_n n\, \|J_n(D^X)\|^2 < \infty$, which implies $D^*X \in \mathcal{D}_{1,2}$ by Theorem 21.8. Moreover, Theorem 21.7 yields

$$\begin{aligned} D_t(D^*X) &= D_t \sum\nolimits_n \zeta^{n+1} \tilde{f}_n \\ &= \sum\nolimits_n (n+1)!\, \zeta^n \tilde{f}_n(\cdot, t) \\ &= \sum\nolimits_n \zeta^n f_n(\cdot, t) + \sum\nolimits_{n,i} \zeta^n f_n\Big(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_n, t_i\Big) \\ &= X_t + \sum_{n \ge 0} \zeta^n g_n(\cdot, t), \end{aligned}$$

where

$$g_n\Big(t_1, \ldots, t_n, t\Big) = \sum_{i \le n} f_n\Big(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_n, t_i\Big).$$

Now fix any $s, t \in T$, and conclude from Theorem 21.7 that $D_tX_s = \sum_n \zeta^{n-1} f_n(\cdot, t, s)$. Then Theorem 21.15 yields

$$D^*(D_tX_s) = \sum_{n \ge 0} \zeta^n g'_n(\cdot, t), \quad t \in T, \tag{5}$$

where

$$\begin{aligned} g'_n\Big(t_1, \ldots, t_n, t\Big) &= \sum_{i \le n} f_n\Big(t_1, \ldots, t_{i-1}, t_n, t_{i+1}, \ldots, t_{n-1}, t, t_i\Big) \\ &= \sum_{i \le n} f_n\Big(t_1, \ldots, t_{i-1}, t, t_{i+1}, \ldots, t_{n-1}, t_n, t_i\Big) \\ &= g_n(\cdot, t), \end{aligned}$$

since f_n is symmetric in the first n variables. Hence, (5) holds with g'_n replaced by g_n, and the asserted relation follows. □

To state the next result, it is suggestive to write for any $X, Y \in \hat{\mathcal{D}}_{1,2}$

$$\mathrm{tr}(DX \cdot DY) = \iint (D_sX_t)(D_tY_s)\, \mu^{\otimes 2}(ds\, dt).$$

Corollary 21.21 *(covariance) For any $X, Y \in \hat{\mathcal{D}}_{1,2}$, we have*[5]

$$\Big\langle D^*X, D^*Y \Big\rangle = E\langle X, Y\rangle + E\, \mathrm{tr}(DX \cdot DY).$$

[5]Throughout this chapter, the inner product $\langle \cdot, \cdot \rangle$ must not be confused with the predictable covariation in Chapter 20.

Proof: By a simple approximation, we may assume that X, Y have finite chaos decompositions. Then $D^*X, D^*Y \in \mathcal{D}_{1,2}$, and we see from the D/D^*-duality, used twice, and Theorem 21.20 that

$$\begin{aligned}\left\langle D^*X, D^*Y \right\rangle &= E\left\langle X, D(D^*Y) \right\rangle \\ &= E\langle X, Y\rangle + E\left\langle X, D^*(DY) \right\rangle \\ &= E\langle X, Y\rangle + E\,\mathrm{tr}(DX \cdot DY),\end{aligned}$$

where the last equality is clear if we write out the duality explicitly in terms of μ-integrals. □

An operator T on a space of random variables ξ is said to be *local* if $\xi = 0$ a.s. on a set $A \in \mathcal{F}$ implies $T\xi = 0$ a.e. on A. We show that D^* is local on $\hat{\mathcal{D}}_{1,2}$ and D is local on $\mathcal{D}_{1,1}$. To make this more precise, we note that for any statement on $\Omega \times T$, the qualification 'a.e.' is understood to be with respect to $P \otimes \mu$.

Theorem 21.22 *(local properties) For any set $A \subset \Omega$ in $\mathcal{F}$, we have*

(i) $X \in \hat{\mathcal{D}}_{1,2},\ 1_A X = 0$ *a.e.* $\Rightarrow\ 1_A(D^*X) = 0$ *a.s.,*

(ii) $\xi \in \mathcal{D}_{1,1},\ 1_A \xi = 0$ *a.s.* $\Rightarrow\ 1_A(D\xi) = 0$ *a.e.*

Proof: (i) Choose a smooth function $f : \mathsf{R} \to [0,1]$ with $f(0) = 1$ and support in $[-1,1]$, and define $f_n(x) = f(nx)$. Let $\eta \in \mathcal{S}_c$ be arbitrary. We can show that $\eta\, f_n(\|X\|^2) \in \mathcal{D}_{1,2}$, and use the product and chain rules to get

$$D\left\{\eta\, f_n(\|X\|^2)\right\} = f_n(\|X\|^2)\, D\eta + 2\,\eta\, f_n'(\|X\|^2)\, DX \cdot X.$$

Hence, the D/D^*-duality yields

$$\begin{aligned}E\left\{\eta\,(D^*X)\, f_n(\|X\|^2)\right\} &= E\left\{f_n(\|X\|^2)\,\langle X, D\eta\rangle\right\} \\ &\quad + 2\,E\left\{\eta\, f_n'(\|X\|^2)\left\langle DX \cdot X,\, X\right\rangle\right\}. \qquad (6)\end{aligned}$$

As $n \to \infty$, we note that

$$\begin{aligned}\eta\,(D^*X)\, f_n(\|X\|^2) &\to \eta\,(D^*X)\, 1\{X = 0\}, \\ f_n(\|X\|^2)\,\langle X, D\eta\rangle &\to 0, \\ \eta\, f_n'(\|X\|^2)\left\langle DX \cdot X,\, X\right\rangle &\to 0,\end{aligned}$$

and (6) yields formally

$$E\left(\eta\, D^*X;\ X = 0\right) = 0. \qquad (7)$$

Since $\mathcal{S}_c$ is dense in $L^2(\zeta)$, this implies $D^*X = 0$ a.s. on $A \subset \{X = 0\}$, as required.

To justify (7), we note that

$$\left| f_n(\|X\|^2)\,\langle X, D\eta\rangle \right| \le \|f\|_\infty\, \|X\|_H\, \|D\eta\|_H,$$

and
$$\Big|\eta\, f_n'(\|X\|^2)\,\Big\langle DX\cdot X,\,X\Big\rangle\Big| \le |\eta|\,\Big|f_n'(\|X\|^2)\Big|\;\|DX\|_{H^2}\,\|X\|^2$$
$$\le 2\,|\eta|\,\|f'\|_\infty\,\|DX\|_{H^2},$$

where both bounds are integrable. Hence, (7) follows from (6) by dominated convergence.

(ii) It is enough to prove the statement for bounded variables $\xi\in\mathcal{D}_{1,1}$, since it will then follow for general ξ by the chain rule for D, applied to the variable $\arctan\xi$. For f and f_n as in (i), define $g_n(t)=\int_{-\infty}^t f_n(r)\,dr$, and note that $\|g_n\|_\infty\le n^{-1}\|f\|_1$, and also $Dg_n(\xi)=f_n(\xi)D\xi$ by the chain rule.

Now let $Y\in\mathcal{S}_b(H)$ be arbitrary, and note that the D/D^*-duality remains valid in the form $E(\eta\,D^*Y)=E\langle D\eta,Y\rangle$, for any bounded variables $\eta\in\mathcal{D}_{1,1}$. Using the chain rule for D, the extended D/D^*-duality, and some easy estimates, we obtain

$$\begin{aligned}\Big|E\Big\{f_n(\xi)\,\langle D\xi,Y\rangle\Big\}\Big| &= \Big|E\Big\langle D(g_n\circ\xi),Y\Big\rangle\Big|\\ &= \Big|E\Big\{g_n(\xi)\,D^*Y\Big\}\Big|\\ &\le n^{-1}\|f\|_1E|D^*Y|\to 0,\end{aligned}$$

and so by dominated convergence

$$E\Big(\langle D\xi,Y\rangle;\,\xi=0\Big)=0,$$

which extends by suitable approximation to $D\xi=0$ a.e. on $A\subset\{\xi=0\}$. □

Since the operators D and D^* are local, we may extend their domains as follows. Say that a random variable ξ lies *locally* in $\mathcal{D}_{1,p}$ and write $\xi\in\mathcal{D}_{\rm loc}^{1,p}$, if there exist some measurable sets $A_n\uparrow\Omega$ with associated random variables $\xi_n\in\mathcal{D}_{1,p}$, such that $\xi=\xi_n$ on A_n. The definitions of $\mathcal{D}_{\rm loc}^{m,p}$ and $\hat{\mathcal{D}}_{\rm loc}^{1,2}$ are similar.

To motivate our introduction of the third basic operator of the subject, let X be a *standard Ornstein–Uhlenbeck process* on H, defined as a centered Gaussian process on $\mathsf{R}\times H$ with covariance function

$$E(X_sh)(X_tk)=e^{-|t-s|}\,\langle h,k\rangle,\quad h,k\in H,\;\; s,t\in\mathsf{R}.$$

This makes X a stationary Markov process on H, such that X_t is isonormal Gaussian on H for every $t\in\mathsf{R}$, whereas X_th is a real OU-process in t for each $h\in H$ with $\|h\|=1$. For the construction of X, fix any ONB $h_1,h_2,\ldots\in H$, let the X_th_j be independent OU-processes in R, and extend to H by linearity and continuity[6].

For the present purposes, we may put $X_0=\zeta$, and define the associated transition operators T_t directly on $L^2(\zeta)$ by

$$T_t(f\circ\zeta)=E\Big(f\circ X_t\,\Big|\,\zeta\Big),\quad t\ge 0,$$

[6]Here the possible path continuity of X plays no role.

for measurable functions f on R^H with $f(\zeta) \in L^2$. In fact, any $\xi \in L^2(\zeta)$ has such a representation by Lemma 1.14 or Theorem 14.26, and since $X_t \overset{d}{=} \zeta$, the right-hand side is a.s. independent of the choice of f.

To identify the operators T_t, we note that the Hermite polynomials p_n, normalized as in Theorem 14.25, satisfy the identity

$$\exp\Big(xt - \tfrac{1}{2}\,t^2\Big) = \sum_{n\geq 0} p_n(x)\,\frac{t^n}{n!}, \quad x, t \in \mathsf{R}. \tag{8}$$

From the moving average representation in Chapter 14, we further note that for any $t \geq 0$,

$$X_t = e^{-t}\zeta + \sqrt{1 - e^{-2t}}\,\zeta_t, \tag{9}$$

where $\zeta \overset{d}{=} \zeta_t \perp\!\!\!\perp \zeta$.

Theorem 21.23 *(Ornstein–Uhlenbeck semi-group) For a standard OU-process X on H with $X_0 = \zeta$, the transition operators T_t on $L^2(\zeta)$ are given by*

$$T_t\xi = \sum_{n\geq 0} e^{-nt} J_n\xi, \quad \xi \in L^2(\zeta),\ t \geq 0.$$

Proof: By Proposition 4.2 and Corollary 6.5, it is enough to consider random variables of the form $\xi = f(\zeta h)$ with $\|h\| = 1$. By the uniqueness theorem for moment-generating functions, we may take $f(x) = \exp(rx - \frac{1}{2}\,r^2)$ for an arbitrary $r \in \mathsf{R}$. Using the representation (9) with $a = e^{-t}$ and $b = \sqrt{1 - e^{-2t}}$, we obtain

$$\begin{aligned} f(X_t) &= \exp\Big(rX_th - \tfrac{1}{2}\,r^2\Big) \\ &= \exp\Big(ra\,\zeta h + rb\,\zeta_t - \tfrac{1}{2}\,r^2\Big) \\ &= \exp\Big(ra\,\zeta h - \tfrac{1}{2}\,r^2a^2\Big)\exp\Big(rb\,\zeta_t h - \tfrac{1}{2}\,r^2b^2\Big). \end{aligned}$$

In particular, we get for $t = 0$ by (8) and Theorem 14.25

$$\begin{aligned} \xi = f(\zeta) &= \exp\Big(r\,\zeta h - \tfrac{1}{2}\,r^2\Big) \\ &= \sum_{n\geq 0} p_n(\zeta h)\,\frac{r^n}{n!} \\ &= \sum_{n\geq 0} \zeta^n h^{\otimes h}\,\frac{r^n}{n!}, \end{aligned}$$

which shows that $J_n\xi = \zeta^n h^{\otimes n}\, r^n/n!$. For general t, we note that $\zeta_t h$ is $N(0,1)$ with $\zeta_t \perp\!\!\!\perp \zeta$, and conclude by the same calculation that

$$\begin{aligned} E\Big\{f(X_t)\,\Big|\,\zeta\Big\} &= \exp\Big(ra\,\zeta h - \tfrac{1}{2}\,r^2a^2\Big) \\ &= \sum_{n\geq 0} \zeta^n h^{\otimes h}\,\frac{(ra)^n}{n!} \\ &= \sum_{n\geq 0} e^{-nt} J_n\xi. \end{aligned}$$

□

The T_t clearly form a contraction semi-group on $L^2(\zeta)$, and we proceed to determine the associated generator, in the sense of convergence in L^2.

Theorem 21.24 *(Ornstein–Uhlenbeck generator) For a standard OU-process X on H with $X_0 = \zeta$, the generator L and its domain $\mathcal{L} \subset L^2(\zeta)$ are given by*

(i) $\xi \in \mathcal{L} \;\Leftrightarrow\; \sum_{n>0} n^2 \|J_n \xi\|^2 < \infty$,

(ii) $L\xi = \sum_{n>0} L(J_n \xi) = -\sum_{n>0} n\, J_n \xi$ *in* L^2,

(iii) $\|L\xi\|^2 = \sum_{n>0} \|L(J_n\xi)\|^2 = \sum_{n>0} n^2 \|J_n \xi\|^2$.

Proof: Let an operator L with domain $\mathcal{L}$ be given by (i)−(ii), and let G denote the generator of (T_t) with domain $\mathcal{G}$. First let $\xi \in \mathcal{L}$. Since J_n and T_t commute on $L^2(\zeta)$ and $n \ge t^{-1}(1 - e^{-nt}) \to n$ as $t \to 0$ for fixed n, we get by dominated convergence as $t \to 0$

$$\begin{aligned} \left\| t^{-1}\Big(T_t \xi - \xi\Big) - L\xi \right\|^2 &= \textstyle\sum_n \left\| n\, J_n \xi - t^{-1}\Big(1 - e^{-nt}\Big)\, J_n \xi \right\|^2 \\ &= \textstyle\sum_n \left| n - t^{-1}\Big(1 - e^{-nt}\Big) \right|^2 \|J_n\|^2 \to 0, \end{aligned}$$

and so $\xi \in \mathcal{G}$ with $G\xi = L\xi$, proving that $G = L$ on $\mathcal{L} \subset \mathcal{G}$.

To show that $\mathcal{G} \subset \mathcal{L}$, let $\xi \in \mathcal{G}$ be arbitrary. Since the J_n act continuously on $L^2(\zeta)$, we get as before

$$J_n(G\xi) \leftarrow t^{-1}\Big\{T_t(J_n\xi) - J_n\xi\Big\} \to -n\, J_n \xi,$$

and so

$$\begin{aligned} \textstyle\sum_n n^2 \|J_n \xi\|^2 &= \textstyle\sum_n \Big\| J_n(G\xi) \Big\|^2 \\ &= \|G\xi\|^2 < \infty, \end{aligned}$$

which shows that indeed $\xi \in \mathcal{L}$. □

We may now establish a remarkable relationship between the three basic operators of Malliavin calculus:

Theorem 21.25 *(operator relation) The operators D, D^*, L are related by*

(i) $\xi \in \mathcal{L} \;\Leftrightarrow\; \xi \in \mathcal{D}_{1,2},\ D\xi \in \mathcal{D}^*$,

(ii) $L = -D^*D$ *on* $\mathcal{L}$.

Proof: First let $\xi \in \mathcal{D}_{1,2}$ and $D\xi \in \mathcal{D}^*$. Then for any $\eta \in \mathcal{H}_n$, we get by the D/D^*-duality and Theorem 21.7

$$\begin{aligned} E\,\eta(D^*D\,\xi) &= E\langle D\,\eta, D\,\xi\rangle \\ &= n\,E\,\eta(J_n\xi), \end{aligned}$$

and so $J_n(D^*D\,\xi) = n\,J_n\xi$, which shows that $\xi \in \mathcal{L}$ with $D^*D\xi = -L\xi$.

It remains to show that $\mathcal{L} \subset \mathrm{dom}(D^*D)$, so let $\xi \in \mathcal{L}$ be arbitrary. Then Theorem 21.24 yields

$$\textstyle\sum_n n\,\|J_n\xi\|^2 \le \sum_n n^2 \|J_n \xi\|^2 < \infty,$$

and so $\xi \in \mathcal{D}_{1,2}$ by Theorem 21.8. Next, Theorems 21.8 and 21.24 yield for any $\eta \in \mathcal{D}_{1,2}$

$$E\langle D\eta, D\xi\rangle = \textstyle\sum_n n\, E(J_n\eta)(J_n\xi)$$
$$= -E\,\eta(L\xi),$$

and so $D\xi \in \mathcal{D}^*$ by Theorem 21.14. Thus, we have indeed $\xi \in \mathrm{dom}(D^*D)$. □

We proceed to show that L behaves like a second order differential operator, when acting on smooth random variables. For comparison, recall from Chapter 32 that an OU-process in R^d solves the SDE $dX_t = \sqrt{2}\,dB_t - X_t\,dt$, and hence that its generator A agrees formally with the expression for L below. See also Theorem 17.24 for an equivalent formula in the context of general Feller diffusions.

Theorem 21.26 *(L as elliptic operator) For smooth random variables $\xi = f(\zeta\bar{h})$ with $\bar{h} = (h_1, \ldots, h_n) \in H^n$, we have*

$$L\xi = \sum_{i,j\le n} \partial^2_{ij} f(\zeta\bar{h})\,\langle h_i, h_j\rangle - \sum_{i\le n} \partial_i f(\zeta\bar{h})\,\zeta h_i.$$

Proof: Since $\xi \in \mathcal{D}_{1,2}$, we have by (2)

$$D\xi = \textstyle\sum_n \partial_i f(\zeta\bar{h})\,h_i,$$

so that in particular $D\xi \in \mathcal{S}_H \subset \mathcal{D}^*$. Using (2) again, along with Lemma 21.17, we obtain

$$D^*(D\xi) = \sum_{i\le n} \partial_i f(\zeta\bar{h})\,\zeta h_i - \sum_{i,j\le n} \partial^2_{ij} f(\zeta\bar{h})\,\langle h_i, h_j\rangle,$$

and the assertion follows by Theorem 21.25. □

We may extend the last result to a chain rule for the OU-operator L, which may be regarded as a second order counterpart of the first order chain rule (3) for D.

Corollary 21.27 *(second order chain rule) Let $\xi_1, \ldots, \xi_n \in \mathcal{D}_{2,4}$, and let $f \in C^2(\mathsf{R}^n)$ with bounded first and second order partial derivatives. Then $f(\xi) \in \mathcal{L}$ with $\xi = (\xi_1, \ldots, \xi_n)$, and we have*

$$L(f\circ\xi) = \sum_{i,j\le n} \partial^2_{ij} f(\xi)\,\Big\langle D\xi_i, D\xi_j\Big\rangle + \sum_{i\le n} \partial_i f(\xi)\,L\xi_i.$$

Proof: Approximate ξ in the norm $\|\cdot\|_{2,4}$ by smooth random variables, and f by smooth functions in $\hat{C}(\mathsf{R}^n)$. The result then follows by the continuity of L in the norm $\|\cdot\|_{2,2}$. We omit the details. □

We now come to the remarkable fact that the divergence operator D^*, here introduced as the adjoint of the differentiation operator D, is a non-anticipating extension of the Itô integral of stochastic calculus. This is why it is often regarded as an integral and may even be denoted by an integral sign.

Here we take ζ to be an isonormal Gaussian process on $H = L^2(\mathsf{R}_+, \lambda)$, generating a Brownian motion B with $B_t = \zeta[0,t]$ a.s. for all $t \geq 0$. Letting $\mathcal{F}$ be the filtration induced by B, we write $L^2_{\mathcal{F}}(P \otimes \lambda)$ for the class of $\mathcal{F}$-progressive processes X on R_+ with $E \int X_t^2\, dt < \infty$. For such a process X, we know from Chapter 18 that the Itô integral $\int X_t\, dB_t$ exists.

Theorem 21.28 *(D^* as extended Itô integral) Let ζ be an isonormal process on $L^2(\mathsf{R}_+, \lambda)$, generating a Brownian motion B with induced filtration $\mathcal{F}$. Then*

(i) $L^2_{\mathcal{F}}(P \otimes \lambda) \subset \mathcal{D}^*$,

(ii) $D^* X = \int_0^\infty X_t\, dB_t$ *a.s.*, $\quad X \in L^2_{\mathcal{F}}(P \otimes \lambda)$.

Proof: First let X be a predictable step process of the form

$$X_t = \sum_{i \leq m} \xi_i 1_{(t_i, t_{i+1}]}(t), \quad t \geq 0,$$

where $0 \leq t_1 < \cdots < t_{m+1}$ and each ξ_i is an $\mathcal{F}_{t_i}$-measurable L^2-random variable. By Corollary 21.19, we have $X \in \mathcal{D}^*$ with

$$\begin{aligned} \int_0^\infty X_t\, dB_t &= \sum_{i \leq m} \xi_i \Big(B_{t_{i+1}} - B_{t_i} \Big) \\ &= \sum_{i \leq m} D^* \Big(\xi_i 1_{(t_i, t_{i+1}]} \Big) = D^* X. \end{aligned} \tag{10}$$

For general $X \in L^2_{\mathcal{F}}(P \otimes \lambda)$, Lemma 18.23 yields some predictable step processes X^n as above, such that $X^n \to X$ in $L^2(P \otimes \lambda)$. Applying (10) to the X^n and using Theorem 18.11, we get

$$D^* X^n = \int_0^\infty X_t^n\, dB_t \to \int_0^\infty X_t\, dB_t \ \text{ in } L^2.$$

Hence, Theorem 21.14 yields[7] $X \in \mathcal{D}^*$ and $D^* X = \int_0^\infty X_t\, dB_t$. □

We proceed with a striking differentiation property of the Itô integral.

Theorem 21.29 *(derivative of Itô integral, Pardoux & Peng) Let ζ be an isonormal Gaussian process on $L^2([0,1], \lambda)$ generating a Brownian motion B, and consider a B-adapted process $X \in L^2(P \otimes \lambda)$. Then*

(i) $X \in \hat{\mathcal{D}}_{1,2} \iff (X \cdot B)_1 \in \mathcal{D}_{1,2}$,

in which case

(ii) $X \cdot B \in \hat{\mathcal{D}}_{1,2}$, *and* $(X \cdot B)_t \in \mathcal{D}_{1,2}$, $t \in [0,1]$,

(iii) $D_s \int_0^t X_r\, dB_r = X_s 1_{s \leq t} + \int_s^t (D_s X_r)\, dB_r$ *a.e.*, $\quad s, t \in [0,1]$.

---

[7]Indeed, the adjoint of a closed operator is always closed.

Proof: First let $X \in \hat{\mathcal{D}}_{1,2}$. Since for fixed t the process $s \mapsto D_t X_s$ is square integrable by definition and adapted by Corollary 21.13, we have $X \in \mathcal{D}^*$ by Theorem 21.28 (i). Furthermore, the isometry of the Itô integral yields

$$E\int_0^1 dt \left| \int_t^1 D_t X_s \, dB_s \right|^2 = \int_0^1 dt \int_t^1 E(D_t X_s)^2 ds < \infty,$$

and so $X \cdot B \in \hat{\mathcal{D}}_{1,2}$. By Theorems 21.20 and 21.28 (ii) we conclude that $(X \cdot B)_1 \in \mathcal{D}_{1.2}$ for all $t \in [0,1]$, and that (iii) holds for $s \le t$. Equation (iii) remains true for $s > t$, since all three terms then vanish by Corollary 21.13 (i).

It remains to prove the implication '$\Leftarrow$' in (i), so assume that $(X \cdot B)_1 \in \mathcal{D}_{1,2}$. Let X_t^n denote the projection of X_t onto $\mathcal{P}_n = \mathcal{H}_0 \oplus \cdots \oplus \mathcal{H}_n$. Then $(X^n \cdot B)_t$ is the projection of $(X \cdot)_t$ onto $\mathcal{P}_{n+1}$, and so $(X^n \cdot B)_1 \to (X \cdot B)_1$ in the topology of $\mathcal{D}_{1,2}$. Applying (iii) to each process X^n and using the martingale property of the Itô integral, we obtain

$$\begin{aligned}
\int_0^1 E\Big\{D_s (X^n \cdot B)_1\Big\}^2 ds &= \int_0^1 E|X_s^n|^2 \, ds + \int_0^1 ds \int_0^s E(D_r X_s^n)^2 \, dr \\
&\ge \int_0^1 ds \int_0^s E(D_r X_s^n)^2 \, dr \\
&= E\Big\| DX^n \Big\|_{\lambda^2}^2 .
\end{aligned}$$

By a similar estimate for $DX^m - DX^n$ we conclude that DX^n converges in $L^2(\lambda^2)$, and so $X \in \hat{\mathcal{D}}_{1,2}$ by the closure of D. (Alternatively, we may use a version of Lemma 21.6.) □

Next recall from Lemma 19.12 that any B-measurable random variable $\xi \in L^2$ with $E\xi = 0$ can be written as an Itô integral $\int_0^\infty X_t \, dB_t$, for an a.e. unique process $X \in L^2_{\mathcal{F}}(P \otimes \lambda)$. We may now give an explicit formula for the integrand X.

Theorem 21.30 *(Brownian functionals, Clark, Ocone) Let B be a Brownian motion on $[0,1]$ with induced filtration $\mathcal{F}$. Then for any $\xi \in \mathcal{D}_{1,2}$,*

(i) $\xi = E\xi + \int_0^1 E\Big(D_t \xi \,\Big|\, \mathcal{F}_t\Big) \, dB_t,$

(ii) $E|\xi|^p \lesssim |E\xi|^p + \int_0^1 E|D_t \xi|^p \, dt, \quad p \ge 2.$

For the stated formula to make sense, we need to choose a progressively measurable version of the integrand. Such an *optional projection* of $D_t \xi$ is easily constructed by a suitable approximation.

Proof: (i) Assuming $\xi = \sum_n \zeta^n f_n$, we get by Theorem 21.7 and Corollary 14.28

$$\begin{aligned}
E\Big(D_t \xi \,\Big|\, \mathcal{F}_t\Big) &= \sum_{n>0} n \, E\Big\{ \zeta^{n-1} f_n(\cdot, t) \,\Big|\, \mathcal{F}_t \Big\} \\
&= \sum_{n>0} n \, \zeta^{n-1} \Big\{ 1_{[0,t]^{n-1}} f_n(\cdot, t) \Big\} .
\end{aligned}$$

The left-hand side is further adapted and square integrable, hence belongs to $\mathcal{D}^*$ by Theorem 21.28 (i). Applying D^* to the last sum and using Theorems 21.15 and 21.28 (ii), we conclude that

$$\int_0^1 E\Big(D_t\xi \,\Big|\, \mathcal{F}_t\Big)\, dB_t = \sum_{n>0} \zeta^n f_n = \xi - E\xi,$$

where the earlier factor n is canceled out by the symmetrization.

(ii) Using Jensen's inequality (three times), along with Fubini's theorem and the Burkholder inequality in Theorem 18.7, we get from (i)

$$\begin{aligned} E|\,\xi\,|^p - |E\,\xi|^p \;&\lesssim\; E\left|\int_0^1 E\Big(D_t\xi \,\Big|\, \mathcal{F}_t\Big)\, dB_t\right|^p \\ &\lesssim\; E\left|\int_0^1 \Big\{E\Big(D_t\xi \,\Big|\, \mathcal{F}_t\Big)\Big\}^2 dt\right|^{p/2} \\ &\le\; E\int_0^1 \Big|E\Big(D_t\xi \,\Big|\, \mathcal{F}_t\Big)\Big|^p\, dt \\ &\le\; \int_0^1 E|D_t\xi|^p\, dt. \end{aligned}$$

□

The Malliavin calculus can often be used to show the existence and possible regularity of a density. The following simple result illustrates the idea.

Theorem 21.31 *(functionals with smooth density) For an isonormal Gaussian process ζ on H, let $\xi \in L^2(\zeta)$ with*

$$\xi \in \mathcal{D}_{1,2}, \quad D\xi \in \mathcal{D}^*, \quad \|D\xi\| > 0.$$

Then $\mathcal{L}(\xi) \ll \lambda$ with the bounded, continuous density[8]

$$g(x) = E\Big\{D^*\Big(\|D\xi\|^{-2} D\xi\Big);\, \xi > x\Big\}, \quad x \in \mathsf{R}.$$

Proof: Let $f \in \hat{C}(\mathsf{R})$ be arbitrary, and put $F(x) = \int_{-\infty}^x f(y)\, dy$. Then the chain rule in Theorem 21.1 yields $F(\xi) \in \mathcal{D}_{1,2}$ and

$$\Big\langle D(F\circ\xi),\, D\xi\Big\rangle = f(\xi)\,\|D\xi\|^2.$$

Using the D/D^*-duality and Fubini's theorem, we obtain

$$\begin{aligned} Ef(\xi) &= E\,\Big\langle D(F\circ\xi),\, \|D\xi\|^{-2} D\xi\Big\rangle \\ &= E\,D^*\Big(\|D\xi\|^{-2} D\xi\Big)\, F(\xi) \\ &= E\,D^*\Big(\|D\xi\|^{-2} D\xi\Big) \int_{-\infty}^{\xi} f(x)\, dx \\ &= \int E\Big\{D^*\Big(\|D\xi\|^{-2} D\xi\Big);\, \xi > x\Big\}\, f(x)\, dx. \end{aligned}$$

Since f was arbitrary, this gives the asserted density. □

Weaker conditions will often ensure the mere absolute continuity, though we may then be unable to give an explicit formula for the density.

[8] Since $\|D\xi\|^{-2}$ is a random variable, not a constant, it can't be pulled out in front of E or even of D^*.

Theorem 21.32 *(absolute continuity) For an isonormal Gaussian process ζ on H, let ξ be a ζ-measurable random variable. Then*

$$\left.\begin{array}{r}\xi \in \mathcal{D}^{1,1}_{\mathrm{loc}} \\ D\xi \neq 0 \ \ a.s.\end{array}\right\} \Rightarrow \ \mathcal{L}(\xi) \ll \lambda.$$

Proof: By localization, we may let $\xi \in \mathcal{D}_{1,1}$, and we may further assume that $\mu = \mathcal{L}(\xi)$ is supported by $[-1,1]$. We need to show that if $A \in \mathcal{B}_{[-1,1]}$ with $\lambda A = 0$, then $\mu A = 0$. As in Lemma 1.37, we may then choose some smooth functions $f_n : [-1,1] \to [0,1]$ with bounded derivatives, such that $f_n \to 1_A$ a.e. $\lambda + \mu$. Writing $F_n(x) = \int_{-1}^{x} f_n(y)\,dy$, we see by dominated convergence that $F_n(x) \to 0$ for all x. Since the F_n are again smooth with bounded derivatives, the chain rule in Theorem 21.1 yields $F_n(\xi) \in \mathcal{D}_{1,1}$ with

$$\begin{aligned} D(F_n \circ \xi) \ &= \ f_n(\xi)\, D\xi \\ &\to 1\{\xi \in A\}\, D\xi \ \text{ a.s.} \end{aligned}$$

Since also $F_n(\xi) \to 0$, the closure property of D gives $1\{\xi \in A\} D\xi = 0$ a.s., and since $D\xi \neq 0$ a.s. by hypothesis, we obtain $\mu A = P\{\xi \in A\} = 0$. □

Exercises

1. Show that when $H = L^2(T,\mu)$ and $\xi \in \mathcal{D}$, we can choose the process $D_t\xi$ to be product measurable on $T \times \Omega$.

2. Let H^1 be the space of functions $x \in C_0([0,1])$ of the form $x_t = \int_0^t \dot{x}_s\,ds$ with $\dot{x} \in H = L^2([0,1])$. Show that the relation $\langle x, y\rangle_{H^1} = \langle \dot{x}, \dot{y}\rangle_H$ defines an inner product on H^1, making it a Hilbert space isomorphic to H, and that the natural embedding $H^1 \to C_0$ is injective and continuous.

3. Let B be a Brownian motion on $[0,1]$ generated by an isonormal Gaussian process ζ on $L^2([0,1])$, so that $B_t = \zeta 1_{[0,t]}$. For any $f \in C^\infty(\mathsf{R}^d)$ and $t \in [0,1]$, show that the random variable $\xi = f(B_t)$ is smooth.

4. For $\xi = f(B_t)$ as above and $h \in L^2$, show that $\langle D\xi, h\rangle$ agrees with the directional derivative of ξ, in the direction of $h \cdot \lambda \in H^1$. (*Hint:* Write $\langle D\xi, h\rangle = f'(B_t)(h\cdot\lambda)_{t_i} = (d/d\varepsilon)\xi(B + \varepsilon h \cdot \lambda)_{\varepsilon=0}$.)

5. Let B be a Brownian motion generated by an isonormal Gaussian process ζ on $L^2([0,1])$. Show that for any $t \in [0,1]$, $n \in \mathsf{N}$, and $f \in \hat{C}^\infty$, the random variable $\xi = f(B_t^n)$ belongs to $\mathcal{D}_{1,2}$. (*Hint:* Note that B_t^n has a finite chaos expansion. Then use Lemma 21.5 and Corollary 21.8.)

6. For a Brownian motion B, some fixed $t_1,\dots,t_n$, and a suitably smooth function f on R^d, define $\xi = f(B_1,\dots,B_n)$. Use Corollary 21.27 to calculate an expression for $L\xi$.

7. Let B be a Brownian motion generated by an isonormal Gaussian process ζ on $L^2([0,1])$, and fix any $n \in \mathsf{N}$. For $X = B^n$, show that $X \in \mathcal{D}^*$, and calculate an expression for D^*X. Check that X satisfies the conditions in Theorem 21.28, compute the corresponding Itô integral $\int_0^1 X_t\,dB_t$, and compare the two expressions. (*Hint:* Use Theorem 21.15.)

8. For a Brownian motion B on $[0,1]$ generated by the isonormal process ζ, define $\xi = B_t$ for a fixed $t > 0$. Calculate $D_t \xi$, and verify the expression in Theorem 21.30.

9. For a Brownian motion B and a fixed $t > 0$, use Theorem 21.31 to calculate an expression for the density of $\mathcal{L}(B_t)$, and compare with the known formula for the normal density.

VII. Convergence and Approximation

Here our primary objective is to derive functional versions of the major limit theorems of probability theory, through the development of general embedding and compactness arguments. The intuitive Skorohod embedding in Chapter 22 leads easily to functional versions of the central limit theorem and the law of the iterated logarithm, and allows extensions to renewal and empirical processes, as well as to martingales with small jumps. Though the compactness approach in Chapter 23 is more subtle, as it depends on the development of compactness criteria in the underlying function or measure spaces, it has a much wider scope, as it applies even to processes with jumps and to random sets and measures. In Chapter 24, we develop some powerful large deviation principles, providing precise exponential bounds for tail probabilities in a great variety of contexts. The first two chapters contain essential core material, whereas the last chapter is more advanced and might be postponed.

— — —

22. Skorohod embedding and functional convergence. Here the key step is to express a centered random variable as the value of a Brownian motion B at a suitably bounded optional time. Iterating the construction yields an embedding of an entire random walk X into B. By estimating the associated approximation error, we can derive a variety of weak or strong limit theorems for X from the corresponding results for B, leading in particular to functional versions of the central limit theorem and the law of the iterated logarithm, with extensions to renewal and empirical processes. A similar embedding applies to martingales with uniformly small jumps.

23. Convergence in distribution. Here we develop the compactness approach to convergence in distribution in function and measure spaces. The key result is Prohorov's theorem, which shows how the required relative compactness is equivalent to tightness, under suitable conditions on the underlying space. The usefulness of the theory relies on our ability to develop efficient tightness criteria. Here we consider in particular the spaces $D_{\mathsf{R}_+,S}$ of functions in S with only jump discontinuities, and the space $\mathcal{M}_S$ of locally finite measures on S, leading to a wealth of useful convergence criteria.

24. Large deviations. The large deviation principles (LDPs) developed in this chapter can be regarded as refinements of the weak law of large numbers in a variety of contexts. The theory involves some general extension principles, leading to an extremely versatile and powerful body of methods and results, covering a wide range of applications. In particular, we include Schilder's LDP for Brownian motion, Sanov's LDP for empirical distributions, the Freidlin–Wentzel LDP for stochastic dynamical systems, and Strassen's functional law of the iterated logarithm.

Chapter 22

Skorohod Embedding and Functional Convergence

Embedding of random walk, Brownian martingales, moment identities and estimates, rate of continuity, approximations of random walk, functional central limit theorem, laws of the iterated logarithm, arcsine laws, approximation of renewal processes, empirical distribution functions, embedding and approximation of martingales, martingales with small jumps, time-scale comparison

In Chapter 6 we used analytic methods to derive criteria for sums of independent random variables to be approximately Gaussian. Though this may remain the easiest approach to the classical limit theorems, such results are best understood when viewed in the context of some general approximation results for random processes. The aim of this chapter is to develop the purely probabilistic technique of *Skorohod embedding* to derive such functional limit theorems.

The scope of applications of the present method extends far beyond the classical limit theorems. This is because some of the most important functionals of a random process, such as the supremum up to a fixed time or the time at which the maximum occurs, can not be expressed in terms of the values at finitely many times. The powerful new technique allows us to extend some basic result for Brownian motion from Chapter 14, including the acsine laws and the law of the iterated logarithm, to random walks and related sequences. Indeed, the technique of Skorohod embedding extends even to a wide class of martingales with small jumps.

From the statements for random walks, similar results can be deduced for various related processes. In particular, we prove a functional central limit theorem and a law of the iterated logarithm for renewal processes, and show how suitably normalized versions of the empirical distribution functions, based on a sample of i.i.d. random variables, can be approximated by a Brownian bridge.

In the simplest setting, we may consider a random walk (X_n), based on some i.i.d. random variables ξ_k with mean 0 and variance 1. Here we can prove the existence of a Brownian motion B and some optional times $\tau_1 \le \tau_2 \le \cdots$, such that $X_n = B_{\tau_n}$ a.s. for every n. For applications, it is essential to choose the τ_n such that the differences $\Delta\tau_n$ become i.i.d. with finite mean. The path of the step process $X_{[t]}$ will then be close to that of B, and many results for Brownian motion carry over, at least approximately, to the random walk.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_23

The present account depends in many ways on material from previous chapters. Thus, we rely on the basic theory of Brownian motion, as set forth in Chapter 14. We also make frequent use of ideas and results from Chapter 9 on martingales and optional times. Finally, occasional references are made to Chapter 5 for empirical distributions, to Chapter 8 for the transfer theorem, to Chapter 12 for random walks and renewal processes, and to Chapter 15 for Poisson processes.

More general approximations and functional limit theorems will be obtained by different methods in Chapters 23–24 and 30. The present approximations of martingales with small jumps are closely related to the time-change results for continuous local martingales in Chapter 19.

To clarify the basic ideas, we begin with a detailed discussion of the classical Skorohod embedding of random walks. Say that a random variable ξ or its distribution is *centered* if $E\xi = 0$.

Theorem 22.1 *(embedding of random walk, Skorohod) Let $\xi_1, \xi_2, \dots$ be i.i.d., centered random variables, and put $X_n = \xi_1 + \cdots + \xi_n$. Then there exists a filtered probability space with a Brownian motion B and some optional times $0 = \tau_0 \le \tau_1 \le \dots$ with i.i.d. differences $\Delta\tau_n = \tau_n - \tau_{n-1}$, such that*

$$(B_{\tau_n}) \stackrel{d}{=} (X_n), \qquad E\Delta\tau_n = E\xi_1^2, \qquad E(\Delta\tau_n)^2 \le 4E\xi_1^4.$$

Here the moment conditions are crucial, since without them the statement would be trivially true and totally useless. In fact, we could then take $B \perp\!\!\!\perp (\xi_n)$ and choose recursively $\tau_n = \inf\big\{t \ge \tau_{n-1};\, B_t = X_n\big\}$.

Our proof of Theorem 22.1 is based on some lemmas. First we list a few martingales associated with Brownian motion.

Lemma 22.2 *(Brownian martingales) For a Brownian motion B, these processes are martingales:*

$$B_t, \quad B_t^2 - t, \quad B_t^4 - 6\,tB_t^2 + 3\,t^2.$$

Proof: Note that

$$EB_t = EB_t^3 = 0, \qquad EB_t^2 = t, \qquad EB_t^4 = 3\,t^2.$$

Write $\mathcal{F}$ for the filtration induced by B, let $0 \le s \le t$, and recall that the process $\tilde{B}_t = B_{s+t} - B_s$ is again a Brownian motion independent of $\mathcal{F}_s$. Hence,

$$\begin{aligned} E\Big(B_t^2 \,\Big|\, \mathcal{F}_s\Big) &= E\Big(B_s^2 + 2B_s\tilde{B}_{t-s} + \tilde{B}_{t-s}^2 \Big|\, \mathcal{F}_s\Big) \\ &= B_s^2 + t - s. \end{aligned}$$

Moreover,

$$\begin{aligned} E\Big(B_t^4 \,\Big|\, \mathcal{F}_s\Big) &= E\Big(B_s^4 + 4B_s^3\tilde{B}_{t-s} + 6\,B_s^2\tilde{B}_{t-s}^2 + 4B_s\tilde{B}_{t-s}^3 + \tilde{B}_{t-s}^4 \Big|\, \mathcal{F}_s\Big) \\ &= B_s^4 + 6\,(t-s)B_s^2 + 3\,(t-s)^2, \end{aligned}$$

and so

$$E\Big(B_t^4 - 6\,tB_t^2 \Big|\, \mathcal{F}_s\Big) = B_s^4 - 6\,sB_s^2 + 3\,(s^2 - t^2). \qquad \square$$

By optional sampling, we may deduce some useful moment formulas.

Lemma 22.3 *(moment relations) For any Brownian motion B and optional time τ such that B^τ is bounded, we have*

$$EB_\tau = 0, \qquad E\tau = EB_\tau^2, \qquad E\tau^2 \le 4EB_\tau^4. \tag{1}$$

Proof: By optional stopping and Lemma 22.2, we get for any $t \ge 0$

$$EB_{\tau\wedge t} = 0, \qquad E(\tau \wedge t) = EB_{\tau\wedge t}^2, \tag{2}$$

$$3\,E(\tau \wedge t)^2 + EB_{\tau\wedge t}^4 = 6\,E(\tau \wedge t)B_{\tau\wedge t}^2. \tag{3}$$

The first two relations in (1) follow from (2) by dominated and monotone convergence as $t \to \infty$. In particular, we have $E\tau < \infty$. We may then take limits even in (3), and conclude by dominated and monotone convergence, together with Cauchy's inequality, that

$$\begin{aligned} 3\,E\tau^2 + EB_\tau^4 &= 6\,E\tau B_\tau^2 \\ &\le 6\left(E\tau^2 EB_\tau^4\right)^{1/2}. \end{aligned}$$

Writing $r = (E\tau^2/EB_\tau^4)^{1/2}$, we get $3\,r^2 + 1 \le 6\,r$, which yields $3\,(r-1)^2 \le 2$, and finally $r \le 1 + (2/3)^{1/2} < 2$. □

Next we may write the centered distributions[1] as mixtures of suitable two-point distributions. For any $a \le 0 \le b$, there exists a unique probability measure $\nu_{a,b}$ on $\{a, b\}$ with mean 0, given by $\nu_{a,b} = \delta_0$ when $ab = 0$, and otherwise

$$\nu_{a,b} = \frac{b\,\delta_a - a\,\delta_b}{b-a}, \quad a < 0 < b.$$

This defines a probability kernel $\nu\colon \mathsf{R}_- \times \mathsf{R}_+ \to \mathsf{R}$, where for mappings between measure spaces, measurability is defined in terms of the σ-fields generated by the *evaluation maps* $\pi_B : \mu \mapsto \mu B$, for arbitrary sets B in the underlying σ-field.

Lemma 22.4 *(centered distributions as mixtures) For any centered distribution μ on R, there exists a μ-measurable distribution $\tilde\mu$ on $\mathsf{R}_- \times \mathsf{R}_+$, such that*

$$\mu = \int \tilde\mu(dx\,dy)\,\nu_{x,y}.$$

Proof (Chung): Let $\mu_\pm$ denote the restrictions of μ to $\mathsf{R}_\pm \setminus \{0\}$, define $l(x) \equiv x$, and put $c = \int l\,d\mu_+ = -\int l\,d\mu_-$. For any measurable function $f\colon \mathsf{R} \to \mathsf{R}_+$ with $f(0) = 0$, we get

$$\begin{aligned} c\int f d\mu &= \int l\,d\mu_+ \int f d\mu_- - \int l\,d\mu_- \int f d\mu_+ \\ &= \iint (y-x)\,\mu_-(dx)\,\mu_+(dy) \int f\,d\nu_{x,y}, \end{aligned}$$

and so we may choose

$$\tilde\mu(dx\,dy) = \mu\{0\}\,\delta_{0,0}(dx\,dy) + c^{-1}(y-x)\,\mu_-(dx)\,\mu_+(dy).$$

[1] probability measures with mean 0

The measurability of the mapping $\mu \mapsto \tilde{\mu}$ is clear by a monotone-class argument, once we note that $\tilde{\mu}(A \times B)$ is a measurable function of μ for any $A, B \in \mathcal{B}$. □

The embedding in Theorem 22.1 may be constructed recursively, beginning with a single random variable ξ with mean 0.

Lemma 22.5 *(embedding of random variable) For any Brownian motion B and centered distribution μ on R, choose a random pair $(\alpha, \beta) \perp\!\!\!\perp B$ with distribution $\tilde{\mu}$ as in Lemma 22.4, and define a random time τ and filtration $\mathcal{F}$ by*

$$\tau = \inf\Big\{ t \geq 0;\ B_t \in \{\alpha, \beta\} \Big\}, \qquad \mathcal{F}_t = \sigma\Big\{ \alpha, \beta;\ B_s,\ s \leq t \Big\}.$$

Then τ is $\mathcal{F}$-optional with

$$\mathcal{L}(B_\tau) = \mu, \qquad E\,\tau = \int x^2 \mu(dx), \qquad E\,\tau^2 \leq 4 \int x^4 \mu(dx).$$

Proof: The process B is clearly an $\mathcal{F}$-Brownian motion, and τ is $\mathcal{F}$-optional as in Lemma 9.6 (ii). Using Lemma 22.3 and Fubini's theorem, we get

$$\begin{aligned} \mathcal{L}(B_\tau) &= E\,\mathcal{L}(B_\tau \,|\, \alpha, \beta) \\ &= E\,\nu_{\alpha,\beta} = \mu, \\ E\,\tau &= E\,E(\tau \,|\, \alpha, \beta) \\ &= E \int x^2 \nu_{\alpha,\beta}(dx) \\ &= \int x^2 \mu(dx), \\ E\,\tau^2 &= E\,E(\tau^2 \,|\, \alpha, \beta) \\ &\leq 4\,E \int x^4 \nu_{\alpha,\beta}(dx) \\ &= 4 \int x^4 \mu(dx). \end{aligned}$$

□

Proof of Theorem 22.1: Let μ be the common distribution of the ξ_n. Introduce a Brownian motion B and some independent i.i.d. random pairs (α_n, β_n), $n \in \mathsf{N}$, with the distribution $\tilde{\mu}$ of Lemma 22.4, and define recursively some random times $0 = \tau_0 \leq \tau_1 \leq \cdots$ by

$$\tau_n = \inf\Big\{ t \geq \tau_{n-1};\ B_t - B_{\tau_{n-1}} \in \{\alpha_n, \beta_n\} \Big\}, \quad n \in \mathsf{N}.$$

Each τ_n is clearly optional for the filtration $\mathcal{F}_t = \sigma\{\alpha_k, \beta_k,\ k \geq 1;\ B^t\}$, $t \geq 0$, and B is an $\mathcal{F}$-Brownian motion. By the strong Markov property at τ_n, the process $B_t^{(n)} = B_{\tau_n + t} - B_{\tau_n}$ is again a Brownian motion independent of $\mathcal{G}_n = \sigma\{\tau_k, B_{\tau_k};\ k \leq n\}$. Since also $(\alpha_{n+1}, \beta_{n+1}) \perp\!\!\!\perp (B^{(n)}, \mathcal{G}_n)$, we obtain $(\alpha_{n+1}, \beta_{n+1}, B^{(n)}) \perp\!\!\!\perp \mathcal{G}_n$, and so the pairs $(\Delta\tau_n, \Delta B_{\tau_n})$ are i.i.d. The remaining assertions now follow by Lemma 22.5. □

Using the last theorem, we may approximate a centered random walk by a Brownian motion. As before, we assume the underlying probability space to be rich enough to support the required randomization variables.

Theorem 22.6 *(approximation of random walk, Skorohod, Strassen) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with $E\,\xi_i = 0$ and $E\,\xi_i^2 = 1$, and put $X_n = \xi_1 + \cdots + \xi_n$. Then there exists a Brownian motion B, such that*

(i) $\quad t^{-1/2} \sup_{s\le t} \left| X_{[s]} - B_s \right| \xrightarrow{P} 0, \quad t \to \infty,$

(ii) $\quad \lim_{t\to\infty} \dfrac{X_{[t]} - B_t}{\sqrt{2t \log\log t}} = 0 \quad$ *a.s.*

The proof of (ii) requires the following estimate.

Lemma 22.7 *(rate of continuity) For a real Brownian motion B,*

$$\lim_{r\downarrow 1} \limsup_{t\to\infty} \sup_{t\le u\le rt} \frac{|B_u - B_t|}{\sqrt{2t\log\log t}} = 0 \quad a.s.$$

Proof: Write $h(t) = (2t \log\log t)^{1/2}$. It is enough to show that

$$\lim_{r\downarrow 1} \limsup_{n\to\infty} \sup_{r^n\le t\le r^{n+1}} \frac{|B_t - B_{r^n}|}{h(r^n)} = 0 \ \text{ a.s.} \tag{4}$$

Proceeding as in the proof of Theorem 14.18, we get[2] as $n \to \infty$ for fixed $r > 1$ and $c > 0$

$$\begin{aligned} P\Big\{ \sup_{t\in[r^n, r^{n+1}]} \Big| B_t - B_{r^n} \Big| > c h(r^n) \Big\} &\lesssim P\Big\{ B\{r^n(r-1)\} > c h(r^n) \Big\} \\ &\lesssim n^{-c^2/(r-1)} (\log n)^{-1/2}. \end{aligned}$$

Choosing $c^2 > r - 1$ and using the Borel–Cantelli lemma, we see that the lim sup in (4) is a.s. bounded by c, and the assertion follows as we let $r \to 1$. □

For the main proof, we introduce the *modulus of continuity*

$$w(f, t, h) = \sup_{r,s\le t,\, |r-s|\le h} |f_r - f_s|, \quad t, h > 0.$$

Proof of Theorem 22.6: By Theorems 8.17 and 22.1, we may choose a Brownian motion B and some optional times $0 \equiv \tau_0 \le \tau_1 \le \cdots$, such that $X_n = B_{\tau_n}$ a.s. for all n, and the differences $\tau_n - \tau_{n-1}$ are i.i.d. with mean 1. Then $\tau_n/n \to 1$ a.s. by the law of large numbers, and so $\tau_{[t]}/t \to 1$ a.s. Now (ii) follows by Lemma 22.7.

Next define

$$\delta_t = \sup_{s\le t} \left| \tau_{[s]} - s \right|, \quad t \ge 0,$$

and note that the a.s. convergence $\tau_n/n \to 1$ implies $\delta_t/t \to 0$ a.s. Fixing any t, h, $\varepsilon > 0$, and using the scaling property of B, we get

[2]Recall that $a \lesssim b$ means $a \le c\,b$ for some constant $c > 0$.

$$\begin{aligned}
P\Big\{\, t^{-1/2} \sup_{s\le t} \Big|B_{\tau_{[s]}} - B_s\Big| > \varepsilon \,\Big\}
&\le P\Big\{w\Big(B, t+th, th\Big) > \varepsilon\, t^{1/2}\Big\} + P\{\delta_t > th\}\\
&= P\Big\{w(B, 1+h, h) > \varepsilon\Big\} + P\{t^{-1}\delta_t > h\}.
\end{aligned}$$

Here the right-hand side tends to zero as $t \to \infty$ and then $h \to 0$, and (i) follows. □

For an immediate application of the last theorem, we may extend the law of the iterated logarithm to suitable random walks.

Corollary 22.8 *(law of the iterated logarithm, Hartman & Wintner) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with $E\xi_i = 0$ and $E\xi_i^2 = 1$, and put $X_n = \xi_1 + \cdots + \xi_n$. Then*

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2n \log\log n}} = 1 \quad a.s.$$

Proof: Combine Theorems 14.18 and 22.6. □

The first approximation in Theorem 22.6 yields a classical weak convergence result for random walks. Here we introduce the space $D_{[0,1]}$ of right-continuous functions on $[0,1]$ with left-hand limits (rcll). For the present purposes, we equip $D_{[0,1]}$ with the norm $\|x\| = \sup_t |x_t|$ and the σ-field $\mathcal{D}$ generated by all evaluation maps $\pi_t : x \mapsto x_t$. Since the norm is clearly $\mathcal{D}$-measurable[3], the same thing is true for the open balls $B_x^r = \{y;\ \|x-y\| < r\}$, $x \in D_{[0,1]}$, $r > 0$.

For a process X with paths in $D_{[0,1]}$ and a functional $f : D_{[0,1]} \to \mathsf{R}$, we say that f is a.s. continuous at X if $X \notin D_f$ a.s., where D_f denotes the set of functions[4] $x \in D_{[0,1]}$, such that f is discontinuous at the path x.

We may now state a version of the functional central limit theorem.

Theorem 22.9 *(functional central limit theorem, Donsker) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with $E\xi_i = 0$ and $E\xi_i^2 = 1$, and define*

$$X_t^n = n^{-1/2} \sum_{k\le nt} \xi_k, \quad t \in [0,1],\ n \in \mathsf{N}.$$

Let the functional $f\colon D_{[0,1]} \to \mathsf{R}$ be measurable and a.s. continuous at the paths of a Brownian motion B on $[0,1]$. Then

$$f(X^n) \xrightarrow{d} f(B).$$

This follows immediately from Theorem 22.6, together with the following lemma.

[3]Note, however, that $\mathcal{D}$ is strictly smaller than the Borel σ-field induced by the norm.

[4]The measurability of D_f is irrelevant here, provided we interpret the a.s. condition in the sense of inner measure.

Lemma 22.10 *(approximation and convergence)* *For* $n \in \mathbb{N}$, *let* X_n, Y_n, Y *be rcll processes on* $[0,1]$ *such that*

$$Y_n \stackrel{d}{=} Y, \quad n \in \mathbb{N}; \qquad \|X_n - Y_n\| \xrightarrow{P} 0,$$

and let the functional $f: D_{[0,1]} \to \mathbb{R}$ *be measurable and a.s. continuous at* Y. *Then*

$$f(X_n) \xrightarrow{d} f(Y).$$

Proof: Put $T = \mathbb{Q} \cap [0,1]$. By Theorem 8.17 there exist some processes X'_n on T, such that $(X'_n, Y) \stackrel{d}{=} (X_n, Y_n)$ on T for all n. Then each X'_n is a.s. bounded with finitely many upcrossings of any non-degenerate interval, and so the process $\tilde{X}_n(t) = X'_n(t+)$ exists a.s. with paths in $D_{[0,1]}$. From the right continuity of paths, it is also clear that $(\tilde{X}_n, Y) \stackrel{d}{=} (X_n, Y_n)$ on $[0,1]$ for every n.

To obtain the desired convergence, we note that

$$\left\|\tilde{X}_n - Y\right\| \stackrel{d}{=} \|X_n - Y_n\| \xrightarrow{P} 0,$$

and hence $f(X_n) \stackrel{d}{=} f(\tilde{X}_n) \xrightarrow{P} f(Y)$, as in Lemma 5.3. □

In particular, we may recover the central limit theorem in Proposition 6.10 by taking $f(x) = x_1$ in Theorem 22.9. We may also obtain results that go beyond the classical theory, such as by choosing $f(x) = \sup_t |x_t|$. For a less obvious application, we may extend the arcsine laws of Theorem 14.16 to suitable random walks. Recall that a random variable ξ is said to be *arcsine distributed* if $\xi \stackrel{d}{=} \sin^2 \alpha$, where α is $U(0, 2\pi)$.

Theorem 22.11 *(arcsine laws, Erdös & Kac, Sparre-Andersen)* *Let* $\xi_1, \xi_2, \ldots$ *be i.i.d., non-degenerate random variables, and put* $X_n = \xi_1 + \cdots + \xi_n$. *Define for* $n \in \mathbb{N}$

$$\begin{aligned}
\tau_n^1 &= n^{-1} \sum_{k \le n} 1\{X_k > 0\}, \\
\tau_n^2 &= n^{-1} \min\Big\{ k \ge 0;\ X_k = \max_{i \le n} X_i \Big\}, \\
\tau_n^3 &= n^{-1} \max\Big\{ k \le n;\ X_k X_n \le 0 \Big\},
\end{aligned}$$

and let τ *be arcsine distributed. Then*

(i) $\tau_n^k \xrightarrow{d} \tau$ *for* $k = 1, 2, 3$ *when* $E\xi_i = 0$ *and* $E\xi_i^2 < \infty$,

(ii) $\tau_n^k \xrightarrow{d} \tau$ *for* $k = 1, 2$ *when* $\mathcal{L}(\xi_i)$ *is symmetric.*

For the proof, we introduce on $D_{[0,1]}$ the functionals

$$\begin{aligned}
f_1(x) &= \lambda\Big\{ t \in [0,1];\ x_t > 0 \Big\}, \\
f_2(x) &= \inf\Big\{ t \in [0,1];\ x_t \vee x_{t-} = \sup_{s \le 1} x_s \Big\}, \\
f_3(x) &= \sup\Big\{ t \in [0,1];\ x_t x_1 \le 0 \Big\}.
\end{aligned}$$

The following result is elementary.

Lemma 22.12 *(continuity of functionals) The functionals f_i are measurable, and*

- *f_1 is continuous at x ⇔ $\lambda\{t;\, x_t = 0\} = 0$,*
- *f_2 is continuous at x ⇔ $x_t \vee x_{t-}$ has a unique maximum,*
- *f_3 is continuous at x, if all local extremes of x_t or x_{t-} on $(0,1]$ are $\neq 0$.*

Proof of Theorem 22.11: Clearly, $\tau_n^i = f_i(X^n)$ for $n \in \mathbb{N}$ and $i = 1,2,3$, where

$$X_t^n = n^{-1/2} X_{[nt]}, \quad t \in [0,1],\ n \in \mathbb{N}.$$

(i) By Theorems 14.16 and 22.9, it suffices to show that each f_i is a.s. continuous at B. Thus, we need to verify that B a.s. satisfies the conditions in Lemma 22.12. This is obvious for f_1, since Fubini's theorem yields

$$E\lambda\Big\{t \le 1;\, B_t = 0\Big\} = \int_0^1 P\{B_t = 0\}dt = 0.$$

The conditions for f_2 and f_3 follow easily from Lemma 14.15.

(ii) Since $\tau_n^1 \overset{d}{=} \tau_n^2$ by Corollary 27.8, it is enough to consider τ_n^1. Then introduce an independent Brownian motion B, and define

$$\sigma_n^\varepsilon = n^{-1} \sum_{k \le n} 1\Big\{\varepsilon B_k + (1-\varepsilon) X_k > 0\Big\}, \quad n \in \mathbb{N},\ \varepsilon \in (0,1].$$

By (i) together with Theorem 12.12 and Corollary 27.8, we have $\sigma_n^\varepsilon \overset{d}{=} \sigma_n^1 \xrightarrow{d} \tau$. Since $P\{X_n = 0\} \to 0$, e.g. by Theorem 5.17, we also note that

$$\limsup_{\varepsilon \to 0} \Big|\sigma_n^\varepsilon - \tau_n^1\Big| \le n^{-1} \sum_{k \le n} 1\{X_k = 0\} \xrightarrow{P} 0.$$

Hence, we may choose some $\varepsilon_n \to 0$ with $\sigma_n^{\varepsilon_n} - \tau_n^1 \xrightarrow{P} 0$, and Theorem 5.29 yields $\tau_n^1 \xrightarrow{d} \tau$. □

Results like Theorem 22.9 are called *invariance principles*, since the limiting distribution of $f(X^n)$ is the same[5] for all i.i.d. sequences (ξ_k) with $E\xi_i = 0$ and $E\xi_i^2 = 1$. This is often useful for applications, since a direct computation may be possible for a special choice of distribution, such as for $P\{\xi_k = \pm 1\} = \frac{1}{2}$.

The approximations in Theorem 22.6 yield similar results for renewal processes, here regarded as non-decreasing step processes.

Theorem 22.13 *(approximation of renewal process) Let N be a renewal process based on a distribution μ with mean 1 and variance $\sigma^2 \in (0,\infty)$. Then there exists a Brownian motion B, such that*

(i) $$t^{-1/2} \sup_{s \le t} \Big| N_s - s - \sigma B_s \Big| \xrightarrow{P} 0, \quad t \to \infty,$$

(ii) $$\lim_{t\to\infty} \frac{N_t - t - \sigma B_t}{\sqrt{2t \log\log t}} = 0 \quad a.s.$$

[5]sometimes also referred to as *universality*

Proof: (ii) Let $\tau_0, \tau_1, \ldots$ be the renewal times of N, and introduce the random walk $X_n = n - \tau_n + \tau_0$, $n \in \mathsf{Z}_+$. Choosing a Brownian motion B as in Theorem 22.6, we get as $n \to \infty$

$$\frac{N_{\tau_n} - \tau_n - \sigma B_n}{\sqrt{2n \log \log n}} = \frac{X_n - \sigma B_n}{\sqrt{2n \log \log n}} \to 0 \quad \text{a.s.}$$

Since $\tau_n \sim n$ a.s. by the law of large numbers, we may replace n in the denominator by τ_n, and by Lemma 22.7 we may further replace B_n by B_{τ_n}. Hence,

$$\frac{N_t - t - \sigma B_t}{\sqrt{2t \log \log t}} \to 0 \quad \text{a.s. along } (\tau_n).$$

By Lemma 22.7 it is enough to show that

$$\frac{\tau_{n+1} - \tau_n}{\sqrt{2\tau_n \log \log \tau_n}} \to 0 \quad \text{a.s.,}$$

which is easily seen from Theorem 22.6.

(i) From Theorem 22.6, we further obtain

$$n^{-1/2} \sup_{k \le n} \left| N_{\tau_k} - \tau_k - \sigma B_k \right| = n^{-1/2} \sup_{k \le n} \left| X_k - \tau_0 - \sigma B_k \right| \xrightarrow{P} 0,$$

and by Brownian scaling,

$$n^{-1/2} w(B, n, 1) \stackrel{d}{=} w(B, 1, n^{-1}) \to 0.$$

It is then enough to show that

$$n^{-1/2} \sup_{k \le n} \left| \tau_k - \tau_{k-1} - 1 \right| = n^{-1/2} \sup_{k \le n} \left| X_k - X_{k-1} \right| \xrightarrow{P} 0,$$

which is again clear from Theorem 22.6. □

Proceeding as in Corollary 22.8 and Theorem 22.9, we may deduce an associated law of the iterated logarithm and a weak convergence result.

Corollary 22.14 *(limits of renewal process) Let N be a renewal process based on a distribution μ with mean* 1 *and variance $\sigma^2 < \infty$, and let the functional $f : D_{[0,1]} \to \mathsf{R}$ be measurable and a.s. continuous at a Brownian motion B. Then*

(i) $$\limsup_{t \to \infty} \frac{\pm(N_t - t)}{\sqrt{2t \log \log t}} = \sigma \quad \text{a.s.,}$$

(ii) $f(X^r) \xrightarrow{d} f(B)$ *as $r \to \infty$, where*

$$X_t^r = \frac{N_{rt} - rt}{\sigma \sqrt{r}}, \quad t \in [0,1], \ r > 0.$$

Part (ii) above yields a similar result for empirical distribution functions, based on a sequence of i.i.d. random variables. The asymptotic behavior can then be expressed in terms of a Brownian bridge.

Theorem 22.15 *(approximation of empirical distributions) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with ordinary and empirical distribution functions F and $\hat{F}_1, \hat{F}_2, \ldots$. Then there exist some Brownian bridges $B^1, B^2, \ldots$, such that*

$$\sup_x \Big| n^{1/2}\Big\{\hat{F}_n(x) - F(x)\Big\} - B^n \circ F(x)\Big| \xrightarrow{P} 0, \quad n \to \infty. \tag{5}$$

Proof: Arguing as in the proof of Proposition 5.24, we may reduce to the case where the ξ_n are $U(0,1)$, and $F(t) \equiv t$ on $[0,1]$. Then clearly

$$n^{1/2}\Big\{\hat{F}_n(t) - F(t)\Big\} = n^{-1/2} \sum_{k \le n} \Big(1\{\xi_k \le t\} - t\Big), \quad t \in [0,1].$$

Now introduce for every n an independent Poisson random variable κ_n with mean n, and conclude from Proposition 15.4 that

$$N_t^n = \sum_{k \le \kappa_n} 1\{\xi_k \le t\}, \quad t \in [0,1],$$

is a homogeneous Poisson process on $[0,1]$ with rate n. By Theorem 22.13, there exist some Brownian motions W^n on $[0,1]$ with

$$\sup_{t \le 1} \Big| n^{-1/2}(N_t^n - nt) - W_t^n \Big| \xrightarrow{P} 0.$$

For the associated Brownian bridges $B_t^n = W_t^n - tW_1^n$, we get

$$\sup_{t \le 1} \Big| n^{-1/2}(N_t^n - tN_1^n) - B_t^n \Big| \xrightarrow{P} 0.$$

To deduce (5), it is enough to show that

$$n^{-1/2} \sup_{t \le 1} \Big| \sum_{k \le |\kappa_n - n|} \Big(1\{\xi_k \le t\} - t\Big) \Big| \xrightarrow{P} 0. \tag{6}$$

Here $|\kappa_n - n| \xrightarrow{P} \infty$, e.g. by Proposition 6.10, and so (6) holds by Proposition 5.24 with $n^{1/2}$ replaced by $|\kappa_n - n|$. It remains to note that $n^{-1/2}|\kappa_n - n|$ is tight, since $E(\kappa_n - n)^2 = n$. □

We turn to a martingale version of the Skorohod embedding in Theorem 22.1 with associated approximation in Theorem 22.6.

Theorem 22.16 *(embedding of martingale) For a martingale (M_n) with $M_0 = 0$ and induced filtration $(\mathcal{G}_n)$, there exists a Brownian motion B with associated optional times $0 = \tau_0 \le \tau_1 \le \cdots$, such that a.s. for all $n \in \mathsf{N}$,*

(i) $M_n = B_{\tau_n}$,

(ii) $E\Big(\Delta\tau_n \Big| \mathcal{F}_{n-1}\Big) = E\Big\{(\Delta M_n)^2 \Big| \mathcal{G}_{n-1}\Big\}$,

(iii) $E\Big\{(\Delta\tau_n)^2 \Big| \mathcal{F}_{n-1}\Big\} \le 4\, E\Big\{(\Delta M_n)^4 \Big| \mathcal{G}_{n-1}\Big\}$,

where $(\mathcal{F}_n)$ denotes the filtration induced by the pairs (M_n, τ_n).

Proof: Let $\mu_1, \mu_2, \ldots$ be probability kernels satisfying

$$\mathcal{L}\big(\Delta M_n \big| \mathcal{G}_{n-1}\big) = \mu_n(M_1, \ldots, M_{n-1}; \cdot) \text{ a.s.,} \quad n \in \mathsf{N}. \tag{7}$$

Since the M_n form a martingale, we may assume that $\mu_n(x; \cdot)$ has mean 0 for all $x \in \mathsf{R}^{n-1}$. Define the associated measures $\tilde{\mu}_n(x; \cdot)$ on R^2 as in Lemma 22.4, and conclude from the measurability part of the lemma that $\tilde{\mu}_n$ is a probability kernel from R^{n-1} to R^2. Next choose some measurable functions $f_n\colon \mathsf{R}^n \to \mathsf{R}^2$ as in Lemma 4.22, such that $\mathcal{L}\{f_n(x, \vartheta)\} = \tilde{\mu}_n(x, \cdot)$ when ϑ is $U(0,1)$.

Now fix a Brownian motion B' and some independent i.i.d. $U(0,1)$ random variables $\vartheta_1, \vartheta_2, \ldots$. Let $\tau'_0 = 0$, and define recursively some random variables $\alpha_n, \beta_n, \tau'_n$ for $n \in \mathsf{N}$ by

$$(\alpha_n, \beta_n) = f_n\Big(B'_{\tau'_1}, \ldots, B'_{\tau'_{n-1}}, \vartheta_n\Big), \tag{8}$$

$$\tau'_n = \inf\Big\{t \geq \tau'_{n-1};\, B'_t - B'_{\tau'_{n-1}} \in \{\alpha_n, \beta_n\}\Big\}. \tag{9}$$

Since B' is a Brownian motion for the filtration $\mathcal{B}_t = \sigma\{(B')^t, (\vartheta_n)\}$, $t \geq 0$, and each τ'_n is $\mathcal{B}$-optional, the strong Markov property shows that $B^{(n)}_t = B'_{\tau'_n+t} - B'_{\tau'_n}$ is again a Brownian motion independent of $\mathcal{F}'_n = \sigma\{\tau'_k, B'_{\tau'_k};\, k \leq n\}$. Since also $\vartheta_{n+1} \perp\!\!\!\perp (B^{(n)}, \mathcal{F}'_n)$, we have $(B^{(n)}, \vartheta_{n+1}) \perp\!\!\!\perp \mathcal{F}'_n$. Writing $\mathcal{G}'_n = \sigma\{B'_{\tau'_k};\, k \leq n\}$, we conclude that

$$\Big(\Delta\tau'_{n+1}, \Delta B'_{\tau'_{n+1}}\Big) \underset{\mathcal{G}'_n}{\perp\!\!\!\perp} \mathcal{F}'_n. \tag{10}$$

Using (8) and Theorem 8.5, we get

$$\mathcal{L}\Big(\alpha_n, \beta_n \Big| \mathcal{G}'_{n-1}\Big) = \tilde{\mu}_n\Big(B'_{\tau'_1}, \ldots, B'_{\tau'_{n-1}}; \cdot\Big). \tag{11}$$

Since also $B^{(n-1)} \perp\!\!\!\perp (\alpha_n, \beta_n, \mathcal{G}'_{n-1})$, we have $B^{(n-1)} \perp\!\!\!\perp_{\mathcal{G}'_{n-1}} (\alpha_n, \beta_n)$, and $B^{(n-1)}$ is conditionally a Brownian motion. Applying Lemma 22.5 to the conditional distributions given $\mathcal{G}'_{n-1}$, we get by (9), (10), and (11)

$$\mathcal{L}\Big(\Delta B'_{\tau'_n} \Big| \mathcal{G}'_{n-1}\Big) = \mu_n\Big(B'_{\tau'_1}, \ldots, B'_{\tau'_{n-1}}; \cdot\Big), \tag{12}$$

$$\begin{aligned} E\Big(\Delta\tau'_n \Big| \mathcal{F}'_{n-1}\Big) &= E\Big(\Delta\tau'_n \Big| \mathcal{G}'_{n-1}\Big) \\ &= E\Big\{(\Delta B'_{\tau'_n})^2 \Big| \mathcal{G}'_{n-1}\Big\}, \end{aligned} \tag{13}$$

$$\begin{aligned} E\Big\{(\Delta\tau'_n)^2 \Big| \mathcal{F}'_{n-1}\Big\} &= E\Big\{(\Delta\tau'_n)^2 \Big| \mathcal{G}'_{n-1}\Big\} \\ &\leq 4\, E\Big\{(\Delta B'_{\tau'_n})^4 \Big| \mathcal{G}'_{n-1}\Big\}. \end{aligned} \tag{14}$$

Comparing (7) and (12), we get $(B'_{\tau'_n}) \overset{d}{=} (M_n)$. By Theorem 8.17, we may then choose a Brownian motion B with associated optional times $\tau_1, \tau_2, \ldots$, such that

$$\Big\{B, (M_n), (\tau_n)\Big\} \overset{d}{=} \Big\{B', (B'_{\tau'_n}), (\tau'_n)\Big\}.$$

All a.s. relations between the objects on the right, including their conditional expectations with respect to appropriate induced σ-fields, remain valid for the objects on the left. In particular, (i) holds for all n, and (13)−(14) imply the corresponding formulas (ii)−(iii). □

The last theorem allows us to approximate a martingale with small jumps by a Brownian motion. For discrete-time martingales M, we define the *quadratic variation* $[M]$ and *predictable variation* $\langle M\rangle$ by

$$[M]_n = \sum_{k\le n} (\Delta M_k)^2,$$
$$\langle M\rangle_n = \sum_{k\le n} E\Big\{(\Delta M_k)^2 \Big| \mathcal{F}_{k-1}\Big\},$$

in analogy with the continuous-time versions in Chapters 18 and 20.

Theorem 22.17 *(approximation of martingales with small jumps) For $n\in \mathsf{N}$, let M^n be an $\mathcal{F}_n$-martingale M^n on Z_+ satisfying*

$$M_0^n = 0, \qquad |\Delta M_k^n| \le 1, \qquad \sup\nolimits_k |\Delta M_k^n| \xrightarrow{P} 0, \tag{15}$$

put $\zeta_n = [M^n]_\infty$, and define

$$X_t^n = \textstyle\sum_k \Delta M_k^n \, 1\Big\{[M^n]_k \le t\Big\}, \quad t\in[0,1],\ n\in \mathsf{N}.$$

Then

(i) $(X^n - B^n)^*_{\zeta_n\wedge 1} \xrightarrow{P} 0$ *for some Brownian motions B^n,*

(ii) *claim* (i) *remains true with $[M^n]$ replaced by $\langle M^n\rangle$,*

(iii) *the third condition in* (15) *can be replaced by*

$$\textstyle\sum_k P\Big\{|\Delta M_k^n| > \varepsilon \Big| \mathcal{F}_{k-1}^n\Big\} \xrightarrow{P} 0, \quad \varepsilon > 0. \tag{16}$$

For the proof, we need to show that the time scales given by the sequences (τ_k^n), $[M^n]$, and $\langle M^n\rangle$ are asymptotically equivalent.

Lemma 22.18 *(time-scale comparison) In Theorem* 22.17*, let $M_k^n = B^n(\tau_k^n)$ a.s. for some Brownian motions B^n with associated optional times τ_k^n as in Theorem* 22.16*, and put*

$$\kappa_t^n = \inf\Big\{k;\ [M^n]_k > t\Big\}, \quad t\ge 0,\ n\in\mathsf{N}.$$

Then as $n\to\infty$ for fixed $t>0$,

$$\sup_{k\le \kappa_t^n} \Big\{\Big|\tau_k^n - [M^n]_k\Big| \vee \Big|[M^n]_k - \langle M^n\rangle_k\Big|\Big\} \xrightarrow{P} 0. \tag{17}$$

Proof: By optional stopping, we may assume $[M^n]$ to be uniformly bounded, and take the supremum in (17) over all k. To handle the second difference in (17), we note that $D^n = [M^n] - \langle M^n\rangle$ is a martingale for each n. Using the martingale property, Proposition 9.17, and dominated convergence, we get

$$\begin{aligned}
E(D^n)^{*2} &\lesssim \sup\nolimits_k E(D_k^n)^2 \\
&= \textstyle\sum_k E(\Delta D_k^n)^2 \\
&= \textstyle\sum_k E\,E\Big\{(\Delta D_k^n)^2 \Big| \mathcal{F}_{k-1}^n\Big\} \\
&\le \textstyle\sum_k E\,E\Big\{\Big(\Delta [M^n]_k\Big)^2 \Big| \mathcal{F}_{k-1}^n\Big\} \\
&= E\textstyle\sum_k (\Delta M_k^n)^4 \\
&\lesssim E\,\sup\nolimits_k (\Delta M_k^n)^2 \to 0,
\end{aligned}$$

and so $(D^n)^* \xrightarrow{P} 0$. This clearly remains true if each sequence $\langle M^n \rangle$ is defined in terms of the filtration $\mathcal{G}^n$ induced by M^n.

To complete the proof of (17), it is enough to show, for the latter versions of $\langle M^n \rangle$, that $(\tau^n - \langle M^n \rangle)^* \xrightarrow{P} 0$. Then write $\mathcal{T}^n$ for the filtration induced by the pairs (M_k^n, τ_k^n), $k \in \mathsf{N}$, and conclude from Theorem 22.16 (ii) that

$$\langle M^n \rangle_m = \sum_{k \le m} E\Big(\Delta\tau_k^n \,\Big|\, \mathcal{T}_{k-1}^n\Big), \quad m, n \in \mathsf{N}.$$

Hence, $\tilde{D}^n = \tau^n - \langle M^n \rangle$ is a $\mathcal{T}^n$-martingale. Using Theorem 22.16 (ii)$-$(iii), we get as before

$$\begin{aligned} E(\tilde{D}^n)^{*2} &\lesssim \sup_k E(\tilde{D}_k^n)^2 \\ &= \sum_k E\,E\Big\{(\Delta\tilde{D}_k^n)^2 \Big| \mathcal{T}_{k-1}^n\Big\} \\ &\le \sum_k E\,E\Big\{(\Delta\tau_k^n)^2 \Big| \mathcal{T}_{k-1}^n\Big\} \\ &\lesssim \sum_k E\,E\Big\{(\Delta M_k^n)^4 \Big| \mathcal{G}_{k-1}^n\Big\} \\ &= E\sum_k (\Delta M_k^n)^4 \\ &\lesssim E\,\sup_k (\Delta M_k^n)^2 \to 0. \end{aligned}$$

□

The sufficiency of (16) is a consequence of the following simple estimate.

Lemma 22.19 *(Dvoretzky) For any filtration $\mathcal{F}$ on Z_+ and sets $A_n \in \mathcal{F}_n$, $n \in \mathsf{N}$, we have*

$$P\bigcup_n A_n \le P\Big\{\sum_n P(A_n | \mathcal{F}_{n-1}) > \varepsilon\Big\} + \varepsilon, \quad \varepsilon > 0.$$

Proof: Write $\xi_n = 1_{A_n}$ and $\hat{\xi}_n = P(A_n | \mathcal{F}_{n-1})$, fix any $\varepsilon > 0$, and define $\tau = \inf\{n;\, \hat{\xi}_1 + \cdots + \hat{\xi}_n > \varepsilon\}$. Then $\{\tau \le n\} \in \mathcal{F}_{n-1}$ for each n, and so

$$\begin{aligned} E\sum_{n<\tau} \xi_n &= \sum_n E\Big(\xi_n;\, \tau > n\Big) \\ &= \sum_n E\Big(\hat{\xi}_n;\, \tau > n\Big) \\ &= E\sum_{n<\tau} \hat{\xi}_n \le \varepsilon. \end{aligned}$$

Hence,

$$\begin{aligned} P\bigcup_n A_n &\le P\{\tau < \infty\} + E\sum_{n<\tau} \xi_n \\ &\le P\Big\{\sum_n \hat{\xi}_n > \varepsilon\Big\} + \varepsilon. \end{aligned}$$

□

Proof of Theorem 22.17: To prove the result for the time scales $[M^n]$, we may reduce by optional stopping to the case where $[M^n] \le 2$ for all n. For every $n \in \mathsf{N}$, we may choose a Brownian motion B^n with associated optional times τ_k^n, as in Theorem 22.16. Then

$$(X^n - B^n)^*_{\zeta_n \wedge 1} \le w\Big(B^n, 1 + \delta_n, \delta_n\Big), \quad n \in \mathsf{N},$$

where

$$\delta_n = \sup_k \Big\{\Big|\tau_k^n - [M^n]_k\Big| + (\Delta M_k^n)^2\Big\},$$

and so

$$E\Big\{\Big(X^n - B^n\Big)^*_{\zeta_n \wedge 1} \wedge 1\Big\} \le E\Big\{w\Big(B^n, 1+h, h\Big) \wedge 1\Big\} + P\{\delta_n > h\}.$$

Since $\delta_n \xrightarrow{P} 0$ by Lemma 22.18, the right-hand side tends to zero as $n \to \infty$ and then $h \to 0$, and the assertion follows.

In case of the time scales $\langle M^n \rangle$, define $\kappa_n = \inf\{k; [M^n] > 2\}$. Then $[M^n]_{\kappa_n} - \langle M^n \rangle_{\kappa_n} \xrightarrow{P} 0$ by Lemma 22.18, and so

$$P\Big\{\langle M^n \rangle_{\kappa_n} < 1,\ \kappa_n < \infty\Big\} \to 0.$$

We may then reduce by optional stopping to the case where $[M^n] \le 3$. The proof may now be completed as before. □

Though the Skorohod embedding has no natural extension to higher dimensions, we can still derive some multi-variate approximations by applying the previous results separately in each component. To illustrate the method, we show how suitable random walks in R^d can be approximated by continuous processes with stationary, independent increments. Extensions to more general limits are obtained by different methods in Corollary 16.17 and Theorem 23.14.

Theorem 22.20 *(approximation of random walks in R^d) Let $X^1, X^2, \dots$ be random walks in R^d with $\mathcal{L}(X^n_{m_n}) \xrightarrow{w} N(0, \sigma\sigma')$, for some $d \times d$ matrix σ and integers $m_n \to \infty$, and define $Y^n_t = X^n_{[m_n t]}$. Then there exist some Brownian motions $B^1, B^2, \dots$ in R^d, such that*

$$\Big(Y^n - \sigma B_n\Big)^*_t \xrightarrow{P} 0, \quad t \ge 0.$$

Proof: Since by Theorem 6.16,

$$\max_{k \le m_n t} \Big|\Delta X^n_k\Big| \xrightarrow{P} 0, \quad t \ge 0,$$

we may assume that $|\Delta X^n_k| \le 1$ for all n and k. Subtracting the means, we may further take $E X^n_k \equiv 0$. Applying Theorem 22.17 in each coordinate yields $w(Y^n, t, h) \xrightarrow{P} 0$, as $n \to \infty$ and then $h \to 0$. Furthermore, $w(\sigma B, t, h) \to 0$ a.s. as $h \to 0$.

Applying Theorem 6.16 in both directions yields $Y^n_{t_n} \xrightarrow{d} \sigma B_t$ as $t_n \to t$. Hence, by independence, $(Y^n_{t_1}, \dots, Y^n_{t_m}) \xrightarrow{d} \sigma(B_{t_1}, \dots, B_{t_m})$ for all $n \in \mathsf{N}$ and $t_1, \dots, t_n \ge 0$, and so $Y^n \xrightarrow{d} \sigma B$ on Q_+ by Theorem 5.30. By Theorem 5.31, or even simpler by Corollary 8.19 and Theorem A5.4, there exist some rcll processes $\tilde{Y}^n \stackrel{d}{=} Y^n$ with $\tilde{Y}^n_t \to \sigma B_t$ a.s. for all $t \in \mathsf{Q}_+$. For any $t, h > 0$,

$$\begin{aligned} E\Big\{\Big(\tilde{Y}^n - \sigma B\Big)^*_t \wedge 1\Big\} \le\ & E\Big(\max_{j \le t/h} \Big|\tilde{Y}^n_{jh} - \sigma B_{jh}\Big| \wedge 1\Big) \\ & + E\Big\{w\Big(\tilde{Y}^n, t, h\Big) \wedge 1\Big\} + E\Big\{w(\sigma B, t, h) \wedge 1\Big\}. \end{aligned}$$

Multiplying by e^{-t}, integrating over $t > 0$, and letting $n \to \infty$ and then $h \to 0$ along Q_+, we get by dominated convergence

$$\int_0^\infty e^{-t} E\Big\{\Big(\tilde{Y}^n - \sigma B\Big)_t^* \wedge 1\Big\}\, dt \to 0.$$

By monotonicity, the last integrand tends to zero as $n \to \infty$, and so $(\tilde{Y}^n - \sigma B)_t^* \xrightarrow{P} 0$ for each $t > 0$. Now use Theorem 8.17. □

Exercises

1. Show that Theorem 22.6 yields a purely probabilistic proof of the central limit theorem, with no need of characteristic functions.

2. Given a Brownian motion B, put $\tau_0 = 0$, and define recursively $\tau_n = \inf\{t > \tau_{n-1};\, |B_{\tau_n} - B_{\tau_{n-1}}| = 1\}$. Show that $E\Delta\tau_n = \infty$ for all n. (*Hint:* We may use Proposition 12.14.)

3. Construct some Brownian martingales as in Lemma 22.2 with leading terms B_t^3 and B_t^5. Use multiple Wiener–Itô integrals to give an alternative proof of the lemma. For every $n \in \mathsf{N}$, find a martingale with leading term B_t^n. (*Hint:* Use Theorem 14.25.)

4. For a Brownian motion B with associated optional time $\tau < \infty$, show that $E\tau \geq EB_\tau^2$. (*Hint:* Truncate τ and use Fatou's lemma.)

5. For X_n as in Corollary 22.8, show that the sequence of random variables $(2n \log\log n)^{-1/2} X_n$, $n \geq 3$, is a.s. relatively compact with set of limit points $[-1, 1]$. (*Hint:* Prove the corresponding property for Brownian motion, and use Theorem 22.6.)

6. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random vectors in R^d with mean 0 and covariances δ_{ij}. Show that Corollary 22.8 remains true with X_n replaced by $|X_n|$. More precisely, show that the sequence $(2n \log\log n)^{-1/2} X_n$, $n \geq 3$, is relatively compact in R^d, with all limit points contained in the closed unit ball. (*Hint:* Apply Corollary 22.8 to the projections $u \cdot X_n$ for arbitrary $u \in \mathsf{R}^d$ with $|u| = 1$.)

7. In Theorem 14.18, show that for any $c \in (0, 1)$ we may choose $t_n \to \infty$, such that a.s. the limsup along (t_n) equals c. Conclude that the set of limit points in the preceding exercise agrees with the closed unit ball in R^d.

8. Condition (16) clearly follows from $\sum_k E\Big(|\Delta M_k^n| \wedge 1 \,\Big|\, \mathcal{F}_{n-1}^n\Big) \xrightarrow{P} 0$. Show by an example that the converse implication is false. (*Hint:* Consider a sequence of random walks.)

9. Specialize Lemma 22.18 to random walks, and give a direct proof in that case.

10. For random walks, show that condition (16) is also necessary for the approximation in Theorem 22.17. (*Hint:* Use Theorem 6.16.)

11. Specialize Theorem 22.17 to random walks in R, and derive a corresponding extension of Theorem 22.9. Then derive a functional version of Theorem 6.13.

12. Specialize further to successive renormalizations of a single random walk X_n. Then derive a limit theorem for the values at $t = 1$, and compare with Proposition 6.10.

13. In the second arcsine law of Theorem 22.11, show that the first maximum on $[0,1]$ can be replaced by the last one. Conclude that the associated times σ_n and τ_n satisfy $\tau_n - \sigma_n \xrightarrow{P} 0$. (*Hint:* Use the corresponding result for Brownian motion. Alternatively, use the symmetry of both (X_n) and the arcsine distribution.)

14. Extend Theorem 22.11 to symmetric random walks satisfying a Lindeberg condition. Further extend the results for τ_n^1 and τ_n^2 to random walks based on diffuse, symmetric distributions. Finally, show that the result for τ_n^3 may fail in the latter case. (*Hint:* Consider the n^{-1}-increments of a compound Poisson process based on the uniform distribution on $[-1,1]$, perturbed by a small diffusion term $\varepsilon_n B$, where B is an independent Brownian motion.)

15. In the context of Theorem 22.20 and for a Brownian motion B, show that we can choose $Y^n \overset{d}{=} X^n$ with $(Y^n - \sigma B)_t^* \to 0$ a.s. for all $t \geq 0$. Prove a corresponding version of Theorem 22.17. (*Hint:* Use Theorem 5.31 or Corollary 8.19.)

Chapter 23

Convergence in Distribution

Tightness and relative compactness, convergence and tightness in function spaces, convergence of continuous and rcll processes, functional central limit theorem, moments and tightness, optional equi-continuity and tightness, approximation of random walks, tightness and convergence of random measures, existence via tightness, vague and weak convergence, Cox and thinning continuity, strong Cox continuity, measure-valued processes, convergence of random sets and point processes

Distributional convergence is another of those indispensable topics, constantly used throughout modern probability. The basic notions were introduced already in Chapter 5, and in Chapter 6 we proved some fundamental limit theorems for sums of independent random variables. In the previous chapter we introduced the powerful method of Skorohod embedding, which enabled us to establish some functional versions of those results, where the Gaussian limits are replaced by entire paths of Brownian motion and related processes.

Despite the amazing power of the latter technique, its scope of applicability is inevitably restricted by the limited class of possible limit laws. Here we turn to a systematic study of general weak convergence theory, based on some fundamental principles of compactness in appropriate function and measure spaces. In particular, some functional limit theorems, derived in the last chapter by cumbersome embedding and approximation techniques, will now be accessible by straightforward compactness arguments.

The key result is Prohorov's theorem, which gives the basic connection between tightness and relative distributional compactness. This result enables us to convert some classical compactness criteria into similar probabilistic versions. In particular, the Arzelà–Ascoli theorem yields a corresponding criterion for distributional compactness of continuous processes. Similarly, an optional equi-continuity condition guarantees the appropriate compactness, for right-continuous processes with left-hand limits[1]. We will also derive some general criteria for convergence in distribution of random measures and sets, with special emphasis on the point process case.

The general criteria will be applied to some interesting concrete situations. In addition to some already familiar results from Chapters 22, we will derive convergence criteria for Cox processes and thinnings of point processes. Further applications appear in other chapters, such as a general approximation result

[1] henceforth referred to as *rcll processes*

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_24

for Markov chains in Chapter 17, various limit theorems for exchangeable processes and particle systems in Chapters 27 and 30, and a way of constructing weak solutions to SDEs in Chapter 32. To avoid overloading this chapter with technical detail, we outline some of the underlying analysis in Appendices 5–6.

Beginning with continuous processes, we fix some separable and complete metric spaces (T,d), (S,ρ), where T is also locally compact, and consider the space $C_{T,S}$ of continuous functions from T to S, endowed with the *locally uniform topology*. The associated notion of convergence, written as $x_n \xrightarrow{ul} x$, is given by

$$\sup_{t\in K} \rho(x_{n,t}, x_t) \to 0, \quad K \in \mathcal{K}_T, \tag{1}$$

where $\mathcal{K}_T$ denotes the class of compact sets $K \subset T$. Since T is locally compact, we may choose $K_1, K_2, \ldots \in \mathcal{K}_T$ with $K_n^o \uparrow T$, and it is enough to verify (1) for all K_n. Using such a sequence, we may easily construct a separable and complete metrization of the topology. In Lemma A5.1 we show that the Borel σ-field in $C_{T,S}$ is generated by the evaluation maps $\pi_t : x \mapsto x_t$, $t \in T$. This is significant for our purposes, since the random elements in $C_{T,S}$ are then precisely the continuous processes on T taking values in S.

Given any such processes X and $X_1, X_2, \ldots$, we need to find tractable conditions for the associated convergence in distribution, written as $X_n \xrightarrow{uld} X$. This mode of convergence is clearly strictly stronger than the finite-dimensional convergence $X_n \xrightarrow{fd} X$, defined by[2]

$$(X^n_{t_1}, \ldots, X^n_{t_m}) \xrightarrow{d} (X_{t_1}, \ldots, X_{t_m}), \quad t_1, \ldots, t_m \in T, \;\; m \in \mathsf{N}. \tag{2}$$

This is clear already in the non-random case, since the point-wise convergence $x^n_t \to x_t$ in $C_{T,S}$ may not be locally uniform.

The classical way around this difficulty is to require the functions x_n to be *locally equi-continuous*. Formally, we may impose suitable conditions on the associated *local moduli of continuity*

$$w_K(x,h) = \sup\Big\{\rho(x_s, x_t);\; s,t \in K,\; d(s,t) \le h\Big\}, \quad h > 0, \;\; K \in \mathcal{K}_T. \tag{3}$$

Using the Arzelà–Ascoli compactness criterion, in the version of Theorem A5.2, we may easily characterize convergence in $C_{T,S}$. Though this case is quite elementary, it is discussed in detail below, since it sets the pattern for the more advanced criteria for distributional convergence, considered throughout the remainder of the chapter.

Lemma 23.1 *(convergence in $C_{T,S}$) Let x, $x_1, x_2, \ldots$ be continuous, S-valued functions on T, where S, T are separable, complete metric spaces and T is locally compact with a dense subset T'. Then $x_n \xrightarrow{ul} x$ in $C_{T,S}$ iff*

[2] To avoid double subscripts, we may sometimes write x_n, X_n as x^n, X^n.

(i) $x_t^n \to x_t$ *for all* $t \in T'$,

(ii) $\lim_{h\to 0} \limsup_{n\to\infty} w_K(x_n, h) = 0, \quad K \in \mathcal{K}_T.$

Proof: If $x_n \xrightarrow{ul} x$, then (i) holds by continuity and (ii) holds by Theorem A5.2. Conversely, assume (i)−(ii). Then (x_n) is relatively compact by Theorem A5.2, and so for any sub-sequence $N' \subset \mathbb{N}$ we have convergence $x_n \xrightarrow{ul} y$ along a further sub-sequence N''. Then by continuity, $x_n(t) \to y(t)$ along N'' for all $t \in T$, and so (i) yields $x(t) = y(t)$ for all $t' \in T'$. Since x and y are continuous and T' is dense in T, we get $x = y$. Hence, $x_n \xrightarrow{ul} x$ along N'', which extends to $\mathbb{N}$ since N' was arbitrary. □

The approach for random processes is similar, except that here we need to replace the compactness criterion in Theorem A5.2 by a criterion for *compactness in distribution*. The key result is Prohorov's Theorem 23.2 below, which relates compactness in a given metric space S to *weak compactness*[3] in the associated space $\hat{\mathcal{M}}_S$ of probability measures on S. For most applications of interest, we may choose S to be a suitable function or measure space.

To make this precise, let Ξ be a family of random elements in a metric space S. Generalizing a notion from Chapter 5, we say that Ξ is *tight* if

$$\inf_{K\in\mathcal{K}_S} \sup_{\xi\in\Xi} P\{\xi \notin K\} = 0. \tag{4}$$

We also say that Ξ is *relatively compact in distribution*, if every sequence of random elements $\xi_1, \xi_2, \ldots \in \Xi$ has a sub-sequence converging in distribution toward a random element ξ in S.

We may now state the fundamental equivalence between tightness and relative compactness, for random elements in sufficiently regular metric spaces. This extends the version for Euclidean spaces in Proposition 6.22.

Theorem 23.2 *(tightness and relative compactness, Prohorov) For a set* Ξ *of random elements in a metric space* S, *we have* (i) ⇒ (ii) *with*

(i) Ξ *is tight,*

(ii) Ξ *is relatively compact in distribution,*

and equivalence holds when S *is separable and complete.*

In particular, we see that when S is separable and complete, a single random element ξ in S is *tight*, in the sense that $\inf_K P\{\xi \notin K\} = 0$. For sequences we may then replace the 'sup' in (4) by 'lim sup'.

For the proof of Theorem 23.2, we need a simple lemma. Recall from Lemma 1.6 that a random element in a sub-space of a metric space S may also be regarded as a random element in S.

[3] For all applications in this book, S and $\hat{\mathcal{M}}_S$ are both metric spaces, and compactness in the topological sense is equivalent to sequential compactness.

Lemma 23.3 *(tightness preservation) Tightness is preserved by continuous maps. Thus, if S is a metric space and Ξ is a set of random elements in a subset $A \subset S$, we have*

$$\Xi \text{ is tight in } A \quad \Rightarrow \quad \Xi \text{ is tight in } S.$$

Proof: Compactness is preserved by continuous maps. This applies in particular to the natural embedding $I: A \to S$. □

Proof of Theorem 23.2 *(Varadarajan):* For $S = \mathsf{R}^d$, the result was proved in Proposition 6.22. Turning to the case of $S = \mathsf{R}^\infty$, consider a tight sequence of random elements $\xi^n = (\xi_1^n, \xi_2^n, \ldots)$ in R^∞. Writing $\eta_k^n = (\xi_1^n, \ldots, \xi_k^n)$, we see from Lemma 23.3 that the sequence $(\eta_k^n;\, n \in \mathsf{N})$ is tight in R^k for each $k \in \mathsf{N}$. For any sub-sequence $N' \subset \mathsf{N}$, a diagonal argument yields $\eta_k^n \xrightarrow{d}$ some η_k for all $k \in \mathsf{N}$, as $n \to \infty$ along a further sub-sequence N''. The sequence $\{\mathcal{L}(\eta_k)\}$ is projective, since the coordinate projections are continuous, and so Theorem 8.21 yields a random sequence $\xi = (\xi_1, \xi_2, \ldots)$ with $(\xi_1, \ldots, \xi_k) \stackrel{d}{=} \eta_k$ for all k. But then $\xi^n \xrightarrow{fd} \xi$ along N'', and so Theorem 5.30 gives $\xi^n \xrightarrow{d} \xi$ along the same sequence.

Next let $S \subset \mathsf{R}^\infty$. If (ξ_n) is tight in S, then by Lemma 23.3 it remains tight as a sequence in R^∞. Hence, for any sequence $N' \subset \mathsf{N}$, we have $\xi_n \xrightarrow{d}$ some ξ in R^∞ along a further sub-sequence N''. To extend the convergence to S, it suffices by Lemma 5.26 to check that $\xi \in S$ a.s. Then choose some compact sets $K_m \subset S$ with $\liminf_n P\{\xi_n \in K_m\} \geq 1 - 2^{-m}$ for all $m \in \mathsf{N}$. Since the K_m remain closed in R^∞, Theorem 5.25 yields

$$\begin{aligned} P\{\xi \in K_m\} &\geq \limsup_{n \in N''} P\{\xi_n \in K_m\} \\ &\geq \liminf_{n \to \infty} P\{\xi_n \in K_m\} \\ &\geq 1 - 2^{-m}, \end{aligned}$$

and so $\xi \in \bigcup_m K_m \subset S$ a.s.

Now let S be σ-compact. It is then separable and therefore homeomorphic to a subset $A \subset \mathsf{R}^\infty$. By Lemma 23.3, the tightness of (ξ_n) carries over to the image sequence $(\tilde{\xi}_n)$ in A, and by Lemma 5.26 the possible relative compactness of $(\tilde{\xi}_n)$ implies the same property for (ξ_n). This reduces the discussion to the previous case.

Now turn to the general case. If (ξ_n) is tight, there exist some compact sets $K_m \subset S$ with $\liminf_n P\{\xi_n \in K_m\} \geq 1 - 2^{-m}$. In particular, $P\{\xi_n \in A\} \to 1$ with $A = \bigcup_m K_m$, and so $P\{\xi_n = \eta_n\} \to 1$ for some random elements η_n in A. Then (η_n) is again tight, even as a sequence in A, and since A is σ-compact, we see as before that (η_n) is relatively compact as a sequence in A. By Lemma 5.26 it remains relatively compact in S, and by Theorem 5.29 the relative compactness carries over to (ξ_n).

To prove the converse, let S be separable and complete, and suppose that (ξ_n) is relatively compact. For any $r > 0$, we may cover S by some open balls

$B_1, B_2, \ldots$ of radius r. Writing $G_k = B_1 \cup \cdots \cup B_k$, we claim that

$$\lim_{k\to\infty} \inf_n P\{\xi_n \in G_k\} = 1. \tag{5}$$

If this fails, we may choose some integers $n_k \uparrow \infty$ with $\sup_k P\{\xi_{n_k} \in G_k\} = c < 1$. The relative compactness yields $\xi_{n_k} \xrightarrow{d}$ some ξ along a sub-sequence $N' \subset \mathsf{N}$, and so

$$\begin{aligned} P\{\xi \in G_m\} &\le \liminf_{k\in N'} P\Big\{\xi_{n_k} \in G_m\Big\} \\ &\le c < 1, \qquad m \in \mathsf{N}. \end{aligned}$$

As $m \to \infty$, we get the contradiction $1 < 1$, proving (5).

Now take $r = m^{-1}$, and write G_k^m for the corresponding sets G_k. For any $\varepsilon > 0$, (5) yields some $k_1, k_1, \ldots \in \mathsf{N}$ with

$$\inf_n P\Big\{\xi_n \in G_{k_m}^m\Big\} \ge 1 - \varepsilon\, 2^{-m}, \quad m \in \mathsf{N}.$$

Writing $A = \bigcap_m G_{k_m}^m$, we get $\inf_n P\{\xi_n \in A\} \ge 1 - \varepsilon$. Since $\bar{A}$ is complete and totally bounded, hence compact, it follows that (ξ_n) is tight. $\Box$

For applications of the last theorem, we need to find effective criteria for tightness. Beginning with the space $C_{T,S}$, we may convert the classical Arzelà–Ascoli compactness criterion into a condition for tightness, stated in terms of the local moduli of continuity in (3). Note that $w_K(x,h)$ is continuous in x for fixed $h > 0$, and hence depends measurably on x.

Theorem 23.4 *(tightness in $C_{T,S}$, Prohorov) Let Ξ be a set of continuous, S-valued processes on T, where S, T are separable, complete metric spaces and T is locally compact with a dense subset T'. Then Ξ is tight in $C_{T,S}$ iff*

(i) $\pi_t \Xi$ *is tight in* S, $\quad t \in T'$,

(ii) $\displaystyle\lim_{h\to 0} \sup_{X\in\Xi} E\Big\{w_K(X,h) \wedge 1\Big\} = 0, \quad K \in \mathcal{K}_T,$

in which case

(iii) $\displaystyle\bigcup_{t\in K} \pi_t \Xi$ *is tight in* S, $\quad K \in \mathcal{K}_T$.

Proof: Let Ξ be tight in $C_{T,S}$, so that for every $\varepsilon > 0$ there exists a compact set $A \subset C_{T,S}$ with $P\{X \notin A\} \le \varepsilon$ for all $X \in \Xi$. Then for any $K \in \mathcal{K}_T$ and $X \in \Xi$,

$$P\Big\{\bigcup_{t\in K} \{X_t\} \notin \bigcup_{t\in K} \pi_t A\Big\} \le P\{X \notin A\} \le \varepsilon, \quad X \in \Xi,$$

and (iii) follows since $\bigcup_{t\in K} \pi_t A$ is relatively compact by Theorem A5.2 (iii). The same theorem yields an $h > 0$ with $w_K(x,h) \le \varepsilon$ for all $x \in A$, and so for any $X \in \Xi$,

$$E\Big\{w_K(X,h) \wedge 1\Big\} \le \varepsilon + P\Big\{w_K(X,h) > \varepsilon\Big\} \le 2\varepsilon,$$

which implies (ii), since $\varepsilon > 0$ was arbitrary. Finally, note that (iii) $\Rightarrow$ (i).

Conversely, assume (i)−(ii). Let $t_1, t_2, \ldots \in T'$ be dense in T, and choose $K_1, K_2, \ldots \in \mathcal{K}_T$ with $K_n^o \uparrow T$. For any $\varepsilon > 0$, (i) yields some compact sets $B_1, B_2, \ldots \subset S$ with

$$P\Big\{X(t_m) \notin B_m\Big\} \le 2^{-m}\varepsilon, \quad m \in \mathsf{N}, \; X \in \Xi, \tag{6}$$

and by (ii) there exist some $h_{nk} > 0$ with

$$P\Big\{w_{K_n}(X, h_{nk}) > 2^{-k}\Big\} \le 2^{-n-k}\varepsilon, \quad n, k \in \mathsf{N}, \; X \in \Xi. \tag{7}$$

By Theorem A5.2, the set

$$A_\varepsilon = \bigcap_{m,n,k} \Big\{x \in C_{T,S};\, x(t_m) \in B_m,\, w_{K_n}(x, h_{nk}) \le 2^{-k}\Big\}$$

is relatively compact in $C_{T,S}$, and for any $X \in \Xi$ we get by (6) and (7)

$$P\{X \notin A_\varepsilon\} \le \textstyle\sum_{n,k} 2^{-n-k}\varepsilon + \sum_m 2^{-m}\varepsilon = 2\,\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, Ξ is then tight in $C_{T,S}$. □

We may now characterize convergence in distribution of continuous, S-valued processes on T, generalizing the elementary Lemma 23.1 for non-random functions. Write $X_n \xrightarrow{uld} X$ for convergence in distribution in $C_{T,S}$ with respect to the locally uniform topology.

Corollary 23.5 *(convergence of continuous processes, Prohorov) Let $X, X_1, X_2, \ldots$ be continuous, S-valued processes on T, where S, T are separable, complete metric spaces and T is locally compact with a dense subset T'. Then $X_n \xrightarrow{uld} X$ iff*

(i) $X_n \xrightarrow{fd} X$ *on* T',

(ii) $$\lim_{h\to 0} \limsup_{n\to\infty} E\Big\{w_K(X_n, h) \wedge 1\Big\} = 0, \quad K \in \mathcal{K}_T. \tag{8}$$

Proof: If $X_n \xrightarrow{uld} X$, then (i) holds by Theorem 5.27 and (ii) by Theorems 23.2 and 23.4. Conversely, assume (i)−(ii). Then (X_n) is tight by Theorem 23.4, and so by Theorem 23.2 it is relatively compact in distribution. For any sub-sequence $N' \subset \mathsf{N}$, we get $X_n \xrightarrow{uld}$ some Y along a further sub-sequence N''. Then Theorem 5.27 yields $X_n \xrightarrow{fd} Y$ along N'', and so by (i) we have $X \overset{fd}{=} Y$ on T', which extends to T by Theorem 5.27. Hence, $X \overset{d}{=} Y$ by Proposition 4.2, and so $X_n \xrightarrow{uld} X$ along N'', which extends to N since N' was arbitrary. □

We illustrate the last result by proving a version of Donsker's Theorem 22.9 in R^d. Since Theorem 23.5 applies only to processes with continuous paths, we need to replace the original step processes by suitably interpolated versions. In Theorem 23.14 below, we will see how such an interpolation can be avoided.

Theorem 23.6 *(functional central limit theorem, Donsker) Let $\xi_1, \xi_2, \ldots$ be i.i.d. random vectors in R^d with mean 0 and covariances δ_{ij}, form the continuous processes*

$$X_t^n = n^{-1/2}\Big\{ \sum_{k \le nt} \xi_k + (nt - [nt])\, \xi_{[nt]+1} \Big\}, \quad t \ge 0, \ n \in \mathsf{N},$$

and let B be a Brownian motion in R^d. Then $X^n \xrightarrow{uld} B$ in $C_{\mathsf{R}_+, \mathsf{R}^d}$.

Proof: By Theorem 23.5 it is enough to prove convergence on $[0,1]$. Clearly, $X_n \xrightarrow{fd} X$ by Proposition 6.10 and Corollary 6.5. Combining the former result with Lemma 11.12, we further get the rough estimate

$$\lim_{r\to\infty} r^2 \limsup_{n\to\infty} P\Big\{ X_n^* \ge r\sqrt{n} \Big\} = 0,$$

which implies

$$\lim_{h\to 0} h^{-1} \limsup_{n\to\infty} \sup_{t\le 1} P\Big\{ \sup_{0\le r\le h} |X_{t+r}^n - X_t^n| > \varepsilon \Big\} = 0.$$

Now (8) follows easily, as we split $[0,1]$ into sub-intervals of length $\le h$. □

The Kolmogorov–Chentsov criterion in Theorem 4.23 can be converted into a sufficient condition for tightness in $C_{\mathsf{R}^d, S}$. An important application appears in Theorem 32.9.

Theorem 23.7 *(moment criteria and Hölder continuity) Let $X^1, X^2, \ldots$ be continuous processes on R^d with values in a separable, complete metric space (S, ρ). Then (X^n) is tight in $C_{\mathsf{R}^d, S}$, whenever (X_0^n) is tight in S and*

$$\sup_n E\Big| \rho(X_s^n, X_t^n) \Big|^a \lesssim |s-t|^{d+b}, \quad s, t \in \mathsf{R}^d, \tag{9}$$

for some $a, b > 0$. The limiting processes are then a.s. locally Hölder continuous with exponent c, for any $c \in (0, b/a)$.

Proof: For every X^n, we may define the associated quantities ξ_{nk} as in the proof of Theorem 4.23, and we see that $E\,\xi_{nk}^a \lesssim 2^{-kb}$. Hence, Lemma 1.31 yields for $m, n \in \mathsf{N}$

$$\begin{aligned} \Big\| w_K(X_n, 2^{-m}) \Big\|_a^{a\wedge 1} &\lesssim \sum_{k\ge m} \|\xi_{nk}\|_a^{a\wedge 1} \\ &\lesssim \sum_{k\ge m} 2^{-kb/(a\vee 1)} \\ &\lesssim 2^{-mb/(a\vee 1)}, \end{aligned}$$

which implies (8). Condition (9) extends by Lemma 5.11 to any limiting process X, and the last assertion follows by Theorem 4.23. □

For processes with possible jump discontinuities, the theory is similar but more complicated. Here we consider the space $D_{\mathsf{R}_+, S}$ of rcll[4] functions on R_+,

[4]right-continuous with left-hand limits

taking values in a separable, complete metric space (S,ρ). For any $\varepsilon, t > 0$, such a function x has clearly at most finitely many jumps of size $\rho(x_t, x_{t-}) > \varepsilon$ up to time t.

Though it still makes sense to consider the locally uniform topology with associated mode of convergence $x_n \xrightarrow{ul} x$, it is technically convenient and more useful for applications to allow appropriate time changes, given by increasing bijections on R_+. Note that such functions λ are continuous and strictly increasing with $\lambda_0 = 0$ and $\lambda_\infty = \infty$. We define the *Skorohod convergence* $x_n \xrightarrow{s} x$ by the requirements

$$\lambda_n \xrightarrow{ul} \iota, \qquad x_n \circ \lambda_n \xrightarrow{ul} x,$$

for some increasing bijections λ_n on R_+, where ι denotes the identity map on R_+. This mode of convergence generates a Polish topology on $D_{\mathsf{R}_+,S}$, known as *Skorohod's J_1-topology*, whose basic properties are listed in Lemma A5.3. In particular, the associated Borel σ-field is again generated by the evaluation maps π_t, and so the random elements in $D_{\mathsf{R}_+,S}$ are precisely the rcll processes in S.

The associated compactness criteria in Theorem A5.4 are analogous to the Arzeà–Ascoli conditions in Theorem A5.2, except for being based on the *modified local moduli of continuity*

$$\tilde{w}_t(x,h) = \inf_{(I_k)} \max_k \sup_{r,s \in I_k} \rho(x_r, x_s), \quad t, h > 0, \tag{10}$$

where the infimum extends over all partitions of the interval $[0,t)$ into subintervals $I_k = [u,v)$, such that $v - u \ge h$ when $v < t$. Note that $\tilde{w}_t(x,h) \to 0$ as $h \to 0$ for fixed $x \in D_{\mathsf{R}_+,S}$ and $t > 0$.

In order to derive criteria for convergence in distribution in $D_{\mathsf{R}_+,S}$, corresponding to the conditions of Theorem 23.5 for processes in $C_{T,S}$, we first need a convenient criterion for tightness.

Theorem 23.8 *(tightness in $D_{\mathsf{R}_+,S}$) Let Ξ be a set of rcll processes on R_+ with values in a separable, complete metric space S, and fix a dense set $T \subset \mathsf{R}_+$. Then Ξ is tight in $D_{\mathsf{R}_+,S}$ iff*

(i) $\pi_t \Xi$ *is tight in* S, $\quad t \in T$,

(ii) $\lim_{h\to 0} \sup_{X \in \Xi} E\left\{\tilde{w}_t(X,h) \wedge 1\right\} = 0, \quad t > 0,$

in which case

(iii) $\bigcup_{s \le t} \pi_s \Xi$ *is tight in* S, $\quad t \ge 0$.

Proof: Same as for Theorem 23.4, except that now we need to use Theorem A5.4 in place of Theorem A5.2. □

We may now characterize convergence in distribution of S-valued, rcll processes on R_+. In particular, the conditions simplify and lead to stronger conclusions, when the limiting process is continuous. Write $X_n \xrightarrow{sd} X$ and $X_n \xrightarrow{uld} X$ for convergence in distribution in $D_{\mathsf{R}_+,S}$ with respect to the Skorohod and locally uniform topologies.

Theorem 23.9 *(convergence of rcll processes, Skorohod) Let $X, X_1, X_2, \dots$ be rcll processes on R_+ with values in a separable, complete metric space S. Then*

(i) $X_n \xrightarrow{sd} X$, *iff* $X_n \xrightarrow{fd} X$ *on a dense set* $T \subset \{t \geq 0;\, X_t = X_{t-} \text{ a.s.}\}$ *and*

$$\lim_{h\to 0} \limsup_{n\to\infty} E\Big\{\tilde{w}_t(X_n, h) \wedge 1\Big\} = 0, \quad t > 0, \tag{11}$$

(ii) $X_n \xrightarrow{uld} X$ *with X a.s. continuous, iff* $X_n \xrightarrow{fd} X$ *on a dense set* $T \subset \mathsf{R}_+$ *and*

$$\lim_{h\to 0} \limsup_{n\to\infty} E\Big\{w_t(X_n, h) \wedge 1\Big\} = 0, \quad t > 0,$$

(iii) *for a.s. continuous X,*

$$X_n \xrightarrow{sd} X \quad \Leftrightarrow \quad X_n \xrightarrow{uld} X.$$

The present proof is more subtle than for continuous processes. For motivation and technical convenience, we begin with the non-random case.

Lemma 23.10 *(convergence in $D_{\mathsf{R}_+,S}$) Let $x, x_1, x_2, \dots$ be rcll functions on R_+ with values in a separable, complete metric space S. Then*

(i) $x_n \xrightarrow{s} x$, *iff* $x_t^n \to x_t$ *for all t in a dense set* $T \subset \{t \geq 0;\, x_t = x_{t-}\}$ *and*

$$\lim_{h\to 0} \limsup_{n\to\infty} \tilde{w}_t(x_n, h) = 0, \quad t > 0,$$

(ii) $x_n \xrightarrow{ul} x$ *with x continuous, iff* $x_t^n \to x_t$ *for all t in a dense set* $T \subset \mathsf{R}_+$ *and*

$$\lim_{h\to 0} \limsup_{n\to\infty} w_t(x_n, h) = 0, \quad t > 0,$$

(iii) *for continuous x,*

$$x_n \xrightarrow{s} x \quad \Leftrightarrow \quad x_n \xrightarrow{ul} x.$$

Proof: (i) If $x^n \xrightarrow{s} x$, then $x_t^n \to x_t$ for all t with $x_t = x_{t-}$, and the displayed condition holds by Theorem A5.4. Conversely, assume the stated conditions, and conclude from the same theorem that (x_n) is relatively compact in $D_{\mathsf{R}_+,S}$. Then for any sub-sequence $N' \subset \mathsf{N}$, we have $x^n \xrightarrow{s}$ some y along a further sub-sequence N''. To see that $x = y$, fix any $t > 0$ with $y_t = y_{t-}$, and choose $t_k \downarrow t$ in T with $x_{t_k} = x_{t_k-}$. Then

$$\rho(x_t, y_t) \leq \rho(x_t, x_{t_k}) + \rho\Big(x_{t_k}, x_{t_k}^n\Big) + \rho\Big(x_{t_k}^n, y_t\Big),$$

which tends to 0 as $n \to \infty$ along N'' and then $k \to \infty$, by the convergence $x^n \xrightarrow{s} y$, the point-wise convergence $x^n \to x$ on T, and the right continuity of x. Hence, $x_t = y_t$ on the continuity set of y. Since the latter set is dense in R_+ and both functions are right-continuous, we obtain $x = y$, and so $x^n \xrightarrow{s} x$ along N''. This extends to N, since N' was arbitrary.

(iii) The claim '$\Leftarrow$' is obvious. Now let $x_n \xrightarrow{s} x$, so that $\lambda_n \xrightarrow{ul} \iota$ and $x_n \circ \lambda_n \xrightarrow{ul} x$ for some λ_n. Since x is continuous, we have also $x \circ \lambda_n \xrightarrow{ul} x$, and so by combination $\rho(x_n \circ \lambda_n,\, x \circ \lambda_n) \xrightarrow{ul} 0$, which implies $x_n \xrightarrow{ul} x$.

(ii) Writing $J_t(x) = \sup_{s \le t} \rho(x_s, x_{s-})$, we note that for any $t, h > 0$,

$$\begin{aligned} \tilde{w}_t(x,h) \vee J_t(x) &\le w_t(x,h) \\ &\le 2\,\tilde{w}_t(x,h) + J_t(x). \end{aligned} \tag{12}$$

Now let $x^n \xrightarrow{ul} x$ with x continuous. Then $x_t^n \to x_t$ and $J_t(x^n) \to 0$ for all $t > 0$, and so the stated conditions follow by (i) and (12). Conversely, assume the two conditions. Then by (12) and (i) we have $x^n \xrightarrow{s} x$, and (12) also shows that $J_t(x^n) \to 0$, which yields the continuity of x. By (iii) the convergence may then be strengthened to $x^n \xrightarrow{ul} x$. □

Proof of Theorem 23.9: (i) Let $X_n \xrightarrow{sd} X$. Then Theorem 5.27 yields $X_n \xrightarrow{fd} X$ on $\{t \ge 0;\ X_t = X_{t-} \text{ a.s.}\}$. Furthermore, (X_n) is tight in $D_{\mathsf{R}_+,S}$ by Theorem 23.2, and (11) follows by Theorem 23.8 (ii).

Conversely, assume the stated conditions. Then $(X_{n,t})$ is tight in S for every $t \in T$ by Theorem 23.2, and so (X_n) is tight in $D_{\mathsf{R}_+,S}$ by Lemma 23.8. Assuming T to be countable and writing $X_{n,T} = \{X_{n,t};\ t \in T\}$, we see that even $(X_n, X_{n,T})$ is tight in $D_{\mathsf{R}_+,S} \times S^T$. By Theorem 23.2 the sequence $(X_n, X_{n,T})$ is then relatively compact in distribution, and so for any sub-sequence $N' \subset \mathsf{N}$ we have convergence $(X_n, X_{n,T}) \xrightarrow{sd} (Y, Z)$ along a further sub-sequence N'', for an S-valued process (Y, Z) on $\mathsf{R}_+ \cup T$. Since also $X_{n,T} \xrightarrow{d} X_T$, we may take $Z = X_T$. By Theorem 5.31, we may choose $X'_n \overset{d}{=} X_n$ such that a.s. $X'_n \xrightarrow{s} Y$ and $X'_{n,t} \to X_t$ along N'' for all $t \in T$. Hence, Lemma 23.10 (i) yields $X'_n \xrightarrow{s} X$ a.s., and so $X_n \xrightarrow{sd} X$ along N'', which extends to N since N' was arbitrary.

(iii) Let $X_n \xrightarrow{sd} X$ for a continuous X. By Theorem 5.31 we may choose $X'_n \overset{d}{=} X_n$ with $X'_n \xrightarrow{s} X$ a.s. Then Lemma 23.10 (iii) yields $X'_n \xrightarrow{ul} X$ a.s., which implies $X_n \xrightarrow{uld} X$. The converse claim is again obvious.

(ii) Use (iii) and Theorem 5.31 to reduce to Lemma 23.10 (ii). □

Tightness in $D_{\mathsf{R}_+,S}$ is often verified most easily by means of the following sufficient condition. Given a process X, we say that a random time is X-*optional* if it is optional with respect to the filtration induced by X.

Theorem 23.11 *(optional equi-continuity and tightness, Aldous) Let* X^1, X^2, *… be rcll processes in a metric space* (S, ρ). *Then* (11) *holds, if for any bounded* X^n*-optional times* τ_n *and positive constants* $h_n \to 0$,

$$\rho\Big(X^n_{\tau_n}, X^n_{\tau_n + h_n}\Big) \xrightarrow{P} 0, \quad n \to \infty. \tag{13}$$

Our proof will be based on two lemmas, beginning with a restatement of condition (13).

Lemma 23.12 *(equi-continuity) The condition in Theorem* 23.11 *holds iff*

$$\lim_{h\to 0}\limsup_{n\to\infty}\sup_{\sigma,\tau} E\Big\{\rho(X^n_\sigma, X^n_\tau)\wedge 1\Big\} = 0, \quad t>0, \tag{14}$$

where the supremum extends over all X^n-optional times $\sigma,\tau\le t$ with $\sigma\le\tau\le\sigma+h$.

Proof: Replacing ρ by $\rho\wedge 1$ if necessary, we may assume that $\rho\le 1$. The condition in Theorem 23.11 is then equivalent to

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\tau\le t}\sup_{h\in[0,\delta]} E\rho(X^n_\tau, X^n_{\tau+h}) = 0, \quad t>0,$$

where the outer supremum extends over all X^n-optional times $\tau\le t$. To deduce (14), suppose that $0\le\tau-\sigma\le\delta$. Then $[\tau,\tau+\delta]\subset[\sigma,\sigma+2\delta]$, and so by the triangle inequality and a simple substitution,

$$\begin{aligned}\delta\,\rho(X_\sigma,X_\tau) &\le \int_0^\delta\Big\{\rho(X_\sigma,X_{\tau+h})+\rho(X_\tau,X_{\tau+h})\Big\}\,dh\\ &\le \int_0^{2\delta}\rho(X_\sigma,X_{\sigma+h})\,dh+\int_0^\delta\rho(X_\tau,X_{\tau+h})\,dh.\end{aligned}$$

Thus,

$$\sup_{\sigma,\tau} E\rho(X_\sigma,X_\tau)\le 3\sup_\tau\sup_{h\in[0,2\delta]} E\rho(X_\tau,X_{\tau+h}),$$

where the suprema extend over all optional times $\tau\le t$ and $\sigma\in[\tau-\delta,\tau]$. □

We also need the following elementary estimate.

Lemma 23.13 *(exponential bound) For any random variables $\xi_1,\dots,\xi_n\ge 0$ with sum X_n, we have*

$$E\,e^{-X_n}\le e^{-nc}+\max_{k\le n}P\{\xi_k<c\}, \quad c>0.$$

Proof: Let p be the maximum on the right. By the Hölder and Chebyshev inequalities,

$$\begin{aligned}Ee^{-X_n} &= E\prod\nolimits_k e^{-\xi_k}\\ &\le \prod\nolimits_k\,(Ee^{-n\xi_k})^{1/n}\\ &\le \Big\{(e^{-nc}+p)^{1/n}\Big\}^n\\ &= e^{-nc}+p.\end{aligned}$$

□

Proof of Theorem 23.11: Again we may assume that $\rho\le 1$, and by suitable approximation we may extend condition (14) to weakly optional times σ and τ. For all $n\in\mathsf{N}$ and $\varepsilon>0$, we define recursively the weakly X^n-optional times

$$\sigma^n_{k+1}=\inf\Big\{s>\sigma^n_k;\ \rho(X^n_{\sigma^n_k},X^n_s)>\varepsilon\Big\}, \quad k\in\mathsf{Z}_+,$$

starting with $\sigma^n_0=0$, and note that for $m\in\mathsf{N}$ and $t,h>0$,

$$\tilde w_t(X^n,h)\le 2\varepsilon+\sum_{k<m}1\Big\{\sigma^n_{k+1}-\sigma^n_k<h,\ \sigma^n_k<t\Big\}+1\{\sigma^n_m<t\}. \tag{15}$$

Writing $\nu_n(t,h)$ for the supremum in (14), we get by Chebyshev's inequality and a simple truncation

$$P\Big\{\sigma^n_{k+1} - \sigma^n_k < h,\ \sigma^n_k < t\Big\} \le \varepsilon^{-1}\nu_n(t+h,\,h), \quad k \in \mathsf{N},\ t,h>0, \tag{16}$$

and so by (14) and (15),

$$\lim_{h\to 0}\ \limsup_{n\to\infty} E\tilde{w}_t(X^n,h) \le 2\varepsilon + \limsup_{n\to\infty} P\{\sigma^n_m < t\}. \tag{17}$$

Next we see from (16) and Lemma 23.13 that, for any $c>0$,

$$\begin{aligned} P\Big\{\sigma^n_m < t\Big\} &\le e^t E\Big(e^{-\sigma^n_m};\,\sigma^n_m < t\Big) \\ &\le e^t\Big\{e^{-mc} + \varepsilon^{-1}\nu_n(t+c,c)\Big\}. \end{aligned}$$

By (14) the right-hand side tends to 0 as $m,n\to\infty$ and then $c\to 0$. Hence, the last term in (17) tends to 0 as $m\to\infty$, and (11) follows since ε is arbitrary. □

To illustrate the use of Theorem 23.11, we prove an extension of Theorem 23.6. A related result was obtained by different methods in Corollary 16.17. An extension to Markov chains appears in Theorem 17.28.

Theorem 23.14 *(approximation of random walks, Skorohod) Let $Y^1, Y^2, \ldots$ be random walks in R^d, and form the rcll processes*

$$X^n_t = Y^n_{[m_n t]}, \quad t\ge 0,\ n\in\mathsf{N},$$

where $m_n\to\infty$. Then for any Lévy process X in R^d, we have

$$X^n_1 \xrightarrow{d} X_1 \quad \Leftrightarrow \quad X^n \xrightarrow{sd} X \ \ in\ \ D_{\mathsf{R}_+,\mathsf{R}^d}.$$

When X is a.s. continuous, this holds iff $X^n \xrightarrow{uld} X$.

Proof: By Corollary 7.9 we have $X^n \xrightarrow{fd} X$, and so by Theorem 23.11 it is enough to show that $|X^n_{\tau_n+h_n} - X^n_{\tau_n}| \xrightarrow{P} 0$ for any optional times $\tau_n<\infty$ and constants $h_n\to 0$. By the strong Markov property of Y^n, or alternatively by Theorem 27.7, we may reduce to the case where $\tau_n = 0$ for all n. Thus, it suffices to show that $X^n_{h_n} \xrightarrow{P} 0$ as $h_n\to 0$, which again can be seen from Corollary 7.9. □

Next we consider convergence in distribution of *random measures* on a separable, complete metric space S, defined as locally finite kernels $\xi\colon \Omega\to S$. Thus, $\xi(\omega,B)$ is a measurable function of $\omega\in\Omega$ for fixed B and a measure in $B\in\mathcal{S}$ for fixed ω, such that $\xi B<\infty$ a.s. for all B in the ring $\hat{\mathcal{S}}\subset\mathcal{S}$ of bounded Borel sets. Equivalently, we may regard random measures as random elements in the space $\mathcal{M}_S$ of locally finite measures μ on S, endowed with the σ-field generated by all *evaluation maps* $\pi_B\colon \mu\mapsto\mu B$ with $B\in\mathcal{S}$.

On $\mathcal{M}_S$ we introduce the *vague topology*, induced by the *integration maps* $\pi_f : \mu \mapsto \mu f$, for all f in the space $\hat{C}_S$ of bounded, continuous functions $f \geq 0$ with bounded support[5]. The corresponding notion of *vague convergence* $\mu_n \xrightarrow{v} \mu$ is then given by $\mu_n f \to \mu f$ for all $f \in \hat{C}_S$. For bounded measures μ on S, we also consider the *weak topology* with associated *weak convergence* $\mu_n \xrightarrow{w} \mu$, given by $\mu_n f \to \mu f$ for all bounded, continuous functions $f \geq 0$ on S. The basic properties of the vague and weak topologies are summarized in Lemma A5.5, and associated compactness criteria are given in Theorem A5.6. Using the latter, we may derive corresponding criteria for tightness of random measures:

Theorem 23.15 *(tightness of random measures, Matthes et al.) Let Ξ be a set of random measures on a separable, complete metric space S. Then Ξ is vaguely tight iff*

(i) $\lim_{r\to\infty} \sup_{\xi\in\Xi} P\{\xi B > r\} = 0, \quad B \in \hat{\mathcal{S}},$

(ii) $\inf_{K\in\mathcal{K}} \sup_{\xi\in\Xi} E\Big\{\xi(B\setminus K)\wedge 1\Big\} = 0, \quad B \in \hat{\mathcal{S}}.$

When the $\xi \in \Xi$ are a.s. bounded, the set Ξ is weakly tight iff (i)$-$(ii) *hold with $B = S$.*

Proof: Let Ξ be vaguely tight. Then for any $\varepsilon > 0$, there exists a vaguely compact set $A \subset \mathcal{M}_S$, such that $P\{\xi \notin A\} < \varepsilon$ for all $\xi \in \Xi$. For fixed $B \in \hat{\mathcal{S}}$, Theorem A5.6 yields $r = \sup_{\mu\in A} \mu B < \infty$, and so $\sup_{\xi\in\Xi} P\{\xi B > r\} < \varepsilon$. The same lemma yields a $K \in \mathcal{K}$ with $\sup_{\mu\in A} \mu(B\setminus K) < \varepsilon$, which implies $\sup_{\xi\in\Xi} E\{\xi(B\setminus K)\wedge 1\} \leq 2\varepsilon$. Since ε was arbitrary, this proves the necessity of (i) and (ii).

Now assume (i)$-$(ii). Fix any $s_0 \in S$, and put $B_n = \{s \in S;\, \rho(s, s_0) < n\}$. For any $\varepsilon > 0$, we may choose some constants $r_1, r_2, \ldots > 0$ and sets $K_1, K_2, \ldots \in \mathcal{K}$, such that for all $\xi \in \Xi$ and $n \in \mathsf{N}$,

$$\begin{aligned} P\{\xi B_n > r_n\} &< 2^{-n}\varepsilon, \\ E\Big\{\xi(B_n\setminus K_n)\wedge 1\Big\} &< 2^{-2n}\varepsilon. \end{aligned} \tag{18}$$

Writing A for the set of measures $\mu \in \mathcal{M}_S$ with

$$\mu B_n \leq r_n, \quad \mu(B_n\setminus K_n) \leq 2^{-n}, \qquad n \in \mathsf{N},$$

we see from Theorem A5.6 that A is vaguely relatively compact. Noting that

$$P\Big\{\xi(B_n\setminus K_n) > 2^{-n}\Big\} \leq 2^n E\Big\{\xi(B_n\setminus K_n)\wedge 1\Big\}, \quad n \in \mathsf{N},$$

we get from (18) for any $\xi \in \Xi$

$$P\{\xi \notin A\} \leq \textstyle\sum_n P\{\xi B_n > r_n\} + \sum_n 2^n E\Big\{\xi(B_n\setminus K_n)\wedge 1\Big\} < 2\varepsilon,$$

[5]Thus, we assume the supports of f to be *metrically* bounded. Requiring compact supports yields a weaker topology.

and since ε was arbitrary, we conclude that Ξ is tight. □

We may now derive criteria for convergence in distribution with respect to the vague topology, written as $\xi_n \xrightarrow{vd} \xi$ and defined[6] by $\mathcal{L}(\xi_n) \xrightarrow{vw} \mathcal{L}(\xi)$, where $\xrightarrow{vw}$ denotes weak convergence on $\hat{\mathcal{M}}_{\mathcal{M}_S}$ with respect to the vague topology on $\mathcal{M}_S$. Let $\hat{\mathcal{S}}_\xi$ be the class of sets $B \in \hat{\mathcal{S}}$ with $\xi\partial B = 0$ a.s. A non-empty class $\mathcal{I}$ of sets $I \subset S$ is called a *semi-ring*, if it is closed under finite intersections, and such that every proper difference in $\mathcal{I}$ is a finite union of disjoint $\mathcal{I}$-sets; it is called a *ring* if it is also closed under finite unions. Typical semi-rings in R^d are families of rectangular boxes $I_1 \times \cdots \times I_d$. By $\hat{\mathcal{I}}_+$ we denote the class of simple functions over $\mathcal{I}$.

Theorem 23.16 *(convergence of random measures, Harris, Matthes et al.) Let $\xi, \xi_1, \xi_2, \ldots$ be random measures on a separable, complete metric space S, and fix a dissecting semi-ring $\mathcal{I} \subset \hat{\mathcal{S}}_\xi$. Then these conditions are equivalent:*

(i) $\xi_n \xrightarrow{vd} \xi$,

(ii) $\xi_n f \xrightarrow{d} \xi f$ *for all* $f \in \hat{C}_S$ *or* $\hat{\mathcal{I}}_+$,

(iii) $Ee^{-\xi_n f} \to Ee^{-\xi f}$ *for all* $f \in \hat{C}_S$ *or* $\hat{\mathcal{I}}_+$ *with* $f \le 1$.

Our proof relies on two simple lemmas. For any function $f\colon S \to \mathsf{R}$, let D_f be the set of discontinuity points of f.

Lemma 23.17 *(limits of integrals) Let $\xi, \xi_1, \xi_2, \ldots$ be random measures on S with $\xi_n \xrightarrow{vd} \xi$, and let $f \in \hat{\mathcal{S}}_+$ with $\xi D_f = 0$ a.s. Then $\xi_n f \xrightarrow{d} \xi f$.*

Proof: By Theorem 5.27 it is enough to prove that, if $\mu_n \xrightarrow{v} \mu$ with $\mu D_f = 0$, then $\mu_n f \to \mu f$. By a suitable truncation and normalization, we may take all μ_n and μ to be probability measures. Then the same theorem yields $\mu_n \circ f^{-1} \xrightarrow{w} \mu \circ f^{-1}$, which implies $\mu_n f \to \mu f$ since f is bounded. □

Lemma 23.18 *(continuity sets) Let $\xi, \eta, \xi_1, \xi_2, \ldots$ be random measures on S with $\xi_n \xrightarrow{vd} \eta$, and fix a dissecting ring $\mathcal{U}$ and a constant $r > 0$. Then*

(i) $\hat{\mathcal{S}}_\eta \supset \hat{\mathcal{S}}_\xi$ *whenever*

$$\liminf_{n\to\infty} E\, e^{-r\,\xi_n U} \ge E\, e^{-r\,\xi U}, \quad U \in \mathcal{U},$$

(ii) *for point processes $\xi, \xi_1, \xi_2, \ldots$, we may assume instead that*

$$\liminf_{n\to\infty} P\{\xi_n U = 0\} \ge P\{\xi U = 0\}, \quad U \in \mathcal{U}.$$

[6]This mode of convergence involves the topologies on all three spaces S, $\mathcal{M}_S$, $\hat{\mathcal{M}}_{\mathcal{M}_S}$, which must not be confused.

Proof: Fix any $B \in \hat{\mathcal{S}}_\xi$. Let $\partial B \subset F \subset U$ with $F \in \hat{\mathcal{S}}_\eta$ and $U \in \mathcal{U}$, where F is closed. By (i) and Lemma 23.17,

$$\begin{aligned} Ee^{-r\,\eta\partial B} &\geq Ee^{-r\,\eta F} \\ &= \lim_{n\to\infty} Ee^{-r\,\xi_n F} \\ &\geq \liminf_{n\to\infty} Ee^{-r\,\xi_n U} \\ &\geq Ee^{-r\,\xi U}. \end{aligned}$$

Letting $U \downarrow F$ and then $F \downarrow \partial B$, we get $Ee^{-r\,\eta\partial B} \geq Ee^{-r\,\xi\partial B} = 1$ by dominated convergence. Hence, $\eta\partial B = 0$ a.s., which means that $B \in \hat{\mathcal{S}}_\eta$. The proof of (ii) is similar. □

Proof of Theorem 23.16: Here (i) ⇒ (ii) holds by Lemma 23.17 and (ii) ⇔ (iii) by Theorem 6.3. To prove (ii) ⇒ (i), assume that $\xi_n f \xrightarrow{d} \xi f$ for all $f \in \hat{C}_S$ or $\hat{\mathcal{I}}_+$, respectively. For any $B \in \hat{\mathcal{S}}$ we may choose an f with $1_B \leq f$, and so $\xi_n B \leq \xi_n f$ is tight by the convergence of $\xi_n f$. Thus, (ξ_n) satisfies (i) of Theorem 23.15.

To prove (ii) of the same result, we may assume that $\xi\partial B = 0$ a.s. A simple approximation yields $(1_B\xi_n)f \xrightarrow{d} (1_B\xi)f$ for all $f \in \hat{C}_S$ of $\hat{\mathcal{I}}_+$, and so we may take $S = B$, and let the measures ξ_n and ξ be a.s. bounded with $\xi_n f \xrightarrow{d} \xi f$ for all $f \in C_S$ or $\mathcal{I}_+$. Here Corollary 6.5 yields $(\xi_n f, \xi_n S) \xrightarrow{d} (\xi f, \xi S)$ for all such f, and so $(\xi_n S \vee 1)^{-1}\xi_n f \xrightarrow{d} (\xi S \vee 1)^{-1}\xi f$, which allows us to assume $\|\xi_n\| \leq 1$ for all n. Then $\xi_n f \xrightarrow{d} \xi f$ implies $E\xi_n f \to E\xi f$ for all f as above, and so $E\xi_n \xrightarrow{v} E\xi$ by the definition of vague convergence. Here Lemma A5.6 yields $\inf_{K\in\mathcal{K}} \sup_n E\xi_n K^c = 0$, which proves (ii) of Theorem 23.15 for the sequence (ξ_n).

The latter theorem shows that (ξ_n) is tight. If $\xi_n \xrightarrow{d} \eta$ along a sub-sequence $N' \subset \mathsf{N}$ for a random measure η on S, then Lemma 23.17 yields $\xi_n f \xrightarrow{d} \eta f$ for all $f \in \hat{C}_S$ or $\mathcal{I}_+$. In the former case, $\eta f \stackrel{d}{=} \xi f$ for all $f \in C_S$, which implies $\eta \stackrel{d}{=} \xi$. In the latter case, Lemma 23.18 yields $\hat{\mathcal{S}}_\eta \supset \hat{\mathcal{S}}_\xi \supset \mathcal{I}$, and so $\xi_n f \xrightarrow{d} \eta f$ for all $f \in \hat{\mathcal{I}}_+$, which again implies $\eta \stackrel{d}{=} \xi$. Since N' was arbitrary, we obtain $\xi_n \xrightarrow{vd} \xi$. □

Corollary 23.19 *(existence of limit) Let $\xi_1, \xi_2, \ldots$ be random measures on an lcscH space S, such that $\xi_n f \xrightarrow{d} \alpha_f$ for some random variables α_f, $f \in \hat{C}_S$. Then $\xi_n \xrightarrow{vd} \xi$ for a random measure ξ on S satisfying*

$$\xi f \stackrel{d}{=} \alpha_f, \quad f \in \hat{C}_S.$$

Proof: Condition (ii) of Theorem 23.15 being void, the tightness of (ξ_n) follows already from the tightness of $(\xi_n f)$ for all $f \in \hat{C}_S$. Hence, Theorem 23.2 yields $\xi_n \xrightarrow{vd} \xi$ along a sub-sequence $N' \subset \mathsf{N}$, for a random measure ξ on S, and so $\xi_n f \xrightarrow{d} \xi f$ along N' for every $f \in \hat{C}_S$. The latter convergence extends to N by hypothesis, and so $\xi_n \xrightarrow{vd} \xi$ by Theorem 23.16. □

Theorem 23.20 *(weak and vague convergence) Let $\xi, \xi_1, \xi_2, \ldots$ be a.s. bounded random measures on S with $\xi_n \xrightarrow{vd} \xi$. Then these conditions are equivalent:*

(i) $\xi_n \xrightarrow{wd} \xi$,

(ii) $\xi_n S \xrightarrow{d} \xi S$,

(iii) $\inf_{B \in \hat{\mathcal{S}}} \limsup_{n\to\infty} E\Big(\xi_n B^c \wedge 1\Big) = 0.$

Proof, (i) $\Rightarrow$ (ii): Clear since $1 \in C_S$.

(ii) $\Rightarrow$ (iii): If (iii) fails, then a diagonal argument yields a sub-sequence $N' \subset \mathsf{N}$ with

$$\inf_{B \in \hat{\mathcal{S}}} \liminf_{n \in N'} E\Big(\xi_n B^c \wedge 1\Big) > 0. \tag{19}$$

Under (ii) the sequences (ξ_n) and $(\xi_n S)$ are vaguely tight, whence so is the sequence of pairs $(\xi_n, \xi_n S)$, which yields the convergence $(\xi_n, \xi_n S) \xrightarrow{vd} (\tilde{\xi}, \alpha)$ along a further sub-sequence $N'' \subset N'$. Since clearly $\tilde{\xi} \stackrel{d}{=} \xi$, a transfer argument allows us to take $\tilde{\xi} = \xi$. Then for any $B \in \hat{\mathcal{S}}_\xi$, we get along N''

$$\begin{aligned} 0 \le \xi_n B^c &= \xi_n S - \xi_n B \\ &\xrightarrow{d} \alpha - \xi B, \end{aligned}$$

and so $\xi B \le \alpha$ a.s., which implies $\xi S \le \alpha$ a.s. since B was arbitrary. Since also $\alpha \stackrel{d}{=} \xi S$, we get

$$\begin{aligned} E\Big| e^{-\xi S} - e^{-\alpha}\Big| &= Ee^{-\xi S} - Ee^{-\alpha} \\ &= Ee^{-\xi S} - Ee^{-\xi S} = 0, \end{aligned}$$

proving that $\alpha = \xi S$ a.s. Hence, for any $B \in \hat{\mathcal{S}}_\xi$, we have along N''

$$\begin{aligned} E\Big(\xi_n B^c \wedge 1\Big) &= E\Big\{(\xi_n S - \xi_n B) \wedge 1\Big\} \\ &\to E\Big\{(\xi S - \xi B) \wedge 1\Big\} \\ &= E\Big(\xi B^c \wedge 1\Big), \end{aligned}$$

which tends to 0 as $B \uparrow S$ by dominated convergence. This contradicts (19), and (iii) follows.

(iii) $\Rightarrow$ (i): Writing $K^c \subset B^c \cup (B \setminus K)$ when $K \in \mathcal{K}$ and $B \in \hat{\mathcal{S}}$, and using the additivity of P and ξ_n and the sub-additivity of $x \wedge 1$ for $x \ge 0$, we get

$$\begin{aligned} &\limsup_{n\to\infty} E\Big(\xi_n K^c \wedge 1\Big) \\ &\qquad \le \limsup_{n\to\infty} \Big(E\Big\{\xi_n B^c \wedge 1\Big\} + E\Big\{\xi_n (B \setminus K) \wedge 1\Big\}\Big) \\ &\qquad \le \limsup_{n\to\infty} E\Big(\xi_n B^c \wedge 1\Big) + \limsup_{n\to\infty} E\Big\{\xi_n (B \setminus K) \wedge 1\Big\}. \end{aligned}$$

Taking the infimum over $K \in \mathcal{K}$ and then letting $B \uparrow S$, we get by Theorem 23.15 and (iii)

$$\inf_{K \in \mathcal{K}} \limsup_{n\to\infty} E\Big(\xi_n K^c \wedge 1\Big) = 0. \tag{20}$$

Similarly, we have for any $r > 1$

$$\begin{aligned}\limsup_{n\to\infty} P\Big\{\xi_n S > r\Big\} \\ &\le \limsup_{n\to\infty}\Big(P\Big\{\xi_n B > r-1\Big\} + P\Big\{\xi_n B^c > 1\Big\}\Big) \\ &\le \limsup_{n\to\infty} P\Big\{\xi_n B > r-1\Big\} + \limsup_{n\to\infty} E\Big(\xi_n B^c \wedge 1\Big).\end{aligned}$$

Letting $r \to \infty$ and then $B \uparrow S$, we get by Theorem 23.15 and (iii)

$$\lim_{r\to\infty}\liminf_{n\to\infty} P\Big\{\xi_n S > r\Big\} = 0. \tag{21}$$

By Theorem 23.15, we see from (20) and (21) that (ξ_n) is weakly tight. Assuming $\xi_n \xrightarrow{wd} \eta$ along a sub-sequence, we have also $\xi_n \xrightarrow{vd} \eta$, and so $\eta \overset{d}{=} \xi$. Since the sub-sequence was arbitrary, (i) follows by Theorem 23.2. □

For a first application, we prove some continuity properties of Cox processes and thinnings, extending the simple uniqueness properties in Lemma 15.7. Further limit theorems involving Cox processes and thinnings appear in Chapter 30.

Theorem 23.21 *(Cox and thinning continuity) For $n \in \mathbb{N}$ and a fixed $p \in (0,1]$, let ξ_n be a Cox process on S directed by a random measure η_n or a p-thinning of a point process η_n on S. Then*

$$\xi_n \xrightarrow{vd} \xi \quad \Leftrightarrow \quad \eta_n \xrightarrow{vd} \eta$$

for some ξ, η, where ξ is a Cox process directed by η or a p-thinning of η, respectively. In that case also

$$(\xi_n, \eta_n) \xrightarrow{vd} (\xi, \eta) \ \text{ in } \mathcal{M}_S^2.$$

Proof: Let $\eta_n \xrightarrow{vd} \eta$ for a random measure η on S. Then Lemma 15.2 yields for any $f, g \in \hat{C}_S$

$$\begin{aligned} E\, e^{-\xi_n f - \eta_n g} &= E \exp\Big\{-\eta_n(1 - e^{-f} + g)\Big\} \\ &\to E \exp\Big\{-\eta(1 - e^{-f} + g)\Big\} \\ &= E\, e^{-\xi f - \eta g},\end{aligned}$$

where ξ is a Cox process directed by η. Hence, Theorem 6.3 gives $\xi_n f + \eta_n g \xrightarrow{d} \xi f + \eta g$ for all such f and g, and so $(\xi_n, \eta_n) \xrightarrow{vd} (\xi, \eta)$ by Theorem 23.16.

Conversely, let $\xi_n \xrightarrow{vd} \xi$ for a point process ξ on S. Then the sequence (ξ_n) is vaguely tight, and so Theorem 23.15 yields for any $B \in \hat{\mathcal{S}}$ and $u > 0$

$$\lim_{r\to 0}\inf_{n\ge 1} E e^{-r\,\xi_n B} = \sup_{K\in\mathcal{K}}\inf_{n\ge 1} E \exp\Big\{-u\,\xi_n(B\setminus K)\Big\} = 1.$$

Similar relations hold for (η_n) by Lemma 15.2, with r and u replaced by $r' = 1 - e^{-r}$ and $u' = 1 - e^{-u}$, and so even (η_n) is vaguely tight by Theorem 23.15, which implies that (η_n) is vaguely relatively compact. If $\eta_n \xrightarrow{vd} \eta$

along a sub-sequence, then $\xi_n \xrightarrow{vd} \xi'$ as before, where ξ' is a Cox process directed by η. Since $\xi' \overset{d}{=} \xi$, the distribution of η is unique by Lemma 15.7, and so the convergence $\eta_n \xrightarrow{vd} \eta$ extends to the entire sequence. The proof for p-thinnings is similar. $\Box$

We also consider a stronger, non-topological version of distributional convergence for random measures, defined as follows. Beginning with non-random measures μ and $\mu_1, \mu_2, \ldots \in \mathcal{M}_S$, we write $\mu_n \xrightarrow{u} \mu$ for convergence in total variation and $\mu_n \xrightarrow{ul} \mu$ for the corresponding local version, defined by $1_B \mu_n \xrightarrow{u} 1_B \mu$ for all $B \in \hat{\mathcal{S}}$. For *random* measures ξ and ξ_n on S, we may now define the convergence $\xi_n \xrightarrow{ud} \xi$ by $\mathcal{L}(\xi_n) \xrightarrow{u} \mathcal{L}(\xi)$, and let $\xi_n \xrightarrow{uld} \xi$ be the corresponding local version, given by $1_B\, \xi_n \xrightarrow{ud} 1_B\, \xi$ for all $B \in \hat{\mathcal{S}}$. Finally, we define the convergence $\xi_n \xrightarrow{ulP} \xi$ by $\|\xi_n - \xi\|_B \xrightarrow{P} 0$ for all $B \in \hat{\mathcal{S}}$, and similarly for the global version $\xi_n \xrightarrow{uP} \xi$.

The last result has the following partial analogue for strong convergence.

Theorem 23.22 *(strong Cox continuity) Let ξ, $\xi_1, \xi_2, \ldots$ be Cox processes directed by some random measures η, $\eta_1, \eta_2, \ldots$ on S. Then*

$$\eta_n \xrightarrow{ulP} \eta \quad \Rightarrow \quad \xi_n \xrightarrow{uld} \xi,$$

with equivalence when η, $\eta_1, \eta_2, \ldots$ are non-random.

Proof: We may clearly take S to be bounded. First let $\eta = \lambda$ and all $\eta_n = \lambda_n$ be non-random. For any $n \in \mathbb{N}$, put

$$\hat{\lambda}_n = \lambda \wedge \lambda_n, \qquad \lambda'_n = \lambda - \hat{\lambda}_n, \qquad \lambda''_n = \lambda_n - \hat{\lambda}_n,$$

so that

$$\lambda = \hat{\lambda}_n + \lambda'_n, \qquad \lambda_n = \hat{\lambda}_n + \lambda''_n,$$
$$\|\lambda_n - \lambda\| = \|\lambda'_n\| + \|\lambda''_n\|.$$

Letting $\hat{\xi}_n$, ξ'_n, ξ''_n be independent Poisson processes with intensities $\hat{\lambda}_n$, λ'_n, λ''_n, respectively, we note that

$$\xi \overset{d}{=} \hat{\xi}_n + \xi'_n, \qquad \xi_n \overset{d}{=} \hat{\xi}_n + \xi''_n.$$

Assuming $\lambda_n \xrightarrow{u} \lambda$, we get

$$\begin{aligned}
\left\|\mathcal{L}(\xi) - \mathcal{L}(\xi_n)\right\| &\leq \left\|\mathcal{L}(\xi) - \mathcal{L}(\hat{\xi}_n)\right\| + \left\|\mathcal{L}(\xi_n) - \mathcal{L}(\hat{\xi}_n)\right\| \\
&\lesssim P\{\xi'_n \neq 0\} + P\{\xi''_n \neq 0\} \\
&= \left(1 - e^{-\|\lambda'_n\|}\right) + \left(1 - e^{-\|\lambda''_n\|}\right) \\
&\leq \|\lambda'_n\| + \|\lambda''_n\| \\
&= \|\lambda - \lambda_n\| \to 0,
\end{aligned}$$

which shows that $\xi_n \xrightarrow{ud} \xi$.

Conversely, let $\xi_n \xrightarrow{ud} \xi$. Then for any $B \in \mathcal{S}$ and $n \in \mathbb{N}$,

$$\begin{aligned} P\{\xi_n B = 0\} &\geq P\{\xi_n = 0\} \\ &\to P\{\xi = 0\} \\ &= e^{-\|\lambda\|} > 0. \end{aligned}$$

Since $(\log x)' = x^{-1} \lesssim 1$ on every interval $[\varepsilon, 1]$ with $\varepsilon > 0$, we get

$$\begin{aligned} \|\lambda - \lambda_n\| &\lesssim \sup_B |\lambda B - \lambda_n B| \\ &= \sup_B \Big| \log P\{\xi B = 0\} - \log P\{\xi_n B = 0\} \Big| \\ &\lesssim \sup_B \Big| P\{\xi B = 0\} - P\{\xi_n B = 0\} \Big| \\ &\leq \|\mathcal{L}(\xi) - \mathcal{L}(\xi_n)\| \to 0, \end{aligned}$$

which shows that $\lambda_n \xrightarrow{u} \lambda$.

Now let η and the η_n be random measures satisfying $\eta_n \xrightarrow{uP} \eta$. To obtain $\xi_n \xrightarrow{ud} \xi$, it is enough to show that, for any sub-sequence $N' \subset \mathbb{N}$, the desired convergence holds along a further sub-sequence N''. By Lemma 5.2 we may then assume that $\eta_n \xrightarrow{u} \eta$ a.s. Letting the processes ξ and ξ_n be conditionally independent and Poisson distributed with intensities η and η_n, respectively, we conclude from the previous case that

$$\Big\| \mathcal{L}(\xi_n \,|\, \eta_n) - \mathcal{L}(\xi \,|\, \eta) \Big\| \to 0 \quad \text{a.s.},$$

and so by dominated convergence,

$$\begin{aligned} \Big\| \mathcal{L}(\xi_n) - \mathcal{L}(\xi) \Big\| &= \Big\| E\Big\{ \mathcal{L}(\xi_n \,|\, \eta_n) - \mathcal{L}(\xi \,|\, \eta) \Big\} \Big\| \\ &= \sup_{|f| \leq 1} \Big| E\Big(E\{f(\xi_n) |\, \eta_n\} - E\{f(\xi) |\, \eta\} \Big) \Big| \\ &\leq \sup_{|f| \leq 1} E \Big| E\{f(\xi_n) |\, \eta_n\} - E\{f(\xi) |\, \eta\} \Big| \\ &\leq E \Big\| \mathcal{L}(\xi_n |\, \eta_n) - \mathcal{L}(\xi \,|\, \eta) \Big\| \to 0, \end{aligned}$$

which shows that $\xi_n \xrightarrow{ud} \xi$. □

For measure-valued processes X^n with rcll paths, we may characterize tightness in terms of the real-valued projections

$$X_t^n f = \int f(s)\, X_t^n(ds), \quad f \in \hat{C}^+.$$

Theorem 23.23 *(measure-valued processes) Let $X^1, X^2, \ldots$ be vaguely rcll processes in $\mathcal{M}_S$, where S is lcscH. Then these conditions are equivalent:*

(i) *(X^n) is vaguely tight in $D_{\mathbb{R}_+, \mathcal{M}_S}$,*

(ii) *$(X^n f)$ is tight in $D_{\mathbb{R}_+, \mathbb{R}_+}$ for every $f \in \hat{C}_S^+$.*

Proof: Let $(X^n f)$ be tight for every $f \in \hat{C}^+$, and fix any $\varepsilon > 0$. Let $f_1, f_2, \ldots$ be such as in Theorem A5.7, and choose some compact sets $B_1, B_2, \ldots \subset D_{\mathsf{R}_+,\mathsf{R}_+}$ with

$$P\Big\{X^n f_k \in B_k\Big\} \geq 1 - \varepsilon 2^{-k}, \quad k, n \in \mathsf{N}. \tag{22}$$

Then $A = \bigcap_k \{\mu;\, \mu f_k \in B_k\}$ is relatively compact in $D_{\mathsf{R}_+,\mathcal{M}_S}$, and (22) yields $P\{X^n \in A\} \geq 1 - \varepsilon$. □

Next we consider convergence of random sets. Given an lcscH space S, let $\mathcal{F}$, $\mathcal{G}$, $\mathcal{K}$ be the classes of closed, open, and compact subsets of S. We endow $\mathcal{F}$ with the *Fell topology*, generated by the sets

$$\Big\{F \in \mathcal{F};\, F \cap G \neq \emptyset\Big\}, \quad G \in \mathcal{G},$$
$$\Big\{F \in \mathcal{F};\, F \cap K = \emptyset\Big\}, \quad K \in \mathcal{K}.$$

Some basic properties of this topology are summarized in Theorem A6.1. In particular, $\mathcal{F}$ is compact and metrizable, and the set $\{F;\, F \cap B = \emptyset\}$ is universally measurable for every $B \in \hat{\mathcal{S}}$.

A *random closed set* in S is defined as a random element φ in $\mathcal{F}$. Write $\varphi \cap B = \varphi B$ for any $B \in \hat{\mathcal{S}}$, and note that the probabilities $P\{\varphi B = \emptyset\}$ are well defined. For any random closed set φ, we introduce the class

$$\hat{\mathcal{S}}_\varphi = \Big\{B \in \hat{\mathcal{S}};\, P\{\varphi B^o = \emptyset\} = P\{\varphi \bar{B} = \emptyset\}\Big\},$$

which is separating by Lemma A6.2. We may now state the basic convergence criterion for random sets.

Theorem 23.24 *(convergence of random sets, Norberg, OK) Let $\varphi, \varphi_1, \varphi_2, \ldots$ be random closed sets in an lcscH space S, and fix a separating class $\mathcal{U} \subset \hat{\mathcal{S}}_\varphi$. Then $\varphi_n \xrightarrow{d} \varphi$ iff*

$$P\{\varphi_n U = \emptyset\} \to P\{\varphi U = \emptyset\}, \quad U \in \mathcal{U}. \tag{23}$$

In particular, we may take $\mathcal{U} = \hat{\mathcal{S}}_\varphi$.

Proof: Write

$$h(B) = P\{\varphi B \neq \emptyset\},$$
$$h_n(B) = P\{\varphi_n B \neq \emptyset\}.$$

If $\varphi_n \xrightarrow{d} \varphi$, Theorem 5.25 yields

$$\begin{aligned} h(B^o) &\leq \liminf_{n\to\infty} h_n(B) \\ &\leq \limsup_{n\to\infty} h_n(B) \leq h(\bar{B}), \quad B \in \hat{\mathcal{S}}, \end{aligned}$$

and so for any $B \in \hat{\mathcal{S}}_\varphi$ we have $h_n(B) \to h(B)$.

Now consider a separating class $\mathcal{U}$ satisfying (23). Fix any $B \in \hat{\mathcal{S}}_\varphi$, and conclude from (23) that, for any $U, V \in \mathcal{U}$ with $U \subset B \subset V$,

$$\begin{aligned} h(U) &\le \liminf_{n\to\infty} h_n(B) \\ &\le \limsup_{n\to\infty} h_n(B) \le h(V). \end{aligned} \tag{24}$$

Since $\mathcal{U}$ is separating, we may choose some $U_k, V_k \in \mathcal{U}$ with $U_k \uparrow B^o$ and $\bar{V}_k \downarrow \bar{B}$. Here clearly

$$\begin{aligned} U_k \uparrow B^o &\Rightarrow \{\varphi U_k \neq \emptyset\} \uparrow \{\varphi B^o \neq \emptyset\} \\ &\Rightarrow h(U_k) \uparrow h(B^o) = h(B), \\ \bar{V}_k \downarrow \bar{B} &\Rightarrow \{\varphi \bar{V}_k \neq \emptyset\} \downarrow \{\varphi \bar{B} \neq \emptyset\} \\ &\Rightarrow h(\bar{V}_k) \downarrow h(\bar{B}) = h(B), \end{aligned}$$

where the finite-intersection property was used in the second line. In view of (24), we get $h_n(B) \to h(B)$, and (23) follows with $\mathcal{U} = \hat{\mathcal{S}}_\varphi$.

Since $\mathcal{F}$ is compact, the sequence (φ_n) is relatively compact by Theorem 23.2. For any sub-sequence $N' \subset \mathsf{N}$, we obtain $\varphi_n \xrightarrow{d} \psi$ along a further sub-sequence N'', for a random closed set ψ. Combining the direct statement with (23), we get

$$P\{\varphi B = \emptyset\} = P\{\psi B = \emptyset\}, \quad B \in \hat{\mathcal{S}}_\varphi \cap \hat{\mathcal{S}}_\psi. \tag{25}$$

Since $\hat{\mathcal{S}}_\varphi \cap \hat{\mathcal{S}}_\psi$ is separating by Lemma A6.2, we may approximate as before to extend (25) to any compact set B. The class of sets $\{F;\, F \cap K = \emptyset\}$ with compact K is clearly a π-system, and so a monotone-class argument yields $\varphi \overset{d}{=} \psi$. Since N' is arbitrary, we obtain $\varphi_n \xrightarrow{d} \varphi$ along N. □

Simple point processes allow the dual descriptions as integer-valued random measures or locally finite random sets. The corresponding notions of convergence are different, and we proceed to clarify how they are related. Since the mapping $\mu \mapsto \operatorname{supp} \mu$ is continuous on $\mathcal{N}_S$, the convergence $\xi_n \xrightarrow{vd} \xi$ implies $\operatorname{supp} \xi_n \xrightarrow{d} \operatorname{supp} \xi$. Conversely, if $E\xi$ and $E\xi_n$ are locally finite, then by Theorem 23.24 and Proposition 23.25, we have $\xi_n \xrightarrow{vd} \xi$ whenever $\operatorname{supp} \xi_n \xrightarrow{d} \operatorname{supp} \xi$ and $E\xi_n \xrightarrow{v} E\xi$. Here we give a precise criterion.

Theorem 23.25 *(convergence of point processes) Let $\xi, \xi_1, \xi_2, \ldots$ be point processes on S with ξ simple, and fix any dissecting ring $\mathcal{U} \subset \hat{\mathcal{S}}_\xi$ and semi-ring $\mathcal{I} \subset \mathcal{U}$. Then $\xi_n \xrightarrow{vd} \xi$ iff*

(i) $P\{\xi_n U = 0\} \to P\{\xi U = 0\}, \qquad U \in \mathcal{U}$,

(ii) $\limsup_{n\to\infty} P\{\xi_n I > 1\} \le P\{\xi I > 1\}, \quad I \in \mathcal{I}$.

Proof: The necessity holds by Lemma 23.17. Now assume (i)−(ii). We may choose $\mathcal{U}$ and $\mathcal{I}$ to be countable. Define $\eta(U) = \xi U \wedge 1$ and $\eta_n(U) = \xi_n U \wedge 1$

for $U \in \mathcal{U}$. Taking repeated differences in (i), we get $\eta_n \xrightarrow{d} \eta$ for the product topology on $\mathsf{R}_+^{\mathcal{U}}$. Hence, Theorem 5.31 yields some processes $\tilde\eta \overset{d}{=} \eta$ and $\tilde\eta_n \overset{d}{=} \eta_n$, such that a.s. $\tilde\eta_n(U) \to \tilde\eta(U)$ for all $U \in \mathcal{U}$. By a transfer argument, we may choose $\tilde\xi_n \overset{d}{=} \xi_n$ and $\tilde\xi \overset{d}{=} \xi$, such that a.s.

$$\begin{aligned}\tilde\eta_n(U) &= \tilde\xi_n U \wedge 1,\\ \tilde\eta(U) &= \tilde\xi U \wedge 1, \quad U \in \mathcal{U}.\end{aligned}$$

We may then assume that $\eta_n(U) \to \eta(U)$ for all $U \in \mathcal{U}$ a.s.

Now fix an $\omega \in \Omega$ with $\eta_n(U) \to \eta(U)$ for all $U \in \mathcal{U}$, and let $U \in \mathcal{U}$ be arbitrary. Splitting U into disjoint sets $I_1, \dots, I_m \in \mathcal{I}$ with $\xi I_k \le 1$ for all $k \le m$, we get

$$\begin{aligned}\eta_n(U) \to \eta(U) &\le \xi U\\ &= \textstyle\sum_j \xi I_j = \sum_j \eta(I_j)\\ &\leftarrow \textstyle\sum_j \eta_n(I_j)\\ &\le \textstyle\sum_j \xi_n I_j = \xi_n U,\end{aligned}$$

and so as $n \to \infty$, we have a.s.

$$\begin{aligned}\limsup_{n\to\infty} \big\{\xi_n U \wedge 1\big\} &\le \xi U\\ &\le \liminf_{n\to\infty} \xi_n U, \quad U \in \mathcal{U}.\end{aligned} \tag{26}$$

Next we note that, for any $h, k \in \mathsf{Z}_+$,

$$\begin{aligned}\big\{k \le h \le 1\big\}^c &= \{h > 1\} \cup \big\{h < k \wedge 2\big\}\\ &= \{k > 1\} \cup \big\{h = 0,\ k = 1\big\} \cup \big\{h > 1 \ge k\big\},\end{aligned}$$

where all unions are disjoint. Substituting $h = \xi I$ and $k = \xi_n I$, and using (ii) and (26), we get

$$\lim_{n\to\infty} P\big\{\xi I < \xi_n I \wedge 2\big\} = 0, \quad I \in \mathcal{I}. \tag{27}$$

Letting $U \in \mathcal{U}$ be a union of disjoint sets $I_1, \dots, I_m \in \mathcal{I}$, we get by (27)

$$\begin{aligned}\limsup_{n\to\infty} P\big\{\xi_n U > \xi U\big\} &\le \limsup_{n\to\infty} P \textstyle\bigcup_j \big\{\xi_n I_j > \xi I_j\big\}\\ &\le P \textstyle\bigcup_j \{\xi I_j > 1\}.\end{aligned}$$

By dominated convergence, we can make the right-hand side arbitrarily small, and so $P\{\xi_n U > \xi U\} \to 0$. Combining with (26) gives $P\{\xi_n U \ne \xi U\} \to 0$, which means that $\xi_n U \xrightarrow{P} \xi U$. Hence, $\xi_n f \xrightarrow{P} \xi f$ for every $f \in \mathcal{U}_+$, and Theorem 23.16 (i) yields $\xi_n \xrightarrow{vd} \xi$. □

Exercises

1. Show that the two versions of Donsker's theorem in Theorems 22.9 and 23.6 are equivalent.

2. Extend either version of Donsker's theorem to random vectors in R^d.

3. For any metric space (S,ρ), show that if $x^n \to x$ in $D_{\mathsf{R}_+,S}$ with x continuous, then $\sup_{s\le t}\rho(x_s^n,x_s)\to 0$ for every $t\ge 0$. (*Hint:* Note that x is uniformly continuous on every interval $[0,t]$.)

4. For any separable, complete metric space (S,ρ), show that if $X^n \xrightarrow{d} X$ in $D_{\mathsf{R}_+,S}$ with X continuous, we may choose $Y^n \overset{d}{=} X^n$ with $\sup_{s\le t}\rho(Y_s^n,X_s)\to 0$ a.s. for every $t\ge 0$. (*Hint:* Combine the preceding result with Theorems 5.31 and A5.4.)

5. Give an example where $x_n\to x$ and $y_n\to y$ in $D_{\mathsf{R}_+,\mathsf{R}}$, and yet $(x_n,y_n)\not\to(x,y)$ in $D_{\mathsf{R}_+,\mathsf{R}^2}$.

6. Let f be a continuous map between the metric spaces S,T, where T is separable, complete. Show that if $X^n\xrightarrow{d}X$ in $D_{\mathsf{R}_+,S}$, then $f(X^n)\xrightarrow{d}f(X)$ in $D_{\mathsf{R}_+,T}$. (*Hint:* By Theorem 5.27 it suffices to show that $x_n\to x$ in $D_{\mathsf{R}_+,S}$ implies $f(x_n)\to f(x)$ in $D_{\mathsf{R}_+,T}$. Since $A=\{x,x_1,x_2,\dots\}$ is relatively compact in $D_{\mathsf{R}_+,S}$, Theorem A5.4 shows that $U_t=\bigcup_{s\le t}\pi_sA$ is relatively compact in S for every $t>0$. Hence, f is uniformly continuous on each U_t.)

7. Show by an example that the condition in Theorem 23.11 is not necessary for tightness. (*Hint:* Consider non-random processes X_n.)

8. Show that in Theorem 23.11 it is enough to consider optional times taking finitely many values. (*Hint:* Approximate from the right and use the right-continuity of paths.)

9. Let the process X on R_+ be continuous in probability with values in a separable, complete metric space (S,ρ). Show that if $\rho(X_{\tau_n},X_{\tau_n+h_n})\xrightarrow{P}0$ for any bounded optional times τ_n and constants $h_n\to 0$, then X has an rcll version. (*Hint:* Approximate by suitable step processes, and use Theorems 23.9 and 23.11.)

10. Let $X,X^1,X^2,\dots$ be Lévy processes in R^d. Show that $X^n\xrightarrow{d}X$ in $D_{\mathsf{R}_+,\mathsf{R}^d}$ iff $X_1^n\xrightarrow{d}X_1$ in R^d. Compare with Theorem 16.14.

11. Examine how Theorem 23.15 simplifies when S is lcscH. (*Hint:* We may then omit condition (ii).)

12. Show that (ii)−(iii) of Theorem 23.16 remain sufficient if we replace $\hat{\mathcal{S}}_\xi$ by an arbitrary separating class. (*Hint:* Restate the conditions in terms of Laplace transforms, and extend to $\hat{\mathcal{S}}_\xi$ by a suitable approximation.)

13. Let η, $\eta_1,\eta_2,\dots$ be λ-randomizations of some point processes ξ, $\xi_1,\xi_2,\dots$ on a separable, complete space S. Show that $\xi_n\xrightarrow{d}\xi$ iff $\eta_n\xrightarrow{d}\eta$.

14. For random measures ξ, $\xi_1,\xi_2,\dots$ on S, show that $\xi_n\xrightarrow{uld}\xi$ implies $\xi_n\xrightarrow{vd}\xi$. Further show by an example that the converse implication may fail.

15. Let X and X^n be random elements in $D_{\mathsf{R}_+,\mathsf{R}^d}$ with $X^n\xrightarrow{fd}X$, and such that $uX^n\xrightarrow{d}uX$ in $D_{\mathsf{R}_+,\mathsf{R}}$ for every $u\in\mathsf{R}^d$. Show that $X^n\xrightarrow{d}X$. (*Hint:* Proceed as in Theorems 23.23 and A5.7.)

16. On an lcscH space S, let ξ, $\xi_1, \xi_2, \ldots$ be simple point processes with associated supports φ, $\varphi_1, \varphi_2, \ldots$. Show that $\xi_n \xrightarrow{vd} \xi$ implies $\varphi_n \xrightarrow{d} \varphi$. Further show by an example that the reverse implication may fail.

17. For an lcscH space S, let $\mathcal{U} \subset \hat{\mathcal{S}}$ be separating. Show that if $K \subset G$ with K compact and G open, there exists a $U \in \mathcal{U}$ with $K \subset U^o \subset \bar{U} \subset G$. (*Hint:* First choose $B, C \in \hat{\mathcal{S}}$ with $K \subset B^o \subset \bar{B} \subset C^o \subset \bar{C} \subset G$.)

Chapter 24

Large Deviations

Exponential rates, cumulant-generating function, Legendre–Fenchel transform, rate function, large deviations in R^d*, Schilder's theorem, regularization and uniqueness, large deviation principle, goodness, exponential tightness, weak and functional LDPs, continuous mapping, random sequences, exponential equivalence, perturbed dynamical systems, empirical distributions, relative entropy, functional law of the iterated logarithm*

In many contexts throughout probability theory and its applications, we need to estimate the asymptotic rate of decline of various tail probabilities. Since such problems might be expected to lead to tedious technical calculations, we may be surprised and delighted to learn about the existence of a powerful general theory, providing such asymptotic rates with great accuracy. Indeed, in many cases of interest, the asymptotic rate turns out to be exponential, and the associated rate function can often be obtained explicitly from the underlying distributions.

In its simplest setting, the theory of large deviation provides exponential rates of convergence in the weak law of large numbers. Here we consider some i.i.d. random variables $\xi_1, \xi_2, \ldots$ with mean m and *cumulant-generating function* $\Lambda(u) = \log Ee^{u\xi_i} < \infty$, and write $\bar{\xi}_n = n^{-1}\sum_{k\le n}\xi_k$. For any $x > m$, we show that the tail probabilities $P\{\bar{\xi}_n > x\}$ tend to 0 at an exponential rate $I(x)$, given by the *Legendre–Fenchel transform* Λ^* of Λ. In higher dimensions, it is convenient to state the result in the more general form

$$-n^{-1}\log P\{\bar{\xi}_n \in B\} \to I(B) = \inf_{x\in B} I(x),$$

where B is restricted to a suitable class of continuity sets. In this standard form of a *large-deviation principle with rate function* I, the result extends to a wide range of contexts throughout probability theory.

For a striking example of fundamental importance in statistical mechanics, we may mention *Sanov's theorem*, which provides similar large deviation bounds for the empirical distributions of a sequence of i.i.d. random variables with a common distribution μ. Here the rate function I, defined on the space of probability measures ν on R, agrees with the *relative entropy* $H(\nu|\,\mu)$. Another important case is *Schilder's theorem* for a family of rescaled Brownian motions in R^d, where the rate function becomes $I(x) = \frac{1}{2}\|\dot{x}\|_2^2$ —the squared norm in the Cameron–Martin space first encountered in Chapter 19. The latter result can be used to derive the powerful *Fredlin–Wentzell estimates* for randomly

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_25

perturbed dynamical systems. It also yields a short proof of *Strassen's law of the iterated logarithm*, a stunning extension of the classical Khinchin law from Chapter 14.

Modern proofs of those and other large deviation results rely on some general extension principles, which explains the wide applicability of the present ideas. In addition to some rather elementary and straightforward methods of continuity and approximation, we consider the more sophisticated and extremely powerful principles of *inverse continuous mapping* and *projective limits*, both of which play crucial roles in ensuing applications. We may also stress the significance of the notion of *exponential tightness*, and of the essential equivalence between the setwise and functional formulations of the large-deviation principle.

Large deviation theory is arguably one of the most technical areas of modern probability. For a beginning student, it is then essential not to get distracted by topological subtleties or elaborate computations. Many results are therefore stated under simplifying assumptions. Likewise, we postpone our discussion of the general principles, until the reader has become well acquainted with some basic ideas in a simple and concrete setting. For those reasons, important applications appear both at the beginning and at the end of the chapter, separated by some more abstract discussions of general notions and principles.

Returning to the elementary setting of i.i.d. random variables $\xi, \xi_1, \xi_2, \dots$, we write $S_n = \sum_{k\le n}\xi_k$ and $\bar\xi_n = S_n/n$. If $m = E\xi$ exists and is finite, then $P\{\bar\xi_n > x\} \to 0$ for all $x > m$ by the weak law of large numbers. Under stronger moment conditions, the rate of convergence turns out to be exponential and can be estimated with great accuracy. This elementary but quite technical observation, along with its multi-dimensional counterpart, lie at the core of large-deviation theory, and provide both a general pattern and a point of departure for the more advanced developments. For motivation, we begin with some simple observations.

Lemma 24.1 *(tail rate) For i.i.d. random variables* $\xi, \xi_1, \xi_2, \dots$,

(i) $n^{-1}\log P\{\bar\xi_n \ge x\} \to \sup_n n^{-1}\log P\{\bar\xi_n \ge x\} \equiv -h(x),\ \ x \in \mathsf{R}$,

(ii) *h is $[0,\infty]$-valued, non-decreasing, and convex,*

(iii) $h(x) < \infty \ \Leftrightarrow\ P\{\xi \ge x\} > 0$.

Proof: (i) Writing $p_n = P\{\bar\xi_n \ge x\}$, we get for any $m, n \in \mathsf{N}$

$$\begin{aligned} p_{m+n} &= P\Big\{S_{m+n} \ge (m+n)x\Big\} \\ &\ge P\Big\{S_m \ge mx,\ S_{m+n} - S_m \ge nx\Big\} = p_m\, p_n. \end{aligned}$$

Taking logarithms, we conclude that the sequence $-\log p_n$ is sub-additive, and the assertion follows by Lemma 25.19.

(ii) The first two assertions are obvious. To prove the convexity, let $x, y \in \mathsf{R}$ be arbitrary, and proceed as before to get

$$P\Big\{S_{2n} \geq n(x+y)\Big\} \geq P\{S_n \geq nx\}\, P\{S_n \geq ny\}.$$

Taking logarithms, dividing by $2n$, and letting $n \to \infty$, we obtain

$$h\Big\{\tfrac{1}{2}\,(x+y)\Big\} \leq \tfrac{1}{2}\Big\{h(x)+h(y)\Big\}, \quad x,y>0.$$

(iii) If $P\{\xi \geq x\} = 0$, then $P\{\bar{\xi}_n \geq x\} = 0$ for all n, and so $h(x) = \infty$. Conversely, (i) yields $\log P\{\xi \geq x\} \leq -h(x)$, and so $h(x) = \infty$ implies $P\{\xi \geq x\} = 0$. □

To identify the limit in Lemma 24.1, we need some further notation, here given for convenience directly in d dimensions. For any random vector ξ in R^d, we introduce the *cumulant-generating function*

$$\Lambda(u) = \Lambda_\xi(u) = \log E e^{u\xi}, \quad u \in \mathsf{R}^d. \tag{1}$$

Note that Λ is convex, since by Hölder's inequality we have for any $u, v \in \mathsf{R}^d$ and $p, q > 0$ with $p + q = 1$

$$\begin{aligned}\Lambda(pu+qv) &= \log E \exp\Big\{(pu+qv)\,\xi\Big\}\\ &\leq \log\Big\{(Ee^{u\xi})^p\,(Ee^{v\xi})^q\Big\}\\ &= p\,\Lambda(u) + q\,\Lambda(v).\end{aligned}$$

The surface $z = \Lambda(u)$ in R^{d+1} is determined by the class of supporting hyper-planes[1] of different slopes, and we note that the plane with slope $x \in \mathsf{R}^d$ (or normal vector $(1, -x)$) has equation

$$z + \Lambda^*(x) = xu, \quad u \in \mathsf{R}^d,$$

where Λ^* is the *Legendre–Fenchel transform* of Λ, given by

$$\Lambda^*(x) = \sup_{u\in\mathsf{R}^d}\Big\{ux - \Lambda(u)\Big\}, \quad x \in \mathsf{R}^d. \tag{2}$$

We can often compute Λ^* explicitly. Here we consider two simple cases, both needed below. The results may be proved by elementary calculus.

Lemma 24.2 *(Gaussian and Bernoulli distributions)*

(i) *When* $\xi = (\xi_1, \ldots, \xi_d)$ *is standard Gaussian in* R^d,

$$\Lambda_\xi^*(x) \equiv \tfrac{1}{2}\,|x|^2, \quad x \in \mathsf{R}^d,$$

(ii) *when* $\xi \in \{0,1\}$ *with* $P\{\xi = 1\} = p \in (0,1)$,

$$\Lambda_\xi^*(x) = \begin{cases} x\,\log\frac{x}{p} + (1-x)\,\log\frac{1-x}{1-p}, & x \in [0,1],\\ \infty, & x \notin [0,1].\end{cases}$$

[1] d-dimensional affine subspaces

The function Λ^* is again convex, since for any $x, y \in \mathsf{R}^d$ and for p, q as before,

$$\begin{aligned}\Lambda^*(px+qy) &= \sup\nolimits_u\Big(p\left\{ux - \Lambda(u)\right\} + q\left\{uy - \Lambda(u)\right\}\Big)\\ &\le p\,\sup\nolimits_u\Big\{ux - \Lambda(u)\Big\} + q\,\sup\nolimits_u\Big\{uy - \Lambda(u)\Big\}\\ &= p\,\Lambda^*(x) + q\,\Lambda^*(y).\end{aligned}$$

If $\Lambda < \infty$ near the origin, then $m = E\xi$ exists and agrees with the gradient $\nabla\Lambda(0)$. Thus, the surface $z = \Lambda(u)$ has tangent hyper-plane $z = mu$ at 0, and we conclude that $\Lambda^*(m) = 0$ and $\Lambda^*(x) > 0$ for $x \ne m$. If ξ is truly d-dimensional, then Λ is strictly convex at 0, and Λ^* is finite and continuous near m. For $d = 1$, we sometimes need the corresponding one-sided statements, which are easily derived by dominated convergence.

The following key result identifies the function h in Lemma 24.1. For simplicity, we assume that $m = E\xi$ exists in $[-\infty, \infty)$.

Theorem 24.3 *(rate function, Cramér, Chernoff) Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. random variables with $m = E\xi < \infty$. Then*

$$n^{-1}\log P\Big\{\bar\xi_n \ge x\Big\} \to -\Lambda^*(x), \quad x \ge m. \tag{3}$$

Proof: Using Chebyshev's inequality and (1), we get for any $u > 0$

$$\begin{aligned}P\Big\{\bar\xi_n \ge x\Big\} &= P\Big\{e^{uS_n} \ge e^{nux}\Big\}\\ &\le e^{-nux}Ee^{uS_n}\\ &= \exp\Big\{n\Lambda(u) - nux\Big\},\end{aligned}$$

and so

$$n^{-1}\log P\Big\{\bar\xi_n \ge x\Big\} \le \Lambda(u) - ux.$$

This remains true for $u \le 0$, since in that case $\Lambda(u) - ux \ge 0$ for $x \ge m$. Hence, by (2), we have the upper bound

$$n^{-1}\log P\Big\{\bar\xi_n \ge x\Big\} \le -\Lambda^*(x), \quad x \ge m,\ n \in \mathsf{N}. \tag{4}$$

To derive a matching lower bound, we first assume that $\Lambda < \infty$ on R_+. Then Λ is smooth on $(0, \infty)$ with $\Lambda'(0+) = m$ and $\Lambda'(\infty) = \operatorname{ess\,sup}\xi \equiv b$, and so for any $a \in (m, b)$ we can choose a $u > 0$ with $\Lambda'(u) = a$. Let $\eta, \eta_1, \eta_2, \ldots$ be i.i.d. with distribution

$$P\{\eta \in B\} = e^{-\Lambda(u)}E\Big(e^{u\xi};\, \xi \in B\Big), \quad B \in \mathcal{B}. \tag{5}$$

Then $\Lambda_\eta(r) = \Lambda_\xi(r+u) - \Lambda_\xi(u)$, and so $E\eta = \Lambda'_\eta(0) = \Lambda'_\xi(u) = a$. For any $\varepsilon > 0$, we get by (5)

$$\begin{aligned}P\Big\{\Big|\,\bar\xi_n - a\,\Big| < \varepsilon\Big\} &= e^{n\Lambda(u)}E\Big\{\exp(-nu\bar\eta_n);\, |\bar\eta_n - a| < \varepsilon\Big\}\\ &\ge \exp\Big\{n\Lambda(u) - nu(a+\varepsilon)\Big\}\,P\Big\{|\bar\eta_n - a| < \varepsilon\Big\}.\end{aligned} \tag{6}$$

Here the last probability tends to 1 by the law of large numbers, and so by (2)

$$\liminf_{n\to\infty} n^{-1}\log P\Big\{\Big|\,\bar\xi_n - a\,\Big| < \varepsilon\Big\} \geq \Lambda(u) - u(a+\varepsilon) \\ \geq -\Lambda^*(a+\varepsilon).$$

Fixing any $x \in (m,b)$ and putting $a = x+\varepsilon$, we get for small enough $\varepsilon > 0$

$$\liminf_{n\to\infty} n^{-1}\log P\Big\{\bar\xi_n \geq x\Big\} \geq -\Lambda^*(x+2\varepsilon).$$

Since Λ^* is continuous on (m,b) by convexity, we may let $\varepsilon \to 0$ and combine with (4) to obtain (3).

The result for $x > b$ is trivial, since in that case both sides of (3) equal $-\infty$. If instead $x = b < \infty$, then both sides equal $\log P\{\xi = b\}$, the left side by a simple computation and the right side by an elementary estimate. Finally, let $x = m > -\infty$. Since the statement is trivial when $\xi = m$ a.s., we may assume that $b > m$. For any $y \in (m,b)$, we have

$$\begin{aligned} 0 &\geq n^{-1}\log P\Big\{\bar\xi_n \geq m\Big\} \\ &\geq n^{-1}P\Big\{\bar\xi_n \geq y\Big\} \\ &\to -\Lambda^*(y) > -\infty. \end{aligned}$$

Here $\Lambda^*(y) \to \Lambda^*(m) = 0$ by continuity, and (3) follows for $x = m$. This completes the proof when $\Lambda < \infty$ on R_+.

The case where $\Lambda(u) = \infty$ for some $u > 0$ may be handled by truncation. Thus, for any $r > m$ we consider the random variables $\xi_k^r = \xi_k \wedge r$. Writing Λ_r and Λ_r^* for the associated functions Λ and Λ^*, we get for $x \geq m \geq E\xi^r$

$$\begin{aligned} n^{-1}\log P\Big\{\bar\xi_n \geq x\Big\} &\geq n^{-1}\log P\Big\{\bar\xi_n^r \geq x\Big\} \\ &\to -\Lambda_r^*(x). \end{aligned} \tag{7}$$

Now $\Lambda_r(u) \uparrow \Lambda(u)$ by monotone convergence as $r \to \infty$, and by Dini's theorem[2] the convergence is uniform on every compact interval where $\Lambda < \infty$. Since also Λ' is unbounded on the set where $\Lambda < \infty$, it follows easily that $\Lambda_r^*(x) \to \Lambda^*(x)$ for all $x \geq m$. The required lower bound is now immediate from (7). □

We may supplement Lemma 24.1 with a criterion for exponential decline of the tail probabilities $P\{\bar\xi_n \geq x\}$.

Corollary 24.4 *(exponential rate) Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. random variables with $m = E\xi < \infty$ and $b = \operatorname{ess\,sup}\xi$. Then*

(i) *for any $x \in (m,b)$, the probabilities $P\Big\{\bar\xi_n \geq x\Big\}$ decrease exponentially iff $\Lambda(\varepsilon) < \infty$ for some $\varepsilon > 0$,*

(ii) *the exponential decline extends to $x = b$ iff $P\{\xi = b\} \in (0,1)$.*

[2] When $f_n \uparrow f$ for some continuous functions f_n, f on a metric space, the convergence is uniform on compacts.

Proof: If $\Lambda(\varepsilon) < \infty$ for some $\varepsilon > 0$, then $\Lambda'(0+) = m$ by dominated convergence, and so $\Lambda^*(x) > 0$ for all $x > m$. If instead $\Lambda = \infty$ on $(0, \infty)$, then $\Lambda^*(x) = 0$ for all $x \geq m$. The statement for $x = b$ is trivial. □

The large deviation estimates in Theorem 24.3 are easily extended from intervals $[x, \infty)$ to arbitrary open or closed sets, which leads to a *large-deviation principle* for i.i.d. sequences in R. To fulfill our needs in subsequent applications and extensions, we consider a version of the same result in R^d. Motivated by the last result, and to avoid some technical complications, we assume that $\Lambda(u) < \infty$ for all u. Write B^o and $\bar{B} = B^-$ for the interior and closure of a set B.

Theorem 24.5 *(large deviations in R^d, Varadhan) Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. random vectors in R^d with $\Lambda = \Lambda_\xi < \infty$. Then for any $B \in \mathcal{B}^d$,*

$$\begin{aligned} -\inf_{x \in B^o} \Lambda^*(x) &\leq \liminf_{n \to \infty} n^{-1} \log P\Big\{\bar{\xi}_n \in B\Big\} \\ &\leq \limsup_{n \to \infty} n^{-1} \log P\Big\{\bar{\xi}_n \in B\Big\} \leq -\inf_{x \in \bar{B}} \Lambda^*(x). \end{aligned}$$

Proof: To derive the upper bound, fix any $\varepsilon > 0$. By (2) there exists for every $x \in \mathsf{R}^d$ some $u_x \in \mathsf{R}^d$, such that

$$u_x x - \Lambda(u_x) > \Big\{\Lambda^*(x) - \varepsilon\Big\} \wedge \varepsilon^{-1},$$

and by continuity we may choose an open ball B_x around x, such that

$$u_x y > \Lambda(u_x) + \Big\{\Lambda^*(x) - \varepsilon\Big\} \wedge \varepsilon^{-1}, \quad y \in B_x.$$

By Chebyshev's inequality and (1), we get for any $n \in \mathsf{N}$

$$\begin{aligned} P\Big\{\bar{\xi}_n \in B_x\Big\} &\leq E \exp\Big\{u_x S_n - n \inf\Big(u_x y;\, y \in B_x\Big)\Big\} \\ &\leq \exp\Big\{-n\Big(\{\Lambda^*(x) - \varepsilon\} \wedge \varepsilon^{-1}\Big)\Big\}. \end{aligned} \tag{8}$$

Further note that $\Lambda < \infty$ implies $\Lambda^*(x) \to \infty$ as $|x| \to \infty$, at least when $d = 1$. By Lemma 24.1 and Theorem 24.3, we may then choose $r > 0$ so large that

$$n^{-1} \log P\Big\{|\bar{\xi}_n| > r\Big\} \leq -1/\varepsilon, \quad n \in \mathsf{N}. \tag{9}$$

Now let $B \subset \mathsf{R}^d$ be closed. Then the set $\{x \in B;\, |x| \leq r\}$ is compact and may be covered by finitely many balls $B_{x_1}, \ldots, B_{x_m}$ with centers $x_i \in B$. By (8) and (9), we get for any $n \in \mathsf{N}$

$$\begin{aligned} P\Big\{\bar{\xi}_n \in B\Big\} &\leq \sum_{i \leq m} P\Big\{\bar{\xi}_n \in B_{x_i}\Big\} + P\Big\{|\bar{\xi}_n| > r\Big\} \\ &\leq \sum_{i \leq m} \exp\Big\{-n\Big(\{\Lambda^*(x_i) - \varepsilon\} \wedge \varepsilon^{-1}\Big)\Big\} + e^{-n/\varepsilon} \\ &\leq (m+1) \exp\Big\{-n\Big(\{\Lambda^*(B) - \varepsilon\} \wedge \varepsilon^{-1}\Big)\Big\}, \end{aligned}$$

where $\Lambda^*(B) = \inf_{x\in B} \Lambda^*(x)$. Hence,

$$\limsup_{n\to\infty} n^{-1} \log P\Big\{\bar{\xi}_n \in B\Big\} \le -\Big\{\Lambda^*(B) - \varepsilon\Big\} \wedge \varepsilon^{-1},$$

and the upper bound follows since ε was arbitrary.

Turning to the lower bound, we first assume that $\Lambda(u)/|u| \to \infty$ as $|u| \to \infty$. Fix any open set $B \subset \mathsf{R}^d$ and a point $x \in B$. By compactness and the smoothness of Λ, there exists a $u \in \mathsf{R}^d$ with $\nabla\Lambda(u) = x$. Let $\eta, \eta_1, \eta_2, \ldots$ be i.i.d. random vectors with distribution (5), and note as before that $E\eta = x$. For $\varepsilon > 0$ small enough, we get as in (6)

$$\begin{aligned} P\Big\{\bar{\xi}_n \in B\Big\} &\ge P\Big\{\Big|\,\bar{\xi}_n - x\,\Big| < \varepsilon\Big\} \\ &\ge \exp\Big\{n\Lambda(u) - nux - n\varepsilon|u|\Big\}\, P\Big\{|\bar{\eta}_n - x| < \varepsilon\Big\}. \end{aligned}$$

Hence, by the law of large numbers and (2),

$$\begin{aligned} \liminf_{n\to\infty} n^{-1} \log P\Big\{\bar{\xi}_n \in B\Big\} &\ge \Lambda(u) - ux - \varepsilon|u| \\ &\ge -\Lambda^*(x) - \varepsilon|u|. \end{aligned}$$

It remains to let $\varepsilon \to 0$ and take the supremum over $x \in B$.

To eliminate the growth condition on Λ, let $\zeta, \zeta_1, \zeta_2, \ldots$ be i.i.d. standard Gaussian random vectors independent of ξ and the ξ_n. Then for any $\sigma > 0$ and $u \in \mathsf{R}^d$, we have by Lemma 24.2 (i)

$$\begin{aligned} \Lambda_{\xi+\sigma\zeta}(u) &= \Lambda_\xi(u) + \Lambda_\zeta(\sigma u) \\ &= \Lambda_\xi(u) + \tfrac{1}{2}\,\sigma^2|u|^2 \\ &\ge \Lambda_\xi(u), \end{aligned}$$

and in particular $\Lambda^*_{\xi+\sigma\zeta} \le \Lambda^*_\xi$. Since also $\Lambda_{\xi+\sigma\zeta}(u)/|u| \ge \sigma^2|u|/2 \to \infty$, we note that the previous bound applies to $\bar{\xi}_n + \sigma\bar{\zeta}_n$. Now fix any $x \in B$ as before, and choose $\varepsilon > 0$ small enough that B contains a 2ε-ball around x. Then

$$\begin{aligned} P\Big\{\Big|\bar{\xi}_n + \sigma\bar{\zeta}_n - x\Big| < \varepsilon\Big\} &\le P\Big\{\bar{\xi}_n \in B\Big\} + P\Big\{\sigma|\bar{\zeta}_n| \ge \varepsilon\Big\} \\ &\le 2\left(P\Big\{\bar{\xi}_n \in B\Big\} \vee P\Big\{\sigma|\bar{\zeta}_n| \ge \varepsilon\Big\}\right). \end{aligned}$$

Applying the lower bound to the variables $\bar{\xi}_n + \sigma\bar{\zeta}_n$ and the upper bound to $\bar{\zeta}_n$, we get by Lemma 24.2 (i)

$$\begin{aligned} -\Lambda^*_\xi(x) &\le -\Lambda^*_{\xi+\sigma\zeta}(x) \\ &\le \liminf_{n\to\infty} n^{-1} \log P\Big\{\Big|\bar{\xi}_n + \sigma\bar{\zeta}_n - x\Big| < \varepsilon\Big\} \\ &\le \liminf_{n\to\infty} n^{-1} \log \left(P\Big\{\bar{\xi}_n \in B\Big\} \vee P\Big\{\sigma|\bar{\zeta}_n| \ge \varepsilon\Big\}\right) \\ &\le \liminf_{n\to\infty} n^{-1} \log P\Big\{\bar{\xi}_n \in B\Big\} \,\vee\, (-\varepsilon^2/2\sigma^2). \end{aligned}$$

The desired lower bound now follows, as we let $\sigma \to 0$ and then take the supremum over all $x \in B$. $\square$

We can also establish large-deviation results in function spaces. The following theorem is basic and sets the pattern for more complex results. For convenience, we may write $C = C_{[0,1],\mathsf{R}^d}$ and $C_0^k = \{x \in C^k;\, x_0 = 0\}$. We also introduce the *Cameron–Martin space* H_1, consisting of all absolutely continuous functions $x \in C_0$ with a Radon–Nikodym derivative $\dot{x} \in L^2$, so that $\|\dot{x}\|_2^2 = \int_0^1 |\dot{x}_t|^2 dt < \infty$.

Theorem 24.6 *(large deviations of Brownian motion, Schilder) Let X be a d-dimensional Brownian motion on $[0,1]$. Then for any Borel set $B \subset C_{[0,1],\mathsf{R}^d}$, we have*

$$-\inf_{x\in B^o} I(x) \le \liminf_{\varepsilon\to 0}\, \varepsilon^2 \log P\{\varepsilon X \in B\}$$
$$\le \limsup_{\varepsilon\to 0}\, \varepsilon^2 \log P\{\varepsilon X \in B\} \le -\inf_{x\in \bar{B}} I(x),$$

where

$$I(x) = \begin{cases} \frac{1}{2}\,\|\dot{x}\|_2^2, & x \in H_1, \\ \infty, & x \notin H_1. \end{cases}$$

The proof requires a simple topological fact.

Lemma 24.7 *(level sets) The level sets L_r below are compact in $C_{[0,1],\mathsf{R}^d}$:*

$$L_r = I^{-1}[0,r] = \left\{x \in H_1;\, \|\dot{x}\|_2^2 \le 2r\right\}, \quad r > 0.$$

Proof: Cauchy's inequality gives

$$|x_t - x_s| \le \int_s^t |\dot{x}_u|\, du$$
$$\le (t-s)^{1/2}\, \|\dot{x}\|_2, \quad 0 \le s < t \le 1,\ x \in H_1.$$

By the Arzelà–Ascoli Theorem A5.2, the set L_r is then relatively compact in C. It is also weakly compact in the Hilbert space H_1 with norm $\|x\| = \|\dot{x}\|_2$. Thus, every sequence $x_1, x_2, \ldots \in L_r$ has a sub-sequence that converges in both C and H_1, say with limits $x \in C$ and $y \in L_r$, respectively. For every $t \in [0,1]$, the sequence $x_n(t)$ then converges in R^d to both $x(t)$ and $y(t)$, and we get $x = y \in L_r$. □

Proof of Theorem 24.6: To establish the lower bound, fix any open set $B \subset C$. Since $I = \infty$ outside H_1, it suffices to prove that

$$-I(x) \le \liminf_{\varepsilon\to 0}\, \varepsilon^2 \log P\{\varepsilon X \in B\}, \quad x \in B \cap H_1. \tag{10}$$

Then note as in Lemma 1.37 that C_0^2 is dense in H_1, and also that $\|x\|_\infty \le \|\dot{x}\|_1 \le \|\dot{x}\|_2$ for any $x \in H_1$. Hence, for every $x \in B \cap H_1$, there exist some functions $x_n \in B \cap C_0^2$ with $I(x_n) \to I(x)$, and it suffices to prove (10) for $x \in B \cap C_0^2$.

For small enough $h > 0$, Theorem 19.23 yields

$$P\{\varepsilon X \in B\} \ge P\left\{\|\varepsilon X - x\|_\infty < h\right\}$$
$$= E\left(\mathcal{E}\left\{-(\dot{x}/\varepsilon)\cdot X\right\}_1;\, \|\varepsilon X\|_\infty < h\right). \tag{11}$$

Integrating by parts gives

$$\begin{aligned}\log \mathcal{E}\Big\{-(\dot{x}/\varepsilon)\cdot X\Big\}_1 &= -\varepsilon^{-1}\int_0^1 \dot{x}_t\, dX_t - \varepsilon^{-2}I(x)\\ &= -\varepsilon^{-1}\dot{x}_1 X_1 + \varepsilon^{-1}\int_0^1 \ddot{x}_t\, X_t\, dt - \varepsilon^{-2}I(x),\end{aligned}$$

and so by (11),

$$\varepsilon^2 \log P\{\varepsilon X \in B\} \geq -I(x) - h|\dot{x}_1| - h\|\ddot{x}\|_1 + \varepsilon^2 \log P\Big\{\|\varepsilon X\|_\infty < h\Big\}.$$

Now (10) follows as we let $\varepsilon \to 0$ and then $h \to 0$.

Turning to the upper bound, fix any closed set $B \subset C$, and let B_h be the closed h-neighborhood of B. Letting X_n be the n-segment, polygonal approximation of X with $X_n(k/n) = X(k/n)$ for $k \leq n$, we note that

$$P\{\varepsilon X \in B\} \leq P\Big\{\varepsilon X_n \in B_h\Big\} + P\Big\{\varepsilon\|X - X_n\| > h\Big\}. \tag{12}$$

Writing $I(B_h) = \inf\{I(x);\, x \in B_h\}$, we obtain

$$P\Big\{\varepsilon X_n \in B_h\Big\} \leq P\Big\{I(\varepsilon X_n) \geq I(B_h)\Big\}.$$

Here $2I(X_n)$ is a sum of nd variables ξ_{ik}^2, where the ξ_{ik} are i.i.d. $N(0,1)$, and so by Lemma 24.2 (i) and an interpolated version of Theorem 24.5,

$$\limsup_{\varepsilon\to 0}\, \varepsilon^2 \log P\Big\{\varepsilon X_n \in B_h\Big\} \leq -I(B_h). \tag{13}$$

Next, we get by Proposition 14.13 and some elementary estimates

$$\begin{aligned}P\Big\{\varepsilon\|X - X_n\| > h\Big\} &\leq n\, P\Big\{\varepsilon\|X\| > h\sqrt{n}/2\Big\}\\ &\leq 2\, nd\, P\Big\{\varepsilon^2\xi^2 > h^2 n/4d\Big\},\end{aligned}$$

where ξ is $N(0,1)$. Applying Theorem 24.5 and Lemma 24.2 (i) again, we obtain

$$\limsup_{\varepsilon\to 0}\, \varepsilon^2 \log P\Big\{\varepsilon\|X - X_n\| > h\Big\} \leq -h^2 n/8d. \tag{14}$$

Combining (12), (13), and (14) gives

$$\limsup_{\varepsilon\to 0}\, \varepsilon^2 \log P\{\varepsilon X \in B\} \leq -I(B_h) \wedge (h^2 n/8d),$$

and as $n \to \infty$ we obtain the upper bound $-I(B_h)$.

It remains to show that $I(B_h) \uparrow I(B)$ as $h \to 0$. Then fix any $r > \sup_h I(B_h)$. For every $h > 0$, we may choose an $x_h \in B_h$ with $I(x_h) \leq r$, and by Lemma 24.7 we may extract a convergent sequence $x_{h_n} \to x$ with $h_n \to 0$, such that even $I(x) \leq r$. Since also $x \in \bigcap_h B_h = B$, we obtain $I(B) \leq r$, as required. $\square$

The last two theorems suggest the following abstraction. Letting ξ_ε, $\varepsilon > 0$, be random elements in a metric space S with Borel σ-field $\mathcal{S}$, we say that the family (ξ_ε) satisfies a *large-deviation principle (LDP) with rate function* $I: S \to [0,\infty]$, if for any $B \in \mathcal{S}$,

$$\begin{aligned} -\inf_{x \in B^o} I(x) &\le \liminf_{\varepsilon \to 0} \varepsilon \log P\{\xi_\varepsilon \in B\} \\ &\le \limsup_{\varepsilon \to 0} \varepsilon \log P\{\xi_\varepsilon \in B\} \le -\inf_{x \in \bar{B}} I(x). \end{aligned} \tag{15}$$

For sequences $\xi_1, \xi_2, \ldots$, we require the same condition with the normalizing factor ε replaced by n^{-1}. It is often convenient to write $I(B) = \inf_{x \in B} I(x)$. Writing $\mathcal{S}_I$ for the class $\{B \in \mathcal{S};\, I(B^o) = I(\bar{B})\}$ of *I-continuity sets*, we note that (15) yields the convergence

$$\lim_{\varepsilon \to 0} \varepsilon \log P\{\xi_\varepsilon \in B\} = -I(B), \quad B \in \mathcal{S}_I. \tag{16}$$

If $\xi, \xi_1, \xi_2, \ldots$ are i.i.d. random vectors in R^d with $\Lambda(u) = Ee^{u\xi} < \infty$ for all u, then by Theorem 24.5 the averages $\bar{\xi}_n$ satisfy the LDP in R^d with rate function Λ^*. If instead X is a d-dimensional Brownian motion on $[0,1]$, then Theorem 24.6 shows that the processes $\varepsilon^{1/2} X$ satisfy the LDP in $C_{[0,1],\mathsf{R}^d}$, with rate function $I(x) = \frac{1}{2}\|\dot{x}\|_2^2$ for $x \in H_1$ and $I(x) = \infty$ otherwise.

We show that the rate function I is essentially unique.

Lemma 24.8 *(regularization and uniqueness) Let (ξ_ε) satisfy an LDP in a metric space S with rate function I. Then*

(i) *I can be chosen to be lower semi-continuous,*

(ii) *the choice of I in* (i) *is unique.*

Proof: (i) Assume (15) for some I. Then the function

$$J(x) = \liminf_{y \to x} I(y), \quad x \in S,$$

is clearly lower semi-continuous with $J \le I$. It is also easy to verify that $J(G) = I(G)$ for all open sets $G \subset S$. Thus, (15) remains true with I replaced by J.

(ii) Suppose that (15) holds for each of the lower semi-continuous functions I and J, and let $I(x) < J(x)$ for some $x \in S$. By the semi-continuity of J, we may choose a neighborhood G of x with $J(\bar{G}) > I(x)$. Applying (15) to both I and J, we get the contradiction

$$\begin{aligned} -I(x) &\le -I(G) \\ &\le \liminf_{\varepsilon \to 0} \varepsilon \log P\{\xi_\varepsilon \in G\} \\ &\le -J(\bar{G}) < -I(x). \end{aligned}$$ □

Justified by the last result, we may henceforth take the lower semi-continuity to be part of our definition of a rate function. (An arbitrary function I satisfying (15) will then be called a *raw* rate function.) No regularization is needed in Theorems 24.5 and 24.6, since the associated rate functions Λ^* and I are already lower semi-continuous, the former as the supremum of a family of continuous functions, the latter by Lemma 24.7.

It is sometimes useful to impose a slightly stronger regularity condition on the function I. Thus, we say that I is *good*, if the level sets $I^{-1}[0,r] = \{x \in S; I(x) \le r\}$ are compact (rather than just closed). Note that the infimum $I(B) = \inf_{x\in B} I(x)$ is then attained for every closed set $B \neq \emptyset$. The rate functions in Theorems 24.5 and 24.6 are clearly both good.

A related condition on the family (ξ_ε) is the *exponential tightness*

$$\inf_K \limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \notin K\Big\} = -\infty, \tag{17}$$

where the infimum extends over all compact sets $K \subset S$. We actually need only the slightly weaker condition of *sequential exponential tightness*, where (17) is only required along sequences $\varepsilon_n \to 0$. To simplify our exposition, we often omit the sequential qualification from our statements, and carry out the proofs under the stronger non-sequential hypothesis.

We finally say that (ξ_ε) satisfies the *weak LDP* with rate function I, if the lower bound in (15) holds as stated, while the upper bound is only required for compact sets B. We list some relations between the mentioned properties.

Lemma 24.9 *(goodness, exponential tightness, and weak LDP) Let ξ_ε, $\varepsilon > 0$, be random elements in a metric space S. Then*

(i) *if (ξ_ε) satisfies an LDP with rate function I, then* (16) *holds, and the two conditions are equivalent when I is good,*

(ii) *if the ξ_ε are exponentially tight and satisfy a weak LDP with rate function I, then I is good, and (ξ_ε) satisfies the full LDP,*

(iii) *if S is Polish and (ξ_ε) satisfies an LDP with rate function I, then I is good iff (ξ_ε) is sequentially exponentially tight.*

Proof: (i) Let I be good and satisfy (16). Write B^h for the closed h-neighborhood of $B \in \mathcal{S}$. Since I is non-increasing on $\mathcal{S}$, we have $B^h \notin \mathcal{S}_I$ for at most countably many $h > 0$. Hence, (16) yields for almost every $h > 0$

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log P\{\xi_\varepsilon \in B\} &\le \lim_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in B^h\Big\} \\ &= -I(B^h).\end{aligned}$$

To see that $I(B^h) \uparrow I(\bar{B})$ as $h \to 0$, suppose that instead $\sup_h I(B^h) < I(\bar{B})$. Since I is good, we may choose for every $h > 0$ some $x_h \in B^h$ with $I(x_h) = I(B^h)$, and then extract a convergent sequence $x_{h_n} \to x \in \bar{B}$ with $h_n \to 0$. Then the lower semi-continuity of I yields the contradiction

$$\begin{aligned}I(\bar{B}) \le I(x) &\le \liminf_{n\to\infty} I(x_{h_n}) \\ &\le \sup_{h>0} I(B^h) < I(\bar{B}),\end{aligned}$$

proving the upper bound. Next let $x \in B^o$, and conclude from (16) that, for almost all sufficiently small $h > 0$,

$$\begin{aligned}-I(x) &\le -I(\{x\}^h)\\ &= \lim_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in \{x\}^h\Big\}\\ &\le \liminf_{\varepsilon\to 0} \varepsilon \log P\{\xi_\varepsilon \in B\}.\end{aligned}$$

The lower bound now follows as we take the supremum over $x \in B^o$.

(ii) By (17), we may choose some compact sets K_r satisfying

$$\limsup_{\varepsilon\to 0} \varepsilon \log P\{\xi_\varepsilon \notin K_r\} < -r, \quad r > 0. \tag{18}$$

For any closed set $B \subset S$, we have

$$P\{\xi_\varepsilon \in B\} \le 2\Big(P\Big\{\xi_\varepsilon \in B \cap K_r\Big\} \vee P\Big\{\xi_\varepsilon \notin K_r\Big\}\Big), \quad r > 0,$$

and so, by the weak LDP and (18),

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log P\{\xi_\varepsilon \in B\} &\le -I(B \cap K_r) \wedge r\\ &\le -I(B) \wedge r.\end{aligned}$$

The upper bound now follows as we let $r \to \infty$. Applying the lower bound and (18) to the sets K_r^c gives

$$-I(K_r^c) \le \limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \notin K_r\Big\} < -r, \quad r > 0,$$

and so $I^{-1}[0, r] \subset K_r$ for all $r > 0$, which shows that I is good.

(iii) The sufficiency follows from (ii), applied to an arbitrary sequence $\varepsilon_n \to 0$. Now let S be separable and complete, and assume that the rate function I is good. For any $k \in \mathsf{N}$, we may cover S by some open balls $B_{k1}, B_{k2}, \ldots$ of radius $1/k$. Putting $U_{km} = \bigcup_{j\le m} B_{kj}$, we have $\sup_m I(U_{km}^c) = \infty$, since any level set $I^{-1}[0, r]$ is covered by finitely many sets B_{kj}. Now fix any sequence $\varepsilon_n \to 0$ and constant $r > 0$. By the LDP upper bound and the fact that $P\{\xi_{\varepsilon_n} \in U_{km}^c\} \to 0$ as $m \to \infty$ for fixed n and k, we may choose $m_k \in \mathsf{N}$ so large that

$$P\Big\{\xi_{\varepsilon_n} \in U_{k,m_k}^c\Big\} \le \exp\left(-rk/\varepsilon_n\right), \quad n, k \in \mathsf{N}.$$

Summing a geometric series, we obtain

$$\limsup_{n\to\infty} \varepsilon_n \log P\Big\{\xi_{\varepsilon_n} \in \textstyle\bigcup_k U_{k,m_k}^c\Big\} \le -r.$$

The asserted exponential tightness now follows, since the set $\bigcap_k U_{k,m_k}$ is totally bounded and hence relatively compact. □

By analogy with weak convergence theory, we may look for a version of (16) for continuous functions.

Theorem 24.10 *(functional LDP, Varadhan, Bryc) Let ξ_ε, $\varepsilon > 0$, be random elements in a metric space S.*

(i) *If (ξ_ε) satisfies an LDP with rate function I, and $f: S \to \mathsf{R}$ is continuous and bounded above, then*

$$\Lambda_f \equiv \lim_{\varepsilon \to 0} \varepsilon \log E \exp\Big\{f(\xi_\varepsilon)/\varepsilon\Big\} = \sup_{x \in S} \Big\{f(x) - I(x)\Big\}.$$

(ii) *If the ξ_ε are exponentially tight and the limits Λ_f in* (i) *exist for all $f \in C_b$, then (ξ_ε) satisfies an LDP with the good rate function*

$$I(x) = \sup_{f \in C_b} \Big\{f(x) - \Lambda_f\Big\}, \quad x \in S.$$

Proof: (i) For every $n \in \mathsf{N}$, we can choose finitely many closed sets $B_1, \ldots, B_m \subset S$, such that $f \leq -n$ on $\bigcap_j B_j^c$, and the oscillation of f on each B_j is at most n^{-1}. Then

$$\begin{aligned}
\limsup_{\varepsilon \to 0} \varepsilon \log E e^{f(\xi_\varepsilon)/\varepsilon} &\leq \max_{j \leq m} \limsup_{\varepsilon \to 0} \varepsilon \log E\Big(e^{f(\xi_\varepsilon)/\varepsilon}; \xi_\varepsilon \in B_j\Big) \vee (-n) \\
&\leq \max_{j \leq m} \Big\{\sup_{x \in B_j} f(x) - \inf_{x \in B_j} I(x)\Big\} \vee (-n) \\
&\leq \max_{j \leq m} \sup_{x \in B_j} \Big\{f(x) - I(x) + n^{-1}\Big\} \vee (-n) \\
&= \sup_{x \in S} \Big\{f(x) - I(x) + n^{-1}\Big\} \vee (-n).
\end{aligned}$$

The upper bound now follows as we let $n \to \infty$. Next we fix any $x \in S$ with a neighborhood G, and write

$$\begin{aligned}
\liminf_{\varepsilon \to 0} \varepsilon \log E e^{f(\xi_\varepsilon)/\varepsilon} &\geq \liminf_{\varepsilon \to 0} \varepsilon \log E\Big(e^{f(\xi_\varepsilon)/\varepsilon}; \xi_\varepsilon \in G\Big) \\
&\geq \inf_{y \in G} f(y) - \inf_{y \in G} I(y) \\
&\geq \inf_{y \in G} f(y) - I(x).
\end{aligned}$$

Here the lower bound follows as we let $G \downarrow \{x\}$, and then take the supremum over $x \in S$.

(ii) Note that I is lower semi-continuous, as the supremum over a family of continuous functions. Since $\Lambda_f = 0$ for $f = 0$, it is also clear that $I \geq 0$. By Lemma 24.9 (ii), it remains to show that (ξ_ε) satisfies the weak LDP with rate function I. Then fix any $\delta > 0$. For every $x \in S$, we may choose a function $f_x \in C_b$ satisfying

$$f_x(x) - \Lambda_{f_x} > \Big\{I(x) - \delta\Big\} \wedge \delta^{-1},$$

and by continuity there exists a neighborhood B_x of x such that

$$f_x(y) > \Lambda_{f_x} + \Big\{I(x) - \delta\Big\} \wedge \delta^{-1}, \quad y \in B_x.$$

By Chebyshev's inequality, we get for any $\varepsilon > 0$

$$\begin{aligned}
P\{\xi_\varepsilon \in B_x\} &\leq E \exp\Big\{\varepsilon^{-1}\Big(f_x(\xi_\varepsilon) - \inf\{f_x(y); y \in B_x\}\Big)\Big\} \\
&\leq E \exp\Big\{\varepsilon^{-1}\Big(f_x(\xi_\varepsilon) - \Lambda_{f_x} - \{I(x) - \delta\} \wedge \delta^{-1}\Big)\Big\},
\end{aligned}$$

and so by the definition of Λ_{f_x},

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in B_x\Big\}\\
&\quad \le \lim_{\varepsilon\to 0} \varepsilon \log E \exp\Big\{f_x(\xi_\varepsilon)/\varepsilon\Big\} - \Lambda_{f_x} - \Big\{I(x)-\delta\Big\} \wedge \delta^{-1}\\
&\quad = -\Big\{I(x)-\delta\Big\} \wedge \delta^{-1}.
\end{aligned}$$

Now fix any compact set $K \subset S$, and choose $x_1,\dots,x_m \in K$ such that $K \subset \bigcup_i B_{x_i}$. Then

$$\begin{aligned}
\limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in K\Big\} &\le \max_{i\le m} \limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in B_{x_i}\Big\}\\
&\le -\min_{i\le m}\Big\{I(x_i)-\delta\Big\} \wedge \delta^{-1}\\
&\le -\Big\{I(K)-\delta\Big\} \wedge \delta^{-1}.
\end{aligned}$$

The upper bound now follows as we let $\delta \to 0$.

Next consider any open set G and element $x \in G$. For any $n \in \mathsf{N}$, we may choose a continuous function $f_n\colon S \to [-n, 0]$ such that $f_n(x) = 0$ and $f_n = -n$ on G^c. Then

$$\begin{aligned}
-I(x) &= \inf_{f\in C_b}\Big\{\Lambda_f - f(x)\Big\}\\
&\le \Lambda_{f_n} - f_n(x) = \Lambda_{f_n}\\
&= \lim_{\varepsilon\to 0} \varepsilon \log E \exp\Big\{f_n(\xi_\varepsilon)/\varepsilon\Big\}\\
&\le \liminf_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in G\Big\} \vee (-n).
\end{aligned}$$

The lower bound now follows as we let $n \to \infty$, and then take the supremum over all $x \in G$. □

We proceed to show how the LDP is preserved by continuous mappings. The following results are often referred to as *direct and inverse contraction principles*. Given a rate function I on S and a map $f\colon S \to T$, we define the image $J = I \circ f^{-1}$ on T as the function

$$\begin{aligned}
J(y) &= I\Big(f^{-1}\{y\}\Big)\\
&= \inf\Big\{I(x);\ f(x) = y\Big\}, \quad y \in T.
\end{aligned} \tag{19}$$

Note that the corresponding set functions are related by

$$\begin{aligned}
J(B) &\equiv \inf_{y\in B} J(y)\\
&= \inf\Big\{I(x);\ f(x) \in B\Big\}\\
&= I(f^{-1}B), \quad B \subset T.
\end{aligned}$$

Theorem 24.11 *(continuous mapping) Let f be a continuous map between two metric spaces S, T, and let ξ_ε be random elements in S.*

(i) *If (ξ_ε) satisfies an LDP in S with rate function I, then the images $f(\xi_\varepsilon)$ satisfy an LDP in T with the raw rate function $J = I \circ f^{-1}$. Moreover, J is a good rate function on T, whenever the function I is good on S.*

(ii) *(Ioffe) Let (ξ_ε) be exponentially tight in S, let f be injective, and let the images $f(\xi_\varepsilon)$ satisfy a weak LDP in T with rate function J. Then (ξ_ε) satisfies an LDP in S with the good rate function $I = J \circ f$.*

Proof: (i) Since f is continuous, the set $f^{-1}B$ is open or closed whenever the corresponding property holds for B. Using the LDP for (ξ_ε), we get for any $B \subset T$

$$\begin{aligned} -I(f^{-1}B^o) &\le \liminf_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in f^{-1}B^o\Big\} \\ &\le \limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in f^{-1}\bar{B}\Big\} \le -I(f^{-1}\bar{B}), \end{aligned}$$

which proves the LDP for $\{f(\xi_\varepsilon)\}$ with the raw rate function $J = I \circ f^{-1}$.

When I is good, we claim that

$$J^{-1}[0,r] = f\Big(I^{-1}[0,r]\Big), \quad r \ge 0. \tag{20}$$

To see this, fix any $r \ge 0$, and let $x \in I^{-1}[0,r]$. Then

$$\begin{aligned} J \circ f(x) &= I \circ f^{-1} \circ f(x) \\ &= \inf\Big\{I(u);\ f(u) = f(x)\Big\} \\ &\le I(x) \le r, \end{aligned}$$

which means that $f(x) \in J^{-1}[0,r]$. Conversely, let $y \in J^{-1}[0,r]$. Since I is good and f is continuous, the infimum in (19) is attained at some $x \in S$, and we get $y = f(x)$ with $I(x) \le r$. Thus, $y \in f(I^{-1}[0,r])$, which completes the proof of (20). Since continuous maps preserve compactness, (20) shows that the goodness of I carries over to J.

(ii) Here I is again a rate function, since the lower semi-continuity of J is preserved by composition with the continuous map f. By Lemma 24.9 (ii), it is then enough to show that (ξ_ε) satisfies a weak LDP in S. To prove the upper bound, fix any compact set $K \subset S$, and note that the image set $f(K)$ is again compact, since f is continuous. Hence, the weak LDP for $\{f(\xi_\varepsilon)\}$ yields

$$\begin{aligned} \limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in K\Big\} &= \limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{f(\xi_\varepsilon) \in f(K)\Big\} \\ &\le -J\{f(K)\} = -I(K). \end{aligned}$$

Next we fix any open set $G \subset S$, and let $x \in G$ be arbitrary with $I(x) = r < \infty$. Since (ξ_ε) is exponentially tight, we may choose a compact set $K \subset S$ such that

$$\limsup_{\varepsilon\to 0} \varepsilon \log P\Big\{\xi_\varepsilon \notin K\Big\} < -r. \tag{21}$$

The continuous image $f(K)$ is compact in T, and so by (21) and the weak LDP for $\{f(\xi_\varepsilon)\}$,

$$\begin{aligned}
-I(K^c) &= -J\Big\{f(K^c)\Big\}\\
&\le -J\Big(\{f(K)\}^c\Big)\\
&\le \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{f(\xi_\varepsilon)\notin f(K)\Big\}\\
&\le \limsup_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\xi_\varepsilon \notin K\Big\} < -r.
\end{aligned}$$

Since $I(x) = r$, we conclude that $x \in K$.

As a continuous bijection from the compact set K onto $f(K)$, the function f is a homeomorphism between the two sets with their subset topologies. Then Lemma 1.6 yields an open set $G' \subset T$, such that $f(x) \in f(G\cap K) = G'\cap f(K)$. Noting that

$$P\Big\{f(\xi_\varepsilon)\in G'\Big\} \le P\Big\{\xi_\varepsilon \in G\Big\} + P\Big\{\xi_\varepsilon \notin K\Big\},$$

and using the weak LDP for $\{f(\xi_\varepsilon)\}$, we get

$$\begin{aligned}
-r &= -I(x) = -J\{f(x)\}\\
&\le \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{f(\xi_\varepsilon)\in G'\Big\}\\
&\le \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\xi_\varepsilon\in G\Big\} \;\vee\; \limsup_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\xi_\varepsilon \notin K\Big\}.
\end{aligned}$$

Hence, by (21),

$$-I(x) \le \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\xi_\varepsilon\in G\Big\}, \quad x\in G,$$

and the lower bound follows as we take the supremum over all $x \in G$. □

We turn to the powerful method of *projective limits*. The following sequential version is sufficient for our needs, and will enable us to extend the LDP to a variety of infinite-dimensional contexts. Some general background on projective limits is provided by Appendix 6.

Theorem 24.12 *(random sequences, Dawson & Gärtner) For any metric spaces $S_1, S_2, \dots$, let $\xi_\varepsilon = (\xi_\varepsilon^n)$ be random elements in $S^\infty = S_1 \times S_2 \times \cdots$, such that for each $n \in \mathsf{N}$ the vectors $(\xi_\varepsilon^1, \dots, \xi_\varepsilon^n)$ satisfy an LDP in $S^n = S_1\times\cdots\times S_n$ with the good rate function I_n. Then (ξ_ε) satisfies an LDP in S^∞ with the good rate function*

$$I(x) = \sup\nolimits_n I_n(x_1,\dots,x_n), \quad x = (x_1, x_2, \dots) \in S^\infty. \tag{22}$$

Proof: For any $m \le n$, we introduce the natural projections $\pi_n\colon S^\infty \to S^n$ and $\pi_{mn} : S^n \to S^m$. Since the π_{mn} are continuous and the I_n are good, Theorem 24.11 yields $I_m = I_n \circ \pi_{mn}^{-1}$ for all $m \le n$, and so $\pi_{mn}(I_n^{-1}[0,r]) \subset I_m^{-1}[0,r]$ for all $r \ge 0$ and $m \le n$. Hence, for each $r \ge 0$ the level sets $I_n^{-1}[0,r]$ form a projective sequence. Since they are also compact by hypothesis, and by (22),

$$I^{-1}[0,r] = \bigcap\nolimits_n \pi_n^{-1} I_n^{-1}[0,r], \quad r \ge 0, \tag{23}$$

the sets $I^{-1}[0,r]$ are compact by Lemma A6.4. Thus, I is again a good rate function.

Now fix any closed set $A \subset S$, and put $A_n = \pi_n A$, so that $\pi_{mn} A_n = A_m$ for all $m \le n$. Since the π_{mn} are continuous, we have also $\pi_{mn} A_n^- \subset A_m^-$ for $m \le n$, which means that the sets A_n^- form a projective sequence. We claim that

$$A = \bigcap_n \pi_n^{-1} A_n^-. \tag{24}$$

Here the relation $A \subset \pi_n^{-1} A_n^-$ is obvious. Now let $x \notin A$. By the definition of the product topology, we may choose a $k \in \mathsf{N}$ and an open set $U \subset S^k$, such that $x \in \pi_k^{-1} U \subset A^c$. It follows easily that $\pi_k x \in U \subset A_k^c$. Since U is open, we have even $\pi_k x \in (A_k^-)^c$. Thus, $x \notin \bigcap_n \pi_n^{-1} A_n^-$, which completes the proof of (24). The projective property carries over to the intersections $A_n^- \cap I_n^{-1}[0,r]$, and formulas (23) and (24) combine into the relation

$$A \cap I^{-1}[0,r] = \bigcap_n \pi_n^{-1} \Big\{ A_n^- \cap I_n^{-1}[0,r] \Big\}, \quad r \ge 0. \tag{25}$$

Now let $I(A) > r \in \mathsf{R}$. Then $A \cap I^{-1}[0,r] = \emptyset$, and by (25) and Lemma A6.4 we get $A_n^- \cap I_n^{-1}[0,r] = \emptyset$ for some $n \in \mathsf{N}$, which implies $I_n(A_n^-) \ge r$. Noting that $A \subset \pi_n^{-1} A_n$ and using the LDP in S^n, we conclude that

$$\begin{aligned} \limsup_{\varepsilon \to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in A\Big\} &\le \limsup_{\varepsilon \to 0} \varepsilon \log P\Big\{\pi_n \xi_\varepsilon \in A_n\Big\} \\ &\le -I_n(A_n^-) \le -r, \end{aligned}$$

The upper bound now follows as we let $r \uparrow I(A)$.

Finally, fix an open set $G \subset S^\infty$, and let $x \in G$ be arbitrary. By the definition of the product topology, we may choose $n \in \mathsf{N}$ and an open set $U \subset S^n$ such that $x \in \pi_n^{-1} U \subset G$. The LDP in S^n yields

$$\begin{aligned} \liminf_{\varepsilon \to 0} \varepsilon \log P\Big\{\xi_\varepsilon \in G\Big\} &\ge \liminf_{\varepsilon \to 0} \varepsilon \log P\Big\{\pi_n \xi_\varepsilon \in U\Big\} \\ &\ge -I_n(U) \\ &\ge -I_n \circ \pi_n(x) \\ &\ge -I(x), \end{aligned}$$

and the lower bound follows as we take the supremum over all $x \in G$. □

We consider yet another basic extension principle for the LDP, involving suitable approximations. The families of random elements ξ_ε and η_ε in a common separable metric space (S,d) are said to be *exponentially equivalent*, if

$$\lim_{\varepsilon \to 0} \varepsilon \log P\Big\{d(\xi_\varepsilon, \eta_\varepsilon) > h\Big\} = -\infty, \quad h > 0. \tag{26}$$

The separability of S is only needed to ensure measurability of the pairwise distances $d(\xi_\varepsilon, \eta_\varepsilon)$. In general, we may replace (26) by a similar condition involving the outer measure.

Lemma 24.13 *(approximation) Let ξ_ε, η_ε be exponentially equivalent random elements in a separable metric space S. Then (ξ_ε) satisfies an LDP with the good rate function I, iff the same LDP holds for (η_ε).*

Proof: Let the LDP hold for (ξ_ε) with rate function I. Fix any closed set $B \subset S$, and write B^h for the closed h-neighborhood of B. Then

$$P\Big\{\eta_\varepsilon \in B\Big\} \le P\Big\{\xi_\varepsilon \in B^h\Big\} + P\Big\{d(\xi_\varepsilon, \eta_\varepsilon) > h\Big\},$$

and so by (26) and the LDP for (ξ_ε),

$$\begin{aligned}
&\limsup_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\eta_\varepsilon \in B\Big\} \\
&\quad \le \limsup_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\xi_\varepsilon \in B^h\Big\} \vee \limsup_{\varepsilon\to 0}\, \varepsilon \log P\Big\{d(\xi_\varepsilon, \eta_\varepsilon) > h\Big\} \\
&\quad \le -I(B^h) \vee (-\infty) \\
&\quad = -I(B^h).
\end{aligned}$$

Since I is good, we have $I(B^h) \uparrow I(B)$ as $h \to 0$, and the required upper bound follows.

Now fix any open set $G \subset S$, and let $x \in G$. If $d(x, G^c) > h > 0$, we may choose a neighborhood U of x such that $U^h \subset G$. Noting that

$$P\Big\{\xi_\varepsilon \in U\Big\} \le P\Big\{\eta_\varepsilon \in G\Big\} + P\Big\{d(\xi_\varepsilon, \eta_\varepsilon) > h\Big\},$$

we get by (26) and the LDP for (ξ_ε)

$$\begin{aligned}
-I(x) &\le -I(U) \\
&\le \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\xi_\varepsilon \in U\Big\} \\
&\le \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\eta_\varepsilon \in G\Big\} \vee \limsup_{\varepsilon\to 0}\, \varepsilon \log P\Big\{d(\xi_\varepsilon, \eta_\varepsilon) > h\Big\} \\
&= \liminf_{\varepsilon\to 0}\, \varepsilon \log P\Big\{\eta_\varepsilon \in G\Big\}.
\end{aligned}$$

The required lower bound now follows, as we take the supremum over all $x \in G$. □

We turn to some important applications, illustrating the power of the general theory. First we study the perturbations of an ordinary differential equation $\dot{x} = b(x)$ by a small noise term. More precisely, we consider the unique solutions X^ε with $X_0^\varepsilon = 0$ of the d-dimensional SDEs[3]

$$dX_t = \varepsilon^{1/2}\, dB_t + b(X_t)\, dt, \quad t \ge 0,\ \varepsilon \ge 0, \tag{27}$$

where B is a Brownian motion in R^d and b is a bounded and uniformly Lipschitz continuous mapping on R^d. Let H_∞ be the set of all absolutely continuous functions $x\colon \mathsf{R}_+ \to \mathsf{R}^d$ with $x_0 = 0$, such that $\dot{x} \in L^2$.

Theorem 24.14 *(perturbed dynamical systems, Freidlin & Wentzell) For any bounded, uniformly Lipschitz continuous function $b\colon \mathsf{R}^d \to \mathsf{R}^d$, the solutions X^ε to* (27) *with $X_0^\varepsilon = 0$ satisfy an LDP in $C_{\mathsf{R}_+,\mathsf{R}^d}$ with the good rate function*

$$I(x) = \tfrac{1}{2} \int_0^\infty \Big|\dot{x}_t - b(x_t)\Big|^2 dt, \quad x \in H_\infty. \tag{28}$$

[3]See Chapter 32 for a detailed discussion of such equations.

Here it is understood that $I(x) = \infty$ when $x \notin H_\infty$. Note that the result for $b = 0$ extends Theorem 24.6 to processes on R_+.

Proof: If B^1 is a Brownian motion on $[0,1]$, then for every $r > 0$, the scaled process $B^r = \Phi(B^1)$ given by $B^r_t = r^{1/2} B^1_{t/r}$ is a Brownian motion on $[0,r]$. Since Φ is continuous from $C_{[0,1]}$ to $C_{[0,r]}$, we see from Theorems 24.6 and 24.11 (i), combined with Lemma 24.7, that the processes $\varepsilon^{1/2} B^r$ satisfy an LDP in $C_{[0,r]}$ with the good rate function $I_r = I_1 \circ \Phi^{-1}$, where $I_1(x) = \frac{1}{2}\, \|\dot{x}\|_2^2$ for $x \in H_1$ and $I_1(x) = \infty$ otherwise. Now Φ maps H_1 onto H_r, and when $y = \Phi(x)$ with $x \in H_1$, we have $\dot{x}_t = r^{1/2}\, \dot{y}_{rt}$. Hence, a simple calculation yields $I_r(y) = \frac{1}{2} \int_0^r |\dot{y}_s|^2 ds = \frac{1}{2}\, \|\dot{y}\|_2^2$, which extends Theorem 24.6 to $[0,r]$. For the further extension to R_+, write $\pi_n x$ for the restriction of a function $x \in C_{\mathsf{R}_+}$ to $[0,n]$, and infer from Theorem 24.12 that the processes $\varepsilon^{1/2} B$ satisfy an LDP in C_{R_+} with the good rate function $I_\infty(x) = \sup_n I_n(\pi_n x) = \frac{1}{2}\, \|\dot{x}\|_2^2$.

By an elementary version of Theorem 32.3, the integral equation

$$x_t = z_t + \int_0^t b(x_s)\, ds, \quad t \geq 0, \tag{29}$$

has a unique solution $x = F(z)$ in $C = C_{\mathsf{R}_+}$ for every $z \in C$. Letting $z^1, z^2 \in C$ be arbitrary, and writing a for the Lipschitz constant of b, we note that the corresponding solutions $x^i = F(z^i)$ satisfy

$$\left| x_t^1 - x_t^2 \right| \leq \left\| z^1 - z^2 \right\| + a \int_0^t \left| x_s^1 - x_s^2 \right| ds, \quad t \geq 0.$$

Hence, Gronwall's Lemma 32.4 yields $\|x^1 - x^2\| \leq \|z^1 - z^2\|\, e^{ar}$ on the interval $[0,r]$, which shows that F is continuous. Using Schilder's theorem on R_+, along with Theorem 24.11 (i), we conclude that the processes X^ε satisfy an LDP in C_{R_+} with the good rate function $I = I_\infty \circ F^{-1}$. Now F is clearly bijective, and by (29) the functions z and $x = F(z)$ lie simultaneously in H_∞, in which case $\dot{z} = \dot{x} - b(x)$ a.e. Thus, I is indeed given by (28). $\square$

Now consider a random element ξ with distribution μ, in an arbitrary metric space S. Introduce the *cumulant-generating functional*

$$\begin{aligned} \Lambda(f) &= \log E e^{f(\xi)} \\ &= \log \mu e^f, \quad f \in C_b(S), \end{aligned}$$

with associated Legendre–Fenchel transform

$$\Lambda^*(\nu) = \sup_{f \in C_b} \left\{ \nu f - \Lambda(f) \right\}, \quad \nu \in \hat{\mathcal{M}}_S, \tag{30}$$

where $\hat{\mathcal{M}}_S$ is the class of probability measures on S, endowed with the topology of weak convergence. Note that Λ and Λ^* are both convex, by the same argument as for R^d.

For any measures $\mu, \nu \in \hat{\mathcal{M}}_S$, we define the *relative entropy* of ν with respect to μ by

$$H(\nu\,|\,\mu) = \begin{cases} \nu \log p = \mu(p \log p), & \nu \ll \mu \text{ with } \nu = p\cdot\mu, \\ \infty, & \nu \not\ll \mu. \end{cases}$$

Since $x\log x$ is convex, the function $H(\nu\,|\,\mu)$ is convex in ν for fixed μ, and Jensen's inequality yields

$$\begin{aligned} H(\nu\,|\,\mu) &\geq \mu p \log \mu p \\ &= \nu S \log \nu S = 0, \quad v \in \hat{\mathcal{M}}_S, \end{aligned}$$

with equality iff $\nu = \mu$.

Now let $\xi_1, \xi_2, \ldots$ be i.i.d. random elements in S. The associated *empirical distributions*

$$\eta_n = n^{-1} \sum_{k\leq n} \delta_{\xi_k}, \quad n \in \mathsf{N},$$

may be regarded as random elements in $\hat{\mathcal{M}}_S$, and we note that

$$\eta_n f = n^{-1} \sum_{k\leq n} f(\xi_k), \quad f \in C_b(S), \ \ n \in \mathsf{N}.$$

In particular, Theorem 24.5 applies to the random vectors $(\eta_n f_1, \ldots, \eta_n f_m)$, for fixed $f_1, \ldots, f_m \in C_b(S)$. The following result may be regarded as an infinite-dimensional version of Theorem 24.5. It also provides an important connection to statistical mechanics, via the entropy function.

Theorem 24.15 *(large deviations of empirical distributions, Sanov) Let ξ_1, $\xi_2, \ldots$ be i.i.d. random elements with distribution μ in a Polish space S, and put $\Lambda(f) = \log \mu\, e^f$. Then the associated empirical distributions $\eta_1, \eta_2, \ldots$ satisfy an LDP in $\hat{\mathcal{M}}_S$ with the good rate function*

$$\Lambda^*(\nu) = H(\nu\,|\,\mu), \quad \nu \in \hat{\mathcal{M}}_S. \tag{31}$$

A couple of lemmas will be needed for the proof.

Lemma 24.16 *(relative entropy, Donsker & Varadhan) For any $\mu, \nu \in \hat{\mathcal{M}}_S$,*

(i) *the supremum in* (30) *may be taken over all bounded, measurable functions $f\colon S \to \mathsf{R}$,*

(ii) *then* (31) *extends to measures on any measurable space S.*

Proof: (i) Use Lemma 1.37 and dominated convergence.

(ii) If $\nu \not\ll \mu$, then $H(\nu\,|\,\mu) = \infty$ by definition. Choosing $B \in \mathcal{S}$ with $\mu B = 0$ and $\nu B > 0$, and taking $f_n = n 1_B$, we obtain $\nu f_n - \log \mu e^{f_n} = n\, \nu B \to \infty$. Thus, even $\Lambda^*(\nu) = \infty$ in this case, and it remains to prove (31) when $\nu \ll \mu$. Assuming $\nu = p \cdot \mu$ and writing $f = \log p$, we note that

$$\begin{aligned} \nu f - \log \mu e^f &= \nu \log p - \log \mu p \\ &= H(\nu\,|\,\mu). \end{aligned}$$

If $f = \log p$ is unbounded, it may approximated by some bounded, measurable functions f_n satisfying $\mu e^{f_n} \to 1$ and $\nu f_n \to \nu f$, and we get $\Lambda^*(\nu) \geq H(\nu\,|\,\mu)$.

To prove the reverse inequality, we first let $\mathcal{S}$ be finite and generated by a partition $B_1, \dots, B_n$ of S. Putting $\mu_k = \mu B_k$, $\nu_k = \nu B_k$, and $p_k = \nu_k/\mu_k$, we may write our claim in the form

$$\begin{aligned} g(x) &\equiv \textstyle\sum_k \nu_k x_k - \log \sum_k \mu_k e^{x_k} \\ &\le \textstyle\sum_k \nu_k \log p_k, \end{aligned}$$

where $x = (x_1, \dots, x_n) \in \mathsf{R}^d$ is arbitrary. Here the function g is concave and satisfies $\nabla g(x) = 0$ for $x = (\log p_1, \dots, \log p_n)$, asymptotically when $p_k = 0$ for some k. Thus,

$$\begin{aligned} \sup\nolimits_x g(x) &= g\Big(\log p_1, \dots, \log p_k\Big) \\ &= \textstyle\sum_k \nu_k \log p_k. \end{aligned}$$

To prove the general inequality $\nu f - \log \mu e^f \le \nu \log p$, we may choose f to be simple. The generated σ-field $\mathcal{F} \subset \mathcal{S}$ is then finite, and we note that $\nu = \mu(p\,|\mathcal{F}) \cdot \mu$ on $\mathcal{F}$. Using the result in the finite case, along with Jensen's inequality for conditional expectations, we obtain

$$\begin{aligned} \nu f - \log \mu e^f &\le \mu\Big\{\mu(p\,|\mathcal{F})\, \log \mu(p\,|\mathcal{F})\Big\} \\ &\le \mu\, \mu\Big(p \log p \,\Big|\, \mathcal{F}\Big) \\ &= \nu \log p. \end{aligned}$$ □

Lemma 24.17 *(exponential tightness) The empirical distributions η_n in Theorem* 24.15 *are exponentially tight in $\hat{\mathcal{M}}_S$.*

Proof: If $B \in \mathcal{S}$ with $P\{\xi \in B\} = p \in (0,1)$, Theorem 24.3 and Lemmas 24.1 and 24.2 yield for any $x \in [p,1]$

$$\sup\nolimits_n n^{-1} \log P\Big\{\eta_n B > x\Big\} \le -x \log \frac{x}{p} - (1-x) \log \frac{1-x}{1-p}\,. \tag{32}$$

In particular, the right-hand side tends to $-\infty$ as $p \to 0$ for fixed $x \in (0,1)$. Now fix any $r > 0$. By (32) and Theorem 23.2, we may choose some compact sets $K_1, K_2, \dots \subset S$ with

$$P\Big\{\eta_n K_k^c > 2^{-k}\Big\} \le e^{-knr}, \quad k, n \in \mathsf{N}.$$

Summing over k gives

$$\limsup_{n\to\infty} n^{-1} \log P \textstyle\bigcup_k \Big\{\eta_n K_k^c > 2^{-k}\Big\} \le -r,$$

and it remains to note that the set

$$M = \textstyle\bigcap_k \Big\{\nu \in \hat{\mathcal{M}}_S;\ \nu K_k^c \le 2^{-k}\Big\}$$

is compact, by another application of Theorem 23.2. □

Proof of Theorem 24.15: By Theorem 1.8, we can embed S as a Borel subset of a compact metric space K. The function space $C_b(K)$ is separable, and we can choose a dense sequence $f_1, f_2, \ldots \in C_b(K)$. For any $m \in \mathbb{N}$, the random vector $\{f_1(\xi), \ldots, f_m(\xi)\}$ has cumulant-generating function

$$\begin{aligned}\Lambda_m(u) &= \log E \exp \sum_{k\le m} u_k f_k(\xi)\\ &= \Lambda \circ \sum_{k\le m} u_k f_k, \quad u \in \mathbb{R}^m,\end{aligned}$$

and so by Theorem 24.5 the random vectors $(\eta_n f_1, \ldots, \eta_n f_m)$ satisfy an LDP in $\mathbb{R}^m$ with the good rate function Λ_m^*. Then by Theorem 24.12, the infinite sequences $(\eta_n f_1, \eta_n f_2, \ldots)$ satisfy an LDP in $\mathbb{R}^\infty$ with the good rate function $J = \sup_m(\Lambda_m^* \circ \pi_m)$, where π_m denotes the natural projection of $\mathbb{R}^\infty$ onto $\mathbb{R}^m$. Since $\hat{\mathcal{M}}_K$ is compact by Theorem 23.2 and the mapping $\nu \mapsto (\nu f_1, \nu f_2, \ldots)$ is a continuous injection of $\hat{\mathcal{M}}_K$ into $\mathbb{R}^\infty$, Theorem 24.11 (ii) shows that the random measures η_n satisfy an LDP in $\hat{\mathcal{M}}_K$ with the good rate function

$$\begin{aligned}I_K(\nu) &= J\big(\nu f_1, \nu f_2, \ldots\big)\\ &= \sup\nolimits_m \Lambda_m^*\big(\nu f_1, \ldots, \nu f_m\big)\\ &= \sup_m \sup_{u\in\mathbb{R}^m}\Big(\sum_{k\le m} u_k \nu f_k - \Lambda \circ \sum_{k\le m} u_k f_k\Big)\\ &= \sup_{f\in\mathcal{F}}\big\{\nu f - \Lambda(f)\big\}\\ &= \sup_{f\in C_b}\big\{\nu f - \Lambda(f)\big\},\end{aligned} \tag{33}$$

where $\mathcal{F}$ is the set of all linear combinations of $f_1, f_2, \ldots$.

Next, we note that the natural embedding $\hat{\mathcal{M}}_S \to \hat{\mathcal{M}}_K$ is continuous, since for any $f \in C_b(K)$ the restriction of f to S belongs to $C_b(S)$. Since it is also trivially injective, we see from Theorem 24.11 (ii) and Lemma 24.17 that the η_n satisfy an LDP even in $\hat{\mathcal{M}}_S$, with a good rate function I_S equal to the restriction of I_K to $\hat{\mathcal{M}}_S$. It remains to note that $I_S = \Lambda^*$ by (33) and Lemma 24.16. □

We conclude with a remarkable application of Schilder's Theorem 24.6. Writing B for a standard Brownian motion in $\mathbb{R}^d$, we introduce for each $t > e$ the scaled process

$$X_s^t = \frac{B_{st}}{\sqrt{2t \log\log t}}, \quad s \ge 0. \tag{34}$$

Theorem 24.18 *(functional law of the iterated logarithm, Strassen) For a Brownian motion B in $\mathbb{R}^d$, define the processes X^t by (34). Then these statements are equivalent and hold outside a fixed P-null set:*

(i) *the set of paths X^t with $t \ge 3$ is relatively compact in $C_{\mathbb{R}_+,\mathbb{R}^d}$, with set of limit points as $t \to \infty$*

$$K = \big\{x \in H_\infty;\ \|\dot{x}\|_2 \le 1\big\},$$

(ii) *for any continuous function $F\colon C_{\mathbb{R}_+,\mathbb{R}^d} \to \mathbb{R}$,*

$$\limsup_{t\to\infty} F(X^t) = \sup_{x\in K} F(x).$$

In particular, we may recover the classical law of the iterated logarithm in Theorem 14.18 by choosing $F(x) = x_1$. Using Theorem 22.6, we can easily derive a correspondingly strengthened version for random walks.

Proof: The equivalence (i) $\Leftrightarrow$ (ii) being elementary, it is enough to prove (i). Noting that $X^t \stackrel{d}{=} (2\log\log t)^{-1/2}B$ and using Theorem 24.6, we get for any measurable set $A \subset C_{\mathsf{R}_+,\mathsf{R}^d}$ and constant $r > 1$

$$\limsup_{n\to\infty} \frac{\log P\Big\{X^{r^n}\in A\Big\}}{\log n} \le \limsup_{t\to\infty} \frac{\log P\{X^t\in A\}}{\log\log t} \le -2\,I(\bar{A}),$$
$$\liminf_{n\to\infty} \frac{\log P\Big\{X^{r^n}\in A\Big\}}{\log n} \ge \liminf_{t\to\infty} \frac{\log P\{X^t\in A\}}{\log\log t} \ge -2\,I(A^o),$$

where $I(x) = \frac{1}{2}\,\|\dot{x}\|_2^2$ for $x \in H_\infty$ and $I(x) = \infty$ otherwise. Hence,

$$\sum\nolimits_n P\Big\{X^{r^n}\in A\Big\}\begin{cases} < \infty, & 2\,I(\bar{A}) > 1,\\ = \infty, & 2\,I(A^o) < 1.\end{cases} \tag{35}$$

Now fix any $r > 1$, and let $G \supset K$ be open. Note that $2\,I(G^c) > 1$ by Lemma 24.7. By the first part of (35) and the Borel–Cantelli lemma, we have $P\{X^{r^n}\notin G \text{ i.o.}\} = 0$, or equivalently $1_G(X^{r^n}) \to 1$ a.s. Since G was arbitrary, it follows that $\rho(X^{r^n}, K) \to 0$ a.s. for any metrization ρ of $C_{\mathsf{R}_+,\mathsf{R}^d}$. In particular, this holds with any $c > 0$ for the metric

$$\rho_c(x,y) = \int_0^\infty \Big\{(x-y)_s^* \wedge 1\Big\}\,e^{-cs}\,ds, \quad x,y \in C_{\mathsf{R}_+,\mathsf{R}^d}.$$

To extend the convergence to the entire family $\{X^t\}$, fix a path of B with $\rho_1(X^{r^n}, K) \to 0$, and choose some functions $y^{r^n} \in K$ with $\rho_1(X^{r^n}, y^{r^n}) \to 0$. For any $t \in [r^n, r^{n+1})$, the paths X^{r^n} and X^t are related by

$$X^t(s) = X^{r^n}(t\,r^{-n}s)\left(\frac{r^n\log\log r^n}{t\log\log t}\right)^{1/2}, \quad s > 0.$$

Defining y^t in the same way in terms of y^{r^n}, we note that also $y^t \in K$, since $I(y^t) \le I(y^{r^n})$. (The two H_∞-norms would agree if the logarithmic factors were omitted.) Furthermore,

$$\begin{aligned}\rho_r(X^t, y^t) &= \int_0^\infty \Big\{\Big(X^t - y^t\Big)_s^* \wedge 1\Big\}\,e^{-rs}\,ds\\ &\le \int_0^\infty \Big\{\Big(X^{r^n} - y^{r^n}\Big)_{rs}^* \wedge 1\Big\}\,e^{-rs}\,ds\\ &= r^{-1}\rho_1\Big(X^{r^n}, y^{r^n}\Big) \to 0.\end{aligned}$$

Thus, $\rho_r(X^t, K) \to 0$. Since K is compact, we conclude that $\{X^t\}$ is relatively compact, and that all its limit points as $t \to \infty$ belong to K.

Now fix any $y \in K$ and $u \ge \varepsilon > 0$. By the established part of the theorem and Cauchy's inequality, we have a.s.

$$\begin{aligned}\limsup_{t\to\infty}\left(X^t-y\right)^*_\varepsilon &\le \sup_{x\in K}(x-y)^*_\varepsilon\\ &\le \sup_{x\in K}x^*_\varepsilon+y^*_\varepsilon\\ &\le 2\,\varepsilon^{1/2}.\end{aligned}\tag{36}$$

Write $x^*_{\varepsilon,u}=\sup_{s\in[\varepsilon,u]}|x_s-x_\varepsilon|$, and choose $r>u/\varepsilon$ to ensure independence between the variables $(X^{r^n}-y)^*_{\varepsilon,u}$. Applying the second case of (35) to the open set $A=\{x;\,(x-y)^*_{\varepsilon,u}<\varepsilon\}$, and using the Borel–Cantelli lemma together with (36), we obtain a.s.

$$\begin{aligned}\liminf_{t\to\infty}\left(X^t-y\right)^*_u &\le \limsup_{t\to\infty}\left(X^t-y\right)^*_\varepsilon+\liminf_{n\to\infty}\left(X^{r^n}-y\right)^*_{\varepsilon,u}\\ &\le 2\,\varepsilon^{1/2}+\varepsilon.\end{aligned}$$

As $\varepsilon\to 0$, we get $\liminf_t(X^t-y)^*_u=0$ a.s., and so $\liminf_t\rho_1(X^t,y)\le e^{-u}$ a.s. Letting $u\to\infty$, we obtain $\liminf_t\rho_1(X^t,y)=0$ a.s. Applying this result to a dense sequence $y_1,y_2,\ldots\in K$, we see that a.s. every element of K is a limit point as $t\to\infty$ of the family $\{X^t\}$. □

Exercises

1. For any random vector ξ and constant a in R^d, show that $\Lambda_{\xi-a}(u)=\Lambda_\xi(u)-ua$ and $\Lambda^*_{\xi-a}(x)=\Lambda^*_\xi(x+a)$.

2. For any random vector ξ in R^d and non-singular $d\times d$ matrix a, show that $\Lambda_{a\xi}(u)=\Lambda_\xi(ua)$ and $\Lambda^*_{a\xi}(x)=\Lambda^*_\xi(a^{-1}x)$.

3. For any random vectors $\xi\perp\!\!\!\perp\eta$, show that $\Lambda_{\xi,\eta}(u,v)=\Lambda_\xi(u)+\Lambda_\eta(v)$ and $\Lambda^*_{\xi,\eta}(x,y)=\Lambda^*_\xi(x)+\Lambda^*_\eta(y)$.

4. Prove the claims of Lemma 24.2.

5. For a Gaussian vector ξ in R^d with mean $m\in\mathsf{R}^d$ and covariance matrix a, show that $\Lambda^*_\xi(x)=\frac{1}{2}\,(x-m)'a^{-1}(x-m)$. Explain the interpretation when a is singular.

6. Let ξ be a standard Gaussian random vector in R^d. Show that the family $\varepsilon^{1/2}\xi$ satisfies an LDP in R^d with the good rate function $I(x)=\frac{1}{2}\,|x|^2$. (*Hint:* Deduce the result along the sequence $\varepsilon_n=n^{-1}$ from Theorem 24.5, and extend by monotonicity to general $\varepsilon>0$.)

7. Use Theorem 24.11 (i) to deduce the preceding result from Schilder's theorem. (*Hint:* For $x\in H_1$, note that $|x_1|\le\|\dot{x}\|_2$, with equality iff $x_t\equiv tx_1$.)

8. Prove Schilder's theorem on $[0,T]$ by the same argument as for $[0,1]$.

9. Deduce Schilder's theorem in $C_{[0,n],\mathsf{R}^d}$ from the version in $C_{[0,1],\mathsf{R}^{nd}}$.

10. Let B be a Brownian bridge in R^d. Show that the processes $\varepsilon^{1/2}B$ satisfy an LDP in $C_{[0,1],\mathsf{R}^d}$ with the good rate function $I(x)=\frac{1}{2}\,\|\dot{x}\|_2^2$ for $x\in H_1$ with $x_1=0$ and $I(x)=\infty$ otherwise. (*Hint:* Write $B_t=X_t-tX_1$, where X is a Brownian motion in R^d, and use Theorem 24.11. Check that $\|\dot{x}-a\|_2$ is minimized for $a=x_1$.)

11. Show that the exponential tightness and its sequential version are preserved by continuous mappings.

12. Show that if the processes X^ε, Y^ε in $C_{\mathsf{R}_+,\mathsf{R}^d}$ are exponentially tight, then so is any linear combination $aX^\varepsilon + bY^\varepsilon$. (*Hint:* Use the Arzelà–Ascoli theorem.)

13. Prove directly from (27) that the processes X^ε in Theorem 24.14 are exponentially tight. (*Hint:* Use Lemmas 24.7 and 24.9 (iii), along with the Arzelà–Ascoli theorem.) Derive the same result from the stated theorem.

14. Let ξ_ε be random elements in a locally compact metric space S, satisfying an LDP with the good rate function I. Show that the ξ_ε are exponentially tight, even in the non-sequential sense. (*Hint:* For any $r > 0$, there exists a compact set $K_r \subset S$ with $I^{-1}[0,r] \subset K_r^o$. Now apply the LDP upper bound to the closed sets $(K_r^o)^c \supset K_r^c$.)

15. For any metric space S and lcscH space T, let X^ε be random elements in $C_{T,S}$, whose restrictions X_K^ε to an arbitrary compact set $K \subset T$ satisfy an LDP in $C_{K,S}$ with the good rate function I_K. Show that the X^ε satisfy an LDP in $C_{T,S}$ with the good rate function $I = \sup_K (I_K \circ \pi_K)$, where π_K denotes the restriction map $C_{T,S} \to C_{K,S}$.

16. Let ξ_{kj} be i.i.d. random vectors in R^d with $\Lambda(u) = Ee^{u\xi_{kj}} < \infty$ for all $u \in \mathsf{R}^d$. Show that the sequences $\bar\xi_n = n^{-1}\sum_{k\le n}(\xi_{k1}, \xi_{k2}, \ldots)$ satisfy an LDP in $(\mathsf{R}^d)^\infty$ with the good rate function $I(x) = \sum_j \Lambda^*(x_j)$. Also derive an LDP for the associated random walks in R^d.

17. Let ξ be a sequence of i.i.d. $N(0,1)$ random variables. Use the preceding result to show that the sequences $\varepsilon^{1/2}\xi$ satisfy an LDP in R^∞ with the good rate function $I(x) = \frac{1}{2}\|x\|^2$ for $x \in l^2$ and $I(x) = \infty$ otherwise. Also derive the statement from Schilder's theorem.

18. Let $\xi_1, \xi_2, \ldots$ be i.i.d. random probability measures on a Polish space S. Derive an LDP in $\hat{\mathcal{M}}_S$ for the averages $\bar\xi_n = n^{-1}\sum_{k\le n}\xi_k$. (*Hint:* Define $\Lambda(f) = \log Ee^{\xi_k f}$, and proceed as in the proof of Sanov's theorem.)

19. Show how the classical LIL[4] in Theorem 14.18 follows from Theorem 24.18. Also use the latter result to derive an LIL for the variables $\xi_t = |B_{2t} - B_t|$, where B is a Brownian motion in R^d.

20. Use Theorem 24.18 to derive a corresponding LIL in $C_{[0,1],\mathsf{R}^d}$.

21. Use Theorems 22.6 and 24.18 to derive a functional LIL for random walks, based on i.i.d. random variables with mean 0 and variance 1. (*Hint:* For the result in $C_{\mathsf{R}_+,\mathsf{R}}$, replace the summation process $S_{[t]}$ by its linearly interpolated version, as in Corollary 23.6.)

22. Use Theorems 22.13 and 24.18 to derive a functional LIL for suitable renewal processes.

23. Let $B^1, B^2, \ldots$ be independent Brownian motions in R^d. Show that the sequence of paths $X_t^n = (2\log n)^{-1/2}B_t^n$, $n \ge 2$, is a.s. relatively compact in $C_{\mathsf{R}_+,\mathsf{R}^d}$ with set of limit points $K = \{x \in H_\infty;\ \|\dot x\|_2 \le 1\}$.

[4]law of the iterated logarithm

VIII. Stationarity, Symmetry and Invariance

Stationary and symmetric processes represent the third basic dependence structure of probability theory. In Theorem 25 we prove the pointwise and mean ergodic theorems in discrete and continuous time, along with various multi-variate and sub-additive extensions. Some main results admit extensions to transition semi-groups of Markov processes, which leads in Chapter 26 to some powerful limit theorems for the latter. Here we also study weak and strong ergodicity as well as Harris recurrence. In Chapter 27 we prove the basic representations and limit theorems for exchangeable processes, as well as the predictable sampling and mapping theorems. The final Chapter 28 provides representations of contractable, exchangeable, and rotatable arrays. For the novice we recommend especially Chapter 25 plus selected material from Chapters 26–27. Chapter 28 is more advanced and might be postponed.

— — —

25. Stationary processes and ergodic theorems. After a general discussion of stationary and ergodic sequences and processes, we prove the basic discrete- and continuous-time ergodic theorems, along with various multivariate and sub-additive extensions. Next we consider the basic ergodic decomposition and prove some powerful coupling theorems. Among applications we note some limit theorems for entropy and information and for random matrices. We also include a very general version of the classical ballot theorem.

26. Ergodic properties of Markov processes. Here we begin with the a.e. ergodic theorem for positive L^1-L^∞ contractions on an abstract measure space, along with a related ratio ergodic theorem, leading to some powerful limit theorems for Markov processes. Next we characterize weak and strong ergodicity, and study the notion of Harris recurrence and the existence of invariant measures for regular Feller processes, using tools from potential theory.

27. Symmetric distributions and predictable maps. Here the central result is the celebrated de Finetti-type representation of exchangeable and contractable sequences, along with various continuous-time extensions, allowing applications to sampling from a finite population. We further note the powerful predictable sampling and mapping theorems, where the original invariance property is extended to predictable times and mappings.

28. Multi-variate arrays and symmetries. Here we establish the Aldous–Hoover-type coding representations of jointly exchangeable and contractable arrays of arbitrary dimension, and identify pairs of functions that can be used to represent the same array. After highlighting some special cases, we proceed to a representation of rotatable arrays in terms of Gaussian random variables, and conclude with an extension of Kingman's paintbox representation for exchangeable and related partitions.

Chapter 25

Stationary Processes and Ergodic Theory

Stationarity and invariance, two-sided extension, invariance and almost invariance, ergodicity, maximal ergodic lemma, discrete and continuous-time ergodic theorems, varying functions, moment and maximum inequalities, multi-variate ergodic theorem, commuting maps, convex averages, mean and sub-additive ergodic theorems, products of random matrices, ergodic decomposition, shift coupling and mixing criteria, group coupling, ballot theorems, entropy and information

We now come to the subject of stationarity, the third basic dependence structure of modern probability, beside those of martingales and Markov processes[1]. Apart from their useful role in probabilistic modeling and in providing a natural setting for many general theorems, stationary processes often arise as limiting processes in a wide variety of contexts throughout probability. In particular, they appear as steady-state versions of various Markov and renewal-type processes. Their importance is further due to the existence of a range of fundamental limit theorems and maximum inequalities, belonging to the basic tool kit of every probabilist.

A process is said to be *stationary* if its distribution is invariant[2] under shifts. A key result is Birkhoff's ergodic theorem, which may be regarded as a strong law of large numbers for stationary sequences and processes. After proving the classical ergodic theorems in discrete and continuous time, we turn to the multi-variate versions of Zygmund and Wiener, the former in a setting for non-commutative maps on rectangular regions, the latter in the commutative case and involving averages over increasing families of convex sets. We further consider a version of Kingman's sub-additive ergodic theorem, and discuss a basic application to random matrices.

In all the mentioned results, the limit is a random variable, measurable with respect to an appropriate invariant σ-field $\mathcal{I}$. Of special importance is the ergodic case, where $\mathcal{I}$ is trivial and the limit reduces to a constant. For general stationary processes, we consider a decomposition of the distribution into ergodic components. We further consider some basic criteria for a suitable shift coupling of two processes, expressed in terms of the tail and invariant σ-fields $\mathcal{T}$ and $\mathcal{I}$, respectively. Those results will be helpful to prove some ergodic

[1]not to mention the mere independence

[2]Stationarity and invariance are often confused. The a.s. invariance $T\xi = \xi$ is clearly much stronger than the stationarity $T\xi \overset{d}{=} \xi$ or $\mathcal{L}(T\xi) = \mathcal{L}(\xi)$.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_26

theorems in Chapters 27 and 31. We conclude with some general versions of the classical ballot theorem, and prove the basic limit theorem for entropy and information.

Our treatment of stationary sequences and processes is continued in Chapters 27 and 31 with some important applications and extensions of the present theory, including various ergodic theorems for Palm distributions. In Chapter 26, we show how the basic ergodic theorems allow extensions to suitable contraction operators, which leads to a profound unification of the present theory with the ergodic theory for Markov transition operators. In this context, we also prove some powerful ratio ergodic theorems.

Returning to the basic notions of stationarity and invariance, we fix a general measurable space $(S, \mathcal{S})$, often taken to be Borel. Given a measure μ and a measurable transformation T on S, we say that T preserves μ or is *μ-preserving* if $\mu \circ T^{-1} = \mu$. Thus, if ξ is a random element in S with distribution μ, then T is measure-preserving iff $T\xi \overset{d}{=} \xi$. In particular, we may consider a random sequence $\xi = (\xi_0, \xi_1, \ldots)$ in a measurable space $(S', \mathcal{S}')$, and let θ denote the *shift* on $S = (S')^\infty$, given by $\theta(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$. Then ξ is said to be *stationary*[3] if $\theta\,\xi \overset{d}{=} \xi$. We show that the general situation is equivalent to this special case.

Lemma 25.1 *(stationarity and invariance) For any random element ξ in S and measurable map T on S, we have*

(i) *$T\xi \overset{d}{=} \xi$ iff the sequence $(T^n\xi)$ is stationary,*

(ii) *for ξ as in* (i), *$(f \circ T^n\xi)$ is stationary for every measurable function f,*

(iii) *any stationary random sequence in S can be represented as in* (ii).

Proof: Assuming $T\xi \overset{d}{=} \xi$, we get

$$\begin{aligned} \theta(f \circ T^n\xi) &= (f \circ T^{n+1}\xi) \\ &= (f \circ T^n T\xi) \\ &\overset{d}{=} (f \circ T^n\xi), \end{aligned}$$

and so $(f \circ T^n\xi)$ is stationary. Conversely, if $\eta = (\eta_0, \eta_1, \ldots)$ is stationary, we may write $\eta_n = \pi_0(\theta^n\eta)$ with $\pi_0(x_0, x_1, \ldots) = x_0$, and note that $\theta\eta \overset{d}{=} \eta$ by the stationarity of η. □

In particular, when $\xi_0, \xi_1, \ldots$ is a stationary sequence of random elements in a measurable space S, and f is a measurable map of S^∞ into a measurable space S', we see that the random sequence

$$\eta_n = f(\xi_n, \xi_{n+1}, \ldots), \quad n \in \mathsf{Z}_+,$$

[3]The existence of a P-preserving transformation T on the underlying probability space Ω is often postulated as part of the definition, so that any random element ξ generates a stationary sequence $\xi_n = \xi \circ T^n$. Here we prefer a purely probabilistic setting, where stationarity is defined directly in terms of the distributions of (ξ_n). Though the resulting statements are more general, most proofs remain essentially the same.

is again stationary.

The definition of stationarity extends in an obvious way to random sequences indexed by $\mathsf{Z} = \{\dots, -1, 0, 1, \dots\}$. A technical advantage of the two-sided version is that the associated shift operators are invertible and form a group, rather than just a semi-group. We show that the two cases are essentially equivalent. Here we assume the existence of appropriate randomization variables, as explained in Chapter 8.

Lemma 25.2 *(two-sided extension) Every stationary random sequence $\xi_0, \xi_1, \dots$ in a Borel space can be extended to a two-sided stationary sequence*

$$(\dots, \xi_{-1}, \xi_0, \xi_1, \xi_2, \dots).$$

Proof: Letting $\vartheta_1, \vartheta_2, \dots$ be i.i.d. $U(0,1)$ and independent of $\xi = (\xi_0, \xi_1, \dots)$, we may construct the ξ_{-n} recursively as functions of ξ and $\vartheta_1, \dots, \vartheta_n$, such that $(\xi_{-n}, \xi_{-n+1}, \dots) \stackrel{d}{=} \xi$ for all n. In fact, once $\xi_{-1}, \dots, \xi_{-n}$ have been chosen, the existence of ξ_{-n-1} is clear from Theorem 8.17, if we note that $(\xi_{-n}, \xi_{-n+1}, \dots) \stackrel{d}{=} \theta\xi$. Finally, the extended sequence is stationary by Proposition 4.2. $\square$

Now fix a measurable transformation T on a measure space $(S, \mathcal{S}, \mu)$, and let $\mathcal{S}^\mu$ denote the μ-completion of $\mathcal{S}$. We say that a set $I \subset S$ is *invariant* if $T^{-1}I = I$ and *almost invariant* if $T^{-1}I = I$ a.e. μ, in the sense that $\mu(T^{-1}I \triangle I) = 0$. Since inverse maps preserve the basic set operations, the classes $\mathcal{I}$ and $\mathcal{I}'$ of invariant sets in $\mathcal{S}$ and almost invariant sets in $\mathcal{S}^\mu$ form σ-fields in S, called the *invariant* and *almost invariant σ-fields*, respectively.

A measurable function f on S is said to be *invariant* if $f \circ T \equiv f$, and *almost invariant* if $f \circ T = f$ a.e. μ. We show how invariant or almost invariant sets and functions are related.

Lemma 25.3 *(invariant sets and functions) Fix a measure μ and a measurable transformation T on S, and let f be a measurable map of S into a Borel space U. Then*

(i) *f is invariant iff it is $\mathcal{I}$-measurable,*

(ii) *f is almost invariant iff it is $\mathcal{I}'$-measurable.*

Proof: Since U is Borel, we may take $U = \mathsf{R}$. If f is invariant or almost invariant, then so is the set $I_x = f^{-1}(-\infty, x)$ for any $x \in \mathsf{R}$, and so $I_x \in \mathcal{I}$ or $\mathcal{I}'$, respectively. Conversely, if f is measurable with respect to $\mathcal{I}$ or $\mathcal{I}'$, then $I_x \in \mathcal{I}$ or $\mathcal{I}'$, respectively, for every $x \in \mathsf{R}$. Hence, the function $f_n(s) = 2^{-n}[2^n f(s)]$, $s \in S$, is invariant or almost invariant for every $n \in \mathsf{N}$, and the invariance or almost invariance carries over to the limit f. $\square$

We proceed to clarify the relationship between the invariant and almost invariant σ-fields. Here $\mathcal{I}^\mu$ denotes the μ-completion of $\mathcal{I}$ in $\mathcal{S}^\mu$, the σ-field generated by $\mathcal{I}$ and the μ-null sets in $\mathcal{S}^\mu$.

Lemma 25.4 *(almost invariance) For any distribution μ and μ-preserving transformation T on S, the invariant and almost invariant σ-fields $\mathcal{I}$ and $\mathcal{I}'$ are related by $\mathcal{I}' = \mathcal{I}^\mu$.*

Proof: If $J \in \mathcal{I}^\mu$, there exists an $I \in \mathcal{I}$ with $\mu(I \triangle J) = 0$. Since T is μ-preserving, we get

$$\begin{aligned}\mu\Big(T^{-1}J \triangle J\Big) &\le \mu\Big(T^{-1}J \triangle T^{-1}I\Big) + \mu\Big(T^{-1}I \triangle I\Big) + \mu(I \triangle J)\\ &= \mu \circ T^{-1}(J \triangle I)\\ &= \mu(J \triangle I) = 0,\end{aligned}$$

which shows that $J \in \mathcal{I}'$. Conversely, given any $J \in \mathcal{I}'$, we may choose a $J' \in \mathcal{S}$ with $\mu(J \triangle J') = 0$, and put $I = \bigcap_n \bigcup_{k \ge n} T^{-n} J'$. Then clearly $I \in \mathcal{I}$ and $\mu(I \triangle J) = 0$, and so $J \in \mathcal{I}^\mu$. □

A measure-preserving map T on a probability space $(S, \mathcal{S}, \mu)$ is said to be *ergodic* for μ or simply μ*-ergodic*, if the invariant σ-field $\mathcal{I}$ is μ*-trivial*, in the sense that $\mu I = 0$ or 1 for every $I \in \mathcal{I}$. Depending on viewpoint, we may prefer to say that μ is ergodic for T, or T-ergodic. The terminology carries over to any random element ξ with distribution μ, which is said to be ergodic whenever this is true for T or μ. Thus, ξ is ergodic iff $P\{\xi \in I\} = 0$ or 1 for any $I \in \mathcal{I}$, i.e., if the σ-field $\mathcal{I}_\xi = \xi^{-1}\mathcal{I}$ in Ω is P-trivial. In particular, a stationary sequence $\xi = (\xi_n)$ is ergodic, if the shift-invariant σ-field is trivial for the distribution of ξ.

We show how the ergodicity of a random element ξ is related to the ergodicity of the generated stationary sequence.

Lemma 25.5 *(ergodicity) Let ξ be a random element in S with distribution μ, and let T be a μ-preserving map on S. Then*

$$\xi \text{ is } T\text{-ergodic} \quad \Leftrightarrow \quad (T^n\xi) \text{ is } \theta\text{-ergodic},$$

in which case $\eta = (f \circ T^n \xi)$ is θ-ergodic for every measurable function f on S.

Proof: Fix any measurable map $f\colon S \to S'$, and define $F = \{f \circ T^n;\, n \ge 0\}$, so that $F \circ T = \theta \circ F$. If $I \subset (S')^\infty$ is θ-invariant, then $T^{-1}F^{-1}I = F^{-1}\theta^{-1}I = F^{-1}I$, and so $F^{-1}I$ is T-invariant in S. Assuming ξ to be ergodic, we obtain $P\{\eta \in I\} = P\{\xi \in F^{-1}I\} = 0$ or 1, which shows that even η is ergodic.

Conversely, let the sequence $(T^n\xi)$ be ergodic, and fix any T-invariant set I in S. Put $F = \{T^n;\, n \ge 0\}$, and define $A = \{s \in S^\infty;\, s_n \in I \text{ i.o.}\}$. Then $I = F^{-1}A$ and A is θ-invariant. Hence, $P\{\xi \in I\} = P\{(T^n\xi) \in A\} = 0$ or 1, which means that even ξ is ergodic. □

We may now state the fundamental a.s. and mean ergodic theorem for stationary sequences of random variables. Recall that $(S, \mathcal{S})$ denotes an arbitrary measurable space, and write $\mathcal{I}_\xi = \xi^{-1}\mathcal{I}$ for convenience.

Theorem 25.6 *(ergodic theorem, Birkhoff) Let ξ be a random element in S with distribution μ, and let T be a μ-preserving map*[4] *on S with invariant σ-field $\mathcal{I}$. Then for any measurable function $f \geq 0$ on S,*

$$n^{-1}\sum_{k<n} f(T^k\xi) \to E\Big\{f(\xi)\,\Big|\,\mathcal{I}_\xi\Big\} \quad a.s. \tag{1}$$

When $f \in L^p(\mu)$ for a $p \geq 1$, the convergence extends to L^p.

The proof is based on a simple but ingenious inequality.

Lemma 25.7 *(maximal ergodic lemma, Hopf) Let $\xi = (\xi_k)$ be a stationary sequence of integrable random variables, and put $S_n = \xi_1 + \cdots + \xi_n$. Then*

$$E\Big(\xi_1;\, \sup_{n\geq 1} S_n > 0\Big) \geq 0.$$

Proof (Garsia): Put $M_n = S_1 \vee \cdots \vee S_n$. Assuming ξ to be defined on the canonical space R^∞, we get

$$\begin{aligned} S_k &= \xi_1 + S_{k-1}\circ\theta \\ &\leq \xi_1 + (M_n\circ\theta)_+, \quad k = 1,\dots,n. \end{aligned}$$

Taking maxima yields $M_n \leq \xi_1 + (M_n\circ\theta)_+$ for all $n \in \mathsf{N}$, and so by stationarity

$$\begin{aligned} E\Big(\xi_1;\, M_n > 0\Big) &\geq E\Big\{M_n - (M_n\circ\theta)_+;\, M_n > 0\Big\} \\ &\geq E\Big\{(M_n)_+ - (M_n\circ\theta)_+\Big\} = 0. \end{aligned}$$

Since $M_n \uparrow \sup_n S_n$, the assertion follows by dominated convergence. □

Proof of Theorem 25.6 *(Yosida & Kakutani):* First let $f \in L^1$, and put $\eta_k = f(T^{k-1}\xi)$ for convenience. Since $E(\eta_1|\mathcal{I}_\xi)$ is an invariant function of ξ by Lemmas 1.14 and 25.3, the sequence $\zeta_k = \eta_k - E(\eta_1|\mathcal{I}_\xi)$ is again stationary. Writing $S_n = \zeta_1 + \cdots + \zeta_n$, we define for any $\varepsilon > 0$

$$A_\varepsilon = \Big\{\limsup_{n\to\infty}(S_n/n) > \varepsilon\Big\}, \qquad \zeta_n^\varepsilon = (\zeta_n - \varepsilon)1_{A_\varepsilon},$$

and note that the sums $S_n^\varepsilon = \zeta_1^\varepsilon + \cdots + \zeta_n^\varepsilon$ satisfy

$$\begin{aligned} \Big\{\sup\nolimits_n S_n^\varepsilon > 0\Big\} &= \Big\{\sup\nolimits_n (S_n^\varepsilon/n) > 0\Big\} \\ &= \Big\{\sup\nolimits_n (S_n/n) > \varepsilon\Big\} \cap A_\varepsilon = A_\varepsilon. \end{aligned}$$

Since $A_\varepsilon \in \mathcal{I}_\xi$, the sequence (ζ_n^ε) is stationary, and Lemma 25.7 yields

$$\begin{aligned} 0 &\leq E\Big(\zeta_1^\varepsilon;\, \sup\nolimits_n S_n^\varepsilon > 0\Big) \\ &= E\Big(\zeta_1 - \varepsilon;\, A_\varepsilon\Big) \\ &= E\Big\{E(\zeta_1|\mathcal{I}_\xi);\, A_\varepsilon\Big\} - \varepsilon P A_\varepsilon \\ &= -\varepsilon P A_\varepsilon, \end{aligned}$$

[4]When T is a P-preserving transformation directly on Ω, the result reduces to $n^{-1}\sum_{k<n}(\xi\circ T^k) \to E(\xi\,|\,\mathcal{I})$ a.s. for all random variables $\xi \geq 0$. The proof is of course the same. Similar remarks apply to all subsequent ergodic theorems.

which implies $PA_\varepsilon = 0$. Thus, $\limsup_n (S_n/n) \le \varepsilon$ a.s., and ε being arbitrary, we obtain $\limsup_n (S_n/n) \le 0$ a.s. Applying the same result to $-S_n$ yields $\liminf_n (S_n/n) \ge 0$ a.s., and so by combination $S_n/n \to 0$ a.s.

Now let $f \in L^p$ for a $p \ge 1$. Using Jensen's inequality and the stationarity of $T^k\xi$, we get for any $A \in \mathcal{A}$ and $r > 0$

$$\begin{aligned} E\, 1_A \Big| n^{-1} \sum_{k<n} f(T^k\xi) \Big|^p &\le n^{-1} \sum_{k<n} E\Big\{ \big| f(T^k\xi)\big|^p; A \Big\} \\ &\le r^p PA + E\Big\{ |f(\xi)|^p;\, |f(\xi)| > r \Big\}, \end{aligned}$$

which tends to 0 as $PA \to 0$ and then $r \to \infty$. Hence, by Lemma 5.10, the p-th powers on the left are uniformly integrable, and the asserted L^p-convergence follows by Proposition 5.12.

Finally, let $f \ge 0$ be arbitrary, and put $E\{f(\xi)|\mathcal{I}_\xi\} = \bar\eta$. Conditioning on the event $\{\bar\eta \le r\}$ for an arbitrary $r > 0$, we see that (1) holds a.s. on $\{\bar\eta < \infty\}$. Next, we have a.s. for any $r > 0$

$$\begin{aligned} \liminf_{n\to\infty} n^{-1} \sum_{k\le n} f(T^k\xi) &\ge \lim_{n\to\infty} n^{-1} \sum_{k \le n} \Big\{ f(T^k\xi) \wedge r \Big\} \\ &= E\Big\{ f(\xi) \wedge r \,\Big|\, \mathcal{I}_\xi \Big\}. \end{aligned}$$

As $r \to \infty$, the right-hand side tends a.s. to $\bar\eta$, by the monotone convergence property of conditional expectations. In particular, the left-hand side is a.s. infinite on $\{\bar\eta = \infty\}$, as required. □

Write $\mathcal{I}$ and $\mathcal{T}$ for the shift-invariant and tail σ-fields in R^∞, respectively, and note that $\mathcal{I} \subset \mathcal{T}$. Thus, for any sequence of random variables $\xi = (\xi_1, \xi_2, \ldots)$, we have $\mathcal{I}_\xi = \xi^{-1}\mathcal{I} \subset \xi^{-1}\mathcal{T}$. By Kolmogorov's 0−1 law, the latter σ-field is trivial when the ξ_n are independent. If they are even i.i.d. and integrable, then Theorem 25.6 yields $n^{-1}(\xi_1 + \cdots + \xi_n) \to E\xi_1$ a.s. and in L^1, conforming with Theorem 5.23. Hence, the last theorem contains the strong law of large numbers.

We may often allow the function $f = f_{n,k}$ in Theorem 25.6 to depend on n or k. Here we consider a slightly more general situation.

Corollary 25.8 *(varying functions, Maker) Let ξ be a random element in S with distribution μ, let T be a μ-preserving map on S with invariant σ-field $\mathcal{I}$, and let f and $f_{m,k}$ be measurable functions on S. Then*

(i) *if $f_{m,k} \to f$ a.s. and $\sup_{m,k} |f_{m,k}| \in L^1$, we have as $m, n \to \infty$*

$$n^{-1} \sum_{k<n} f_{m,k}(T^k\xi) \to E\Big\{ f(\xi) \,\Big|\, \mathcal{I}_\xi \Big\} \quad a.s.,$$

(ii) *when $f_{m,k} \to f$ in L^p for a $p \ge 1$, the convergence extends to L^p.*

Proof: (i) By Theorem 25.6 we may take $f = 0$. Then put $g_r = \sup_{m,k>r} |f_{m,k}|$, and conclude from the same result that a.s.

$$\begin{aligned} \limsup_{m,n\to\infty} \Big| n^{-1} \sum_{k<n} f_{m,k}(T^k\xi) \Big| &\le \lim_{n\to\infty} n^{-1} \sum_{k<n} g_r(T^k\xi) \\ &= E\Big\{ g_r(\xi) \,\Big|\, \mathcal{I}_\xi \Big\}. \end{aligned}$$

Here $g_r(\xi) \to 0$ a.s., and so by dominated convergence $E\{g_r(\xi) \,|\, \mathcal{I}_\xi\} \to 0$ a.s.

(ii) Assuming $f = 0$, we get by Minkowski's inequality and the invariance of μ
$$\Big\| n^{-1} \sum_{k<n} f_{m,k} \circ T^k \Big\|_p \le n^{-1} \sum_{k<n} \|f_{m,k}\|_p \to 0. \qquad \Box$$

To extend the ergodic theorem to continuous time, consider a family of transformations T_t on S, $t \ge 0$, satisfying the *semi-group* property $T_{s+t} = T_s T_t$. The latter is called a *flow* if it is also *measurable*, in the sense that the map $(x,t) \mapsto T_t x$ is product measurable from $S \times \mathsf{R}_+$ to S. The *invariant σ-field* $\mathcal{I}$ now consists of all sets $I \in \mathcal{S}$ with $T_t^{-1} I = I$ for all t. A random element ξ in S is said to be (T_t)*-stationary* if $T_t \xi \overset{d}{=} \xi$ for all $t \ge 0$.

Corollary 25.9 *(continuous-time ergodic theorem) Let ξ be a random element in S with distribution μ, and let (T_s) be a μ-preserving flow*[5] *on S with invariant σ-field $\mathcal{I}$. Then for any measurable function $f \ge 0$ on S,*
$$\lim_{t\to\infty} t^{-1} \int_0^t f(T_s \xi)\, ds = E\Big\{ f(\xi) \,\Big|\, \mathcal{I}_\xi \Big\} \quad a.s. \tag{2}$$
When $f \in L^p(\mu)$ for a $p \ge 1$, the convergence extends to L^p.

Proof: In both cases we may assume that $f \ge 0$. Writing $X_s = f(T_s \xi)$, we get by Jensen's inequality and Fubini's theorem
$$\begin{aligned} E\left| t^{-1} \int_0^t X_s\, ds \right|^p &\le E\, t^{-1} \int_0^t X_s^p\, ds \\ &= t^{-1} \int_0^t E X_s^p\, ds \\ &= E X_0^p < \infty. \end{aligned}$$
The required convergence now follows, as we apply Theorem 25.6 to the function $g(x) = \int_0^1 f(T_s x)\, ds$ and the discrete shift $T = T_1$.

To identify the limit, take $f \in L^1$, and introduce the invariant version
$$\bar{f}(\xi) = \lim_{r\to\infty} \limsup_{n\to\infty} n^{-1} \int_r^{r+n} f(T_s \xi)\, ds,$$
which is also $\mathcal{I}_\xi$-measurable. The stationarity of $T_s \xi$ yields $E^{\mathcal{I}_\xi} f(T_s \xi) = E^{\mathcal{I}_\xi} f(\xi)$ a.s. for all $s \ge 0$. Using Fubini's theorem, the L^1-convergence in (2), and the contraction property of conditional expectations, we get as $t \to \infty$
$$\begin{aligned} E^{\mathcal{I}_\xi} f(\xi) &= t^{-1} E^{\mathcal{I}_\xi} \int_0^t f(T_s \xi)\, ds \\ &\overset{P}{\to} E^{\mathcal{I}_\xi} \bar{f}(\xi) \\ &= \bar{f}(\xi), \end{aligned}$$
as required. The result extends as before to arbitrary $f \ge 0$. $\Box$

[5]When (T_s) is a P-preserving flow directly on Ω, the result reduces to $t^{-1}\int_0^t (\xi \circ T_s)\, ds \to E(\xi|\mathcal{I})$ a.s. for any random variable $\xi \ge 0$. The present version is more general.

We return to the case of a stationary sequence $\xi_1, \xi_2, \ldots$ of integrable random variables, and put $S_n = \sum_{k\le n} \xi_k$. Since S_n/n converges a.s. by Theorem 25.6, the maximum $M = \sup_n(S_n/n)$ is a.s. finite. The following result, relating the moments of ξ and M, is known as the *dominated ergodic theorem.* Here we write $\log_+ x = \log(x \vee 1)$, for convenience.

Proposition 25.10 *(maximum inequalities, Hardy & Littlewood, Wiener) Let $\xi = (\xi_k)$ be a stationary sequence of random variables, and put $S_n = \sum_{k\le n} \xi_k$ and $M = \sup_n(S_n/n)$. Then*

(i) $E|M|^p \lesssim E|\xi_1|^p, \ p > 1,$

(ii) $E\Big(|M| \log_+^m |M|\Big) \lesssim 1 + E\Big(|\xi_1| \log_+^{m+1} |\xi_1|\Big), \ m \ge 0.$

The proof requires a simple estimate related to Lemma 25.7.

Lemma 25.11 *(tail estimate) When $\xi = (\xi_k)$ is stationary in L^1,*

$$r\,P\Big\{\sup_n (S_n/n) > 2r\Big\} \le E\Big(\xi_1;\, \xi_1 > r\Big), \quad r > 0.$$

Proof: For any $r > 0$, put $\xi_k^r = \xi_k 1\{\xi_k > r\}$, and note that $\xi_k \le \xi_k^r + r$. Assuming ξ to be defined on the canonical space R^∞, and writing $A_n = S_n/n$, we get

$$\begin{aligned} A_n - 2r &= A_n \circ (\xi - 2r) \\ &\le A_n \circ (\xi^r - r), \end{aligned}$$

which implies $M - 2r \le M \circ (\xi^r - r)$. Applying Lemma 25.7 to the sequence $\xi^r - r$, we obtain

$$\begin{aligned} r\,P\{M > 2r\} &\le r\,P\Big\{M \circ (\xi^r - r) > 0\Big\} \\ &\le E\Big\{\xi_1^r;\, M \circ (\xi^r - r) > 0\Big\} \\ &\le E\xi_1^r = E\Big(\xi_1;\, \xi_1 > r\Big). \end{aligned}$$

□

Proof of Proposition 25.10: We may clearly assume that $\xi_1 \ge 0$ a.s.

(i) By Lemma 25.11, Fubini's theorem, and calculus,

$$\begin{aligned} E\,M^p &= p\,E\int_0^M r^{p-1} dr \\ &= p\int_0^\infty P\{M > r\}\, r^{p-1} dr \\ &\le 2p\int_0^\infty E\Big(\xi_1;\, 2\,\xi_1 > r\Big)\, r^{p-2}\, dr \\ &= 2p\,E\,\xi_1 \int_0^{2\xi_1} r^{p-2}\, dr \\ &= 2p\,(p-1)^{-1} E\Big\{\xi_1 (2\,\xi_1)^{p-1}\Big\} \\ &\lesssim E\,\xi_1^p. \end{aligned}$$

(ii) For $m = 0$, we may write

$$\begin{aligned}
EM - 1 &\le E(M-1)_+ \\
&= \int_1^\infty P\{M > r\}\,dr \\
&\le 2\int_1^\infty E\Big(\xi_1;\ 2\,\xi_1 > r\Big)\,r^{-1}dr \\
&= 2\,E\,\xi_1 \int_1^{2\,\xi_1 \vee 1} r^{-1}dr \\
&= 2\,E\Big(\xi_1 \log_+ 2\xi_1\Big) \\
&\le e + 2E\Big(\xi_1 \log 2\,\xi_1;\ 2\,\xi_1 > e\Big) \\
&\lesssim 1 + E\Big(\xi_1 \log_+ \xi_1\Big).
\end{aligned}$$

For $m > 0$, we may write instead

$$\begin{aligned}
E\Big(M \log_+^m M\Big) &= \int_0^\infty P\Big\{M \log_+^m M > r\Big\}\,dr \\
&= \int_1^\infty P\{M > t\}\,\Big(m \log^{m-1} t + \log^m t\Big)\,dt \\
&\le 2\int_1^\infty E\Big(\xi_1;\ 2\,\xi_1 > t\Big)\Big(m \log^{m-1} t + \log^m t\Big)\,t^{-1}dt \\
&= 2\,E\,\xi_1 \int_0^{\log_+ 2\xi_1} \Big(mx^{m-1} + x^m\Big)\,dx \\
&= 2\,E\,\xi_1 \left(\log_+^m 2\,\xi_1 + \frac{\log_+^{m+1} 2\,\xi_1}{m+1}\right) \\
&\le 2e + 4\,E\Big(\xi_1 \log^{m+1} 2\,\xi_1;\ 2\,\xi_1 > e\Big) \\
&\lesssim 1 + E\Big(\xi_1 \log_+^{m+1} \xi_1\Big).
\end{aligned}$$

□

Given a measure space $(S, \mathcal{S}, \mu)$, we introduce for any $m \ge 0$ the class $L \log^m L(\mu)$ of measurable functions f on S satisfying $\int |f| \log_+^m |f|\,d\mu < \infty$. Note in particular that $L \log^0 L = L^1$. Using the maximum inequalities of Proposition 25.10, we may prove the following multi-variate version of Theorem 25.6, for possibly non-commuting, measure-preserving transformations $T_1, \dots, T_d$.

Theorem 25.12 *(multi-variate ergodic theorem, Zygmund) Let ξ be a random element in S with distribution μ, let $T_1, \dots, T_d$ be μ-preserving maps on S with invariant σ-fields $\mathcal{I}_1, \dots, \mathcal{I}_d$, and put $\mathcal{J}_k = \xi^{-1}\mathcal{I}_k$. Then for any $f \in L \log^{d-1} L(\mu)$, we have as $n_1, \dots, n_d \to \infty$*

$$(n_1 \cdots n_d)^{-1} \sum_{k_1 < n_1} \cdots \sum_{k_d < n_d} f\Big(T_1^{k_1} \cdots T_d^{k_d} \xi\Big) \to E^{\mathcal{J}_d} \cdots E^{\mathcal{J}_1} f(\xi) \quad a.s. \tag{3}$$

When $f \in L^p(\mu)$ for a $p \ge 1$, the convergence extends to L^p.

Proof: Since $E\{f(\xi) \,|\, \mathcal{J}_k\} = \mu(f \,|\, \mathcal{I}_k) \circ \xi$ a.s., e.g. by Theorem 25.6, we may choose ξ to be the identity map on S. For $d = 1$, the result reduces to Theorem 25.6. Now assume the statement to be true up to dimension d. Proceeding by induction, consider any μ-preserving maps $T_1, \dots, T_{d+1}$ on S, and let $f \in L \log^d L$. By the induction hypothesis, the d-dimensional version of (3) holds as stated, and we may write the result in the form $f_m \to \bar{f}$ a.s., where $m = (n_1, \dots, n_d)$. Iterating Proposition 25.10, we also note that $\mu \sup_m |f_m| < \infty$. Hence, Corollary 25.8 (i) yields as $m, n \to \infty$

$$n^{-1} \sum_{k<n} f_m \circ T_{d+1}^k \to \mu\left(\bar{f} \,\middle|\, \mathcal{I}_{d+1}\right) \text{ a.s.},$$

as required. The proof of the L^p-version is similar. □

In the commutative case, the last result yields an interesting relationship between the associated conditional expectations. Let $L^1(\xi)$ denote the set of all integrable, ξ-measurable random variables.

Corollary 25.13 *(commuting maps and expectations) Let the maps $T_1, \dots, T_d$ in Theorem* 25.12 *commute, and put $\mathcal{J} = \bigcap_k \mathcal{J}_k$. Then*

$$E^{\mathcal{J}_1} \cdots E^{\mathcal{J}_d} = E^{\mathcal{J}} \quad on \ \ L^1(\xi).$$

Proof: Since even $T_1^{k_1}, \dots, T_d^{k_d}$ commute for arbitrary $k_1, \dots, k_d \in \mathsf{Z}_+$, Theorem 25.12 yields

$$E^{\mathcal{J}_1} \cdots E^{\mathcal{J}_d} f(\xi) = E^{\mathcal{J}_{p_1}} \cdots E^{\mathcal{J}_{p_d}} f(\xi) \text{ a.s.} \tag{4}$$

for any measurable function $f \geq 0$ on S and permutation $p_1, \dots, p_d$ of $1, \dots, d$. In particular, the expression in (4) is a.s. $\mathcal{J}_k$-measurable for every k, and therefore $\mathcal{J}$-measurable. It remains to note that

$$E\left\{E^{\mathcal{J}_1} \cdots E^{\mathcal{J}_d} f(\xi)\,; A\right\} = E\left\{f(\xi)\,; A\right\}, \quad A \in \mathcal{J}.$$

□

For commuting maps $T_1, \dots, T_d$ on S, the compositions $T^k = T_1^{k_1} \cdots T_d^{k_d}$ form a d-dimensional semi-group indexed by Z_+^d. Similarly, when $(T_1^s), \dots, (T_d^s)$ are commuting flows on S, the compositions $T^s = T_1^{s_1} \cdots T_d^{s_d}$ form a d-dimensional, measurable semi-group or flow indexed by R_+^d. In the continuous parameter case, it may be more natural to consider flows indexed by R^d, corresponding to the case of stationary processes on R^d. In this context, one may also want to average over more general sets than rectangles. Here we consider a basic ergodic theorem for increasing sequences of convex sets. Given such a set B, we define the *inner radius* $r(B)$ as the radius of the largest open ball contained in B.

Theorem 25.14 *(convex averages, Wiener) Consider a random element ξ in S with distribution μ and a flow of μ-preserving maps T_s on S, $s \in \mathsf{R}^d$, with invariant σ-field $\mathcal{I}$, and let $B_1 \subset B_2 \subset \cdots$ be bounded, convex sets in $\mathcal{B}^d$ with $r(B_n) \to \infty$. Then for any measurable function $f \geq 0$ on S,*

$$(\lambda^d B_n)^{-1} \int_{B_n} f(T_s \xi)\, ds \to E\left\{f(\xi) \,\middle|\, \mathcal{I}_\xi\right\} \quad a.s.$$

When $f \in L^p(\mu)$ for a $p \geq 1$, the convergence extends to L^p.

Several lemmas will be needed for the proof, beginning with a simple geometric estimate.

Lemma 25.15 *(space filling) For any bounded, convex sets $B_1 \subset \cdots \subset B_m$ in $\mathcal{B}^d$ with $\lambda^d B_1 > 0$, a bounded set $K \in \mathcal{B}^d$, and a function $h: K \to \{1, \ldots, m\}$, there exists a finite subset $M \subset K$, such that the sets $B_{h(x)} + x$, $x \in M$, are disjoint and satisfy*

$$\lambda^d K \le \binom{2d}{d} \sum_{x \in M} \lambda^d B_{h(x)}.$$

Proof: Put $C_x = B_{h(x)} + x$, and choose $x_1, x_2, \ldots \in K$ recursively, as follows. Once $x_1, \ldots, x_{j-1}$ have been selected, choose $x_j \in K$ with the largest possible $h(x)$, such that $C_{x_i} \cap C_{x_j} = \emptyset$ for all $i < j$. The construction terminates when no such x_j exists. Put $M = \{x_i\}$, and note that the sets C_x with $x \in M$ are disjoint. Now fix any $y \in K$. By the construction of M, we have $C_x \cap C_y \neq \emptyset$ for some $x \in M$ with $h(x) \ge h(y)$, and so

$$\begin{aligned} y &\in B_{h(x)} - B_{h(y)} + x \\ &\subset B_{h(x)} - B_{h(x)} + x. \end{aligned}$$

Hence, $K \subset \bigcup_{x \in M} (B_{h(x)} - B_{h(x)} + x)$, and so by Lemma A6.5 (i),

$$\begin{aligned} \lambda^d K &\le \sum_{x \in M} \lambda^d \Big(B_{h(x)} - B_{h(x)} \Big) \\ &\le \binom{2d}{d} \sum_{x \in M} \lambda^d B_{h(x)}. \end{aligned}$$

□

Next we prove a multi-variate version of Lemma 25.11, stated for convenience in terms of random measures. For motivation, we note that the set function $\eta B = \int_B f(T_s \xi)\, ds$ in Theorem 25.14 is a stationary random measure on R^d, and that the *intensity* m of η, defined by the relation $E\eta = m\lambda^d$, is equal to $Ef(\xi)$.

Lemma 25.16 *(maximum inequality) Let ξ be a stationary random measure on R^d with intensity m, and let $B_1 \subset B_2 \subset \cdots$ be bounded, convex sets in $\mathcal{B}^d$ with $\lambda^d B_1 > 0$. Then*

$$r\, P\Big\{ \sup_k \frac{\xi B_k}{\lambda^d B_k} > r \Big\} \le m \binom{2d}{d}, \quad r > 0.$$

Proof: Fix any $r, a > 0$ and $n \in \mathsf{N}$, and define a process ν on R^d and a random set K in $S_a = \{x \in \mathsf{R}^d;\ |x| \le a\}$ by

$$\begin{aligned} \nu(x) &= \inf \Big\{ k \in \mathsf{N};\ \xi(B_k + x) > r\, \lambda^d B_k \Big\}, \quad x \in \mathsf{R}^d, \\ K &= \Big\{ x \in S_a;\ \nu(x) \le n \Big\}. \end{aligned}$$

By Lemma 25.15 we may choose a finite, random subset $M \subset K$, such that the sets $B_{\nu(x)} + x$, $x \in M$, are disjoint and $\lambda^d K \le \binom{2d}{d} \sum_{x \in M} \lambda^d B_{\nu(x)}$. Writing $b = \sup\{|x|;\ x \in B_n\}$, we get

$$\begin{aligned}\xi S_{a+b} &\geq \sum_{x\in M} \xi\Big(B_{\nu(x)} + x\Big)\\ &\geq r \sum_{x\in M} \lambda^d B_{\nu(x)}\\ &\geq r \binom{2d}{d}^{-1} \lambda^d K.\end{aligned}$$

Taking expectations and using Fubini's theorem and the stationarity and measurability of ν, we obtain

$$\begin{aligned}m \binom{2d}{d} \lambda^d S_{a+b} &\geq rE\lambda^d K\\ &= r\int_{S_a} P\Big\{\nu(x) \leq n\Big\}\, dx\\ &= r\lambda^d S_a\, P\Big\{\max_{k\leq n} \frac{\xi B_k}{\lambda^d B_k} > r\Big\}.\end{aligned}$$

Now divide by $\lambda^d S_a$, and then let $a \to \infty$ and $n \to \infty$ in this order. □

We finally need an elementary Hilbert-space result. By a *contraction* on a Hilbert space H we mean a linear operator T, such that $\|Tx\| \leq \|x\|$ for all $x \in H$. For any linear subspace $M \subset H$, write $M^\perp$ for the orthogonal complement and $\bar{M}$ for the closure of M. The *adjoint* T^* of an operator T is determined by the identity $\langle x, Ty\rangle = \langle T^*x, y\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product in H.

Lemma 25.17 *(invariant subspace) Let $\mathcal{T}$ be a set of contraction operators on a Hilbert space H, and write N for the $\mathcal{T}$-invariant subspace of H. Then*

$$N^\perp \subset \bar{R}, \qquad R = \operatorname{span}\Big\{x - Tx;\ x \in H,\, T \in \mathcal{T}\Big\}.$$

Proof: If $x \perp R$,

$$\langle\, x - T^*x,\, y\rangle = \langle\, x,\, y - Ty\rangle = 0, \quad T \in \mathcal{T},\ \ y \in H,$$

which implies $T^*x = x$ for every $T \in \mathcal{T}$. Hence, for any $T \in \mathcal{T}$, we have $\langle Tx, x\rangle = \langle x, T^*x\rangle = \|x\|^2$, and so by contraction

$$\begin{aligned}0 &\leq \|Tx - x\|^2\\ &= \|Tx\|^2 + \|x\|^2 - 2\langle Tx, x\rangle\\ &\leq 2\|x\|^2 - 2\|x\|^2 = 0,\end{aligned}$$

which implies $Tx = x$. This gives $R^\perp \subset N$, and so $N^\perp \subset (R^\perp)^\perp = \bar{R}$. □

Proof of Theorem 25.14: First let $f \in L^1$, and define

$$\tilde{T}_s f = f \circ T_s, \qquad A_n = (\lambda^d B_n)^{-1} \int_{B_n} \tilde{T}_s ds.$$

For any $\varepsilon > 0$, Lemma 25.17 yields a measurable decomposition

$$f = f^\varepsilon + \sum_{k\leq m} \Big(g_k^\varepsilon - \tilde{T}_{s_k} g_k^\varepsilon\Big) + h^\varepsilon,$$

where $f^\varepsilon \in L^2$ is $\tilde{T}_s$-invariant for all $s \in \mathsf{R}^d$, the functions $g_1^\varepsilon, \dots, g_m^\varepsilon$ are bounded, and $E|h^\varepsilon(\xi)| < \varepsilon$. Here clearly $A_n f^\varepsilon \equiv f^\varepsilon$. Using Lemma A6.5 (ii), we get as $n \to \infty$ for fixed $k \le m$ and $\varepsilon > 0$

$$\begin{aligned}\left\| A_n\left(g_k^\varepsilon - \tilde{T}_{s_k} g_k^\varepsilon\right)\right\| &\le \frac{\lambda^d\left\{(B_n + s_k) \,\triangle\, B_n\right\}}{\lambda^d B_n} \|g_k^\varepsilon\| \\ &\le 2\left\{\left(1 + |s_k|/r(B_n)\right)^d - 1\right\} \|g_k^\varepsilon\| \to 0.\end{aligned}$$

Finally, Lemma 25.16 gives

$$r\, P\left\{\sup\nolimits_n A_n |h^\varepsilon(\xi)| \ge r\right\} \le \binom{2d}{d} E|h^\varepsilon(\xi)| \le \binom{2d}{d}\varepsilon, \quad r, \varepsilon > 0,$$

and so $\sup_n A_n |h^\varepsilon(\xi)| \xrightarrow{P} 0$ as $\varepsilon \to 0$. In particular, $\liminf_n A_n f(\xi) < \infty$ a.s., which ensures that

$$\begin{aligned}\Big(\limsup_{n\to\infty} - \liminf_{n\to\infty}\Big) A_n f(\xi) &= \Big(\limsup_{n\to\infty} - \liminf_{n\to\infty}\Big) A_n h^\varepsilon(\xi) \\ &\le 2 \sup\nolimits_n A_n |h^\varepsilon(\xi)| \xrightarrow{P} 0.\end{aligned}$$

Thus, the left-hand side vanishes a.s., and the required a.s. convergence follows.

When $f \in L^p$ for a $p \ge 1$, the asserted L^p-convergence follows as before, by the uniform integrability of the powers $|A_n f(\xi)|^p$. We may now identify the limit, as in the proof of Corollary 25.9, and the a.s. convergence extends to arbitrary $f \ge 0$, as in case of Theorem 25.6. □

The L^p-version of Theorem 25.14 remains valid under weaker conditions. For a simple extension, say that the distributions μ_n on R^d are *asymptotically invariant* if $\|\mu_n - \mu_n * \delta_s\| \to 0$ for every $s \in \mathsf{R}^d$, where $\|\cdot\|$ denotes the total variation norm. Note that the conclusion of Theorem 25.14 can be written as $\mu_n X \to \bar{X}$ a.s., where

$$\mu_n = \frac{1_{B_n}\lambda^d}{\lambda^d B_n}, \qquad X_s = f(T_s \xi), \qquad \bar{X} = E\left\{f(\xi) \,\middle|\, \mathcal{I}_\xi\right\}.$$

Corollary 25.18 *(mean ergodic theorem) Let X be a stationary, measurable, and L^p-valued process on R^d, where $p \ge 1$. Then for any asymptotically invariant distributions μ_n on R^d,*

$$\mu_n X \to \bar{X} \equiv E\left(X_s \,\middle|\, \mathcal{I}_X\right) \text{ in } L^p.$$

Proof: By Theorem 25.14 we may choose some distributions ν_m on R^d with $\nu_m X \to \bar{X}$ in L^p. Using Minkowski's inequality and its extension in Corollary 1.32, along with the stationarity of X, the invariance of $\bar{X}$, and dominated convergence, we get as $n \to \infty$ and then $m \to \infty$

$$\begin{aligned}\left\| \mu_n X - \bar{X}\right\|_p &\le \left\|\mu_n X - (\mu_n * \nu_m) X\right\|_p + \left\|(\mu_n * \nu_m) X - \bar{X}\right\|_p \\ &\le \left\| \mu_n - \mu_n * \nu_m\right\| \|X\|_p + \int \left\|(\delta_s * \nu_m) X - \bar{X}\right\|_p \mu_n(ds) \\ &\le \|X\|_p \int \left\| \mu_n - \mu_n * \delta_t\right\| \nu_m(dt) + \left\|\nu_m X - \bar{X}\right\|_p \to 0. \qquad \square\end{aligned}$$

We turn to a sub-additive version of Theorem 25.6. For motivation and subsequent needs, we begin with a simple result for non-random sequences. A sequence $c_1, c_2, \ldots \in \mathsf{R}$ is said to be *sub-additive*, if $c_{m+n} \leq c_m + c_n$ for all $m, n \in \mathsf{N}$.

Lemma 25.19 *(sub-additivity) For any sub-additive sequence* $c_1, c_2, \ldots \in \mathsf{R}$,

$$\lim_{n\to\infty} \frac{c_n}{n} = \inf_n \frac{c_n}{n} \in [-\infty, \infty).$$

Proof: Iterating the sub-additivity relation, we get for any $k, n \in \mathsf{N}$

$$\begin{aligned} c_n &\leq [n/k]c_k + c_{n-k[n/k]} \\ &\leq [n/k]c_k + c_0 \vee \cdots \vee c_{k-1}, \end{aligned}$$

where $c_0 = 0$. Noting that $[n/k] \sim n/k$ as $n \to \infty$, we get $\limsup_n (c_n/n) \leq c_k/k$ for all k, and so

$$\begin{aligned} \inf_n \frac{c_n}{n} &\leq \liminf_{n\to\infty} \frac{c_n}{n} \\ &\leq \limsup_{n\to\infty} \frac{c_n}{n} \leq \inf_n \frac{c_n}{n}. \end{aligned}$$ □

For two-dimensional arrays c_{jk}, $0 \leq j < k$, sub-additivity is defined by $c_{0,n} \leq c_{0,m} + c_{m,n}$ for all $m < n$, which reduces to the one-dimensional notion when $c_{jk} = c_{k-j}$ for some sequence c_k. Note that sub-additivity holds automatically for arrays of the form $c_{jk} = a_{j+1} + \cdots + a_k$.

We now extend the ergodic theorem to sub-additive arrays of random variables ξ_{jk}, $0 \leq j < k$. For motivation, recall from Theorem 25.6 that, if $\xi_{jk} = \eta_{j+1} + \cdots + \eta_k$ for a stationary sequence of integrable random variables η_k, then $\xi_{0,n}/n$ converges a.s. and in L^1. A similar result holds for general sub-additive arrays (ξ_{jk}), stationary under simultaneous shifts in the two indices, so that $(\xi_{j+1,\,k+1}) \overset{d}{=} (\xi_{j,k})$. To allow a wider range of applications, we introduce the slightly weaker hypotheses

$$\Big(\xi_{k,2k}, \xi_{2k,3k}, \ldots\Big) \overset{d}{=} \Big(\xi_{0,k}, \xi_{k,2k}, \ldots\Big), \quad k \in \mathsf{N}, \tag{5}$$

$$\Big(\xi_{k,k+1}, \xi_{k,k+2}, \ldots\Big) \overset{d}{=} \Big(\xi_{0,1}, \xi_{0,2}, \ldots\Big), \quad k \in \mathsf{N}. \tag{6}$$

For convenience, we also restate the sub-additivity condition

$$\xi_{0,n} \leq \xi_{0,m} + \xi_{m,n}, \quad 0 < m < n. \tag{7}$$

Theorem 25.20 *(sub-additive ergodic theorem, Kingman, Liggett) Let* (ξ_{jk}) *be a sub-additive array of random variables with* $E\xi_{0,1}^+ < \infty$, *satisfying* (5)−(6). *Then*

(i) $n^{-1}\xi_{0,n} \to \bar\xi$ *a.s. in* $[-\infty, \infty)$, *where* $E\bar\xi = \inf_n (E\xi_{0,n}/n) \equiv c$,

(ii) *the convergence in* (i) *extends to* L^1 *when* $c > -\infty$,

(iii) $\bar\xi$ *is a.s. constant when the sequences in* (5) *are ergodic.*

Proof: Put $\xi_{0,n} = \xi_n$ for convenience. By (6) and (7) we have $E\xi_n^+ \le nE\xi_1^+ < \infty$. First let $c > -\infty$, so that the variables $\xi_{m,n}$ are integrable. Iterating (7) gives

$$\frac{\xi_n}{n} \le \sum_{j=1}^{[n/k]} \xi_{(j-1)k,\,jk}/n + \sum_{j=k[n/k]+1}^{n} \xi_{j-1,\,j}/n, \quad n,k \in \mathsf{N}. \tag{8}$$

By (5) the sequence $\xi_{(j-1)k,\,jk}$, $j \in \mathsf{N}$, is stationary for fixed k, and so Theorem 25.6 yields $n^{-1}\sum_{j\le n}\xi_{(j-1)k,\,jk} \to \bar{\xi}_k$ a.s. and in L^1, where $E\bar{\xi}_k = E\xi_k$. Hence, the first term in (8) tends to $\bar{\xi}_k/k$ a.s. and in L^1. Similarly, $n^{-1}\sum_{j\le n}\xi_{j-1,\,j} \to \bar{\xi}_1$ a.s. and in L^1, and so the second term in (8) tends in the same sense to 0. Thus, the right-hand side converges a.s. and in L^1 toward $\bar{\xi}_k/k$, and since k is arbitrary, we get

$$\begin{aligned}\limsup_{n\to\infty}(\xi_n/n) &\le \inf\nolimits_n\big(\bar{\xi}_n/n\big)\\ &\equiv \bar{\xi} < \infty \;\text{ a.s.}\end{aligned} \tag{9}$$

The variables ξ_n^+/n are uniformly integrable by Proposition 5.12, and moreover

$$\begin{aligned}E\limsup_{n\to\infty}(\xi_n/n) &\le E\bar{\xi}\\ &\le \inf\nolimits_n\big(E\bar{\xi}_n/n\big)\\ &= \inf\nolimits_n(E\xi_n/n) = c.\end{aligned} \tag{10}$$

To derive a lower bound, let $\kappa_n \perp\!\!\!\perp (\xi_{jk})$ be uniformly distributed over $\{1,\dots,n\}$ for each n, and define

$$\begin{aligned}\zeta_k^n &= \xi_{\kappa_n,\,\kappa_n+k},\\ \eta_k^n &= \xi_{\kappa_n+k} - \xi_{\kappa_n+k-1}, \qquad k \in \mathsf{N}.\end{aligned}$$

Then by (6),

$$(\zeta_1^n, \zeta_2^n, \dots) \overset{d}{=} (\xi_1, \xi_2, \dots), \quad n \in \mathsf{N}. \tag{11}$$

Moreover, $\eta_k^n \le \xi_{\kappa_n+k-1,\,\kappa_n+k} \overset{d}{=} \xi_1$ by (6) and (7), and so the variables $(\eta_k^n)^+$ are uniformly integrable. On the other hand, the sequence $E\xi_1, E\xi_2, \dots$ is sub-additive, and so by Lemma 25.19 we have as $n \to \infty$

$$\begin{aligned}E\eta_k^n &= n^{-1}\big(E\xi_{n+k} - E\xi_k\big)\\ &\to \inf\nolimits_n(E\xi_n/n) = c, \quad k \in \mathsf{N}.\end{aligned} \tag{12}$$

In particular, $\sup_n E|\eta_k^n| < \infty$, which shows that the sequence $\eta_k^1, \eta_k^2, \dots$ is tight for each k. Hence, by Theorems 5.30, 6.20, and 8.21, there exist some random variables ζ_k and η_k such that

$$\big(\zeta_1^n, \zeta_2^n, \dots;\, \eta_1^n, \eta_2^n, \dots\big) \xrightarrow{d} \big(\zeta_1, \zeta_2, \dots;\, \eta_1, \eta_2, \dots\big) \tag{13}$$

along a sub-sequence. Here $(\zeta_k) \overset{d}{=} (\xi_k)$ by (11), and so by Theorem 8.17 we may assume that $\zeta_k = \xi_k$ for each k.

The sequence $\eta_1, \eta_2, \dots$ is clearly stationary, and by Lemma 5.11 it is also integrable. From (7) we get

$$\begin{aligned}\eta_1^n + \cdots + \eta_k^n &= \xi_{\kappa_n+k} - \xi_{\kappa_n}\\ &\le \xi_{\kappa_n,\,\kappa_n+k} = \zeta_k^n,\end{aligned}$$

and so in the limit $\eta_1 + \cdots + \eta_k \le \xi_k$ a.s. Hence, Theorem 25.6 yields

$$\xi_n/n \ge n^{-1}\sum_{k\le n}\eta_k \to \bar\eta \ \text{ a.s. and in } L^1,$$

for some $\bar\eta \in L^1$. In particular, the variables ξ_n^-/n are uniformly integrable, and so the same thing is true for ξ_n/n. Using Lemma 5.11 and the uniform integrability of the variables $(\eta_k^n)^+$, together with (10) and (12), we get

$$\begin{aligned}c &= \limsup_{n\to\infty} E\eta_1^n\\ &\le E\eta_1 = E\bar\eta\\ &\le E\liminf_{n\to\infty}(\xi_n/n)\\ &\le E\limsup_{n\to\infty}(\xi_n/n)\\ &\le E\bar\xi \le c.\end{aligned}$$

Thus, ξ_n/n converges a.s., and by (9) the limit equals $\bar\xi$. Furthermore, by Lemma 5.11, the convergence holds even in L^1, and $E\bar\xi = c$. If the sequences in (5) are ergodic, then $\bar\xi_n = E\xi_n$ a.s. for each n, and we get $\bar\xi = c$ a.s.

Now assume instead that $c = -\infty$. Then for each $r \in \mathsf{Z}$, the truncated array $\xi_{m,n} \vee r(n-m)$, $0 \le m < n$, satisfies the hypotheses of the theorem with c replaced by $c^r = \inf_n(E\xi_n^r/n) \ge r$, where $\xi_n^r = \xi_n \vee rn$. Thus, $\xi_n^r/n = (\xi_n/n) \vee r$ converges a.s. toward a random variable $\bar\xi^r$ with mean c^r, and so $\xi_n/n \to \inf_r \bar\xi^r \equiv \bar\xi$. Finally, $E\bar\xi \le \inf_r c^r = c = -\infty$ by monotone convergence. □

For an application of the last result, we derive a celebrated ergodic theorem for products of random matrices.

Theorem 25.21 *(random matrices, Furstenberg & Kesten) Let $X^k = (\xi_{ij}^k)$ be a stationary sequence of random $d\times d$ matrices, such that*

$$\xi_{ij}^k > 0 \ \text{ a.s.}, \quad E\left|\log \xi_{ij}^k\right| < \infty, \qquad i,j \le d.$$

Then there exists a random variable η, independent of i,j, such that as $n\to\infty$,

$$n^{-1}\log\left(X^1\cdots X^n\right)_{ij} \to \eta \ \text{ a.s. and in } L^1, \quad i,j\le d.$$

Proof: First let $i = j = 1$, and define

$$\eta_{m,n} = \log\left(X^{m+1}\cdots X^n\right)_{11}, \quad 0 \le m < n.$$

The array $(-\eta_{m,n})$ is clearly sub-additive and jointly stationary, and $E|\eta_{0,1}| < \infty$ by hypothesis. Further note that

$$\left(X^1\cdots X^n\right)_{11} \le d^{\,n-1}\prod_{k\le n}\max_{i,j}\xi_{ij}^k.$$

Hence,

$$\begin{aligned}\eta_{0,n} - (n-1)\log d &\le \sum_{k\le n} \log \max_{i,j} \xi_{ij}^k \\ &\le \sum_{k\le n} \sum\nolimits_{i,j} \Big| \log \xi_{ij}^k \Big|,\end{aligned}$$

and so

$$n^{-1} E\eta_{0,n} \le \log d + \sum\nolimits_{i,j} E\Big| \log \xi_{ij}^1 \Big| < \infty.$$

Thus, by Theorem 25.20 and its proof, we have $\eta_{0,n}/n \to \eta$ a.s. and in L^1 for an invariant random variable η.

To extend the convergence to arbitrary $i, j \in \{1, \dots, d\}$, we may write for any $n \in \mathsf{N}$

$$\begin{aligned}\xi_{i1}^2 \Big(X^3 \cdots X^n\Big)_{11} \xi_{1j}^{n+1} &\le \Big(X^2 \cdots X^{n+1}\Big)_{ij} \\ &\le \Big(\xi_{1i}^1 \xi_{j1}^{n+2}\Big)^{-1} \Big(X^1 \cdots X^{n+2}\Big)_{11}.\end{aligned}$$

Noting that $n^{-1} \log \xi_{ij}^n \to 0$ a.s. and in L^1 by Theorem 25.6, and using the stationarity of (X^n) and the invariance of ξ, we obtain $n^{-1} \log (X^2 \cdots X^{n+1})_{ij} \to \eta$ a.s. and in L^1. The desired convergence now follows by stationarity. □

We turn to the decomposition of an invariant distribution into ergodic components. For motivation, consider the setting of Theorem 25.6 or 25.14, and let S be Borel, to ensure the existence of regular conditional distributions. Writing $\eta = \mathcal{L}(\xi \,|\, \mathcal{I}_\xi)$, we get

$$\begin{aligned}\mathcal{L}(\xi) &= E\mathcal{L}(\xi \,|\, \mathcal{I}_\xi) = E\eta \\ &= \int m\, P\{\eta \in dm\}.\end{aligned} \tag{14}$$

Furthermore,

$$\begin{aligned}\eta I &= P\Big\{\xi \in I \,\Big|\, \mathcal{I}_\xi\Big\} \\ &= 1\{\xi \in I\} \text{ a.s.,} \quad I \in \mathcal{I},\end{aligned}$$

and so $\eta I = 0$ or 1 a.s. for any fixed $I \in \mathcal{I}$. If we can choose the exceptional null set to be independent of I, then η is a.s. ergodic, and (14) gives the desired ergodic decomposition of $\mu = \mathcal{L}(\xi)$. Though the suggested statement is indeed true, its proof requires a different approach.

Proposition 25.22 *(ergodicity by conditioning, Farrell, Varadarajan) Let ξ be a random element in a Borel space S with distribution μ, and consider a measurable group G of μ-preserving maps T_s on S, $s \in \mathsf{R}^d$, with invariant σ-field $\mathcal{I}$. Then $\eta = \mathcal{L}(\xi \,|\, \mathcal{I}_\xi)$ is a.s. invariant and ergodic under G.*

For the proof, consider an increasing sequence of convex sets $B_n \in \mathcal{B}^d$ with $r(B_n) \to \infty$, and introduce on S the probability kernels

$$\mu_n(x, A) = (\lambda^d B_n)^{-1} \int_{B_n} 1_A(T_s x)\, ds, \quad x \in S,\ A \in \mathcal{S}, \tag{15}$$

along with the *empirical distributions* $\eta_n = \mu_n(\xi, \cdot)$. By Theorem 25.14, we have $\eta_n f \to \eta f$ a.s. for every bounded, measurable function f on S, where $\eta = \mathcal{L}(\xi \,|\, \mathcal{I}_\xi)$. Say that a class $\mathcal{C} \subset \mathcal{S}$ is *measure-determining*, if every distribution on S is uniquely determined by its values on $\mathcal{C}$.

Lemma 25.23 *(degenerate limit) Let $A_1, A_2, \ldots \in \mathcal{S}$ be measure-determining with empirical distributions $\eta_n = \mu_n(\xi, \cdot)$, where the μ_n are given by* (15)*. Then ξ is ergodic whenever*

$$\eta_n A_k \to P\{\xi \in A_k\} \text{ a.s.}, \quad k \in \mathbb{N}.$$

Proof: By Theorem 25.14, we have $\eta_n A \to \eta A \equiv P(\xi \in A \mid \mathcal{I}_\xi)$ a.s. for every $A \in \mathcal{S}$, and so by comparison $\eta A_k = P\{\xi \in A_k\}$ a.s. for all k. Since the A_k are measure-determining, we get $\eta = \mathcal{L}(\xi)$ a.s. Hence, for any $I \in \mathcal{I}$ we have a.s.

$$\begin{aligned} P\{\xi \in I\} = \eta I &= P\Big\{\xi \in I \,\Big|\, \mathcal{I}_\xi\Big\} \\ &= 1_I(\xi) \in \{0,1\}, \end{aligned}$$

which implies $P\{\xi \in I\} = 0$ or 1. □

Proof of Proposition 25.22: By the stationarity of ξ, we have for any $A \in \mathcal{S}$ and $s \in \mathbb{R}^d$

$$\begin{aligned} \eta \circ T_s^{-1} A &= P\Big\{T_s \xi \in A \,\Big|\, \mathcal{I}_\xi\Big\} \\ &= P\Big\{\xi \in A \,\Big|\, \mathcal{I}_\xi\Big\} = \eta A \text{ a.s.} \end{aligned}$$

Since S is Borel, we obtain $\eta \circ T_s^{-1} = \eta$ a.s. for every s. Now put $C = [0,1]^d$, and define $\bar\eta = \int_C (\eta \circ T_s^{-1})\, ds$. Since η is a.s. invariant under shifts in $\mathbb{Z}^d$, the variable $\bar\eta$ is a.s. invariant under arbitrary shifts. Furthermore, Fubini's theorem yields

$$\lambda^d \Big\{s \in [0,1]^d;\ \eta \circ T_s^{-1} = \eta\Big\} = 1 \text{ a.s.},$$

and therefore $\bar\eta = \eta$ a.s. Thus, η is a.s. G-invariant.

Now choose a measure-determining sequence $A_1, A_2, \ldots \in \mathcal{S}$, which exists since S is Borel. Noting that $\eta_n A_k \to \eta A_k$ a.s. for every k by Theorem 25.14, we get by Theorem 8.5

$$\eta \textstyle\bigcap_k \Big\{x \in S;\ \mu_n(x, A_k) \to \eta A_k\Big\} = P^{\mathcal{I}_\xi} \textstyle\bigcap_k \Big\{\eta_n A_k \to \eta A_k\Big\} = 1 \text{ a.s.}$$

Since η is a.s. a G-invariant distribution on S, Lemma 25.23 applies for every $\omega \in \Omega$ outside a P-null set, and we conclude that η is a.s. ergodic. □

In (14) we saw that, for stationary ξ, the distribution $\mu = \mathcal{L}(\xi)$ is a mixture of invariant, ergodic probability measures. We show that this representation is unique[6] and characterizes the ergodic measures as extreme points in the convex set of invariant measures.

To explain the terminology, we say that a subset M of a linear space is *convex* if $c\, m_1 + (1-c)\, m_2 \in M$ for all $m_1, m_2 \in M$ and $c \in (0,1)$. In that case, an element $m \in M$ is *extreme* if for any m_1, m_2, c as above, the relation $m = c\, m_1 + (1-c)\, m_2$ implies $m_1 = m_2 = m$. With any set of measures μ on a measurable space $(S, \mathcal{S})$, we associate the σ-field generated by all evaluation maps $\pi_B : \mu \mapsto \mu B$, $B \in \mathcal{S}$.

[6] Under stronger conditions, a more explicit representation was obtained in Theorem 3.12.

Theorem 25.24 *(ergodic decomposition, Krylov & Bogolioubov) Let $G = \{T_s;\, s \in \mathsf{R}^d\}$ be a measurable group of transformations on a Borel space S. Then*

(i) *the G-invariant distributions on S form a convex set M, whose extreme points are precisely the ergodic measures in M,*

(ii) *every measure $\mu \in M$ has a unique representation $\mu = \int m\,\nu(dm)$, with ν restricted to the set of ergodic measures in M.*

Proof: The set M is clearly convex, and Proposition 25.22 yields for every $\mu \in M$ a representation $\mu = \int m\,\nu(dm)$, where ν is a probability measure on the set of ergodic measures in M. To see that ν is unique, we introduce a regular conditional distribution $\eta = \mu(\,\cdot\,|\,\mathcal{I})$ a.s. μ on S, and note that $\mu_n A \to \eta A$ a.s. μ for all $A \in \mathcal{S}$ by Theorem 25.14. Thus, for any $A_1, A_2, \ldots \in \mathcal{S}$, we have

$$m \bigcap\nolimits_k \Big\{x \in S;\, \mu_n(x, A_k) \to \eta(x, A_k)\Big\} = 1 \text{ a.e. } \nu.$$

The same relation holds with $\eta(x, A_k)$ replaced by mA_k, since ν is restricted to the class of ergodic measures in M. Assuming the sets A_k to be measure-determining, we conclude that $m\{x;\, \eta(x,\cdot) = m\} = 1$ a.e. ν. Hence, for any measurable set $A \subset M$,

$$\begin{aligned} \mu\{\eta \in A\} &= \int m\{\eta \in A\}\,\nu(dm) \\ &= \int 1_A(m)\,\nu(dm) = \nu A, \end{aligned}$$

which shows that $\nu = \mu \circ \eta^{-1}$.

To prove the equivalence of ergodicity and extremality, fix any measure $\mu \in M$ with ergodic decomposition $\int m\,\nu(dm)$. First let μ be extreme. If it is not ergodic, then ν is non-degenerate, and we have $\nu = c\,\nu_1 + (1-c)\,\nu_2$ for some $\nu_1 \perp \nu_2$ and $c \in (0,1)$. Since μ is extreme, we obtain $\int m\,\nu_1(dm) = \int m\,\nu_2(dm)$, and so $\nu_1 = \nu_2$ by the uniqueness of the decomposition. The contradiction shows that μ is ergodic.

Next let μ be ergodic, so that $\nu = \delta_\mu$, and let $\mu = c\,\mu_1 + (1-c)\,\mu_2$ with $\mu_1, \mu_2 \in M$ and $c \in (0,1)$. If $\mu_i = \int m\,\nu_i(dm)$ for $i = 1, 2$, then $\delta_\mu = c\,\nu_1 + (1-c)\,\nu_2$ by the uniqueness of the decomposition. Hence, $\nu_1 = \nu_2 = \delta_\mu$, and so $\mu_1 = \mu_2$, which shows that μ is extreme. □

We turn to some powerful coupling results, needed for the ergodic theory of Markov processes in Chapter 26 and for our discussion of Palm distributions in Chapter 31. First we consider pairs of measurable processes on R_+, with values in an arbitrary measurable space S. In the associated path space, we introduce the invariant σ-field $\mathcal{I}$ and tail σ-field $\mathcal{T} = \bigcap_t \mathcal{T}_t$ with $\mathcal{T}_t = \sigma(\theta_t)$, where clearly $\mathcal{I} \subset \mathcal{T}$. For any signed measure ν, write $\|\nu\|_{\mathcal{A}}$ for the total variation of ν on the σ-field $\mathcal{A}$.

Theorem 25.25 *(coupling on* R_+*, Goldstein, Berbee, Aldous & Thorisson) Let* X, Y *be* S*-valued, product-measurable processes on* R_+*. Then these conditions are equivalent:*

(i) $X \overset{d}{=} Y$ *on* $\mathcal{T}$*,*
(ii) $(\sigma, \theta_\sigma X) \overset{d}{=} (\tau, \theta_\tau Y)$ *for some random times* $\sigma, \tau \geq 0$*,*
(iii) $\left\| \mathcal{L}(\theta_t X) - \mathcal{L}(\theta_t Y) \right\| \to 0$ *as* $t \to \infty$*.*

So are the conditions:

(i′) $X \overset{d}{=} Y$ *on* $\mathcal{I}$*,*
(ii′) $\theta_\sigma X \overset{d}{=} \theta_\tau Y$ *for some random times* $\sigma, \tau \geq 0$*,*
(iii′) $\left\| \int_0^1 \{\mathcal{L}(\theta_{st} X) - \mathcal{L}(\theta_{st} Y)\}\, ds \right\| \to 0$ *as* $t \to \infty$*.*

When the path space is Borel, we may choose $\tilde{Y} \overset{d}{=} Y$ *such that* (ii) *and* (ii′) *become, respectively,*

$$\theta_\tau X = \theta_\tau \tilde{Y}, \qquad \theta_\sigma X = \theta_\tau \tilde{Y} \quad a.s.$$

Proof for (i)−(iii): Let μ_1, μ_2 be the distributions of X, Y, and assume $\mu_1 = \mu_2$ on $\mathcal{T}$. Write $U = S^{\mathsf{R}_+}$, and define a mapping p on $\mathsf{R}_+ \times U$ by $p(s, x) = (s, \theta_s x)$. Let $\mathcal{C}$ be the class of pairs (ν_1, ν_2) of measures on $\mathsf{R}_+ \times U$ with

$$\nu_1 \circ p^{-1} = \nu_2 \circ p^{-1}, \quad \bar{\nu}_1 \leq \mu_1, \quad \bar{\nu}_2 \leq \mu_2, \tag{16}$$

where $\bar{\nu}_i = \nu_i(\mathsf{R}_+ \times \cdot)$, and regard $\mathcal{C}$ as partially ordered under component-wise inequality. Since by Corollary 1.17 every linearly ordered subset has an upper bound in $\mathcal{C}$, Zorn's lemma[7] yields a maximal pair (ν_1, ν_2).

To see that $\bar{\nu}_1 = \mu_1$ and $\bar{\nu}_2 = \mu_2$, define $\mu_i' = \mu_i - \bar{\nu}_i$, and conclude from the equality in (16) that

$$\begin{aligned} \left\| \mu_1' - \mu_2' \right\|_{\mathcal{T}} &= \left\| \bar{\nu}_1 - \bar{\nu}_2 \right\|_{\mathcal{T}} \\ &\leq \left\| \bar{\nu}_1 - \bar{\nu}_2 \right\|_{\mathcal{T}_n} \\ &\leq 2\, \nu_1 \Big\{ (n, \infty) \times U \Big\} \to 0, \end{aligned} \tag{17}$$

which implies $\mu_1' = \mu_2'$ on $\mathcal{T}$. Next, Corollary 2.13 yields some measures $\mu_i^n \leq \mu_i'$ satisfying

$$\mu_1^n = \mu_2^n = \mu_1' \wedge \mu_2' \ \text{ on } \ \mathcal{T}_n, \quad n \in \mathsf{N}.$$

Writing $\nu_i^n = \delta_n \otimes \mu_i^n$, we get $\bar{\nu}_i^n \leq \mu_i'$ and $\nu_1^n \circ p^{-1} = \nu_2^n \circ p^{-1}$, and so $(\nu_1 + \nu_1^n, \nu_2 + \nu_2^n) \in \mathcal{C}$. Since (ν_1, ν_2) is maximal, we obtain $\nu_1^n = \nu_2^n = 0$, and so Corollary 2.9 gives $\mu_1' \perp \mu_2'$ on $\mathcal{T}_n$ for all n. Thus, $\mu_1' A_n = \mu_2' A_n^c = 0$ for some sets $A_n \in \mathcal{T}_n$. Then also $\mu_1' A = \mu_2' A^c = 0$, where $A = \limsup_n A_n \in \mathcal{T}$. Since the μ_i' agree on $\mathcal{T}$, we obtain $\mu_1' = \mu_2' = 0$, which means that $\bar{\nu}_i = \mu_i$. Hence, Theorem 8.17 yields some random variables $\sigma, \tau \geq 0$, such that the pairs (σ, X) and (τ, Y) have distributions ν_1, ν_2, and the desired coupling follows from the equality in (16).

[7] We don't hesitate to use the Axiom of Choice, whenever it leads to short and transparent proofs.

The remaining claims are easy. Thus, the relation $(\sigma, \theta_\sigma X) \stackrel{d}{=} (\tau, \theta_\tau Y)$ implies $\|\mu_1 - \mu_2\|_{\mathcal{T}_n} \to 0$ as in (17), and the latter condition yields $\mu_1 = \mu_2$ on $\mathcal{T}$. When the path space is Borel, the asserted a.s. coupling follows from the distributional version by Theorem 8.17. □

To avoid repetitions, we postpone the proof for (i′)−(iii′) until after the proof of the next theorem, involving groups G of transformations on an arbitrary measurable space $(S, \mathcal{S})$.

Theorem 25.26 *(group coupling, Thorisson) Let G be an lcscH group, acting measurably on a space S with G-invariant σ-field $\mathcal{I}$. Then for any random elements ξ, η in S, these conditions are equivalent:*

(i) $\xi \stackrel{d}{=} \eta$ *on* $\mathcal{I}$,

(ii) $\gamma\xi \stackrel{d}{=} \eta$ *for a random element* γ *in* G.

Proof (OK): Let μ_1, μ_2 be the distributions of ξ, η. Define $p\colon G \times S \to S$ by $p(r, s) = rs$, and let $\mathcal{C}$ be the class of pairs (ν_1, ν_2) of measures on $G \times S$ satisfying (16) with $\bar{\nu}_i = \nu_i(G \times \cdot)$. As before, Zorn's lemma yields a maximal element (ν_1, ν_2), and we claim that $\bar{\nu}_i = \mu_i$ for $i = 1, 2$.

To see this, let λ be a right-invariant Haar measure on G, which exists by Theorem 3.8. Since λ is σ-finite, we may choose a probability measure $\tilde{\lambda} \sim \lambda$, and define

$$\begin{aligned} \mu_i' &= \mu_i - \bar{\nu}_i, \\ \chi_i &= \tilde{\lambda} \otimes \mu_i', \qquad i = 1, 2. \end{aligned}$$

By Corollary 2.13, there exist some measures $\nu_i' \le \chi_i$ satisfying

$$\begin{aligned} \nu_1' \circ p^{-1} &= \nu_2' \circ p^{-1} \\ &= \chi_1 \circ p^{-1} \wedge \chi_2 \circ p^{-1}. \end{aligned}$$

Then $\bar{\nu}_i' \le \mu_i'$ for $i = 1, 2$, and so $(\nu_1 + \nu_1',\ \nu_2 + \nu_2') \in \mathcal{C}$. The maximality of (ν_1, ν_2) yields $\nu_1' = \nu_2' = 0$, and so $\chi_1 \circ p^{-1} \perp \chi_2 \circ p^{-1}$ by Corollary 2.9. Thus, there exists a set $A_1 = A_2^c \in \mathcal{S}$ with $\chi_i \circ p^{-1} A_i = 0$ for $i = 1, 2$. Since $\lambda \ll \tilde{\lambda}$, Fubini's theorem gives

$$\int_S \mu_i'(ds) \int_G 1_{A_i}(gs)\, \lambda(dg) = (\lambda \otimes \mu_i') \circ p^{-1} A_i = 0. \tag{18}$$

By the right invariance of λ, the inner integral on the left is G-invariant and therefore $\mathcal{I}$-measurable in $s \in S$. Since also $\mu_1' = \mu_2'$ on $\mathcal{I}$ by (16), equation (18) remains true with A_i replaced by A_i^c. Adding the two versions gives $\lambda \otimes \mu_i' = 0$, and so $\mu_i' = 0$. Thus, $\bar{\nu}_i = \mu_i$ for $i = 1, 2$.

Since G is Borel, Theorem 8.17 yields some random elements σ, τ in G, such that (σ, ξ) and (τ, η) have distributions ν_1, ν_2. By (16) we get $\sigma\xi \stackrel{d}{=} \tau\eta$, and so the same theorem yields a random element $\tilde{\tau}$ in G with $(\tilde{\tau}, \sigma\xi) \stackrel{d}{=} (\tau, \tau\eta)$. But then $\tilde{\tau}^{-1}\sigma\xi \stackrel{d}{=} \tau^{-1}\tau\eta = \eta$, which proves the desired relation with $\gamma = \tilde{\tau}^{-1}\sigma$. □

Proof of Theorem 25.25 (i′)−(iii′): In the last proof, replace S by the path space $U = S^{\mathsf{R}_+}$, G by the semi-group of shifts θ_t, $t \geq 0$, and λ by Lebesgue measure on R_+. Assuming $X \overset{d}{=} Y$ on $\mathcal{I}$, we may proceed as before up to equation (18), which now takes the form

$$\int_U \mu_i'(dx) \int_0^\infty 1_{A_i}(\theta_t x)\, dt = (\lambda \otimes \mu_i') \circ p^{-1} A_i = 0. \tag{19}$$

Writing $f_i(x)$ for the inner integral on the left, we note that for any $h > 0$,

$$\begin{aligned} f_i(\theta_h x) &= \int_h^\infty 1_{A_i}(\theta_t x)\, dt \\ &= \int_0^\infty 1_{\theta_h^{-1} A_i}(\theta_t x)\, dt. \end{aligned} \tag{20}$$

Hence, (19) remains true with A_i replaced by $\theta_h^{-1} A_i$, and then also for the θ_1-invariant sets

$$\tilde{A}_1 = \limsup_{n\to\infty} \theta_n^{-1} A_1,$$
$$\tilde{A}_2 = \liminf_{n\to\infty} \theta_n^{-1} A_2,$$

where $n \to \infty$ along N. Since $\tilde{A}_1 = \tilde{A}_2^c$, we may henceforth take the A_i in (19) to be θ_1-invariant. Then so are the functions f_i in view of (20). By the monotonicity of $f_i \circ \theta_h$, the f_i are then θ_h-invariant for all $h > 0$ and therefore $\mathcal{I}$-measurable. Arguing as before, we see that $\theta_\sigma X \overset{d}{=} \theta_\tau Y$ for some random variables $\sigma, \tau \geq 0$. The remaining assertions are again routine. □

We turn to a striking relationship between the *sample intensity* $\bar{\xi} = E\big(\xi[0,1]\,\big|\,\mathcal{I}_\xi\big)$ of a stationary, a.s. singular random measure ξ on R_+ and the corresponding maximum over increasing intervals. It may be compared with the more general but less precise maximum inequalities in Proposition 25.10 and Lemmas 25.11 and 25.16.

For the needs of certain applications, we also consider random measures ξ on $[0,1)$. Here $\bar{\xi} = \xi[0,1)$ by definition, and stationarity is defined in the cyclic sense of shifts θ_t on $[0,1)$, where $\theta_t s = s + t$ (mod 1), and correspondingly for sets and measures. Recall that ξ is said to be *singular* if its absolutely continuous component vanishes, which holds in particular for purely atomic measures ξ.

Theorem 25.27 *(ballot theorem, Whitworth, Takács, OK) Let ξ be a stationary, a.s. singular random measure on R_+ or $[0,1)$. Then there exists a $U(0,1)$ random variable $\sigma \perp\!\!\!\perp \mathcal{I}_\xi$, such that*

$$\sigma \sup_{t>0} t^{-1} \xi[0,t] = \bar{\xi} \quad a.s. \tag{21}$$

To justify the statement, we note that the singularity of a measure μ is a measurable property. Indeed, by Proposition 2.24, it is equivalent that the function $F_t = \mu[0,t]$ be singular. Now it is easy to check that the singularity of F can be described by countably many conditions, each involving the increments of F over finitely many intervals with rational endpoints.

Proof: If ξ is stationary on $[0,1)$, the periodic continuation $\eta = \sum_{n\le 0}\theta_n\xi$ is clearly stationary on R_+, and moreover $\mathcal{I}_\eta = \mathcal{I}_\xi$ and $\bar\eta = \bar\xi$. Furthermore, the elementary inequality

$$\frac{x_1+\cdots+x_n}{t_1+\cdots+t_n} \le \max_{k\le n}\frac{x_k}{t_k}, \quad n\in\mathsf{N},$$

valid for arbitrary $x_1, x_2, \ldots \ge 0$ and $t_1, t_2, \ldots > 0$, shows that $\sup_t t^{-1}\eta[0,t] = \sup_t t^{-1}\xi[0,t]$. Thus, it suffices to consider random measures on R_+.

Then put $X_t = \xi(0,t]$, and define

$$\begin{aligned} A_t &= \inf_{s\ge t}(s - X_s), \\ \alpha_t &= 1\big\{A_t = t - X_t\big\}, \qquad t\ge 0. \end{aligned} \tag{22}$$

Noting that $A_t \le t - X_t$ and using the monotonicity of X, we get for any $s<t$

$$\begin{aligned} A_s &= \inf_{r\in[s,t)}(r - X_r)\wedge A_t \\ &\ge (s - X_t)\wedge A_t \\ &\ge (s - t + A_t)\wedge A_t \\ &= s - t + A_t. \end{aligned}$$

If A_0 is finite, then so is A_t for every t, and we get by subtraction

$$0 \le A_t - A_s \le t - s \ \text{ on } \{A_0 > -\infty\}, \quad s<t. \tag{23}$$

Thus, A is non-decreasing and absolutely continuous on $\{A_0 > -\infty\}$.

Now fix a singular path of X such that A_0 is finite, and let $t\ge 0$ be such that $A_t < t - X_t$. Then $A_t + X_{t\pm} < t$ by monotonicity, and so by the left and right continuity of A and X, there exists an $\varepsilon > 0$ such that

$$A_s + X_s < s - 2\varepsilon, \quad |s-t|<\varepsilon.$$

Then (23) yields

$$\begin{aligned} s - X_s &> A_s + 2\varepsilon \\ &> A_t + \varepsilon, \quad |s-t|<\varepsilon, \end{aligned}$$

and by (22) it follows that $A_s = A_t$ for $|s-t|<\varepsilon$. In particular, A has derivative $A_t' = 0 = \alpha_t$ at t.

On the complementary set $D = \{t\ge 0;\ A_t = t - X_t\}$, Theorem 2.15 shows that both A and X are differentiable a.e., the latter with derivative 0, and we may form a set D' by excluding the corresponding null sets. We may also exclude the countable set of isolated points of D. Then for any $t\in D'$ we may choose some $t_n\to t$ in $D\setminus\{t\}$. By the definition of D,

$$\frac{A_{t_n} - A_t}{t_n - t} = 1 - \frac{X_{t_n} - X_t}{t_n - t}, \quad n\in\mathsf{N},$$

and as $n\to\infty$ we get $A_t' = 1 = \alpha_t$. Combining this with the result in the previous case, we see that $A' = \alpha$ a.e. Since A is absolutely continuous, Theorem 2.15 yields

$$A_t - A_0 = \int_0^t \alpha_s ds \ \text{ on } \{A_0 > -\infty\}, \quad t \ge 0. \tag{24}$$

Now $X_t/t \to \bar{\xi}$ a.s. as $t \to \infty$, by Corollary 30.10 below. When $\bar{\xi} < 1$, (22) gives $-\infty < A_t/t \to 1 - \bar{\xi}$ a.s. Furthermore,

$$\begin{aligned} A_t + X_t - t &= \inf_{s \ge t} \Big\{ (s-t) - (X_s - X_t) \Big\} \\ &= \inf_{s \ge 0} \Big\{ s - \theta_t \, \xi(0,s] \Big\}, \end{aligned}$$

and hence

$$\alpha_t = 1\Big\{ \inf_{s \ge 0} \Big(s - \theta_t \, \xi(0,s] \Big) = 0 \Big\}, \quad t \ge 0.$$

Dividing (24) by t and using Corollary 25.9, we get a.s. on $\{\bar{\xi} < 1\}$

$$\begin{aligned} P\Big\{ \sup_{t>0} (X_t/t) \le 1 \Big| \mathcal{I}_\xi \Big\} &= P\Big\{ \sup_{t \ge 0} (X_t - t) = 0 \Big| \mathcal{I}_\xi \Big\} \\ &= P\Big\{ A_0 = 0 \Big| \mathcal{I}_\xi \Big\} \\ &= E\Big(\alpha_0 \Big| \mathcal{I}_\xi \Big) = 1 - \bar{\xi}. \end{aligned}$$

Replacing ξ by $r\xi$ and taking complements, we obtain more generally

$$P\Big\{ r \sup_{t>0} (X_t/t) > 1 \Big| \mathcal{I}_\xi \Big\} = r\bar{\xi} \wedge 1 \ \text{ a.s.}, \quad r \ge 0, \tag{25}$$

where the result for $r\bar{\xi} \in [1, \infty)$ follows by monotonicity.

When $\bar{\xi} \in (0, \infty)$, we may define σ by (21). If instead $\bar{\xi} = 0$ or ∞, we take $\sigma = \vartheta$, where $\vartheta \perp\!\!\!\perp \xi$ is $U(0,1)$. In the latter case, (21) clearly remains true, since $\xi = 0$ a.s. on $\{\bar{\xi} = 0\}$ and $X_t/t \to \infty$ a.s. on $\{\bar{\xi} = \infty\}$. To verify the distributional claim, we conclude from (25) and Theorem 8.5 that, on $\{\bar{\xi} \in (0,\infty)\}$,

$$\begin{aligned} P\Big\{ \sigma < r \Big| \mathcal{I}_\xi \Big\} &= P\Big\{ r \sup_{t>0} (X_t/t) > \bar{\xi} \Big| \mathcal{I}_\xi \Big\} \\ &= r \wedge 1 \ \text{ a.s.}, \quad r \ge 0. \end{aligned}$$

Since the same relation holds trivially when $\bar{\xi} = 0$ or ∞, σ is conditionally $U(0,1)$ given $\mathcal{I}_\xi$, hence unconditionally $U(0,1)$ and independent of $\mathcal{I}_\xi$. □

The last theorem yields a similar result in discrete time. Here (21) holds only with inequality and will be supplemented by an equality similar to (25). For a stationary sequence $\xi = (\xi_1, \xi_2, \ldots)$ in R_+ with invariant σ-field $\mathcal{I}_\xi$, we define $\bar{\xi} = E(\xi_1 | \mathcal{I}_\xi)$ a.s. On $\{1, \ldots, n\}$ we define stationarity in the obvious way in terms of addition modulo n, and put $\bar{\xi} = n^{-1} \sum_k \xi_k$.

Corollary 25.28 *(discrete-time ballot theorem) Let $\xi = (\xi_1, \xi_2, \ldots)$ be a finite or infinite, stationary random sequence in R_+, and put $S_n = \sum_{k \le n} \xi_k$. Then*

(i) *there exists a $U(0,1)$ random variable $\sigma \perp\!\!\!\perp \mathcal{I}_\xi$, such that*

$$\sigma \sup_{n>0} (S_n/n) \le \bar{\xi} \ \text{ a.s.},$$

(ii) *when the* ξ_k *are* Z_+*-valued, we have also*

$$P\Big\{\sup_{n>0}(S_n - n) \ge 0 \,\Big|\, \mathcal{I}_\xi\Big\} = \bar\xi \wedge 1 \quad a.s.$$

Proof: (i) Arguing by periodic continuation, as before, we may reduce to the case of infinite sequences ξ. Now let $\vartheta \perp\!\!\!\perp \xi$ be $U(0,1)$, and define $X_t = S_{[t+\vartheta]}$. Then X has stationary increments, and we note that $\mathcal{I}_X = \mathcal{I}_\xi$ and $\bar X = \bar\xi$. By Theorem 25.27 there exists a $U(0,1)$ random variable $\sigma \perp\!\!\!\perp \mathcal{I}_X$, such that a.s.

$$\begin{aligned}\sup_{n>0}(S_n/n) &= \sup_{t>0}(S_{[t]}/t)\\ &\le \sup_{t>0}(X_t/t) = \bar\xi/\sigma.\end{aligned}$$

(ii) If the ξ_k are Z_+-valued, the same result yields a.s.

$$\begin{aligned}P\Big\{\sup_{n>0}(S_n - n) \ge 0 \,\Big|\, \mathcal{I}_\xi\Big\} &= P\Big\{\sup_{t\ge 0}(X_t - t) > 0 \,\Big|\, \mathcal{I}_\xi\Big\}\\ &= P\Big\{\sup_{t>0}(X_t/t) > 1 \,\Big|\, \mathcal{I}_\xi\Big\}\\ &= P\Big\{\bar\xi > \sigma \,\Big|\, \mathcal{I}_\xi\Big\} = \bar\xi \wedge 1.\end{aligned}$$

□

To state the next result, let ξ be a random element in a countable space S, and put $p_j = P\{\xi = j\}$. For any σ-field $\mathcal{F}$, we define the *information* $I(j)$ and *conditional information* $I(j\,|\mathcal{F})$ by

$$\begin{aligned}I(j) &= -\log p_j,\\ I(j\,|\,\mathcal{F}) &= -\log P\Big\{\xi = j \,\Big|\, \mathcal{F}\Big\}, \qquad j \in S.\end{aligned}$$

For motivation, we note the additivity

$$I\Big(\xi_1, \dots, \xi_n\Big) = I(\xi_1) + I(\xi_2|\,\xi_1) + \cdots + I\Big(\xi_n\Big|\,\xi_1, \dots, \xi_{n-1}\Big), \tag{26}$$

valid for any random elements $\xi_1, \dots, \xi_n$ in S. Next, we form the associated *entropy* $H(\xi) = E\,I(\xi)$ and *conditional entropy* $H(\xi|\mathcal{F}) = E\,I(\xi|\mathcal{F})$, and note that

$$H(\xi) = E\,I(\xi) = -\textstyle\sum_j p_j \log p_j.$$

By (26), even H is additive, in the sense that

$$H\Big(\xi_1, \dots, \xi_n\Big) = H(\xi_1) + H(\xi_2|\,\xi_1) + \cdots + H\Big(\xi_n\Big|\,\xi_1, \dots, \xi_{n-1}\Big). \tag{27}$$

When the sequence (ξ_n) is stationary and ergodic with $H(\xi_0) < \infty$, we show that the averages in (26) and (27) converge toward a common limit.

Theorem 25.29 *(entropy and information, Shannon, McMillan, Breiman, Ionescu Tulcea) Let* $\xi = (\xi_k)$ *be a stationary, ergodic sequence in a countable space* S*, such that* $H(\xi_0) < \infty$*. Then*

$$n^{-1} I\Big(\xi_1, \dots, \xi_n\Big) \to H\Big(\xi_0 \,\Big|\, \xi_{-1}, \xi_{-2}, \dots\Big) \quad a.s. \text{ and in } L^1.$$

Note that $H(\xi_0) < \infty$ holds automatically when the state space is finite. Our proof will be based on a technical estimate.

Lemma 25.30 *(maximum inequality, Chung, Neveu) For any countably valued random variable ξ and discrete filtration $(\mathcal{F}_n)$, we have*

$$E \sup_n I(\xi \,|\, \mathcal{F}_n) \leq H(\xi) + 1.$$

Proof: Write $p_j = P\{\xi = j\}$ and $\eta = \sup_n I(\xi \,|\mathcal{F}_n)$. For fixed $r > 0$, we introduce the optional times

$$\begin{aligned} \tau_j &= \inf\Big\{n;\ I(j\,|\mathcal{F}_n) > r\Big\} \\ &= \inf\Big\{n;\ P\Big\{\xi = j \,\Big|\, \mathcal{F}_n\Big\} < e^{-r}\Big\}, \quad j \in S. \end{aligned}$$

By Lemma 8.3,

$$\begin{aligned} P\Big\{\eta > r,\ \xi = j\Big\} &= P\Big\{\tau_j < \infty,\ \xi = j\Big\} \\ &= E\Big(P\Big\{\xi = j \,\Big|\, \mathcal{F}_{\tau_j}\Big\};\ \tau_j < \infty\Big) \\ &\leq e^{-r} P\{\tau_j < \infty\} \leq e^{-r}. \end{aligned}$$

Since the left-hand side is also bounded by p_j, Lemma 4.4 yields

$$\begin{aligned} E\,\eta &= \textstyle\sum_j E\Big(\eta;\ \xi = j\Big) \\ &= \textstyle\sum_j \displaystyle\int_0^\infty P\Big\{\eta > r,\ \xi = j\Big\}\, dr \\ &\leq \textstyle\sum_j \displaystyle\int_0^\infty \Big(e^{-r} \wedge p_j\Big)\, dr \\ &= \textstyle\sum_j p_j \Big(1 - \log p_j\Big) \\ &= H(\xi) + 1. \end{aligned}$$

□

Proof of Theorem 25.29 *(Breiman):* We may choose ξ to be defined on the canonical space S^∞. Then introduce the functions

$$\begin{aligned} g_k(\xi) &= I\Big(\xi_0 \,\Big|\, \xi_{-1}, \dots, \xi_{-k+1}\Big), \\ g(\xi) &= I\Big(\xi_0 \,\Big|\, \xi_{-1}, \xi_{-2}, \dots\Big). \end{aligned}$$

By (26), we may write the assertion as

$$n^{-1} \sum_{k \leq n} g_k(\theta^k \xi) \to E\, g(\xi) \ \text{ a.s. and in } L^1. \tag{28}$$

Here $g_k(\xi) \to g(\xi)$ a.s. by martingale convergence, and $E \sup_k g_k(\xi) < \infty$ by Lemma 25.30. Hence, (28) follows by Corollary 25.8. □

Exercises

1. State and prove some continuous-time, two-sided, and higher-dimensional versions of Lemma 25.1.

2. For a stationary random sequence $\xi = (\xi_1, \xi_2, \ldots)$, show that the ξ_n are i.i.d. iff $\xi_1 \perp\!\!\!\perp (\xi_2, \xi_3, \ldots)$.

3. For a Borel space S, let X be a stationary array of random elements in S indexed by N^d. Prove the existence of a stationary array Y indexed by Z^d such that $X = Y$ a.s. on N^d.

4. Let X be a stationary process on R_+ with values in a Borel space S. Prove the existence of a stationary process Y on R with $X \overset{d}{=} Y$ on R_+. Strengthen this to an a.s. equality, when S is a complete metric space and X is right-continuous.

5. Let ξ be a two-sided, stationary random sequence with restriction η to N. Show that ξ, η are simultaneously ergodic. (*Hint:* For any measurable, invariant set $I \in S^{\mathsf{Z}}$, there exists a measurable, invariant set $I' \in S^{\mathsf{N}}$ with $I = S^{\mathsf{Z}_-} \times I'$ a.s. $\mathcal{L}(\xi)$.)

6. Establish two-sided and higher-dimensional versions of Lemmas 25.4 and 25.5, as well as of Theorem 25.9.

7. Say that a measure-preserving transformation T on a probability space $(S, \mathcal{S}, \mu)$ is *mixing* if $\mu(A \cap T^{-n}B) \to \mu A \cdot \mu B$ for all $A, B \in \mathcal{S}$. Prove a counterpart of Lemma 25.5 for mixing. Also show that mixing transformations are ergodic. (*Hint:* For the latter assertion, take $A = B$ to be invariant.)

8. Show that it suffices to verify the mixing property for sets in a generating π-system. Using this, prove that i.i.d. sequences are mixing under shifts.

9. For any $a \in \mathsf{R}$, define $Ts = s + a \pmod 1$ on $[0, 1]$. Show that T fails to be mixing but is ergodic iff $a \notin \mathsf{Q}$. (*Hint:* For the ergodicity, let $I \subset [0, 1]$ be T-invariant. Then so is the measure $1_I\lambda$. Since the points ka are dense in $[0, 1]$, it follows that $1_I\lambda$ is invariant. Now use Theorem 2.6.)

10. (*Bohl, Sierpiński, Weyl*) For any $a \notin \mathsf{Q}$, put $\mu_n = n^{-1}\sum_{k \le n} \delta_{ka}$, where $ka \in [0, 1]$ is defined modulo 1. Show that $\mu_n \xrightarrow{w} \lambda$. (*Hint:* Apply Theorem 25.6 to the map in the previous exercise.)

11. Show that the transformation $Ts = 2s \pmod 1$ on $[0, 1]$ is mixing. Also show how the map of Lemma 4.20 can be generated as in Lemma 25.1 by means of T.

12. Note that Theorem 25.6 remains true for invertible shifts T, with averages taken over increasing index sets $[a_n, b_n]$ with $b_n - a_n \to \infty$. Show by an example that the a.s. convergence may fail without the monotonicity. (*Hint:* Consider an i.i.d. sequence (ξ_n) and disjoint intervals $[a_n, b_n]$, and use the Borel–Cantelli lemma.)

13. Consider a one- or two-sided, stationary random sequence (ξ_n) in a measurable space $(S, \mathcal{S})$, and fix any $B \in \mathcal{S}$. Show that a.s. either $\xi_n \in B^c$ for all n or $\xi_n \in B$ i.o. (*Hint:* Use Theorem 25.6.)

14. (*von Neumann*) Give a direct proof of the L^2-version of Theorem 25.6. (*Hint:* Define a unitary operator U on $L^2(S)$ by $Uf = f \circ T$. Let M denote the U-invariant subspace of L^2 and put $A = I - U$. Check that $M^\perp = \bar{R}_A$, the closed range of A. By Theorem 1.35 it is enough to take $f \in M$ or $f \in R_A$.) Deduce the general L^p-version, and extend the argument to higher dimensions.

15. In the context of Theorem 25.24, show that the set of ergodic measures in M is measurable. (*Hint:* Use Lemma 3.2, Proposition 5.32, and Theorem 25.14.)

16. Prove a continuous-time version of Theorem 25.24.

17. Deduce Theorem 5.23 for $p \le 1$ from Theorem 25.20. (*Hint:* Take $X_{m,n} = |S_n - S_m|^p$, and note that $E|S_n|^p = o(n)$ when $p < 1$.)

18. Let $\xi = (\xi_1, \xi_2, \ldots)$ be a stationary sequence of random variables, fix any $B \in \mathcal{B}_{\mathsf{R}^d}$, and let κ_n be the number of indices $k \in \{1, \ldots, n-d\}$ with $(\xi_k, \ldots, \xi_d) \in B$. Use Theorem 25.20 to show that κ_n/n converges a.s. Deduce the same result from Theorem 25.6, by considering suitable sub-sequences.

19. Strengthen the inequality in Lemma 25.7 to $E\Big\{\xi_1;\, \sup_n (S_n/n) \ge 0\Big\} \ge 0$. (*Hint:* Apply the original result to the variables $\xi_k + \varepsilon$, and let $\varepsilon \to 0$.)

20. Extend Proposition 25.10 to stationary processes on Z^d.

21. Extend Theorem 25.14 to averages over arbitrary rectangles $A_n = [0, a_{n1}] \times \cdots \times [0, a_{nd}]$, such that $a_{nj} \to \infty$ and $\sup_n (a_{ni}/a_{nj}) < \infty$ for all $i \ne j$. (*Hint:* Note that Lemma 25.16 extends to this case.)

22. Derive a version of Theorem 25.14 for stationary processes X on Z^d. (*Hint:* By a suitable randomization, construct an associated stationary process $\tilde{X}$ on R^d, apply Theorem 25.14 to $\tilde{X}$, and estimate the error term as in Corollary 30.10.)

23. Give an example of two processes X, Y on R_+, such that $X \overset{d}{=} Y$ on $\mathcal{I}$ but not on $\mathcal{T}$.

24. Derive a version of Theorem 25.25 for processes on Z_+. Also prove versions for processes on R_+^d and Z_+^d.

25. Show that Theorem 25.25 (ii) implies a corresponding result for processes on R. (*Hint:* Apply Theorem 25.25 to the processes $\tilde{X}_t = \theta_t X$ and $\tilde{Y} = \theta_t Y$.) Also show how the two-sided statement follows from Theorem 25.26.

26. For processes X on R_+, define $\tilde{X}_t = (X_t, t)$, and let $\tilde{\mathcal{I}}$ be the associated invariant σ-field. Assuming X, Y to be measurable, show that $X \overset{d}{=} Y$ on $\mathcal{T}$ iff $\tilde{X} \overset{d}{=} \tilde{Y}$ on $\tilde{\mathcal{I}}$. (*Hint:* Use Theorem 25.25.)

27. Show by an example that the conclusion of Theorem 25.27 may fail when ξ is not singular.

28. Give an example where the inequality in Corollary 25.28 is a.s. strict. (*Hint:* Examine the proof.)

29. (*Bertrand, André*) Show that if two candidates A, B in an election get the proportions p and $1-p$ of the votes, then the probability that A will lead throughout the ballot count equals $(2p-1)_+$. (*Hint:* Apply Corollary 25.28, or use a direct combinatorial argument based on the reflection principle.)

30. Prove the second claim in Corollary 25.28 by a martingale argument, when $\xi_1, \ldots, \xi_n$ are Z_+-valued and exchangeable. (*Hint:* We may assume that S_n is non-random. Then the variables $M_k = S_k/k$ form a reverse martingale, and the result follows by optional sampling.)

31. Prove that the convergence in Theorem 25.29 holds in L^p for arbitrary $p > 0$ when S is finite. (*Hint:* Show as in Lemma 25.30 that $\|\sup_n I(\xi\,|\mathcal{F}_n)\|_p < \infty$ when ξ is S-valued, and use Corollary 25.8 (ii).)

32. Show that $H(\xi,\eta) \le H(\xi) + H(\eta)$ for any ξ, η. (*Hint:* Note that $H(\eta\,|\,\xi) \le H(\eta)$ by Jensen's inequality.)

33. Give an example of a stationary Markov chain (ξ_n), such that $H(\xi_1) > 0$ but $H(\xi_1|\,\xi_0) = 0$.

34. Give an example of a stationary Markov chain (ξ_n), such that $H(\xi_1) = \infty$ but $H(\xi_1|\,\xi_0) < \infty$. (*Hint:* Choose the state space Z_+ and transition probabilities $p_{ij} = 0$, unless $j \in \{0,\, i+1\}$.)

Chapter 26

Ergodic Properties of Markov Processes

Transition and contraction operators, operator ergodic theorem, maximum inequalities, ratio ergodic theorem, filling operators and functionals, ratio limit theorem, space-time invariance, strong and weak ergodicity and mixing, tail and invariant triviality, Harris recurrence and ergodicity, potential and resolvent operators, convergence dichotomy, recurrence dichotomy, invariant measures, distributional and pathwise limits

The primary aim of this chapter is to study the asymptotic behavior of broad classes of Markov processes. First, we will see how the ergodic theorems of the previous chapter admit some powerful extensions to suitable contraction operators, leading to some pathwise and distributional ratio limit theorems for fairly general Markov processes. Next, we establish some basic criteria for strong or weak ergodicity of conservative Markov semi-groups, and show that Harris recurrent Feller processes are strongly ergodic. Finally, we show that any regular Feller process is either Harris recurrent or uniformly transient, examine the existence of invariant measures, and prove some associated distributional and pathwise limit theorems. The latter analysis relies crucially on some profound tools from potential theory.

For a more detailed description, we first extend the basic ergodic theorem of Chapter 25 to suitable contraction operators on an arbitrary measure space, and establish a general operator version of the ratio ergodic theorem. The relevance of those results for the study of Markov processes is due to the fact that their transition operators are positive $L^1 - L^\infty$-contractions with respect to any invariant measure λ on the state space S. The mentioned results cover both the positive-recurrent case where $\lambda S < \infty$, and the null-recurrent case where $\lambda S = \infty$. Even more remarkably, the same ergodic theorems apply to both the transition probabilities and the sample paths, in both cases giving conclusive information about the asymptotic behavior.

Next we prove, for an arbitrary Markov process, that a certain condition of strong ergodicity is equivalent to the triviality of the tail σ-field, the constancy of all bounded, space-time invariant functions, and a uniform mixing condition. We also consider a similar result where all four conditions are replaced by suitably averaged versions. For both sets of equivalences, one gets very simple and transparent proofs by applying the general coupling results of Chapter 25.

In order to apply the mentioned theorems to specific Markov processes, we need to find regularity conditions ensuring the existence of an invariant

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_27

measure or the triviality of the tail σ-field. Here we consider a general class of Feller processes satisfying either a strong recurrence or a uniform transience condition. In the former case, we prove the existence of an invariant measure, required for the application of the mentioned ergodic theorems, and show that the space-time invariant functions are constant, which implies the mentioned strong ergodicity. Our proofs of the latter results depend on some potential theoretic tools, related to those developed in Chapter 17.

To set the stage for the technical developments, we consider a Markov transition operator T on an arbitrary measurable space $(S, \mathcal{S})$. Note that T is *positive*, in the sense that $f \geq 0$ implies $Tf \geq 0$, and also that $T1 = 1$. As before, we write P_x for the distribution of a Markov process on Z_+ with transition operator T, starting at $x \in S$. More generally, we define $P_\mu = \int P_x\, \mu(dx)$ for any measure μ on S. A measure λ on S is said to be *invariant* if $\lambda Tf = \lambda f$ for any measurable function $f \geq 0$. Writing θ for the shift on the path space S^∞, we define the associated operator $\tilde{\theta}$ by $\tilde{\theta} f = f \circ \theta$.

For any $p \geq 1$, an operator T on a measure space $(S, \mathcal{S}, \mu)$ is called an *L^p-contraction*, if $\|Tf\|_p \leq \|f\|_p$ for every $f \in L^p$. We further say that T is an *$L^1\!-\!L^\infty$-contraction*, if it is an L^p-contraction for every $p \in [1, \infty]$. The following result shows the relevance of the mentioned notions for the theory of Markov processes.

Lemma 26.1 *(transition and contraction operators) Let T be a Markov transition operator on $(S, \mathcal{S})$ with invariant measure λ. Then*

(i) *T is a positive $L^1\!-\!L^\infty$-contraction on (S, λ),*
(ii) *$\tilde{\theta}$ is a positive $L^1\!-\!L^\infty$-contraction on (S^∞, P_λ).*

Proof: (i) Applying Jensen's inequality to the transition kernel $\mu(x, B) = T1_B(x)$ and using the invariance of λ, we get for any $p \in [1, \infty)$ and $f \in L^p$

$$\begin{aligned} \|Tf\|_p^p &= \lambda|\mu f|^p \\ &\leq \lambda\mu|f|^p \\ &= \lambda|f|^p = \|f\|_p^p. \end{aligned}$$

The result for $p = \infty$ is obvious.

(ii) Proceeding as in Lemma 11.11, we see that θ is a measure-preserving transformation on (S^∞, P_λ). Hence, for any measurable function $f \geq 0$ on S^∞ and constant $p \geq 1$, we have

$$\begin{aligned} P_\lambda\big|\tilde{\theta} f\big|^p &= P_\lambda\big|f \circ \theta\big|^p \\ &= \big(P_\lambda \circ \theta^{-1}\big)|f|^p \\ &= P_\lambda|f|^p. \end{aligned}$$

The contraction property for $p = \infty$ is again obvious. □

Some crucial results from Chapter 25 carry over to the context of positive $L^1 - L^\infty$-contractions on an arbitrary measure space. First we consider an operator version of Birkhoff's ergodic theorem. To simplify the writing, we introduce the operators

$$S_n = \sum_{k<n} T^k, \qquad A_n = S_n/n, \qquad Mf = \sup_n A_n f.$$

Say that f is *T-invariant* if $Tf = f$.

Theorem 26.2 *(operator ergodic theorem, Hopf, Dunford & Schwartz) Let T be a positive L^1-L^∞-contraction on a measure space $(S, \mathcal{S}, \mu)$. Then for any $f \in L^1$ there exists a T-invariant function $Af \in L^1$, such that*

$$A_n f \to Af \ \ a.e., \quad f \in L^1.$$

For the proof, we need to extend the inequalities in Lemmas 25.7 and 25.11 and Proposition 25.10 (i) to an operator setting.

Lemma 26.3 *(maximum inequalities) For a positive L^1-contraction T on a measure space $(S, \mathcal{S}, \mu)$, we have*

(i) $\mu\big(f;\, Mf > 0\big) \geq 0,\ f \in L^1.$

If T is even an L^1-L^∞-contraction, then also

(ii) $r\,\mu\{Mf > 2r\} \leq \mu(f;\, f > r),\ f \in L^1,\ r > 0,$

(iii) $\|Mf\|_p \lesssim \|f\|_p,\ f \in L^p,\ p > 1.$

Proof: (i) For any $f \in L^1$, write $M_n f = S_1 f \vee \cdots \vee S_n f$, and conclude that by positivity

$$\begin{aligned} S_k f &= f + TS_{k-1} f \\ &\leq f + T(M_n f)_+, \quad k = 1, \dots, n. \end{aligned}$$

Hence, $M_n f \leq f + T(M_n f)_+$ for all n, and so by positivity and contractivity,

$$\begin{aligned} \mu\big(f;\, M_n f > 0\big) &\geq \mu\big\{M_n f - T(M_n f)_+;\, M_n f > 0\big\} \\ &\geq \mu\big\{(M_n f)_+ - T(M_n f)_+\big\} \\ &= \big\|(M_n f)_+\big\|_1 - \big\|T(M_n f)_+\big\|_1 \geq 0. \end{aligned}$$

As before, it remains to let $n \to \infty$.

(ii) Put $f_r = f1\{f > r\}$. By the L^∞-contractivity and positivity of A_n,

$$\begin{aligned} A_n f - 2r &\leq A_n(f - 2r) \\ &\leq A_n(f_r - r), \quad n \in \mathsf{N}, \end{aligned}$$

which implies $Mf - 2r \leq M(f_r - r)$. Hence, by part (i),

$$\begin{aligned} r\,\mu\{Mf > 2r\} &\leq r\,\mu\big\{M(f_r - r) > 0\big\} \\ &\leq \mu\big\{f_r;\, M(f_r - r) > 0\big\} \\ &\leq \mu f_r = \mu\big(f;\, f > r\big). \end{aligned}$$

(iii) Here the earlier proof applies with only notational changes. □

Proof of Theorem 26.2: Fix any $f \in L^1$. By dominated convergence, f may be approximated in L^1 by functions $\tilde{f} \in L^1 \cap L^\infty \subset L^2$. By Lemma 25.17 we may next approximate $\tilde{f}$ in L^2 by functions of the form $\hat{f} + (g - Tg)$, where $\hat{f}, g \in L^2$ and $T\hat{f} = \hat{f}$. Finally, we may approximate g in L^2 by functions $\tilde{g} \in L^1 \cap L^\infty$. Since T contracts L^2, the functions $\tilde{g} - T\tilde{g}$ will then approximate $g - Tg$ in L^2. Combining the three approximations, we have for any $\varepsilon > 0$

$$f = f_\varepsilon + (g_\varepsilon - Tg_\varepsilon) + h_\varepsilon + k_\varepsilon, \tag{1}$$

where $f_\varepsilon \in L^2$ with $Tf_\varepsilon = f_\varepsilon$, $g_\varepsilon \in L^1 \cap L^\infty$, and $\|h_\varepsilon\|_2 \vee \|k_\varepsilon\|_1 < \varepsilon$.

Since f_ε is invariant, we have $A_n f_\varepsilon \equiv f_\varepsilon$. Next, we note that

$$\begin{aligned} \Big\| A_n(g_\varepsilon - Tg_\varepsilon) \Big\|_\infty &= n^{-1} \Big\| g_\varepsilon - T^n g_\varepsilon \Big\|_\infty \\ &\le 2n^{-1} \|g_\varepsilon\|_\infty \to 0. \end{aligned} \tag{2}$$

Hence,

$$\limsup_{n\to\infty} A_n f \le f_\varepsilon + Mh_\varepsilon + Mk_\varepsilon < \infty \quad \text{a.e.},$$

and similarly for $\liminf_n A_n f$. Combining the two estimates gives

$$\Big(\limsup_{n\to\infty} - \liminf_{n\to\infty}\Big) A_n f \le 2M|h_\varepsilon| + 2M|k_\varepsilon|.$$

Now Lemma 26.3 yields for any $\varepsilon, r > 0$

$$\begin{aligned} \Big\| M|h_\varepsilon| \Big\|_2 \;&\lesssim\; \|h_\varepsilon\|_2 \;<\; \varepsilon, \\ \mu\Big\{ M|k_\varepsilon| > 2r \Big\} \;&\le\; r^{-1}\|k_\varepsilon\|_1 \;<\; \varepsilon/r, \end{aligned}$$

and so $M|h_\varepsilon| + M|k_\varepsilon| \to 0$ a.e. as $\varepsilon \to 0$ along a suitable sequence. Thus, $A_n f$ converges a.e. toward a limit Af.

To prove that Af is T-invariant, we see from (1) and (2) that the a.e. limits Ah_ε and Ak_ε exist and satisfy $TAf - Af = (TA - A)(h_\varepsilon + k_\varepsilon)$. By the contraction property and Fatou's lemma, the right-hand side tends a.e. to 0 as $\varepsilon \to 0$ along a sequence, and we get $TAf = Af$ a.e. □

The limit Af in the last theorem may be 0, in which case the a.s. convergence $A_n f \to Af$ gives little information about the asymptotic behavior of $A_n f$. For example, this happens when $\mu S = \infty$ and T is the operator induced by a μ-preserving and ergodic transformation θ on S. Then Af is a constant, and the condition $Af \in L^1$ implies $Af = 0$. To get around this difficulty, we may instead compare the asymptotic behavior of $S_n f$ with that of $S_n g$ for a suitable reference function $g \in L^1$. This idea leads to a far-reaching and powerful extension of Birkhoff's theorem.

Theorem 26.4 *(ratio ergodic theorem, Chacon & Ornstein) Let T be a positive L^1-contraction on a measure space $(S, \mathcal{S}, \mu)$, and let $f \in L^1$ and $g \in L^1_+$. Then*

$$\frac{S_n f}{S_n g} \quad \textit{converges a.e. on } \{S_\infty g > 0\}.$$

Three lemmas are needed for the proof.

Lemma 26.5 *(individual terms) In the context of Theorem 26.4,*

$$\frac{T^n f}{S_{n+1}\, g} \to 0 \ \ \textit{a.e. on} \ \{S_\infty g > 0\}.$$

Proof: We may take $f \geq 0$. Fix any $\varepsilon > 0$, and define

$$\begin{aligned} h_n &= T^n f - \varepsilon\, S_{n+1}\, g, \\ A_n &= \{h_n > 0\}, \qquad n \geq 0. \end{aligned}$$

By positivity,

$$\begin{aligned} h_n &= T h_{n-1} - \varepsilon g \\ &\leq T h_{n-1}^+ - \varepsilon g, \quad n \geq 1. \end{aligned}$$

Examining separately the cases A_n and A_n^c, we conclude that

$$h_n^+ \leq T h_{n-1}^+ - \varepsilon 1_{A_n} g, \quad n \geq 1,$$

and so by contractivity,

$$\begin{aligned} \varepsilon\, \mu(g; A_n) &\leq \mu\Big(T h_{n-1}^+\Big) - \mu h_n^+ \\ &\leq \mu h_{n-1}^+ - \mu h_n^+. \end{aligned}$$

Summing over n gives

$$\begin{aligned} \varepsilon\, \mu \sum_{n\geq 1} 1_{A_n} g &\leq \mu h_0^+ = \mu(f - \varepsilon g) \\ &\leq \mu f < \infty, \end{aligned}$$

which implies $\mu(g;\, A_n \text{ i.o.}) = 0$ and hence $\limsup_n (T^n f / S_{n+1} g) \leq \varepsilon$ a.e. on $\{g > 0\}$. Since ε was arbitrary, we obtain $T^n f / S_{n+1} g \to 0$ a.e. on $\{g > 0\}$. Applying this result to the functions $T^m f$ and $T^m g$ gives the same convergence on $\{S_{m-1} g = 0 < S_m g\}$, for arbitrary $m > 1$. □

To state the next result, we introduce the non-linear *filling operator* F on L^1, given by $Fh = Th_+ - h_-$. We may think of the sequence $F^n h$ as arising from successive attempts to fill a hole h_-, where in each step we are only mapping matter that has not yet fallen into the hole. We also define $M_n h = S_1 h \vee \cdots \vee S_n h$.

Lemma 26.6 *(filling operator) For any $h \in L^1$ and $n \in \mathsf{N}$,*

$$F^{n-1} h \geq 0 \ \ \textit{on} \ \{M_n h > 0\}.$$

Proof: Writing $h_k = h_+ + (Fh)_+ + \cdots + (F^k h)_+$, we claim that

$$h_k \geq S_{k+1} h, \quad k \geq 0. \tag{3}$$

This is clear for $k = 0$ since $h_+ = h + h_- \geq h$. Proceeding by induction, let (3) be true for $k = m \geq 0$. Using the induction hypothesis and the definitions of S_k, h_k, and F, we get for $m + 1$

$$\begin{aligned}
S_{m+2}h &= h + TS_{m+1}h \\
&\leq h + Th_m \\
&= h + \sum_{k\leq m} T(F^k h)_+ \\
&= h + \sum_{k\leq m} \Big\{ F^{k+1}h + (F^k h)_- \Big\} \\
&= h + \sum_{k\leq m} \Big\{ (F^{k+1}h)_+ - (F^{k+1}h)_- + (F^k h)_- \Big\} \\
&= h + h_{m+1} - h_+ + h_- - (F^{m+1}h)_- \leq h_{m+1}.
\end{aligned}$$

This completes the proof of (3).

If $M_n h > 0$ at some point in S, then $S_k h > 0$ for some $k \leq n$, and so by (3) we have $h_k > 0$ for some $k < n$. But then $(F^k h)_+ > 0$, and therefore $(F^k h)_- = 0$ for the same k. Since $(F^k h)_-$ is non-increasing, it follows that $(F^{n-1}h)_- = 0$, and hence $F^{n-1}h \geq 0$. $\square$

To state our third and crucial lemma, write $g \in \mathcal{T}_1(f)$ for a given $f \in L^1_+$, if there exists a decomposition $f = f_1 + f_2$ with $f_1, f_2 \in L^1_+$ such that $g = Tf_1 + f_2$. Note in particular that $f, g \in L^1_+$ implies $U(f-g) = f' - g$ for some $f' \in \mathcal{T}_1(f)$. The classes $\mathcal{T}_n(f)$ are defined recursively by $\mathcal{T}_{n+1}(f) = \mathcal{T}_1\{\mathcal{T}_n(f)\}$, and we put $\mathcal{T}(f) = \bigcup_n \mathcal{T}_n(f)$. Now introduce the functionals

$$\psi_B f = \sup\Big\{\mu(g; B);\ g \in \mathcal{T}(f)\Big\}, \quad f \in L^1_+,\ B \in \mathcal{S}.$$

Lemma 26.7 *(filling functionals) For any $f, g \in L^1_+$ and $B \in \mathcal{S}$,*

$$B \subset \Big\{\limsup_{n\to\infty} S_n(f-g) > 0\Big\} \quad \Rightarrow \quad \psi_B f \geq \psi_B g.$$

Proof: Fix any $g' \in \mathcal{T}(g)$ and $c > 1$. First we show that

$$\Big\{\limsup_{n\to\infty} S_n(f-g) > 0\Big\} \subset \Big\{\limsup_{n\to\infty} S_n(cf - g') > 0\Big\} \text{ a.e.} \tag{4}$$

Here we may assume that $g' \in \mathcal{T}_1(g)$, since the general result then follows by iteration in finitely many steps. Letting $g' = r + Ts$ for some $r, s \in L^1_+$ with $r + s = g$, we obtain

$$\begin{aligned}
S_n(cf - g') &= \sum_{k<n} T^k(cf - r - Ts) \\
&= S_n(f-g) + (c-1)S_n f + s - T^n s.
\end{aligned}$$

Since

$$\frac{T^n s}{S_n f} = \frac{T^{n-1}Ts}{S_n f} \to 0 \text{ a.e. on } \{S_\infty f > 0\}$$

by Lemma 26.5, we conclude that eventually $S_n(cf - g') \geq S_n(f-g)$ a.e. on the same set, and (4) follows.

Combining the given hypothesis with (4), we obtain the a.e. relation $B \subset \{M(cf - g') > 0\}$. Now Lemma 26.6 yields

$$F^{n-1}(cf - g') \geq 0 \text{ on } B_n \equiv B \cap \Big\{M_n(cf - g') > 0\Big\}, \quad n \in \mathsf{N}.$$

Since $B_n \uparrow B$ a.e. and $F^{n-1}(cf - g') = f' - g'$ for some $f' \in \mathcal{T}(cf)$, we get

$$\begin{aligned} 0 &\le \mu\Big\{F^{n-1}(cf - g'); B_n\Big\} \\ &= \mu(f'; B_n) - \mu(g'; B_n) \\ &\le \psi_B(cf) - \mu(g'; B_n) \\ &\to c\,\psi_B f - \mu(g'; B), \end{aligned}$$

and so $c\,\psi_B f \ge \mu(g'; B)$. It remains to let $c \to 1$ and take the supremum over $g' \in \mathcal{T}(g)$. □

Proof of Theorem 26.4: We may take $f \ge 0$. On $\{S_\infty g > 0\}$, put

$$\alpha = \liminf_{n\to\infty} \frac{S_n f}{S_n g} \le \limsup_{n\to\infty} \frac{S_n f}{S_n g} = \beta,$$

and define $\alpha = \beta = 0$ otherwise. Since $S_n g$ is non-decreasing, we have for any $c > 0$

$$\begin{aligned} \{\beta > c\} &\subset \Big\{ \limsup_{n\to\infty} \frac{S_n(f - cg)}{S_n g} > 0,\ S_\infty g > 0\Big\} \\ &\subset \Big\{ \limsup_{n\to\infty} S_n(f - cg) > 0\Big\}. \end{aligned}$$

Writing $B = \{\beta = \infty,\, S_\infty g > 0\}$, we see from Lemma 26.7 that

$$\begin{aligned} c\,\psi_B g &= \psi_B(cg) \le \psi_B f \\ &\le \mu f < \infty, \end{aligned}$$

and as $c \to \infty$ we get $\psi_B g = 0$. But then $\mu(T^n g; B) = 0$ for all $n \ge 0$, and therefore $\mu(S_\infty g; B) = 0$. Since $S_\infty g > 0$ on B, we obtain $\mu B = 0$, which means that $\beta < \infty$ a.e.

Now define $C = \{\alpha < a < b < \beta\}$ for fixed $b > a > 0$. As before,

$$C \subset \Big\{ \limsup_{n\to\infty} S_n(f - bg) \wedge \limsup_{n\to\infty} S_n(ag - f) > 0\Big\},$$

and so by Lemma 26.7,

$$\begin{aligned} b\,\psi_C g &= \psi_C(bg) \le \psi_C f \\ &\le \psi_C(ag) \\ &= a\,\psi_C g < \infty, \end{aligned}$$

which implies $\psi_C g = 0$, and therefore $\mu C = 0$. Hence,

$$\mu\{\alpha < \beta\} \le \sum_{a<b} \mu\Big\{\alpha < a < b < \beta\Big\} = 0,$$

where the summation extends over all rational $a < b$, and so $\alpha = \beta$ a.e., which proves the asserted convergence. □

To illustrate the use of the last theorem, we consider a striking application to discrete-time Markov processes. For such a process X on S and a measurable function f on S^∞, we define $S_n f = \sum_{k<n} f(\theta_k X)$.

Corollary 26.8 *(ratio limit theorem) Let λ be an invariant measure of a discrete-time Markov process in S, and let $f \in L^1(P_\lambda)$ and $g \in L^1_+(P_\lambda)$. Then*

(i) $\dfrac{S_n f}{S_n g}$ *converges a.e. P_λ on* $\big\{y \in S^\infty;\ S_\infty g(y) > 0\big\}$,

(ii) $\dfrac{E_x S_n f}{E_x S_n g}$ *converges a.e. λ on* $\big\{x \in S;\ E_x S_\infty g > 0\big\}$.

Proof: (i) By Lemma 26.1 (ii), we may apply Theorem 26.4 to the L^1–L^∞-contraction $\tilde\theta$ on (S^∞, P_λ) induced by the shift θ, and the result follows.

(ii) Writing $\tilde f(x) = E_x f(X)$ and using the Markov property at k, we get

$$\begin{aligned} E_x f(\theta_k X) &= E_x E_{X_k} f(X) \\ &= E_x \tilde f(X_k) \\ &= T^k \tilde f(x), \end{aligned}$$

and so

$$\begin{aligned} E_x S_n f &= \sum_{k<n} E_x f(\theta_k X) \\ &= \sum_{k<n} T^k \tilde f(x) \\ &= S_n \tilde f(x), \end{aligned}$$

where $S_n = T^0 + \cdots + T^{n-1}$ on the right. We also note that

$$\begin{aligned} \lambda \tilde f &= \int \tilde f(x)\, \lambda(dx) \\ &= \int E_x f(X)\, \lambda(dx) \\ &= E_\lambda f(X) = P_\lambda f. \end{aligned}$$

Now Lemma 26.1 (i) shows that T is a positive L^1–L^∞-contraction on (S, λ). By Theorem 26.4, we conclude that $S_n \tilde f(x)/S_n \tilde g(x)$ converges a.e. λ on the set $\big\{x \in S;\ S_\infty \tilde g(x) > 0\big\}$, which translates immediately into the asserted statement. □

Now consider a conservative, continuous-time Markov process on an arbitrary state space $(S, \mathcal{S})$ with distributions P_x and associated expectation operators E_x. On the canonical path space $\Omega = S^{\mathbb{R}_+}$, we introduce the shift operators θ_t and filtration $\mathcal{F} = (\mathcal{F}_t)$. A bounded function $f : S \to \mathbb{R}$ is said to be *invariant* or *harmonic*, if it is measurable and such that

$$\begin{aligned} f(x) &= T_t f(x) \\ &= E_x f(X_t), \quad x \in S,\ t \ge 0. \end{aligned}$$

More generally, we say that a bounded function $f : S \times \mathbb{R}_+ \to \mathbb{R}$ is *space-time invariant* or *harmonic*, if it is measurable and satisfies

$$f(x, s) = E_x f(X_t, s + t), \quad x \in S,\ s, t \ge 0. \tag{5}$$

For motivation, we note that f is then invariant for the associated *space-time process* $\tilde X_t = (X_t, s + t)$ in $S \times \mathbb{R}_+$, where the second component is deterministic, apart from a possibly random initial value $s \geq 0$. Note that $\tilde{X}$ is again a time-homogeneous Markov process with transition operators $\tilde{T}_t f(x,s) = E_x f(X_t, s+t)$. We need the following useful martingale connection.

Lemma 26.9 *(space-time invariance) For any bounded, measurable function $f\colon S \times \mathsf{R}_+ \to \mathsf{R}$, these conditions are equivalent:*

(i) *f is space-time invariant,*
(ii) *$M_t = f(X_t, s+t)$ is a P_μ-martingale for every μ and $s \geq 0$.*

Proof, (i) ⇒ (ii): Assume (i), and let $s, t, h \geq 0$. Using the Markov property of X, we get

$$\begin{aligned} E_\mu\Big(M_{t+h}\,\Big|\,\mathcal{F}_t\Big) &= E_\mu\Big\{f\Big(X_{t+h},\, s+t+h\Big)\Big|\,\mathcal{F}_t\Big\} \\ &= E_{X_t} f\Big(X_h,\, s+t+h\Big) \\ &= f(X_t, s+t) = M_t. \end{aligned}$$

(ii) ⇒ (i): By (ii) with $\mu = \delta_x$, we have for any $x \in S$ and $s, t \geq 0$

$$\begin{aligned} E_x f(X_t, s+t) &= E_x M_t \\ &= E_x M_0 \\ &= f(x,s). \end{aligned}$$

□

The *tail σ-field* on Ω is defined as $\mathcal{T} = \bigcap_t \mathcal{T}_t$, where $\mathcal{T}_t = \sigma(\theta_t) = \sigma\{X_s;\, s \geq t\}$. A σ-field $\mathcal{G}$ on Ω is said to be *P_μ-trivial*, if $P_\mu A = 0$ or 1 for every $A \in \mathcal{G}$. Writing $P_\mu^B = P_\mu(\,\cdot\,|B)$, we say that P_μ is *mixing*, if

$$\lim_{t\to\infty} \Big\| P_\mu \circ \theta_t^{-1} - P_\mu^B \circ \theta_t^{-1} \Big\| = 0, \quad B \in \mathcal{F}_\infty \text{ with } P_\mu B > 0.$$

The following properties define the notion of *strong ergodicity*, as opposed to the weak ergodicity in Theorem 26.11.

Theorem 26.10 *(strong ergodicity and mixing, Orey) For a conservative Markov semi-group on R_+ or Z_+ with distributions P_μ, these conditions are equivalent:*

(i) *the tail σ-field $\mathcal{T}$ is P_μ-trivial for every μ,*
(ii) *P_μ is mixing for every μ,*
(iii) *every bounded, space-time invariant function is a constant,*
(iv) *$\Big\| P_\mu \circ \theta_t^{-1} - P_\nu \circ \theta_t^{-1} \Big\| \to 0$ as $t \to \infty$ for all μ and ν.*

First proof: By Theorem 25.25 (i), conditions (ii) and (iv) are equivalent to respectively

(ii′) $P_\mu = P_\mu^B$ on $\mathcal{T}$ for all μ and B,
(iv′) $P_\mu = P_\nu$ on $\mathcal{T}$ for all μ, ν.

It is then enough to prove that (ii′) ⇔ (i) ⇒ (iv′) and (iv) ⇒ (iii) ⇒ (i).

(i) ⇔ (ii′): If $P_\mu A = 0$ or 1, then clearly also $P_\mu^B A = 0$ or 1, which shows that (i) ⇒ (ii′). Conversely, let $A \in \mathcal{T}$ be arbitrary with $P_\mu A > 0$. Taking $B = A$ in (ii′) gives $P_\mu A = (P_\mu A)^2$, which implies $P_\mu A = 1$.

(i) ⇒ (iv′): Applying (i) to the distribution $\frac{1}{2}(\mu + \nu)$ gives $P_\mu A + P_\nu A = 0$ or 2 for every $A \in \mathcal{T}$, which implies $P_\mu A = P_\nu A = 0$ or 1.

(iv) ⇒ (iii): Let f be bounded and space-time invariant. Using (iv) with $\mu = \delta_x$ and $\nu = \delta_y$ gives

$$\begin{aligned}\big|f(x,s) - f(y,s)\big| &= \big|E_x f(X_t,\, s+t) - E_y f(X_t,\, s+t)\big| \\ &\le \|f\|\, \big\|P_x \circ \theta_t^{-1} - P_y \circ \theta_t^{-1}\big\| \to 0,\end{aligned}$$

which shows that $f(x,s) = f(s)$ is independent of x. Then $f(s) = f(s+t)$ by (5), and so f is a constant.

(iii) ⇒ (i): Fix any $A \in \mathcal{T}$. Since $A \in \mathcal{T}_t = \sigma(\theta_t)$ for every $t \ge 0$, we have $A = \theta_t^{-1} A_t$ for some sets $A_t \in \mathcal{F}_\infty$, which are unique since the θ_t are surjective. For any $s, t \ge 0$, we have

$$\begin{aligned}\theta_t^{-1}\theta_s^{-1} A_{s+t} &= \theta_{s+t}^{-1} A_{s+t} \\ &= A = \theta_t^{-1} A_t,\end{aligned}$$

and so $\theta_s^{-1} A_{s+t} = A_t$. Putting $f(x,t) = P_x A_t$ and using the Markov property at time s, we get

$$\begin{aligned}E_x f(X_s, s+t) &= E_x P_{X_s} A_{s+t} \\ &= E_x P_x\big\{\theta_s^{-1} A_{s+t} \,\big|\, \mathcal{F}_s\big\} \\ &= P_x A_t = f(x,t).\end{aligned}$$

Thus, f is space-time invariant and hence a constant $c \in [0,1]$. By the Markov property at t and martingale convergence as $t \to \infty$, we have a.s.

$$\begin{aligned}c &= f(X_t, t) = P_{X_t} A_t \\ &= P_\mu\big\{\theta_t^{-1} A_t \,\big|\, \mathcal{F}_t\big\} \\ &= P_\mu(A \,|\, \mathcal{F}_t) \to 1_A,\end{aligned}$$

which implies $P_\mu A = c \in \{0,1\}$. This shows that $\mathcal{T}$ is P_μ-trivial. □

Second proof: We can avoid to use the rather deep Theorem 25.25 by giving direct proofs of the implications (i) ⇒ (ii) ⇒ (iv).

(i) ⇒ (ii): Assuming (i), we get by reverse martingale convergence

$$\begin{aligned}\big\|P_\mu(\cdot \cap B) - P_\mu(\cdot)\, P_\mu(B)\big\|_{\mathcal{T}_t} &= \big\|E_\mu\big\{P_\mu(B \,|\, \mathcal{T}_t) - P_\mu B;\, \cdot\big\}\big\|_{\mathcal{T}_t} \\ &\le E_\mu\big|P_\mu(B \,|\, \mathcal{T}_t) - P_\mu B\big| \to 0.\end{aligned}$$

(ii) ⇒ (iv): Let $\mu' - \nu'$ be the Hahn decomposition of $\mu - \nu$, and choose $B \in \mathcal{S}$ with $\mu' B^c = \nu B = 0$. Writing $\chi = \mu' + \nu'$ and $A = \{X_0 \in B\}$, we get by (ii)

$$\left\| P_\mu \circ \theta_t^{-1} - P_\nu \circ \theta_t^{-1} \right\| = \|\mu'\| \left\| P_\chi^A \circ \theta_t^{-1} - P_\chi^{A^c} \circ \theta_t^{-1} \right\| \to 0. \qquad \square$$

The *invariant σ-field* $\mathcal{I}$ on Ω consists of all events $A \subset \Omega$ with $\theta_t^{-1} A = A$ for all $t \geq 0$. Note that a random variable ξ on Ω is $\mathcal{I}$-measurable iff $\xi \circ \theta_t = \xi$ for all $t \geq 0$. The invariant σ-field $\mathcal{I}$ is clearly contained in the tail σ-field $\mathcal{T}$. We say that P_μ is *weakly mixing* if

$$\lim_{t \to \infty} \left\| \int_0^1 \left(P_\mu - P_\mu^B \right) \circ \theta_{st}^{-1}\, ds \right\| = 0, \quad B \in \mathcal{F}_\infty \text{ with } P_\mu B > 0,$$

with the understanding that $\theta_s = \theta_{[s]}$ when the time scale is discrete. We may now state a weak counterpart of Theorem 26.10.

Theorem 26.11 *(weak ergodicity and mixing) For a conservative Markov semi-group on R_+ or Z_+ with distributions P_μ, these conditions are equivalent:*

(i) *the invariant σ-field $\mathcal{I}$ is P_μ-trivial for every μ,*
(ii) *P_μ is weakly mixing for every μ,*
(iii) *every bounded, invariant function is a constant,*
(iv) $\left\| \int_0^1 (P_\mu - P_\nu) \circ \theta_{st}^{-1} ds \right\| \to 0$ *as $t \to \infty$ for all μ, ν.*

Proof: By Theorem 25.25 (ii), we note that (ii) and (iv) are equivalent to the conditions

(ii′) $P_\mu = P_\mu^B$ on $\mathcal{I}$ for all μ and B,
(iv′) $P_\mu = P_\nu$ on $\mathcal{I}$ for all μ, ν.

Here (ii′) ⇔ (i) ⇒ (iv′) may be established as before, and it only remains to show that (iv) ⇒ (iii) ⇒ (i).

(iv) ⇒ (iii): Let f be bounded and invariant. Then $f(x) = E_x f(X_t) = T_t f(x)$, and therefore $f(x) = \int_0^1 T_{st} f(x)\, ds$. Using (iv) gives

$$\begin{aligned} \left| f(x) - f(y) \right| &= \left| \int_0^1 \left\{ T_{st} f(x) - T_{st} f(y) \right\} ds \right| \\ &\leq \|f\| \left\| \int_0^1 (P_x - P_y) \circ \theta_{st}^{-1}\, ds \right\| \to 0, \end{aligned}$$

which shows that f is constant.

(iii) ⇒ (i): Fix any $A \in \mathcal{I}$, and define $f(x) = P_x A$. Using the Markov property at t and the invariance of A, we get

$$\begin{aligned} E_x f(X_t) &= E_x P_{X_t} A \\ &= E_x P_x \left\{ \theta_t^{-1} A \,\middle|\, \mathcal{F}_t \right\} \\ &= P_x \theta_t^{-1} A \\ &= P_x A = f(x), \end{aligned}$$

which shows that f is invariant. By (iii) it follows that f equals a constant $c \in [0,1]$. Hence, by the Markov property and martingale convergence, we have a.s.

$$\begin{aligned} c = f(X_t) &= P_{X_t} A \\ &= P_\mu\Big\{ \theta_t^{-1} A \Big| \mathcal{F}_t \Big\} \\ &= P_\mu(A \,|\, \mathcal{F}_t) \to 1_A, \end{aligned}$$

which implies $P_\mu A = c \in \{0,1\}$. Thus, $\mathcal{I}$ is P_μ-trivial. $\square$

We now specialize to conservative Feller processes X with distributions P_x, defined on an lcscH[1] space S with Borel σ-field $\mathcal{S}$. Say that the process is *regular*, if there exist a locally finite measure ρ on S and a continuous function $(x,y,t) \mapsto p_t(x,y) > 0$ on $S^2 \times (0,\infty)$, such that

$$P_x\{X_t \in B\} = \int_B p_t(x,y)\, \rho(dy), \quad x \in S,\ B \in \mathcal{S},\ t > 0.$$

The *supporting measure* ρ is then unique up to an equivalence, and $\operatorname{supp} \rho = S$ by the Feller property. A Feller process is said to be *Harris recurrent*, if it is regular with a supporting measure ρ satisfying

$$\int_0^\infty 1_B(X_t)\, dt = \infty \ \text{ a.s. } P_x, \quad x \in S,\ B \in \mathcal{S} \text{ with } \rho B > 0. \tag{6}$$

Theorem 26.12 *(Harris recurrence and ergodicity, Orey) For a Feller process X, we have*

$$X \textit{ is Harris recurrent} \quad \Rightarrow \quad X \textit{ is strongly ergodic.}$$

Proof: By Theorem 26.10 it suffices to prove that any bounded, space-time invariant function $f\colon S \times \mathsf{R}_+ \to \mathsf{R}$ is a constant. First we show for fixed $x \in S$ that $f(x,t)$ is independent of t. Then assume that instead $f(x,h) \neq f(x,0)$ for some $h > 0$, say $f(x,h) > f(x,0)$. Recall from Lemma 26.9 that $M_t^s = f(X_t, s+t)$ is a P_y-martingale for any $y \in S$ and $s \geq 0$. In particular, the limit M_∞^s exists a.s. along $h\mathsf{Q}$, and we get a.s. P_x

$$\begin{aligned} E_x\Big(M_\infty^h - M_\infty^0 \Big| \mathcal{F}_0\Big) &= M_0^h - M_0^0 \\ &= f(x,h) - f(x,0) > 0, \end{aligned}$$

which implies $P_x\{M_\infty^h > M_\infty^0\} > 0$. We may then choose some constants $a < b$ such that

$$P_x\Big\{M_\infty^0 < a < b < M_\infty^h\Big\} > 0. \tag{7}$$

We also note that

$$\begin{aligned} M_t^{s+h} \circ \theta_s &= f\Big(X_{s+t},\ s+t+h\Big) \\ &= M_{s+t}^h, \quad s,t,h \geq 0. \end{aligned} \tag{8}$$

[1] locally compact, second countable, and Hausdorff

Restricting s, t to $h\,\mathsf{Q}_+$, we define

$$g(y,s) = P_y \bigcap_{t\geq 0} \left\{ M_t^s \leq a < b \leq M_t^{s+h} \right\}, \quad y \in S, \;\; s \geq 0.$$

Using (8) and the Markov property at s, we get a.s. P_x for any $r \leq s$

$$\begin{aligned} g(X_s, s) &= P_{X_s} \bigcap_{t\geq 0} \left\{ M_t^s \leq a < b \leq M_t^{s+h} \right\} \\ &= P_x^{\mathcal{F}_s} \bigcap_{t\geq 0} \left\{ M_t^s \circ \theta_s \leq a < b \leq M_t^{s+h} \circ \theta_s \right\} \\ &= P_x^{\mathcal{F}_s} \bigcap_{t\geq s} \left\{ M_t^0 \leq a < b \leq M_t^h \right\} \\ &\geq P_x^{\mathcal{F}_s} \bigcap_{t\geq r} \left\{ M_t^0 \leq a < b \leq M_t^h \right\}. \end{aligned}$$

By martingale convergence we get, a.s. as $s \to \infty$ along $h\mathsf{Q}$ and then $r \to \infty$,

$$\begin{aligned} \liminf_{s\to\infty} g(X_s, s) &\geq \liminf_{t\to\infty} 1\Big\{ M_t^0 \leq a < b \leq M_t^h \Big\} \\ &\geq 1\Big\{ M_\infty^0 < a < b < M_\infty^h \Big\}, \end{aligned}$$

and so by (7),

$$P_x\Big\{ g(X_s, s) \to 1 \Big\} \geq P_x\Big\{ M_\infty^0 < a < b < M_\infty^h \Big\} > 0. \tag{9}$$

Now fix any non-empty, bounded, open set $B \subset S$. Using (6) and the right-continuity of X, we note that $\limsup_s 1_B(X_s) = 1$ a.s. P_x, and so by (9)

$$P_x\Big\{ \limsup_{s\to\infty} 1_B(X_s)\, g(X_s, s) = 1 \Big\} > 0. \tag{10}$$

Furthermore, we have by regularity

$$p_h(u,v) \wedge p_{2h}(u,v) \geq \varepsilon > 0, \quad u, v \in B, \tag{11}$$

for some $\varepsilon > 0$. By (10) we may choose some $y \in B$ and $s \geq 0$ with $g(y,s) > 1 - \frac{1}{2}\,\varepsilon\,\rho B$. Define for $i = 1, 2$

$$B_i = B \setminus \Big\{ u \in S;\; f(u,\, s + ih) \leq a < b \leq f\Big(u,\, s + (i+1)h\Big) \Big\}.$$

Using (11), the definitions of B_i, M^s, g, and the properties of y and s, we get

$$\begin{aligned} \varepsilon\,\rho B_i &\leq P_y\{X_{ih} \in B_i\} \\ &\leq 1 - P_y\Big\{ f\Big(X_{ih},\, s + ih\Big) \leq a < b \leq f\Big(X_{ih},\, s + (i+1)h\Big) \Big\} \\ &= 1 - P_y\Big\{ M_{ih}^s \leq a < b \leq M_{ih}^{s+h} \Big\} \\ &\leq 1 - g(y,s) \\ &< \tfrac{1}{2}\,\varepsilon\,\rho B. \end{aligned}$$

Thus, $\rho B_1 + \rho B_2 < \rho B$, and there exists a $u \in B \setminus (B_1 \cup B_2)$. But this yields the contradiction $a < b \leq f(u, s + 2h) \leq a$, showing that $f(x,t) = f(x)$ is indeed independent of t.

To see that $f(x)$ is also independent of x, assume that instead

$$\rho\Big\{x;\, f(x) \le a\Big\} \wedge \rho\Big\{x;\, f(x) \ge b\Big\} > 0$$

for some $a < b$. Then by (6), the martingale $M_t = f(X_t)$ satisfies

$$\int_0^\infty 1\{M_t \le a\}\, dt = \int_0^\infty 1\{M_t \ge b\}\, dt = \infty \text{ a.s.} \tag{12}$$

Writing $\tilde{M}$ for the right-continuous version of M, which exists by Theorem 9.28 (ii), we get by Fubini's theorem for any $x \in S$

$$\begin{aligned}
\sup_{u>0} E_x\Big| \int_0^u 1\{M_t \le a\}\, dt - \int_0^u 1\{\tilde{M}_t \le a\}\, dt\Big| \\
\le \int_0^\infty E_x\Big|1\{M_t \le a\} - 1\{\tilde{M}_t \le a\}\Big|\, dt \\
\le \int_0^\infty P_x\Big\{M_t \ne \tilde{M}_t\Big\}\, dt = 0,
\end{aligned}$$

and similarly for the events $M_t \ge b$ and $\tilde{M}_t \ge b$. Thus, the integrals on the left agree a.s. P_x for all x, and so (12) remains true with M replaced by $\tilde{M}$. In particular,

$$\liminf_{t\to\infty} \tilde{M}_t \le a < b \le \limsup_{t\to\infty} \tilde{M}_t \text{ a.s. } P_x.$$

But this is impossible, since $\tilde{M}$ is a bounded, right-continuous martingale and hence converges a.s. The contradiction shows that $f(x) = c$ a.e. ρ for some constant $c \in \mathsf{R}$. Then for any $t > 0$,

$$\begin{aligned}
f(x) &= E_x f(X_t) \\
&= \int f(y)\, p_t(x,y)\, \rho(dy) = c, \quad x \in S,
\end{aligned}$$

and so $f(x)$ is indeed independent of x. □

Our further analysis of regular Feller processes requires some potential theory. For any measurable functions $f, h \ge 0$ on S, we define the h-*potential* $U_h f$ of f by

$$U_h f(x) = E_x \int_0^\infty e^{-A_t^h} f(X_t)\, dt, \quad x \in S,$$

where A^h denotes the elementary additive functional

$$A_t^h = \int_0^t h(X_s)\, ds, \quad t \ge 0.$$

When h is a constant $a \ge 0$, we note that $U_h = U_a$ agrees with the resolvent operator R_a of the semigroup (T_t), in which case

$$\begin{aligned}
U_a f(x) &= E_x \int_0^\infty e^{-at} f(X_t)\, dt \\
&= \int_0^\infty e^{-at}\, T_t f(x)\, dt, \quad x \in S.
\end{aligned}$$

The classical resolvent equation extends to general h-potentials, as follows.

Lemma 26.13 *(resolvent equation) Let $f \ge 0$ and $h \ge k \ge 0$ be measurable functions on S, where h is bounded. Then $U_h h \le 1$, and*

$$\begin{aligned} U_k f &= U_h f + U_h(h-k)\, U_k f \\ &= U_h f + U_k(h-k)\, U_h f. \end{aligned}$$

Proof: Define $F = f(X)$, $H = h(X)$, and $K = k(X)$ for convenience. By Itô's formula for continuous functions of bounded variation,

$$e^{-A_t^h} = 1 - \int_0^t e^{-A_s^h}\, H_s\, ds, \quad t \ge 0, \tag{13}$$

which implies $U_h h \le 1$. We also see from the Markov property of X that a.s.

$$\begin{aligned} U_k f(X_t) &= E_{X_t} \int_0^\infty e^{-A_s^k}\, F_s\, ds \\ &= E_x^{\mathcal{F}_t} \int_0^\infty e^{-A_s^k \circ \theta_t}\, F_{s+t}\, ds \\ &= E_x^{\mathcal{F}_t} \int_t^\infty e^{-A_u^k + A_t^k}\, F_u\, du. \end{aligned} \tag{14}$$

Using (13) and (14) together with Fubini's theorem, we get

$$\begin{aligned} U_h(h-k)\, U_k f(x) &= E_x \int_0^\infty e^{-A_t^h}\, (H_t - K_t)\, dt \int_t^\infty e^{-A_u^k + A_t^k}\, F_u\, du \\ &= E_x \int_0^\infty e^{-A_u^k}\, F_u\, du \int_0^u e^{-A_t^h + A_t^k}\, (H_t - K_t)\, dt \\ &= E_x \int_0^\infty e^{-A_u^k} F_u \Big(1 - e^{-A_u^h + A_u^k}\Big)\, du \\ &= E_x \int_0^\infty \Big(e^{-A_u^k} - e^{-A_u^h}\Big)\, F_u\, du \\ &= U_k f(x) - U_h f(x). \end{aligned}$$

A similar calculation yields the same expression for $U_k(h-k)\, U_h f(x)$. □

For a simple application of the resolvent equation, we show that any bounded potential function $U_h f$ is continuous.

Lemma 26.14 *(boundedness and continuity) For a regular Feller process on S and any bounded, measurable functions $f, h \ge 0$ on S, we have*

$$U_h f \text{ is bounded} \quad \Rightarrow \quad U_h f \text{ is continuous.}$$

Proof: Using Fatou's lemma and the continuity of $p_t(\cdot, y)$, we get for any time $t > 0$ and sequence $x_n \to x$ in S

$$\begin{aligned} \liminf_{n\to\infty} T_t f(x_n) &= \liminf_{n\to\infty} \int p_t(x_n, y) f(y)\, \rho(dy) \\ &\ge \int p_t(x, y) f(y)\, \rho(dy) \\ &= T_t f(x). \end{aligned}$$

If $f \le c$, the same relation applies to the function $T_t(c-f) = c - T_t f$, and so by combination $T_t f$ is continuous. By dominated convergence, $U_a f$ is then continuous for every $a > 0$.

Now let $h \le a$. Applying the previous result to the bounded, measurable function $(a-h)\, U_h f \ge 0$, we see that even $U_a(a-h)\, U_h f$ is continuous. The continuity of $U_h f$ now follows from Lemma 26.13, with h and k replaced by a and h. □

We proceed with some useful estimates.

Lemma 26.15 *(lower bounds) For a regular Feller process on S, there exist some continuous functions $h, k : S \to (0,1]$, such that for any measurable function $f \ge 0$ on S,*

(i) $U_2 f(x) \ge \rho(kf)\, h(x)$,

(ii) $U_h f(x) \ge \rho(kf)\, U_h h(x)$.

Proof: Fix any compact sets $K \subset S$ and $T \subset (0,\infty)$ with $\rho K > 0$ and $\lambda T > 0$. Define

$$u_a^T(x,y) = \int_T e^{-at}\, p_t(x,y)\, dt, \quad x, y \in S,$$

and note that for any measurable function $f \ge 0$ on S,

$$U_a f(x) \ge \int u_a^T(x,y)\, f(y)\, \rho(dy), \quad x \in S.$$

Applying Lemma 26.13 to the constant functions 4 and 2 gives

$$\begin{aligned} U_2 f(x) &= U_4 f(x) + 2\, U_4 U_2 f(x) \\ &\ge 2 \int_K u_4^T(x,y)\, \rho(dy) \int u_2^T(y,z)\, f(z)\, \rho(dz), \end{aligned}$$

and (i) follows with

$$\begin{aligned} h(x) &= 2 \int_K u_4^T(x,y)\, \rho(dy) \wedge 1, \\ k(x) &= \inf_{y \in K} u_2^T(y,x) \wedge 1. \end{aligned}$$

To deduce (ii), we may combine (i) with Lemma 26.13 for the functions 2 and h, to obtain

$$\begin{aligned} U_h f(x) &\ge U_h(2-h)\, U_2 f(x) \\ &\ge U_h U_2 f(x) \\ &\ge U_h h(x)\, \rho(kf). \end{aligned}$$

The continuity of h is clear by dominated convergence. For the same reason, the function u_2^T is jointly continuous on S^2. Since K is compact, the functions $u_2^T(y,\cdot)$, $y \in K$, are then equi-continuous on S, which yields the required continuity of k. Finally, the relation $h > 0$ is obvious, whereas $k > 0$ holds by the compactness of K. □

Fixing a function h as in Lemma 26.15, we introduce the kernel

$$Q_x B = U_h(h 1_B)(x), \quad x \in S, \ B \in \mathcal{S}, \tag{15}$$

and note that $Q_x S = U_h h(x) \le 1$ by Lemma 26.13.

Lemma 26.16 *(convergence dichotomy) For h as in Lemma* 26.15*, define Q by* (15)*. Then*

(i) *when $U_h h \not\equiv 1$, there exists an $r \in (0,1)$ with*
$$\|Q^n S\| \le r^{n-1}, \quad n \in \mathsf{N},$$

(ii) *when $U_h h \equiv 1$, there exists a Q-invariant distribution $\nu \sim \rho$ with*
$$\left\|Q_x^n - \nu\right\| \to 0, \quad x \in S,$$
and every σ-finite, Q-invariant measure on S is proportional to ν.

Proof: (i) Choose k as in Lemma 26.15, fix any $a \in S$ with $U_h h(a) < 1$, and define
$$r = 1 - h(a)\,\rho\Big\{hk\,(1 - U_h h)\Big\}.$$

Note that $\rho\{hk\,(1-U_h h)\} > 0$, since $U_h h$ is continuous by Lemma 26.14. Using Lemma 26.15 (i), we obtain
$$\begin{aligned} 0 &< 1 - r \\ &\le h(a)\,\rho k \\ &\le U_2 1(a) = \tfrac{1}{2}. \end{aligned}$$

Next, Lemma 26.15 (ii) yields
$$\begin{aligned} (1-r)\,U_h h &= h(a)\,\rho\Big\{hk\,(1 - U_h h)\Big\}\,U_h h \\ &\le U_h h(1 - U_h h). \end{aligned}$$

Hence,
$$\begin{aligned} Q^2 S &= U_h h\, U_h h \\ &\le r\, U_h h \\ &= r\, QS, \end{aligned}$$

and so by iteration
$$\begin{aligned} Q^n S &\le r^{n-1} QS \\ &\le r^{n-1}, \quad n \in \mathsf{N}. \end{aligned}$$

(ii) Introduce a measure $\tilde\rho = hk \cdot \rho$ on S. Since $U_h h = 1$, Lemma 26.15 (ii) yields
$$\begin{aligned} \tilde\rho B &= \rho(hk\,1_B) \\ &\le U_h(h\,1_B)(x) \\ &= Q_x B, \quad B \in \mathcal{S}. \end{aligned} \tag{16}$$

Regarding $\tilde\rho$ as a kernel, we have for any $x, y \in S$ and $m, n \in \mathsf{Z}_+$

$$\Big(Q_x^m - Q_y^n\Big)\tilde\rho^k = \Big(Q_x^m S - Q_y^n S\Big)\tilde\rho^k = 0.$$

Iterating this and using (16), we get as $n \to \infty$

$$\begin{aligned} \left\| Q_x^n - Q_y^{n+k} \right\| &= \left\| (\delta_x - Q_y^k) Q^n \right\| \\ &= \left\| (\delta_x - Q_y^k)(Q - \tilde{\rho})^n \right\| \\ &\le \left\| \delta_x - Q_y^k \right\| \sup_z \left\| Q_z - \tilde{\rho} \right\|^n \\ &\le 2\,(1 - \tilde{\rho}S)^n \to 0. \end{aligned}$$

Hence, $\sup_x \|Q_x^n - \nu\| \to 0$ for some set function ν on S, and we note that ν is a Q-invariant probability measure. By Fubini's theorem, we have $Q_x \ll \rho$ for all x, and so $\nu = \nu Q \ll \rho$. Conversely, Lemma 26.15 (ii) yields

$$\begin{aligned} \nu B &= \nu Q(B) \\ &= \nu \Big\{ U_h(h 1_B) \Big\} \\ &\ge \rho(hk1_B), \end{aligned}$$

which shows that even $\rho \ll \nu$.

Now consider any σ-finite, Q-invariant measure μ on S. By Fatou's lemma, we get for any $B \in \mathcal{S}$

$$\begin{aligned} \mu B &= \liminf_{n\to\infty} \mu Q^n B \\ &\ge \mu \nu B \\ &= \mu S\, \nu B. \end{aligned}$$

Choosing B with $\mu B < \infty$ and $\nu B > 0$, we obtain $\mu S < \infty$. By dominated convergence, we conclude that $\mu = \mu Q^n \to \mu\nu = \mu S\, \nu$, which proves the asserted uniqueness of ν. □

We may now establish the basic recurrence dichotomy for regular Feller processes. Write $U = U_0$, and say that X is *uniformly transient* if

$$\|U 1_K\| = \sup_x E_x \int_0^\infty 1_K(X_t)\, dt < \infty, \quad K \subset S \text{ compact}.$$

Theorem 26.17 *(recurrence dichotomy) A regular Feller process is either Harris recurrent or uniformly transient.*

Proof: Choose h, k as in Lemma 26.15 and Q, r, ν as in Lemma 26.16. First assume $U_h h \not\equiv 1$. Letting $a \in (0, \|h\|]$, we note that $ah \le (h \wedge a)\|h\|$, and hence

$$a\, U_{h\wedge a} h \le U_{h\wedge a}(h \wedge a)\, \|h\| \le \|h\|. \tag{17}$$

Furthermore, Lemma 26.13 yields

$$\begin{aligned} U_{h\wedge a} h &\le U_h h + U_h h U_{h\wedge a} h \\ &= Q\Big(1 + U_{h\wedge a} h\Big). \end{aligned}$$

Iterating this relation and using Lemma 26.16 (i) and (17), we get

$$\begin{aligned} U_{h\wedge a} h &\le \sum_{k\le n} Q^k 1 + Q^n U_{h\wedge a} h \\ &\le \sum_{k\le n} r^{k-1} + r^{n-1} \Big\| U_{h\wedge a} h \Big\| \\ &\le (1 - r)^{-1} + r^{n-1} \|h\|/a. \end{aligned}$$

Letting $n \to \infty$ and then $a \to 0$, we conclude by dominated and monotone convergence that $Uh \le (1-r)^{-1}$. Now fix any compact set $K \subset S$. Since $b \equiv \inf_K h > 0$, we get for all $x \in S$

$$\begin{aligned} U1_K(x) &\le b^{-1}\,Uh(x) \\ &\le b^{-1}(1-r)^{-1} < \infty, \end{aligned}$$

which shows that X is uniformly transient.

Now assume that instead $U_h h \equiv 1$. Fix any measurable function f on S with $0 \le f \le h$ and $\rho f > 0$, and put $g = 1 - U_f f$. By Lemma 26.13, we get

$$\begin{aligned} g &= 1 - U_f f \\ &= U_h h - U_h f - U_h(h-f)U_f f \\ &= U_h(h-f)(1-U_f f) \\ &= U_h(h-f)g \\ &\le U_h hg = Qg. \end{aligned} \tag{18}$$

Iterating this relation and using Lemma 26.16 (ii), we obtain $g \le Q^n g \to \nu g$, where $\nu \sim \rho$ is the unique Q-invariant distribution on S. Inserting this into (18) gives $g \le U_h(h-f)\nu g$, and so by Lemma 26.15 (ii),

$$\begin{aligned} \nu g &\le \nu\Big\{U_h(h-f)\Big\}\,\nu g \\ &\le \Big\{1-\rho(kf)\Big\}\,\nu g. \end{aligned}$$

Since $\rho(kf) > 0$, we obtain $\nu g = 0$, and so $U_f f = 1 - g = 1$ a.e. $\nu \sim \rho$. Since $U_f f$ is continuous by Lemma 26.14 and $\operatorname{supp}\rho = S$, we obtain $U_f f \equiv 1$. Taking expected values in (13), we conclude that $A^f_\infty = \infty$ a.s. P_x for every $x \in S$. Now fix any compact set $K \subset S$ with $\rho K > 0$. Since $b \equiv \inf_K h > 0$, we may choose $f = b\,1_K$, and the desired Harris recurrence follows. □

A measure λ on S is said to be *invariant* for the semi-group (T_t), if $\lambda(T_t f) = \lambda f$ for all $t > 0$ and every measurable function $f \ge 0$ on S. In the Harris recurrent case, the existence of an invariant measure λ can be inferred from Lemma 26.16.

Theorem 26.18 *(invariant measure, Harris, Watanabe) For a Harris recurrent Feller process on S with supporting measure ρ,*

(i) *there exists a locally finite, invariant measure $\lambda \sim \rho$,*

(ii) *every σ-finite, invariant measure agrees with λ up to a normalization.*

To prepare for the proof, we first express the required invariance in terms of the resolvent operators.

Lemma 26.19 *(invariance equivalence) Let (T_t) be a Feller semi-group on S with resolvent (U_a), and fix a locally finite measure λ on S and a constant $c > 0$. Then these conditions are equivalent:*

(i) *λ is T_t-invariant for every $t > 0$,*
(ii) *λ is $a\,U_a$-invariant for every $a \geq c$.*

Proof, (i) $\Rightarrow$ (ii): Assuming (i), we get by Fubini's theorem for any measurable function $f \geq 0$ and constant $a > 0$

$$\begin{aligned} \lambda(U_a f) &= \int_0^\infty e^{-at}\lambda(T_t f)\,dt \\ &= \int_0^\infty e^{-at}\lambda f\,dt \\ &= a^{-1}\lambda f. \end{aligned} \tag{19}$$

(ii) $\Rightarrow$ (i): Assume (ii). Then for any measurable function $f \geq 0$ on S with $\lambda f < \infty$, the integrals in (19) agree for all $a \geq c$. Hence, by Theorem 6.3 the measures $\lambda(T_t f)\,e^{-ct}\,dt$ and $\lambda f\,e^{-ct}\,dt$ agree on R_+, which implies $\lambda(T_t f) = \lambda f$ for almost every $t \geq 0$. By the semi-group property and Fubini's theorem, we then obtain for any $t \geq 0$

$$\begin{aligned} \lambda(T_t f) &= c\,\lambda U_c(T_t f) \\ &= c\,\lambda \int_0^\infty e^{-cs}\,T_s T_t f\,ds \\ &= c\int_0^\infty e^{-cs}\lambda(T_{s+t} f)\,ds \\ &= c\int_0^\infty e^{-cs}\lambda f\,ds = \lambda f. \end{aligned}$$

□

Proof of Theorem 26.18: (i) Let h, Q, ν be such as in Lemmas 26.15 and 26.16, and put $\lambda = h^{-1}\cdot\nu$. Using the definition of λ (twice), the Q-invariance of ν (three times), and Lemma 26.13, we get for any constant $a \geq \|h\|$ and bounded, measurable function $f \geq 0$ on S

$$\begin{aligned} a\,\lambda U_a f &= a\,\nu(h^{-1}U_a f) \\ &= a\,\nu U_h U_a f \\ &= \nu\Big(U_h f - U_a f + U_h h U_a f\Big) \\ &= \nu U_h f \\ &= \nu(h^{-1}f) = \lambda f, \end{aligned}$$

which shows that λ is $a\,U_a$-invariant for every such a. By Lemma 26.19 it follows that λ is also (T_t)-invariant.

(ii) Consider any σ-finite, (T_t)-invariant measure λ' on S. By Lemma 26.19, λ' is even $a\,U_a$-invariant for every $a \geq \|h\|$. Now define $\nu' = h\cdot\lambda'$. Letting $f \geq 0$ be bounded and measurable on S and using Lemma 26.13, we get as before

$$\begin{aligned} \nu' U_h(hf) &= \lambda'(hU_h(hf)) \\ &= a\,\lambda' U_a h U_h(hf) \\ &= a\,\lambda'\Big\{U_a(hf) - U_h(hf) + a\,U_a U_h(hf)\Big\} \\ &= a\,\lambda' U_a(hf) \\ &= \lambda'(hf) = \nu' f, \end{aligned}$$

which shows that ν' is Q-invariant. Hence, the uniqueness in Lemma 26.16 (ii) yields $\nu' = c\,\nu$ for some constant $c \geq 0$, which implies $\lambda' = c\,\lambda$. □

A Harris recurrent Feller process is said to be *positive-recurrent* if the invariant measure λ is bounded, and *null-recurrent* otherwise. In the former case, we may choose λ to be a probability measure on S. For any process X in S, the divergence $X_t \to \infty$ a.s. or $X_t \xrightarrow{P} \infty$ means that $1_K(X_t) \to 0$ in the same sense for every compact set $K \subset S$.

Theorem 26.20 *(distributional limits) Let X be a regular Feller process in S. Then for any distribution μ on S, we have as $t \to \infty$*

(i) *when X is positive-recurrent with invariant distribution λ,*

$$\left\| P_\mu^A \circ \theta_t^{-1} - P_\lambda \right\| \to 0, \quad A \in \mathcal{F}_\infty \ \textit{with} \ P_\mu A > 0,$$

(ii) *when X is null-recurrent or transient,*

$$X_t \xrightarrow{P_\mu} \infty.$$

Proof: (i) Since $P_\lambda \circ \theta_t^{-1} = P_\lambda$ by Lemma 11.11, the assertion follows from Theorem 26.12, together with properties (ii) and (iv) of Theorem 26.10.

(ii) (*null-recurrent case*) For any compact set $K \subset S$ and constant $\varepsilon > 0$, define

$$B_t = \left\{ x \in S;\ T_t 1_K(x) \geq \mu T_t 1_K - \varepsilon \right\}, \quad t > 0,$$

and note that for any invariant measure λ,

$$\begin{aligned} \left(\mu T_t 1_K - \varepsilon\right) \lambda B_t &\leq \lambda(T_t 1_K) \\ &= \lambda K < \infty. \end{aligned} \tag{20}$$

Since $\mu T_t 1_K - T_t 1_K(x) \to 0$ for all $x \in S$ by Theorem 26.12, we have $\liminf_t B_t = S$, and so $\lambda B_t \to \infty$ by Fatou's lemma. Hence, (20) yields $\limsup_t \mu T_t 1_K \leq \varepsilon$, and ε being arbitrary, we obtain $P_\mu\{X_t \in K\} = \mu T_t 1_K \to 0$.

(ii) (*transient case*) Fix any compact set $K \subset S$ with $\rho K > 0$, and conclude from the uniform transience of X that $U 1_K$ is bounded. Hence, by the Markov property at t and dominated convergence,

$$\begin{aligned} E_\mu U 1_K(X_t) &= E_\mu E_{X_t} \int_0^\infty 1_K(X_s)\, ds \\ &= E_\mu \int_t^\infty 1_K(X_s)\, ds \to 0, \end{aligned}$$

which shows that $U 1_K(X_t) \xrightarrow{P_\mu} 0$. Since $U 1_K$ is strictly positive and also continuous by Lemma 26.14, we conclude that $X_t \xrightarrow{P_\mu} \infty$. □

We conclude with a pathwise limit theorem for regular Feller processes. Here 'almost surely' means a.s. P_μ for every initial distribution μ on S.

Theorem 26.21 *(pathwise limits) For a regular Feller process X in S, we have as $t \to \infty$*

(i) *when X is positive-recurrent with invariant distribution λ,*

$$t^{-1}\int_0^t f(\theta_s X)\,ds \to E_\lambda f(X) \ \text{ a.s.}, \quad f \text{ bounded, measurable},$$

(ii) *when X is null-recurrent,*

$$t^{-1}\int_0^t 1_K(X_s)\,ds \to 0 \ \text{ a.s.}, \quad K \subset S \text{ compact},$$

(iii) *when X is transient,*

$$X_t \to \infty \ \text{ a.s.}$$

Proof: (i) From Lemma 11.11 and Theorems 26.10 (i) and 26.12, we note that P_λ is stationary and ergodic, and so the assertion holds a.s. P_λ by Corollary 25.9. Since the stated convergence is a tail event and $P_\mu = P_\lambda$ on $\mathcal{T}$ for any μ, the general result follows.

(ii) Since P_λ is shift-invariant with $P_\lambda\{X_s \in K\} = \lambda K < \infty$, the left-hand side converges a.e. P_λ by Theorem 26.2. By Theorems 26.10 and 26.12, the limit is a.e. a constant $c \geq 0$. Using Fatou's lemma and Fubini's theorem, we get

$$\begin{aligned} E_\lambda c &\leq \liminf_{t\to\infty} t^{-1}\int_0^t P_\lambda\{X_s \in K\}\,ds \\ &= \lambda K < \infty, \end{aligned}$$

which implies $c = 0$ since $\|P_\lambda\| = \|\lambda\| = \infty$. The general result follows from the fact that $P_\mu = P_\nu$ on $\mathcal{T}$ for any distributions μ and ν.

(iii) Fix any compact set $K \subset S$ with $\rho K > 0$, and conclude from the Markov property at $t > 0$ that a.s. P_μ,

$$\begin{aligned} U1_K(X_t) &= E_{X_t}\int_0^\infty 1_K(X_r)\,dr \\ &= E_\mu^{\mathcal{F}_t}\int_t^\infty 1_K(X_r)\,dr. \end{aligned}$$

Using the tower property of conditional expectations, we get for any $s < t$

$$\begin{aligned} E_\mu\Big\{U1_K(X_t)\,\Big|\,\mathcal{F}_s\Big\} &= E_\mu^{\mathcal{F}_s}\int_t^\infty 1_K(X_r)\,dr \\ &\leq E_\mu^{\mathcal{F}_s}\int_s^\infty 1_K(X_r)\,dr \\ &= U1_K(X_s), \end{aligned}$$

which shows that $U1_K(X_t)$ is a super-martingale. Since it is also non-negative and right-continuous, it converges a.s. P_μ as $t \to \infty$, and the limit equals 0 a.s., since $U1_K(X_t) \xrightarrow{P_\mu} 0$ by the previous proof. Since $U1_K$ is strictly positive and continuous, it follows that $X_t \to \infty$ a.s. P_μ. $\Box$

Exercises

1. For a measure space $(S, \mathcal{S}, \mu)$, let T be a positive, linear operator on $L^1 \cap L^\infty$. Show that if T is both an L^1-contraction and an L^∞-contraction, it is also an L^p-contraction for every $p \in [1, \infty]$. (*Hint:* Prove a Hölder-type inequality for T.)

2. Extend Lemma 25.3 to any transition operators T on a measurable space $(S, \mathcal{S})$. Thus, letting $\mathcal{I}$ be the class of sets $B \in \mathcal{S}$ with $T1_B = 1_B$, show that an $\mathcal{S}$-measurable function $f \geq 0$ is T-invariant iff it is $\mathcal{I}$-measurable.

3. Prove a continuous-time version of Theorem 26.2, for measurable semi-groups of positive $L^1 - L^\infty$-contractions. (*Hint:* Interpolate in the discrete-time result.)

4. Let (T_t) be a measurable, discrete- or continuous-time semi-group of positive $L^1 - L^\infty$-contractions on $(S, \mathcal{S}, \nu)$, let $\mu_1, \mu_2, \ldots$ be asymptotically invariant distributions on Z_+ or R_+, and define $A_n = \int T_t \, \mu_n(dt)$. Show that $A_n f \xrightarrow{\nu} Af$ for any $f \in L^1(\lambda)$, where $\xrightarrow{\nu}$ denotes convergence in measure. (*Hint:* Proceed as in Theorem 26.2, using the contractivity together with Minkowski's and Chebyshev's inequalities to estimate the remainder terms.)

5. Prove a continuous-time version of Theorem 26.4. (*Hint:* Use Lemma 26.5 to interpolate in the discrete-time result.)

6. Derive Theorem 25.6 from Theorem 26.4. (*Hint:* Take $g \equiv 1$, and proceed as in Corollary 25.9 to identify the limit.)

7. Show that when $f \geq 0$, the limit in Theorem 26.4 is strictly positive on the set $\{S_\infty f \wedge S_\infty g > 0\}$.

8. Show that the limit in Theorem 26.4 is invariant, at least when T is induced by a measure-preserving map on S.

9. Derive Lemma 26.3 (i) from Lemma 26.6. (*Hint:* Note that if $g \in \mathcal{T}(f)$ with $f \in L^1_+$, then $\mu g \leq \mu f$. Conclude that for any $h \in L^1$, $\mu\{h;\, M_n h > 0\} \geq \mu\{U^{n-1}h;\, M_n h > 0\} \geq 0$.)

10. Show that Brownian motion X in R^d is regular and strongly ergodic for every $d \in \mathsf{N}$, with an invariant measure that is unique up to a constant factor. Also show that X is Harris recurrent for $d \leq 2$, uniformly transient for $d \geq 3$.

11. Let X be a Markov process with associated space-time process $\tilde{X}$. Show that X is strongly ergodic in the sense of Theorem 26.10 iff $\tilde{X}$ is weakly ergodic in the sense of Theorem 26.11. (*Hint:* Note that a function is space-time invariant for X iff it is invariant for $\tilde{X}$.)

12. For a Harris recurrent process on R_+ or Z_+, every tail event is clearly a.s. invariant. Show by an example that the statement may fail in the transient case.

13. State and prove discrete-time versions of Theorems 26.12, 26.17, and 26.18. (*Hint:* The continuous-time arguments apply with obvious changes.)

14. Derive discrete-time versions of Theorems 26.17 and 26.18 from the corresponding continuous-time results.

15. Give an example of a regular Markov process that is weakly but not strongly ergodic. (*Hint:* For any strongly ergodic process, the associated space-time process has the stated property. For a less trivial example, consider a suitable super-critical branching process.)

16. Give examples of non-regular Markov processes with no invariant measure, with exactly one (up to a normalization), and with more than one.

17. Show that a discrete-time Markov process X and the corresponding pseudo-Poisson process Y have the same invariant measures. Further show that if X is regular then so is Y, but not conversely.

Chapter 27

Symmetric Distributions and Predictable Maps

Asymptotically invariant sampling, exchangeable and contractable, mixed and conditionally i.i.d. sequences, coding representation and equivalence criteria, urn sequences and factorial measures, strong stationarity, prediction sequence, predictable sampling, optional skipping, sojourns and maxima, exchangeable and mixed Lévy processes, compensated jumps, sampling processes, hyper-contraction and tightness, rotatable sequences and processes, predictable mapping

Classical results in probability typically require both an independence and an invariance condition[1]. It is then quite remarkable that, under special circumstances, the independence follows from the invariance assumption alone. We have already seen some notable instances of this phenomenon, such as in Theorems 11.5 and 11.14 for Markov processes and in Theorems 10.27 and 19.8 for martingales. Here and in the next chapter, we will study some broad classes of random phenomena, where a symmetry or invariance condition alone yields some very precise structural information, typically involving properties of independence.

For a striking example, recall that an infinite sequence of random variables $\xi = (\xi_1, \xi_2, \ldots)$ is said to be *stationary* if $\theta_n \xi \overset{d}{=} \xi$ for every $n \in \mathsf{Z}_+$, where θ_n denotes the n-step shift operator on R^∞. Replacing n by an arbitrary optional time τ in Z_+ yields the notion of *strong stationarity*, where $\theta_\tau \xi \overset{d}{=} \xi$. We will show that ξ is strongly stationary iff it is conditionally i.i.d., given a suitable σ-field $\mathcal{I}$. This statement is essentially equivalent to *de Finetti's theorem*—the fact that ξ is *exchangeable*, defined as invariance in distribution under finite permutations, iff it is mixed i.i.d. Thus, apart from a randomization, a simple symmetry condition leads us back to the elementary setting of i.i.d. sequences.

This is by modern standards a quite elementary result that can be proved in just a few lines. The same argument yields the remarkable extension by Ryll-Nardzewski—the fact that the mere *contractability*[2], where all sub-sequences are assumed to have the same distribution, suffices for ξ to be mixed i.i.d. Further highlights of the theory include the *predictable sampling theorem*, extending of the exchangeablity property of a finite or infinite sequence to arbitrary

[1] We may think of i.i.d., such as in the CLT, LLN, and LIL.

[2] short for contraction invariance in distribution

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_28

predictable permutations. In Chapters 14, 16, and 22, we saw how the latter result yields short proofs of the classical arcsine laws.

All the mentioned results have continuous-time counterparts. In particular, a process on R_+ with exchangeable or contractable increments is a mixture of Lévy processes. The more subtle representation of exchangeable processes on $[0,1]$ will be derived from some asymptotic properties for sequences obtained by sampling without replacement from a finite population. Similarly, the predictable sampling theorem turns into the quite subtle *predictable mapping theorem*—the invariance in distribution of an exchangeable process on R_+ or $[0,1]$ under *predictable* and pathwise measure-preserving transformations. Here some predictable mapping properties from Chapter 19, involving ordinary or discounted compensators, will be helpful for the proof.

The material in this chapter is related in many ways to other parts of the book. In particular, we note some links to various applications and extensions in Chapters 14–16, for results involving exchangeable sequences or processes. The predictable sampling theorem is further related to results for random time change in Chapters 15–16 and 19.

To motivate our first main topic, we consider a simple limit theorem for multi-variate sampling from a stationary process. Here we consider a measurable process X on an index set T, taking values in a space S, and let $\tau = (\tau_1, \tau_2, \dots)$ be an independent sequence of random elements in T with joint distribution μ. Define the associated *sampling sequence* $\xi = X \circ \tau$ in S^∞ by

$$\xi = (\xi_1, \xi_2, \dots) = (X_{\tau_1}, X_{\tau_2}, \dots)$$

henceforth referred to as a μ*-sample* from X. The sampling distributions $\mu_1, \mu_2, \dots$ on T^∞ are said to be *asymptotically invariant*, if their projections onto T^k are asymptotically invariant for every $k \in \mathsf{N}$. Recall that $\mathcal{I}_X$ denotes the invariant σ-field of X, and note that the conditional distribution $\eta = \mathcal{L}(X_0 \,|\, \mathcal{I}_X)$ exists by Theorem 8.5 when S is Borel.

Lemma 27.1 *(asymptotically invariant sampling) Let X be a stationary, measurable process on $T = \mathsf{R}$ or Z with values in a Polish space S. For $n \in \mathsf{N}$ let ξ_n be a μ_n-sample from X, where $\mu_1, \mu_2, \dots$ are asymptotically invariant distributions on T^∞. Then*

$$\xi_n \xrightarrow{d} \xi \ \text{ in } \ S^\infty, \qquad \mathcal{L}(\xi) = E\eta^\infty \ \text{ with } \ \eta = \mathcal{L}(X_0 | \mathcal{I}_X).$$

Proof: Write $\xi = (\xi^k)$ and $\xi_n = (\xi_n^k)$. Fix any asymptotically invariant distributions $\nu_1, \nu_2, \dots$ on T, and let $f_1, \dots, f_m$ be measurable functions on S bounded by ± 1. Proceeding as in the proof of Corollary 25.18 (i), we get

$$\begin{aligned}
&\Big| E \textstyle\prod_k f_k(\xi_n^k) - E \prod_k f_k(\xi^k) \Big| \\
&\quad \le E \Big| \mu_n \textstyle\bigotimes_k f_k(X) - \prod_k \eta f_k \Big| \\
&\quad \le \Big\| \mu_n - \mu_n * \nu_r^m \Big\| + \displaystyle\int E \Big| (\nu_r^m * \delta_t) \textstyle\bigotimes_k f_k(X) - \prod_k \eta f_k \Big| \, \mu_n(dt) \\
&\quad \le \displaystyle\int \Big\| \mu_n - \mu_n * \delta_t \Big\| \, \nu_r^m(dt) + \textstyle\sum_k \sup_t E \Big| (\nu_r * \delta_t) f_k(X) - \eta f_k \Big|.
\end{aligned}$$

Using the asymptotic invariance of μ_n and ν_r, together with Corollary 25.18 (i) and dominated convergence, we see that the right-hand side tends to 0 as $n \to \infty$ and then $r \to \infty$. The assertion now follows by Theorem 5.30. □

The last result leads immediately to a version of *de Finetti's theorem*. For a precise statement, consider a finite or infinite random sequence $\xi = (\xi_1, \xi_2, \ldots)$ with index set I, and say that ξ is *exchangeable* if

$$(\xi_{k_1}, \xi_{k_2}, \ldots) \stackrel{d}{=} (\xi_1, \xi_2, \ldots), \tag{1}$$

for any finite[3] permutation $(k_1, k_2, \ldots)$ of I. For infinite sequences ξ, we also consider the seemingly weaker property of *contractability*[4], where $I = \mathbb{N}$ and (1) is only required for increasing sequences $k_1 < k_2 < \cdots$. Then ξ is stationary, and every sub-sequence has the same distribution as ξ. By Lemma 27.1 we obtain $\mathcal{L}(\xi) = E\eta^\infty$ with $\eta = \mathcal{L}(\xi_1 | \mathcal{I}_\xi)$, here called the *directing random measure* of ξ. We may also give a stronger conditional statement. Recall that, for any random measure η on a measurable space $(S, \mathcal{S})$, the associated σ-field is generated by the random variables ηB for arbitrary $B \in \mathcal{S}$.

Theorem 27.2 *(exchangeable sequences, de Finetti, Ryll-Nardzewski) For an infinite random sequence ξ in a Borel space S, these conditions are equivalent:*

(i) *ξ is exchangeable,*
(ii) *ξ is contractable,*
(iii) *ξ is mixed i.i.d.,*
(iv) *$\mathcal{L}(\xi \,|\, \eta) = \eta^\infty$ a.s. for a random distribution η on S.*

Here η is a.s. unique and equal to $\mathcal{L}(\xi_1 | \mathcal{I}_\xi)$.

Since η^∞ is a.s. the distribution of an i.i.d. sequence in S based on the measure η, (iv) means that ξ is *conditionally i.i.d. η*. Taking expectations of both sides in (iv), we obtain the seemingly weaker condition $\mathcal{L}(\xi) = E\eta^\infty$, stating that ξ is *mixed i.i.d.* The latter condition implies that ξ is exchangeable, and so the two versions of (iii) are indeed equivalent.

First proof of Theorem 27.2 (OK): Since S is Borel, we may take $S = [0,1]$. Write μ_n for the uniform distribution on the product set $\prod_k \{(k-1)n + 1, \ldots, kn\}$. Assuming (ii), we see from Lemma 27.1 that $\mathcal{L}(\xi) = E\eta^\infty$. More generally, we may extend (1) to

$$\left\{1_A(\xi), \xi_{k_1}, \xi_{k_2}, \ldots\right\} \stackrel{d}{=} \left\{1_A(\xi), \xi_1, \xi_2, \ldots\right\}, \quad k_1 < k_2 < \cdots,$$

for any invariant set $A \in \mathcal{S}^\infty$. Applying Lemma 27.1 to the sequence of pairs $\{\xi_k, 1_A(\xi)\}$, we get as before $\mathcal{L}(\xi; \xi \in A) = E(\eta^\infty; \xi \in A)$, and since η is $\mathcal{I}_\xi$-measurable, it follows that $\mathcal{L}(\xi \,|\, \eta) = \eta^\infty$ a.s.

[3] one affecting only finitely many elements

[4] short for contraction invariance in distribution, also called *sub-sequence invariance*, *spreading invariance*, or *spreadability*

To see that η is a.s. unique, we may use the law of large numbers and Theorem 8.5 to obtain

$$n^{-1}\sum_{k\le n} 1_B(\xi_k) \to \eta B \ \text{ a.s.}, \quad B \in \mathcal{S}. \qquad \square$$

We can also prove de Finetti's theorem by a simple martingale argument:

Second proof of Theorem 27.2 (Aldous): If ξ is contractable, then

$$\begin{aligned}(\xi_m, \theta_m\xi) &\overset{d}{=} (\xi_k, \theta_m\xi)\\ &\overset{d}{=} (\xi_k, \theta_n\xi), \quad k \le m \le n.\end{aligned}$$

Now write $\mathcal{T}_\xi = \bigcap_n \sigma(\theta_n\xi)$. Using Lemma 8.10 followed by Theorem 9.24 as $n \to \infty$, we get a.s. for a fixed set $B \in \mathcal{S}$

$$\begin{aligned}\mathcal{L}(\xi_m \,|\, \theta_m\xi) &= \mathcal{L}(\xi_k \,|\, \theta_m\xi)\\ &= \mathcal{L}(\xi_k \,|\, \theta_n\xi)\\ &\to \mathcal{L}(\xi_k \,|\, \mathcal{T}_\xi),\end{aligned}$$

which shows that the extreme members agree a.s. In particular,

$$\begin{aligned}\mathcal{L}(\xi_m \,|\, \theta_m\xi) &= \mathcal{L}(\xi_m \,|\, \mathcal{T}_\xi)\\ &= \mathcal{L}(\xi_1 | \, \mathcal{T}_\xi) \ \text{ a.s.}\end{aligned}$$

Here the first relation gives $\xi_m \perp\!\!\!\perp_{\mathcal{T}_\xi} \theta_m\xi$ for all $m \in \mathsf{N}$, and so by iteration $\xi_1, \xi_2, \ldots$ are conditionally independent given $\mathcal{T}_\xi$. The second relation shows that the conditional distributions agree a.s. This proves that $\mathcal{L}(\xi \,|\, \mathcal{T}_\xi) = \nu^\infty$ with $\nu = \mathcal{L}(\xi_1 | \, \mathcal{T}_\xi)$. $\square$

The nature of de Finetti's criterion is further clarified by the following functional representation, here included to prepare for the much more subtle higher-dimensional versions in Chapter 28.

Corollary 27.3 *(coding representation) Let $\xi = (\xi_j)$ be an infinite random sequence in a Borel space S. Then ξ is exchangeable iff*

$$\xi_j = f(\alpha, \zeta_j), \quad j \in \mathsf{N},$$

for a measurable function $f\colon [0,1]^2 \to S$ and some i.i.d. $U(0,1)$ random variables $\alpha, \zeta_1, \zeta_2, \ldots$. The directing random measure ν of ξ is then given by

$$\nu B = \int \Big\{1_B \circ f(\alpha, x)\Big\}\, dx, \quad B \in \mathcal{S}.$$

Proof: Let ξ be conditionally i.i.d. with directing random measure ν. For any i.i.d. $U(0,1)$ random variables $\tilde\alpha$ and $\tilde\zeta_1, \tilde\zeta_2, \ldots$, a repeated use of Theorem 8.17 yields some measurable functions g, h, such that

$$\begin{aligned}\tilde\nu &\equiv g(\tilde\alpha) \overset{d}{=} \nu,\\ (\tilde\xi_1, \tilde\nu) &\equiv \Big\{h(\tilde\nu, \tilde\zeta_1), \tilde\nu\Big\} \overset{d}{=} (\xi_1, \nu).\end{aligned} \qquad (2)$$

Writing $f(s,x) = h\{g(s),x\}$ gives

$$\begin{aligned}\tilde{\xi}_1 &= h(\tilde{\nu},\tilde{\zeta}_1)\\ &= f\Big\{g(\tilde{\alpha}),\tilde{\zeta}_1\Big\}\\ &= f(\tilde{\alpha},\tilde{\zeta}_1),\end{aligned}$$

and we may define more generally

$$\begin{aligned}\tilde{\xi}_j &= f(\tilde{\alpha},\tilde{\zeta}_j)\\ &= h(\tilde{\nu},\tilde{\zeta}_j), \quad j \in \mathbb{N}.\end{aligned}$$

The last expression shows that the $\tilde{\xi}_j$ are conditionally independent given $\tilde{\nu}$, and by (2) their common conditional distribution equals $\tilde{\nu}$. Hence, the sequence $\tilde{\xi} = (\tilde{\xi}_j)$ satisfies $(\tilde{\nu},\tilde{\xi}) \stackrel{d}{=} (\nu,\xi)$, and Corollary 8.18 yields a.s.

$$\nu = g(\alpha), \qquad \xi_j = h(\nu,\zeta_j), \quad j \in \mathbb{N}, \tag{3}$$

for some i.i.d. $U(0,1)$ variables α and $\zeta_1,\zeta_2,\ldots$.

To identify ν, let $B \in \mathcal{S}$ be arbitrary, and use (3), Fubini's theorem, and the definition of ν to get

$$\begin{aligned}\nu B &= P\Big\{\xi_j \in B \,\Big|\, \nu\Big\}\\ &= P\Big\{h(\nu,\zeta_j) \in B \,\Big|\, \nu\Big\}\\ &= \int\Big\{1_B \circ h(\nu,x)\Big\}\,dx\\ &= \int\Big\{1_B \circ f(\alpha,x)\Big\}\,dx.\end{aligned}$$ □

The representing function f in Corollary 27.3 is far from unique. Here we state some basic equivalence criteria, motivating the two-dimensional versions in Theorem 28.3. The present proof is the same, apart from some obvious simplifications. To simplify the writing, put $I = [0,1]$.

Proposition 27.4 *(equivalent coding functions, Hoover, OK) Fix any measurable functions $f, g\colon I^2 \to S$, and let $\alpha, \xi_1, \xi_2, \ldots$ and $\alpha', \xi'_1, \xi'_2, \ldots$ be i.i.d. $U(0,1)$. Then these conditions are equivalent*[5]*:*

(i) *we may choose $(\tilde{\alpha},\tilde{\xi}_1,\tilde{\xi}_2,\ldots) \stackrel{d}{=} (\alpha,\xi_1,\xi_2,\ldots)$ with*
$$f(\alpha,\xi_i) = g(\tilde{\alpha},\tilde{\xi}_i) \ \text{a.s.}, \quad i \ge 1,$$

(ii) $\Big\{f(\alpha,\xi_i);\, i \ge 1\Big\} \stackrel{d}{=} \Big\{g(\alpha,\xi_i);\, i \ge 1\Big\}$,

(iii) *there exist some measurable functions $T, T'\colon I \to I$ and $U, U'\colon I^2 \to I$, preserving λ in the highest order arguments, such that*
$$f\Big\{T(\alpha),\, U(\alpha,\xi_i)\Big\} = g\Big\{T'(\alpha),\, U'(\alpha,\xi_i)\Big\} \ \text{a.s.}, \ i \ge 1,$$

[5]This result is often misunderstood: it is not enough in (iii) to consider λ-preserving transformations U, U' of a single variable, nor is (iv) true without the extra randomization variables α', ξ'_i.

(iv) *there exist some measurable functions* $T : I^2 \to I$ *and* $U : I^4 \to I$, *mapping* λ^2 *into* λ *in the highest order arguments, such that*

$$f(\alpha, \xi_i) = g\Big\{T(\alpha, \alpha'),\, U(\alpha, \alpha', \xi_i, \xi_i')\Big\} \text{ a.s.}, \; i \ge 1.$$

Theorem 27.2 fails for finite sequences. Here we need instead to replace the inherent i.i.d. sequences by suitable *urn sequences*, generated by successive drawing without replacement from a finite set. To make this precise, fix a measurable space S, and consider a measure of the form $\mu = \sum_{k \le n} \delta_{s_k}$ with $s_1, \dots, s_n \in S$. As in Lemma 2.20, we introduce on S^n the associated *factorial measure*

$$\mu^{(n)} = \sum_{p \in \mathcal{P}_n} \delta_{s \circ p}\,,$$

where $\mathcal{P}_n$ is the set of permutations $p = (p_1, \dots, p_n)$ of $1, \dots, n$, and $s \circ p = (s_{p_1}, \dots, s_{p_n})$. Note that $\mu^{(n)}$ is independent of the order of $s_1, \dots, s_n$ and depends measurably on μ.

Lemma 27.5 *(finite exchangeable sequences) Let* $\xi = (\xi_1, \dots, \xi_n)$ *be a finite random sequence in* S, *and put* $\eta = \sum_k \delta_{\xi_k}$. *Then* ξ *is exchangeable iff*

$$\mathcal{L}(\xi \,|\, \eta) = \frac{\eta^{(n)}}{n!} \text{ a.s.}$$

Proof: Since η is invariant under permutations of $\xi_1, \dots, \xi_n$, we have $(\xi \circ p, \eta) \overset{d}{=} (\xi, \eta)$ for any $p \in \mathcal{P}_n$. Now introduce an exchangeable permutation $\pi \perp\!\!\!\perp \xi$ of $1, \dots, n$. Using Fubini's theorem twice, we get for any measurable sets A and B in appropriate spaces

$$\begin{aligned} P\Big\{\xi \in B,\, \eta \in A\Big\} &= P\Big\{\xi \circ \pi \in B,\, \eta \in A\Big\} \\ &= E\Big\{P\Big(\xi \circ \pi \in B \,\Big|\, \xi\Big);\, \eta \in A\Big\} \\ &= E\Big(\eta^{(n)} B / n!\,;\, \eta \in A\Big). \end{aligned}$$

□

Just as for martingales and Markov processes, we may relate the defining properties to a filtration $\mathcal{F} = (\mathcal{F}_n)$. Thus, a finite or infinite sequence of random elements $\xi = (\xi_1, \xi_2, \dots)$ is said to be $\mathcal{F}$-*exchangeable*, if it is $\mathcal{F}$-adapted and such that, for every $n \ge 0$, the shifted sequence $\theta_n \xi = (\xi_{n+1}, \xi_{n+2}, \dots)$ is conditionally exchangeable given $\mathcal{F}_n$. For infinite sequences ξ, the definition of $\mathcal{F}$-*contractability* is similar.[6] When $\mathcal{F}$ is the filtration induced by ξ, the stated properties reduce to the unqualified versions considered earlier.

An infinite, adapted sequence ξ is said to be *strongly stationary* or $\mathcal{F}$-*stationary* if $\theta_\tau \xi \overset{d}{=} \xi$ for every optional time $\tau < \infty$. By the *prediction sequence* of ξ we mean the set of conditional distributions

$$\pi_n = \mathcal{L}(\,\theta_n \xi \,|\, \mathcal{F}_n), \quad n \in \mathbb{Z}_+. \tag{4}$$

[6]Since both definitions can be stated without reference to regular conditional distributions, no restrictions are needed on S.

The random probability measures $\pi_0, \pi_1, \ldots$ on S are said to form a *measure-valued martingale*, if $(\pi_n B)$ is a real-valued martingale for every measurable set $B \subset S$. We show that strong stationarity is equivalent to exchangeability,[7] and exhibit an interesting martingale connection.

Corollary 27.6 *(strong stationarity) Let ξ be an infinite, $\mathcal{F}$-adapted random sequence with prediction sequence π, taking values in a Borel space S. Then these conditions are equivalent:*

(i) *ξ is $\mathcal{F}$-exchangeable,*

(ii) *ξ is $\mathcal{F}$-contractable,*

(iii) *ξ is $\mathcal{F}$-stationary,*

(iv) *π is a measure-valued $\mathcal{F}$-martingale.*

Proof: Conditions (i) and (ii) are equivalent by Theorem 27.2. Assuming (ii), we get a.s. for any $B \in \mathcal{S}^\infty$ and $n \in \mathsf{Z}_+$

$$\begin{aligned} E\Big(\pi_{n+1}B \,\Big|\, \mathcal{F}_n\Big) &= P\Big\{\theta_{n+1}\xi \in B \,\Big|\, \mathcal{F}_n\Big\} \\ &= P\Big\{\theta_n\xi \in B \,\Big|\, \mathcal{F}_n\Big\} = \pi_n B, \end{aligned} \tag{5}$$

which proves (iv). Conversely, (ii) follows by iteration from the second equality in (5), and so (ii) and (iv) are equivalent.

Next we note that (4) extends by Lemma 8.3 to

$$\pi_\tau B = P\Big\{\theta_\tau\xi \in B \,\Big|\, \mathcal{F}_\tau\Big\} \quad \text{a.s.}, \qquad B \in \mathcal{S}^\infty,$$

for any finite optional time τ. By Lemma 9.14, condition (iv) is then equivalent to

$$\begin{aligned} P\Big\{\theta_\tau\xi \in B\Big\} &= E\pi_\tau B = E\pi_0 B \\ &= P\{\xi \in B\}, \qquad B \in \mathcal{S}^\infty, \end{aligned}$$

which is in turn equivalent to (iii). □

The exchangeability property extends to a broad class of *random* permutations. Say that an integer-valued random variable τ is *predictable* with respect to a given filtration $\mathcal{F}$, if the shifted time $(\tau-1)_+$ is $\mathcal{F}$-optional.

Theorem 27.7 *(predictable sampling) Let $\xi = (\xi_1, \xi_2, \ldots)$ be a finite or infinite, $\mathcal{F}$-exchangeable random sequence indexed by I. Then for any a.s. distinct, $\mathcal{F}$-predictable times $\tau_1, \ldots, \tau_n$ in I, we have*

$$(\xi_{\tau_1}, \ldots, \xi_{\tau_n}) \overset{d}{=} (\xi_1, \ldots, \xi_n). \tag{6}$$

[7] This remains true for finite sequences $\xi = (\xi_1, \ldots, \xi_n)$, if we define $\theta_k\xi = (\xi_{k+1}, \ldots, \xi_n, \xi_k, \ldots, \xi_1)$.

Note that we do not require ξ to be infinite or the τ_k to be increasing. A simple special case is that of *optional skipping*, where $I = \mathsf{N}$ and $\tau_1 < \tau_2 < \cdots$. If $\tau_k \equiv \tau + k$ for some optional time $\tau < \infty$, then (6) reduces to the strong stationarity in Proposition 27.6. For both applications and proof, it is useful to consider the inverse *allocation sequence*

$$\alpha_j = \inf\Big\{k;\, \tau_k = j\Big\}, \quad j \in I.$$

When α_j is finite, it gives the position of j in the permuted sequence (τ_k). The times τ_k are clearly predictable iff the sequence (α_j) is predictable in the sense of Chapters 9–10.

Proof of Theorem 27.7: First let ξ be indexed by $I_n = \{1, \dots, n\}$, so that $(\tau_1, \dots, \tau_n)$ and $(\alpha_1, \dots, \alpha_n)$ are mutually inverse random permutations of I_n. For each $m \in \{0, \dots, n\}$, put $\alpha_j^m = \alpha_j$ for all $j \le m$, and define recursively

$$\alpha_{j+1}^m = \min\Big(I_n \setminus \{\alpha_1^m, \dots, \alpha_j^m\}\Big), \quad m \le j \le n.$$

Then $(\alpha_1^m, \dots, \alpha_n^m)$ is a predictable, $\mathcal{F}_{m-1}$-measurable permutation of I_n. Since also $\alpha_j^m = \alpha_j^{m-1} = \alpha_j$ whenever $j < m$, Theorem 8.5 yields for any measurable functions $f_1, \dots, f_n \ge 0$ on S

$$\begin{aligned}
E \textstyle\prod_j f_{\alpha_j^m}(\xi_j) &= E\, E\Big\{ \textstyle\prod_j f_{\alpha_j^m}(\xi_j) \,\Big|\, \mathcal{F}_{m-1}\Big\} \\
&= E \prod_{j<m} f_{\alpha_j^m}(\xi_j)\, E\Big\{ \prod_{j \ge m} f_{\alpha_j^m}(\xi_j) \,\Big|\, \mathcal{F}_{m-1}\Big\} \\
&= E \prod_{j<m} f_{\alpha_j^{m-1}}(\xi_j)\, E\Big\{ \prod_{j \ge m} f_{\alpha_j^{m-1}}(\xi_j) \,\Big|\, \mathcal{F}_{m-1}\Big\} \\
&= E \textstyle\prod_j f_{\alpha_j^{m-1}}(\xi_j).
\end{aligned}$$

Summing over $m \in \{1, \dots, n\}$ and noting that $\alpha_j^n = \alpha_j$ and $\alpha_j^0 = j$ for all j, we get

$$E \prod_{k \le n} f_k(\xi_{\tau_k}) = E \prod_{j \le n} f_{\alpha_j}(\xi_j) = E \prod_{k \le n} f_k(\xi_k),$$

which extends to (6) by a monotone-class argument.

Next let $I = I_m$ with $m > n$. The sequence (τ_k) may then be extended recursively to I_m, through the formula

$$\tau_{k+1} = \min\Big(I_m \setminus \{\tau_1, \dots, \tau_k\}\Big), \quad k \ge n, \tag{7}$$

so that $\tau_1, \dots, \tau_m$ form a random permutation of I_m. By induction based on (7), the times $\tau_{n+1}, \dots, \tau_m$ are again predictable. Hence, the previous case applies, and (6) follows.

Finally, let $I = \mathsf{N}$. For each $m \in \mathsf{N}$, consider the predictable times

$$\tau_k^m = \tau_k 1\{\tau_k \le m\} + (m + k) 1\{\tau_k > m\}, \quad k = 1, \dots, n,$$

and conclude from the previous version of (6) that

$$\left(\xi_{\tau_1^m}, \ldots, \xi_{\tau_n^m}\right) \stackrel{d}{=} (\xi_1, \ldots, \xi_n). \tag{8}$$

As $m \to \infty$ we get $\tau_k^m \to \tau_k$, and (6) follows from (8) by dominated convergence. □

The last result yields a simple proof of yet another striking property of random walks in $\mathbb{R}$, relating the first maximum to the number of positive values. This leads in turn to simple proofs of the arcsine laws in Theorems 14.16 and 22.11.

Corollary 27.8 *(sojourns and maxima, Sparre-Andersen) Let $\xi_1, \ldots, \xi_n$ be exchangeable random variables, and put $S_k = \xi_1 + \cdots + \xi_k$. Then*

$$\sum_{k \le n} 1\{S_k > 0\} \stackrel{d}{=} \min\Big\{ k \ge 0;\ S_k = \max_{j \le n} S_j \Big\}.$$

Proof (OK): Put $\tilde{\xi}_k = \xi_{n-k+1}$ for $k = 1, \ldots, n$, and note that the $\tilde{\xi}_k$ remain exchangeable for the filtration $\mathcal{F}_k = \sigma\{S_n, \tilde{\xi}_1, \ldots, \tilde{\xi}_k\}$, $k = 0, \ldots, n$. Write $\tilde{S}_k = \tilde{\xi}_1 + \cdots + \tilde{\xi}_k$, and introduce the predictable permutation

$$\alpha_k = \sum_{j \in 0}^{k-1} 1\{\tilde{S}_j < S_n\} + (n - k + 1) 1\{\tilde{S}_{k-1} \ge S_n\}, \quad k = 1, \ldots, n.$$

Define $\xi'_k = \sum_j \tilde{\xi}_j 1\{\alpha_j = k\}$ for $k = 1, \ldots, n$, and conclude from Theorem 27.7 that $(\xi'_k) \stackrel{d}{=} (\xi_k)$. Writing $S'_k = \xi'_1 + \cdots + \xi'_k$, we further note that

$$\begin{aligned} &\min\Big\{ k \ge 0;\ S'_k = \max\nolimits_j S'_j \Big\} \\ &\qquad = \sum_{j=0}^{n-1} 1\{\tilde{S}_j < S_n\} = \sum_{k=1}^{n} 1\{S_k > 0\}. \end{aligned}$$ □

Turning to the continuous-time case, we say that a process X in a topological space is *continuous in probability*[8] if $X_s \xrightarrow{P} X_t$ as $s \to t$. An $\mathbb{R}^d$-valued process X on $[0, 1]$ or $\mathbb{R}_+$ is said to be *exchangeable* or *contractable*, if it is continuous in probability with $X_0 = 0$, and such that the *increments* $X_t - X_s$ over disjoint intervals $(s, t]$ of equal length form exchangeable or contractable sequences. Finally, we say that X has *conditionally stationary, independent increments*, given a σ-field $\mathcal{I}$, if the stated property holds conditionally for any finite collection of intervals.

The following continuous-time version of Theorem 27.2 characterizes the exchangeable processes on $\mathbb{R}_+$; the harder finite-interval case is treated in Theorem 27.10. The point process case was already considered separately, by different methods, in Theorem 15.14.

Proposition 27.9 *(exchangeable processes on $\mathbb{R}_+$, Bühlmann) Let X be an $\mathbb{R}^d$-valued process on $\mathbb{R}_+$, continuous in probability with $X_0 = 0$. Then these conditions are equivalent:*

[8] also said to be *stochastically continuous*

(i) *X has exchangeable increments,*
(ii) *X has contractable increments,*
(iii) *X is a mixed Lévy process,*
(iv) *X is conditionally Lévy, given a σ-field $\mathcal{I}$.*

Proof: The sufficiency of (iii)−(iv) being obvious, it suffices to show that the conditions are also necessary. Thus, let X be contractable. The increments ξ_{nk} over the dyadic intervals $I_{nk} = 2^{-n}(k-1,k]$ are then contractable for fixed n, and so by Theorem 27.2 they are conditionally i.i.d. η_n for some random probability measure η_n on R. Using Corollary 4.12 and the uniqueness in Theorem 27.2, we obtain

$$\eta_n^{*2^{n-m}} = \eta_m \ \text{ a.s.}, \quad m < n. \tag{9}$$

Thus, for any $m < n$, the increments ξ_{mk} are conditionally i.i.d. η_m, given η_n. Since the σ-fields $\sigma(\eta_n)$ are a.s. non-decreasing by (9), Theorem 9.24 shows that the ξ_{mk} remain conditionally i.i.d. η_m, given $\mathcal{I} \equiv \sigma\{\eta_0, \eta_1, \ldots\}$.

Now fix any disjoint intervals $I_1, \ldots, I_n$ of equal length with associated increments $\xi_1, \ldots, \xi_n$, and approximate by disjoint intervals $I_1^m, \ldots, I_n^m$ of equal length with dyadic endpoints. For each m, the associated increments ξ_k^m are conditionally i.i.d., given $\mathcal{I}$. Thus, for any bounded, continuous functions $f_1, \ldots, f_n$,

$$E^{\mathcal{I}} \prod_{k \le n} f_k(\xi_k^m) = \prod_{k \le n} E^{\mathcal{I}} f_k(\xi_k^m) = \prod_{k \le n} E^{\mathcal{I}} f_k(\xi_1^m). \tag{10}$$

Since X is continuous in probability, we have $\xi_k^m \xrightarrow{P} \xi_k$ for each k, and so (10) extends by dominated convergence to the original variables ξ_k. By suitable approximation and monotone-class arguments, we may finally extend the relations to any measurable indicator functions $f_k = 1_{B_k}$. □

We turn to the more difficult case of exchangeable processes on $[0,1]$. To avoid some technical complications[9], we will prove the following result only for $d = 1$.

Theorem 27.10 *(exchangeable processes on $[0,1]$) Let X be an R^d-valued process on $[0,1]$, continuous in probability with $X_0 = 0$. Then X is exchangeable iff it has a version*

$$X_t = \alpha\, t + \sigma B_t + \sum_{j \ge 1} \beta_j \Big(1\{\tau_j \le t\} - t \Big), \quad t \in [0,1], \tag{11}$$

for a Brownian bridge B, some independent i.i.d. $U(0,1)$ random variables $\tau_1, \tau_2, \ldots$, and an independent set[10] of coefficients α, σ, $\beta_1, \beta_2, \ldots$ with $\sum_j |\beta_j|^2 < \infty$ a.s. The sum in (11) converges a.s., uniformly on $[0,1]$, toward an rcll process on $[0,1]$.

---

[9] The complications for $d > 1$ are similar to those in Theorem 7.7.

[10] Thus, $B, \tau_1, \tau_2, \ldots, \{\alpha, \sigma, (\beta_j)\}$ are independent; no claim is made about the dependence between $\alpha, \sigma, \beta_1, \beta_2, \ldots$.

In particular, a simple point process on $[0,1]$ is symmetric with respect to Lebesgue measure λ iff it is a mixed binomial process based on λ, as noted already in Theorem 15.14. Combining the present result with Theorem 27.9, we see that a continuous process X on R_+ or $[0,1]$ with $X_0 = 0$ is exchangeable iff $X_t \equiv \alpha t + \sigma B_t$ a.s., where B is a Brownian motion or bridge, respectively, and (α, σ) is an independent pair of random variables.

The coefficients α, σ, $\beta_1, \beta_2, \ldots$ in (11) are clearly a.s. unique and X-measurable, apart from the order of the jump sizes β_j. Thus, X determines a.s. the pair (α, ν), where ν is the a.s. bounded random measure on R given by

$$\nu = \sigma^2 \delta_0 + \textstyle\sum_j \beta_j^2 \, \delta_{\beta_j}. \tag{12}$$

Since $\mathcal{L}(X)$ is also uniquely determined by $\mathcal{L}(\alpha, \nu)$, we say that a process X as in (11) is *directed by* (α, ν).

First we need to establish the convergence of the sum in (11).

Lemma 27.11 *(uniform convergence) Let Y_t^n be the n-th partial sum of the series in* (11). *Then these conditions are equivalent:*

(i) $\sum_j \beta_j^2 < \infty$ *a.s.,*

(ii) Y_t^n *converges a.s. for every* $t \in [0,1]$*,*

(iii) *there exists an rcll process* Y *on* $[0,1]$*, such that*

$$(Y^n - Y)^* \to 0 \text{ a.s.}$$

Proof: We may clearly take the variables β_j to be non-random. Further note that trivially (iii) $\Rightarrow$ (ii).

(i) $\Leftrightarrow$ (ii): Here we may take the β_j to be non-random. Then for fixed $t \in (0,1)$ the terms are independent and bounded with mean 0 and variance $\beta_j^2\, t(1-t)$, and so by Theorem 5.18 the series converges iff $\sum_j \beta_j^2 < \infty$.

(i) $\Rightarrow$ (iii): The processes $M_t^n = (1-t)^{-1} Y_t^n$ are L^2-martingales on $[0,1)$, with respect to the filtration induced by the processes $1\{\tau_j \le t\}$. By Doob's inequality, we have for any $m < n$ and $t \in [0,1)$

$$\begin{aligned} E(Y^n - Y^m)_t^{*2} &\le E(M^n - M^m)_t^{*2} \\ &\le 4\, E\Big(M_t^n - M_t^m\Big)^2 \\ &= 4\,(1-t)^{-2} E\Big(Y_t^n - Y_t^m\Big)^2 \\ &\le 4\, t(1-t)^{-1} \sum_{j>m} \beta_j^2, \end{aligned}$$

which tends to 0 as $m \to \infty$ for fixed t. By symmetry we have the same convergence for the time-reversed processes Y_{1-t}^n, and so by combination

$$(Y^m - Y^n)_1^* \xrightarrow{P} 0, \quad m, n \to \infty.$$

Statement (iii) now follows by Lemma 11.13. $\square$

For the main proof, we consider first a continuity theorem for exchangeable processes of the form (11). Once the general representation is established, this yields a continuity theorem for general exchangeable processes on $[0,1]$, which is clearly of independent interest. The processes X are regarded as random elements in the function space $D_{[0,1]}$ of rcll functions on $[0,1]$, endowed with an obvious modification of the Skorohod topology, discussed in Chapter 23 and Appendix 5. We further regard the a.s. bounded random measures ν as random elements in the measure space $\mathcal{M}_{\mathsf{R}}$ endowed with the weak topology, equivalent to the vague topology on $\mathcal{M}_{\bar{\mathsf{R}}}$.

Theorem 27.12 *(limits of exchangeable processes) Let $X_1, X_2, \ldots$ be exchangeable processes as in* (11) *directed by the pairs (α_n, ν_n). Then*

$$X^n \xrightarrow{sd} X \text{ in } D_{[0,1]} \quad \Leftrightarrow \quad (\alpha_n, \nu_n) \xrightarrow{wd} (\alpha, \nu) \text{ in } \mathsf{R} \times \mathcal{M}_{\mathsf{R}},$$

where X is exchangeable and directed by (α, ν).

Here and below, we need an elementary but powerful tightness criterion.

Lemma 27.13 *(hyper-contraction and tightness) Let the random variables $\xi_1, \xi_2, \ldots \geq 0$ and σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be such that*

$$E\Big(\xi_n^2 \,\Big|\, \mathcal{F}_n\Big) \leq c\Big\{E(\xi_n |\, \mathcal{F}_n)\Big\}^2 < \infty \quad a.s., \qquad n \in \mathsf{N},$$

for a constant $c > 0$, and put $\eta_n = E(\xi_n | \mathcal{F}_n)$. Then

$$(\xi_n) \text{ is tight} \quad \Rightarrow \quad (\eta_n) \text{ is tight.}$$

Proof: By Lemma 5.9, we need to show that $r_n \eta_n \xrightarrow{P} 0$ whenever $0 \leq r_n \to 0$. Then conclude from Lemma 5.1 that, for any $p \in (0,1)$ and $\varepsilon > 0$,

$$\begin{aligned} 0 &< (1-p)^2\, c^{-1} \\ &\leq P\Big\{\xi_n \geq p\,\eta_n \,\Big|\, \mathcal{F}_n\Big\} \\ &\leq P\Big\{r_n\,\xi_n \geq p\,\varepsilon \,\Big|\, \mathcal{F}_n\Big\} + 1\{r_n\,\eta_n < \varepsilon\}. \end{aligned}$$

Here the first term on the right tends in probability to 0, since $r_n \xi_n \xrightarrow{P} 0$ by Lemma 5.9. Hence, $1\{r_n \eta_n < \varepsilon\} \xrightarrow{P} 1$, which means that $P\{r_n \eta_n \geq \varepsilon\} \to 0$. Since ε is arbitrary, we get $r_n \eta_n \xrightarrow{P} 0$. □

Proof of Theorem 27.12: First let $(\alpha_n, \nu_n) \xrightarrow{wd} (\alpha, \nu)$. To show that $X^n \xrightarrow{sd} X$ for the corresponding processes in (11), it suffices by Lemma 5.28 to take all α_n and ν_n to be non-random. Thus, we may restrict our attention to processes X^n with constant coefficients α_n, σ_n, and β_{nj}, $j \in \mathsf{N}$.

To prove that $X^n \xrightarrow{fd} X$, we begin with four special cases. First we note that, if $\alpha_n \to \alpha$, then trivially $\alpha_n t \to \alpha t$ uniformly on $[0,1]$. Similarly, $\sigma_n \to \sigma$ implies $\sigma_n B \to \sigma B$ in the same sense. Next suppose that $\alpha_n = \sigma_n = 0$ and $\beta_{n,m+1} = \beta_{n,m+2} = \cdots = 0$ for a fixed $m \in \mathsf{N}$. We may then assume that even

$\alpha = \sigma = 0$ and $\beta_{m+1} = \beta_{m+2} = \cdots = 0$, and that $\beta_{nj} \to \beta_j$ for all j. Here the convergence $X^n \to X$ is obvious. We finally assume that $\alpha_n = \sigma_n = 0$ and $\alpha = \beta_1 = \beta_2 = \cdots = 0$. Then $\max_j |\beta_{nj}| \to 0$, and for any $s \le t$,

$$\begin{aligned} E(X_s^n X_t^n) &= s\,(1-t) \textstyle\sum_k \beta_{nk}^2 \\ &\to s\,(1-t)\,\sigma^2 \\ &= E(X_s X_t). \end{aligned} \tag{13}$$

In this case, $X^n \xrightarrow{fd} X$ by Theorem 6.12 and Corollary 6.5. By independence, we may combine the four special cases into $X^n \xrightarrow{fd} X$, whenever $\beta_j = 0$ for all but finitely many j. To extend this to the general case, we may use Theorem 5.29, where the required uniform error estimate is obtained as in (13).

To strengthen the convergence to $X^n \xrightarrow{sd} X$ in $D_{[0,1]}$, it is enough to verify the tightness criterion in Theorem 23.11. Thus, for any X^n-optional times τ_n and positive constants $h_n \to 0$ with $\tau_n + h_n \le 1$, we need to show that $X^n_{\tau_n + h_n} - X^n_{\tau_n} \xrightarrow{P} 0$. By Theorem 27.7 and a simple approximation, it is equivalent that $X^n_{h_n} \xrightarrow{P} 0$, which is clear since

$$E(X^n_{h_n})^2 = h_n^2\,\alpha_n^2 + h_n\,(1-h_n)\,\|\nu_n\| \to 0.$$

To obtain the reverse implication, let $X^n \xrightarrow{sd} X$ in $D_{[0,1]}$ for some process X. Since $\alpha_n = X_1^n \xrightarrow{d} X_1$, the sequence (α_n) is tight. Next define for $n \in \mathbb{N}$

$$\begin{aligned} \eta_n &= 2\,X^n_{1/2} - X^n_1 \\ &= 2\,\sigma_n B_{1/2} + 2 \textstyle\sum_j \beta_{nj} \Big(1\{\tau_j \le \frac{1}{2}\} - \frac{1}{2}\Big). \end{aligned}$$

Then

$$\begin{aligned} E(\eta_n^2 \,|\, \nu_n) &= \sigma_n^2 + \textstyle\sum_j \beta_{nj}^2 \\ &= \|\nu_n\|, \\ E(\eta_n^4 \,|\, \nu_n) &= 3\,\Big(\sigma_n^2 + \textstyle\sum_j \beta_{nj}^2\Big)^2 - 2 \textstyle\sum_j \beta_{nj}^4 \\ &\le 3\,\|\nu_n\|^2. \end{aligned}$$

Since (η_n) is tight, even (ν_n) is tight by Lemmas 23.15 and 27.13, and so the same thing is true for the sequence of pairs (α_n, ν_n).

The tightness implies relative compactness in distribution, and so every sub-sequence contains a further sub-sequence converging in $\mathbb{R} \times \mathcal{M}_{\bar{\mathbb{R}}}$ toward some random pair (α, ν). Since the measures in (12) form a vaguely closed subset of $\mathcal{M}_{\bar{\mathbb{R}}}$, the limit ν has the same form for suitable σ and $\beta_1, \beta_2, \ldots$. Then the direct assertion yields $X^n \xrightarrow{sd} Y$ with Y as in (11), and therefore $X \overset{d}{=} Y$. Since the coefficients in (11) are measurable functions of Y, the distribution of (α, ν) is uniquely determined by that of X. Thus, the limiting distribution is independent of sub-sequence, and the convergence $(\alpha_n, \nu_n) \xrightarrow{wd} (\alpha, \nu)$ remains valid along $\mathbb{N}$. Finally, we may use Corollary 8.18 to transfer the representation (22) to the original process X. □

Next we prove a functional limit theorem for partial sums of exchangeable random variables, extending the classical limit theorems for random walks in Theorems 22.9 and 23.14. Then for every $n \in \mathsf{N}$ we consider some exchangeable random variables ξ_{nj}, $j \le m_n$, and form the summation processes

$$X_t^n = \sum_{j \le m_n t} \xi_{nj}, \quad t \in [0,1], \ n \in \mathsf{N}. \tag{14}$$

Assuming $m_n \to \infty$, we show that the X^n can be approximated by exchangeable processes as in (11). The convergence criteria may be stated in terms of the random variables and measures

$$\alpha_n = \sum\nolimits_j \xi_{nj}, \quad \nu_n = \sum\nolimits_j \xi_{nj}^2 \, \delta_{\xi_{nj}}, \qquad n \in \mathsf{N}. \tag{15}$$

When the α_n and ν_n are non-random, this reduces to a functional limit theorem for sampling without replacement from a finite population, which is again of independent interest.

Theorem 27.14 *(limits of sampling processes, Hagberg, OK) For $n \in \mathsf{N}$, let ξ_{nj}, $j \le m_n \to \infty$, be exchangeable random variables, and define the processes X^n with associated pairs (α_n, ν_n) by (14) and (15). Then*

$$X^n \xrightarrow{sd} X \ \textit{in} \ D_{[0,1]} \quad \Leftrightarrow \quad (\alpha_n, \nu_n) \xrightarrow{wd} (\alpha, \nu) \ \textit{in} \ \mathsf{R} \times \mathcal{M}_{\mathsf{R}},$$

where X is exchangeable on $[0,1]$ and directed by (α, ν).

Proof: Let $\tau_1, \tau_2, \ldots$ be i.i.d. $U(0,1)$ and independent of all ξ_{nj}, and define

$$\begin{aligned} Y_t^n &= \sum\nolimits_j \xi_{nj} 1\{\tau_j \le t\} \\ &= \alpha_n t + \sum\nolimits_j \xi_{nj} \Big(1\{\tau_j \le t\} - t \Big), \quad t \in [0,1]. \end{aligned}$$

Writing $\tilde{\xi}_{nk}$ for the k-th jump from the left of Y^n (including possible zero jumps when $\xi_{nj} = 0$), we note that $(\tilde{\xi}_{nj}) \overset{d}{=} (\xi_{nj})$ by exchangeability. Thus, $\tilde{X}^n \overset{d}{=} X^n$, where $\tilde{X}_t^n$ is defined as in (14). Furthermore, $d(\tilde{X}^n, Y^n) \to 0$ a.s. by Proposition 5.24, where d is the metric in Theorem A5.4. Hence, by Theorem 5.29, it is equivalent to replace X^n by Y^n. The assertion then follows by Theorem 27.12. □

We are now ready to prove the main representation theorem for exchangeable processes on $[0,1]$. As noted before, we consider only the case of $d = 1$.

Proof of Theorem 27.10: The sufficiency being obvious, it is enough to prove the necessity. Then let X have exchangeable increments. Introduce the step processes

$$X_t^n = X\Big(2^{-n} [2^n t] \Big), \quad t \in [0,1], \ n \in \mathsf{N},$$

define ν_n as in (15) in terms of the jump sizes of X^n, and put $\alpha_n \equiv X_1$. If the sequence (ν_n) is tight, then $(\alpha_m, \nu_n) \xrightarrow{wd} (\alpha, \nu)$ along a sub-sequence, and Theorem 27.14 yields $X^n \xrightarrow{sd} Y$ along the same sub-sequence, where Y

has a representation as in (11). In particular, $X^n \xrightarrow{fd} Y$, and so the finite-dimensional distributions of X and Y agree for dyadic times. The agreement extends to arbitrary times, since both processes are continuous in probability. Then Lemma 4.24 shows that X has a version in $D_{[0,1]}$, whence Corollary 8.18 yields the desired representation.

To prove the required tightness of (ν_n), let ξ_{nj} denote the increments of X^n, put $\zeta_{nj} = \xi_{nj} - 2^{-n}\alpha_n$, and note that

$$\begin{aligned} \|\nu_n\| &= \textstyle\sum_j \xi_{nj}^2 \\ &= \textstyle\sum_j \zeta_{nj}^2 + 2^{-n}\alpha_n^2. \end{aligned} \tag{16}$$

Writing $\eta_n = 2X^n_{1/2} - X^n_1 = 2X_{1/2} - X_1$, and noting that $\sum_j \zeta_{nj} = 0$, we get the elementary estimates

$$\begin{aligned} E(\eta_n^4 \,|\, \nu_n) &\lesssim \sum_{j\geq 1} \zeta_{nj}^4 + \sum_{i\neq j} \zeta_{ni}^2\, \zeta_{nj}^2 \\ &= \Big(\textstyle\sum_j \zeta_{nj}^2\Big)^2 \\ &\lesssim \Big\{E(\eta_n^2 \,|\, \nu_n)\Big\}^2. \end{aligned}$$

Since η_n is independent of n, the sequence of sums $\sum_j \zeta_{nj}^2$ is tight by Lemma 27.13, and the tightness of (ν_n) follows by (16). □

Next we extend Theorem 14.3 to continuous time and higher dimensions. Say that a random sequence $\xi = (\xi_k)$ in R^d is *rotatable*[11] if $\xi^n U \stackrel{d}{=} \xi^n$ for every finite sub-sequence ξ^n and orthogonal matrix U. A process X on R_+ or $[0,1]$ is said to be rotatable, if it is continuous in probability with $X_0 = 0$, and such that the increments over disjoint intervals of equal lengths are rotatable.

Proposition 27.15 *(rotatable sequences and processes, Schoenberg, Freedman)*

(i) *An infinite random sequence $X = (\xi_n^i)$ in R^d is rotatable iff*

$$\xi_n^i = \sigma_k^i\, \zeta_n^k \ \ a.s., \quad i \leq d,\ n \in \mathsf{N},$$

for a random $d \times d$ matrix σ and some i.i.d. $N(0,1)$ random variables ζ_n^k, $k \leq d$, $n \in \mathsf{N}$.

(ii) *An R^d-valued process $X = (X_t^i)$ on $I = [0,1]$ or R_+ is rotatable iff*

$$X_t^i = \sigma_k^i\, B_t^k \ \ a.s., \quad i \leq d,\ t \in I,$$

for a random $d \times d$ matrix σ and an R^d-valued Brownian motion B on I.

In both cases, $\sigma\sigma'$ is X-measurable and $\mathcal{L}(X)$ is determined by $\mathcal{L}(\sigma\sigma')$.

Proof: (i) As in Theorem 14.3, the ξ_n are conditionally i.i.d. with a common distribution ν on R^d, such that all one-dimensional projections are Gaussian. Letting ρ be the covariance matrix of ν, we may choose σ to be the non-negative square root of ρ, which yields the required representation. The last assertions are now obvious.

[11] short for rotation-invariant in distribution, also called *spherical*

(ii) We may take $I = [0,1]$. Then let $h_1, h_2, \ldots \in L^2[0,1]$ be the Haar system of ortho-normal functions on $[0,1]$, and introduce the associated random vectors $Y_n = \int h_n\, dX$ in R^d, $n \in \mathsf{N}$. Then the sequence $Y = (Y_n)$ is rotatable and hence representable as in (i), which translates into a representation of X as in (ii), for a Brownian motion B' on the dyadic subset of $[0,1]$. By Theorem 8.17, we may extend B' to a Brownian motion B on the entire interval $[0,1]$, and since X is continuous in probability, the stated representation remains a.s. valid for all $t \in [0,1]$. □

We turn to a continuous-time counterpart of the predictable sampling Theorem 27.7. Here we consider predictable mappings V on the index set I that are a.s. measure-preserving, in the sense that $\lambda \circ V^{-1} = \lambda$ a.s. for Lebesgue measure λ on $I = [0,1]$ or R_+. The associated transformations of a process X on I are given by

$$(X \circ V^{-1})_t = \int_I 1\{V_s \le t\}\, dX_s, \quad t \in I,$$

where the right-hand side is a stochastic integral of the predictable process $U_s = 1\{V_s \le t\}$ with respect to the semi-martingale X. Note that if $X_t = \xi[0,t]$ for a random measure ξ on I, the map $t \mapsto (X \circ V^{-1})_t$ reduces to the distribution function of the transformed measure $\xi \circ V^{-1}$, so that $X \circ V^{-1} \overset{d}{=} X$ becomes equivalent to $\xi \circ V^{-1} \overset{d}{=} \xi$.

Theorem 27.16 *(predictable invariance) Let X be an R^d-valued, $\mathcal{F}$-exchangeable process on $I = [0,1]$ or R_+, and let V be an $\mathcal{F}$-predictable mapping on I. Then*

$$\lambda \circ V^{-1} = \lambda \ \text{a.s.} \quad \Rightarrow \quad X \circ V^{-1} \overset{d}{=} X.$$

This holds by definition when V is non-random. The result applies in particular to Lévy processes X. For a Brownian motion or bridge, we have the stronger statements in Theorem 19.9. With some effort, we can prove the result from the discrete-time version in Theorem 27.7 by a suitable approximation. Here we use instead some general time-change results from Chapter 19, which were in turn based on ideas from Chapters 10 and 15. For clarity, we treat the cases of $I = \mathsf{R}_+$ or $[0,1]$ separately.

Proof for $I = \mathsf{R}_+$: Extending the original filtration if necessary, we may take the characteristics of X to be $\mathcal{F}_0$-measurable. First let X have isolated jumps. Then

$$X_t = \alpha\, t + \sigma B_t + \int_0^t \!\! \int x\, \xi(ds\, dx), \quad t \ge 0,$$

for a Cox process ξ on $\mathsf{R}_+ \times \mathsf{R}^d$ directed by $\lambda \otimes \nu$ and an independent Brownian motion B. We may also choose the triple (α, σ, ν) to be $\mathcal{F}_0$-measurable and the pair (B, ξ) to be $\mathcal{F}$-exchangeable.

Since ξ has compensator $\hat{\xi} = \lambda \otimes \nu$, we obtain $\hat{\xi} \circ V^{-1} = \lambda \otimes \nu$, which implies $\xi \circ V^{-1} \overset{d}{=} \xi$ by Lemma 19.14. Next define for every $t \ge 0$ a predictable process $U_{t,r} = 1\{V_r \le t\}$, $r \ge 0$, and note that

$$\int_0^\infty U_{s,r}U_{t,r}\,d[B^i,B^j]_r = \delta_{ij}\int_0^\infty 1\Big\{V_r \le s\wedge t\Big\}\,dr$$
$$= (s\wedge t)\,\delta_{ij}.$$

Using Lemma 19.14 again, we conclude that the R^d-valued process

$$(B\circ V^{-1})_t = \int_0^\infty U_{t,r}\,dB_r, \quad t\ge 0,$$

is Gaussian with the same covariance function as B. The same result shows that $\xi\circ V^{-1}$ and $B\circ V^{-1}$ are conditionally independent given $\mathcal{F}_0$. Hence,

$$\Big\{(\xi,\lambda,B)\circ V^{-1},\,\alpha,\,\sigma,\,\nu\Big\} \stackrel{d}{=} \Big(\xi,\lambda,B,\,\alpha,\,\sigma,\,\nu\Big),$$

which implies $X\circ V^{-1}\stackrel{d}{=}X$.

In the general case, we may write

$$X_t = M_t^\varepsilon + (X_t - M_t^\varepsilon), \quad t\ge 0,$$

where M^ε is the purely discontinuous local martingale formed by all compensated jumps of modulus $\le\varepsilon$. Then $(X-M^\varepsilon)\circ V^{-1}\stackrel{d}{=}(X-M^\varepsilon)$ as above, and it suffices to show that $M_t^\varepsilon\xrightarrow{P}0$ and $(M^\varepsilon\circ V^{-1})_t\xrightarrow{P}0$ as $\varepsilon\to 0$ for fixed $t\ge 0$. In the one-dimensional case, we may use the isometry property of stochastic L^2-integrals in Theorem 20.2, along with the measure-preserving property of V, to see that

$$\begin{aligned} E^{\mathcal{F}_0}(M^\varepsilon\circ V^{-1})_t^2 &= E^{\mathcal{F}_0}\Big(\int_0^\infty 1\{V_s\le t\}\,dM_s^\varepsilon\Big)^2 \\ &= E^{\mathcal{F}_0}\int_0^\infty 1\{V_s\le t\}\,d\langle M^\varepsilon\rangle_s, \\ &= E^{\mathcal{F}_0}\int_0^\infty 1\{V_s\le t\}\,ds\int_{|x|\le\varepsilon}x^2\nu(dx) \\ &= t\int_{|x|\le\varepsilon}x^2\nu(dx)\to 0. \end{aligned}$$

Hence, by Jensen's inequality and dominated convergence,

$$E\Big\{\Big(M^\varepsilon\circ V^{-1}\Big)_t^2\wedge 1\Big\} \le E\Big\{E^{\mathcal{F}_0}\Big(M^\varepsilon\circ V^{-1}\Big)_t^2\wedge 1\Big\}\to 0,$$

which implies $(M^\varepsilon\circ V^{-1})_t\xrightarrow{P}0$. Specializing to $V_s\equiv s$, we get in particular $M_t^\varepsilon\xrightarrow{P}0$. □

Proof for $I=[0,1]$: Extending the original filtration if necessary, we may assume that the coefficients in the representation of X are $\mathcal{F}_0$-measurable. First truncate the sum of centered jumps in the representation of Theorem 27.10. Let X_n denote the remainder after the first n terms, and write $X_n = M_n+\hat{X}_n$. Noting that $\mathrm{tr}[M_n]_1=\sum_{j>n}|\beta_j|^2\to 0$, and using the BDG inequalities in Theorem 20.12, we obtain $(M_n\circ V^{-1})_t\xrightarrow{P}0$ for every $t\in[0,1]$. Using the martingale property of $(X_1-X_t)/(1-t)$ and the symmetry of X, we further see that

$$\begin{aligned} E\Big(\int_0^1 |d\hat{X}_n| \,\Big|\, \mathcal{F}_0\Big) &\lesssim E(X_n^* | \mathcal{F}_0) \\ &\lesssim \big(\mathrm{tr}[X_n]\big)_1^{1/2} \\ &= \Big(\sum_{j>n} |\beta_j|^2\Big)^{1/2} \to 0. \end{aligned}$$

Hence, $(X_n \circ V^{-1})_t \xrightarrow{P} 0$ for all $t \in [0,1]$, which reduces the proof to the case of finitely many jumps.

It is then enough to consider a jointly exchangeable pair, consisting of a marked point process ξ on $[0,1]$ and a Brownian bridge B in R^d. By a simple randomization, we may further reduce to the case where ξ has a.s. distinct marks. It is then equivalent to consider finitely many optional times $\tau_1, \dots, \tau_m$, such that B and the point processes $\xi_j = \delta_{\tau_j}$ are jointly exchangeable.

To apply Proposition 19.15, we may express the continuous martingale M as an integral with respect to the Brownian motion in Lemma 19.10, with the predictable process V replaced by the associated indicator functions $U_r^t = 1\{V_r \le t\}$, $r \in [0,1]$. Then the covariations of the lemma become

$$\begin{aligned} \int_0^1 \big(U_r^s - \bar{U}_r^s\big)\big(U_r^t - \bar{U}_r^t\big)\, dr &= \int_0^1 U_r^s U_r^t \, dr - \lambda U^s \cdot \lambda U^t \\ &= \lambda U^{s\wedge t} - \lambda U^s \cdot \lambda U^t \\ &= s \wedge t - st \\ &= E(B_s B_t), \end{aligned}$$

as required.

As for the random measures $\xi_j = \delta_{\tau_j}$, Proposition 10.23 yields the associated compensators

$$\begin{aligned} \eta_j[0,t] &= \int_0^{t\wedge\tau_j} \frac{ds}{1-s} \\ &= -\log(1 - t\wedge\tau_j), \quad t \in [0,1]. \end{aligned}$$

Since the η_j are diffuse, we get by Theorem 10.24 (ii) the corresponding discounted compensators

$$\begin{aligned} \zeta_j[0,t] &= 1 - \exp\big(-\eta_j[0,t]\big) \\ &= 1 - (1 - t\wedge\tau_j) \\ &= t \wedge \tau_j, \quad t \in [0,1]. \end{aligned}$$

Thus, $\zeta_j = \lambda([0,\tau_j] \cap \cdot) \le \lambda$ a.s., and so $\zeta_j \circ V^{-1} \le \lambda \circ V^{-1} = \lambda$ a.s. Hence, Proposition 19.15 yields

$$\big(B \circ V^{-1}, V_{\tau_1}, \dots, V_{\tau_m}\big) \stackrel{d}{=} (B, \tau_1, \dots, \tau_m),$$

which implies $X \circ V^{-1} \stackrel{d}{=} X$. □

Exercises

1. Let μ_c be the joint distribution of $\tau_1 < \tau_2 < \cdots$, where $\xi = \sum_j \delta_{\tau_j}$ is a stationary Poisson process on R_+ with rate $c > 0$. Show that the μ_c are asymptotically invariant as $c \to 0$.

2. Let the random sequence ξ be conditionally i.i.d. η. Show that ξ is ergodic iff η is a.s. non-random.

3. Let ξ, η be random probability measures on a Borel space such that $E\xi^\infty = E\eta^\infty$. Show that $\xi \stackrel{d}{=} \eta$. (*Hint:* Use the law of large numbers.)

4. Let $\xi_1, \xi_2, \ldots$ be contractable random elements in a Borel space S. Prove the existence of a measurable function $f: [0,1]^2 \to S$ and some i.i.d. $U(0,1)$ random variables $\vartheta_0, \vartheta_1, \ldots$, such that $\xi_n = f(\vartheta_0, \vartheta_n)$ a.s. for all n. (*Hint:* Use Lemma 4.22, Proposition 8.20, and Theorems 8.17 and 27.2.)

5. Let $\xi = (\xi_1, \xi_2, \ldots)$ be an $\mathcal{F}$-contractable random sequence in a Borel space S. Prove the existence of a random measure η, such that for any $n \in \mathsf{Z}_+$ the sequence $\theta^n \xi$ is conditionally i.i.d. η, given $\mathcal{F}_n$ and η.

6. Describe the representation of Lemma 27.5, in the special case where $\xi_1, \ldots, \xi_n$ are conditionally i.i.d. with directing random measure η.

7. Give an example of a finite, exchangeable sequence that is not mixed i.i.d. Also give an example of an exchangeable process on $[0,1]$ that is not mixed Lévy.

8. For a simple point process or diffuse random measure ξ on $[0,1]$, show that ξ is exchangeable iff it is contractable.

9. Say that a finite random sequence in S is contractable if all subsequences of equal length have the same distribution. Give an example of a finite random sequence that is contractable but not exchangeable. (*Hint:* We may take $|S| = 3$ and $n = 2$.)

10. Show that for $|S| = 2$, a finite or infinite sequence ξ in S is exchangeable iff it is contractable. (*Hint:* Regard ξ as a simple point process on $\{1, \ldots, n\}$ or N, and use Theorem 15.8.)

11. State and prove a continuous-time version of Lemma 27.6. (If no regularity conditions are imposed on the exchangeable processes in Theorem 27.9, we need to consider optional times taking countably many values.)

12. Let $\xi_1, \ldots, \xi_n$ be exchangeable random variables, fix a Borel set B, and let $\tau_1 < \cdots < \tau_\nu$ be the indices $k \in \{1, \ldots, n\}$ with $\sum_{j<k} \xi_j \in B$. Construct a random vector $(\eta_1, \ldots, \eta_n) \stackrel{d}{=} (\xi_1, \ldots, \xi_n)$ with $\xi_{\tau_k} = \eta_k$ a.s. for all $k \le \nu$. (*Hint:* Extend the sequence (τ_k) to $k \in (\nu, n]$, and apply Theorem 27.7.)

13. Prove a version of Corollary 27.8 for the *last* maximum.

14. Use Proposition 27.16 to give direct proofs of the relation $\tau_1 \stackrel{d}{=} \tau_2$ in Theorems 14.16 and 14.17. (*Hint:* Proceed as in Theorem 27.8.)

15. Show that any R^d-valued, contractable process on R_+ has a version with rcll paths. (*Hint:* Use the corresponding regularity property of Lévy processes in Chapter 16. Alternatively, use the the result for one-dimensional, exchangeable processes on $[0,1]$ in Theorem 27.10.)

16. Let $\xi_1, \ldots, \xi_n$ be exchangeable random variables, and define $X_t = \sum_k \xi_k 1\{\tau_k \le t\}$, $t \in [0,1]$, where $\tau_1 < \cdots, \tau_n$ form a uniform binomial process on $[0,1]$. Show that X is an exchangeable process on $[0,1]$, write it is the form of Theorem 27.10, and identify the characteristics of X.

17. Let X be a real, continuous, exchangeable process on R_+ with $t^{-1}X_t \to 0$ a.s. Show that X is rotatable.

18. Describe the representation of Theorem 27.10, in the special case where X is a Lévy process with characteristics (a, b, ν).

19. Specialize Theorem 27.14 to suitably normalized sequences of i.i.d. random variables, and compare with Corollary 23.6.

20. For continuous exchangeable processes X on $[0, 1]$ or R_+, show that Theorem 27.16 follows from Theorem 19.9. In what sense is the latter result more general? (*Hint:* Note that in this case, $X_t = \alpha t + \sigma B_t$ for a Brownian motion or bridge B.)

21. For simple exchangeable point processes ξ on $[0, 1]$ or R_+, show that Theorem 27.16 follows from Theorem 10.27 or 15.15. (*Hint:* Note that ξ is exchangeable iff it is a mixed Poisson or binomial process based on λ.)

22. Let M be a continuous local martingale with $[M]_\infty = \infty$, and let ξ be a ql-continuous simple point process on R_+ with unbounded compensator $\tau\xi$. Time-change M and ξ as in Theorems 15.15 and 19.2 to form a Brownian motion B and a Poisson process η. Show that $B \perp\!\!\!\perp \eta$. (*Hint:* Use Lemma 19.14.)

23. Let $\mathcal{F}$ be the filtration induced by a Brownian motion B and some independent i.i.d. $U(0, 1)$ times $\tau_1, \tau_2, \ldots$, and let $V_1, V_2, \ldots$ be $\mathcal{F}$-predictable, a.s. measure-preserving maps on $[0, 1]$. Show that the variables $\sigma_k = V_k \circ \tau_k$ are i.i.d. $U(0, 1)$ and independent of B. (*Hint:* Use Lemma 19.15.)

Chapter 28

Multi-variate Arrays and Symmetries

Separately or jointly exchangeable arrays, contraction and extension properties, coding representations, equivalent coding functions, coupling, conditional independence, shell σ-field, independent entries, invariant coding, symmetric functions, inversion, conditioning and independence, shell-measurable and dissociated arrays, separate or joint rotatability, totally rotatable arrays, symmetric partitions

Our survey of modern probability would be incomplete without a discussion of higher-dimensional random arrays, which appear frequently in a wide range of applications. The most obvious use may be to describe the interactions between the nodes in a random graph or network, where X_{ij} represents the interaction between nodes i and j. Here even higher order interactions such as X_{ijk} are conceivable. The interactions may take values in an abstract space S, and we may allow the one-sided interactions $i \mapsto j$ and $j \mapsto i$ to be different, in general. The simplest example is the adjacency matrix between the nodes of a random graph.

For a totally different application, consider a Hilbert-space valued random process ξ_h, $h \in H$, with associated inner products $\rho_{h,k} = \langle \xi_h, \xi_k \rangle$, for arbitrary $h, k \in H$. Fixing an ortho-normal basis (ONB) $h_1, h_2, \ldots$ in H, we obtain a discrete array $X_{ij} = \rho_{h_i,h_j}$. We may also think of $X = (X_{ij})$ as the coordinate representation of an operator on H, representing an observable quantity in quantum mechanics. More mundane examples of random arrays appear frequently in statistics, such as in the context of U-statistics. We may further mention the case of random partitions of a fixed, countable set.

In all those applications, it is natural to impose suitable symmetry conditions. For the graph interactions, we may think of all nodes as equivalent, or we may enumerate them in random order, which leads to a condition of *joint exchangeability* of the interaction array X. If X instead codifies the interactions between two different kinds of nodes, those may be numbered separately, which suggests an assumption of *separate exchangeability*. For processes on a Hilbert space, we may think of all ortho-normal bases as equivalent, which leads to a condition of *joint rotatability*. Indeed, for operators in quantum mechanics there is no preferred ONB, simply because the individual coordinate representations are not observable.

In this chapter, we derive representations of multi-variate random arrays with exchangeable, contractable, or rotatable symmetries. In the exchangeable case, there are no simple mixing representations in terms of i.i.d. arrays, and all higher-dimensional representations are instead extensions of the coding

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_29

representation in Corollary 27.3. They yield in particular a remarkable extension property, generalizing the classical equivalence between exchangeable and contractable sequences.

The higher-dimensional proofs are quite intricate, and involve a subtle use of notions and results for conditional independence from Chapter 8. The rotatable case is even harder, as it leads in general to representations in terms of multiple Wiener-Itô integrals such as in Chapter 14. Here we discuss only the relatively elementary representations in two dimensions, which can be stated in terms of i.i.d. Gaussian random variables. We finally include a version of Kingman's paintbox representation for exchangeable and related partitions.

To prepare for our more precise technical discussions, we say that a random array $X = (X_k)$, indexed by vectors $k = (k_1, \dots, k_d) \in \mathsf{N}^d$, is *separately exchangeable*, if its distribution is invariant under arbitrary permutations $p^1, \dots, p^d$ in the d indices. We also consider the weaker property of *joint exchangeability*, where the same invariance is only required for a single permutation $p = (p_1, p_2, \dots)$ of N, so that

$$X_{k_1,\dots,k_d} \stackrel{d}{=} X_{p(k_1),\dots,p(k_d)}, \quad k \in \mathsf{N}^d. \tag{1}$$

In the latter case, it is equivalent and more natural to consider arrays X indexed by the non-diagonal set $\mathsf{N}^{(d)}$, where all indices $k_1, \dots, k_d$ are different.

Replacing the permutations in (1) by sub-sequences $p_1 < p_2 < \cdots$ yields the even weaker notion of *(joint) contractability*[1]. Here it is equivalent and often preferable to consider arrays indexed by the *tetrahedral* set $\mathsf{N}^{\uparrow d}$ of increasing sequences $k_1 < \ldots < k_d$, which may be identified with the corresponding *sets* of components $\tilde{k} = \{k_1, \dots, k_d\}$. Define

$$\hat{\mathsf{N}}^d = \bigcup_{k \le d} \mathsf{N}^k, \qquad \tilde{\mathsf{N}}^d = \bigcup_{k \le d} \mathsf{N}^{\uparrow k}, \qquad \hat{\mathsf{N}}^{(d)} = \bigcup_{k \le d} \mathsf{N}^{(k)}.$$

By a *U-array* we mean an indexed set of i.i.d. $U(0,1)$ random variables[2]. For any U-array ξ on $\tilde{\mathsf{N}}^d$ and vector $k \in \mathsf{N}^{(d)}$ or set $J \in \mathsf{N}^{\uparrow d}$, we write

$$\hat{\xi}_k = \{\xi_{\tilde{h}};\, h \subset k\}, \qquad \hat{\xi}_J = \{\xi_I;\, I \subset J\},$$

where $h \subset k$ means that h is a sub-sequence of k, and $\tilde{h}$ denotes the set of components of the vector h. Here the order of enumeration, important for subsequent representations, is determined by the order of the k-components. Thus, the index set for $d = 2$ consists of the sets $\emptyset, \{k_1\}, \{k_2\}, \{k_1, k_2\}$, whereas for $d = 3$ we have the index set

$$\emptyset, \{k_1\}, \{k_2\}, \{k_3\}, \{k_1, k_2\}, \{k_1, k_3\}, \{k_2, k_3\}, \{k_1, k_2, k_3\}.$$

The jointly exchangeable and contractable arrays can be characterized by the following coding representations, extending the one-dimensional version in Corollary 27.3. Its proof will take up much of the present chapter.

[1] Here the qualification 'joint' can be dropped, since for infinite arrays the notions of separate exchangeability or contractability are equivalent by Ryll-Nardzewski's theorem.

[2] There is nothing special about the uniform distribution, and we could use any other diffuse distribution on a Borel space.

Theorem 28.1 *(jointly exchangeable and contractable arrays, Aldous, Hoover, OK) For a d-dimensional random array X in a Borel space S, we have*

(i) *$X = (X_k)$ is jointly exchangeable on $\hat{\mathsf{N}}^{(d)}$ iff there exist a measurable function $f\colon [0,1]^{2^d} \to S$ and a U-array $\xi = (\xi_J)$ on $\tilde{\mathsf{N}}^d$, such that*

$$X_k = f(\hat{\xi}_k) \ \ a.s., \quad k \in \hat{\mathsf{N}}^{(d)},$$

(ii) *$X = (X_J)$ is contractable on $\tilde{\mathsf{N}}^d$ iff it can be extended to a jointly exchangeable array on $\hat{\mathsf{N}}^{(d)}$, and hence can be represented as in* (i),

(iii) *the array in* (ii) *is exchangeable on $\tilde{\mathsf{N}}^d$ iff we can choose the representing function f to be symmetric.*

When $d = 2$, the representation in (i) becomes

$$X_{ij} = f(\alpha, \xi_i, \xi_j, \zeta_{ij}), \quad i \neq j \text{ in } \mathsf{N}, \tag{2}$$

for a measurable function $f\colon [0,1]^4 \to S$ and some i.i.d. $U(0,1)$ random variables α, ξ_i, and $\zeta_{ij} = \zeta_{ji}$. Part (i) yields a similar representation of separately exchangeable arrays, where the representing U-array is now indexed by $\hat{\mathsf{N}}^d$, and $\hat{\xi}_k$ is indexed by all sub-sequences $h \subset k$. Thus, the index set for $d = 2$ has elements $\emptyset, k_1, k_2, (k_1, k_2)$, where $\emptyset$ denotes the empty sequence, whereas for $d = 3$ we have the index set

$$\emptyset,\ k_1, k_2, k_3, (k_1, k_2), (k_1, k_3), (k_2, k_3), (k_1, k_2, k_3).$$

The representation in the separately exchangeable case is the same as before, apart from the dependencies between coding variables.

Corollary 28.2 *(separately exchangeable arrays) An S-valued array $X = (X_k)$ on N^d is separately exchangeable iff there exist a measurable function $f\colon [0,1]^{2^d} \to S$ and a U-array $\xi = (\xi_h)$ on $\hat{\mathsf{N}}^d$, such that*

$$X_k = f(\hat{\xi}_k) \ \ a.s., \quad k \in \mathsf{N}^d. \tag{3}$$

Thus, for $d = 2$ we get the representation

$$X_{ij} = f(\alpha, \xi_i, \eta_j, \zeta_{ij}), \quad i, j \in \mathsf{N}, \tag{4}$$

for a measurable function $f\colon [0,1]^4 \to S$ and some i.i.d. $U(0,1)$ random variables α, ξ_i, η_j, ζ_{ij}.

Proof: If X is separately exchangeable, it is also jointly exchangeable and hence can be represented as in Theorem 28.1 (i), in terms of a U-array $\xi = (\xi_J)$ on $\tilde{\mathsf{N}}^d$. Now choose any disjoint, countable sets $N_1, \ldots, N_d \subset \mathsf{N}$, and let Y be the restriction of X to $N_1 \times \cdots \times N_d$. Then the representation of Y reduces to the form (3), and since $X \overset{d}{=} Y$ apart from index sets, Theorem 8.17 yields the same representation for X. $\square$

Before proceeding to the proof of Theorem 28.1, we note that the representing functions are far from unique. It then becomes important to determine when two functions f and g can be used to represent the same array. The answer differs in the three cases, and for convenience we consider only jointly exchangeable arrays of dimension 2, where the more explicit statements may clarify the basic ideas. See also the one-dimensional version in Lemma 27.4. Put $I = [0,1]$ for convenience.

Theorem 28.3 *(equivalent coding functions, Hoover, OK) For any measurable functions $f, g\colon I^4 \to S$ and i.i.d. $U(0,1)$ random variables α, ξ_i, $\zeta_{ij} = \zeta_{ji}$ and α', ξ'_i, $\zeta'_{ij} = \zeta'_{ji}$, these conditions are equivalent*[3]*:*

(i) *there exist some variables $\{\tilde\alpha, \tilde\xi_i, \tilde\zeta_{ij};\, i \neq j\} \overset{d}{=} \{\alpha, \xi_i, \zeta_{ij};\, i \neq j\}$, such that*
$$f(\alpha, \xi_i, \xi_j, \zeta_{ij}) = g(\tilde\alpha, \tilde\xi_i, \tilde\xi_j, \tilde\zeta_{ij}) \text{ a.s.},\ i \neq j,$$

(ii) $$\Big\{f(\alpha, \xi_i, \xi_j, \zeta_{ij});\ i \neq j\Big\} \overset{d}{=} \Big\{g(\alpha, \xi_i, \xi_j, \zeta_{ij});\ i \neq j\Big\},$$

(iii) *there exist some measurable functions*
$$T, T'\colon I \to I, \qquad U, U'\colon I^2 \to I, \qquad V, V'\colon I^4 \to I,$$
where V, V' are symmetric and all functions preserve λ in the highest order arguments, such that
$$\begin{aligned} &f\Big\{T(\alpha),\, U(\alpha,\xi_i),\, U(\alpha,\xi_j),\, V(\alpha,\xi_i,\xi_j,\zeta_{ij})\Big\} \\ &\qquad = g\Big\{T'(\alpha),\, U'(\alpha,\xi_i),\, U'(\alpha,\xi_j),\, V'(\alpha,\xi_i,\xi_j,\zeta_{ij})\Big\}, \end{aligned}$$

(iv) *there exist some measurable functions*
$$T\colon I^2 \to I, \qquad U\colon I^4 \to I, \qquad V\colon I^8 \to I,$$
where V is symmetric and all functions map λ^2 into λ in the highest order arguments, such that
$$\begin{aligned} f(\alpha,\xi_i,\xi_j,\zeta_{ij}) = g\Big\{&T(\alpha,\alpha'),\, U(\alpha,\alpha',\, \xi_i,\xi'_i),\, U(\alpha,\alpha',\, \xi_j,\xi'_j), \\ &V(\alpha,\alpha',\, \xi_i,\xi'_i,\, \xi_j,\xi'_j,\, \zeta_{ij},\zeta'_{ij})\Big\}. \end{aligned}$$

Note that (iii) is expressed in terms of six functions T, T', U, U', V, V', whereas (iv) requires only the three functions T, U, V. On the other hand, we need[4] in (iv) (but not in (iii)) some external randomization variables α', ξ'_i, $\zeta'_{ij} = \zeta'_{ji}$.

[3]This result is often misunderstood: it is not enough in (iii) to consider measure-preserving transformations U, V or U', V' of a single variable, nor is (iv) true without the additional randomization variables $\alpha', \xi'_i, \zeta'_{ij}$.

[4]This is essentially because a permutation on N is invertible, whereas a measure-preserving transformation on $[0,1]$ is not in general, which is why we can't solve for f in (iii) without a suitable randomization.

To prepare for the proofs of Theorems 28.1 and 28.3, we begin with some general lemmas of independent interest.

Lemma 28.4 *(coupling) Let X, Y, Z be S-valued arrays with $X \perp\!\!\!\perp_Y Z$. Then*
(i) *if (X, Y), (Y, Z) are exchangeable on $\hat{\mathsf{N}}^{(d)}$, so is (X, Y, Z),*
(ii) *if (X, Y), (Y, Z) are contractable on $\tilde{\mathsf{Q}}^d$, so is (X, Y, Z).*

Proof: (i) Writing $\mu(Y, \cdot) = \mathcal{L}(X \mid Y)$, we get for any measurable sets A, B and permutation p on N

$$\begin{aligned}
P&\Big\{(X, Y, Z) \circ p \in A \times B\Big\} \\
&= E\Big(P\Big\{X \circ p \in A \,\Big|\, (Y, Z) \circ p\Big\};\ (Y, Z) \circ p \in B\Big) \\
&= E\Big(P\Big\{X \circ p \in A \,\Big|\, Y \circ p\Big\};\ (Y, Z) \circ p \in B\Big) \\
&= E\Big\{\mu(Y \circ p, A);\ (Y, Z) \circ p \in B\Big\} \\
&= E\Big\{\mu(Y, A);\ (Y, Z) \in B\Big\} \\
&= E\Big(P\{X \in A \mid Y, Z\};\ (Y, Z) \in B\Big) \\
&= P\Big\{(X, Y, Z) \in A \times B\Big\},
\end{aligned}$$

which extends immediately to $(X, Y, Z) \circ p \overset{d}{=} (X, Y, Z)$.

(ii) For any $a < b$ in Q_+ and $t \in \mathsf{Q}$, we define on Q the functions

$$\begin{aligned}
p_{a,t}(x) &= x + a\, 1\{x > t\}, \qquad p_a = p_{a,0}, \\
p_a^b(x) &= x + a\,(1 - b^{-1}x)\, 1\{0 < x < b\},
\end{aligned}$$

and note that $p_a^b \circ p_b = p_b$. The contractability of (X, Y) yields $(X, Y) \circ p_a^b \overset{d}{=} (X, Y)$, and so

$$\begin{aligned}
(X \circ p_b, Y) &\overset{d}{=} \Big(X \circ p_a^b \circ p_b,\ Y \circ p_a^b\Big) \\
&= \Big(X \circ p_b,\ Y \circ p_a^b\Big).
\end{aligned}$$

By Corollary 8.10, we get

$$(X \circ p_b) \underset{Y \circ p_a^b}{\perp\!\!\!\perp} Y,$$

and since $\sigma(Y \circ p_a^b) = \sigma(Y \circ p_a)$ and $\sigma(Y \circ p_b;\ b > a) = \sigma(Y \circ p_a)$, a monotone-class argument yields the extended relation

$$(X \circ p_a) \underset{Y \circ p_a}{\perp\!\!\!\perp} Y.$$

Since also $(X \circ p_a) \perp\!\!\!\perp_Y (Z \circ p_a)$ by hypothesis, Theorem 8.12 gives

$$(X \circ p_a) \underset{Y \circ p_a}{\perp\!\!\!\perp} (Z \circ p_a).$$

Combining this with the hypothetical relations

$$X \perp\!\!\!\perp_Y Z, \qquad \begin{cases} (X, Y) \circ p_a \overset{d}{=} (X, Y), \\ (Y, Z) \circ p_a \overset{d}{=} (Y, Z), \end{cases}$$

we see as in case (i) that

$$(X, Y, Z) \circ p_a \overset{d}{=} (X, Y, Z), \quad a \in \mathsf{Q}_+.$$

Relabeling Q if necessary, we get the same relation for the more general mappings $p_{a,t}$. For any finite sets $I, J \subset \mathsf{Q}$ with $|I| = |J|$, we may choose compositions p and q of such functions satisfying $p(I) = q(J)$, and the required contractability follows. □

Lemma 28.5 *(conditional independence) Let (X, ξ) be a contractable array on $\tilde{\mathsf{Z}}^d$ with restriction (Y, η) to $\tilde{\mathsf{N}}^d$, where ξ has independent entries. Then*

$$Y \underset{\eta}{\perp\!\!\!\perp} \xi.$$

Proof: Putting

$$p_n(k) = k - n\, 1\{k \le 0\}, \quad k \in \mathsf{Z}, \; n \in \mathsf{N},$$

we get by the contractability of (X, ξ)

$$(Y, \, \xi \circ p_n) \overset{d}{=} (Y, \xi), \quad n \in \mathsf{N}.$$

Since also $\sigma\{\xi \circ p_n\} \subset \sigma\{\xi\}$, Corollary 8.10 yields

$$Y \underset{\xi \circ p_n}{\perp\!\!\!\perp} \xi, \quad n \in \mathsf{N}.$$

Further note that $\bigcap_n \sigma\{\xi \circ p_n\} = \sigma\{\eta\}$ a.s. by Corollary 9.26. The assertion now follows by martingale convergence as $n \to \infty$. □

For any S-valued random array X on N^d, we introduce the associated *shell σ-field*

$$\mathcal{S}(X) = \bigcap\nolimits_n \sigma\Big\{X_k; \, \max\nolimits_i k_i \ge n\Big\}.$$

Proposition 28.6 *(shell σ-field, Aldous, Hoover) Let the array X on Z_+^d be separately exchangeable under permutations on N, and write X^+ for the restriction of X to N^d. Then*

$$X^+ \underset{\mathcal{S}(X^+)}{\perp\!\!\!\perp} (X \setminus X^+).$$

Proof: When $d = 1$, X is just a random sequence $(\eta, \, \xi_1, \xi_2, \ldots) = (\eta, \xi)$, and $\mathcal{S}(X^+)$ reduces to the tail σ-field $\mathcal{T}$. Applying de Finetti's theorem to both ξ and (ξ, η) gives

$$\mathcal{L}(\xi \,|\, \mu) = \mu^\infty, \qquad \mathcal{L}(\xi \,|\, \nu, \eta) = \nu^\infty,$$

for some random probability measures μ and ν on $\mathcal{S}$. By the law of large numbers, μ and ν are $\mathcal{T}$-measurable with $\mu = \nu$ a.s., and so $\xi \perp\!\!\!\perp_{\mathcal{T}} \eta$ by Corollary 8.10.

Assuming the truth in dimension $d-1$, we turn to the d-dimensional case. Define an S^∞-valued array Y on Z_+^{d-1} by

$$Y_m = \Big(X_{m,k};\ k \ge 0\Big), \quad m \in \mathsf{Z}_+^{d-1},$$

and note that Y is separately exchangeable under permutations on N. Writing Y^+ for the restriction of Y to N^{d-1} and X^0 for the restriction of X to $\mathsf{N}^{d-1} \times \{0\}$, we get by the induction hypothesis

$$\begin{aligned}(X^+, X^0) &= Y^+ \underset{\mathcal{S}(Y^+)}{\perp\!\!\!\perp} (Y \setminus Y^+) \\ &= X \setminus (X^+, X^0).\end{aligned}$$

Since also

$$\begin{aligned}\mathcal{S}(Y^+) &\subset \mathcal{S}(X^+) \vee \sigma(X^0) \\ &\subset \sigma(Y^+),\end{aligned}$$

we obtain

$$X^+ \underset{\mathcal{S}(X^+),\, X^0}{\perp\!\!\!\perp} (X \setminus X^+). \tag{5}$$

Next we apply the result for $d = 1$ to the S^∞-valued sequence

$$Z_k = \Big(X_{m,k};\ m \in \mathsf{N}^{d-1}\Big), \quad k \ge 0,$$

to obtain

$$X^+ = (Z \setminus Z_0) \underset{\mathcal{S}(Z)}{\perp\!\!\!\perp} Z_0 = X^0.$$

Since clearly

$$\mathcal{S}(Z) \subset \mathcal{S}(X^+) \subset \sigma(X^+),$$

we conclude that

$$X^+ \underset{\mathcal{S}(X^+)}{\perp\!\!\!\perp} X^0.$$

Now combine with (5) and use Theorem 8.12 to get the desired relation. □

We need only the following corollary. Given an array ξ on Z_+^d and a set $I \subset \mathsf{N}_d = \{1, \ldots, d\}$, we write ξ^I for the sub-array on the index set $\mathsf{N}^I \times \{0\}^{I^c}$, and put $\hat{\xi}^J = (\xi^I;\ I \subset J)$. Further define $\xi^m = (\xi^I;\ |I| = m)$ and $\hat{\xi}^m = (\xi^I;\ |I| \le m)$, and similarly for arrays on $\hat{\mathsf{N}}^{\uparrow d}$. Recall that $\mathsf{N}^{\uparrow d} = (J \subset \mathsf{N};\ |J| = d)$ and $\tilde{\mathsf{N}}^d = (J \subset \mathsf{N};\ |J| \le d)$.

Corollary 28.7 *(independent entries)*

(i) *Let ξ be a random array on Z_+^d with independent entries, separately exchangeable under permutations on N, and define*

$$\eta_k = f(\hat{\xi}_k), \quad k \in \mathsf{Z}_+^d,$$

for a measurable function f. Then even η has independent entries iff

$$\eta_k \perp\!\!\!\perp (\hat{\xi}_k \setminus \xi_k), \quad k \in \mathsf{Z}_+^d, \tag{6}$$

in which case

$$\eta^I \perp\!\!\!\perp (\xi \setminus \xi^I), \quad I \subset \mathsf{N}_d. \tag{7}$$

(ii) *Let ξ be a contractable array on $\tilde{\mathsf{N}}^d$ with independent entries, and define*

$$\eta_J = f(\hat{\xi}_J), \quad J \in \tilde{\mathsf{N}}^d, \tag{8}$$

for a measurable function f. Then even η has independent entries iff

$$\eta_J \perp\!\!\!\perp (\hat{\xi}_J \setminus \xi_J), \quad J \in \tilde{\mathsf{N}}^d, \tag{9}$$

in which case

$$\eta^k \perp\!\!\!\perp (\xi \setminus \xi^k), \quad k \le d. \tag{10}$$

Proof: (i) Assume (6). Since also $\hat{\xi}_k \perp\!\!\!\perp (\xi \setminus \hat{\xi}_k)$, we get $\eta_k \perp\!\!\!\perp (\xi \setminus \xi_k)$, and so

$$\eta_k \perp\!\!\!\perp \left(\xi \setminus \xi_k,\, \hat{\eta}^{|k|} \setminus \eta_k\right), \quad k \in \mathsf{Z}_+^d,$$

where $|k|$ is the number of k-components in N. As in Lemma 4.8, this yields both (7) and the independence of the η-entries.

Conversely, let η have independent entries. Then $\mathcal{S}(\eta^I)$ is trivial for $\emptyset \neq I \subset \mathsf{N}_d$ by Kolmogorov's 0−1 law. Applying Lemma 28.6 to the Z_+^I-indexed array $(\eta^I, \hat{\xi}^I \setminus \xi^I)$, which is clearly separately exchangeable on N^I, we obtain $\eta^I \perp\!\!\!\perp (\hat{\xi}^I \setminus \xi^I)$, and (6) follows.

(ii) Assume (9). Since also $\hat{\xi}_J \perp\!\!\!\perp (\xi \setminus \hat{\xi}_J)$, we obtain $\eta_J \perp\!\!\!\perp (\xi \setminus \xi_J)$, and so

$$\eta_J \perp\!\!\!\perp \left(\xi \setminus \xi^d,\, \hat{\eta}^d \setminus \eta^J\right), \quad J \in \tilde{\mathsf{N}}^d.$$

Thus, η has independent entries and satisfies (10).

Conversely, let η have independent entries. Extend ξ to a contractable array on $\mathsf{Q}^{\uparrow d}$, and extend η accordingly to $\mathsf{Q}^{\uparrow d}$ by means of (8). Define

$$(\xi'_k, \eta'_k) = (\xi, \eta)_{r(1,k_1),\dots,r(d,k_d)}, \quad k \in \mathsf{Z}_+^d,$$

where $r(i,k) = i + k(k+1)^{-1}$. Then ξ' is separately exchangeable on N^d with independent entries, and (8) yields $\eta'_k = f(\hat{\xi}'_k)$ for all $k \in \mathsf{Z}_+^d$. Since $\eta' \subset \eta$ has again independent entries, (i) yields $\eta'_k \perp\!\!\!\perp (\hat{\xi}'_k \setminus \xi'_k)$ for all $k \in \mathsf{N}^d$, or equivalently $\eta_J \perp\!\!\!\perp (\hat{\xi}_J \setminus \xi_J)$ for suitable $J \in \mathsf{Q}^{\uparrow d}$. The general relation (9) now follows by the contractability of ξ. □

Lemma 28.8 *(coding) Let $\xi = (\xi_j)$ and $\eta = (\eta_j)$ be random arrays on a countable index set I, satisfying*

$$(\xi_i, \eta_i) \overset{d}{=} (\xi_j, \eta_j), \quad i, j \in I, \tag{11}$$

$$\xi_j \underset{\eta_j}{\perp\!\!\!\perp} (\xi \setminus \xi_j, \eta), \quad j \in I. \tag{12}$$

Then there exist some measurable functions f, g, such that for any U-array $\vartheta \perp\!\!\!\perp (\xi, \eta)$ on I, the random variables

$$\zeta_j = g(\xi_j, \eta_j, \vartheta_j), \quad j \in I, \tag{13}$$

form a U-array $\zeta \perp\!\!\!\perp \eta$ satisfying

$$\xi_j = f(\eta_j, \zeta_j) \;\; a.s., \quad j \in I. \tag{14}$$

Proof: For any $i \in I$, Theorem 8.17 yields a measurable function f, such that

$$(\xi_i, \eta_i) \overset{d}{=} \left\{f(\eta_i, \vartheta_i),\, \eta_i\right\}.$$

The same theorem provides some measurable functions g and h, such that the random elements

$$\zeta_i = g(\xi_i, \eta_i, \vartheta_i), \qquad \tilde{\eta}_i = h(\xi_i, \eta_i, \vartheta_i)$$

satisfy

$$(\xi_i, \eta_i) = \left\{ f(\tilde{\eta}_i, \zeta_i),\, \tilde{\eta}_i \right\} \text{ a.s.,}$$
$$(\tilde{\eta}_i, \zeta_i) \overset{d}{=} (\eta_i, \vartheta_i).$$

In particular $\tilde{\eta}_i = \eta_i$ a.s., and so (14) holds for $j = i$ with ζ_i as in (13). Furthermore, ζ_i is $U(0,1)$ and independent of η_i. The two statements extend by (11) to arbitrary $j \in I$.

Combining (12) with the relations $\vartheta \perp\!\!\!\perp (\xi, \eta)$ and $\vartheta_j \perp\!\!\!\perp (\vartheta \setminus \vartheta_j)$, we get

$$(\xi_j, \eta_j, \vartheta_j) \underset{\eta_j}{\perp\!\!\!\perp} \left(\xi \setminus \xi_j,\, \eta,\, \vartheta \setminus \vartheta_j\right), \quad j \in I,$$

which implies

$$\zeta_j \underset{\eta_j}{\perp\!\!\!\perp} (\zeta \setminus \zeta_j,\, \eta), \quad j \in I.$$

Since also $\zeta_j \perp\!\!\!\perp \eta_j$, Theorem 8.12 yields $\zeta_j \perp\!\!\!\perp (\zeta \setminus \zeta_j,\, \eta)$. Then Lemma 4.8 shows that the array η and the elements ζ_j, $j \in I$, are all independent. Thus, ζ is indeed a U-array independent of η. □

Proposition 28.9 *(invariant coding) Let G be a finite group acting measurably on a Borel space S, and let ξ, η be random elements in S satisfying*

$$r(\xi, \eta) \overset{d}{=} (\xi, \eta), \quad r\eta \neq \eta \ \text{ a.s.}, \qquad r \in G.$$

Then for any $U(0,1)$ random variable $\vartheta \perp\!\!\!\perp (\xi, \eta)$,

(i) *there exists a measurable function $f\colon S \times [0,1] \to S$ and a $U(0,1)$ random variable $\zeta \perp\!\!\!\perp \eta$, such that a.s.*

$$r\xi = f(r\eta, \zeta), \quad r \in G,$$

(ii) *we may choose $\zeta = b(\xi, \eta, \vartheta)$, for a measurable function $b\colon S^2 \times [0,1] \to [0,1]$ satisfying*

$$b(rx, ry, t) = b(x, y, t), \quad x, y \in S, \ r \in G, \ t \in [0,1].$$

Our proof relies on a simple algebraic fact:

Lemma 28.10 *(invariant functions) Let G be a finite group acting measurably on a Borel space S, such that $rs \neq s$ for all $r \in G$ and $s \in S$. Then there exists a measurable function $h\colon S \to G$, such that*

(i) $h(rs)\, r = h(s), \quad r \in G, \ s \in S,$

(ii) *for any map $b\colon S \to S$, the function*

$$f(s) = h_s^{-1}\, b(h_s\, s), \quad s \in S,$$

satisfies

$$f(rs) = rf(s), \quad r \in G, \ s \in S.$$

Proof: (i) We may take $S \subset \mathsf{R}$. For any $s \in S$, the elements rs are all different, and we may choose h_s to be the element $r \in G$ maximizing rs. Then h_{rs} is the element $p \in G$ maximizing $p(rs)$, so that pr maximizes $(pr)s$, and hence $h_{rs}r = p\,r = h_s$.

(ii) Using (i) and the definition of f, we get for any $r \in G$ and $s \in S$

$$\begin{aligned} f(rs) &= h_{rs}^{-1}\, b(h_{rs}\, rs) \\ &= h_{rs}^{-1}\, b(h_s\, s) \\ &= h_{rs}^{-1}\, h_s\, f(s) = r\, f(s). \end{aligned}$$

□

Proof of Lemma 28.9: (i) Assuming $r\eta \neq \eta$ identically, we may define h as in Lemma 28.10. Writing $\tilde{\xi} = h_\eta\, \xi$ and $\tilde{\eta} = h_\eta\, \eta$, we have for any $r \in G$

$$\begin{aligned} (\tilde{\xi}, \tilde{\eta}) &= h_\eta(\xi, \eta) \\ &= h_{r\eta}\, r(\xi, \eta). \end{aligned}$$

Letting $\gamma \perp\!\!\!\perp (\xi, \eta)$ be uniformly distributed in G, we get by Fubini's theorem

$$\begin{aligned} \gamma(\xi, \eta) &\overset{d}{=} (\xi, \eta), \\ (\gamma h_\eta,\, \xi, \eta) &\overset{d}{=} (\gamma, \xi, \eta). \end{aligned}$$

Combining those relations and using Lemma 28.10 (i), we obtain

$$\begin{aligned} (\eta, \tilde{\xi}, \tilde{\eta}) &= \Big(\eta,\, h_\eta\, \xi,\, h_\eta\, \eta\Big) \\ &\overset{d}{=} \Big(\gamma\eta,\, h_{\gamma\eta}\gamma\, \xi,\, h_{\gamma\eta}\gamma\, \eta\Big) \\ &= \Big(\gamma\eta,\, h_\eta\, \xi,\, h_\eta\, \eta\Big) \\ &\overset{d}{=} \Big(\gamma h_\eta\, \eta,\, h_\eta\, \xi,\, h_\eta\, \eta\Big) \\ &= (\gamma\tilde{\eta},\, \tilde{\xi},\, \tilde{\eta}). \end{aligned}$$

Since also $\gamma\tilde{\eta} \perp\!\!\!\perp_{\tilde{\eta}} \tilde{\xi}$ by the independence $\gamma \perp\!\!\!\perp (\xi, \eta)$, we get $\eta \perp\!\!\!\perp_{\tilde{\eta}} \tilde{\xi}$. Hence, Proposition 8.20 yields a measurable function $g\colon S \times [0,1] \to S$ and a $U(0,1)$ random variable $\zeta \perp\!\!\!\perp \eta$, such that a.s.

$$\begin{aligned} h_\eta\, \xi = \tilde{\xi} &= g(\tilde{\eta}, \zeta) \\ &= g(h_\eta\, \eta,\, \zeta). \end{aligned}$$

This shows that $\xi = f(\eta, \zeta)$ a.s., where

$$f(s,t) = h_s^{-1} g(h_s s, t), \quad s \in S,\ t \in [0,1],$$

for a measurable extension of h to S. Using Lemma 28.10 (ii), we obtain a.s.

$$\begin{aligned} r\xi &= r f(\eta, \zeta) \\ &= f(r\eta, \zeta), \quad r \in G. \end{aligned}$$

(ii) Defining $rt = t$ for $r \in G$ and $t \in [0,1]$, we have by (i) and the independence $\zeta \perp\!\!\!\perp \eta$

$$r(\xi,\eta,\zeta) = (r\xi,\, r\eta,\, \zeta) \\ \stackrel{d}{=} (\xi,\eta,\zeta), \quad r\in G.$$

Proceeding as in (i), we may choose a measurable function $b\colon S^2\times[0,1]\to[0,1]$ and a $U(0,1)$ random variable $\vartheta\perp\!\!\!\perp(\xi,\eta)$, such that a.s.

$$\zeta = b(r\xi,\, r\eta,\, \vartheta) \;\text{ a.s.}, \quad r\in G.$$

Averaging over G, we may assume relation (ii) to hold identically. Letting $\vartheta'\perp\!\!\!\perp(\xi,\eta)$ be $U(0,1)$ and putting $\zeta' = b(\xi,\eta,\vartheta')$, we get $(\xi,\eta,\zeta')\stackrel{d}{=}(\xi,\eta,\zeta)$, and so (i) remains true with ζ replaced by ζ'. □

We proceed with some elementary properties of symmetric functions. Note that if two arrays ξ, η are related by $\eta_J = f(\hat\xi_J)$, then $\hat\eta_J = \hat f(\hat\xi_J)$ with

$$\hat f\Big(x_I;\, I\subset \mathsf{N}_d\Big) = \Big\{f(x_{J\circ I};\, I\subset \mathsf{N}_{|J|});\, J\subset\mathsf{N}_d\Big\}.$$

Lemma 28.11 *(symmetric functions) Let ξ be a random array on $\tilde{\mathsf{N}}^d$, and let f,g be measurable functions between suitable spaces. Then*

(i) $\eta_J = f(\hat\xi_J)$ *on* $\tilde{\mathsf{N}}^d \Rightarrow \hat\eta_J = \hat f(\hat\xi_J)$ *on* $\tilde{\mathsf{N}}^d$,

(ii) $\eta_k = f(\hat\xi_k)$ *on* $\hat{\mathsf{N}}^{(d)}$ *with f symmetric* $\Rightarrow \hat\eta_k = \hat f(\hat\xi_k)$ *on* $\hat{\mathsf{N}}^{(d)}$,

(iii) *when f,g are symmetric, so is $f\circ\hat g$,*

(iv) *for exchangeable ξ and symmetric f, the array $\eta_J = f(\hat\xi_J)$ on $\mathsf{N}^{\uparrow d}$ is again symmetric.*

Proof: (i) This just restates the definition.

(ii) Let $k\in\mathsf{N}^{(d)}$ and $J\subset\mathsf{N}_d$. By the symmetry of f and the definition of $\hat\xi$,

$$f\Big(\xi_{k\circ(J\circ I)};\, I\subset\mathsf{N}_{|J|}\Big) = \Big(\xi_{(k\circ J)\circ I};\, I\subset\mathsf{N}_{|J|}\Big) \\ = f(\hat\xi_{k\circ I}).$$

Hence, by the definitions of $\hat\xi$, $\hat f$, η, $\hat\eta$,

$$\begin{aligned}\hat f(\hat\xi_k) &= \hat f\Big(\xi_{k\circ I};\, I\subset\mathsf{N}_d\Big)\\ &= \Big\{f\Big(\xi_{k\circ(J\circ I)};\, I\subset\mathsf{N}_{|J|}\Big);\, J\subset\mathsf{N}_d\Big\}\\ &= \Big\{f(\hat\xi_{k\circ J});\, J\subset\mathsf{N}_d\Big\}\\ &= \Big(\eta_{k\circ J};\, J\subset\mathsf{N}_d\Big) = \hat\eta_k.\end{aligned}$$

(iii) Put $\eta_k = g(\hat\xi_k)$ and $h = f\circ\hat g$. Applying (ii) to g and using the symmetry of f, we get for any vector $k\in\mathsf{N}^{(d)}$ with associated subset $\tilde k\in\mathsf{N}^{\uparrow d}$

$$\begin{aligned}h(\hat\xi_k) &= f\circ\hat g(\hat\xi_k)\\ &= f(\hat\eta_k) = f(\hat\eta_{\tilde k})\\ &= f\circ\hat g(\hat\xi_{\tilde k}) = h(\hat\xi_{\tilde k}).\end{aligned}$$

(iv) Fix any $J \in \mathsf{N}^{\uparrow d}$ and a permutation p of N_d. Using the symmetry of f, we get as in (ii)

$$\begin{aligned}(\eta \circ p)_J &= \eta_{p\circ J} = f(\hat{\xi}_{p\circ J})\\ &= f\Big(\xi_{(p\circ J)\circ I};\, I \subset \mathsf{N}_d\Big)\\ &= f\Big(\xi_{p\circ (J\circ I)};\, I \subset \mathsf{N}_d\Big)\\ &= f\Big\{(\xi \circ p)_{J\circ I};\, I \subset \mathsf{N}_d\Big\}\\ &= f\Big\{(\xi\circ p)\hat{}_J\Big\}.\end{aligned}$$

Thus, $\eta \circ p = g(\xi \circ p)$ for a suitable function g, and so the exchangeability of ξ carries over to η. □

Proposition 28.12 *(inversion) Let ξ be a contractable array on $\tilde{\mathsf{N}}^d$ with independent entries, and fix a measurable function f, such that the array*

$$\eta_J = f(\hat{\xi}_J), \quad J \in \tilde{\mathsf{N}}^d, \tag{15}$$

has independent entries. Then

(i) *there exist some measurable functions g, h, such that for any U-array $\vartheta \perp\!\!\!\perp \xi$ on $\tilde{\mathsf{N}}^d$, the variables*

$$\zeta_J = g(\hat{\xi}_J, \vartheta_J), \quad J \in \tilde{\mathsf{N}}^d, \tag{16}$$

form a U-array $\zeta \perp\!\!\!\perp \eta$ satisfying

$$\xi_J = h(\hat{\eta}_J, \hat{\zeta}_J) \;\; a.s., \quad J \in \tilde{\mathsf{N}}^d, \tag{17}$$

(ii) *when f is symmetric, we can choose g, h with the same property.*

Proof: (i) Write ξ^d, $\hat{\xi}^d$ for the restrictions of ξ to $\mathsf{N}^{\uparrow d}$ and $\tilde{\mathsf{N}}^d$, respectively, and similarly for η, ζ, ϑ. For $d = 0$, Lemma 28.8 applied to the pair $(\xi_\emptyset, \eta_\emptyset)$ yields some functions g, h, such that when $\vartheta_\emptyset \perp\!\!\!\perp \xi$ is $U(0,1)$, the variable $\zeta_\emptyset = g(\xi_\emptyset, \vartheta_\emptyset)$ becomes $U(0,1)$ with $\zeta_\emptyset \perp\!\!\!\perp \eta_\emptyset$ and $\xi_\emptyset = h(\eta_\emptyset, \zeta_\emptyset)$ a.s.

Proceeding by recursion on d, let g, h be such that (16) defines a U-array $\hat{\zeta}^{d-1} \perp\!\!\!\perp \hat{\eta}^{d-1}$ on $\tilde{\mathsf{N}}^{d-1}$ satisfying (17) on $\tilde{\mathsf{N}}^{d-1}$. To extend the construction to dimension d, we note that

$$\xi_J \underset{\hat{\xi}'_J}{\perp\!\!\!\perp} (\hat{\xi}^d \setminus \xi_J), \quad J \in \mathsf{N}^{\uparrow d},$$

since ξ has independent entries. Then (15) yields

$$\xi_J \underset{\hat{\xi}'_J,\, \eta_J}{\perp\!\!\!\perp} \Big(\hat{\xi}^d \setminus \xi_J,\, \hat{\eta}^d\Big), \quad J \in \mathsf{N}^{\uparrow d},$$

and by contractability we see that $(\hat{\xi}_J, \eta_J) = (\xi_J, \hat{\xi}'_J, \eta_J)$ has the same distribution for all $J \in \mathsf{N}^{\uparrow d}$. Hence, Lemma 28.8 yields some measurable functions g_d, h_d, such that the random variables

$$\zeta_J = g_d\Big(\hat{\xi}_J, \eta_J, \vartheta_J\Big), \quad J \in \mathsf{N}^{\uparrow d}, \tag{18}$$

form a U-array ζ^d on $\mathsf{N}^{\uparrow d}$ satisfying

$$\zeta^d \perp\!\!\!\perp \Big(\hat{\xi}^{d-1},\, \eta^d\Big), \tag{19}$$

$$\xi_J = h_d\left(\hat{\xi}'_J, \eta_J, \zeta_J\right) \text{ a.s.,} \quad J \in \mathsf{N}^{\uparrow d}. \tag{20}$$

Inserting (15) into (18), and (17) with $J \in \tilde{\mathsf{N}}^{d-1}$ into (20), we obtain

$$\begin{aligned} \zeta_J &= g_d\left\{\hat{\xi}_J, f(\hat{\xi}_J), \vartheta_J\right\}, \\ \xi_J &= h_d\left\{\hat{h}'(\hat{\eta}'_J, \hat{\zeta}'_J),\ \eta_J,\ \zeta_J\right\} \text{ a.s.,} \quad J \in \mathsf{N}^{\uparrow d}, \end{aligned}$$

for a measurable function $\hat{h}'$. Thus, (16) and (17) remain valid on $\tilde{\mathsf{N}}^d$ for suitable extensions of g, h.

To complete the recursion, we need to show that $\hat{\zeta}^{d-1}$, ζ^d, $\hat{\eta}^d$ are independent. Then note that $\hat{\vartheta}^{d-1} \perp\!\!\!\perp (\hat{\xi}^d, \vartheta^d)$, since ϑ is a U-array independent of ξ. Hence, (15) and (16) yield

$$\hat{\vartheta}^{d-1} \perp\!\!\!\perp \left(\hat{\xi}^{d-1}, \eta^d, \zeta^d\right).$$

Combining this with (19) gives

$$\zeta^d \perp\!\!\!\perp \left(\hat{\xi}^{d-1}, \eta^d, \hat{\vartheta}^{d-1}\right),$$

and so by (15) and (16) we have $\zeta^d \perp\!\!\!\perp (\hat{\eta}^d, \hat{\zeta}^{d-1})$.

Now only $\hat{\eta}^d \perp\!\!\!\perp \hat{\zeta}^{d-1}$ remains to be proved. Since ξ and η are related by (15) and have independent entries, Lemma 28.7 (ii) yields $\eta^d \perp\!\!\!\perp \hat{\xi}^{d-1}$, which extends to $\eta^d \perp\!\!\!\perp (\hat{\xi}^{d-1}, \hat{\vartheta}^{d-1})$ since $\vartheta \perp\!\!\!\perp (\xi, \eta)$. Hence, $\eta^d \perp\!\!\!\perp (\hat{\eta}^{d-1}, \hat{\zeta}^{d-1})$ by (15) and (16). Since also $\hat{\eta}^{d-1} \perp\!\!\!\perp \hat{\zeta}^{d-1}$ by the induction hypothesis, we have indeed $\hat{\eta}^d \perp\!\!\!\perp \hat{\zeta}^{d-1}$.

(ii) Here a similar recursive argument applies, except that now we need to invoke Lemmas 28.9 and 28.11 in each step, to ensure the desired symmetry of g, h. We omit the details. □

For the next result, we say that an array is *representable*, if it can be represented as in Theorem 28.1.

Lemma 28.13 *(augmentation) Suppose that all contractable arrays on $\tilde{\mathsf{N}}^d$ are representable, and let (X, ξ) be a contractable array on $\tilde{\mathsf{N}}^d$, where ξ has independent entries. Then there exist a measurable function f and a U-array $\eta \perp\!\!\!\perp \xi$ on $\tilde{\mathsf{N}}^d$, such that*

$$X_J = f\left(\hat{\xi}_J, \hat{\eta}_J\right) \text{ a.s.,} \quad J \in \tilde{\mathsf{N}}^d.$$

Proof: Since (X, ξ) is representable, there exist some measurable functions g, h and a U-array ζ, such that a.s.

$$\begin{aligned} X_J &= g(\hat{\zeta}_J), \\ \xi_J &= h(\hat{\zeta}_J), \quad J \in \tilde{\mathsf{N}}^d. \end{aligned} \tag{21}$$

Then Lemma 28.12 yields a function k and a U-array $\eta \perp\!\!\!\perp \xi$, such that

$$\zeta_J = k\left(\hat{\xi}_J, \hat{\eta}_J\right) \text{ a.s.,} \quad J \in \tilde{\mathsf{N}}^d.$$

Inserting this into (21) gives the desired representation of X with $f = g \circ \hat{k}$. □

— — —

We are now ready to prove the main results, beginning with Theorem 28.1 (i). Here our proof is based on the following lemma.

Lemma 28.14 *(recursion) Suppose that all exchangeable arrays on* $\hat{\mathsf{N}}^{(d-1)}$ *are representable. Then any exchangeable array* X *on* $\hat{\mathsf{N}}^{(d)}$ *can be represented on* $\hat{\mathsf{N}}^{(d-1)}$ *in terms of a U-array* ξ *on* $\tilde{\mathsf{N}}^{d-1}$*, such that the pair* (X,ξ) *is exchangeable and the arrays* $\tilde{X}_k = (X_h;\, h \sim k)$ *satisfy*

$$\tilde{X}_k \underset{\hat{\xi}_k}{\perp\!\!\!\perp} \Big(X \setminus \tilde{X}_k,\, \xi\Big), \quad k \in \mathsf{N}^{(d)}. \tag{22}$$

Proof: Let $\bar{X}$ be a stationary extension of X to $\hat{\mathsf{Z}}^{(d)}$. For $r = (r_1, \dots, r_m) \in \hat{\mathsf{Z}}^{(d)}$, let r_+ be the sub-sequence of elements $r_j > 0$. For $k \in \hat{\mathsf{N}}^{(d-1)}$ and $r \in \hat{\mathsf{Z}}^{(d)}$ with $r_+ \sim (1, \dots, |k|)$, write $k \circ r = (k_{r_1}, \dots, k_{r_m})$, where $k_{r_j} = r_j$ when $r_j \le 0$. On $\hat{\mathsf{N}}^{(d-1)}$, we introduce the array

$$Y_k = \Big\{\bar{X}_{k\circ r};\, r \in \hat{\mathsf{Z}}^{(d)},\, r_+ \sim (1, \dots, |k|)\Big\}, \quad k \in \hat{\mathsf{N}}^{(d-1)},$$

so that informally $Y_k = \{\bar{X}_h;\, h_+ \sim k\}$ with a specified order of enumeration. The exchangeability of $\bar{X}$ extends to the pair (X, Y).

For $k = (k_1, \dots, k_d) \in \mathsf{N}^{(d)}$, write $\hat{Y}_k$ for the restriction of Y to sequences in $\tilde{k}$ of length $< d$, and note that

$$(\tilde{X}_k, \hat{Y}_k) \overset{d}{=} \Big(\tilde{X}_k,\, X \setminus \tilde{X}_k,\, Y\Big), \quad k \in \mathsf{N}^{(d)},$$

since the two sides are restrictions of the same exchangeable array $\bar{X}$ to sequences in $\mathsf{Z}_- \cup \{\tilde{k}\}$ and Z, respectively. Since also $\sigma(\hat{Y}_k) \subset \sigma(X \setminus \tilde{X}_k,\, Y)$, Lemma 8.10 yields

$$\tilde{X}_k \underset{\hat{Y}_k}{\perp\!\!\!\perp} \Big(X \setminus \tilde{X}_k,\, Y\Big), \quad k \in \mathsf{N}^{(d)}. \tag{23}$$

Since exchangeable arrays on $\hat{\mathsf{N}}^{(d-1)}$ are representable, there exist a U-array ξ on $\tilde{\mathsf{N}}^{d-1}$ and a measurable function f, such that

$$(X_k, Y_k) = f(\hat{\xi}_k), \quad k \in \hat{\mathsf{N}}^{(d-1)}. \tag{24}$$

By Theorems 8.17 and 8.20 we may choose $\xi \perp\!\!\!\perp_{X',Y} X$, where X' denotes the restriction of X to $\hat{\mathsf{N}}^{(d-1)}$, in which case (X,ξ) is again exchangeable by Lemma 28.4. The same conditional independence yields

$$\tilde{X}_k \underset{Y,\, X \setminus \tilde{X}_k}{\perp\!\!\!\perp} \xi, \quad k \in \mathsf{N}^{(d)}. \tag{25}$$

Using Theorem 8.12, we may combine (23) and (25) into

$$\tilde{X}_k \underset{\hat{Y}_k}{\perp\!\!\!\perp} \Big(X \setminus \tilde{X}_k,\, \xi\Big), \quad k \in \mathsf{N}^{(d)},$$

and (22) follows since $\hat{Y}_k$ is $\hat{\xi}_k$-measurable by (24). □

Proof of Theorem 28.1 (i): The result for $d = 0$ is elementary and follows from Lemma 28.8 with $\eta = 0$. Proceeding by induction, suppose that all exchangeable arrays on $\hat{\mathsf{N}}^{(d-1)}$ are representable, and consider an exchangeable array X on $\hat{\mathsf{N}}^{(d)}$. By Lemma 28.14, we may choose a U-array ξ on $\tilde{\mathsf{N}}^{d-1}$ and a measurable function f satisfying

$$X_k = f(\hat{\xi}_k) \;\text{ a.s.}, \quad k \in \hat{\mathsf{N}}^{(d-1)}, \tag{26}$$

and such that (X, ξ) is exchangeable and the arrays $\tilde{X}_k = (X_h;\, h \sim k)$ with $k \in \mathsf{N}^{(d)}$ satisfy (22).

Letting $\mathcal{P}_d$ be the group of permutations on N_d, we may write

$$\begin{aligned}\tilde{X}_k &= \Big(X_{k\circ p};\, p \in \mathcal{P}_d\Big),\\ \hat{\xi}_k &= \Big(\xi_{k\circ I};\, I \subset \mathsf{N}_d\Big), \qquad k \in \mathsf{N}^{(d)},\end{aligned}$$

where $k \circ I = (k_i;\, i \in I)$. The pairs $(\tilde{X}_k, \hat{\xi}_k)$ are equally distributed by the exchangeability of (X, ξ), and in particular

$$\Big(\tilde{X}_{k\circ p}, \hat{\xi}_{k\circ p}\Big) \stackrel{d}{=} (\tilde{X}_k, \hat{\xi}_k), \quad p \in \mathcal{P}_d,\ k \in \mathsf{N}^{(d)}.$$

Since ξ is a U-array, the arrays $\hat{\xi}_{k\circ p}$ are a.s. different for fixed k, and so the associated symmetry groups are a.s. trivial. Hence, Lemma 28.9 yields some $U(0,1)$ random variables $\zeta_k \perp\!\!\!\perp \hat{\xi}_k$ and a measurable function h, such that

$$X_{k\circ p} = h\Big(\hat{\xi}_{k\circ p}, \zeta_k\Big) \text{ a.s.}, \quad p \in \mathcal{P}_d,\ k \in \mathsf{N}^{(d)}. \tag{27}$$

Now introduce a U-array $\vartheta \perp\!\!\!\perp \xi$ on $\mathsf{N}^{\uparrow d}$, and define

$$\begin{aligned}Y_k &= h(\hat{\xi}_k, \vartheta_{\tilde{k}}),\\ \tilde{Y}_k &= \Big(Y_{k\circ p};\, p \in \mathcal{P}_d\Big), \qquad k \in \mathsf{N}^{(d)}.\end{aligned} \tag{28}$$

Comparing with (27) and using the independence properties of (ξ, ϑ), we get

$$\begin{aligned}(\tilde{Y}_k, \hat{\xi}_k) &\stackrel{d}{=} (\tilde{X}_k, \hat{\xi}_k),\\ \tilde{Y}_k &\underset{\hat{\xi}_k}{\perp\!\!\!\perp} \Big(Y \setminus \tilde{Y}_k,\, \xi\Big), \qquad k \in \mathsf{N}^{(d)}.\end{aligned}$$

By (22) it follows that $(\tilde{Y}_k, \xi) \stackrel{d}{=} (\tilde{X}_k, \xi)$ for all $k \in \mathsf{N}^{(d)}$, and moreover

$$\begin{aligned}\tilde{X}_k &\perp\!\!\!\perp_\xi (X \setminus \tilde{X}_k),\\ \tilde{Y}_k &\perp\!\!\!\perp_\xi (Y \setminus \tilde{Y}_k), \qquad k \in \mathsf{N}^{(d)}.\end{aligned}$$

Thus, the arrays $\tilde{X}_k$ with different $\tilde{k}$ are conditionally independent given ξ, and similarly for the arrays $\tilde{Y}_k$. Letting X' be the restriction of X to $\hat{\mathsf{N}}^{d-1}$, which is ξ-measurable by (26), we obtain $(X, \xi) \stackrel{d}{=} (X', Y, \xi)$. Using (28) along with Corollary 8.18, we conclude that

$$X_k = h(\hat{\xi}_k, \eta_k) \text{ a.s.}, \quad k \in \mathsf{N}^{(d)},$$

for a U-array $\eta \perp\!\!\!\perp \xi$ on $\mathsf{N}^{\uparrow d}$. Combining this with (26) yields the desired representation of X. □

— — —

Next we prove the equivalence criteria for jointly exchangeable arrays.

Proof of Theorem 28.3, (i) ⇔ (ii): The implication (i) ⇒ (ii) is obvious, and the converse follows from Theorem 8.17.

(iv) ⇒ (i): We may write condition (iv) as

$$f\Big(\alpha,\xi_i,\xi_j,\zeta_{ij}\Big) = g\Big(\alpha'',\xi_i'',\xi_j'',\zeta_{ij}''\Big) \text{ a.s.,} \quad i\neq j, \tag{29}$$

where

$$\alpha'' = T(\bar\alpha), \qquad \xi_i'' = U(\bar\alpha,\bar\xi_i), \qquad \zeta_{ij}'' = V\Big(\bar\alpha,\bar\xi_i,\bar\xi_j,\bar\zeta_{ij}\Big),$$

with

$$\bar\alpha = (\alpha,\alpha'), \quad \bar\xi_i = (\xi_i,\xi_i'), \quad \bar\zeta_{ij} = (\zeta_{ij},\zeta_{ij}').$$

The symmetry of V and $\bar\zeta$ yields

$$\begin{aligned}\zeta_{ij}'' &= V\Big(\bar\alpha,\bar\xi_i,\bar\xi_j,\bar\zeta_{ij}\Big)\\ &= V\Big(\bar\alpha,\bar\xi_j,\bar\xi_i,\bar\zeta_{ji}\Big) = \zeta_{ji}'', \quad i\neq j.\end{aligned}$$

Using the mapping property $\lambda^2\mapsto\lambda$ of the functions T, U, V, and conditioning first on the pair $\bar\alpha$ and then on the array $(\bar\alpha,\bar\xi)$, we see that the variables $\alpha'',\xi_i'',\zeta_{ij}''$ are again i.i.d. $U(0,1)$. Condition (i) now follows from (29).

(i) ⇒ (iii): Assume (i), so that

$$\begin{aligned}X_{ij} &= f\Big(\alpha,\xi_i,\xi_j,\zeta_{ij}\Big)\\ &= g\Big(\alpha',\xi_i',\xi_j',\zeta_{ij}'\Big) \text{ a.s.,} \quad i\neq j,\end{aligned} \tag{30}$$

for some U-arrays (α,ξ,ζ) and (α',ξ',ζ') on $\tilde{\mathsf{N}}^2$. Since the latter may be chosen to be conditionally independent given X, the combined array $(\bar\alpha,\bar\xi,\bar\zeta)$ is again jointly exchangeable by Lemma 28.4 (i). Hence, it may be represented as in Theorem 28.1 (i) in terms of a U-array (α'',ξ'',ζ'') on $\tilde{\mathsf{N}}^2$, so that a.s. for all $i\neq j$

$$\bar\alpha = \bar T(\alpha''), \qquad \bar\xi_i = \bar U(\alpha'',\xi_i''), \qquad \bar\zeta_{ij} = \bar V\Big(\alpha'',\xi_i'',\xi_j'',\zeta_{ij}''\Big),$$

for some functions $\bar T = (T,T')$, $\bar U = (U,U')$, and $\bar V = (V,V')$ between suitable spaces. Since $\bar\zeta_{ij} = \bar\zeta_{ji}$, we may choose $\bar V$ to be symmetric, and by Lemma 28.7 (ii) we may choose T, U,V and T', U',V' to preserve λ in the highest order arguments. Now (iii) follows by substitution into (30).

(iii) ⇒ (iv): Condition (iii) may be written as

$$f\Big(\alpha',\xi_i',\xi_j',\zeta_{ij}'\Big) = g\Big(\alpha'',\xi_i'',\xi_j'',\zeta_{ij}''\Big) \text{ a.s.,} \quad i\neq j, \tag{31}$$

where

$$\alpha' = T'(\alpha), \quad \xi_i' = U'(\alpha,\xi_i), \quad \zeta_{ij}' = V'(\alpha,\xi_i,\xi_j,\zeta_{ij}), \tag{32}$$

$$\alpha'' = T''(\alpha), \quad \xi_i'' = U''(\alpha,\xi_i), \quad \zeta_{ij}'' = V''(\alpha,\xi_i,\xi_j,\zeta_{ij}). \tag{33}$$

By the measure-preserving property of T', U',V' and T'', U'',V'' along with Lemma 28.7, the triples (α',ξ',ζ') and (α'',ξ'',ζ'') are again U-arrays on $\tilde{\mathsf{N}}^2$.

By Lemma 28.12 we may solve for (α,ξ,ζ) in (32) to obtain

$$\alpha = T(\bar\alpha'), \quad \xi_i = U(\bar\alpha',\bar\xi_i'), \quad \zeta_{ij} = V\Big(\bar\alpha',\bar\xi_i',\bar\xi_j',\bar\zeta_{ij}'\Big), \tag{34}$$

for some measurable functions T, U,V between suitable spaces, where

$$\bar\alpha' = (\alpha',\beta), \qquad \bar\xi_i' = (\xi_i',\eta_i), \qquad \bar\zeta_{ij}' = (\zeta_{ij}',\vartheta_{ij}),$$

for an independent U-array (β,η,ϑ). Substituting (34) into (33) and using (31), we get

$$f\Big(\alpha', \xi'_i, \xi'_j, \zeta'_{ij}\Big) = g\Big\{\,\bar{T}(\bar{\alpha}'),\, \bar{U}(\bar{\alpha}', \bar{\xi}'_i),\, \bar{U}(\bar{\alpha}', \bar{\xi}'_j),\, \bar{V}\Big(\bar{\alpha}', \bar{\xi}'_i, \bar{\xi}'_j, \bar{\zeta}'_{ij}\Big)\Big\},$$

where $\bar{T}$, $\bar{U}, \bar{V}$ result from composition of (T, U, V) into (T'', U'', V''). By Lemma 28.7 (ii) we can modify $\bar{T}$, $\bar{U}, \bar{V}$ to achieve the desired mapping property $\lambda^2 \mapsto \lambda$, and by Lemma 28.12 (ii) we can choose V to be symmetric, so that $\bar{V}$ becomes symmetric as well, by Lemma 28.11 (iii). □

— — —

Next we prove Theorem 28.1 (ii) by establishing a representation[5] as in (i). Here the following construction plays a key role for the main proof. Define $\tilde{T}' = \tilde{T} \setminus \{\emptyset\}$ and $\mathsf{Z}_- = -\mathsf{Z}_+ = \mathsf{Z} \setminus \mathsf{N}$.

Lemma 28.15 *(key construction) Let X be a contractable array on $\tilde{\mathsf{Z}}$, and define for $J \in \tilde{\mathsf{N}}$ and $n \in \mathsf{N}$*

$$Y_J = \Big(X_{I \cup J};\, I \in \tilde{\mathsf{Z}}'_-\Big), \qquad \begin{aligned} X^n_J &= X_{\{n\} \cup (J+n)}, \\ Y^n_J &= \Big(Y_{J+n}, Y_{\{n\} \cup (J+n)}\Big). \end{aligned}$$

Then the pairs (X, Y) and (X^n, Y^n) are contractable on $\tilde{\mathsf{N}}$, and the latter are equally distributed with

$$X^n \underset{Y^n}{\perp\!\!\!\perp} (X \setminus X^n), \quad n \in \mathsf{N}. \tag{35}$$

Proof: It is enough to prove (35), the remaining assertions being obvious. Then write $X = (Q^n, R^n)$ for all $n \in \mathsf{N}$, where Q^n denotes the restriction of X to $\tilde{\mathsf{N}} + n - 1$ and $R^n = X \setminus Q^n$. Defining

$$p_n(k) = k - (n-1)\, 1\{k < n\}, \quad k \in \mathsf{Z},\;\; n \in \mathsf{N},$$

and using the contractability of X, we obtain

$$\begin{aligned} (Q^n, R^n) &= X \overset{d}{=} X \circ p_n \\ &= (Q^n,\ R^n \circ p_n), \end{aligned}$$

and so Lemma 8.10 yields $R^n \perp\!\!\!\perp_{R^n \circ p_n} Q^n$, which amounts to

$$\Big(Y, X^1, \ldots, X^{n-1}\Big) \underset{Y^n}{\perp\!\!\!\perp} \Big(X^n, X^{n+1}, \ldots; X_\emptyset\Big).$$

Replacing n by $n+1$ and noting that $Y^{n+1} \subset Y^n$, we see that also

$$\Big(Y, X^1, \ldots, X^n\Big) \underset{Y^n}{\perp\!\!\!\perp} \Big(X^{n+1}, X^{n+2}, \ldots; X_\emptyset\Big).$$

Combining the last two relations, we get by Lemma 4.8

$$X^n \underset{Y^n}{\perp\!\!\!\perp} \Big(Y, X^1, \ldots, X^{n-1}, X^{n+1}, \ldots; X_\emptyset\Big) = X \setminus X^n.$$ □

[5] No direct proof of the extension property is known.

Lemma 28.16 *(recursion) Suppose that all contractable arrays on* $\tilde{\mathsf{N}}^{d-1}$ *are representable, and let* X *be a contractable array on* $\tilde{\mathsf{N}}^d$*. Then there exists a U-array* η *on* $\tilde{\mathsf{N}}^{d-1}$*, such that* (X,η) *is contractable with* $\sigma(X_\emptyset)\subset\sigma(\eta_\emptyset)$*, and the arrays*

$$X_J^n = X_{\{n\}\cup(J+n)}, \quad \eta_J^n = \eta_{J+n}, \qquad J\in\tilde{\mathsf{N}}^{d-1},$$

satisfy

$$X^n \underset{\eta^n}{\perp\!\!\!\perp} \big(X\setminus X^n,\,\eta\big), \quad n\in\mathsf{N}. \tag{36}$$

Proof: Defining Y as in Lemma 28.15 in terms of a contractable extension of X to $\tilde{\mathsf{Z}}$, we note that the pair $(X_\emptyset, Y)$ is contractable on $\tilde{\mathsf{N}}^{d-1}$. Then by hypothesis it has a representation

$$X_\emptyset = f(\eta_\emptyset); \qquad Y_J = g(\hat\eta_J), \quad J\in\tilde{\mathsf{N}}^{d-1}, \tag{37}$$

for some measurable functions f,g and a U-array η on $\tilde{\mathsf{N}}^{d-1}$. By Theorems 8.17 and 8.21, we may extend the pairs (X,Y) and (Y,η) to contractable arrays on $\tilde{\mathsf{Q}}$. Choosing

$$\eta \underset{X_\emptyset,\,Y}{\perp\!\!\!\perp} X, \tag{38}$$

we see that (X,Y,η) becomes contractable on $\tilde{\mathsf{Q}}$ by Lemma 28.4.

By Theorem 8.12 and Lemma 28.15, we have

$$X^n \underset{X_\emptyset,\,Y}{\perp\!\!\!\perp} (X\setminus X^n), \quad n\in\mathsf{N}.$$

Combining with (38) and using a conditional version of Lemma 4.8, we get

$$X^n \underset{X_\emptyset,\,Y}{\perp\!\!\!\perp} \big(X\setminus X^n,\,\eta\big), \quad n\in\mathsf{N},$$

and since $X_\emptyset$ and Y are η-measurable by (37), it follows that

$$X^n \underset{\eta}{\perp\!\!\!\perp} \big(X\setminus X^n,\,\eta\big), \quad n\in\mathsf{N}. \tag{39}$$

By the contractability of (X,η), we may replace $\tilde{\mathsf{Q}}$ by the index sets $\tilde T_\varepsilon$ with $T_\varepsilon = \bigcup_{k>0}(k-\varepsilon,k]$, $\varepsilon>0$. Letting $\varepsilon\to 0$, we see from Corollary 9.26 that (39) remains true with η restricted to $\tilde{\mathsf{N}}^{d-1}$. Since also $X^n \perp\!\!\!\perp_{\eta^n} \eta$ by Lemma 28.5, (36) follows by Theorem 8.12. □

Proof of Theorem 28.1 (ii): For $d=0$, the result is elementary and follows from Lemma 28.8 with $\eta=0$. Now suppose that all contractable arrays on $\tilde{\mathsf{N}}^{d-1}$ are representable, and let X be a contractable array on $\tilde{\mathsf{N}}^d$. Choose a U-array η on $\tilde{\mathsf{N}}^{d-1}$ as in Lemma 28.16, and extend η to $\tilde{\mathsf{N}}^d$ by attaching some independent $U(0,1)$ variables η_J with $|J|=d$. Then (X,η) remains contractable, the element $X_\emptyset$ is $\eta_\emptyset$-measurable, and the arrays[6]

$$\begin{aligned} X_J^n &= X_{\{n\}\cup(J+n)},\\ \eta_J^n &= \big(\eta_{J+n},\,\eta_{\{n\}\cup(J+n)}\big), \qquad J\in\tilde{\mathsf{N}}^{d-1},\ n\in\mathsf{N}, \end{aligned}$$

satisfy

$$X^n \underset{\eta^n}{\perp\!\!\!\perp} \big(X\setminus X^n,\,\eta\big), \quad n\in\mathsf{N}. \tag{40}$$

[6]Note that these definitions differ slightly from those in Lemma 28.16.

The pairs (X^n, η^n) are equally distributed and inherit the contractability from (X, η). Furthermore, the elements of η^n are independent for fixed n and uniformly distributed on $[0,1]^2$. Hence, Corollary 28.13 yields a measurable function G and some U-arrays $\zeta^n \perp\!\!\!\perp \eta^n$ on $\tilde{\mathsf{N}}^{d-1}$, such that

$$X_J^n = G(\hat{\eta}_J^n, \hat{\zeta}_J^n) \text{ a.s.,} \quad J \in \tilde{\mathsf{N}}^{d-1}, \; n \in \mathsf{N}. \tag{41}$$

By Theorem 8.17, we may choose

$$\zeta^n = h\Big(X^n, \eta^n, \vartheta_n\Big), \quad n \in \mathsf{N}, \tag{42}$$

for a U-sequence $\vartheta = (\vartheta_n) \perp\!\!\!\perp (X, \eta)$ and a measurable function h, so that by Proposition 8.20,

$$\zeta^n \underset{X^n, \eta^n}{\perp\!\!\!\perp} (\eta, \zeta \setminus \zeta^n), \quad n \in \mathsf{N}.$$

Using (40) and (42), along with the independence of the ϑ_n, we get

$$X^n \underset{\eta^n}{\perp\!\!\!\perp} (\eta, \zeta \setminus \zeta^n), \quad n \in \mathsf{N}.$$

Combining the last two relations, we get by Theorem 8.12

$$\zeta^n \underset{\eta^n}{\perp\!\!\!\perp} (\eta, \zeta \setminus \zeta^n), \quad n \in \mathsf{N}.$$

Since also $\zeta^n \perp\!\!\!\perp \eta^n$, the same result yields

$$\zeta^n \perp\!\!\!\perp (\eta, \zeta \setminus \zeta^n), \quad n \in \mathsf{N}.$$

By Lemma 4.8, the ζ^n are then mutually independent and independent of η.

Now combine the arrays ζ^n into a single U-array $\zeta \perp\!\!\!\perp \eta$ on $\tilde{\mathsf{N}}^d \setminus \{\emptyset\}$, given by

$$\zeta_{\{n\} \cup (J+n)} = \zeta_J^n, \quad J \in \tilde{\mathsf{N}}^{d-1}, \; n \in \mathsf{N},$$

and extend ζ to $\tilde{\mathsf{N}}^d$ by attaching an independent $U(0,1)$ variable $\zeta_\emptyset$. Then (41) becomes

$$X_J = F(\hat{\eta}_J, \hat{\zeta}_J) \text{ a.s.,} \quad J \in \tilde{\mathsf{N}}^d \setminus \{\emptyset\}, \tag{43}$$

where F is given for $k = 1, \dots, d$ by

$$F\Big\{(y,z)_I;\, I \subset \mathsf{N}_k\Big\} = G\Big\{y_{I+1}, (y,z)_{\{1\} \cup (I+1)};\, I \subset \mathsf{N}_{k-1}\Big\}.$$

Since $X_\emptyset$ is $\eta_\emptyset$-measurable, we may extend (43) to $J = \emptyset$, through a suitable extension of F to $[0,1]^2$, as in Lemma 1.14. By Lemma 28.8 with $\eta = 0$, we may finally choose a U-array ξ on $\tilde{\mathsf{N}}^d$ and a measurable function $b\colon [0,1] \to [0,1]^2$, such that

$$(\eta_J, \zeta_J) = b(\xi_J) \text{ a.s.,} \quad J \in \tilde{\mathsf{N}}^d.$$

Then $(\hat{\eta}_J, \hat{\zeta}_J) = \hat{b}(\hat{\xi}_J)$ a.s. for all $J \in \tilde{\mathsf{N}}^d$, and a substitution into (43) yields the desired representation $X_J = f(\hat{\xi}_J)$ a.s. with $f = F \circ \hat{b}$. □

— — —

We turn to a more detailed study of separately exchangeable arrays X on N^2. Define the shell σ-field of X as in Lemma 28.6. Say that X is *dissociated* if

$$\Big(X_{ij};\, i \le m,\, j \le n\Big) \perp\!\!\!\perp \Big(X_{ij};\, i > m,\, j > n\Big), \quad m, n \in \mathsf{N}.$$

Theorem 28.17 *(conditioning and independence) Let X be the separately exchangeable array on N^2 given by (4), write $\mathcal{S}$ for the shell σ-field of X, and let X^- be an exchangeable extension of X to Z_-^2. Then*

(i) $\mathcal{S} \subset \sigma(\alpha, \xi, \eta)$ *a.s.,*

(ii) $X \perp\!\!\!\perp_{\mathcal{S}} (\alpha, \xi, \eta)$,

(iii) *the X_{ij} are conditionally independent given $\mathcal{S}$,*

(iv) *X is i.i.d.* $\Leftrightarrow$ $X \perp\!\!\!\perp (\alpha, \xi, \eta)$,

(v) *X is dissociated* $\Leftrightarrow$ $X \perp\!\!\!\perp \alpha$,

(vi) *X is $\mathcal{S}$-measurable* $\Leftrightarrow$ $X \perp\!\!\!\perp \zeta$,

(vii) $\mathcal{L}(X \mid X^-) = \mathcal{L}(X \mid \alpha)$ *a.s.*

Proof: (i) Use Corollary 9.26.

(ii) Apply Lemma 28.6 to the array X on N^2, extended to Z_+^2 by

$$X_{0,0} = \alpha, \quad X_{i,0} = \xi_i, \quad X_{0,j} = \eta_j, \qquad i, j \in \mathsf{N}.$$

(iii) By Lemma 8.7, the X_{ij} are conditionally independent given (α, ξ, η). Now apply (i)$-$(ii) and Theorem 8.9.

(iv) If X is i.i.d., then $\mathcal{S}$ is trivial by Kolmogorov's 0$-$1 law, and (ii) yields $X \perp\!\!\!\perp (\alpha, \xi, \eta)$. Conversely, $X \perp\!\!\!\perp (\alpha, \xi, \eta)$ implies $\mathcal{L}(X) = \mathcal{L}(X \mid \alpha, \xi, \eta)$ by Lemma 8.6, and so Lemma 8.7 yields $X \overset{d}{=} Y$ with

$$Y_{ij} = g(\zeta_{ij}) = \lambda^3 f(\cdot, \cdot, \cdot, \zeta_{ij}), \quad i, j \in \mathsf{N},$$

which shows that the Y_{ij}, and then also the X_{ij}, are i.i.d.

(v) Choose some disjoint, countable sets $N_1, N_2, \ldots \subset \mathsf{N}$, and write X^n for the restriction of X to N_n^2. If X is dissociated, the sequence (X^n) is i.i.d. Applying Theorem 27.2 to the exchangeable sequence of pairs (α, X^n), we obtain a.s.

$$\begin{aligned} \mathcal{L}(X) &= \mathcal{L}(X^1) \\ &= \mathcal{L}(X^1 | \, \alpha) \\ &= \mathcal{L}(X | \, \alpha), \end{aligned}$$

and so $X \perp\!\!\!\perp \alpha$ by Lemma 8.6. Conversely, $X \perp\!\!\!\perp \alpha$ implies $\mathcal{L}(X) = \mathcal{L}(X \mid \alpha)$ a.s. by the same lemma, and so Lemma 8.7 yields $X \overset{d}{=} Y$ with

$$\begin{aligned} Y_{ij} &= g(\xi_i, \eta_j, \zeta_{ij}) \\ &= \lambda f(\cdot, \xi_i, \eta_j, \zeta_{ij}), \quad i, j \in \mathsf{N}, \end{aligned}$$

which shows that Y, and then also X, is dissociated.

(vi) We may choose X to take values in $[0, 1]$. If X is $\mathcal{S}$-measurable, then by (i)$-$(ii) and Lemma 8.7,

$$\begin{aligned} X &= E(X|\,\mathcal{S}) \\ &= E(X|\,\alpha,\xi,\eta), \end{aligned}$$

and so X is (α,ξ,η)-measurable, which implies $X \perp\!\!\!\perp \zeta$. Conversely, $X \perp\!\!\!\perp \zeta$ implies $\mathcal{L}(X) = \mathcal{L}(X \mid \zeta)$ a.s. by Lemma 8.6, and so Lemma 8.7 yields $X \overset{d}{=} Y$ with

$$\begin{aligned} Y_{ij} &= g(\alpha,\xi_i,\eta_j) \\ &= \lambda f(\alpha,\xi_i,\eta_j,\cdot), \quad i,j \in \mathsf{N}, \end{aligned}$$

which shows that Y is (α,ξ,η)-measurable. Applying (i)−(ii) to Y yields

$$\begin{aligned} Y &= E(Y|\,\alpha,\xi,\eta) \\ &= E(Y|\,\mathcal{S}_Y), \end{aligned}$$

which shows that Y is $\mathcal{S}_Y$-measurable. Hence, $X \overset{d}{=} Y$ is $\mathcal{S}_X$-measurable.

(vii) Define the arrays X^n as in (v), and use X^- to extend the sequence (X^n) to Z_-, so that the entire sequence is exchangeable on Z. Since X^- is invariant under permutations on N, the exchangeability on N is preserved by conditioning on X^-, and the X^n with $n > 0$ are conditionally i.i.d. by the law of large numbers. This remains true under conditioning on α in the sequence of pairs (X^n,α). By the uniqueness in Theorem 27.2, the two conditional distributions agree a.s., and so by exchangeability

$$\begin{aligned} \mathcal{L}(X|\,X^-) &= \mathcal{L}(X^1|\,X^-) \\ &= \mathcal{L}(X^1|\,\alpha) \\ &= \mathcal{L}(X|\,\alpha). \end{aligned}$$ □

Corollary 28.18 *(representations in special cases, Aldous) Let X be a separately exchangeable array on N^2 with shell σ-field $\mathcal{S}$. Then*

(i) *X is dissociated iff it has an a.s. representation*

$$X_{ij} = f(\xi_i,\eta_j,\zeta_{ij}), \quad i,j \in \mathsf{N},$$

(ii) *X is $\mathcal{S}$-measurable iff it has an a.s. representation*

$$X_{ij} = f(\alpha,\xi_i,\eta_j), \quad i,j \in \mathsf{N},$$

(iii) *X is dissociated and $\mathcal{S}$-measurable iff it has an a.s. representation*

$$X_{ij} = f(\xi_i,\eta_j), \quad i,j \in \mathsf{N}.$$

Proof: (i) Use Theorem 28.17 (v).

(ii) Proceed as in the proof of Theorem 28.17 (iv).

(iii) Combine (i) and (ii). □

− − −

We turn to the subject of rotational symmetries. Given a real-valued random array $X = (X_{ij})$ indexed by N^2, we introduce the rotated version

$$Y_{ij} = \sum_{h,k} U_{ih} V_{jk}\, X_{hk}, \quad i, j \in \mathsf{N},$$

where U, V are unitary arrays on N^2 affecting only finitely many coordinates, hence orthogonal matrices of finite order n, extended to $i \vee j > n$ by $U_{ij} = V_{ij} = \delta_{ij}$. Writing $Y = (U \otimes V)X$ for the mapping $X \mapsto Y$, we say that X is *separately rotatable* if $(U \otimes V)X \overset{d}{=} X$ for all U, V, and *jointly rotatable* if the same condition holds with $U = V$.

The representations of separately and jointly rotatable arrays may be stated in terms of *G-arrays*, defined as arrays of independent $N(0,1)$ random variables.

Theorem 28.19 *(rotatable arrays, Aldous, OK) Let $X = (X_{ij})$ be a real-valued random array on N^2. Then*

(i) *X is separately rotatable iff a.s.*

$$X_{ij} = \sigma\, \zeta_{ij} + \sum_k \alpha_k\, \xi_{ki}\, \eta_{kj}, \quad i, j \in \mathsf{N},$$

for some independent G-arrays $(\xi_{ki}), (\eta_{kj}), (\zeta_{ij})$ and an independent set[7] *of random variables σ, $\alpha_1, \alpha_2, \ldots$ with $\sum_k \alpha_k^2 < \infty$ a.s.,*

(ii) *X is jointly rotatable iff a.s.*

$$X_{ij} = \rho\, \delta_{ij} + \sigma\, \zeta_{ij} + \sigma' \zeta_{ji} + \sum_{h,k} \alpha_{hk} \Big(\xi_{hi}\, \xi_{kj} - \delta_{ij}\, \delta_{hk} \Big), \quad i, j \in \mathsf{N},$$

for some independent G-arrays $(\xi_{hi}), (\zeta_{ij})$ and an independent set[8] *of random variables ρ, σ, σ', and α_{hk} with $\sum_{h,k} \alpha_{hk}^2 < \infty$ a.s.*

In case (i), we note the remarkable symmetry $(X_{ij}) \overset{d}{=} (X_{ji})$. The centering term $\delta_{ij}\, \delta_{hk}$ in (ii) is needed to ensure convergence of the double sum[9]. The simpler representation in (i) is essentially a special case. Here we prove only part (i), which still requires several lemmas.

Lemma 28.20 *(conditional moments) Let X be a separately rotatable array on N^2 with rotatable extension X^- to Z_-^2. Then*

$$E\Big(|X_{ij}|^p \,\Big|\, X^- \Big) < \infty \ \text{ a.s.}, \quad p > 0, \ i, j \in \mathsf{N}.$$

Proof: For fixed $i \in \mathsf{Z}$, we have $X_{ij} = \sigma_i\, \zeta_{ij}$ for all j by Theorem 14.3, where $\sigma_i \geq 0$ and the ζ_{ij} are independent of σ_i and i.i.d. $N(0,1)$. By the stronger

[7] This means that the arrays $\xi, \eta, \zeta, (\sigma, \alpha)$ are mutually independent; no independence is claimed between the variables $\sigma, \alpha_1, \alpha_2, \ldots$.

[8] Here $\xi, \zeta, (\rho, \sigma, \sigma', \alpha)$ are independent; the dependence between $\rho, \sigma, \sigma', \alpha$ is arbitrary.

[9] Comparing (ii) with Theorem 14.25, we recognize a double Wiener–Itô integral. To see the general pattern, we need to go to higher dimensions.

Theorem 27.15, we have in fact $(\zeta_{ij}) \perp\!\!\!\perp X^-$ for $i \le 0$ and $j > 0$, and since σ_i is X^--measurable by the law of large numbers, we get

$$E\Big(|X_{ij}|^p \,\Big|\, X^- \Big) = \sigma_i^p \, E|\zeta_{ij}|^p < \infty \text{ a.s.}, \quad i \le 0, \; j \in \mathsf{N}.$$

Hence, by rotatability,

$$E\Big(|X_{ij}|^p \,\Big|\, X^- \Big) < \infty \text{ a.s.}, \quad i\,j < 0, \; p > 0. \tag{44}$$

Taking $i = 1$, we put $\zeta_{1j} = \zeta_j$ and $X_{1j} = \sigma_1 \zeta_j = \xi_j$. Writing $E(\,\cdot\,|X^-) = E^{X^-}$ and letting $n \in \mathsf{N}$, we get by Cauchy's inequality

$$\begin{aligned} E^{X^-}|\xi_j|^p &= E^{X^-}\Big(\zeta_j^2 \, \sigma_1^2 \Big)^{p/2} \\ &= E^{X^-}\bigg(\zeta_j^2 \, \frac{\xi_{-1}^2 + \cdots + \xi_{-n}^2}{\zeta_{-1}^2 + \cdots + \zeta_{-n}^2} \bigg)^{p/2} \\ &\le \bigg\{ E^{X^-}\bigg(\frac{\zeta_j^2}{\zeta_{-1}^2 + \cdots + \zeta_{-n}^2} \bigg)^p E^{X^-}\Big(\xi_{-1}^2 + \cdots + \xi_{-n}^2 \Big)^p \bigg\}^{1/2}. \end{aligned}$$

For the second factor on the right, we get a.s. by Minkowski's inequality and (44)

$$E^{X^-}\Big(\xi_{-1}^2 + \cdots + \xi_{-n}^2 \Big)^p \le n^p \, E^{X^-} |\xi_{-1}|^{2p} < \infty.$$

The first factor is a.s. finite for $n > 2p$, since its expected value equals

$$\begin{aligned} E\bigg(\frac{\zeta_j^2}{\zeta_{-1}^2 + \cdots + \zeta_{-n}^2} \bigg)^p &= E|\zeta_j|^{2p} \, E\Big(\zeta_{-1}^2 + \cdots + \zeta_{-n}^2 \Big)^{-p} \\ &\lesssim \int_0^\infty r^{-2p} \, e^{-r^2/2} \, r^{n-1} \, dr < \infty. \end{aligned}$$

□

Lemma 28.21 *(arrays of product type, Aldous) Let X be a separately rotatable array of the form*

$$X_{ij} = f(\xi_i, \eta_j), \quad i, j \in \mathsf{N},$$

for a measurable function f on $[0,1]^2$ and some independent U-sequences (ξ_i) and (η_j). Then a.s.

$$X_{ij} = \textstyle\sum_k c_k \, \varphi_{ki} \, \psi_{kj}, \quad i, j \in \mathsf{N}, \tag{45}$$

for some independent G-arrays (φ_{ki}), (ψ_{kj}) on N^2 and constants c_k with $\sum_k c_k^2 < \infty$.

Proof: Let $\bar{X}$ be a stationary extension of X to Z^2 with restriction X^- to Z_-^2, and note that $X^- \perp\!\!\!\perp X$ by the independence of all ξ_i and η_j. Hence, Lemma 28.20 yields $EX_{ij}^2 < \infty$ for all i, j, so that $f \in L^2(\lambda^2)$. By a standard Hilbert space result[10], there exist some ortho-normal sequences (g_k), (h_k) in $L^2(\lambda)$ and constants $c_k > 0$ with $\sum_k c_k^2 < \infty$, such that

$$f(x, y) = \textstyle\sum_k c_k \, g_k(x) \, h_k(y) \text{ in } L^2, \quad x, y \in [0,1],$$

and (45) follows with

[10]Cf. Lemma A3.1 in K(05), p. 450.

$$\begin{aligned}\varphi_{ki} &= g_k(\xi_i),\\ \psi_{kj} &= h_k(\eta_j), \qquad i,j,k \in \mathbb{N}.\end{aligned}$$

We need to show that the variables $\varphi_k = g_k(\xi)$ and $\psi_k = h_k(\eta)$ are i.i.d. $N(0,1)$, where $\xi_1 = \xi$ and $\eta_1 = \eta$.

By Fubini's theorem,

$$\int_0^1 dx \sum\nolimits_k c_k^2 \big\{g_k(x)\big\}^2 = \sum\nolimits_k c_k^2 \, \|g_k\|^2 < \infty,$$

and so the sum on the left converges a.e. λ, and we may modify the g_k (along with f) on a null set, to ensure that

$$\sum\nolimits_k c_k^2 \left\{g_k(x)\right\}^2 < \infty, \quad x \in [0,1].$$

We may then define

$$\begin{aligned} Y(x) &= f(x,\eta)\\ &= \sum\nolimits_k c_k \, g_k(x)\, \psi_k, \quad x \in [0,1], \end{aligned}$$

with convergence in L^2. The array $X = (X_{ij})$ remains rotatable in j under conditioning on the invariant variables ξ_i, and so by Lemma 8.7 (i) and Theorem 27.15 the variables $Y(x_1), \dots, Y(x_n)$ are jointly centered Gaussian for $x_1, \dots, x_n \in [0,1]$ a.e. λ^n. Defining a measurable map $G: [0,1] \to l^2$ by $G(x) = \{c_k \, g_k(x)\}$ and putting $\mu = \lambda \circ G^{-1}$, we see from Lemma 1.19 that

$$\lambda\Big\{x \in [0,1];\, G(x) \in \operatorname{supp}\mu\Big\} = \mu(\operatorname{supp}\mu) = 1.$$

We can then modify the functions g_k (along with Y and f) on a null set, to make $G(x) \in \operatorname{supp}\mu$ hold identically. Since the set of Gaussian distributions on $\mathbb{R}^n$ is weakly closed, we conclude that Y is centered Gaussian.

Now let H_1 be the linear subspace of l^2 spanned by the sequences $\{c_k \, g_k(x)\}$, so that for any $(y_k) \in H_1$ the sum $\sum_k y_k \, \psi_k$ converges in L^2 to a centered Gaussian limit. If instead $(y_k) \perp H_1$ in l^2, we have $\sum_k c_k y_k \, g_k(x) \equiv 0$, which implies $\sum_k c_k y_k \, g_k = 0$ in L^2. Hence, $c_k y_k \equiv 0$ by ortho-normality, and so $y_k \equiv 0$. This gives $\bar{H}_1 = l^2$, and so the sums $\sum_k y_k \, \psi_k$ are centered Gaussian for all $(y_k) \in l^2$. The ψ_k are then jointly centered Gaussian by Corollary 6.5. For the remaining properties, interchange the roles of i and j, and use the ortho-normality and independence. □

Lemma 28.22 *(totally rotatable arrays, Aldous) Let X be a dissociated, separately rotatable array on $\mathbb{N}^2$ with shell σ-field $\mathcal{S}$, such that $E(X|\,\mathcal{S}) = 0$. Then the X_{ij} are i.i.d. centered Gaussian.*

Proof (OK): For any rotation R of rows or columns and indices $h \neq k$, we get by Lemma 8.8, Proposition 28.17 (iii), and the centering assumption

$$\begin{aligned} E\Big\{(RX)_k \,\Big|\, \mathcal{S}\Big\} &\overset{d}{=} E\Big(X_k \,\Big|\, \mathcal{S}\Big) = 0,\\ E\Big\{(RX)_h (RX)_k \,\Big|\, \mathcal{S}\Big\} &\overset{d}{=} E\Big(X_h X_k \,\Big|\, \mathcal{S}\Big)\\ &= E(X_h \,|\, \mathcal{S})\, E(X_k \,|\, \mathcal{S}) = 0. \end{aligned}$$

Letting $\mathcal{T}$ be the tail σ-field in either index, and noting that $\mathcal{T} \subset \mathcal{S}$, we get by the tower property of conditioning

$$E\Big\{(RX)_k \,\Big|\, \mathcal{T}\Big\} = 0,$$
$$E\Big\{(RX)_h (RX)_k \,\Big|\, \mathcal{T}\Big\} = 0.$$

Now $\mathcal{L}(X \,|\, \mathcal{T})$ is a.s. Gaussian by Theorem 27.15. Since the Gaussian property is preserved by rotations, Lemma 14.1 yields $\mathcal{L}(RX \,|\, \mathcal{T}) = \mathcal{L}(X \,|\, \mathcal{T})$, and so X is conditionally rotatable in both indices, hence i.i.d. $N(0,1)$. Thus, $X_{ij} = \sigma\, \zeta_{ij}$ a.s. for a G-array (ζ_{ij}) and an independent random variable $\sigma \geq 0$.

Since X is dissociated, we get for $h, k \in \mathsf{N}^2$ differing in both indices

$$\begin{aligned} E\,\sigma^2\, E|\zeta_h| \cdot E|\zeta_k| &= E\Big|X_h X_k\Big| \\ &= E|X_h| \cdot E|X_k| \\ &= (E\,\sigma)^2\, E|\zeta_h| \cdot E|\zeta_k|, \end{aligned}$$

and so $E\sigma^2 = (E\sigma)^2$, which yields $\mathrm{Var}(\sigma) = 0$, showing that σ is a constant. □

Proof of Theorem 28.19 (i): If X is separately rotatable on N^2, it is also separately exchangeable and hence representable as in (4), in terms of a U-array $(\alpha, \xi, \eta, \zeta)$. The rotatability on N^2 is preserved by conditioning on an exchangeable extension X^- to Z_-^2, and by Proposition 28.17 (vii) it is equivalent to condition on α, which makes X dissociated. By Theorem A1.3, it is enough to prove the representation under this conditioning, and so we may assume that X is representable as in Corollary 28.18 (i). Then $E|X_{ij}|^p < \infty$ for all $p > 0$ by Lemma 28.20.

Letting $\mathcal{S}$ be the shell σ-field of X, we may decompose X into the arrays

$$X' = E(X|\,\mathcal{S}), \qquad X'' = X - E(X|\,\mathcal{S}),$$

which are again separately rotatable by Lemma 8.8. Writing $g(a, x, y) = \lambda f(a, x, y, \cdot)$, we have by Propositions 8.7 (i), 8.9, and 28.17 (i)–(ii) the a.s. representations

$$\begin{aligned} X'_{ij} &= g(\xi_i, \eta_j), \\ X''_{ij} &= f(\xi_i, \eta_j, \zeta_{ij}) - g(\xi_i, \eta_j), \qquad i, j \in \mathsf{N}, \end{aligned}$$

and in particular X' and X'' are again dissociated. By Lemma 28.21, X' is representable as in (45). Furthermore, the X''_{ij} are i.i.d. centered Gaussian by Lemma 28.22, since clearly $E(X''|\,\mathcal{S}) = 0$. Then $X' \perp\!\!\!\perp X''$ by Proposition 28.17 (iv), and we may choose the underlying G-arrays to be independent. □

— — —

Exchangeable arrays can be used to model the interactions between nodes in a random graph or network, in which case X_{ij} represents the directed interaction between nodes i, j. For an *adjacency array* we have $X_{ij} \in \{0, 1\}$, where $X_{ij} = 1$ when there is a directed link from i to j.

We turn to the related subject of *random partitions* of a countable set I, represented by arrays

$$X_{ij} = \kappa_i \, 1\{i \sim j\}, \quad i, j \in I,$$

where $i \sim j$ whenever i, j belong to the same random subset of I, and $\kappa_i = \kappa_j$ is an associated random mark, taking values in a Borel space S. Given a class $\mathcal{P}$ of injective maps on I, we say that X is $\mathcal{P}$-*symmetric* if $X \circ p \overset{d}{=} X$ for all $p \in \mathcal{P}$, where $(X \circ p)_{ij} = X_{p_i, p_j}$. In particular, the partition is *exchangeable* if the array $X = (X_{ij})$ is jointly exchangeable.

We give an extended version of the celebrated *paintbox representation.*

Theorem 28.23 *(symmetric partitions, Kingman, OK) Let $\mathcal{P}$ be a class of injective maps on a countable set I, and let the array $X = (X_{ij})$ on I^2 represent a random partition of I with marks in a Borel space S. Then X is $\mathcal{P}$-symmetric iff*

$$X_{ij} = b(\xi_i) 1\{\xi_i = \xi_j\}, \quad i, j \in I, \tag{46}$$

for a measurable function $b \colon \mathsf{R} \to S$ and a $\mathcal{P}$-symmetric sequence of random variables $\xi_1, \xi_2, \ldots$.

Though contractable arrays on N^2 may not be exchangeable in general, the two properties are equivalent when X is the indicator array of a random partition on N. The equivalence fails for partitions of N^2.

Corollary 28.24 *(contractable partitions) Let X be a marked partition of N. Then*

$$X \text{ is exchangeable} \quad \Leftrightarrow \quad X \text{ is contractable.}$$

Proof: Use Theorem 28.23, together with the equivalence of exchangeability and contractability for a single sequence (ξ_n). □

Our proof of Theorem 28.23 is based on an elementary algebraic fact, where for any partition of N with associated *indicator array* $r_{ij} = 1\{i \sim j\}$, we define a mapping $m(r) \colon \mathsf{N} \to \mathsf{N}$ by

$$m_j(r) = \min\Big\{ i \in \mathsf{N};\ i \sim j \Big\}, \quad j \in \mathsf{N}.$$

Lemma 28.25 *(lead elements) For any indicator array r on N^2 and injective map p on N, there exists an injection q on N, depending measurably on r and p, such that*

$$q \circ m_j(r \circ p) = m_{p_j}(r), \quad j \in \mathsf{N}.$$

Proof: Fix any indicator array r with associated partition classes $A_1, A_2, \ldots$, listed in the order of first entry, and define

$$B_k = p^{-1} A_k, \qquad K = \Big\{ k \in \mathsf{N};\ B_k \neq \emptyset \Big\},$$
$$a_k = \min A_k, \quad b_k = \min B_k, \quad q(b_k) = a_k, \qquad k \in K.$$

For any $j \in \mathsf{N}$, we note that $j \in B_k = p^{-1}A_k$ and hence $p_j \in A_k$ for some $k \in K$, and so

$$q\Big\{m_j(r \circ p)\Big\} = q(b_k) = a_k = m_{p_j}(r).$$

To extend q to an injective map on N, we need to check that $|J^c| \le |I^c|$, where

$$I = \Big(a_k;\, k \in K\Big), \qquad J = \Big(b_k;\, k \in K\Big).$$

Noting that $|B_k| \le |A_k|$ for all k since p is injective, we get

$$\begin{aligned} |J^c| &= \bigcup_{k \in K} \Big(B_k \setminus \{b_k\}\Big) \\ &\le \bigcup_{k \in K} \Big(A_k \setminus \{a_k\}\Big) \le |I^c|. \end{aligned}$$

$\square$

Proof of Theorem 28.23: We may take $I = \mathsf{N}$. The sufficiency is clear from (46), since

$$\begin{aligned} (X \circ p)_{ij} &= X_{p_i, p_j} \\ &= b(\xi_{p_i})\, 1\{\xi_{p_i} = \xi_{p_j}\}, \quad i, j \in \mathsf{N}. \end{aligned}$$

To prove the necessity, let X be $\mathcal{P}$-symmetric, put $k_j(r) = r_{jj}$, choose $\vartheta_1, \vartheta_2, \ldots \perp\!\!\!\perp X$ to be i.i.d. $U(0,1)$, and define

$$\begin{aligned} \eta_j &= \vartheta \circ m_j(X), \\ \kappa_j &= k_j(X) = X_{jj}, \quad j \in \mathsf{N}. \end{aligned} \tag{47}$$

For any $p \in \mathcal{P}$, Lemma 28.25 yields a random injection $q(X)$ on N, such that

$$q(X) \circ m_j(X \circ p) = m(X) \circ p_j, \quad j \in \mathsf{N}. \tag{48}$$

Using (47) and (48), the $\mathcal{P}$-symmetry of X, the independence $\vartheta \perp\!\!\!\perp X$, and Fubini's theorem (four times), we get for any measurable function $f \ge 0$ on $([0,1] \times S)^\infty$

$$\begin{aligned} Ef\Big(\eta \circ p,\ \kappa \circ p\Big) &= Ef\Big\{\vartheta \circ m(X) \circ p,\ k(X) \circ p\Big\} \\ &= Ef\Big\{\vartheta \circ q(X) \circ m(X \circ p),\, k(X \circ p)\Big\} \\ &= E\Big|Ef\Big\{\vartheta \circ q(r) \circ m(r \circ p),\, k(r \circ p)\Big\}\Big|_{r=X} \\ &= E\Big|Ef\Big\{\vartheta \circ m(r \circ p),\, k(r \circ p)\Big\}\Big|_{r=X} \\ &= Ef\Big\{\vartheta \circ m(X \circ p),\, k(X \circ p)\Big\} \\ &= E\Big|Ef\Big\{t \circ m(X \circ p),\, k(X \circ p)\Big\}\Big|_{t=\vartheta} \\ &= E\Big|Ef\Big\{t \circ m(X),\, k(X)\Big\}\Big|_{t=\vartheta} \\ &= Ef\Big\{\vartheta \circ m(X),\, k(X)\Big\} \\ &= Ef(\eta, \kappa), \end{aligned}$$

which shows that $(\eta, \kappa) \circ p \overset{d}{=} (\eta, \kappa)$.

Now choose a Borel isomorphism $h\colon [0,1]\times S\to B\in\mathcal{B}$ with inverse $(b,c)\colon B\to[0,1]\times S$, and note that the B-valued random variables $\xi_j = h(\kappa_j,\eta_j)$ are again $\mathcal{P}$-symmetric. Now (46) follows from

$$\begin{aligned}\xi_i=\xi_j \quad&\Leftrightarrow\quad (\kappa_i,\eta_i)=(\kappa_j,\eta_j)\\ &\Leftrightarrow\quad \eta_i=\eta_j\\ &\Leftrightarrow\quad m_i(X)=m_j(X)\\ &\Leftrightarrow\quad i\sim j.\end{aligned}$$

□

Exercises

1. Show how the de Finetti–Ryll-Nardzewski theorem can be obtained as a special case of Theorem 28.1.

2. Explain the difference between the representations in (2) and (4). Then write out explicitly the corresponding three-dimensional representations, and again explain in what way they differ.

3. State a coding representation of a single random element in a Borel space, and specify when two coding functions f, g give rise to random elements with the same distribution.

4. Show how the one-dimensional coding representation with associated equivalence criteria follow from the corresponding two-dimensional results.

5. Give an example of a measure-preserving map f on $[0,1]$ that is not invertible. Then show how we can construct an inversion by introducing an extra randomization variable. (*Hint:* For a simple example, let $f(x) = 2\,x \pmod 1$ for $x\in[0,1]$. Then introduce a Bernoulli variable to choose between the intervals $[0,\frac{1}{2}]$ and $(\frac{1}{2},1]$.)

6. Let $X_{ij} = f(\alpha,\xi_i,\eta_j)$ for some i.i.d. random variables α, $\xi_1,\xi_2,\ldots$, and $\eta_1,\eta_2,\ldots$. Show that the X_{ij} cannot be i.i.d., non-degenerate. (*Hint:* This can be shown from either Corollary 28.18 or Theorem 28.3.)

7. Express the notions of separate and joint rotatability as invariance under unitary transformations on a Hilbert space H. Then write the double sum in Theorem 28.19 (ii) in terms of a double Wiener-Itô integral of a suitable isonormal Gaussian process. Similarly, write the double sum in part (i) as a tensor product of two independent isonormal Gaussian processes on H.

8. Show that all sets in an exchangeable partition have cardinality 1 or ∞, and give the probabilities for a fixed element to belong to a set of each type. Further give the probability for two fixed elements to belong to the same set. (*Hint:* Use the paintbox representation.)

9. Say that a graph is *exchangeable* if the associated adjacency matrix is jointly exchangeable. Show that in an infinite, exchangeable graph, a.s. every node is either disconnected or has infinitely many links to other nodes. (*Hint:* Use Lemma 30.9.)

10. Say that two nodes in a graph are *connected*, if one can be reached from the other by a sequence of links. Show that this defines an equivalence relation between sites. For an infinite, exchangeable graph, show that the associated partition is exchangeable. (*Hint:* Note that the partition is exchangeable and hence has a paintbox representation.)

IX. Random Sets and Measures

In Chapter 29 we show how the entire excursion structure of a regenerative process can be described by a Poisson process on the time scale given by an associated local time process. We further study the local time as a function of the space variable, and initiate the study of continuous additive functionals. In Chapter 30 we discuss the weak and strong Poisson or Cox convergence under superposition and scattering of point processes, highlighting the relationship between the asymptotic properties of the processes and smoothing properties of the underlying transforms. Finally, in Chapter 31 we explore the basic notion of Palm measures, first for stationary point processes on the real line, and then for general random measures on an abstract Borel space. In both cases, we prove some local approximations and develop a variety of duality relations. A duality argument also leads to the notion of Gibbs kernel, related to ideas in statistical mechanics. For a first encounter, we recommend especially a careful study of the semi-martingale local time and excursion theory in Chapter 29, along with some basic limit theorems from Chapter 30.

— — —

29. Local time, excursions, and additive functionals. This chapter provides three totally different but essentially equivalent approaches to local time. Our first approach is based on Tanaka's formula in stochastic calculus and leads to an interpretation of local time as an occupation density. The second approach extends to a fundamental representation of the entire excursion structure of a regenerative process in terms of a Poisson process on the local time scale. The third approach is based on the potential theory of continuous additive functionals.

30. Random measures, smoothing and scattering. Here we begin with the weak or strong Poisson convergence, for superpositions of independent, uniformly small point processes. Next, we consider the Cox convergence of dissipative transforms of suitable point processes, depending on appropriate averaging and smoothing properties of the transforms. Here a core result is the celebrated theorem of Dobrushin, based on conditions of asymptotic invariance. We finally explore some Cox criteria of stationary line and flat processes, and study the basic cluster dichotomy for spatial branching processes.

31. Palm and Gibbs kernels, local approximation. Here we begin with a study of Palm measures of stationary point processes on Euclidean spaces, highlighting the dual relationship between stationarity in discrete and continuous time. Next, we consider the Palm measures of random measures on an abstract Borel space, and prove some basic local approximations. Further highlights include a conditioning approach to Palm measures and a dual approach via conditional densities. Finally, we discuss the dual notions of Gibbs and Papangelou kernels, of significance in statistical mechanics.

Chapter 29

Local Time, Excursions, and Additive Functionals

Semi-martingale local time, Tanaka's formula, space-time regularity, occupation measure and density, extended Itô formula, regenerative sets and processes, excursion law, excursion local time and Poisson process, approximations of local time, inverse local time as a subordinator, Brownian excursion, Ray–Knight theorem, continuous additive functionals, Revuz measure, potentials, excessive functions, additive-functional local time, additive functionals of Brownian motion

Local time is yet another indispensable notion of modern probability, playing key roles in both stochastic calculus and excursion theory, and also needed to represent continuous additive functionals of suitable Markov processes. Most remarkably, it appears as the occupation density of continuous semi-martingales, and it provides a canonical time scale for the family of excursions of a regenerative process.

In each of the mentioned application areas, there is an associated method of construction, which leads to three different but essentially equivalent versions of the local time process. We begin with the semi-martingale local time, constructed from Tanaka's formula, and leading to a useful extension of Itô's formula, and to an interpretation of local time as an occupation density. Next we develop excursion theory for processes that are regenerative at a fixed state, and prove the powerful Itô representation, involving a Poisson process of excursions on the time scale given by an associated local time. Among its many applications, we consider a version of the Ray–Knight theorem, describing the spatial variation of Brownian local time. Finally, we study continuous additive functionals (CAFs) and their potentials, prove the existence of local time at a regular point, and show that any CAF of a one-dimensional Brownian motion is a mixture of local times.

The beginning of the chapter may be regarded as a continuation of the stochastic calculus developed in Chapter 18. The present excursion theory continues the elementary discussion of the discrete-time case in Chapter 12. Though the theory of CAFs is formally developed for Feller processes, few results are needed from Chapter 17, beyond the strong Markov property and its integrated version in Corollary 17.19. Both semi-martingale local time and excursion theory will be useful in Chapter 33 to study one-dimensional SDEs and diffusions. Our discussion of CAFs of Brownian motion with associated potentials will continue at the end of Chapter 34.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_30

For the stochastic calculus approach to local time, let X be a continuous semi-martingale in R. The *semi-martingale local time* L^0 of X at 0 may be defined by *Tanaka's formula*

$$L_t^0 = |X_t| - |X_0| - \int_0^t \operatorname{sgn}(X_s-)\, dX_s, \quad t \ge 0, \tag{1}$$

where $\operatorname{sgn}(x-) = 1_{(0,\infty)}(x) - 1_{(-\infty,0]}(x)$. More generally, we define the local time L_t^x at a point $x \in \mathsf{R}$ as the local time at 0 of the process $X_t - x$. Note that the stochastic integral on the right exists, since the integrand is bounded and progressive. The process L^0 is clearly continuous and adapted with $L_0^0 = 0$. For motivation, we note that if Itô's rule could be applied to the function $f(x) = |x|$, we would formally obtain (1) with $L_t^0 = \int_{s\le t} \delta(X_s)\, d[X]_s$.

We state the basic properties of local time at a fixed point. Say that a non-decreasing function f is *supported* by a Borel set A, if the associated measure μ satisfies $\mu A^c = 0$. The *support* of f is the smallest closed set with this property.

Theorem 29.1 *(semi-martingale local time) Let X be a continuous semi-martingale with local time L^0 at 0. Then a.s.*

(i) *L^0 is non-decreasing, continuous, and supported by*

$$\Xi = \big\{t \ge 0;\, X_t = 0\big\},$$

(ii) $$L_t^0 = \Big\{-|X_0| - \inf_{s\le t} \int_0^s \operatorname{sgn}(X-)\, dX\Big\} \vee 0, \quad t \ge 0.$$

The proof of (ii) depends on an elementary observation.

Lemma 29.2 *(supporting function, Skorohod) Let f be a continuous function f on R_+ with $f_0 \ge 0$. Then*

$$g_t = -\inf_{s\le t} f_s \wedge 0 = \sup_{s\le t}(-f_s) \vee 0, \quad t \ge 0,$$

is the unique non-decreasing, continuous function g on R_+ with $g_0 = 0$, satisfying

$$h \equiv f + g \ge 0, \qquad \int 1\{h > 0\}\, dg = 0,$$

Proof: The given function has clearly the stated properties. To prove the uniqueness, suppose that both g and g' have the stated properties, and put $h = f + g$ and $h' = f + g'$. If $g_t < g'_t$ for some $t > 0$, define $s = \sup\{r < t;\, g_r = g'_r\}$, and note that $h' \ge h' - h = g' - g > 0$ on $(s, t]$. Hence, $g'_s = g'_t$, and so $0 < g'_t - g_t \le g'_s - g_s = 0$, a contradiction. □

Proof of Theorem 29.1: (i) For any constant $h > 0$, we may choose a convex function $f_h \in C^2$ with

$$f_h(x) = \begin{cases} -x, & x \le 0, \\ x - h, & x \ge h. \end{cases}$$

Then Itô's formula gives, a.s. for any $t \geq 0$,

$$\begin{aligned} Y_t^h &\equiv f_h(X_t) - f_h(X_0) - \int_0^t f_h'(X_s)\, dX_s \\ &= \tfrac{1}{2} \int_0^t f_h''(X_s)\, d[X]_s. \end{aligned}$$

As $h \to 0$, we have $f_h(x) \to |x|$ and $f_h' \to \operatorname{sgn}(x-)$. By Corollary 18.13 and dominated convergence, we get $(Y^h - L^0)_t^* \xrightarrow{P} 0$ for every $t > 0$. Now (i) follows, since the processes Y^h are non-decreasing with

$$\int_0^\infty 1\Big\{X_s \notin [0,h]\Big\}\, dY_s^h = 0 \text{ a.s.}, \quad h > 0.$$

(ii) This is clear from Lemma 29.2. □

This leads in particular to a basic relationship between a Brownian motion, its maximum process, and its local time at 0. The result extends the more elementary Proposition 14.13.

Corollary 29.3 *(local time and maximum process, Lévy) Let L^0 be the local time at 0 of a Brownian motion B, and define $M_t = \sup_{s \leq t} B_s$. Then*

$$\Big(L^0, |B|\Big) \stackrel{d}{=} (M, M - B).$$

Proof: Define

$$B_t' = -\int_{s \leq t} \operatorname{sgn}(B_s-)\, dB_s, \qquad M_t' = \sup_{s \leq t} B_s',$$

and conclude from (1) and Theorem 29.1 (ii) that $L^0 = M'$ and $|B| = L^0 - B' = M' - B'$. Further note that $B' \stackrel{d}{=} B$ by Theorem 19.3. □

For any continuous local martingale M, the associated local time L_t^x has a jointly continuous version. More generally, we have for any continuous semimartingale:

Theorem 29.4 *(spatial regularity, Trotter, Yor) Let X be a continuous semimartingale with canonical decomposition $M + A$. Then the local time $L = (L_t^x)$ of X has a version that is rcll in x, uniformly for bounded t, and satisfies*

$$L_t^x - L_t^{x-} = 2 \int_0^t 1\{X_s = x\}\, dA_s, \quad x \in \mathsf{R},\ t \in \mathsf{R}_+.$$

Proof: By the definition of L, we have for any $x \in \mathsf{R}$ and $t \geq 0$

$$\begin{aligned} L_t^x = &\; |X_t - x| - |X_0 - x| \\ &- \int_0^t \operatorname{sgn}\Big(X_s - x - \Big)\, dM_s - \int_0^t \operatorname{sgn}\Big(X_s - x - \Big)\, dA_s. \end{aligned}$$

By dominated convergence, the last term has the required continuity properties, with discontinuities given by the displayed formula. Since the first two terms are trivially continuous in x and t, it remains to show that the first integral term, henceforth denoted by I_t^x, has a jointly continuous version.

By localization we may then assume that the processes $X - X_0$, $[M]^{1/2}$, and $\int |dA|$ are all bounded by some constant c. Fixing any $p > 2$ and using Theorem 18.7, we get[1] for any $x < y$

$$\begin{aligned} E\Big(I^x - I^y\Big)_t^{*p} &\le 2^p\, E\Big\{1_{(x,y]}(X) \cdot M\Big\}_t^{*p} \\ &\lesssim E\Big\{1_{(x,y]}(X) \cdot [M]\Big\}_t^{p/2}. \end{aligned}$$

To estimate the integral on the right, put $y - x = h$, and choose $f \in C^2$ with $f'' \ge 2 \cdot 1_{(x,y]}$ and $|f'| \le 2h$. Then by Itô's formula,

$$\begin{aligned} 1_{(x,y]}(X) \cdot [M] &\le \tfrac{1}{2} f''(X) \cdot [X] \\ &= f(X) - f(X_0) - f'(X) \cdot X \\ &\le 4\,c\,h + \Big|f'(X) \cdot M\Big|, \end{aligned}$$

and another use of Theorem 18.7 yields

$$\begin{aligned} E\Big\{f'(X) \cdot M\Big\}_t^{*p/2} &\lesssim E\Big(\{f'(X)\}^2 \cdot [M]\Big)_t^{p/4} \\ &\le (2\,c\,h)^{p/2}. \end{aligned}$$

Combining the last three estimates gives $E(I^x - I^y)_t^{*p} \lesssim (c\,h)^{p/2}$, and since $p/2 > 1$, the desired continuity follows by Theorem 4.23. □

We may henceforth take L to be a regularized version of the local time. Given a continuous semi-martingale X, write ξ_t for the associated *occupation measures*, defined as below with respect to the quadratic variation process $[X]$. We show that $\xi_t \ll \lambda$ a.s. for every $t \ge 0$, and that the set of local times L_t^x provides the associated densities. This also leads to a simultaneous extension of the Itô and Tanaka formulas.

Since every convex function f on R has a non-decreasing and left-continuous left derivative $f'_-(x) = f'(x-)$, Theorem 2.14 yields an associated measure μ_f on R satisfying

$$\mu_f[x, y) = f'_-(y) - f'_-(x) \quad x \le y \text{ in } \mathsf{R}.$$

More generally, when f is the difference of two convex functions, μ_f becomes a locally finite, signed measure.

Theorem 29.5 *(occupation density, Meyer, Wang) Let X be a continuous semi-martingale with right-continuous local time L. Then outside a fixed P-null set, we have*

(i) *for any measurable function $f \ge 0$ on R,*

$$\int_0^t f(X_s)\, d[X]_s = \int_{-\infty}^{\infty} f(x)\, L_t^x\, dx, \quad t \ge 0,$$

[1] Recall the notation $(X \cdot Y)_t = \int_0^t X_s\, dY_s$.

(ii) *when f is the difference between two convex functions,*

$$f(X_t) - f(X_0) = \int_0^t f'_-(X)\, dX + \tfrac{1}{2} \int_{-\infty}^{\infty} L_t^x \, \mu_f(dx), \quad t \ge 0.$$

In particular, Theorem 18.18 *extends to functions $f \in C^1_{\mathsf{R}}$, such that f' is absolutely continuous with density f''.*

Note that (ii) remains valid for the left-continuous version of L, provided we replace $f'_-(X)$ by the right derivative $f'_+(X)$.

Proof: (ii) For $f(x) \equiv |x-a|$, this reduces to the definition of L_t^a. Since the formula is also trivially true for affine functions $f(x) \equiv ax + b$, it extends by linearity to the case where μ_f is supported by a finite set. By linearity and suitable truncation, it remains to prove the formula when μ_f is positive with bounded support and $f(-\infty) = f'(-\infty) = 0$. Then for any $n \in \mathsf{N}$, we define the functions

$$\begin{aligned} g_n(x) &= f'\Big(2^{-n}[2^n x]-\Big), \\ f_n(x) &= \int_{-\infty}^x g_n(u)\, du, \qquad x \in \mathsf{R}, \end{aligned}$$

and note that (ii) holds for all f_n. As $n \to \infty$ we get $f'_n(x-) = g_n(x-) \uparrow f'_-(x)$, and so by Corollary 18.13 we have $f'_n(X-) \cdot X \xrightarrow{P} f'_-(X) \cdot X$. Further note that $f_n \to f$ by monotone convergence. It remains to show that $\int L_t^x \, \mu_{f_n}(dx) \to \int L_t^x \, \mu_f(dx)$. Then let h be any bounded, right-continuous function on R, and note that $\mu_{f_n} h = \mu_f h_n$ with $h_n(x) = h(2^{-n}[2^n x + 1])$. Since $h_n \to h$, we get $\mu_f h_n \to \mu_f h$ by dominated convergence.

(i) Comparing (ii) with Itô's formula, we obtain the stated relation for fixed $t \ge 0$ and $f \in C$. Since both sides define random measures on R, a simple approximation yields $\xi_t = L_t \cdot \lambda$, a.s. for fixed $t \ge 0$. By the continuity of each side, we may choose the exceptional null set to be independent of t.

If $f \in C^1$ with f' as stated, then (ii) applies with $\mu_f(dx) = f''(x)\, dx$, and the last assertion follows by (i). $\square$

In particular, the *occupation measure* of X at time t, given by

$$\eta_t A = \int_0^t 1_A(X_s)\, d[X]_s, \quad A \in \mathcal{B}_{\mathsf{R}}, \ t \ge 0, \tag{2}$$

is a.s. absolutely continuous with density L_t. This leads to a spatial approximation of L, which may be compared with the temporal approximation in Proposition 29.12 below.

Corollary 29.6 *(spatial approximation) Let X be a continuous semi-martingale in R with local time L and occupation measures η_t. Then outside a fixed P-null set, we have*

$$L_t^x = \lim_{h \to 0} h^{-1} \eta_t [x, x+h), \quad t \ge 0, \ x \in \mathsf{R}.$$

Proof: Use Theorem 29.5 and the right-continuity of L. □

Local time processes also arise naturally in the context of regenerative processes. Here we consider an rcll process X in a Polish space S, adapted to a right-continuous and complete filtration $\mathcal{F}$. Say that X is *regenerative* at a state $a \in S$, if it satisfies the strong Markov property at a, so that for any optional time τ,

$$\mathcal{L}\Big(\theta_\tau X \,\Big|\, \mathcal{F}_\tau\Big) = P_a \text{ a.s. on } \Big\{\tau < \infty,\, X_\tau = a\Big\},$$

for some probability measure P_a on the path space $D_{\mathsf{R}_+,S}$ of X. In particular, a strong Markov process is regenerative at every point.

The associated random set $\Xi = \{t \geq 0;\, X_t = a\}$ is regenerative in the same sense. By the right-continuity of X, we note that $\Xi \ni t_n \downarrow t$ implies $t \in \Xi$, which means that all points of $\bar{\Xi} \setminus \Xi$ are isolated from the right. In particular, the first-entry times $\tau_r = \inf\{t \geq r;\, t \in \Xi\}$ lie in Ξ, a.s. on $\{\tau_r < \infty\}$. Since $(\bar{\Xi})^c$ is open and hence a countable union of disjoint intervals (u, v), it also follows that Ξ^c is a countable union of intervals (u, v) or $[u, v)$. If we are only interested in regeneration at a fixed point a, we may further assume that $X_0 = a$ or $0 \in \Xi$ a.s. The mentioned properties will be assumed below, even when we make no reference to an underlying process X.

We first classify the regenerative sets into five different categories. Recall that a set $B \subset \mathsf{R}_+$ is said to be *nowhere dense* if $(\bar{B})^o = \emptyset$, and *perfect* if it is closed with no isolated points.

Theorem 29.7 *(regenerative sets) For a regenerative set Ξ, exactly one of these cases occurs a.s.:*

(i) *Ξ is locally finite,*

(ii) $\Xi = \mathsf{R}_+$,

(iii) *Ξ is a locally finite union of disjoint intervals with i.i.d., exponentially distributed lengths,*

(iv) *Ξ is nowhere dense with no isolated points, and $\bar{\Xi} = \operatorname{supp}(1_\Xi \lambda)$,*

(v) *Ξ is nowhere dense with no isolated points, and $\lambda\Xi = 0$.*

Case (iii) *is excluded when Ξ is closed.*

In case of (i), Ξ is a generalized renewal process based on a distribution μ on $(0, \infty]$. In cases (ii)−(iv), Ξ is known as a *regenerative phenomenon*, whose distribution is determined by the *p-function* $p(t) = P\{t \in \Xi\}$. Our proof of Theorem 29.7 is based on the following dichotomies:

Lemma 29.8 *(local dichotomies) For a regenerative set Ξ,*

(i) *either $(\bar{\Xi})^o = \emptyset$ a.s., or $\overline{\Xi^o} = \bar{\Xi}$ a.s.,*

(ii) *either a.s. all points of Ξ are isolated, or a.s. none of them is,*

(iii) *either $\lambda\Xi = 0$ a.s., or* $\operatorname{supp}(\Xi \cdot \lambda) = \bar{\Xi}$ *a.s.*

Proof: We may take $\mathcal{F}$ to be the right-continuous, complete filtration induced by Ξ, which allows us to use a canonical notation. For any optional time τ, the regenerative property yields

$$\begin{aligned} P\{\tau = 0\} &= E\Big\{P(\tau = 0 \,|\, \mathcal{F}_0);\ \tau = 0\Big\} \\ &= (P\{\tau = 0\})^2, \end{aligned}$$

and so $P\{\tau = 0\} = 0$ or 1. If σ is another optional time, then $\tau' = \sigma + \tau \circ \theta_\sigma$ is again optional by Proposition 11.8, and so for any $h \geq 0$,

$$\begin{aligned} P\Big\{\tau' - h \leq \sigma \in \Xi\Big\} &= P\Big\{\tau \circ \theta_\sigma \leq h,\ \sigma \in \Xi\Big\} \\ &= P\{\tau \leq h\}\, P\{\sigma \in \Xi\}, \end{aligned}$$

which implies $\mathcal{L}(\tau' - \sigma \,|\, \sigma \in \Xi) = P \circ \tau^{-1}$. In particular, $\tau = 0$ a.s. gives $\tau' = \sigma$ a.s. on $\{\sigma \in \Xi\}$.

(i) The previous observations apply to the optional times $\tau = \inf \Xi^c$ and $\sigma = \tau_r$. If $\tau > 0$ a.s., then $\tau \circ \theta_{\tau_r} > 0$ a.s. on $\{\tau_r < \infty\}$, and so $\tau_r \in \overline{\Xi^o}$ a.s. on the same set. Since the set $\{\tau_r;\ r \in \mathsf{Q}_+\}$ is dense in $\bar{\Xi}$, we obtain $\bar{\Xi} = \overline{\Xi^o}$ a.s. If instead $\tau = 0$ a.s., then $\tau \circ \theta_{\tau_r} = 0$ a.s. on $\{\tau_r < \infty\}$, and so $\tau_r \in \overline{\Xi^c}$ a.s. on the same set. Hence, $\bar{\Xi} \subset \overline{\Xi^c}$ a.s., and so $\overline{\Xi^c} = \mathsf{R}_+$ a.s. It remains to note that $\overline{\Xi^c} = \overline{(\bar{\Xi})^c}$, since Ξ^c is a disjoint union of intervals (u, v) or $[u, v)$.

(ii) Put $\tau = \inf(\Xi \setminus \{0\})$. If $\tau = 0$ a.s., then $\tau \circ \theta_{\tau_r} = 0$ a.s. on $\{\tau_r < \infty\}$. Since every isolated point of Ξ equals τ_r for some $r \in \mathsf{Q}_+$, it follows that Ξ has a.s. no isolated points. If instead $\tau > 0$ a.s., we define recursively some optional times σ_n by $\sigma_{n+1} = \sigma_n + \tau \circ \theta_{\sigma_n}$, starting with $\sigma_1 = \tau$. Then the σ_n form a renewal process based on $\mathcal{L}(\tau)$, and so $\sigma_n \to \infty$ a.s. by the law of large numbers. Thus, $\Xi = \{\sigma_n < \infty;\ n \in \mathsf{N}\}$ a.s. with $\sigma_0 = 0$, and a.s. all points of Ξ are isolated.

(iii) Let $\tau = \inf\{t > 0;\ (1_\Xi \cdot \lambda)_t > 0\}$. If $\tau = 0$ a.s., then $\tau \circ \theta_{\tau_r} = 0$ a.s. on $\{\tau_r < \infty\}$, and so $\tau_r \in \operatorname{supp}(1_\Xi \lambda)$ a.s. on the same set. Hence, $\bar{\Xi} \subset \operatorname{supp}(1_\Xi \lambda)$ a.s., and the two sets agree a.s. If instead $\tau > 0$ a.s., then $\tau = \tau + \tau \circ \theta_\tau > \tau$ a.s. on $\{\tau < \infty\}$, which implies $\tau = \infty$ a.s., and so $\lambda \Xi = 0$ a.s. $\Box$

Proof of Theorem 29.7: Using Lemma 29.8 (ii), we may eliminate case (i) where Ξ is a renewal process, which is clearly a.s. locally finite. Next, we may use parts (i) and (iii) of the same lemma to separate out cases (iv) and (v). Excluding (i) and (iv)$-$(v), we see that Ξ is a.s. a locally finite union of intervals of i.i.d. lengths. Writing $\gamma = \inf \Xi^c$ and using the regenerative property at τ_s, we get

$$\begin{aligned} P\{\gamma > s + t\} &= P\Big\{\gamma \circ \theta_s > t,\ \gamma > s\Big\} \\ &= P\{\gamma > s\}\, P\{\gamma > t\}, \quad s, t \geq 0. \end{aligned}$$

Hence, the monotone function $p_t = P\{\gamma > t\}$ satisfies the Cauchy equation $p_{s+t} = p_s\, p_t$ with initial condition $p_0 = 1$, and so $p_t \equiv e^{-ct}$ for some constant

$c \geq 0$. Here $c = 0$ corresponds to case (ii) and $c > 0$ to case (iii). If Ξ is closed, then $\gamma \in \Xi$ on $\{\gamma < \infty\}$, and the regenerative property at γ yields $\gamma = \gamma + \gamma \circ \theta_\gamma > \gamma$ a.s. on $\{\gamma < \infty\}$, which implies $\gamma = \infty$ a.s., corresponding to case (ii). $\square$

The complement $(\bar{\Xi})^c$ is a countable union of disjoint intervals (σ, τ), each supporting an *excursion* of X away from a. The shifted excursions $Y_{\sigma,\tau} = \theta_\sigma X^\tau$ are rcll processes in their own right. Let D_0 be the space of possible excursion paths, write $l(x)$ for the length of excursion $x \in D_0$, and put $D_h = \{x \in D_0;\ l(x) > h\}$. Let κ_h be the number of excursions of X longer than h. We need some basic facts.

Lemma 29.9 *(long excursions) Fix any $h > 0$, and allow even $h \geq 0$ when the recurrence time is positive. Then exactly one of these cases occurs:*

(i) $\kappa_h = 0$ *a.s.,*

(ii) $\kappa_h = \infty$ *a.s.,*

(iii) κ_h *is geometrically distributed with mean* $m_h \in [1, \infty)$.

The excursions Y_j^h are i.i.d., and in case (iii) *we have* $l(Y_{\kappa_h}^h) = \infty$.

Proof: Let κ_h^t be the number of excursions in D_h completed at time $t \in [0, \infty]$, and note that $\kappa_h^{\tau_t} > 0$ when $\tau_t = \infty$. Writing $p_h = P\{\kappa_h > 0\}$, we get for any $t > 0$

$$\begin{aligned} p_h &= P\{\kappa_h^{\tau_t} > 0\} + P\Big\{\kappa_h^{\tau_t} = 0,\ \kappa_h \circ \theta_{\tau_t} > 0\Big\} \\ &= P\{\kappa_h^{\tau_t} > 0\} + P\{\kappa_h^{\tau_t} = 0\}\, p_h. \end{aligned}$$

As $t \to \infty$, we obtain $p_h = p_h + (1 - p_h)\, p_h$, which implies $p_h = 0$ or 1.

When $p_h = 1$, let $\sigma = \sigma_1$ be the right endpoint of the first D_h-excursion, and define recursively $\sigma_{n+1} = \sigma_n + \sigma_1 \circ \theta_{\sigma_n}$, starting with $\sigma_0 = 0$. Writing $q = P\{\sigma < \infty\}$ and using the regenerative property at each σ_n, we obtain $P\{\kappa_h > n\} = q^n$. Thus, $q = 1$ yields $\kappa_h = \infty$ a.s., whereas for $q < 1$ the variable κ_h is geometrically distributed with mean $m_h = (1-q)^{-1}$, proving the first assertion. Even the last claim holds by regeneration at each σ_n. $\square$

Write $\hat{h} = \inf\{h > 0;\ \kappa_h = 0 \text{ a.s.}\}$. For any $h < \hat{h}$, let ν_h be the common distribution of all excursions in D_h. The measures ν_h may be combined into a single σ-finite measure ν:

Lemma 29.10 *(excursion law, Itô) Let X be regenerative at a with $\bar{\Xi}$ a.s. perfect. Then*

(i) *there exists a measure ν on D_0, such that for every $h \in (0, \hat{h})$,*

$$\nu D_h \in (0, \infty), \qquad \nu_h = \nu(\,\cdot\,|D_h),$$

(ii) *ν is unique up to a normalization,*

(iii) *ν is bounded iff the recurrence time is a.s. positive.*

Proof (OK): (i) Fix any $h \le k$ in $(0, \hat{h})$, and let $Y_h^1, Y_h^2, \ldots$ be such as in Lemma 29.9. Then the first excursion in D_k is the first process Y_h^j belonging to D_k, and since the Y_h^j are i.i.d. ν_h, we have

$$\nu_k = \nu_h(\,\cdot\,|D_k), \quad 0 < h \le k < \hat{h}. \tag{3}$$

Now fix any $k \in (0, \hat{h})$, and define $\tilde{\nu}_h = \nu_h / \nu_h D_k$ for all $h \in (0, k]$. Then (3) yields $\tilde{\nu}_{h'} = \tilde{\nu}_h(\cdot \cap D_{h'})$ for any $h \le h' \le k$, and so $\tilde{\nu}_h$ increases as $h \to 0$ toward a measure ν on D_0 with $\nu(\cdot \cap D_h) = \tilde{\nu}_h$ for all $h \le k$. For any $h \in (0, \hat{h})$, we get

$$\begin{aligned} \nu(\,\cdot\,|D_h) &= \tilde{\nu}_{h \wedge k}(\,\cdot\,|D_h) \\ &= \nu_{h \wedge k}(\,\cdot\,|D_h) = \nu_h. \end{aligned}$$

(ii) If ν' is another measure with the stated properties, then

$$\frac{\nu(\cdot \cap D_h)}{\nu D_k} = \frac{\nu_h}{\nu_h D_k} = \frac{\nu'(\cdot \cap D_h)}{\nu' D_k}, \quad h \le k < \hat{h}.$$

Letting $h \to 0$ for fixed k, we get $\nu = r\nu'$ with $r = \nu D_k / \nu' D_k$.

(iii) If the recurrence time is positive, then (3) remains true for $h = 0$, and we may take $\nu = \nu_0$. Otherwise, let $h \le k$ in $(0, \hat{h})$, and write $\kappa_{h,k}$ for the number of D_h-excursions up to the first completed excursion in D_k. For fixed k we have $\kappa_{h,k} \to \infty$ a.s. as $h \to 0$, since $\bar{\Xi}$ is perfect and nowhere dense. Noting that $\kappa_{h,k}$ is geometrically distributed with mean

$$\begin{aligned} E\kappa_{h,k} &= (\nu_h D_k)^{-1} \\ &= \big\{\nu(D_k|D_h)\big\}^{-1} \\ &= \nu D_h / \nu D_k, \end{aligned}$$

we get $\nu D_h \to \infty$, which shows that ν is unbounded. □

We turn to the fundamental theorem of excursion theory, describing the excursion structure of X in terms of a Poisson process ξ on $\mathsf{R}_+ \times D_0$ and a diffuse random measure ζ on R_+. The measure ζ, with associated cumulative process $L_t = \zeta[0, t]$, is unique up to a normalization and will be referred to as the *excursion local time* of X at a, whereas ξ is referred to as the *excursion point process* of X at a.

Theorem 29.11 *(excursion local time and point process, Lévy, Itô) Let the process X be regenerative at a with $\bar{\Xi}$ a.s. perfect. Then there exist a non-decreasing, continuous, adapted process L on R_+ with support $\bar{\Xi}$ a.s., a Poisson process ξ on $\mathsf{R}_+ \times D_0$ with $E\xi = \lambda \otimes \nu$, and a constant $c \ge 0$, such that*

(i) $\Xi \cdot \lambda = c\,L$ *a.s.,*

(ii) *the excursions of X with associated L-values are given a.s. by the restriction of ξ to $[0, L_\infty]$.*

The product $\nu \cdot L$ is a.s. unique.

Proof (OK, beginning): When $E\gamma = c > 0$, we define $\nu = \nu_0/c$, and introduce a Poisson process ξ on $\mathsf{R}_+ \times D_0$ with intensity $\lambda \otimes \nu$, say with points $(\sigma_j, \tilde{Y}_j)$, $j \in \mathsf{N}$. Putting $\sigma_0 = 0$, we see from Theorem 13.6 that the differences $\tilde{\gamma}_j = \sigma_j - \sigma_{j-1}$ are i.i.d. exponentially distributed with mean m. By Proposition 15.3 (i), the processes $\tilde{Y}_j$ are further independent of the σ_j and i.i.d. ν_0. Writing $\tilde{\kappa} = \inf\{j;\ l(\tilde{Y}_j) = \infty\}$, we see from Theorem 29.7 and Lemma 29.9 that

$$\Big(\gamma_j, Y_j;\ j \le \kappa\Big) \stackrel{d}{=} \Big(\tilde{\gamma}_j, \tilde{Y}_j;\ j \le \tilde{\kappa}\Big), \tag{4}$$

where the quantities on the left are the holding times and subsequent excursions of X. By Theorem 8.17 we may redefine ξ to make (4) hold a.s. Then the stated conditions become fulfilled with $L = 1_\Xi \cdot \lambda$.

When $E\gamma = 0$, we may again choose ξ to be Poisson $\lambda \otimes \nu$, but now with ν as in Lemma 29.10. For any $h \in (0, \hat{h})$, we may enumerate the points of ξ in $\mathsf{R}_+ \times D_h$ as $(\sigma_h^j, \tilde{Y}_j^h)$, $j \in \mathsf{N}$, and define $\tilde{\kappa}_h = \inf\{j;\ l(Y_j^h) = \infty\}$. The processes $\tilde{Y}_j^h$ are i.i.d. ν_h, and Lemma 29.9 yields

$$\Big(\gamma_j^h, Y_j^h;\ j \le \kappa_h\Big) \stackrel{d}{=} \Big(\tilde{\gamma}_j^h, \tilde{Y}_j^h;\ j \le \tilde{\kappa}_h\Big), \quad h \in (0, \hat{h}). \tag{5}$$

Since the excursions in D_h form nested sub-arrays, the entire collections have the same finite-dimensional distributions, and so by Theorem 8.17 we may redefine ξ to make (5) hold a.s.

Let τ_h^j be the right endpoint of the j-th excursion in D_h, and define

$$L_t = \inf\Big\{\sigma_h^j;\ h, j > 0,\ \tau_h^j \ge t\Big\}, \quad t \ge 0.$$

Then for any $t \ge 0$ and $h, j > 0$,

$$\begin{aligned} L_t < \sigma_h^j \ &\Rightarrow\ t \le \tau_h^j \\ &\Rightarrow\ L_t \le \sigma_h^j. \end{aligned} \tag{6}$$

To see that L is continuous, we may assume that (5) holds identically. Since ν is unbounded, we may further assume that the set $\{\sigma_h^j;\ h, j > 0\}$ is dense in the interval $[0, L_\infty]$. If $\Delta L_t > 0$, there exist some $i, j, h > 0$ with $L_{t-} < \sigma_h^i < \sigma_h^j < L_{t+}$. By (6) we get $t - \varepsilon \le \tau_h^i < \tau_h^j \le t + \varepsilon$ for every $\varepsilon > 0$, which is impossible. Thus, $\Delta L_t = 0$ for all t.

To prove $\bar{\Xi} \subset \operatorname{supp} L$ a.s., we may further take $\bar{\Xi}_\omega$ to be perfect and nowhere dense for each $\omega \in \Omega$. If $t \in \bar{\Xi}$, then for every $\varepsilon > 0$ there exist some $i, j, h > 0$ with $t - \varepsilon < \tau_h^i < \tau_h^j < t + \varepsilon$. By (6) we get $L_{t-\varepsilon} \le \sigma_h^i < \sigma_h^j \le L_{t+\varepsilon}$, and so $L_{t-\varepsilon} < L_{t+\varepsilon}$ for all $\varepsilon > 0$, which implies $t \in \operatorname{supp} L$. □

In the perfect case, it remains to establish the a.s. relation $\Xi \cdot \lambda = c\,L$ for a constant $c \ge 0$, and to show that L is a.s. unique and adapted. The former claim is a consequence of Theorem 29.13 below, whereas the latter statement follows immediately from the next theorem.

Most standard constructions of L are special cases of the following result, where $\eta_t A$ denotes the number of excursions in a set $A \in D_0$ completed by time $t \ge 0$, so that η is an adapted, measure-valued process on D_0. The result may be compared with the global construction in Corollary 29.6.

Proposition 29.12 *(approximation) For sets $A_1, A_2, \ldots \in \mathcal{D}_0$ with finite measures $\nu A_n \to \infty$, we have*

(i) $\sup_{t \le u} \left| \dfrac{\eta_t A_n}{\nu A_n} - L_t \right| \xrightarrow{P} 0, \quad u \ge 0,$

(ii) *the convergence holds a.s. when the A_n are increasing.*

In particular, we have $\eta_t D_h / \nu D_h \to L_t$ a.s. as $h \to 0$ for fixed t, which shows that L is a.s. determined by the regenerative set Ξ.

Proof: Choose ξ as in Theorem 29.11, and put $\xi_s = \xi([0,s] \times \cdot)$. If the A_n are increasing, then for any unit rate Poisson process N on R_+, we have $(\xi_s A_n) \stackrel{d}{=} (N_{s\,\nu A_n})$ for all $s \ge 0$. Since $t^{-1} N_t \to 1$ a.s. by the law of large numbers, we get

$$\frac{\xi_s A_n}{\nu A_n} \to s \quad \text{a.s.}, \qquad s \ge 0.$$

By the monotonicity of each side, this may be strengthen to

$$\sup_{s \le r} \left| \frac{\xi_s A_n}{\nu A_n} - s \right| \to 0 \quad \text{a.s.}, \qquad r \ge 0.$$

The desired convergence now follows, by the continuity of L and the fact that $\xi_{L_t -} \le \eta_t \le \xi_{L_t}$ for all $t \ge 0$.

Dropping the monotonicity assumption, but keeping the condition $\nu A_n \uparrow \infty$ for convenience, we may choose a non-decreasing sequence $A'_n \in D_0$ such that $\nu A_n = \nu A'_n$ for all n. Then clearly $(\xi_s A_n) \stackrel{d}{=} (\xi_s A'_n)$ for fixed n, and so

$$\sup_{s \le r} \left| \frac{\xi_s A_n}{\nu A_n} - s \right| \stackrel{d}{=} \sup_{s \le r} \left| \frac{\xi_s A'_n}{\nu A'_n} - s \right| \to 0 \quad \text{a.s.}, \qquad r \ge 0,$$

which implies the convergence $\xrightarrow{P} 0$ on the left. The asserted convergence now follows as before. □

The excursion local time L may be described most conveniently in terms of its right-continuous inverse

$$T_s = L_s^{-1} = \inf\left\{t \ge 0;\, L_t > s\right\}, \qquad s \ge 0,$$

which will be shown to be a *generalized subordinator*, defined as a non-decreasing Lévy process with possibly infinite jumps. To state the next result, form $\Xi' \subset \Xi$ by omitting all points of Ξ that are isolated from the right. Recall that $l(u)$ denotes the length of an excursion path $u \in D_0$.

Theorem 29.13 *(inverse local time) Let L, ξ, ν, c be such as in Theorem* 29.11. *Then a.s.*

(i) $T = L^{-1}$ *is a generalized subordinator with characteristics $(c, \nu \circ l^{-1})$ and range Ξ' in R_+,*

(ii) $T_s = c\,s + \displaystyle\int_0^{s+} \!\!\int l(u)\, \xi(dr\, du), \quad s \ge 0.$

Proof: We may discard the P-null set where L fails to be continuous with support $\bar{\Xi}$. If $T_s < \infty$ for some $s \geq 0$, then $T_s \in \operatorname{supp} L = \bar{\Xi}$ by the definition of T, and since L is continuous, we get $T_s \notin \bar{\Xi} \setminus \Xi'$. Thus, $T(\mathsf{R}_+) \subset \Xi' \cup \{\infty\}$ a.s. Conversely, let $t \in \Xi'$. Then for any $\varepsilon > 0$ we have $L_{t+\varepsilon} > L_t$, and so $t \leq T \circ L_t \leq t + \varepsilon$. As $\varepsilon \to 0$, we get $T \circ L_t = t$. Thus, $\Xi' \subset T(\mathsf{R}_+)$ a.s.

The times T_s are optional by the right-continuity of $\mathcal{F}$. Furthermore, Proposition 29.12 shows that, as long as $T_s < \infty$, the process $\theta_s T - T_s$ is obtainable from $\theta_{T_s} X$ by a measurable map independent of s. Using the regenerative property at each T_s, we conclude from Lemma 16.13 that T is a generalized subordinator, hence admitting a representation as in Theorem 16.3. Since the jumps of T agree with the interval lengths in $(\bar{\Xi})^c$, we obtain (ii) for a constant $c \geq 0$.

The double integral in (ii) equals $\int x\,(\xi_s \circ l^{-1})(dx)$, which shows that T has Lévy measure $E(\xi_1 \circ l^{-1}) = \nu \circ l^{-1}$. Taking $s = L_t$ in (ii), we get a.s. for any $t \in \Xi'$

$$\begin{aligned} t &= T \circ L_t \\ &= c\,L_t + \int_0^{L_t+}\!\!\int l(u)\,\xi(dr\,du) \\ &= c\,L_t + (\Xi^c \cdot \lambda)_t, \end{aligned}$$

and so by subtraction $c\,L_t = (1_\Xi \cdot \lambda)_t$ a.s., which extends by continuity to arbitrary $t \geq 0$. Noting that a.s. $T^{-1}[0,t] = [0, L_t]$ or $[0, L_t)$ for all $t \geq 0$, since T is strictly increasing, we have a.s.

$$\xi[0,t] = L_t = \lambda \circ T^{-1}[0,t], \quad t \geq 0,$$

which extends to $\xi = \lambda \circ T^{-1}$ a.s. by Lemma 4.3. □

To justify our terminology, we show that the semi-martingale and excursion local times agree whenever both exist.

Proposition 29.14 *(reconciliation) Let X be a continuous semi-martingale with local time L, and let X be regenerative at some $a \in \mathsf{R}$ with $P\{L^a_\infty \neq 0\} > 0$. Then*

(i) $\Xi = \{t;\, X_t = a\}$ *is a.s. perfect and nowhere dense,*

(ii) L^a *is a version of the excursion local time at a.*

Proof: The regenerative set $\Xi = \{t;\, X_t = a\}$ is closed by the continuity of X. If Ξ were a countable union of closed intervals, then L^a would vanish a.s., contrary to our hypothesis, and so by Theorem 29.7 it is perfect and nowhere dense. Let L be a version of the excursion local time at a, and put $T = L^{-1}$. Define $Y_s = L^a \circ T_s$ for $s < L_\infty$, and put $Y_s = \infty$ otherwise. The continuity of L^a yields $Y_{s\pm} = L^a \circ T_{s\pm}$ for every $s < L_\infty$. If $\Delta T_s > 0$, then $L^a \circ T_{s-} = L^a \circ T_s$, since (T_{s-}, T_s) is an excursion interval of X and L^a is continuous with support in Ξ. Thus, Y is a.s. continuous on $[0, L_\infty)$.

By Corollary 29.6 and Proposition 29.12, the processes $\theta_s Y - Y_s$ can be obtained from $\theta_{T_s} X$ for all $s < L_\infty$ by a common measurable map. By the regenerative property at a, Y is then a generalized subordinator, and so by Theorem 16.3 and the continuity of Y, we have $Y_s \equiv c\,s$ a.s. on $[0, L_\infty)$ for a constant $c \geq 0$. For $t \in \Xi'$ we have a.s. $T \circ L_t = t$, and therefore

$$\begin{aligned} L_t^a &= L^a \circ (T \circ L_t) \\ &= (L^a \circ T) \circ L_t = c\,L_t, \end{aligned}$$

which extends to R_+ since both extremes are continuous with support in Ξ. □

For Brownian motion it is convenient to normalize local time by Tanaka's formula, which leads to a corresponding normalization of the excursion law ν. By the spatial homogeneity of Brownian motion, we may restrict our attention to excursions from 0. Then excursions of different lengths have the same distribution, apart from a scaling. For a precise statement, we define the scaling operators S_r on D by

$$(S_r f)_t = r^{1/2} f_{t/r}, \quad t \geq 0, \;\; r > 0, \;\; f \in D.$$

Theorem 29.15 *(Brownian excursion) Let ν be the normalized excursion law of Brownian motion. Then there exists a unique distribution $\hat{\nu}$ on the set of excursions of length one, such that*

$$\nu = (2\pi)^{-1/2} \int_0^\infty \left(\hat{\nu} \circ S_r^{-1}\right) r^{-3/2} dr. \tag{7}$$

Proof: By Theorem 29.13, the inverse local time L^{-1} is a subordinator with Lévy measure $\nu \circ l^{-1}$, where $l(u)$ denotes the length of u. Furthermore, $L \overset{d}{=} M$ by Corollary 29.3, where $M_t = \sup_{s \leq t} B_s$, and so by Theorem 16.12 the measure $\nu \circ l^{-1}$ has density $(2\pi)^{-1/2}\, r^{-3/2}$, $r > 0$. As in Theorem 8.5 there exists a probability kernel $(\nu_r)\colon (0,\infty) \to D_0$, such that $\nu_r \circ l^{-1} \equiv \delta_r$ and

$$\nu = (2\pi)^{-1/2} \int_0^\infty \nu_r r^{-3/2} dr, \tag{8}$$

and we note that the measures ν_r are unique a.e. λ.

For any $r > 0$, the process $\tilde{B} = S_r B$ is again a Brownian motion, and by Corollary 29.6 the local time of $\tilde{B}$ equals $\tilde{L} = S_r L$. If B has an excursion u ending at time t, then the corresponding excursion $S_r u$ of $\tilde{B}$ ends at rt, and the local time for $\tilde{B}$ at the new excursion equals $\tilde{L}_{rt} = r^{1/2} L_t$. Thus, the excursion process $\tilde{\xi}$ for $\tilde{B}$ is obtained from the process ξ for B through the mapping $T_r : (s, u) \mapsto (r^{1/2} s,\, S_r u)$. Since $\tilde{\xi} \overset{d}{=} \xi$, each T_r leaves the intensity measure $\lambda \otimes \nu$ invariant, and we get

$$\nu \circ S_r^{-1} = r^{1/2} \nu, \quad r > 0. \tag{9}$$

Combining (8) and (9), we get for any $r > 0$

$$\begin{aligned} \int_0^\infty \left(\nu_x \circ S_r^{-1}\right) x^{-3/2}\, dx &= r^{1/2} \int_0^\infty \nu_x\, x^{-3/2}\, dx \\ &= \int_0^\infty \nu_{rx}\, x^{-3/2}\, dx, \end{aligned}$$

and the uniqueness in (8) yields

$$\nu_x \circ S_r^{-1} = \nu_{rx}, \quad x > 0 \ \text{ a.e. } \lambda, \ r > 0.$$

By Fubini's theorem, we may then fix an $x = c > 0$ with

$$\nu_c \circ S_r^{-1} = \nu_{cr}, \quad r > 0 \ \text{ a.s. } \lambda.$$

Defining $\hat{\nu} = \nu_c \circ S_{1/c}^{-1}$, we conclude that for almost every $r > 0$,

$$\begin{aligned} \nu_r &= \nu_{c(r/c)} \\ &= \nu_c \circ S_{r/c}^{-1} \\ &= \nu_c \circ S_{1/c}^{-1} \circ S_r^{-1} \\ &= \hat{\nu} \circ S_r^{-1}. \end{aligned}$$

Substituting this into (8) yields equation (7).

If μ is another probability measure with the stated properties, then for almost every $r > 0$ we have $\mu \circ S_r^{-1} = \hat{\nu} \circ S_r^{-1}$, and hence

$$\begin{aligned} \mu &= \mu \circ S_r^{-1} \circ S_{1/r}^{-1} \\ &= \hat{\nu} \circ S_r^{-1} \circ S_{1/r}^{-1} = \hat{\nu}. \end{aligned}$$

Thus, $\hat{\nu}$ is unique. □

By the continuity of paths, an excursion of Brownian motion is either positive or negative, and by symmetry the two possibilities have the same probability $\frac{1}{2}$ under $\hat{\nu}$. This leads to the further decomposition $\hat{\nu} = \frac{1}{2}(\hat{\nu}_+ + \hat{\nu}_-)$. A process with distribution $\hat{\nu}_+$ is called a *(normalized) Brownian excursion.*

For subsequent needs, we insert a simple computational result.

Lemma 29.16 *(height distribution) Let ν be the excursion law of Brownian motion. Then*

$$\nu\Big\{u \in D_0;\ \sup_t u_t > h\Big\} = (2\,h)^{-1}, \quad h > 0.$$

Proof: By Tanaka's formula, the process

$$\begin{aligned} M &= 2B \vee 0 - L^0 \\ &= B + |B| - L^0 \end{aligned}$$

is a martingale, and so for $\tau = \inf\{t \geq 0;\ B_t = h\}$ we get

$$E\, L^0_{\tau\wedge t} = 2\, E\Big(B_{\tau\wedge t} \vee 0\Big), \quad t \geq 0.$$

Hence, by monotone and dominated convergence, $EL^0_\tau = 2E(B_\tau \vee 0) = 2\,h$. On the other hand, Theorem 29.11 shows that L^0_τ is exponentially distributed with mean $(\nu A_h)^{-1}$, where $A_h = \{u;\ \sup_t u_t \geq h\}$. □

We proceed to show that Brownian local time has some quite amazing spatial properties.

Theorem 29.17 *(space dependence, Ray, Knight) Let B be a Brownian motion with local time L, and put $\tau = \inf\{t > 0;\, B_t = 1\}$. Then the process $S_t = L_\tau^{1-t}$ on $[0,1]$ is a squared Bessel process of order* 2.

Several proofs are known. Here we derive the result from the previously developed excursion theory.

Proof (Walsh): Fix any $u \in [0,1]$, put $\sigma = L_\tau^u$, and let $\xi^\pm$ denote the Poisson processes of positive and negative excursions from u. Write Y for the process B, stopped when it first hits u. Then $Y \perp\!\!\!\perp (\xi^+, \xi^-)$ and $\xi^+ \perp\!\!\!\perp \xi^-$, and so $\xi^+ \perp\!\!\!\perp (\xi^-, Y)$. Since σ is ξ^+-measurable, we obtain $\xi^+ \perp\!\!\!\perp_\sigma (\xi^-, Y)$, and hence $\xi_\sigma^+ \perp\!\!\!\perp_\sigma (\xi_\sigma^-, Y)$, which implies the Markov property of L_τ^x at $x = u$.

To find the corresponding transition kernels, fix any $x \in [0, u)$, and write $h = u - x$. Put $\tau_0 = 0$, and let $\tau_1, \tau_2, \ldots$ be the right endpoints of excursions from x exceeding u. Further define $\zeta_k = L_{\tau_{k+1}}^x - L_{\tau_k}^x$, $k \geq 0$, so that $L_\tau^x = \zeta_0 + \cdots + \zeta_\kappa$ with $\kappa = \sup\{k;\, \tau_k \leq \tau\}$. By Lemma 29.16, the variables ζ_k are i.i.d. and exponentially distributed with mean $2h$. Since κ agrees with the number of completed u-excursions before time τ reaching x, and since $\sigma \perp\!\!\!\perp \xi^-$, we further see that κ is conditionally Poisson $\sigma/2h$, given σ.

Next we show that $(\sigma, \kappa) \perp\!\!\!\perp (\zeta_0, \zeta_1, \ldots)$. Then define $\sigma_k = L_{\tau_k}^u$. Since ξ^- is Poisson, we have $(\sigma_1, \sigma_2, \ldots) \perp\!\!\!\perp (\zeta_1, \zeta_2, \ldots)$, and so $(\sigma, \sigma_1, \sigma_2, \ldots) \perp\!\!\!\perp (Y, \zeta_1, \zeta_2, \ldots)$. The desired relation now follows, since κ is a measurable function of $(\sigma, \sigma_1, \sigma_2, \ldots)$, and ζ_0 depends measurably on Y.

For any $s \geq 0$, we may now compute

$$\begin{aligned} E\Big\{\exp(-sL_\tau^{u-h})\Big|\,\sigma\Big\} &= E\Big\{\Big(Ee^{-s\zeta_0}\Big)^{\kappa+1}\Big|\,\sigma\Big\} \\ &= E\Big\{(1+2sh)^{-\kappa-1}\Big|\,\sigma\Big\} \\ &= (1+2sh)^{-1}\exp\left(\frac{-s\,\sigma}{1+2sh}\right). \end{aligned}$$

By the Markov property of L_τ^x, the last relation is equivalent, via the substitutions $u = 1-t$ and $2s = (a-t)^{-1}$, to the martingale property, for every $a > 0$, of the process

$$M_t = (a-t)^{-1}\exp\left\{\frac{-L_\tau^{1-t}}{2(a-t)}\right\}, \quad t \in [0,a).$$

Now let X be a squared Bessel process of order 2, and note that $L_\tau^1 = X_0 = 0$ by Theorem 29.4. By Corollary 14.12, the process X is again Markov. To see that X has the same transition kernel as L_τ^{1-t}, it is enough to show that, for any $a > 0$, the process M remains a martingale when L_τ^{1-t} is replaced by X_t. This is clear from Itô's formula, if we note that X is a weak solution to the SDE $dX_t = 2X_t^{1/2}dB_t + 2dt$. $\square$

For an important application of the last result, we show that the local time is strictly positive on the range of the process.

Corollary 29.18 *(range and support) Let M be a continuous local martingale with local time L. Then outside a fixed P-null set,*

$$\{L_t^x > 0\} = \Big\{ \inf_{s\le t} M_s < x < \sup_{s\le t} M_s \Big\}, \quad x \in \mathsf{R},\ t \ge 0.$$

Proof: By Corollary 29.6 and the continuity of L, we have $L_t^x = 0$ for x outside the stated interval, except on a fixed P-null set. To see that $L_t^x > 0$ otherwise, we may reduce by Theorem 19.3 and Corollary 29.6 to the case where M is a Brownian motion B. Letting $\tau_u = \inf\{t \ge 0;\, B_t = u\}$, we see from Theorems 19.6 (i) and 18.16 that, outside a fixed P-null set,

$$L_{\tau_u}^x > 0, \quad 0 \le x < u \in \mathsf{Q}_+. \tag{10}$$

If $0 \le x < \sup_{s\le t} B_s$ for some t and x, there exists a $u \in \mathsf{Q}_+$ with $x < u < \sup_{s\le t} B_s$. But then $\tau_u < t$, and (10) yields $L_t^x \ge L_{\tau_u}^x > 0$. A similar argument applies to the case where $\inf_{s\le t} B_s < x \le 0$. □

Our third approach to local times is via additive functionals and their potentials. Here we consider a canonical Feller process X with state space S, associated terminal time ζ, probability measures P_x, transition operators T_t, shift operators θ_t, and filtration $\mathcal{F}$. By a *continuous, additive functional (CAF)* of X we mean a non-decreasing, continuous, adapted process A with $A_0 = 0$ and $A_{\zeta\vee t} \equiv A_\zeta$, satisfying

$$A_{s+t} = A_s + A_t \circ \theta_s \ \text{ a.s.}, \quad s, t \ge 0, \tag{11}$$

where the 'a.s.' without qualification means P_x-a.s. for every x. By the continuity of A, we may choose the exceptional null set to be independent of t. If it can also be chosen to be independent of s, then A is said to be *perfect*.

For a simple example, let $f \ge 0$ be a bounded, measurable function on S, and consider the associated *elementary CAF*

$$A_t = \int_0^t f(X_s)\, ds, \quad t \ge 0. \tag{12}$$

More generally, given a CAF A and a function f as above, we may define a new CAF $f \cdot A$ by $(f \cdot A)_t = \int_{s\le t} f(X_s)\, dA_s$, $t \ge 0$. A less trivial example is given by the local time of X at a fixed point x, whenever it exists in either sense discussed above.

For any CAF A and constant $\alpha \ge 0$, we may introduce the associated *α-potential*

$$U_A^\alpha(x) = E_x \int_0^\infty e^{-\alpha t}\, dA_t, \quad x \in S,$$

and put $U_A^\alpha f = U_{f\cdot A}^\alpha$. In the special case where $A_t \equiv t \wedge \zeta$, we shall often write $U^\alpha f = U_A^\alpha f$. Note in particular that $U_A^\alpha = U^\alpha f = R_\alpha f$ for the A in (12). If $\alpha = 0$, we may omit the superscript and write $U = U^0$ and $U_A = U_A^0$. We proceed to show that a CAF is determined by its α-potential, whenever the latter is finite.

Lemma 29.19 *(potentials of additive functionals) Let A, B be continuous, additive functionals of a Feller process X, and fix any $\alpha \geq 0$. Then*

$$U_A^\alpha = U_B^\alpha < \infty \quad \Rightarrow \quad A = B \ \ a.s.$$

Proof: Define $A_t^\alpha = \int_{s \leq t} e^{-\alpha s}\, dA_s$, and conclude from (11) and the Markov property at t that, for any $x \in S$,

$$\begin{aligned} E_x\Big(A_\infty^\alpha \,\Big|\, \mathcal{F}_t\Big) - A_t^\alpha &= e^{-\alpha t} E_x\Big(A_\infty^\alpha \circ \theta_t \,\Big|\, \mathcal{F}_t\Big) \\ &= e^{-\alpha t}\, U_A^\alpha(X_t). \end{aligned} \tag{13}$$

Comparing with the same relation for B, we see that $A^\alpha - B^\alpha$ is a continuous P_x-martingale of finite variation, and so $A^\alpha = B^\alpha$ a.s. P_x by Proposition 18.2. Since x was arbitrary, we get $A = B$ a.s. □

For any CAF A of Brownian motion in R^d, we may introduce the associated *Revuz measure* ν_A, given for any measurable function $g \geq 0$ on R^d by $\nu_A g = \bar{E}(g \cdot A)_1$, where $\bar{E} = \int E_x\, dx$. When A is defined by (12), we get in particular $\nu_A g = \langle f, g \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2(\mathsf{R}^d)$. In general, we need to show that ν_A is σ-finite.

Lemma 29.20 *(Revuz measure) For a continuous, additive functional A of Brownian motion X in R^d, the associated Revuz measure ν_A is σ-finite.*

Proof: Fix any integrable function $f > 0$ on R^d, and define

$$g(x) = E_x \int_0^\infty e^{-t-A_t} f(X_t)\, dt, \quad x \in \mathsf{R}^d.$$

Using Corollary 17.19, the additivity of A, and Fubini's theorem, we get

$$\begin{aligned} U_A^1 g(x) &= E_x \int_0^\infty e^{-t}\, dA_t\, E_{X_t} \int_0^\infty e^{-s-A_s} f(X_s)\, ds \\ &= E_x \int_0^\infty e^{-t}\, dA_t \int_0^\infty e^{-s-A_s \circ \theta_t} f(X_{s+t})\, ds \\ &= E_x \int_0^\infty e^{A_t}\, dA_t \int_t^\infty e^{-s-A_s} f(X_s)\, ds \\ &= E_x \int_0^\infty e^{-s-A_s} f(X_s)\, ds \int_0^s e^{A_t}\, dA_t \\ &= E_x \int_0^\infty e^{-s}\Big(1 - e^{-A_s}\Big) f(X_s)\, ds \\ &\leq E_0 \int_0^\infty e^{-s} f(X_s + x)\, ds. \end{aligned}$$

Hence, by Fubini's theorem,

$$\begin{aligned} e^{-1} \nu_A g &\leq \int U_A^1 g(x)\, dx \\ &\leq \int dx\, E_0 \int_0^\infty e^{-s} f(X_s + x)\, ds \\ &= E_0 \int_0^\infty e^{-s} ds \int f(X_s + x)\, dx \\ &= \int f(x) dx < \infty. \end{aligned}$$

The assertion now follows since $g > 0$. □

Now let $p_t(x)$ denote the transition density $(2\pi t)^{-d/2} e^{-|x|^2/2t}$ of Brownian motion in R^d, and put $u^\alpha(x) = \int_0^\infty e^{-\alpha t} p_t(x)\, dt$. For any measure μ on R^d, we may introduce the associated α*-potential* $U^\alpha \mu(x) = \int u^\alpha(x-y)\, \mu(dy)$. We show that the Revuz measure has the same potential as the underlying CAF.

Theorem 29.21 *(potentials of Revuz measure, Hunt, Revuz) Let A be a continuous, additive functional of Brownian motion in R^d with Revuz measure ν_A. Then*

$$U_A^\alpha = U^\alpha \nu_A, \quad \alpha \ge 0.$$

Proof: By monotone convergence we may take $\alpha > 0$. By Lemma 29.20, we may choose some positive functions $f_n \uparrow 1$ with $\nu_{f_n \cdot A} 1 = \nu_A f_n < \infty$ for each n, and by dominated convergence we have $U_{f_n \cdot A}^\alpha \uparrow U_A^\alpha$ and $U^\alpha \nu_{f_n \cdot A} \uparrow U^\alpha \nu_A$. Thus, we may further assume that ν_A is bounded. Then clearly $U_A^\alpha < \infty$ a.e.

Now fix any bounded, continuous function $f \ge 0$ on R^d, and conclude by dominated convergence that $U^\alpha f$ is again bounded and continuous. Writing $h = n^{-1}$ for an arbitrary $n \in \mathsf{N}$, we get by dominated convergence and the additivity of A

$$\begin{aligned} \nu_A U^\alpha f &= \bar{E} \int_0^1 U^\alpha f(X_s)\, dA_s \\ &= \lim_{n\to\infty} \bar{E} \sum_{j<n} U^\alpha f(X_{jh}) A_h \circ \theta_{jh}. \end{aligned}$$

Noting that the operator U^α is self-adjoint and using the Markov property, we may write the expression on the right as

$$\begin{aligned} \sum_{j<n} \bar{E}\, U^\alpha f(X_{jh}) E_{X_{jh}} A_h &= n \int U^\alpha f(x)\, E_x A_h\, dx \\ &= n \big\langle f,\, U^\alpha E.\, A_h \big\rangle. \end{aligned}$$

To estimate the function $U^\alpha E.\, A_h$ on the right, it is enough to consider arguments x with $U_A^\alpha(x) < \infty$. Using the Markov property of X and the additivity of A, we get

$$\begin{aligned} U^\alpha E.\, A_h(x) &= E_x \int_0^\infty e^{-\alpha s} E_{X_s} A_h\, ds \\ &= E_x \int_0^\infty e^{-\alpha s} (A_h \circ \theta_s)\, ds \\ &= E_x \int_0^\infty e^{-\alpha s} \Big(A_{s+h} - A_s \Big)\, ds \\ &= \Big(e^{\alpha h} - 1 \Big)\, E_x \int_0^\infty e^{-\alpha s} A_s\, ds - e^{\alpha h} E_x \int_0^h e^{-\alpha s} A_s\, ds. \end{aligned} \tag{14}$$

Integrating by parts gives

$$\begin{aligned} E_x \int_0^\infty e^{-\alpha s} A_s\, ds &= \alpha^{-1} E_x \int_0^\infty e^{-\alpha t}\, dA_t \\ &= \alpha^{-1}\, U_A^\alpha(x). \end{aligned}$$

Thus, as $n = h^{-1} \to \infty$, the first term on the right of (14) tends in the limit to $\langle f, U_A^\alpha \rangle$. The second term is negligible, since

$$\Big\langle f, E.\, A_h \Big\rangle \;\lesssim\; \bar{E} A_h \\ = \; h\,\nu_A 1 \to 0.$$

Hence,

$$\Big\langle U^\alpha \nu_A, f \Big\rangle = \nu_A U^\alpha f \\ = \langle\, U_A^\alpha, f \rangle,$$

and since f is arbitrary, we obtain $U_A^\alpha = U^\alpha \nu_A$ a.e.

To extend this to an identity, fix any $h > 0$ and $x \in \mathsf{R}^d$. Using the additivity of A, the Markov property at h, the a.e. version, Fubini's theorem, and the Chapman–Kolmogorov relation, we get

$$\begin{aligned} e^{\alpha h} E_x \int_h^\infty e^{-\alpha s} dA_s &= E_x \int_0^\infty e^{-\alpha s} dA_s \circ \theta_h \\ &= E_x\, U_A^\alpha(X_h) \\ &= E_x\, U^\alpha \nu_A(X_h) \\ &= \int \nu_A(dy)\, E_x u^\alpha(X_h - y) \\ &= e^{\alpha h} \int \nu_A(dy) \int_h^\infty e^{-\alpha s} p_s(x-y)\, ds. \end{aligned}$$

The required identity $U_A^\alpha(x) = U^\alpha \nu_A(x)$ now follows by monotone convergence as $h \to 0$. □

It follows easily that a CAF is determined by its Revuz measure.

Corollary 29.22 *(uniqueness) Let A, B be continuous, additive functionals of a Brownian motion in R^d. Then*

$$A = B \;\; a.s. \quad \Leftrightarrow \quad \nu_A = \nu_B.$$

Proof: By Lemma 29.20, we may take ν_A to be bounded, so that $U_A^\alpha < \infty$ a.e. for all $\alpha > 0$. Now ν_A determines U_A^α by Theorem 29.21, and the proof of Lemma 29.19 shows that U_A^α determines A a.s. P_x, whenever $U_A^\alpha(x) < \infty$. Since $P_x \circ X_h^{-1} \ll \lambda^d$ for each $h > 0$, it follows that $A \circ \theta_h$ is a.s. unique, and it remains to let $h \to 0$. □

We turn to the reverse problem of constructing a CAF associated with a given potential. To motivate the following definition, we may take expected values in (13) to get $e^{-\alpha t}\, T_t U_A^\alpha \le U_A^\alpha$. A function f on S is said to be *uniformly α-excessive*, if it is bounded and measurable with $0 \le e^{-\alpha t}\, T_t f \le f$ for all $t \ge 0$, and such that $\|T_t f - f\| \to 0$ as $t \to 0$, where $\|\cdot\|$ denotes the supremum norm.

Theorem 29.23 *(excessive functions and additive functionals, Volkonsky) Let X be a Feller process in S, fix any $\alpha > 0$, and let $f \ge 0$ be a uniformly α-excessive function on S. Then there exists an a.s. unique, perfect, continuous, additive functional A of X, such that $f = U_A^\alpha$.*

Proof: For any bounded, measurable function g on S, we get by Fubini's theorem and the Markov property of X

$$\begin{aligned}
\tfrac{1}{2} E_x\left|\int_0^\infty e^{-\alpha t} g(X_t)\,dt\right|^2 &= E_x \int_0^\infty e^{-\alpha t} g(X_t)\,dt \int_0^\infty e^{-\alpha(t+h)} g(X_{t+h})\,dh \\
&= E_x \int_0^\infty e^{-2\alpha t} g(X_t)\,dt \int_0^\infty e^{-\alpha h}\, T_h g(X_t)\,dh \\
&= E_x \int_0^\infty e^{-2\alpha t} g\, U^\alpha g(X_t)\,dt \\
&= \int_0^\infty e^{-2\alpha t}\, T_t g\, U^\alpha g(x)\,dt \\
&\le \|U^\alpha g\| \int_0^\infty e^{-\alpha t}\, T_t|g|(x)\,dt \\
&\le \|U^\alpha g\| \left\|U^\alpha |g|\right\|. \qquad (15)
\end{aligned}$$

Now introduce for each $h > 0$ the bounded, non-negative functions

$$\begin{aligned}
g_h &= h^{-1}\left(f - e^{-\alpha h}\, T_h f\right), \\
f_h &= U^\alpha g_h = h^{-1} \int_0^h e^{-\alpha s}\, T_s f\,ds,
\end{aligned}$$

and define

$$\begin{aligned}
A_h(t) &= \int_0^t g_h(X_s)\,ds, \\
M_h(t) &= A_h^\alpha(t) + e^{-\alpha t} f_h(X_t).
\end{aligned}$$

As in (13), we note that the processes M_h are martingales under P_x for every x. Using the continuity of the A_h, we get by Proposition 9.17 and (15), for $x \in S$ and as $h, k \to 0$,

$$\begin{aligned}
E_x\left(A_h^\alpha - A_k^\alpha\right)^{*2} &\lesssim E_x \sup_{t\in\mathbb{Q}_+}\left|M_h(t) - M_k(t)\right|^2 + \|f_h - f_k\|^2 \\
&\lesssim E_x\left|A_h^\alpha(\infty) - A_k^\alpha(\infty)\right|^2 + \|f_h - f_k\|^2 \\
&\lesssim \|f_h - f_k\|\,\|f_h + f_k\| + \|f_h - f_k\|^2 \to 0.
\end{aligned}$$

Hence, there exists a continuous process A independent of x, such that $E_x(A_h^\alpha - A^\alpha)^{*2} \to 0$ for every x.

For a suitable sequence $h_n \to 0$, we have $(A_{h_n}^\alpha \to A^\alpha)^* \to 0$ a.s. P_x for all x, and it follows easily that A is a.s. a perfect CAF. Taking limits in the relation $f_h(x) = E_x A_h^\alpha(\infty)$, we also note that $f(x) = E_x A^\alpha(\infty) = U_A^\alpha(x)$. Thus, A has α-potential f. □

The last result can be used to construct local times. Then say that a CAF A is *supported* by a set $B \subset S$, if its set of increase is a.s. contained in the closure of the set $\{t \ge 0;\, X_t \in B\}$. In particular, a non-zero, perfect CAF supported by a singleton set $\{x\}$ is called a *local time* at x. This terminology is clearly consistent with our earlier definitions of local time. Writing $\tau_x = \inf\{t > 0;\, X_t = x\}$, we say that x is *regular* (for itself) if $\tau_x = 0$ a.s. P_x. By Proposition 29.8, this holds iff, P_x-a.s., the random set $\Xi_x = \{t \ge 0;\, X_t = x\}$ has no isolated points.

Theorem 29.24 *(additive-functional local time, Blumenthal & Getoor) Let X be a Feller process in S, and fix any $a \in S$. Then*

(i) *X has a local time L at a iff a is regular,*

(ii) *L is then a.s. unique up to a normalization,*

(iii) *$U_L^1(x) = U_L^1(a)\, E_x e^{-\tau_a} < \infty, \quad x \in S$.*

Proof: (iii) Let L be a local time at a. Comparing with the renewal process L_n^{-1}, $n \in \mathsf{Z}_+$, we see that $\sup_{x,t} E_x(L_{t+h} - L_t) < \infty$ for every $h > 0$, which implies $U_L^1(x) < \infty$ for all x. By the strong Markov property at $\tau = \tau_a$, we get for any $x \in S$

$$\begin{aligned} U_L^1(x) &= E_x\Big(L_\infty^1 - L_\tau^1\Big) \\ &= E_x\, e^{-\tau}\Big(L_\infty^1 \circ \theta_\tau\Big) \\ &= E_x\, e^{-\tau} E_a L_\infty^1 \\ &= U_L^1(a)\, E_x e^{-\tau}. \end{aligned}$$

(ii) Use Lemma 29.19.

(i) Define $f(x) = E_x e^{-\tau}$, and note that f is bounded and measurable. Since $\tau \le t + \tau \circ \theta_t$, we also see from the Markov property at t that, for any $x \in S$,

$$\begin{aligned} f(x) &= E_x\, e^{-\tau} \\ &\ge e^{-t} E_x\Big(e^{-\tau} \circ \theta_t\Big) \\ &= e^{-t} E_x E_{X_t} e^{-\tau} \\ &= e^{-t} E_x f(X_t) \\ &= e^{-t}\, T_t f(x). \end{aligned}$$

Noting that $\sigma_t = t + \tau \circ \theta_t$ is non-decreasing and tends to 0 a.s. P_a as $t \to 0$, by the regularity of a, we further obtain

$$\begin{aligned} 0 &\le f(x) - e^{-h}\, T_h f(x) \\ &= E_x\Big(e^{-\tau} - e^{-\sigma_h}\Big) \\ &\le E_x\Big(e^{-\tau} - e^{-\sigma_{h+\tau}}\Big) \\ &= E_x\, e^{-\tau} E_a\Big(1 - e^{-\sigma_h}\Big) \\ &\le E_a\Big(1 - e^{-\sigma_h}\Big) \to 0. \end{aligned}$$

Thus, f is uniformly 1-excessive, and so by Theorem 29.23 there exists a perfect CAF L with $U_L^1 = f$.

To see that L is supported by the singleton $\{a\}$, we may write

$$\begin{aligned} E_x\Big(L_\infty^1 - L_\tau^1\Big) &= E_x\, e^{-\tau} E_a L_\infty^1 \\ &= E_x\, e^{-\tau} E_a e^{-\tau} \\ &= E_x\, e^{-\tau} = E_x L_\infty^1, \end{aligned}$$

which implies $L_\tau^1 = 0$ a.s. Hence, $L_\tau = 0$ a.s., and so the Markov property yields $L_{\sigma_t} = L_t$ a.s. for rational t. This shows that L has a.s. no point of increase outside the closure of $\{t \ge 0;\, X_t = a\}$. □

We may finally represent any CAF A of a one-dimensional Brownian motion as a unique mixture of local times. Recall that ν_A denotes the Revuz measure of A.

Theorem 29.25 *(additive functionals of Brownian motion, Volkonsky, McKean & Tanaka) Let X be a Brownian motion in* R *with local time L. Then a process A is a continuous additive functional of X iff a.s.*

$$A_t = \int_{-\infty}^{\infty} L_t^x \, \nu(dx), \quad t \ge 0, \tag{16}$$

for a locally finite measure ν on R*. The latter is then unique and equal to ν_A.*

Proof: For any measure ν, we may define an associated process A as in (16). If ν is locally finite, we see from the continuity of L and dominated convergence that A is a.s. continuous, hence a CAF. In the opposite case, ν is unbounded in every neighborhood of some point $a \in \mathsf{R}$. Under P_a and for any $t > 0$, the process L_t^x is further a.s. continuous and strictly positive near $x = a$. Hence, $A_t = \infty$ a.s. P_a, and A fails to be a CAF.

Next, we see from Fubini's theorem and Theorem 29.5 that

$$\begin{aligned} \bar{E} L_1^x &= \int dy\, E_y L_1^x \\ &= E_0 \int L_1^{x-y} dy = 1. \end{aligned}$$

Since L^x is supported by $\{x\}$, we get for any CAF A as in (16)

$$\begin{aligned} \nu_A f &= \bar{E}(f \cdot A)_1 \\ &= \bar{E} \int \nu(dx) \int_0^1 f(X_t)\, dL_t^x \\ &= \int f(x)\, \nu(dx)\, \bar{E} L_1^x = \nu f, \end{aligned}$$

which shows that $\nu = \nu_A$.

Now consider any CAF A. By Lemma 29.20, there exists a function $f > 0$ with $\nu_A f < \infty$. The process

$$\begin{aligned} B_t &= \int L_t^x \, \nu_{f \cdot A}(dx) \\ &= \int L_t^x \, f(x)\, \nu_A(dx), \quad t \ge 0, \end{aligned}$$

is then a CAF with $\nu_B = \nu_{f \cdot A}$, and Corollary 29.22 yields $B = f \cdot A$ a.s. Thus, $A = f^{-1} \cdot B$ a.s., and (16) follows. □

Exercises

1. Show that the set of increase of Brownian local time at 0 agrees a.s. with the zero set Ξ. Extend this to any continuous local martingale. (*Hint:* Apply Lemma 14.15 to the process $\mathrm{sgn}(B-) \cdot B$ in Theorem 29.1.)

2. (*Lévy*) Let M be the maximum process of a Brownian motion B. Show that B can be measurably recovered from $M - B$. (*Hint:* Use Corollaries 29.3 and 29.6.)

3. Use Corollary 29.3 to give a simple proof of the relation $\tau_2 \stackrel{d}{=} \tau_3$ in Theorem 14.16. (*Hint:* Recall that the maximum is unique by Lemma 14.15.) Also use Proposition 27.16 to give a direct proof of the relation $\tau_1 \stackrel{d}{=} \tau_2$. (*Hint:* Integrate separately over the positive and negative excursions of B, and use Lemma 14.15 to identify the minimum.)

4. Show that for any $c \in (0, \frac{1}{2})$, Brownian local time L_t^x is a.s. Hölder continuous in x with exponent c, uniformly for bounded t. Also show that the bound $c < \frac{1}{2}$ is best possible. (*Hint:* Apply Theorem 4.23 to the estimate in the proof of Theorem 29.4. For the last assertion, use Theorem 29.17.)

5. Consider a continuous local martingale $M = B \circ [M]$ a.s. for a Brownian motion B. Show that if B has local time L_t^x, then the local time of M at x equals $L^x \circ [M]$. (*Hint:* Use Theorem 29.5, and note that $L \circ [M]$ is jointly continuous.)

6. For a continuous semi-martingale X, show that $\int_0^t f(X_s, s)\, d[X]_s = \int dx \times \int_0^t f(x, s)\, dL_s^x$ outside a fixed null set. (*Hint:* Extend Theorem 29.5 by a monotone-class argument.)

7. Let Ξ be the zero set of Brownian motion B. Use Proposition 29.12 and Theorem 29.15 to construct its local time L directly from Ξ. Also use Lemma 29.16 to construct L from the heights of the excursions of B. Finally, use Corollary 29.6 to construct L from the occupation measure of B.

8. Let η be the maximum of a Brownian excursion. Show that $E\eta = (\pi/2)^{1/2}$. (*Hint:* Use Theorem 29.15 and Lemmas 29.16 and 4.4.)

9. Let L be the continuous local time of a continuous local martingale M with $[M]_\infty = \infty$ a.s. Show that a.s. $L_t^x \to \infty$ as $t \to \infty$, uniformly on compacts. (*Hint:* Reduce to the case of Brownian motion. Then use Corollary 29.18, the strong Markov property, and the law of large numbers.)

10. Show that the intersection of two regenerative sets is regenerative.

11. Let L be the local time of a regenerative set, and let τ be an independent, exponentially distributed time. Show that L_τ is again exponentially distributed. (*Hint:* Derive a Cauchy equation for the function $P\{L_\tau > s\}$.)

12. For an unbounded regenerative set Ξ, show that L is a.s. determined by Ξ. (*Hint:* Use the law of large numbers.)

13. Let Ξ be a non-trivial regenerative set. Show that $c\Xi \stackrel{d}{=} \Xi$ for all $c > 0$ iff the inverse local time is strictly stable.

14. Let X be a Feller process in R, and put $M_t = \sup_{s \le t} X_s$. Show that the points of increase of M form a regenerative set. Also prove the same statement for the process $X_t^* = \sup_{s \le t} |X_s|$ when $-X \stackrel{d}{=} X$.

15. Let X be a strictly stable Lévy process, let Ξ be the set of increase of the process $M_t = \sup_{s \le t} X_s$, and write L for the local time of Ξ. Assuming Ξ to be non-trivial, show that L^{-1} is strictly stable. Also prove the corresponding statement for X^* when X is symmetric.

16. Give an explicit construction of the process X in Theorem 29.11, based on the Poisson process ξ and the constant c. (*Hint:* Use Theorem 29.13 to construct the time scale.)

17. Show that the semi-martingale local time is preserved by a change of measure $Q = Z_t \cdot P$. Use this result to extend Corollary 29.18 to Brownian motion with a suitable drift. (*Hint:* Use Proposition 19.21 and Corollary 19.26.)

18. Show that the class of continuous additive functionals is preserved under a suitable change of measure $Q = Z_t \cdot P$. Use this result to extend Theorem 29.25 to a Brownian motion with drift.

Chapter 30

Random Measures, Smoothing and Scattering

Null arrays of random measures, weak and strong Poisson convergence, compound Poisson approximation, dissipative scattering, strong equivalence and tightness, thinning limits, Cox criterion, $0-\infty$ laws, weak and strong averaging, asymptotically invariant smoothing and scattering, convolution powers, constant-velocity transforms, independence and Cox property, line and flat processes, spanning and density criteria, cluster dichotomy, truncation and Palm tree criteria

Much of the probability theory developed so far has dealt with processes on the real line, where the discussion has often centered around Brownian motion and related processes. Here and in the next chapter, we return to the equally important and often neglected[1] theory of discrete point measures, where many results have connections to the class of Poisson and related processes.

The present treatment may be thought of as continuing our discussion of the latter processes in Chapter 15, though there are also strong connections with the elementary Poisson limit theorems in Chapter 6, the ordinary or discounted compensators in Chapter 10, and the weak convergence theory for random measures and point processes in Chapter 23. We have also seen the important role of random measures for the large deviation theory of Chapter 24 and the excursion theory of Chapter 29.

In this chapter, we will see how Poisson and Cox processes arise in various ways as limits of more general point processes. Here we begin with the basic weak and strong limit theorems for null arrays of point processes, extending the results for integer-valued random variables from Chapter 6. Next, we consider the Cox convergence of suitable point processes under dissipative shifts of the individual points. In the stationary case, we will prove versions for random measures of the multi-variate ergodic theorem in Chapter 25, and establish some smoothing limits under convolution, including a general version of Dobrushin's celebrated Cox convergence theorem. We conclude with conditions for the shift invariance of stationary line and flat processes, and study the cluster dichotomy for spatial branching processes, leading to a Palm tree criterion for stability of the generating cluster kernel.

---

[1]The discrete theory is often thought of by the ignorant as easier and more elementary, whereas the opposite is often true. Here and in the next chapter, we will encounter an abundance of deep and powerful theorems.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_31

The natural setting of the present theory is in terms of random measures. Recall that a *random measure* on a localized Borel space S is defined as a locally finite kernel ξ from the basic probability space $(\Omega, \mathcal{A}, P)$ into S. Thus, $\xi(\omega, B)$ is a locally finite measure in $B \in \mathcal{S}$ for fixed $\omega \in \Omega$, and a random variable in $\omega \in \Omega$ for each $B \in \hat{\mathcal{S}}$. Equivalently, we may regard ξ as a random element in the space $\mathcal{M}_S$ of locally finite measures μ on S, endowed with the σ-field generated by all evaluation maps $\mu \mapsto \mu B$ with $B \in \hat{\mathcal{S}}$. It is called a *point process* if $\xi B \in \mathsf{Z}_+$ a.s. for all $B \in \hat{\mathcal{S}}$, in which case $\xi = \sum_i \delta_{\sigma_i}$ a.s. for some random variables $\sigma_1, \sigma_2, \ldots$, and we say that ξ is *simple* if the σ_i are a.s. distinct, so that $\sup_s \xi\{s\} \le 1$ a.s.

The random measures ξ_{nj} with $n, j \in \mathsf{N}$ are said to form a *null array*, if they are independent in j for fixed n and such that, for each $B \in \hat{\mathcal{S}}$, the random variables $\xi_{nj} B$ form a null array in the sense of Chapter 6. As in Chapter 23, we write $\xrightarrow{vd}$ for convergence in distribution with respect to the vague topology on $\mathcal{M}_S$, as defined and studied in Appendix 5. Dissecting systems are defined in Chapter 1 and semi-rings in Chapter 23.

We begin with some criteria for Poisson convergence under independent superpositions, extending the elementary Poisson limit Theorem 6.7.

Theorem 30.1 *(weak Poisson convergence, Grigelionis) Let (ξ_{nj}) be a null array of point processes on S, let ξ be a Poisson process on S with $E\xi = \lambda$, and fix a dissecting semi-ring $\mathcal{I} \subset \hat{\mathcal{S}}_\lambda$. Then $\sum_j \xi_{nj} \xrightarrow{vd} \xi$ iff*

(i) $\sum_j P\{\xi_{nj} I = 1\} \to \lambda I, \quad I \in \mathcal{I},$

(ii) $\sum_j P\{\xi_{nj} B > 1\} \to 0, \quad B \in \hat{\mathcal{S}}.$

Just as for the main results in Chapter 7, the easiest and most transparent proof is via a compound Poisson approximation. Given a random measure ξ on S, we introduce an *associated compound Poisson process* $\tilde{\xi} = \int \mu\, \eta(d\mu)$ on S, where η is a Poisson process on $\mathcal{M}_S$ with intensity $E\eta = \mathcal{L}(\xi) = \lambda$. Writing $\pi_f : \mu \mapsto \mu f$, we note that $\tilde{\xi} f = \eta\, \pi_f$, and so by Lemma 15.2 (ii),

$$\begin{aligned} Ee^{-\tilde{\xi} f} &= Ee^{-\eta\,\pi_f} \\ &= \exp\Bigl\{-\lambda(1 - e^{-\pi_f})\Bigr\} \\ &= \exp\Bigl(Ee^{-\xi f} - 1\Bigr). \end{aligned} \tag{1}$$

For a null array (ξ_{nj}) of random measures, we choose the associated compound Poisson processes $\tilde{\xi}_{nj}$ to be independent in j for each n. Then $\tilde{\xi}_n = \sum_j \tilde{\xi}_{nj}$ is again compound Poisson with characteristic measure $\lambda_n = \sum_j \lambda_{nj}$, where $\lambda_{nj} = \mathcal{L}(\xi_{nj})$. By $\overset{vd}{\sim}$ we mean that, whenever either side converges along a sub-sequence $N' \subset \mathsf{N}$, so does the other side, and the two limits agree.

Lemma 30.2 *(compound Poisson approximation) Let (ξ_{nj}) be a null array of random measures on S with associated compound Poisson array $(\tilde{\xi}_{nj})$. Then*

$$\textstyle\sum_j \xi_{nj} \overset{vd}{\sim} \sum_j \tilde{\xi}_{nj}.$$

Proof: Write $\xi_n = \sum_j \xi_{nj}$ and $\tilde{\xi}_n = \sum_j \tilde{\xi}_{nj}$. By Theorem 23.16, it is enough to show that $Ee^{-\xi_n f} \sim Ee^{-\tilde{\xi}_n f}$ for every $f \in \hat{C}_S$, in the sense that convergence of either side implies that both sides converge to the same limit. By (1) we have

$$\begin{aligned} Ee^{-\tilde{\xi}_n f} &= \prod\nolimits_j Ee^{-\tilde{\xi}_{nj} f} \\ &= \prod\nolimits_j \exp\Big(Ee^{-\xi_{nj} f} - 1\Big). \end{aligned}$$

Writing $p_{nj} = 1 - Ee^{-\xi_{nj} f}$, we need to show that $\prod_j e^{-p_{nj}} \sim \prod_j (1 - p_{nj})$. Since $\sup_j p_{nj} \to 0$ by the 'null' property of $(\xi_{nj} f)$ and Lemma 6.6, this holds by a first order Taylor expansion of $\sum_j \log(1 - p_{nj})$. □

Proof of Theorem 30.1: Fix any disjoint sets $I_1, \dots, I_m \in \mathcal{I}$, and put $U = \bigcup_k I_k$. Writing $\rho_n = \sum_j \mathcal{L}(\xi_{nj}) = E\sum_j \tilde{\xi}_{nj}$, we get for any $k \le m$

$$\begin{aligned} \rho_n\Big\{\mu I_k = \mu U = 1\Big\} &= \rho_n\{\mu I_k = 1\} - \rho_n\Big\{\mu I_k = 1 < \mu U\Big\} \to \lambda I_k, \\ \rho_n\Big\{\sum\nolimits_k \mu I_k > 1\Big\} &= \rho_n\{\mu U > 1\} \to 0, \end{aligned}$$

which shows that $\rho_n \xrightarrow{v} \lambda \otimes \delta_1$. Since ξ and all $\sum_j \tilde{\xi}_{nj}$ are compound Poisson, we obtain $\sum_j \tilde{\xi}_{nj} \xrightarrow{d} \xi$, and $\sum_j \xi_{nj} \xrightarrow{d} \xi$ follows by Lemma 30.2. The reverse argument yields the necessity of (i) and (ii). □

The last theorem may be compared with the following strong version, where we use notations like $\xrightarrow{ul}$, $\xrightarrow{uld}$, $\xrightarrow{ulP}$, …, as explained in Chapter 23.

Theorem 30.3 *(strong Poisson convergence, Matthes et al.) Let (ξ_{nj}) be a null array of point processes on S, and let η be a Poisson process on S with $E\eta = \lambda$. Then $\sum_j \xi_{nj} \xrightarrow{uld} \eta$ whenever*

(i) $\sum_j E\xi_{nj} \xrightarrow{ul} \lambda$,

(ii) $\sum_j E\Big(\xi_{nj}B;\, \xi_{nj}B > 1\Big) \to 0, \quad B \in \hat{\mathcal{S}}$.

Lemma 30.4 *(Poisson approximation) Let (ξ_{nj}) be a null array of point processes on a bounded space S, and consider some Poisson processes η_n on S with intensities*

$$\lambda_n = \sum\nolimits_j E\Big(\xi_{nj};\, \|\xi_{nj}\| = 1\Big), \quad n \in \mathsf{N}.$$

Then $\sum_j \xi_{nj} \overset{ud}{\sim} \eta_n$ whenever

$$\sum\nolimits_j P\Big\{\|\xi_{nj}\| > 1\Big\} \to 0, \qquad \limsup_{n\to\infty} \|\lambda_n\| < \infty.$$

Proof: Putting $\lambda_{nj} = E(\xi_{nj};\, \|\xi_{nj}\| = 1)$, we may choose some independent Poisson processes η_{nj} on S with $E\eta_{nj} = \lambda_{nj}$, so that $\eta_n = \sum_j \eta_{nj}$ is again Poisson with $E\eta_n = \lambda_n$. Clearly

$$\begin{aligned} P\Big\{\|\eta_{nj}\| = 1\Big\} &= \|\lambda_{nj}\|\, e^{-\|\lambda_{nj}\|} \\ &= \|\lambda_{nj}\| + O(\|\lambda_{nj}\|^2), \\ P\Big\{\|\eta_{nj}\| > 1\Big\} &= 1 - (1 + \|\lambda_{nj}\|)\, e^{-\|\lambda\|} \\ &= O(\|\lambda_{nj}\|^2), \end{aligned}$$

and by Theorem 15.4,

$$E\Big(\eta_{nj}\,\Big|\,\|\eta_{nj}\|=1\Big)=E\Big(\xi_{nj}\,\Big|\,\|\xi_{nj}\|=1\Big).$$

Hence,

$$\begin{aligned}\Big\|\,\mathcal{L}(\xi_{nj})-\mathcal{L}(\eta_{nj})\,\Big\| &\lesssim \Big|P\Big\{\|\xi_{nj}\|=1\Big\}-P\Big\{\|\eta_{nj}\|=1\Big\}\Big|\\ &\qquad + P\Big\{\|\xi_{nj}\|>1\Big\}+P\Big\{\|\eta_{nj}\|>1\Big\}\\ &\lesssim P\Big\{\|\xi_{nj}\|>1\Big\}+\|\lambda_{nj}\|^2,\end{aligned}$$

and so by independence,

$$\begin{aligned}\Big\|\mathcal{L}\Big(\textstyle\sum_j\xi_{nj}\Big)-\mathcal{L}(\eta_n)\Big\| &\le \textstyle\sum_j\Big\|\mathcal{L}(\xi_{nj})-\mathcal{L}(\eta_{nj})\Big\|\\ &\lesssim \textstyle\sum_j\Big(P\Big\{\|\xi_{nj}\|>1\Big\}+\|\lambda_{nj}\|^2\Big)\\ &\le \textstyle\sum_j P\Big\{\|\xi_{nj}\|>1\Big\}+\|\lambda_{nj}\|\sup_j P\{\xi_{nj}\neq 0\},\end{aligned}$$

which tends to 0 as $n\to\infty$, by the stated hypotheses. □

Proof of Theorem 30.3: We may choose S to be bounded, since for general S we may then consider the restrictions to an arbitrary bounded set $B\in\hat{\mathcal{S}}$. Put $\xi_n=\sum_j\xi_{nj}$, and choose some Poisson processes η_n with intensities λ_n, as in Lemma 30.4. Then (ii) yields

$$\Big\|\textstyle\sum_j E\xi_{nj}-\lambda_n\Big\|\le\textstyle\sum_j E\Big(\|\xi_n\|;\,\|\xi_{nj}\|>1\Big)\to 0,$$

and so $\lambda_n\xrightarrow{u}\lambda$ by (i), which implies $\eta_n\xrightarrow{ud}\eta$ by Theorem 23.21 (ii). Furthermore, the hypotheses of Lemma 30.4 hold by (i) and (ii), and so $\xi_n\overset{ud}{\sim}\eta_n$. Hence, by combination,

$$\Big\|\mathcal{L}(\xi_n)-\mathcal{L}(\eta)\Big\|\le\Big\|\mathcal{L}(\xi_n)-\mathcal{L}(\eta_n)\Big\|+\Big\|\mathcal{L}(\eta_n)-\mathcal{L}(\eta)\Big\|\to 0,$$

which means that $\xi_n\xrightarrow{ud}\eta$. □

We turn to some limits under independent dispersal of the individual points. Say that the kernels $\nu_n: T\to S$ are *dissipative* if $\sup_t\nu_n(t,B)\to 0$ for all $B\in\hat{\mathcal{S}}$. Define $\xi_n\xrightarrow{ulP}\xi$ by $\|\xi_n-\xi\|_B\xrightarrow{P}0$ for all $B\in\hat{\mathcal{S}}$.

Theorem 30.5 *(dissipative scatter) For $n\in\mathsf{N}$, let ξ_n be a ν_n-transform of a point process η_n on T, for some dissipative kernels $\nu_n: T\to S$. Then for suitable ξ, η,*

(i) $\quad\eta_n\nu_n\xrightarrow{vd}\eta \;\Leftrightarrow\; \xi_n\xrightarrow{vd}\xi,$

(ii) $\quad\eta_n\nu_n\xrightarrow{ulP}\eta \;\Rightarrow\; \xi_n\xrightarrow{uld}\xi,$

in which case ξ is distributed as a Cox process directed by η.

The proof of part (ii) requires a lemma.

Lemma 30.6 *(strong equivalence and tightness) For η_n, ν_n as in Theorem 30.5, let ζ_n be Cox processes directed by the random measures $\eta_n\nu_n$. Then*

(i) *(ξ_n), (ζ_n), $(\eta_n\nu_n)$ are simultaneously tight,*

(ii) *$\xi_n \overset{uld}{\sim} \zeta_n$ holds whenever either side is tight.*

Proof: (i) For $t_n > 0$, write $s_n = 1 - e^{-t_n}$, so that $t_n = -\log(1 - s_n)$. By Lemma 15.2, we have for any $B \in \hat{\mathcal{S}}$

$$\begin{aligned} E\exp\Big(-t_n\zeta_n B\Big) &= E\exp\Big(-s_n\eta_n\,\nu_n B\Big),\\ E\exp\Big(-t_n\,\xi_n B\Big) &= E\exp\Big\{\eta_n\log\Big(1 - s_n\nu_n B\Big)\Big\}. \end{aligned}$$

When $t_n \to 0$, which holds iff $s_n \to 0$, all three conditions

$$t_n(\xi_n B) \xrightarrow{P} 0, \qquad t_n(\zeta_n B) \xrightarrow{P} 0, \qquad s_n(\eta_n\nu_n B) \xrightarrow{P} 0,$$

are equivalent, and the assertion follows by Lemma 5.9.

(ii) We may take S to be bounded, so that the sequence $\|\eta_n\nu_n\|$ is tight. To prove the equivalence $\xi_n \overset{ud}{\sim} \zeta_n$, it suffices for any sub-sequence $N' \subset \mathsf{N}$ to prove the required convergence along a further sub-sequence N''. By tightness we may then assume that $\|\eta_n\nu_n\|$ converges in distribution, and by Theorems 5.31 and 8.17 we may even let this convergence hold a.s. In particular, $\|\eta_n\nu_n\|$ is then a.s. bounded.

Assuming the appropriate conditional independence, we may now apply Lemma 30.4 to the conditional distributions of the sequences (ξ_n) and (ζ_n), given (η_n), to obtain

$$\Big\|\mathcal{L}(\xi_n|\,\eta_n) - \mathcal{L}(\zeta_n|\,\eta_n)\Big\| \to 0 \ \text{ a.s.},$$

which implies $\xi_n \overset{ud}{\sim} \zeta_n$, as in the proof of Theorem 23.21 (ii). □

Proof of Theorem 30.5: (i) By Lemma 15.2 (iii),

$$Ee^{-\xi_n f} = E\exp\Big(\eta_n\log\nu_n e^{-f}\Big), \quad f \in \mathcal{S}_+,\ n \in \mathsf{N}.$$

Now let f be supported by a fixed set $B \in \hat{\mathcal{S}}$, and write $g = 1 - e^{-f}$. Noting that for $t \in [0,1)$,

$$t \le -\log(1-t) \le t + O(t^2),$$

we get for any $n \in \mathsf{N}$

$$\begin{aligned} E\exp\Big(-\rho_n\eta_n\,\nu_n g\Big) &\le E\exp(-\xi_n f)\\ &\le E\exp\Big(-\eta_n\,\nu_n g\Big), \end{aligned}$$

where $1 \le \rho_n \to 1$ a.s. If $\eta_n\nu_n \xrightarrow{vd} \eta$, then even $\rho_n\eta_n\nu_n \xrightarrow{vd} \eta$, and so $Ee^{-\xi_n f} \to Ee^{-\eta g}$. Since f was arbitrary, Theorem 23.16 yields $\xi_n \xrightarrow{vd} \xi$, where ξ is a Cox process directed by η.

Conversely, let $\xi_n \xrightarrow{vd} \xi$ for a point process ξ on S. By Theorem 23.15 and Lemma 15.2, we get

$$\lim_{t\to 0}\inf_n Ee^{-t\,\xi_n B} = \sup_{K\in\mathcal{K}}\inf_n Ee^{-\xi_n(B\setminus K)} = 1, \quad B \in \hat{\mathcal{S}}.$$

Since $Ee^{-t'\eta_n\nu_n B} \geq Ee^{-t\,\xi_n B}$ with $t' = 1 - e^{-t}$ by Lemma 15.2, similar conditions hold for the random measures $\eta_n\nu_n$, which are then tight by Theorem 23.15. If $\eta_n\nu_n \xrightarrow{vd} \eta$ along a sub-sequence $N' \subset \mathsf{N}$, the direct assertion yields $\xi_n \xrightarrow{vd} \xi'$ along N', where ξ' is a Cox process directed by η. Since $\xi' \overset{d}{=} \xi$, the distribution of η is unique by Lemma 15.7, and so the convergence $\eta_n\nu_n \xrightarrow{vd} \eta$ extends to the original sequence.

(ii) If $\eta_n\nu_n \xrightarrow{ulP} \eta$, then $(\eta_n\nu_n)$ is tight, and so by Lemma 30.6 (ii) it suffices to show that $\xi_n \xrightarrow{uld} \xi$, which holds by Theorem 23.21 (ii). □

In particular, we get by Theorem 30.5 (i):

Corollary 30.7 *(thinning limits) For $n \in \mathsf{N}$, let ξ_n be a p_n-thinning of a point process η_n on S, where $p_n \to 0$. Then for suitable ξ, η,*

$$\xi_n \xrightarrow{vd} \xi \quad \Leftrightarrow \quad p_n\eta_n \xrightarrow{vd} \eta$$

in which case ξ is distributed as a Cox process directed by η.

This yields an interesting characterization of Cox processes.

Corollary 30.8 *(Cox criterion, Mecke) For a point process ξ on S, these conditions are equivalent:*

(i) *ξ is a Cox process directed by a random measure η,*

(ii) *for every $p \in (0,1)$, ξ is distributed as a p-thinning of a point process ξ_p.*

Then ξ_p is again Cox and directed by a random measure $\eta_p \overset{d}{=} \eta/p$.

Proof: If ξ and ξ_p are Cox processes directed by η and η/p, respectively, then Proposition 15.3 shows that ξ is distributed as a p-thinning of ξ_p. Conversely, if the stated condition holds for every $p \in (0,1)$, then ξ is Cox by Theorem 30.7. □

A random measure ξ on R^d is said to be *stationary* if $\theta_s\xi \overset{d}{=} \xi$ for every $s \in \mathsf{R}^d$, where the *shift operators* θ_s on $\mathcal{M}_{\mathsf{R}^d}$ are defined by $(\theta_s\mu)B = \mu(B - s)$ or $(\theta_s\mu)f = \mu(f \circ \theta_s)$. We begin with a general property of stationary random sets and measures.

Lemma 30.9 *($0-\infty$ laws)*

(i) *For a stationary random measure ξ on R or Z, we have a.s.*

$$\xi \neq 0 \quad \Leftrightarrow \quad \xi\mathsf{R}_\pm = \infty.$$

(ii) *For a stationary, random closed set φ in* R *or* Z, *we have a.s.*

$$\varphi \neq \emptyset \quad \Leftrightarrow \quad \sup(\pm\varphi) = \infty.$$

Proof: (i) By the stationarity of ξ and Fatou's lemma, we have for any $t \in \mathsf{R}$ and $h, \varepsilon > 0$

$$\begin{aligned} P\Big\{\xi[t, t+h) > \varepsilon\Big\} &= \limsup_{n\to\infty} P\Big\{\,\xi[(n-1)h, nh) > \varepsilon\Big\} \\ &\leq P\Big\{\,\xi[(n-1)h, nh) > \varepsilon \text{ i.o.}\Big\} \\ &\leq P\{\,\xi\mathsf{R}_+ = \infty\,\}. \end{aligned}$$

Letting $\varepsilon \to 0$, $h \to \infty$, and $t \to -\infty$ in this order, we get $P\{\xi \neq 0\} \leq P\{\xi\mathsf{R}_+ = \infty\}$. Since $\xi\mathsf{R}_+ = \infty$ implies $\xi \neq 0$, the two events agree a.s. By symmetry we have also $\xi\mathsf{R}_- = \infty$ a.s. on $\{\xi \neq 0\}$.

(ii) Here we may write instead

$$\begin{aligned} P\Big\{\varphi \cap [t, t+h) \neq \emptyset\Big\} &= \limsup_{n\to\infty} P\Big\{\varphi \cap [(n-1)h, nh) \neq \emptyset\Big\} \\ &\leq P\Big\{\varphi \cap [(n-1)h, nh) \neq \emptyset \text{ i.o.}\Big\} \\ &\leq P\{\sup \varphi = \infty\}. \end{aligned}$$

The assertion follows as before, as we let $h \to \infty$ and then $t \to -\infty$. □

For a stationary random measure ξ on R^d, we define the *invariant σ-field* by $\mathcal{I}_\xi = \xi^{-1}\mathcal{I}$, where $\mathcal{I}$ denotes the σ-field of shift-invariant, measurable sets in $\mathcal{M}_{\mathsf{R}^d}$. We further define the *sample intensity* of ξ as the $\bar{\mathsf{R}}_+$-valued random variable $\bar{\xi} = E(\xi B \,|\, \mathcal{I}_\xi)/\lambda^d B$, where $B \in \mathcal{B}^d$ with $\lambda^d B \in (0, \infty)$. This expression is independent of B, by the stationarity of ξ and Theorem 2.6.

We state a version for random measures of the multi-variate ergodic Theorem 25.14. Recall that, for a convex set $B \subset \mathsf{R}^d$, $r(B)$ denotes the radius of the largest open ball included in B. Put $I_1 = [0, 1]^d$.

Theorem 30.10 *(strong averaging, Nguyen & Zessin) Let ξ be a stationary random measure on R^d, and let $B_1 \subset B_2 \subset \cdots$ be bounded, convex sets in $\mathcal{B}^d$ with $r(B_n) \to \infty$. Then*

$$\frac{\xi B_n}{\lambda^d B_n} \to \bar{\xi} \text{ a.s.}$$

For any $p \geq 1$, the convergence extends to L^p whenever $\xi I_1 \in L^p$.

Proof: By Fubini's theorem, we have for any $A, B \in \mathcal{B}^d$

$$\begin{aligned} \int_B (\theta_s\xi)A\, ds &= \int_B ds \int 1_A(t-s)\, \xi(dt) \\ &= \int \xi(dt) \int_B 1_A(t-s)\, ds \\ &= \xi(1_A * 1_B). \end{aligned}$$

Assuming $|A| = 1$ and $A \subset S_a = \{s; |s| < a\}$, and putting $B^+ = B + S_a$ and $B^- = (B^c + S_a)^c$, we note that also $1_A * 1_{B^-} \le 1_B \le 1_A * 1_{B^+}$. Applying this to the sets $B = B_n$ gives

$$\frac{\lambda^d B_n^-}{\lambda^d B_n} \cdot \frac{\xi(1_A * 1_{B_n^-})}{\lambda^d B_n^-} \le \frac{\xi B_n}{\lambda^d B_n} \le \frac{\lambda^d B_n^+}{\lambda^d B_n} \cdot \frac{\xi(1_A * 1_{B_n^+})}{\lambda^d B_n^+}.$$

Since $r(B_n) \to \infty$, Lemma A6.5 (ii) yields $\lambda^d B_n^\pm / \lambda^d B_n \to 1$. Applying Theorem 25.14 to the function $f(\mu) = \mu A$ and the convex sets $B_n^\pm$, we obtain

$$\frac{\xi(1_A * 1_{B_n^\pm})}{\lambda^d B_n^\pm} \to E^{\mathcal{I}_\xi} \xi A = \bar{\xi},$$

in the appropriate sense. □

Turning to some weak ergodic theorems, we define a *weight function* $f : \mathsf{R}^d \to \mathsf{R}_+$ as a probability density on R^d. A sequence of such functions f_n is said to be *asymptotically invariant*, if the corresponding property holds for the measures $f_n \cdot \lambda^d$, in the sense of Chapter 3.

Theorem 30.11 *(weak averaging) Let ξ be a stationary random measure on R^d, let $f_1, f_2, \ldots$ be asymptotically invariant weight functions on R^d. Then for any $p > 1$,*

(i) $\xi f_n \xrightarrow{P} \bar{\xi}$,

(ii) $\xi f_n \to \bar{\xi}$ *in* L^1 $\Leftrightarrow$ $E\,\xi I_1 < \infty$,

(iii) $\xi f_n \to \bar{\xi}$ *in* L^p $\Leftrightarrow$ *the* $(\xi f_n)^p$ *are uniformly integrable.*

In particular, (iii) *holds when $\xi I_1 \in L^p$ and $f_n \precsim 1_{B_n}/\lambda^d B_n$ for some sets B_n as in Theorem 30.10.*

Proof: (ii) By Theorem 30.10, we may choose some weight functions g_m such that $\xi g_m \to \bar{\xi}$ in L^1. Using Minkowski's inequality, the stationarity of ξ, the invariance of $\bar{\xi}$, and dominated convergence, we get as $n \to \infty$ and then $m \to \infty$

$$\begin{aligned}\big\| \xi f_n - \bar{\xi} \big\|_1 &\le \big\| \xi f_n - \xi(f_n * g_m) \big\|_1 + \big\| \xi(f_n * g_m) - \bar{\xi} \big\|_1 \\ &\le E\,\xi \big| f_n - f_n * g_m \big| + \int \big\| \xi(g_m \circ \theta_s) - \bar{\xi} \big\|_1 f_n(s)\,ds \\ &\le E\,\bar{\xi} \int \lambda^d \big| f_n - f_n \circ \theta_t \big| \, g_m(t)\,dt + \big\| \xi g_m - \bar{\xi} \big\|_1 \to 0.\end{aligned}$$

(i) The point processes $\xi^r = 1\{\bar{\xi} \le r\}\xi$ with $r > 0$ are clearly stationary with

$$\begin{aligned}E\,\xi^r I_1 &= E\,\bar{\xi}^r \\ &= E\big(\bar{\xi};\, \bar{\xi} \le r\big) \le r, \quad r > 0,\end{aligned}$$

and so by (ii) we get $\xi^r f_n \xrightarrow{P} \bar{\xi}^r$, which implies $\xi f_n \xrightarrow{P} \bar{\xi}$ on $\{\bar{\xi} < \infty\}$. Next, we introduce the product-measurable processes

$$X_k(s) = 1\big\{\xi B_s^1 \le k\big\}, \quad s \in \mathsf{R}^d,\ k \in \mathsf{N},$$

and note that the random measures $\xi_k = X_k \cdot \xi$ are again stationary with $\bar{\xi} \precsim k$ a.s. As before, we get

$$\xi f_n \geq \xi_k f_n \xrightarrow{P} \bar{\xi}_k, \quad k \in \mathbb{N}. \tag{2}$$

Since $X_k \uparrow 1$, we have $\xi_k \uparrow \xi$ and even $\bar{\xi}_k \uparrow \bar{\xi}$ a.s., by successive monotone convergence. For any sub-sequence $N' \subset \mathbb{N}$, the convergence in (2) holds a.s. along a further sub-sequence N'', and so $\liminf_n \xi f_n \geq \bar{\xi}$ a.s. along N''. In particular, $\xi f_n \xrightarrow{P} \bar{\xi}$ remains true on $\{\bar{\xi} = \infty\}$.

(iii) This follows from (i) by Theorem 5.12. □

Next we consider smoothing by convolution. For the notions of weak and strong asymptotic invariance, see Chapter 3.

Theorem 30.12 *(invariant smoothing) Let ξ be a stationary random measure on $\mathbb{R}^d$ with $\bar{\xi} < \infty$ a.s., and let $\nu_1, \nu_2, \ldots$ be distributions on $\mathbb{R}^d$. Then*

(i) $\xi * \nu_n \xrightarrow{vP} \bar{\xi}\lambda^d$ *when the ν_n are weakly asymptotically invariant,*

(ii) $\xi * \nu_n \xrightarrow{ulP} \bar{\xi}\lambda^d$ *when the ν_n are asymptotically invariant.*

Proof: (i) By a simple approximation, we can choose a measure-determining sequence of continuous and compactly supported weight functions $f_k = \lambda_h * g_k$, where λ_h denotes the uniform distribution on I_h. By a compactness argument based on Theorem A5.6, we may ensure that the f_k are even convergence determining for the limit λ^d. Since the functions $\nu_n * \lambda_h * g_k$ are asymptotically invariant in n for fixed h and k, we get by Theorem 30.11 (i)

$$\begin{aligned}(\xi * \nu_n) f_k &= \big(\xi * \nu_n * \lambda_h\big) g_k \\ &= \xi\big(\nu_n * \lambda_h * g_k\big) \\ &\xrightarrow{P} \bar{\xi}\lambda^d g_k = \bar{\xi},\end{aligned}$$

which implies $\xi * \nu_n \xrightarrow{vP} \bar{\xi}\lambda^d$ by a sub-sequence argument.

(ii) Let the ν_n be asymptotically invariant. Then they are also weakly asymptotically invariant, and so by (i) we have $\xi * \nu_n \xrightarrow{vP} \bar{\xi}\lambda^d$. To see that even $\xi * \nu_n \xrightarrow{ulP} \bar{\xi}\lambda^d$, it suffices to prove, for any sub-sequence $N' \subset \mathbb{N}$, that $\xi * \nu_n \xrightarrow{ul} \bar{\xi}\lambda^d$ a.s. along a further sub-sequence $N'' \subset N'$. It is then enough to show that $\xi * \nu_n \xrightarrow{v} \bar{\xi}\lambda^d$ a.s. implies $\xi * \nu_n \xrightarrow{ul} \bar{\xi}\lambda^d$ a.s. along a sub-sequence. Replacing ξ by $\xi/\bar{\xi}$, we may further assume that $\bar{\xi} = 1$.

Now let $\xi * \nu_n \xrightarrow{v} \lambda^d$ a.s. Since the ν_n are asymptotically invariant, Lemma 3.17 yields $\|\nu_n - \nu_n * \rho\| \to 0$ for every probability measure ρ on $\mathbb{R}^d$. Choosing $\rho = f \cdot \lambda^d$ for a continuous function $f \geq 0$ with bounded support C and writing $\gamma_h = |\nu_n - \nu_n * \rho|$, we get for any $B \in \mathcal{B}^d$

$$\begin{aligned}E\big\|\, \xi * \nu_n - \xi * \nu_n * \rho \,\big\|_B &= E\big(\xi * \gamma_n\big) B \\ &= \|\gamma_n\|\, \lambda^d B \\ &= \big\|\, \nu_n - \nu_n * \rho \,\big\|\, \lambda^d B \to 0,\end{aligned}$$

and so $\|\xi * \nu_n - \xi * \nu_n * \rho\|_B \to 0$ a.s. as $n \to \infty$ along a sub-sequence.

By the assumptions on f, the a.s. convergence $\xi * \nu_n \xrightarrow{v} \lambda^d$ yields a.s.

$$\left(\xi * \nu_n * f\right)_x \to (\lambda^d * f)_x = 1, \quad x \in \mathsf{R}^d.$$

Furthermore, we have a.s. for any compact set $B \subset \mathsf{R}^d$

$$\begin{aligned}\limsup_{n\to\infty} \sup_{x\in B} \left(\xi * \nu_n * f\right)_x &\le \|f\| \limsup_{n\to\infty} (\xi * \nu_n)(B - C)\\ &\le \|f\| \lambda^d(B - C) < \infty.\end{aligned}$$

Hence, by dominated convergence,

$$\left\| \xi * \nu_n * \rho - \lambda^d \right\|_B = \int_B \left| \left(\xi * \nu_n * f\right)_x - 1 \right| dx \to 0.$$

Combining the two estimates, we get a.s. for any $B \in \mathcal{B}^d$

$$\left\| \xi * \nu_n - \lambda^d \right\|_B \le \left\| \xi * \nu_n - \xi * \nu_n * \rho \right\|_B + \left\| \xi * \nu_n * \rho - \lambda^d \right\|_B \to 0,$$

which means that $\xi * \nu_n \xrightarrow{ul} \lambda^d$ a.s. □

Combining Theorems 30.5 and 30.12, we obtain a fundamental limit theorem for asymptotically invariant transforms of point process.

Corollary 30.13 *(invariant scatter, Dobrushin) Let ξ be a stationary point process on R^d with $\bar{\xi} < \infty$ a.s., form some ν_n-transforms ξ_n of ξ based on distributions ν_n on R^d, and let ζ be a Cox process directed by $\bar{\xi}\lambda^d$. Then*

(i) *$\xi_n \xrightarrow{vd} \zeta$ when the ν_n are weakly asymptotically invariant,*

(ii) *$\xi_n \xrightarrow{uld} \zeta$ when the ν_n are asymptotically invariant.*

Proof: (i) Use Theorems 30.5 (i) and 30.12 (i).

(ii) Use Theorems 30.5 (ii) and 30.12 (ii). □

In particular we may let the points of ξ, thought of as particles in R^d, perform independent random walks based on a common distribution μ. Here we may give precise criteria for weak and strong asymptotic invariance. Let μ^{*n} denote the n-th convolution power of μ. Say that μ is of *lattice type*, if it is supported by a shifted, proper sub-group of R^d. By $\mu \perp \nu$ we mean that the measures μ and ν are mutually singular.

Theorem 30.14 *(convolution powers, Maruyama) For a distribution μ on R^d, we have*

(i) *the μ^{*n} are weakly asymptotically invariant $\Leftrightarrow$ μ is non-lattice,*

(ii) *the μ^{*n} are asymptotically invariant $\Leftrightarrow$ $\mu^{*m} \not\perp \lambda^d$ for some $m \in \mathsf{N}$.*

Proof: (i) If μ is non-lattice, we may write $\mu = \frac{1}{2}(\mu_1 + \mu_2)$, where μ_1 is non-lattice with bounded support. Then

$$\begin{aligned}\mu^{*n} &= 2^{-n} \sum_{k=0}^{n} \binom{n}{k} \mu_1^{*k} * \mu_2^{*(n-k)} \\ &= \mu_1^{*m} * \alpha_{n,m} + \beta_{n,m},\end{aligned}$$

where $1 - \|\alpha_{n,m}\| = \|\beta_{n,m}\| \to 0$ as $n \to \infty$ for fixed m. Since $\|\mu * \nu\| \le \|\mu\| \, \|\nu\|$ for any signed measures μ and ν, any weak asymptotic invariance of μ_1^{*n} would extend to μ^{*n}. This reduces the proof to the case of measures μ with bounded support. By a linear transformation combined with a shift, we may further assume that μ has mean 0 and covariances δ_{ij}.

By Theorem 6.11 and the estimate $|\partial_i \varphi^{*n}| \lesssim n^{-1-d/2}$, we have

$$n^{d/2} \left\| \mu^{*n} * \delta_x * p_h - \varphi^{*n} \right\| \to 0, \quad x \in \mathsf{R}^d, \; h > 0,$$

and so for any finite convex combination p of functions $\delta_x * p_h$,

$$n^{d/2} \left\| \left(\mu^{*n} * \delta_x - \mu^{*n} \right) * p \right\| \to 0, \quad x \in \mathsf{R}^d.$$

Writing $\rho = p \cdot \lambda^d$, we have for any $r > 0, \; x \in \mathsf{R}^d, \;$ and $n \in \mathsf{N}$

$$\begin{aligned}\left\| \left(\mu^{*n} * \delta_x - \mu^{*n} \right) * \rho \right\| \; &\lesssim \; r^d \, n^{d/2} \left\| \left(\mu^{*n} * \delta_x - \mu^{*n} \right) * p \right\| \\ &\quad + \left\{ \mu^{*n} * (\rho * \delta_x + \rho) \right\} I^c_{r\sqrt{n}} \, .\end{aligned}$$

Applying the central limit theorem with $\nu = \varphi \cdot \lambda^d$, we get

$$\limsup_{n\to\infty} \left\| \left(\mu^{*n} * \delta_x - \mu^{*n} \right) * \rho \right\| \lesssim 2 \, \nu I^c_r,$$

and $r > 0$ being arbitrary, we conclude that $\|(\mu^{*n} * \delta_x - \mu^{*n}) * \rho\| \to 0$.

For any $h > 0$, let λ_h be the uniform distribution on $[0,h]^d$. By an elementary approximation, we may choose some measures ρ_m as above with $\|\rho_m - \lambda_h\| \to 0$, and so

$$\left\| \left(\mu^{*n} * \delta_x - \mu^{*n} \right) * \lambda_h \right\| \le \left\| \left(\mu^{*n} * \delta_x - \mu^{*n} \right) * \rho_m \right\| + 2 \, \|\rho_m - \lambda_h\|,$$

which tends to 0 as $n \to \infty$ and then $m \to \infty$. The weak asymptotic invariance of μ^{*n} now follows by Lemma 3.18. The necessity is obvious.

(ii) Write $\mu^{*n} = \mu'_n + \mu''_n$ with $\mu'_n \ll \lambda^d \perp \mu''_n$, and note that $\|\mu'_n\|$ is non-decreasing, since $\mu'_n * \mu \ll \lambda^d$. If $\|\mu'_k\| > 0$ for some k, then $\|\mu''_{nk}\| \le \|\mu''_k\|^n \to 0$, and so $\|\mu''_n\| \to 0$. In particular, the μ^{*n} are non-lattice. For any $x \in \mathsf{R}^d$,

$$\begin{aligned}\left\| \mu^{*n} - \delta_x * \mu^{*n} \right\| &= \left\| \mu^{*k} * \left(\mu^{*(n-k)} - \delta_x * \mu^{*(n-k)} \right) \right\| \\ &\le \left\| \mu'_k * \left(\mu^{*(n-k)} - \delta_x * \mu^{*(n-k)} \right) \right\| + 2 \, \|\mu''_k\|,\end{aligned}$$

which tends to 0 by part (i), as $n \to \infty$ and then $k \to \infty$, proving the sufficiency of the stated condition. The necessity holds by Lemma 3.17. □

We turn to the special case where the individual points move independently with constant velocities. For $d = 1$, we may think of this as modeling the traffic flow along a highway. In a space-time diagram, the evolution is described by a line process in R^{d+1}, which gives a basic connection to stochastic geometry.

Corollary 30.15 *(constant velocities, Breiman, Stone) Let ξ be a stationary point process on R^d with $\bar{\xi} < \infty$ a.s., form some processes ξ_t, $t \geq 0$, by independently moving the points of ξ with fixed velocities of distribution μ, and let ζ be a Cox process directed by $\bar{\xi}\lambda^d$. Then as $t \to \infty$,*

(i) *the ξ_t are vaguely tight,*

(ii) *μ is diffuse $\Rightarrow$ all distributional limits are Cox,*

(iii) *μ is locally invariant $\Rightarrow$ $\xi_t \xrightarrow{vd} \zeta$,*

(iv) *$\mu \ll \lambda^d$ $\Rightarrow$ $\xi_t \xrightarrow{uld} \zeta$.*

Proof: For a random variable γ with distribution μ, the measures $\mu_t = \mathcal{L}(\gamma t)$ are weakly asymptotically invariant iff μ is locally invariant. Furthermore, Lemma 3.19 shows that the μ_t are strictly asymptotically invariant iff $\mu \ll \lambda^d$. Now (i) and (ii) follow by Theorem 30.13.

For general μ, all ξ_t have clearly the same sample intensity $\bar{\xi}$, and so the stated tightness follows by Theorem 23.15 and Chebyshev's inequality. If μ is diffuse, a simple compactness argument shows that the μ_t are dissipative as $t \to \infty$, and so all distributional limits are Cox by Theorem 30.5. $\square$

When ξ is Poisson, the independence property of the last theorem is clearly preserved at all times $t \geq 0$, since the entire particle system is then time stationary. The preservation fails in general:

Corollary 30.16 *(stationarity and independence) Let ξ be a stationary point process on R^d, and form the processes (ξ_t) as in Corollary 30.15 for a non-lattice distribution μ. Then for a fixed $t > 0$, these conditions are equivalent:*

(i) *the velocities at time t are i.i.d. and independent of ξ_t,*

(ii) *ξ is a Cox process directed by $\bar{\xi}\lambda^d$,*

(iii) *the line process (ξ_s) is space-time stationary.*

Proof, (i) $\Rightarrow$ (ii): Assuming (i) for a fixed $t > 0$, let ν_t' be the distribution of the associated displacements, and let ν_t'' be the distribution of the inverse shifts leading from ξ_t to ξ. Then the distribution of ξ is invariant under independent shifts with distribution $\mu = \nu_t' * \nu_t''$, which is again non-lattice. The invariance extends by iteration to displacements with distributions μ^{*n}, which are weakly asymptotically invariant by Theorem 30.14 (i). Now (ii) follows by Corollary 30.13 (i).

(ii) $\Rightarrow$ (iii): Use Theorem 15.3.

(iii) $\Rightarrow$ (i): Note that the independence is preserved by temporal shifts. $\square$

The previous results give a connection between particle systems in R^d and line processes in R^{d+1}. We turn to a study of more general flat processes. By a k*-flat* in R^d we mean an affine subspace x of dimension k, obtained by translation of a k-dimensional linear subspace πx of R^d, referred to as the *direction* of x. Write F_k for the space of k-flats in R^d and Φ_k for the subspace of flats through the origin.

Any reasonable parametrization makes F_k a localized Borel space, which allows us to consider point processes or more general random measures on F_k, where the former *flat processes* represent countable families of random flats in R^d. We assume such processes to give a.s. finite mass to the set of flats intersecting an arbitrary bounded set $A \subset \mathsf{R}^d$.

Translations on R^d induce shifts θ_x on F_k, and we say that a random measure η on F_k is *stationary* if $\theta_x \eta \overset{d}{=} \eta$ for all x, and *a.s. invariant* if $\theta_x \eta \equiv \eta$ a.s. Under suitable regularity conditions, we show that a stationary random measure on F_k is a.s. invariant.[2] By Theorem 31.18 below, this yields the remarkable conclusion that any stationary and sufficiently regular flat process is Cox. We consider two basic instances of the former statement.

Theorem 30.17 *(spanning criterion, Davidson, Krickeberg, OK) Let η be a stationary random measure on the space F_k of k-flats in R^d, where $k < d$. Then η is a.s. invariant whenever*[3]

(i) $\operatorname{span}(\pi x, \pi y) = \mathsf{R}^d$, $(x, y) \in F_k^2$ *a.e.* $E\eta^2$,

(ii) $E\eta^2$ *is locally finite.*

Proof: Choose the parametrization $x = (p, r) \in \Phi_k \times \mathsf{R}^{d-k}$ of F_k, where p denotes the direction πx and r is the perpendicular shift bringing πx to x, as specified by an appropriate sign convention. By (ii), Theorem 3.4 applies and yields a disintegration $E\eta^2 = \nu \otimes \mu$, in terms of a σ-finite measure ν on Φ_k^2 and a kernel $\mu \colon \Phi_k^2 \to \mathsf{R}^{2(d-k)}$. Thus, for measurable $f \geq 0$ on F_k^2,

$$E\eta^2 f = \iint \nu(dp\,dq) \iint \mu_{p,q}(dr\,ds) f(p, q, r, s).$$

Since η is stationarity, $E\eta^2$ is jointly invariant under shifts in F_k. Since the kernel μ is a.e. unique and the invariance is determined by countably many conditions, the joint invariance of $E\eta^2$ carries over to $\mu_{p,q}$, for ν-almost all pairs $(p, q) \in \Phi_k^2$. Condition (i) yields $\operatorname{span}(p, q) = \mathsf{R}^d$ a.e. ν.

For non-exceptional pairs (p, q), any shift $x \in (p \cap q)^\perp$ can be written uniquely as $x_p + x_q$, with $x_p \in p$ and $x_q \in q$. Thus, $\mu_{p,q}$ is a.e. invariant under the shifts $x_q \otimes x_p$, hence for all separate shifts in the two coordinates, and so $\mu_{p,q} = c_{p,q}\, \lambda^{2(d-k)}$ by Theorem 2.6. Absorbing the constants $c_{p,q}$ into ν gives

---

[2]Stationarity and invariance are often confused; here the distinction is of course crucial.

[3]With some effort, we can replace (ii) by a first-order moment condition; it is not known whether it can be omitted altogether.

the simplified disintegration $E\eta^2 = \nu \otimes \lambda^{2(d-k)}$. Thus, for any bounded Borel set $B \subset F_k$ and shift $x \in \mathsf{R}^d$,

$$\begin{aligned} E\eta^2\Big(B \times \theta_x^{-1} B\Big) &= \iint \nu(dp\, dq)\, \mu_{p,q}\Big(B_p \times \theta_x^{-1} B_q\Big) \\ &= \iint \nu(dp\, dq)\, \mu_{p,q}(B_p \times B_q) \\ &= E\eta^2 B^2, \end{aligned}$$

where B_p, B_q are the sections of B at p, q, respectively. Combining this with the joint invariance of $E\eta^2$, we obtain a.s.

$$\begin{aligned} &E\Big\{\eta B - \eta\Big(\theta_x^{-1} B\Big)\Big\}^2 \\ &\qquad = E\eta^2 B^2 + E\eta^2\Big(\theta_x^{-1} B\Big)^2 - 2E\eta^2\Big(B \times \theta_x^{-1} B\Big) = 0, \end{aligned}$$

and so $\eta B = \eta(\theta_x^{-1} B)$ a.s., which extends to $\eta = \theta_x \eta$ a.s., showing that η is a.s. invariant. □

The last criterion can only be fulfilled when $2\,k \geq d$, which excludes the important case of line processes in R^d for $d \geq 3$. The following criterion is more restrictive but applies to arbitrary k and d. Here we write λ_k for the unique, rotation invariant probability measure on the homogeneous space Φ_k. For any $u \in \Phi_h$, we say that a random measure η or process X is u*-stationary*, if it is stationary under shifts in u.

Theorem 30.18 *(density criterion, Papangelou, OK) Let η be a u-stationary random measure on the space F_k of k-flats in R^d, where $k < d$ and $u \in \Phi_{d-k+1}$ are fixed. Then η is a.s. invariant whenever*

$$\eta \circ \pi^{-1} \ll \lambda_k \;\; a.s..$$

This follows from a similar result for processes $X \geq 0$ on F_k:

Lemma 30.19 *(stationarity and invariance) Let X be a u-stationary, product-measurable process on F_k, for fixed $k < d$ and $u \in \Phi_{d-k+1}$. Then*

$$X_a = X_b \;\; a.s., \quad a, b \in \pi^{-1} t, \;\; t \in \Phi_k \;\; a.e.\ \lambda_k.$$

Proof: For any $t = t_0 \in \Phi_k$, we may choose some ortho-normal vectors $x_1, \ldots, x_{d-k} \in \mathsf{R}^d$ satisfying

$$\mathrm{span}\Big(x_1, \ldots, x_{d-k},\, t\Big) = \mathsf{R}^d, \qquad x_1, \ldots, x_{d-k} \not\perp t. \tag{3}$$

For fixed $x_1, \ldots, x_{d-k}$, the set G of flats $t \in \Phi_k$ satisfying (3) is clearly open with full λ_k-measure, containing in particular a compact neighborhood C of t_0. The compact space Φ_k is covered by finitely many such sets C with possibly different $x_1, \ldots, x_{d-k}$, and it is enough to consider the restriction of X to one of the sets $\pi^{-1}C$. Equivalently, we may take X to be supported by $\pi^{-1}C$ for a fixed C.

Fixing $x_1, \ldots, x_{d-k}$ accordingly, and letting $a, b \in \pi^{-1}t$ with $t \in C$, we have $a = b + x$ for a linear combination x of $x_1, \ldots, x_{d-k}$, and by iteration we may assume $x = r\, x_i$ for some $r \in \mathsf{R}$ and $i \leq d - k$. By symmetry we may take $i = 1$, and we may also choose $b = t$, so that $a = t + r\, x_1$.

The flats $t \in G$ may now be parametrized as follows, relative to the vector x_1. Writing $\hat{x}_1$ for the projection of x_1 onto t, we put $p = |\hat{x}_1| \in (0, 1)$. By continuity, p is bounded away from 0 and 1 on the compact set C. Next, we introduce the unique unit vector $y \in \mathrm{span}(x_1, \hat{x}_1)$ satisfying $y \perp x_1$ and $\langle y, \hat{x}_1 \rangle > 0$. Finally, let $s \in \Phi_{k-1}$ be the orthogonal complement of $\hat{x}_1$ in t. The map $t \mapsto (p, y, s)$ is $1{-}1$ and measurable on G, and by symmetry the vectors p, y, s are independent under λ_k, say with joint distribution μ.

The image $\hat{\lambda}_k = \lambda_k \circ p^{-1}$ is absolutely continuous on $[0, 1]$ with a positive density, and so for any $\varepsilon \in (0, \varepsilon_0]$ with a fixed $\varepsilon_0 > 0$, there exists a continuous function $g_\varepsilon \colon C \to \Phi_k$ that maps each $t = (p, y, s)$ into a flat $t' = (p', y, s)$ with $p' > p$ and $\hat{\lambda}_k(p, p') = \varepsilon$. By independence,

$$1_C \lambda_k \circ g_\varepsilon^{-1} = 1_{g_\varepsilon C}\, \lambda_k, \quad \varepsilon \in (0, \varepsilon_0). \tag{4}$$

Let p_0 be the maximum of p on C, and define $B_0 = \{(p, y, s);\, p \leq p_0\}$. By Lemma 1.37, any function $f \in L^2(\mu)$ supported by B_0 can be approximated in L^2 by continuous functions f_n with the same support. Combining with (4), we get as $n \to \infty$ for fixed $\varepsilon \in (0, \varepsilon_0)$

$$\begin{aligned} \left\| f \circ g_\varepsilon - f_n \circ g_\varepsilon \right\|_2^2 &= (\mu \circ g_\varepsilon^{-1})(f - f_n)^2 \\ &= \|f - f_n\|_2^2 \to 0. \end{aligned}$$

Furthermore, we get by continuity and dominated convergence

$$\lim_{\varepsilon \to 0} \left\| f_n - f_n \circ g_\varepsilon \right\|_2 \to 0, \quad n \in \mathsf{N}.$$

Using Minkowski's inequality, and letting $\varepsilon \to 0$ and then $n \to \infty$, we obtain

$$\left\| f - f \circ g_\varepsilon \right\|_2 \leq \left\| f_n - f_n \circ g_\varepsilon \right\|_2 + 2\, \|f - f_n\|_2 \to 0. \tag{5}$$

Truncating if necessary, we may take X to be bounded. Since it is also product-measurable, it is measurable on F_k for fixed $\omega \in \Omega$ by Lemma 1.28, and so by (5)

$$\lim_{\varepsilon \to 0} \int \mu(dt) \left| X(t) - X \circ g_\varepsilon(t) \right|^2 = 0.$$

Hence, by Fubini's theorem and dominated convergence,

$$\begin{aligned} &\int \mu(dt)\, E\left| X(t) - X \circ g_\varepsilon(t) \right|^2 \\ &\qquad = E \int \mu(dt) \left| X(t) - X \circ g_\varepsilon(t) \right|^2 \to 0, \end{aligned}$$

and so we may choose some $\varepsilon_n \to 0$ with

$$E\Big|X(t) - X \circ g_{\varepsilon_n}(t)\Big|^2 \to 0, \quad t \in C \text{ a.e. } \mu, \tag{6}$$

say for $t \in C'$.

For any $t \in C$ and $r \in \mathsf{R}$, the flats $g_\varepsilon(t)$ and $a = t + r\,x_1$ are non-random and lie in the $(k+1)$-dimensional span of $x_1, y,\, s$, and so their intersection is non-empty. Thus, the flat $t_n = g_{\varepsilon_n}(t)$ intersects both t and a, with flats of intersection parallel to s.

Now choose recursively some unit vectors $s_1, \ldots, s_{k-1}$, each uniformly distributed in the orthogonal complement of the preceding ones. The generated subspace s' has the same distribution as s under μ, and it is further a.s. linearly independent of u, which implies

$$\text{span}(u, s) = \mathsf{R}^d, \quad t \in C' \text{ a.e. } \mu,$$

say for $t \in C''$. For such a t, every flat in $\pi^{-1}s$ equals $s + x$ for an $x \in u$, and so by (6) and the u-stationarity of X,

$$\begin{aligned} E\Big|X(a) - X(t_n)\Big|^2 &= E\Big|X(t) - X(t_n)\Big|^2 \\ &= E\Big|X(t) - X \circ g_{\varepsilon_n}(t)\Big|^2 \to 0. \end{aligned}$$

Hence, Minkowski's inequality yields in $L^2(P)$

$$\Big\|X(a) - X(t)\Big\|_2 \le \Big\|X(a) - X(t_n)\Big\|_2 + \Big\|X(t_n) - X(t)\Big\|_2 \to 0,$$

and so $X(a) = X(t)$ a.s., as long as $t \in C''$. □

Proof of Theorem 30.18: Every flat $x \in F_k$ can be written uniquely as $x = \pi x + \pi' x$, where $\pi x \in \Phi_k$ and $\pi' x \in (\pi x)^\perp$, the orthogonal complement of πx in R^d. We may think of the pair $(\pi x, \pi' x)$ as a parametrization of x.

Now consider a random flat $\varphi \in \Phi_{d-k}$ with distribution λ_{d-k}. Since φ is generated by some ortho-normal vectors $\alpha_1, \ldots, \alpha_{d-k}$, each uniformly distributed in the orthogonal complement of the preceding ones, we have a.s.

$$\begin{aligned} \dim \text{span}(u^\perp, \varphi) &= \dim u^\perp + \dim \varphi \\ &= (k-1) + (d-k) = d-1, \end{aligned}$$

and so the intersection $u^\perp \cap \varphi$ has a.s. dimension 0. If a vector $y \in \varphi$ is not in the φ-projection of u, then $x \perp y$ for every $x \in u$, which means that $y \in u^\perp$. Assuming $y \in u^\perp \cap \varphi = \{0\}$, we obtain $y = 0$, which shows that the φ-projection of u agrees a.s. with φ itself. In particular, any measure $\nu \ll \lambda^{d-k+1}$ on u has a φ-projection $\nu' \ll \lambda^{d-k}$ a.s. on φ.

Let ν_ε be the uniform distribution on the ε-ball in u around 0. For any $r \in \Phi_k$, write ν_ε^r for the orthogonal projection of ν_ε onto the subspace $r^\perp \in \Phi_{d-k}$, so that $\nu_\varepsilon^r \ll \lambda^{d-k}$ on Φ_{d-k} for $r \in \Phi_k$ a.e. λ_k. Since $\eta \circ \pi^{-1} \ll \mu$ a.s., η admits a disintegration $\int \lambda_k(dr)\, \eta_r$ a.s., and so

$$\eta * \nu_\varepsilon = \int \lambda_k(dr)\left(\eta_r * \nu_\varepsilon^r\right) \text{ a.s.}, \quad \varepsilon > 0,$$

with convolutions in R^d on the left and in $\pi^{-1}r$ on the right, for each $r \in \Phi_k$. The u-stationarity of η carries over to $\eta * \nu_\varepsilon$, whereas the absolute continuity of ν_ε^r carries over to $\eta_r * \nu_\varepsilon^r$, and then also to the mixture $\eta * \nu_\varepsilon$.

For fixed $\varepsilon > 0$, a version of Theorem 2.15 yields a u-stationary, product-measurable process $X^\varepsilon \geq 0$ on F_k, such that $\eta * \nu_\varepsilon = X^\varepsilon \cdot (\lambda_k \otimes \lambda^{d-k})$ in a suitable parameter space $\Phi_k \times \mathsf{R}^{d-k}$. By Lemma 30.19,

$$X_a^\varepsilon = X_b^\varepsilon \text{ a.s.}, \quad a, b \in \pi^{-1}r, \;\; r \in \Phi_k \text{ a.e. } \lambda_k.$$

Since $\eta \circ \pi^{-1} \ll \lambda_k$, we may extend this to all $r \in \Phi_k$ by redefining $X = 0$ on the exceptional λ_k-set, which gives $X_a = X_b$ a.s. whenever $\pi a = \pi b$.

Now fix any bounded, measurable sets $I, J \in F_k$ with $J = I + h$ for some translation $h \in \mathsf{R}^{d-k}$ in the parameter space. Then

$$\begin{aligned} E\left|(\eta * \nu_\varepsilon)I - (\eta * \nu_\varepsilon)J\right| &= E\left|\int_I (X_s - X_{s+h})\,ds\right| \\ &\leq E\int_I \left|X_s - X_{s+h}\right| ds \\ &= \int_I E\left|X_s - X_{s+h}\right| ds = 0, \end{aligned}$$

with all integrations with respect to $\lambda_k \otimes \lambda^{d-k}$, and so $(\eta * \nu_\varepsilon)I = (\eta * \nu_\varepsilon)J$ a.s. Since I was arbitrary, $\eta * \nu_\varepsilon$ is then a.s. invariant under every fixed shift. Since also $\eta * \nu_\varepsilon \xrightarrow{v} \eta$ as $\varepsilon \to 0$, the measure η is a.s. h-invariant for fixed h. Applying this to a dense sequence of shifts $h_1, h_2, \ldots$, and using the vague continuity of $\eta * \delta_h$, we may extend the invariance to arbitrary h, for $\omega \in \Omega$ outside a fixed P-null set. □

We return to the subject of spatial branching processes from Chapter 13, except that now we allow an arbitrary dependence between branching and spatial motion, as codified by a *cluster kernel* on S, defined as a probability kernel $\nu\colon S \to \mathcal{N}_S$, where ν_s represents the offspring distribution of a particle at $s \in S$. Iterating ν, we obtain a discrete-time branching process in S, where the individuals in each generation give rise to independent sets of progeny distributed according to ν. This clearly defines a time-homogeneous Markov chain $\xi = (\xi_n)$ in $\mathcal{N}_S$. The kernel ν is said to be *critical* if $E_{\delta_s}\|\xi_1\| = 1$ for all $s \in S$, where $\|\xi_n\| = \xi_n S$.

For measurable groups $S = G$, we note as before that translations of G induce shifts on $\mathcal{N}_G$. If ξ_0 is stationary on G and ν is shift invariant, then each ξ_n is again stationary under shifts in G, as is in fact that entire branching tree $\xi = (\xi_n)$. In this case, the generation sizes $\|\xi_n\|$, when finite, form a classical Bienaymé process.

We also consider the *cluster invariance* $\xi_1 \overset{d}{=} \xi_0$, which by Lemma 11.11 makes the entire process ξ a space-time stationary Markov chain in $\mathcal{N}_S$. For processes ξ generated by a critical and shift-invariant cluster kernel ν on $S =$

$\mathbb{R}^d$, we may distinguish between two radically different kinds of asymptotic behavior, depending on the nature of ν. We say that a point process ξ on $\mathbb{R}^d$ is *cluster-invariant*, if the generated cluster process (ξ_n) is stationary in n. Note that (ξ_n) is vaguely tight when $E\bar{\xi} < \infty$, since $E\xi_n \equiv E\xi$.

Theorem 30.20 *(cluster dichotomy, Matthes et al.) Given a critical, invariant cluster kernel ν on $\mathbb{R}^d$, exactly one of these cases occurs, for all stationary point process ξ on $\mathbb{R}^d$ with $E\bar{\xi} < \infty$ generating some cluster processes (ξ_n):*

(i) *whenever $\xi_n \xrightarrow{vd} \eta$ along a sub-sequence*[4]*, we have $E\eta = E\xi$,*

(ii) $\xi_n \xrightarrow{vd} 0$.

Furthermore, (i) *is equivalent to*

(iii) *for every $c > 0$ there exists a cluster-invariant process ξ with $E\bar{\xi} = c$.*

We say that ν is *stable* if (i) holds, and otherwise *unstable*. For an intuitive understanding, we note that the branching tends to give rise to increasingly heavier but rarer 'clumps', which in turn get dispersed by the spatial motion. Here the scattering effect wins when ν is stable, whereas the clumping wins when ν is unstable. Further note that (i) holds by Lemma 5.11 iff $(\xi_n B)$ is uniformly integrable for every $B \in \mathcal{B}^d$.

Our proof requires some preliminaries. Given an invariant cluster kernel ν on $\mathbb{R}^d$, write ν^{*n} for the n-th iterate of ν, and consider a point process χ_n with distribution ν^{*n}, representing a cluster of age n rooted at the origin. For any point process ξ_0 on $\mathbb{R}^d$, let (ξ_n) be the generated branching process, and note that ξ_n is a sum of independent clusters of age n rooted at the points of ξ_0. Let κ_n^r be the number of such clusters in ξ_n charging the ball B_0^r. If ξ_0 is stationary with intensity c, then clearly

$$E\kappa_n^r = c\int P\Big\{\chi_n B_x^r > 0\Big\}\,dx < \infty, \quad n \in \mathbb{N},\ r > 0. \tag{7}$$

Lemma 30.21 *(monotonicity) For a critical, invariant cluster kernel ν on $\mathbb{R}^d$, we have*

(i) *$E\kappa_n^r$ is non-increasing in n for fixed $r > 0$,*

(ii) *the limit in* (i) *is either 0 for all r or > 0 for all r.*

Proof: (i) We may form χ_{n+1} by iterated clustering in n steps, starting from χ. Assuming χ to be simple, write χ_n^u for the associated cluster rooted at u. Using the conditional independence, Fubini's theorem, the invariance of λ^d, and the criticality of ν, we obtain

$$\begin{aligned}\int P\Big\{\chi_{n+1}B_x^r > 0\Big\}\,dx &= \int P\Big\{\int \chi(du)\,\chi_n^u B_x^r > 0\Big\}\,dx\\ &\le \int dx\,E\int \chi(du)\,1\Big\{\chi_n^u B_x^r > 0\Big\}\end{aligned}$$

[4]Under stronger regularity conditions, we have even $\xi_n \xrightarrow{vd} \eta$ for some cluster-invariant process η with $E\eta = E\xi$.

$$
\begin{aligned}
&= \int dx \int E\chi(du)\, P\Big\{\chi_n B^r_{x-u} > 0\Big\} \\
&= \int E\chi(du) \int P\Big\{\chi_n B^r_{x-u} > 0\Big\}\, dx \\
&= \int E\chi(du) \int P\Big\{\chi_n B^r_{x} > 0\Big\}\, dx \\
&= \int P\Big\{\chi_n B^r_{x} > 0\Big\}\, dx.
\end{aligned}
$$

For general χ, the same argument applies to any uniform randomizations of the processes χ_n.

(ii) This is clear since $E\chi_n^r$ is non-decreasing in r and B_0^r is covered by finitely many balls B_x^1. □

To prove Theorem 30.20, we may relate the stated properties to the limits

$$L_r = \lim_{n\to\infty} \int P\Big\{\chi_n B_x^r > 0\Big\}\, dx, \quad r > 0. \tag{8}$$

With stability interpreted for the moment as property (iii) of Theorem 30.20, we have the following criterion:

Lemma 30.22 *(hitting criterion, Matthes et al.) Let ν be a critical, invariant cluster kernel on R^d. Then*

$$\nu \text{ is stable} \quad \Leftrightarrow \quad L_r > 0, \ \ r > 0,$$

Partial proof of Theorem 30.20: Our first aim is to show that

$$\text{(ii)} \ \Leftrightarrow\ L_r \equiv 0, \qquad \text{(iii)} \ \Leftrightarrow\ L_r > 0.$$

Since (ii) and (iii) are incompatible, it is then enough to prove the implications to the left, which will also show that the two properties are exhaustive, establishing the dichotomy of Theorem 30.20 with (i) replaced by (iii).

$(L_r \equiv 0) \Rightarrow$ (ii): Consider a stationary point process ξ on R^d with intensity $c \in (0, \infty)$, generating a cluster process (ξ_n). If $L_r \equiv 0$, then (7) yields $\kappa_n^r \xrightarrow{P} 0$, and so $\xi_n \xrightarrow{vd} 0$, proving (ii).

$(L_r \neq 0) \Rightarrow$ (iii): Letting ξ be Poisson with intensity 1, we introduce some point processes η_n with distributions

$$\mathcal{L}(\eta_n) = n^{-1} \sum_{k=1}^{n} \mathcal{L}(\xi_k), \quad n \in \mathsf{N},$$

which are again stationary with intensity 1. By Theorem 23.15, we have convergence $\eta_n \xrightarrow{vd} \zeta$ along a sub-sequence $N' \subset \mathsf{N}$, where ζ is again stationary, and $E\zeta \le \lambda^d$ by Lemma 5.11. Since

$$\Big\| \mathcal{L}(\eta_n) - \mathcal{L}(\eta_{n-1}) \Big\| \le (n-1)\left(\frac{1}{n-1} - \frac{1}{n}\right) + \frac{1}{n} = \frac{2}{n} \to 0,$$

we have even $\eta_{n-1} \xrightarrow{vd} \zeta$ along N', which shows that ζ is ν-invariant.

Now let $L_r > 0$ for all $r > 0$. Fix an $r > 0$ with $\zeta\, \partial B_0^r = 0$ a.s., and write $p = L_r > 0$. Then Theorem 15.3 (i) yields

$$P\{\xi_n B_0^r = 0\} = \exp\Big(-\int P\{\chi_n B_x^r > 0\}\,dx\Big) \to e^{-p},$$

and so

$$P\{\eta_n B_0^r = 0\} = n^{-1}\sum_{k=1}^{n} P\{\xi_k B_0^r = 0\} \to e^{-p}.$$

Letting $n \to \infty$ along N', we obtain $P\{\zeta B_0^r = 0\} = e^{-p} < 1$, which shows that $E\zeta \neq 0$. Since the intensities of ξ_0 and ζ are proportional, property (iii) follows. □

To complete the proof of Theorem 30.20, it remains to show that (i) ⇔ (iii). Since (i) and (ii) are incompatible, we have in fact (i) ⇒ (iii), so we need only prove that (iii) ⇒ (i). Here we need a version of the following technical result.

Lemma 30.23 *(truncation criteria, Matthes et al.)* *Let ν be a critical, invariant cluster kernel on R^d, and fix any $r > 0$. Then*

(i) *ν is stable iff*
$$\lim_{k\to\infty}\inf_{n\geq 1} E\big\|\chi_n^{r,k}\big\| = 1,$$

(ii) *ν is unstable iff*
$$E\big\|\chi_n^{r,k}\big\| \to 0, \quad k > 0.$$

For the moment, we can only prove this with stability understood in the sense of condition (iii) in Theorem 30.20. Once the theorem is fully proven, the lemma will follow in its stated form.

Partial proof: Since the conditions on the right are incompatible while those on the left are exhaustive, it is enough to prove the necessity of the two conditions.

(i) Let ξ be a stationary, cluster-invariant point process on R^d with intensity in $(0,\infty)$, generating a cluster process (ξ_n). By stationarity and truncation of ξ_n or χ_n, we have for any $r,k,n > 0$

$$E\,\xi^{r,k} = E\,\xi_n^{r,k} \leq E\big\|\chi_n^{r,k}\big\|\,E\xi.$$

As $k \to \infty$, we get $E\,\xi^{r,k} \xrightarrow{v} E\xi$ by monotone convergence, and so

$$E\xi \leq \lim_{k\to\infty}\inf_{n\geq 1} E\big\|\chi_n^{r,k}\big\|\,E\xi.$$

Since $E\xi \asymp \lambda^d$, we may finally cancel the factor $E\xi$ on each side.

(ii) Using Fubini's theorem, the invariance of λ^d, and the equivalence of $u \in B_x^r$ and $x \in B_u^r$, we get for any $r,k,n > 0$

$$\begin{aligned}
\int P\{\chi_n B_x^r > 0\}\,dx &\geq \int P\{\chi_n^{r,k} B_x^r > 0\}\,dx \\
&\geq k^{-1}\int E\chi_n^{r,k} B_x^r\,dx \\
&= k^{-1} E\int \chi_n^{r,k}(du)\int 1\{x \in B_u^r\}\,dx \\
&= k^{-1} E\big\|\chi_n^{r,k}\big\|\,\lambda^d B_0^r.
\end{aligned}$$

If ν is unstable, the left-hand side tends to 0 as $n \to \infty$, and the stated condition follows. □

We may now complete the main proof:

Proof of Theorem 30.20, (iii) ⇒ (i): By monotone convergence we may replace ξ by $\xi^{r,k}$ for some fixed $r, k > 0$, and by Lemma 30.23 we may also replace χ_n by $\chi_n^{r,k}$. Letting κ_n be the number of truncated clusters in ξ_n hitting B_0^r, we have $\xi_n B_0^r \leq k\,\kappa_n$, and so it suffices to show that (κ_n) is uniformly integrable. By conditional independence, κ_n is the total mass of a p_n-thinning of the measure $\xi'(dx) = \xi(-dx)$, where

$$p_n^{r,k}(x) = P\Big\{\chi_n^{r,k} B_x^r > 0\Big\}, \quad x \in \mathsf{R}^d,$$

and so

$$E(\kappa_n)^2 \leq E\Big(\xi' p_n^{r,k}\Big)^2 + E\xi' p_n^{r,k}.$$

Writing g_r for the uniform probability density on B_0^r, we note that $p_n^{r,k} \leq p_n^{2r,k} * g_r$, and so

$$\begin{aligned} \xi' p_n^{r,k} &\leq \xi'\Big(p_n^{2r,k} * g_r\Big) \\ &= (\xi' * g_r) p_n^{2r,k} \\ &\lesssim \lambda^d p_n^{2r,k} \lesssim E\,\xi_n B_0^{2r} \\ &\lesssim \lambda^d B_0^{2r}. \end{aligned}$$

Thus, $E(\kappa_n)^2$ is bounded, and the required uniform integrability follows. □

Putting $\nu_n = \mathcal{L}(\chi_n)$, we may write the cluster iteration as

$$\nu_{m+n} = \nu_m \circ \nu_n, \quad m, n \in \mathsf{N}.$$

For every $n \in \mathsf{N}$, we introduce a point process η_n with the centered Palm distribution[5] of ν_n, given by

$$Ef(\eta_n) = E \int f\Big(\theta_{-x}\chi_n\Big)\,\chi_n(dx), \quad n \in \mathsf{N}.$$

The corresponding reduced Palm distributions $\nu_n^0 = \mathcal{L}(\eta_n - \delta_0)$ satisfy a similar recursive property:

Lemma 30.24 *(Palm recursion) Let $\nu_n = \mathcal{L}(\chi_n)$, $n \in \mathsf{N}$, with associated reduced, centered Palm distributions ν_n^0. Then*

$$\nu_{m+n}^0 = \Big(\nu_m^0 \circ \nu_n\Big) * \nu_n^0, \quad m, n \in \mathsf{N}.$$

[5]Though no knowledge of general Palm measures is needed here, some basic ideas from Chapter 31 may be helpful for motivation.

Proof: For any cluster kernels ν_1, ν_2 on R^d, we need to show that

$$(\nu_1 \circ \nu_2)^0 = \left(\nu_1^0 \circ \nu_2\right) * \nu_2^0. \tag{9}$$

When $d = 0$, the measures ν_i are determined by their generating functions $f_i(s) = E\, s^{\chi_i}$ on $[0,1]$, and the kernel composition $\nu = \nu_1 \circ \nu_2$ becomes equivalent to $f = f_1 \circ f_2$. The reduced Palm measure ν_i^0 has generating function $E\,\eta_i s^{\eta_i - 1} = f'(s)$, and (9) holds by the elementary chain rule for differentiation, in the form $(f_1 \circ f_2)' = (f_1' \circ f_2) f_2'$. The general case follows by suitable conditioning. □

Using the last result, we may construct recursively an increasing sequence of Palm trees η_n, along with a random walk $\zeta = (\zeta_n)$ in R^d based on the distribution $\rho = E\chi$, such that each η_n is rooted at $-\zeta_n$ and satisfies $\Delta\eta_n \perp\!\!\!\perp_{\zeta_n} \eta_{n-1}$. We may express the truncation criteria of Lemma 30.23 in terms of the Palm tree $\eta = (\eta_n)$. The results are useful to derive some explicit conditions for stability.[6]

Theorem 30.25 *(Palm tree criteria) Let ν be a critical, invariant cluster kernel on R^d with generated Palm tree (η_n), and fix any $r > 0$. Then*

(i) *ν is stable* $\Leftrightarrow$ $\sup_n \eta_n B_0^r < \infty$ *a.s.,*

(ii) *ν is unstable* $\Leftrightarrow$ $\eta_n B_0^r \to \infty$ *a.s.*

Proof: By the definitions of $\chi_n^{r,k}$ and η_n,

$$\begin{aligned} E\left\|\chi_n^{r,k}\right\| &= E\int 1\left\{\chi_n B_x^r \le k\right\} \chi_n(dx) \\ &= P\left\{\eta_n B_0^r \le k\right\}. \end{aligned}$$

Using the monotonicity of η_n, we get as $k \to \infty$

$$\begin{aligned} \inf_n E\left\|\chi_n^{r,k}\right\| &= \inf_n P\left\{\eta_n B_0^r \le k\right\} \\ &= P \textstyle\bigcap_n \left\{\eta_n B_0^r \le k\right\} \\ &= P\left\{\eta_\infty B_0^r \le k\right\} \\ &\to P\left\{\eta_\infty B_0^r < \infty\right\}. \end{aligned}$$

The stated criteria now follow by Lemma 30.23. □

Exercises

1. Show by examples that Theorem 30.1 fails in both directions if we replace (i) by $\sum_j E\,\xi_{nj} \xrightarrow{v} \lambda$. (*Hint:* We may take $\lambda = \delta_0$ in R.)

[6] From this point on the theory becomes very technical, and we refrain from pursuing those implications.

2. Show by examples that Theorem 30.1 fails in boths directions if we replace (i)−(ii) by the sole condition $\sum_j P\{\xi_{nj} I > 0\} \to \lambda I$ for all $I \in \mathcal{I}$. (*Hint:* Again we may take $\lambda = \delta_0$.)

3. Derive Theorem 30.1 from Theorem 6.7. (*Hint:* First let μ be diffuse and use Theorem 23.25. Then extend to the general case by a suitable randomization.)

4. Show that Theorems 30.1 and 30.3 fail if we drop the null-array assumption.

5. Give an example of a null array (ξ_{nj}) of point processes on S and a Poisson process ξ, such that $\sum_j \xi_{nj} \xrightarrow{vd} \xi$ but not $\sum_j \xi_{nj} \xrightarrow{uld} \xi$. (*Hint:* Take $S = \mathsf{R}$ with $E\xi = \lambda$, and choose the ξ_{nj} to be supported by Q.)

6. Show that the conditions in Theorem 30.5 (i) imply $(\xi_n, \eta_n \nu_n) \xrightarrow{vd} (\xi, \eta)$, with ξ and η related as stated. State and prove a similar extension of Corollary 30.7.

7. For an rcll, exchangeable process X on R_+, use Theorem 27.10 and Corollary 30.8 to show that the point process of jump sizes on $[0, 1]$ is Cox. Also conclude from Theorem 30.7 and the law of large numbers that the point process of jump times and sizes is Cox with directing random measure of the form $\nu \otimes \lambda$.

8. Give a second proof of Corollary 30.8, by approximating in the Laplace transforms of Lemma 15.2.

9. For a stationary sequence of random elements $\xi_1, \xi_2, \ldots$ in S and a set $B \in \mathcal{S}$, show that if $\xi_n \in B$ occurs for some n, then it holds for infinitely many n. (*Hint:* Use Lemma 30.9.)

10. Show that Lemma 30.9 can be strengthened to $\liminf_t t^{-1}\xi[0, t] > 0$ a.s. on $\xi \neq 0$. (*Hint:* Use Corollary 30.10.)

11. Give an example of a stationary, simple point process ξ on R^d with a.s. infinite sample intensity $\bar{\xi}$.

12. Give an example of a stationary random measure ξ and some distributions ν_n on R^d, such that $\xi * \nu_n \xrightarrow{vP} \bar{\xi}\lambda^d$ but not $\xi * \nu_n \xrightarrow{ulP} \bar{\xi}\lambda^d$. (*Hint:* Let ξ be a point process, and choose the ν_n to be weakly asymptotically invariant with support in Q^d.)

13. Give an example of a stationary point process ξ on R^d with ν_n-transforms ξ_n and a Cox process ζ directed by $\bar{\xi}\lambda^d$, such that $\xi_n \xrightarrow{vd} \zeta$ but not $\xi_n \xrightarrow{uld} \zeta$. (*Hint:* Choose the ν_n to be weakly asymptotically invariant with support in Q^d.)

14. In the context of Corollary 30.15, show that we may have $\xi_t \xrightarrow{vd} \zeta$ while $\xi_t \xrightarrow{uld} \zeta$ fails. (*Hint:* Let ξ be of lattice type, and choose μ to be locally invariant with $\mu \perp \lambda^d$.)

15. Show that if the process ξ in Corollary 30.15 is stationary Poisson, then the generated process (ξ_t) is space-time stationary for every choice of μ. (*Hint:* Show that (ξ_t) is again Poisson, using Theorem 15.3.)

16. For the random measure η in Theorem 30.17, show that if $E(\eta \circ \pi^{-1})^2 \ll \lambda_k^2$ for the homogeneous measure λ_k on Φ_k^d, then the spanning condition holds when $2k \geq d$ but fails for $2k < d$. Thus, for line processes in R^d, it holds when $d = 2$ but not for $d \geq 3$. (*Hint:* The set where the condition fails has dimension $< 2k$.)

17. Let (ξ_n) be a cluster process in R^d based on a critical, invariant cluster kernel ν, starting with a stationary Poisson process ξ_0 with intensity 1. Show that the process of ancestors of ξ_n at time 0 is again stationary Poisson, and find its asymptotic rate as $n \to \infty$. Also find the asymptotic size distribution of the cluster components of ξ_n. (*Hint:* Use Theorem 13.18.)

Chapter 31

Palm and Gibbs Kernels, Local Approximation

Palm kernel, supporting measure, invariant Palm disintegration, uniqueness and inversion, time–cycle duality, local approximation, Palm averages, mixed Poisson and binomial processes, Poisson criterion, conditioning and iteration, Palm–density duality, reduced Palm measures, compound Cambell measure, Gibbs and Papangelou kernels, invariance and Cox property, Palm–Gibbs duality, inner and outer conditioning

Palm measures may be regarded as extensions of regular conditional distributions, which explains their fundamental importance. Their study leads in particular to an amazingly rich theory of conditioning in point processes. We consider first the case of stationary random measures ξ, where there is only a single Palm measure, describing the local behavior of the entire process. Here a highlight of the theory is the fundamental theorem of Kaplan, giving a remarkable connection between stationary processes in discrete and continuous time.

Dropping the condition of stationarity, we get a whole family of Palm measures, one for each point in the underlying space S, combining into a kernel on S formed by disintegration of the underlying Campbell measure[1]. Among the key results in the non-stationary case, we note the local approximations in Theorem 31.9, justifying the interpretation of the Palm distribution of a point process ξ at a point $s \in S$ as the conditional distribution, given that ξ has a point at s. We further note the iteration approach in Theorem 31.12, where the Palm construction may be simplified through a preliminary conditioning, and the duality relation in Theorem 31.13, which enables us to derive smoothness properties of the Palm kernel from the possible smoothness of an associated conditional intensity.

We also call attention to the higher order Palm measures of a point process, obtained by disintegration of a *compound Campbell measure*, and allowing an interpretation in terms of *inner* conditioning. This leads naturally to the dual notion of *Gibbs kernel*, describing the corresponding *outer* conditioning. We further derive some basic characterizations involving the *Papangelou kernel*, obtained by restricting the Gibbs kernel to individual point masses.

[1]Note the analogy with the disintegration approach to conditional distributions in Theorem 8.5.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_32

Random measures have already appeared in various contexts in earlier chapters, including the more comprehensive treatments in Chapters 15 and 30. Here we just recall the definition of a random measure on S as a locally finite kernel from the basic probability space Ω into S, and a point process as an integer-valued random measure, supported by a locally finite random set. Given a random measure ξ on S and a random element η in T, we define the associated *Campbell measure* $C_{\xi,\eta}$ on $S \times T$ by

$$C_{\xi,\eta} f = E \int \xi(ds)\, f(s, \eta), \quad f \geq 0.$$

Theorem 31.1 *(Palm disintegration) Consider a random measure ξ on S and a random element η in T, such that $C_{\xi,\eta}$ is σ-finite. Then*

(i) *there exist a σ-finite measure $\nu \sim E\xi$ on S and a kernel $\mathcal{L}(\eta \,\|\, \xi)\colon S \to T$, such that*
$$C_{\xi,\eta} = \nu \otimes \mathcal{L}(\eta \,\|\, \xi),$$

(ii) *when $E\xi = \nu$ is σ-finite, we can choose $\mathcal{L}(\eta \,\|\, \xi)$ to be a probability kernel.*

Proof: Use Theorem 3.4. □

The σ-finiteness of the Campbell measure is not a serious restriction, since we can always replace η by the pair (ξ, η), for which it holds automatically. Putting $\mu = \mathcal{L}(\eta \,\|\, \xi)$, we can write the Palm disintegration as

$$E \int \xi(ds)\, f(s, \eta) = \int \nu(ds) \int \mu_s(dt)\, f(s, t), \quad f \geq 0,$$

just as for the conditional distributions in Theorem 8.5. Here $\mathcal{L}(\eta \,\|\, \xi)$ is called the *Palm kernel* of η with respect to ξ, and its values $\mathcal{L}(\eta \,\|\, \xi)_s$ are known as *Palm measures* or, in case of (ii), as *Palm distributions* at the points s. Furthermore, ν is called a *supporting measure* of ξ. When $\xi = \delta_\tau$ for a random element τ in S, we may choose $\nu = E\xi = \mathcal{L}(\tau)$ to obtain

$$\mathcal{L}(\eta \,\|\, \xi)_s = \mathcal{L}(\eta \,|\, \tau)_s, \quad s \in S \text{ a.e. } \mathcal{L}(\tau),$$

which justifies our notation and suggests that we think of the Palm measures as extensions of the regular conditional distributions from Chapter 8. Further justifications are provided by Theorems 31.5 and 31.9 below.

Now let G be a measurable group with Haar measure λ, acting measurably on the Borel spaces S and T. Given a random measure ξ on S and a random element η in T, we say that the pair (ξ, η) is G*-stationary* if $\theta_r(\xi, \eta) \stackrel{d}{=} (\xi, \eta)$ for all $r \in G$, where $\theta_r \eta = r\eta$ and $\theta_r \xi = \xi \circ \theta_r^{-1}$, so that

$$(\theta_r \xi) B = \xi(\theta_r^{-1} B), \qquad (\theta_r \xi) f = \xi(f \circ \theta_r).$$

If $S = G$, we assume $\xi B < \infty$ a.s. when $\lambda B < \infty$, which makes λ an invariant supporting measure for ξ. In particular we may choose $\eta = \xi$, or replace η by the pair (ξ, η).

If the pair (ξ, η) is jointly stationary with respect to a group G, we may choose the Palm disintegration to be G-invariant. When $S = G$, the entire Palm kernel $\mathcal{L}(\eta \,\|\, \xi)$ is determined by the single measure $\mathcal{L}(\eta \,\|\, \xi)_\iota$, then referred to as *the Palm measure* of η with respect to ξ.

Theorem 31.2 *(invariant Palm disintegration) Let G be a measurable group with Haar measure λ, acting properly on S and measurably on T, where S, T are Borel. Consider a random measure ξ on S and a random element η in T, where (ξ, η) is stationary and $C_{\xi,\eta}$ is σ-finite. Then*

(i) *there exist a σ-finite, G-invariant measure $\nu \sim E\xi$ on S and a G-invariant kernel $\mathcal{L}(\eta \,\|\, \xi)\colon S \to T$, such that*

$$C_{\xi,\eta} = \nu \otimes \mathcal{L}(\eta \,\|\, \xi),$$

(ii) *when $S = G$ we may choose $\nu = \lambda$, in which case $\mathcal{L}(\eta \,\|\, \xi)$ is given, for any $B \in \mathcal{G}$ with $\lambda B = 1$, by*

$$E\Big\{f(\eta) \,\Big\|\, \xi\Big\}_r = E\int_B f\Big(rp^{-1}\eta\Big)\,\xi(dp), \quad r \in G, \;\; f \geq 0. \tag{1}$$

Proof: Use Theorem 3.14. □

When $G = \mathsf{R}^d$ with Lebesgue measure λ^d, the Palm definition in Theorem 31.2 (ii) is essentially a $1-1$ correspondence, and we proceed to derive some useful inversion formulas. If ξ is a simple point process, then so is any 'Palm version' $\tilde{\xi}$ with distribution $\mathcal{L}(\xi \,\|\, \xi)_0$, and furthermore $\tilde{\xi}\{0\} = 1$. The inversion may then be stated in terms of the *Voronoi cell* around 0, given by

$$V_\mu = \Big\{x \in \mathsf{R}^d;\; \mu B_x^{|x|} = 0\Big\}, \quad \mu \in \mathcal{N}_{\mathsf{R}^d},$$

where B_x^r denotes the open ball around x of radius $r > 0$. When $d = 1$, the supporting points of μ may be enumerated in increasing order as $t_n(\mu)$, subject to the convention $t_0(\mu) \leq 0 < t_1(\mu)$.

To avoid repetitions, we assume throughout that R or R^d acts measurably on the Borel space T. We further write $E^\xi = E(\cdot \,\|\, \xi)_0$, for convenience.

Theorem 31.3 *(uniqueness and inversion) For a stationary pair of a random measure ξ on R^d and a random element η in T, the measure $\mathcal{L}(\xi, \eta;\, \xi \neq 0)$ is determined by $\mathcal{L}(\xi, \eta \,\|\, \xi)_0$, and for any measurable functions $f \geq 0$ on T and $g > 0$ on R^d with $\xi g < \infty$ a.s., we have*

(i)
$$E\Big\{f(\eta);\, \xi \neq 0\Big\} = E^\xi \int \frac{f(\theta_x \eta)}{\xi(g \circ \theta_x)}\, g(x)\, dx,$$

(ii) *when ξ is a simple point process on R^d,*

$$E\Big\{f(\eta);\, \xi \neq 0\Big\} = E^\xi \int_{V_0(\xi)} f(\theta_{-x}\eta)\, dx,$$

(iii) *when ξ is a simple point process on* R,

$$E\Big\{f(\eta);\, \xi \neq 0\Big\} = E^{\xi}\int_0^{t_1(\xi)} f(\theta_{-r}\eta)\,dr.$$

Replacing η by (ξ,η) yields inversion formulas for the Palm measures $\mathcal{L}(\xi,\eta\,\|\,\xi)_0$. A similar remark applies to subsequent theorems.

Proof: We may prove the stated formulas with η replaced by $\zeta = (\xi,\eta)$.

(i) Multiplying (1) by $\lambda^d B$ and extending by a monotone-class argument, we get for any measurable function $f \geq 0$

$$E^{\xi}\int f(\zeta,x)\,dx = E\int f\Big(\theta_{-x}\zeta, x\Big)\,\xi(dx),$$

and so by substitution,

$$E^{\xi}\int f\Big(\theta_x\zeta, x\Big)\,dx = E\int f(\zeta,x)\,\xi(dx). \tag{2}$$

In particular, we have for measurable $f, g \geq 0$

$$E^{\xi}\int f(\theta_x\zeta)\,g(x)\,dx = Ef(\zeta)\,\xi g.$$

Now choose $g > 0$ with $\xi g < \infty$ a.s., so that $\xi g > 0$ iff $\xi \neq 0$. The assertion follows as we replace f by the function

$$h(\mu,t) = \frac{f(\mu,t)}{\mu g}\,1\{\mu g > 0\}, \quad \mu \in \mathcal{M}_{\mathsf{R}^d},\ t \in T.$$

(ii) Apply (2) to the function

$$h(\mu,t,x) = f(\mu,t)\,1\Big\{\mu B_0^{|x|} = 0\Big\}.$$

Since $(\theta_x\mu)B_0^{|x|} = 0$ iff $-x \in V_0(\mu)$, and since a stationary point process $\xi \neq 0$ has an a.s. unique point closest to 0, we get

$$\begin{aligned} E^{\xi}\int_{V_0(\xi)} f(\theta_{-x}\zeta)\,dx &= E^{\xi}\int f(\theta_x\zeta)\,1\Big\{(\theta_x\xi)B_0^{|x|} = 0\Big\}\,dx \\ &= E\,f(\zeta)\int 1\Big\{\xi B_0^{|x|} = 0\Big\}\,\xi(dx) \\ &= E\Big\{f(\zeta);\, \xi \neq 0\Big\}. \end{aligned}$$

(iii) Apply (2) to the function

$$h(\mu,t,r) = f(\mu,t)\,1\Big\{t_0(\mu) = r\Big\}.$$

Since $t_0(\theta_r\xi) = r$ iff $-r \in [0, t_1(\xi))$, and since $|t_0(\xi)| < \infty$ a.s. when $\xi \neq 0$ by Theorem 30.9, we get

$$\begin{aligned} E^{\xi}\int_0^{t_1(\xi)} f(\theta_{-r}\zeta)\,dr &= E^{\xi}\int f(\theta_r\zeta)\,1\Big\{t_0(\theta_r\xi) = r\Big\}\,dr \\ &= E\,f(\zeta)\int 1\Big\{t_0(\xi) = r\Big\}\,\xi(dr) \\ &= E\Big\{f(\zeta);\, \xi \neq 0\Big\}. \end{aligned}$$

□

For pairs of a simple point process ξ on R and a random element η in T, we prove a striking relationship between stationarity in discrete and continuous time. Assuming $\xi\,\mathsf{R}_\pm = \infty$ a.s., we say that the pair (ξ,η) is *cycle-stationary* if $\theta_{-\tau_n}(\xi,\eta) \overset{d}{=} (\xi,\eta)$ for every $n \in \mathsf{Z}$, where $\tau_n = t_n(\xi)$. Informally, the excursions of (ξ,η) between the points of ξ then form a stationary sequence. To make the distinction clear, we may now refer to ordinary stationarity on R as *time stationarity.*

Theorem 31.4 *(time–cycle duality, Kaplan) For simple point processes ξ, $\tilde\xi$ on R and random elements η, $\tilde\eta$ in T, put $\zeta = (\xi,\eta)$ and $\tilde\zeta = (\tilde\xi,\tilde\eta)$. Then the relations*

(i) $$\tilde E f(\tilde\zeta) = E\int_0^1 f(\theta_{-r}\zeta)\,\xi(dr),$$

(ii) $$E f(\zeta) = \tilde E\int_0^{t_1(\tilde\xi)} f(\theta_{-r}\tilde\zeta)\,dr,$$

provide a $1{-}1$ *correspondence between all pseudo-distributions of time-stationary pairs ζ with $\xi \neq 0$ and cycle-stationary ones $\tilde\zeta$ with $\tilde\xi\{0\} = 1$ a.s. and $\tilde E\, t_1(\tilde\xi) = 1$.*

Proof: Let ζ be stationary with $\xi \neq 0$ a.s., put $\sigma_k = t_k(\xi)$, and define $\mathcal{L}(\tilde\zeta)$ by (i). Then for $n \in \mathsf{N}$ and any bounded, measurable function $f \geq 0$,

$$\begin{aligned} n\,\tilde E f(\tilde\zeta) &= E\int_0^n f(\theta_{-r}\zeta)\,\xi(dr) \\ &= E\sum_{\sigma_k \in (0,n)} f\big(\theta_{-\sigma_k}\zeta\big). \end{aligned}$$

Writing $\tau_k = t_k(\tilde\xi)$, we get by a suitable substitution

$$n\,\tilde E f\big(\theta_{-\tau_1}\tilde\zeta\big) = E\sum_{\sigma_k \in (0,n)} f\big(\theta_{-\sigma_{k+1}}\zeta\big),$$

and so by subtraction,

$$\Big|\tilde E f\big(\theta_{-\tau_1}\tilde\zeta\big) - \tilde E f(\tilde\zeta)\Big| \leq 2\,n^{-1}\|f\|.$$

As $n \to \infty$ we get $\tilde E f(\theta_{-\tau_1}\tilde\zeta) = \tilde E f(\tilde\zeta)$, and so $\theta_{-\tau_1}\tilde\zeta \overset{d}{=} \tilde\zeta$, which means that $\tilde\zeta$ is cycle-stationary. In this case, (ii) holds by Theorem 31.3 (iii). Taking $f \equiv 1$ gives $\tilde E t_1(\tilde\xi) = 1$.

Conversely, let $\tilde\zeta$ be cycle-stationary with $\tilde E\, t_1(\tilde\xi) = 1$, and define $\mathcal{L}(\zeta)$ by (ii). Then for n and f as above,

$$n\,E f(\zeta) = E\int_0^{\tau_n} f(\theta_{-r}\zeta)\,dr,$$

and so for any $h \in \mathsf{R}$,

$$\begin{aligned} n\,E f(\theta_{-h}\zeta) &= E\int_0^{\tau_n} f\big(\theta_{-r-h}\tilde\zeta\big)\,dr \\ &= E\int_h^{\tau_n+h} f\big(\theta_{-r}\tilde\zeta\big)\,dr, \end{aligned}$$

whence by subtraction,

$$\Big|E f(\theta_{-h}\zeta) - E f(\zeta)\Big| \leq 2\,n^{-1}|h|\,\|f\|.$$

As $n\to\infty$, we get $Ef(\theta_{-h}\zeta)=Ef(\zeta)$, and so $\theta_{-h}\zeta\overset{d}{=}\zeta$, which means that ζ is stationary.

Now choose $\tilde\zeta'=(\tilde\xi',\tilde\eta')$ as in (i), with associated expectation operator $\tilde E'$ satisfying

$$\tilde E' f(\tilde\zeta')=E\int_0^1 f(\theta_{-r}\zeta)\,\xi(dr). \tag{3}$$

Using Theorem 31.3 (iii) and comparing with (ii), we obtain

$$\tilde E'\int_0^{t_1(\tilde\xi')} f\Big(\theta_{-r}\tilde\zeta'\Big)\,dr=\tilde E\int_0^{t_1(\tilde\xi)} f\Big(\theta_{-r}\tilde\zeta\Big)\,dr.$$

Replacing $f(\mu,t)$ by $f\Big\{\theta_{-t_0(\mu)}(\mu,t)\Big\}$ and noting that $-t_0(\theta_{-r}\mu)=r$ when $\mu\{0\}=1$ and $r\in[0,t_1(\mu))$, we get

$$\tilde E' t_1(\tilde\xi')\,f(\tilde\zeta')=\tilde E\,t_1(\tilde\xi)\,f(\tilde\zeta).$$

A further substitution yields $\tilde E' f(\tilde\zeta')=\tilde E f(\tilde\zeta)$, and so (i) holds by (3). □

When ξ is a simple point process on R^d with locally finite intensity $E\xi$, we may think of the Palm distribution of (ξ,η) as the conditional distribution given that $\xi\{0\}=1$, which further justifies our notation $\mathcal{L}(\xi,\eta\,\|\,\xi)_0$. The stated interpretation is suggested by the following result, which also provides an asymptotic formula for the hitting probabilities of small Borel sets. By $B_n\to\{0\}$ we mean that $\sup\big(|s|;\,s\in B_n\big)\to 0$, and $\xrightarrow{u}$ denotes convergence in total variation norm $\|\cdot\|$. The 'hat' in $\hat{\mathcal{L}}$ or $\hat E$ denotes normalization.

Theorem 31.5 *(approximations at* 0*, Korolyuk, Ryll-Nardzewski, König & Matthes) Consider a stationary pair of a simple point process ξ on R^d and a random element η in T, and let $B_1,B_2,\ldots\in\mathcal{B}^d$ with $\lambda^d B_n>0$, $B_n\to\{0\}$, and $E\xi B_1<\infty$. When $\xi B_n=1$, define σ_n by $1_{B_n}\xi=\delta_{\sigma_n}$. Then*

(i) $P\{\xi B_n=1\}\sim P\{\xi B_n>0\}\sim E\xi B_n$,

(ii) $\mathcal{L}\Big(\theta_{-\sigma_n}\eta\,\Big|\,\xi B_n=1\Big)\xrightarrow{u}\hat{\mathcal{L}}(\eta\,\|\,\xi)_0$,

(iii) *for any bounded, measurable, and shift-continuous function f on $\mathcal{N}_d$,*

$$E\Big\{f(\eta)\,\Big|\,\xi B_n>0\Big\}\to\hat E\Big\{f(\eta)\,\Big\|\,\xi\Big\}_0.$$

Proof: (i) Write $\mathcal{L}(\xi,\eta\,\|\,\xi)_0=\mathcal{L}(\tilde\xi,\tilde\eta)$ for convenience. Since $\tilde\xi\{0\}=1$ a.s., we have $(\theta_x\tilde\xi)B_n>0$ iff $x\in B_n$, and so by Theorem 31.3 (ii),

$$\begin{aligned}\frac{P\{\xi B_n>0\}}{E\,\xi[0,1]^d}&=\hat E^{\,\xi}\int_{V_0(\xi)}1\Big\{(\theta_{-x}\xi)B_n>0\Big\}\,dx\\&\geq \hat E^{\,\xi}\lambda^d\Big\{V_0(\xi)\cap(-B_n)\Big\}.\end{aligned}$$

Dividing by $\lambda^d B_n$ and using Fatou's lemma, we get

$$\begin{aligned}\liminf_{n\to\infty}\frac{P\{\xi B_n>0\}}{E\,\xi B_n}&\geq\liminf_{n\to\infty}\frac{\hat E^{\,\xi}\lambda^d\Big\{V_0(\xi)\cap(-B_n)\Big\}}{\lambda^d B_n}\\&\geq\hat E^{\,\xi}\liminf_{n\to\infty}\frac{\lambda^d\Big\{V_0(\xi)\cap(-B_n)\Big\}}{\lambda^d B_n}=1.\end{aligned}$$

The assertions now follow from the elementary relations

$$\begin{aligned} 2 \cdot 1\{k > 0\} - k \le 1\{k = 1\} \\ \le 1\{k > 0\} \le k, \quad k \in \mathsf{Z}_+. \end{aligned}$$

(ii) On T we introduce the measures

$$\begin{aligned} \mu_n &= E \int_{B_n} 1\Big\{\theta_{-x}\eta \in \cdot\Big\}\, \xi(dx), \\ \nu_n &= \mathcal{L}\Big(\theta_{-\sigma_n}\eta;\, \xi B_n = 1\Big), \end{aligned}$$

and put $m_n = E\,\xi B_n$ and $p_n = P\{\xi B_n = 1\}$. By (1) the stated total variation equals

$$\begin{aligned} \left\| \frac{\nu_n}{p_n} - \frac{\mu_n}{m_n} \right\| &\le \left\| \frac{\nu_n}{p_n} - \frac{\nu_n}{m_n} \right\| + \left\| \frac{\nu_n}{m_n} - \frac{\mu_n}{m_n} \right\| \\ &\le p_n \left| \frac{1}{p_n} - \frac{1}{m_n} \right| + \frac{1}{m_n} |p_n - m_n| = 2 \left| 1 - \frac{p_n}{m_n} \right|, \end{aligned}$$

which tends to 0 in view of (i).

(iii) Here we write

$$\begin{aligned} &\Big| E\Big\{ f(\eta) \,\Big|\, \xi B_n > 0 \Big\} - \hat{E}\Big\{ f(\eta) \,\Big\|\, \xi \Big\}_0 \Big| \\ &\quad \le \Big| E\Big\{ f(\eta) \,\Big|\, \xi B_n > 0 \Big\} - E\Big\{ f(\eta) \,\Big|\, \xi B_n = 1 \Big\} \Big| \\ &\quad\quad + \Big| E\Big\{ f(\eta) - f(\theta_{-\sigma_n}\eta) \,\Big|\, \xi B_n = 1 \Big\} \Big| \\ &\quad\quad + \Big| E\Big\{ f(\theta_{-\sigma_n}\eta) \,\Big|\, \xi B_n = 1 \Big\} - \hat{E}\Big\{ f(\eta) \,\Big\|\, \xi \Big\}_0 \Big|. \end{aligned}$$

By (i) and (ii), the first and last terms on the right tend to 0 as $n \to \infty$. To estimate the second term, we introduce on T the bounded, measurable functions

$$g_\varepsilon(t) = \sup_{|x| < \varepsilon} \Big| f(\theta_{-x} t) - f(t) \Big|, \quad \varepsilon > 0,$$

and conclude from (ii) that, for n large enough,

$$\begin{aligned} \Big| E\Big\{ f(\eta) - f(\theta_{-\sigma_n}\eta) \,\Big|\, \xi B_n = 1 \Big\} \Big| &\le E\Big\{ g_\varepsilon(\theta_{-\sigma_n}\eta) \,\Big|\, \xi B_n = 1 \Big\} \\ &\to \hat{E}\Big\{ g_\varepsilon(\eta) \,\Big\|\, \xi \Big\}_0. \end{aligned}$$

Since $\hat{E}\, g_\varepsilon(\eta) \to 0$ as $\varepsilon \to 0$ by continuity and dominated convergence, the asserted convergence follows. □

We turn to some general ergodic theorems for Palm distributions, showing how the mappings $\mathcal{L}(\eta) \leftrightarrow \mathcal{L}(\tilde{\eta})$ can be achieved by suitable averaging. Given a function $f \in \mathcal{T}_+$ and a bounded measure $\nu \ne 0$ on R^d, we introduce the average

$$\bar{f}_\nu(t) = \|\nu\|^{-1} \int f(\theta_x t)\, \nu(dx), \quad t \in T.$$

Say that the *weight functions* (probability densities) $g_1, g_2, \ldots$ on R^d are *asymptotically invariant*, if the corresponding property holds for the associated measures $g_n \cdot \lambda^d$. We often omit the dot in $f \cdot \mu$.

Since such results apply to the proper Palm distributions, only when the sample intensity $\bar\xi$ is a.s. constant, we also need to introduce the *modified Palm distributions* $\bar{\mathcal{L}}(\eta \,\|\, \xi)_0$, defined for measurable functions $f \geq 0$ by

$$\bar E\Big\{f(\eta)\,\Big\|\,\xi\Big\}_0 = E\,\bar\xi^{-1}\!\int_{I_1} f(\theta_{-x}\eta)\,\xi(dx),$$

whenever $0 < \bar\xi < \infty$ a.s. For ergodic ξ we have $\bar\xi = E\bar\xi$ a.s., and $\bar{\mathcal{L}}(\eta \,\|\, \xi)_0$ agrees with $\mathcal{L}(\eta \,\|\, \xi)_0$. We may think of $\bar{\mathcal{L}}(\eta \,\|\, \xi)_0$ as the distribution of η when the pair (ξ, η) is shifted to a 'typical' point of ξ, an interpretation that fails for $\mathcal{L}(\eta \,\|\, \xi)_0$ in general. The pair $(\tilde\xi, \tilde\eta)$ is called a *modified Palm version* of (ξ, η) if $\mathcal{L}(\tilde\xi, \tilde\eta) = \bar{\mathcal{L}}(\xi, \eta \,\|\, \xi)_0$.

Theorem 31.6 *(Palm averaging) For a stationary pair of a random measure ξ on R^d with $\bar\xi \in (0, \infty)$ a.s. and a random element η in T, consider a modified Palm version $(\tilde\xi, \tilde\eta)$. Then for any asymptotically invariant distributions μ_n or weight functions p_n on R^d and for bounded, measurable functions f on T, we have*

(i) $\bar f_{\mu_n}(\tilde\eta) \xrightarrow{P} Ef(\eta)$,

(ii) $\bar f_{p_n\xi}(\eta) \xrightarrow{P} Ef(\tilde\eta)$.

The convergence holds a.s. when $\mu_n = 1_{B_n}\lambda^d$ or $p_n = 1_{B_n}$, respectively, for some bounded, convex, increasing sets $B_n \in \mathcal{B}^d$ with $r(B_n) \to \infty$.

Our proofs, here and below, are based on a remarkable coupling property.

Lemma 31.7 *(shift coupling, Thorisson) For pairs (ξ, η) and $(\tilde\xi, \tilde\eta)$ as in Theorem 31.6, there exist some random elements σ, τ in R^d with*

$$\eta \overset{d}{=} \theta_\sigma\tilde\eta, \qquad \tilde\eta \overset{d}{=} \theta_\tau\eta.$$

Proof: Let $\mathcal{I}$ be the invariant σ-field in $\mathcal{M}_{\mathsf{R}^d} \times T$, put $I_1 = [0,1]^d$, and note that $E(\xi I_1 \,|\, \mathcal{I}_{\xi,\eta}) = \bar\xi$ a.s. Then for any $I \in \mathcal{I}$,

$$\begin{aligned} P\{\tilde\eta \in I\} &= E\,\bar\xi^{-1}\!\int_{I_1} 1\Big\{\theta_{-x}\eta \in I\Big\}\,\xi(dx) \\ &= E\Big(\xi I_1/\bar\xi;\ \eta \in I\Big) \\ &= P\{\eta \in I\}, \end{aligned}$$

which shows that $\eta \overset{d}{=} \tilde\eta$ on $\mathcal{I}$. The assertions now follow by Theorem 25.26. □

Proof of Theorem 31.6: (i) By Lemma 31.7, we may assume that $\tilde\eta = \theta_\tau\eta$ for some random vector τ in R^d. Using Theorem 25.18 (i) and the asymptotic invariance of the μ_n, we get

$$\Big|\bar f_{\mu_n}(\tilde\eta) - Ef(\eta)\Big| \leq \Big\|\mu_n - \theta_\tau\mu_n\Big\|\,\|f\| + \Big|\bar f_{\mu_n}(\eta) - Ef(\eta)\Big| \xrightarrow{P} 0.$$

The a.s. version follows in the same way from Theorem 25.14.

(ii) For fixed f, we may define a stationary random measure ξ_f on R^d by

$$\xi_f B = \int_B f(\theta_{-x}\eta)\,\xi(dx), \quad B \in \mathcal{B}^d. \tag{4}$$

Applying Theorem 25.18 to both ξ and ξ_f and using the representation in Theorem 31.2, we get with $I_1 = [0,1]^d$

$$\begin{aligned}\bar f_{p_n\xi}(\eta) &= \frac{\xi_f p_n}{\lambda^d p_n} \cdot \frac{\lambda^d p_n}{\xi p_n} \xrightarrow{P} \frac{\bar\xi_f}{\bar\xi} \\ &= \frac{E\,\xi_f I_1}{E\,\xi I_1} = Ef(\tilde\eta).\end{aligned}$$

The a.s. version follows by a similar argument, based on Theorem 30.10. □

Taking expected values in Theorem 31.6, we get for bounded f the formulas

$$\begin{aligned} E\bar f_{\mu_n}(\tilde\eta) &\to Ef(\eta), \\ E\bar f_{p_n\xi}(\eta) &\to Ef(\tilde\eta),\end{aligned}$$

which may be regarded as limit theorems for suitable space averages of the distributions of η and $\tilde\eta$. We show that both statements hold uniformly for bounded f. For a striking formulation, we may introduce the possibly defective distributions $\bar{\mathcal{L}}_\mu(\tilde\eta)$ and $\bar{\mathcal{L}}_{p\,\xi}(\eta)$, given for measurable functions $f \geq 0$ by

$$\begin{aligned}\bar{\mathcal{L}}_\mu(\tilde\eta)f &= E\bar f_\mu(\tilde\eta), \\ \bar{\mathcal{L}}_{p\,\xi}(\eta)f &= E\bar f_{p\,\xi}(\eta).\end{aligned}$$

Theorem 31.8 *(distributional averages, Slivnyak, Zähle) For $(\xi,\eta), (\tilde\xi,\tilde\eta)$ as in Theorem 31.6, and for any asymptotically invariant distributions μ_n or weight functions p_n on R^d, we have*

(i) $\bar{\mathcal{L}}_{\mu_n}(\tilde\eta) \xrightarrow{u} \mathcal{L}(\eta)$,

(ii) $\bar{\mathcal{L}}_{p_n\xi}(\eta) \xrightarrow{u} \mathcal{L}(\tilde\eta)$.

Proof: (i) By Lemma 31.7, we may assume that $\tilde\eta = \theta_\tau\eta$ for some random vector τ. Using Fubini's theorem and the stationarity of η, we get for any measurable function $f \geq 0$

$$\begin{aligned}\bar{\mathcal{L}}_{\mu_n}(\eta)f &= \int Ef(\theta_x\eta)\,\mu_n(dx) \\ &= Ef(\eta) = \mathcal{L}(\eta)f.\end{aligned}$$

Hence, by Fubini's theorem and dominated convergence,

$$\begin{aligned}\left\| \bar{\mathcal{L}}_{\mu_n}(\tilde\eta) - \mathcal{L}(\eta)\right\| &= \left\| \bar{\mathcal{L}}_{\mu_n}(\theta_\tau\eta) - \bar{\mathcal{L}}_{\mu_n}(\eta)\right\| \\ &\leq E\left\| \int 1\{\theta_x\eta \in \cdot\}\,\big(\mu_n - \theta_\tau\mu_n\big)(dx)\right\| \\ &\leq E\left\| \mu_n - \theta_\tau\mu_n\right\| \to 0.\end{aligned}$$

(ii) Defining ξ_f by (4) with $0 \le f \le 1$, we get

$$\xi_f p_n = \int f(\theta_{-x}\eta)\, p_n(x)\, \xi(dx) \le \xi p_n,$$

and so

$$\begin{aligned} \left| \bar{\mathcal{L}}_{p_n\xi}(\eta) f - \mathcal{L}(\tilde{\eta}) f \right| &= \left| E \bar{f}_{p_n\xi}(\eta) - E f(\tilde{\eta}) \right| \\ &\le E \left| \frac{\xi_f p_n}{\xi p_n} - \frac{\xi_f p_n}{\bar{\xi}} \right| \le E \left| 1 - \frac{\xi p_n}{\bar{\xi}} \right|, \end{aligned}$$

where $0/0 = 0$. Since $\xi p_n / \bar{\xi} \xrightarrow{P} 1$ by Theorem 25.18, and

$$\begin{aligned} E\big(\xi p_n / \bar{\xi}\big) &= E\Big\{ E\big(\xi p_n \big| \mathcal{I}_{X,\xi}\big) / \bar{\xi} \Big\} \\ &= E(\,\bar{\xi}/\bar{\xi}\,) = 1, \end{aligned}$$

we get $\xi p_n / \bar{\xi} \to 1$ in L^1 by Theorem 5.12, and the assertion follows. □

We now drop the stationarity assumption and return to the setting of general random measures on a localized Borel space S, where the Palm measures are given by Theorem 31.1. Our first aim is to extend the point process approximations of Theorem 31.5 to local approximations at arbitrary points $s \in S$. As before, $a \sim b$ means $a/b \to 1$ whereas $a \approx b$ means $a - b \to 0$, and we write $\overset{u}{\approx}$ for the corresponding uniform approximation. Dissection systems are defined as in Chapter 1.

Theorem 31.9 *(local approximations) Let ξ be a simple point process on a Borel space S with σ-finite intensity $E\xi$, and fix a dissection system*[2] $\mathcal{I} = (I_{nj})$ *in S. Then for $s \in S$ a.e. $E\xi$, we have as $B \downarrow \{s\}$ along $\mathcal{I}$*

(i) $P\{\xi B > 0\} \sim P\{\xi B = 1\} \sim E\xi B$,

(ii) $\mathcal{L}\big(1_B \xi \,\big|\, \xi B > 0\big) \overset{u}{\approx} \mathcal{L}\big(1_B \xi \,\big|\, \xi B = 1\big) \overset{u}{\approx} \mathcal{L}(1_B \xi \,\|\, \xi)_s$,

(iii) *for any random variable $\eta \ge 0$ such that $E\eta\xi$ is σ-finite,*

$E(\eta \,|\, \xi B > 0) \approx E(\eta \,|\, \xi B = 1) \to E(\eta \,\|\, \xi)_s$.

Proof: Putting $I_n(s) = I_{nj}$ when $s \in I_{nj}$, we may define some simple point processes $\xi_n \le \xi$ on S by

$$\begin{aligned} \xi_n B &= \int_B 1\Big\{\xi I_n(s) = 1\Big\}\, \xi(ds) \\ &= \sum_{s \in B} \delta_s\, 1\Big\{\xi\{s\} = \xi I_n(s) = 1\Big\}, \end{aligned}$$

for any $B \in \hat{\mathcal{S}}$. Since ξ is simple and (I_{nj}) is separating, we have $\xi_n \uparrow \xi$, and so by monotone convergence $E\eta\,\xi_n \uparrow E\eta\,\xi$. Since also $E\eta\,\xi_n \le E\eta\,\xi \ll E\xi$, where all three measures are σ-finite, Theorem 2.10 yields $E\eta\,\xi_n = f_n \cdot E\xi$ for

[2]The result remains true with the same proof, if we replace the dissection system $\mathcal{I} = (I_{nj})$ by any *standard differentiation basis*, as defined in K(17).

an increasing sequence of measurable functions $f_n \geq 0$. Assuming $f_n \uparrow f$, we get by monotone convergence $E\eta\xi = f \cdot E\xi$, and so for any $B \in \mathcal{S}$,

$$\begin{aligned}\int_B f(s)\, E\xi(ds) &= (f \cdot E\xi)B = E\,\eta\,\xi B\\ &= \int_B E\xi(ds)\, E(\eta \,\|\, \xi)_s,\end{aligned}$$

which implies $f(s) = E(\eta \,\|\, \xi)_s$ a.e. $E\xi$.

For any $n \in \mathsf{N}$ and $s \in S$ a.e. $E\xi$, we may choose $B = I_n(s)$. Since $\xi_n B \leq 1$ a.s., we get

$$\begin{aligned}E\,\eta\,\xi_n B &= E\big(\eta;\, \xi_n B > 0\big)\\ &\leq E(\eta;\, \xi B > 0)\\ &\leq E\,\eta\,\xi B,\end{aligned}$$

and so

$$\frac{E\,\eta\,\xi_n B}{E\,\xi B} \leq \frac{E(\eta;\, \xi B > 0)}{E\,\xi B} \leq \frac{E\,\eta\,\xi B}{E\,\xi B}\,.$$

For fixed n and $s \in S$ a.e. $E\xi$, the differentiation property gives

$$\begin{aligned}f_n(s) &\leq \liminf_{B\downarrow\{s\}} \frac{E(\eta;\, \xi B > 0)}{E\,\xi B}\\ &\leq \limsup_{B\downarrow\{s\}} \frac{E(\eta;\, \xi B > 0)}{E\,\xi B} \leq f(s).\end{aligned}$$

Since $f_n \uparrow f$, we get as $B \downarrow \{s\}$ along $\mathcal{I}$

$$\frac{E(\eta;\, \xi B > 0)}{E\,\xi B} \approx \frac{E\,\eta\,\xi B}{E\,\xi B} \to f(s), \quad s \in S \text{ a.e. } E\xi. \tag{5}$$

Noting that

$$\begin{aligned}0 &\leq 1\{k > 0\} - 1\{k = 1\}\\ &= \ 1\{k > 1\}\\ &\leq k - 1\{k > 0\}, \quad k \in \mathsf{Z}_+,\end{aligned}$$

we see from (5) that

$$\begin{aligned}0 &\leq \frac{E(\eta;\, \xi B > 0)}{E\,\xi B} - \frac{E(\eta;\, \xi B = 1)}{E\,\xi B}\\ &\leq \frac{E\,\eta\,\xi B}{E\,\xi B} - \frac{E(\eta;\, \xi B > 0)}{E\,\xi B} \to 0,\end{aligned}$$

and using (5) again gives

$$\frac{E(\eta;\, \xi B = 1)}{E\,\xi B} \approx \frac{E(\eta;\, \xi B > 0)}{E\,\xi B} \to f(s), \quad s \in S \text{ a.e. } E\xi.$$

Taking $\eta \equiv 1$ gives (i), and dividing the two versions yields

$$\frac{E(\eta;\, \xi B = 1)}{P\{\xi B = 1\}} \approx \frac{E(\eta;\, \xi B > 0)}{P\{\xi B > 0\}} \to f(s), \quad s \in S \text{ a.e. } E\xi.$$

which is equivalent to (iii).

(ii) Using (i) along with some basic definitions and elementary estimates, we get

$$\begin{aligned}
&\Big\|\mathcal{L}\big(1_B\xi\,\big|\,\xi B=1\big)-\mathcal{L}\big(1_B\xi\,\big\|\,\xi\big)_s\Big\| \\
&\qquad= \Big\|\mathcal{L}\big(\sigma_B\,\big|\,\xi B=1\big)-\mathcal{L}(\tau_B)\Big\| \\
&\qquad= \bigg\|\frac{E\big(1_B\xi;\,\xi B=1\big)}{P(\xi B=1)}-\frac{E\,1_B\xi}{E\,\xi B}\bigg\| \\
&\qquad\le \frac{E(\xi B;\,\xi B>1)}{E\,\xi B}+E\,\xi B\left|\frac{1}{P\{\xi B=1\}}-\frac{1}{E\,\xi B}\right| \\
&\qquad= 2\left(\frac{E\,\xi B}{P\{\xi B=1\}}-1\right)\to 0,
\end{aligned}$$

and

$$\begin{aligned}
&\Big\|\mathcal{L}\big(1_B\xi\,\big|\,\xi B>0\big)-\mathcal{L}\big(1_B\xi\,\big|\,\xi B=1\big)\Big\| \\
&\qquad\le \frac{P\{\xi B>1\}}{P\{\xi B=1\}}+P\{\xi B>0\}\left|\frac{1}{P\{\xi B=1\}}-\frac{1}{P\{\xi B>0\}}\right| \\
&\qquad= 2\left(\frac{P\{\xi B>0\}}{P\{\xi B=1\}}-1\right)\to 0,
\end{aligned}$$

and the assertion follows by combination. □

We consider a special case where the Palm measures can be computed explicitly. Note that if ξ is a mixed Poisson process on S directed by $\rho\,\lambda$ for a random variable $\rho\ge 0$, then

$$P\{\xi B=0\}=\varphi(\lambda B),\quad B\in\mathcal{S},$$

where $\varphi(t)=Ee^{-t\,\rho}$ denotes the Laplace transform of ρ. When $0<\lambda S<\infty$, the same formula holds for a mixed binomial process ξ based on the probability measure $\hat\lambda=\lambda/\lambda S$ and a Z_+-valued random variable κ, provided we choose

$$\varphi(t)=E\,(1-t/\lambda S)^\kappa,\quad 0\le t\le\lambda S.$$

In either case, ξ has Laplace functional

$$Ee^{-\xi f}=\varphi\big\{\lambda(1-e^{-f})\big\},\quad f\in\mathcal{S}_+,\tag{6}$$

and so $\mathcal{L}(\xi)$ is uniquely determined by the pair (λ,φ), which justifies that we write $M(\lambda,\varphi)$ for the distribution of the point process ξ in (6). For convenience, we allow the underlying probability measure P to be σ-finite, so that $M(\lambda,-\varphi')$ makes sense as the pseudo-distribution of a point process on S.

Theorem 31.10 *(mixed Poisson and binomial processes) Let ξ be a mixed Poisson or binomial process on S with distribution $M(\lambda,\varphi)$, where λ is σ-finite. Choosing λ as supporting measure for ξ, we have*

(i) $\mathcal{L}\big(\xi-\delta_s\big\|\,\xi\big)_s=M(\lambda,-\varphi')$, *$s\in S$ a.e. λ,*

(ii) $\mathcal{L}(\xi) \equiv \mathcal{L}\big(\xi - \delta_s \big\| \xi\big)_s \;\Leftrightarrow\; \xi$ *is Poisson* (λ).

Proof: (i) First let $E\xi$ be σ-finite. For any $f \in \mathcal{S}_+$ and $B \in \mathcal{S}$ with $\lambda f < \infty$ and $\lambda B < \infty$, we have by (6)

$$E\, e^{-\xi f - t\,\xi B} = \varphi\Big\{\lambda\Big(1 - e^{-f - t\,1_B}\Big)\Big\}, \quad t \ge 0.$$

Taking right derivatives at $t = 0$, we get by dominated convergence on each side

$$E\Big(\xi B\, e^{-\xi f}\Big) = -\varphi'\Big\{\lambda(1 - e^{-f})\Big\}\,\lambda(1_B\, e^{-f}).$$

Hence, by disintegration,

$$\int_B \lambda(ds)\, E\Big(e^{-\xi f}\,\Big\|\,\xi\Big)_s = -\varphi'\Big\{\lambda(1 - e^{-f})\Big\}\int_B e^{-f(s)}\,\lambda(ds).$$

Since B was arbitrary, we obtain

$$E\Big(e^{-\xi f}\,\Big\|\,\xi\Big)_s = -\varphi'\Big\{\lambda(1 - e^{-f})\Big\}\, e^{-f(s)}, \quad s \in S \;\text{ a.e. } \lambda,$$

which implies

$$\begin{aligned} E\Big(e^{-(\xi - \delta_s) f}\Big\|\,\xi\Big)_s &= E\Big(e^{-\xi f}\,\Big\|\,\xi\Big)_s\, e^{f(s)} \\ &= -\varphi'\Big\{\lambda(1 - e^{-f})\Big\}. \end{aligned}$$

Equation (i) now follows by (6). For general $E\xi$, we may take $f \ge \varepsilon 1_B$ for fixed $\varepsilon > 0$ to justify the differentiation. The resulting formulas extend by monotone convergence to arbitrary $f \ge 0$.

(ii) The equation $\varphi = -\varphi'$ with initial condition $\varphi(0) = 1$ has the unique solution $\varphi(t) = e^{-t}$. □

A similar calculation yields a celebrated Poisson characterization:

Corollary 31.11 *(Poisson criterion, Slivnyak, Kerstan & Matthes, Mecke) Let ξ be a random measure on S with σ-finite intensity $E\xi$. Then ξ is Poisson iff*

$$\mathcal{L}(\xi \,\|\, \xi)_s = \mathcal{L}(\xi + \delta_s), \quad s \in S \;\textit{ a.e. } E\xi.$$

To avoid any reference to Palm measures, we may write the stated condition in the integrated form[3]

$$E\int f(s,\xi)\,\xi(ds) = \int Ef(s,\, \xi + \delta_s)\, E\xi(ds).$$

Proof: The necessity being part of Proposition 31.10, we need only prove the sufficiency. Then assume the stated condition, and put $\lambda = E\xi$. Fixing any $f \ge 0$ with $\lambda f < \infty$ and writing $\varphi(t) = Ee^{-t\,\xi f}$, we get by dominated convergence

[3]This formula is sometimes referred to as the *Mecke equation*.

$$\begin{aligned}
-\varphi'(t) &= E\,\xi f\, e^{-t\,\xi f}\\
&= \int \lambda(ds)\, f(s)\, E\Big(e^{-t\,\xi f}\,\Big\|\,\xi\Big)_s\\
&= \varphi(t)\,\lambda(f\,e^{-tf}).
\end{aligned}$$

Noting that $\varphi(0) = 1$ and $\varphi(1) = Ee^{-\xi f}$, we obtain

$$\begin{aligned}
-\log Ee^{-\xi f} &= \int_0^1 \frac{\varphi'(t)}{\varphi(t)}\,dt = \int_0^1 \lambda(f\,e^{-tf})\,dt\\
&= \lambda \int_0^1 f\,e^{-tf}\,dt\\
&= \lambda(1 - e^{-f}),
\end{aligned}$$

and so ξ is Poisson with intensity λ by Lemma 15.2. □

It is often helpful to calculate the Palm distributions via a preliminary conditioning, similar to the approach in Theorem 8.15. Then consider a random measure ξ on S and some random elements η in T and ζ in U, where S, T, U are Borel and the Campbell measures $C_{\xi,\eta}$ and $C_{\xi,\zeta}$ are σ-finite. Regarding ξ as a random element in the Borel space $\mathcal{M}_S$ and fixing a supporting measure ρ of ξ, we may introduce the Palm measures and conditional distributions

$$\begin{aligned}
\mathcal{L}(\eta,\zeta\,\|\,\xi)_s &= \mu(s,\,\cdot),\\
\mathcal{L}(\xi,\zeta\,|\,\eta)_t &= \mu'(t,\,\cdot),
\end{aligned}$$

in terms of some kernels $\mu\colon\ S \to T \times U$ and $\mu'\colon\ T \to \mathcal{M}_S \times U$. For fixed s, t, we may next form the Palm and conditional measures

$$\begin{aligned}
\mu_s\Big(\tilde\zeta \in \cdot\,\Big|\,\tilde\eta\Big) &= \nu_s(\tilde\eta,\,\cdot),\\
\mu'_t\Big(\tilde\zeta \in \cdot\,\Big\|\,\tilde\xi\Big)_s &= \nu'_t(s,\,\cdot),
\end{aligned}$$

where the former conditioning makes sense, since the measures μ_s are a.e. σ-finite. By Corollary 3.6, we can choose ν, ν' to be product-measurable, hence as kernels on $S \times T$, in which case we may write suggestively

$$\begin{aligned}
P(\cdot\,|\,\eta\,\|\,\xi)_{s,t} &= \Big\{P(\cdot\,\|\,\xi)_s\Big\}(\cdot\,|\,\eta)_t,\\
P(\cdot\,\|\,\xi\,|\,\eta)_{t,s} &= \Big\{P(\cdot\,|\,\eta)_t\Big\}(\cdot\,\|\,\xi)_s.
\end{aligned}$$

Finally, we may take $t = \eta$ and put $\mathcal{F} = \sigma(\eta)$ and $\mathcal{H} = \sigma(\zeta)$, to form on S the product-measurable, measure-valued processes

$$\begin{aligned}
P_{\mathcal{F}}(\cdot\,\|\,\xi)_s &= P(\cdot\,\|\,\xi\,|\,\eta)_{s,\eta},\\
\Big\{P(\cdot\,\|\,\xi)_s\Big\}_{\mathcal{F}} &= \{P(\cdot\,|\,\eta\,\|\,\xi)_{\eta,s}.
\end{aligned}$$

Using this notation, we may state the iteration properties in the following suggestive form. In particular, part (iii) enables us to calculate the Palm kernel via a preliminary conditioning on η, which is especially useful in the contexts of randomizations and Cox processes. Say that a σ-field $\mathcal{F}$ on Ω is *Borel generated*, if $\mathcal{F} = \sigma(\beta)$ for some random element β in a Borel space.

Theorem 31.12 *(iteration) Let ξ be a random measure on a Borel space S, and let $\mathcal{F}$, $\mathcal{H}$ be Borel-generated σ-fields in Ω, such that $C_{\xi,\mathcal{F}}$ and $C_{\xi,\mathcal{H}}$ are σ-finite. Then for a fixed supporting measure $\nu \sim E\xi$, we have on $\mathcal{H}$*

(i) *$E_{\mathcal{F}}\xi$ is a.s. σ-finite,*

(ii) $P_{\mathcal{F}}(\cdot \,\|\, \xi) = \big\{P(\cdot \,\|\, \xi)\big\}_{\mathcal{F}}$ *a.e.* $C_{\xi,\mathcal{F}}$,

(iii) $P(\cdot \,\|\, \xi)_s = E\Big\{P_{\mathcal{F}}(\cdot \,\|\, \xi)_s \Big\| \xi\Big\}_s$, $s \in S$ *a.e.* $E\xi$.

Proof: Let $\mathcal{F} = \sigma(\eta)$ and $\mathcal{H} = \sigma(\zeta)$ for some random elements η, ζ in the Borel spaces T, U, and introduce the measures

$$\mu_1 = \nu, \qquad \mu_2 = \mathcal{L}(\eta), \qquad \mu_3 = \mathcal{L}(\zeta),$$
$$\mu_{12} = C_{\xi,\eta}, \qquad \mu_{13} = C_{\xi,\zeta}, \qquad \mu_{123} = C_{\xi,\eta,\zeta},$$

which are all σ-finite. Since the support conditions of Theorem 3.7 are clearly fulfilled, the result yields

$$\begin{aligned} \mu_{3|2|1} &\overset{\sim}{=} \mu_{3|1|2} \ \text{ a.e. } \mu_{12}, \\ \mu_{3|1} &= \mu_{2|1}\, \mu_{3|1|2} \ \text{ a.e. } \mu_1. \end{aligned} \tag{7}$$

Further note that

$$\begin{aligned} \mu_{12} &= C_{\xi,\eta} \\ &\overset{\sim}{=} \mathcal{L}(\eta) \otimes E_{\mathcal{F}}\xi \\ &= \mu_2 \otimes E_{\mathcal{F}}\xi, \end{aligned}$$

and so the uniqueness in Theorem 3.4 yields $\mu_{1|2} = E_{\mathcal{F}}\xi$ a.s. In particular this implies (i), and the relations (7) are equivalent to (ii) and (iii). □

When looking for versions of the Palm measures with desired regularity properties, the following duality approach may be helpful. Then recall that, for any Borel spaces S, T, a σ-finite measure ρ on $S \times T$ admits the dual disintegrations

$$\rho = \nu \otimes \mu \overset{\sim}{=} \nu' \otimes \mu',$$

for some σ-finite measures ν on S and ν' on T, and some kernels $\mu\colon S \to T$ and $\mu'\colon T \to S$. Here either disintegration may give useful information about the other one, as illustrated by the following simple case:

Theorem 31.13 *(Palm–density duality) Consider a random measure ξ on S and a random element η in T, where S, T are Borel, and fix a supporting measure ν of ξ. Then*

(i) *$E(\xi \,|\, \eta)$ and $\mathcal{L}(\eta \,\|\, \xi)$ exist simultaneously as σ-finite kernels between S and T,*

(ii) *for $E(\xi \,|\, \eta)$ and $\mathcal{L}(\eta \,\|\, \xi)$ as in* (i),

$$E(\xi \,|\, \eta) \ll \nu \ \text{ a.s.} \quad \Leftrightarrow \quad \mathcal{L}(\eta \,\|\, \xi)_s \ll \mathcal{L}(\eta) \ \text{ a.e. } \nu,$$

in which case both kernels have product-measurable a.e. densities,

(iii) *for any measurable function* $f \geq 0$ *on* $S \times T$,

$$E(\xi \,|\, \eta) = f(\cdot, \eta) \cdot \nu \;\; a.s. \quad \Leftrightarrow \quad \mathcal{L}(\eta \,\|\, \xi)_s = f(s, \cdot) \cdot \mathcal{L}(\eta) \;\; a.e. \;\; \nu.$$

Proof: (i) If $E\xi$ is σ-finite, then for any $B \in \mathcal{S}$ and $C \in \mathcal{T}$,

$$\begin{aligned} E\Big(\xi B;\, \eta \in C\Big) &= E\Big\{E(\xi B \,|\, \eta);\, \eta \in C\Big\} \\ &= \int_C P\{\eta \in dt\}\, E(\xi B \,|\, \eta)_t, \end{aligned}$$

which extends immediately to the dual disintegration

$$C_{\xi,\eta} \overset{\sim}{=} \mathcal{L}(\eta) \otimes E(\xi \,|\, \eta). \tag{8}$$

In general, ξ is a countable sum of bounded random measures, and we may use Fubini's theorem to extend (8) to the general case, for a suitable version of $E(\xi \,|\, \eta)$. In particular, the latter kernel has a σ-finite version iff $C_{\xi,\eta}$ is σ-finite. Since this condition is also equivalent to the existence of a σ-finite Palm kernel, the assertion follows.

(iii) Assuming $E(\xi \,|\, \eta) = f(\cdot, \eta) \cdot \nu$ a.s., we get for any $B \in \mathcal{S}$ and $C \in \mathcal{T}$

$$\begin{aligned} E\Big(\xi B;\, \eta \in C\Big) &= E\Big\{E(\xi B \,|\, \eta);\, \eta \in C\Big\} \\ &= E \int_B f(s, \eta)\, \nu(ds)\, 1_C(\eta) \\ &= \int_B \nu(ds)\, E\Big\{f(s, \eta);\, \eta \in C\Big\} \\ &= \int_B \nu(ds) \int_C P\{\eta \in dt\}\, f(s, t), \end{aligned}$$

which implies $\mathcal{L}(\eta \,\|\, \xi)_s = f(s, \cdot) \cdot \mathcal{L}(\eta)$ a.e. ν. Conversely, the latter relation yields

$$\begin{aligned} E\Big(\xi B;\, \eta \in C\Big) &= \int_B \nu(ds)\, P\Big(\eta \in C \,\Big\|\, \xi\Big)_s \\ &= \int_B \nu(ds) \int_C P\{\eta \in dt\}\, f(s, t) \\ &= \int_B \nu(ds)\, E\Big\{f(s, \eta);\, \eta \in C\Big\} \\ &= E \int_B f(s, \eta)\, \nu(ds)\, 1_C(\eta), \end{aligned}$$

which shows that

$$\begin{aligned} E(\xi B \,|\, \eta) &= \int_B f(s, \eta)\, \nu(ds) \\ &= \Big\{f(\cdot, \eta) \cdot \nu\Big\} B \;\; \text{a.s.}, \end{aligned}$$

and therefore $E(\xi \,|\, \eta) = f(\cdot, \eta) \cdot \nu$ a.s.

(ii) Since S, T are Borel, Theorem 9.27 yields a product-measurable density in each case. The assertion now follows from (iii). □

Higher order Palm measures are defined in the obvious way as univariate Palm measures with respect to the product random measures $\xi^n = \xi^{\otimes n}$ on S^n. We are especially interested in the Palm distributions of ξ itself with respect to ξ^n, here written as $\mathcal{L}(\xi \,\|\, \xi^n)$. When ξ is a simple point process with σ-finite moment measures $E\xi^n$, we note that

$$P\Big(\textstyle\prod_i \xi\{s_i\} = 1 \,\Big\|\, \xi^{(n)}\Big)_s = 1, \quad s \in S^{(n)} \text{ a.e. } E\xi^n,$$

where $S^{(n)}$ denotes the non-diagonal part of S^n. This suggests that we consider instead the *reduced Palm measures*

$$\mathcal{L}\Big(\xi' \,\Big\|\, \xi^{(n)}\Big)_s \equiv \mathcal{L}\Big(\xi - \textstyle\sum_i \delta_{s_i} \Big\|\, \xi^{(n)}\Big)_s, \quad s \in S^{(n)},$$

formed by disintegration of the n-th order *reduced Campbell measures*

$$C_\xi^{(n)} f = E \int \xi^{(n)}(ds)\, f\Big(s; \xi - \textstyle\sum_i \delta_{s_i}\Big), \quad f \geq 0 \text{ on } S^{(n)} \times \mathcal{N}_S.$$

Since the latter are symmetric under permutation of coordinates in $S^{(n)}$, we may regard the reduced Palm measures $P(\xi' \,\|\, \xi^{(n)})_s$ as functions of the associated point measures $\sum_i \delta_{s_i}$. By Lemma 2.20 they can then be obtained, simultaneously for all $n \in \mathbb{N}$, by disintegration of the *compound Campbell measure*

$$\begin{aligned} C_\xi f &= \sum_{n\geq 0} C_\xi^{(n)} f/n! \\ &= E \sum_{\mu \leq \xi} f(\mu, \xi - \mu), \end{aligned}$$

where $C_\xi^{(0)} = \mathcal{L}(\xi)$ for consistency, and the last summation extends over all *bounded* point measures $\mu \leq \xi$. More precisely, assuming $\xi = \sum_{i\in J} \delta_{\sigma_i}$ for a countable index set J, we sum[4] over all measures $\mu = \sum_{i \in I} \delta_{\sigma_i}$ with finite $I \subset J$.

We now define the *Gibbs kernel* of ξ as the maximal random measure $\Gamma = G(\xi)$ on $\hat{\mathcal{N}}_S$ satisfying

$$E \int \Gamma(d\mu)\, f(\mu, \xi) \leq E \sum_{\mu \leq \xi} f(\mu, \xi - \mu), \tag{9}$$

for all measurable functions $f \geq 0$ on $\hat{\mathcal{N}}_S \times \mathcal{N}_S$. Such a partial disintegration kernel exists uniquely by Corollary 3.5.

Lemma 31.14 *(boundedness and support) Let ξ be a point process on S with Gibbs kernel Γ. Then a.s.*

(i) $\Gamma\{\mu S = \infty\} = 0$,

(ii) $\Gamma\big\{\mu B = \mu S = n\big\} < \infty, \quad B \in \hat{\mathcal{S}},\ n \in \mathbb{Z}_+,$

(iii) $\int \mu(\operatorname{supp} \xi)\, \Gamma(d\mu) = 0$ *when ξ is simple.*

[4]The sum is clearly independent of the choice of atomic decomposition.

Proof, (i)–(ii): For any $B \in \hat{\mathcal{S}}$ and $m, n \in \mathsf{Z}_+$,

$$\begin{aligned} E\Big(\Gamma\Big\{\mu B = \mu S = n\Big\};\, \xi B = m\Big) &\le C\Big(\Big\{\nu B = m\Big\} \times \Big\{\mu B = \mu S = n\Big\}\Big) \\ &= E\, 1\Big\{\xi B = m + n\Big\} \sum_{\mu \le \xi} 1\Big\{\mu B = \mu S = n\Big\} \\ &= \binom{m+n}{n} P\Big\{\xi B = m + n\Big\} < \infty, \end{aligned}$$

and so $\Gamma\{\nu B = \nu S = n\} < \infty$ a.s. on $\{\xi B = m\}$.

(iii) Approximating $\operatorname{supp} m$ by means of dissections of S and using dominated convergence, we see that the function $f(m, \mu) = \mu(\operatorname{supp} m)$ is product-measurable on $\mathcal{N}_S \times \hat{\mathcal{N}}_S$. Hence, (9) yields for simple ξ

$$E \int \mu(\operatorname{supp} \xi)\, \Gamma(d\mu) \le Cf = E \sum_{\mu \le \xi} \mu\Big\{\operatorname{supp}(\xi - \mu)\Big\} = 0,$$

and so the inner integral on the left vanishes a.s. $\square$

The definition of Γ is justified by some remarkable properties:

Theorem 31.15 *(Gibbs kernel) Let $\Gamma = G(\xi)$ be the Gibbs kernel of a point process ξ on S. Then for any $B \in \hat{\mathcal{S}}$ and measurable $f \ge 0$ on $\hat{\mathcal{N}}_S$,*

(i) $$E \int \Gamma(d\mu)\, f(\xi + \mu)\, 1\Big\{\xi B = \mu B^c = 0\Big\} = E\Big\{f(\xi);\, P(\xi B = 0 \,|\, 1_{B^c}\xi) > 0\Big\},$$

(ii) $$\Gamma(\cdot\,;\, \mu B^c = 0) = \frac{\mathcal{L}\Big(1_B \xi \,\Big|\, 1_{B^c}\xi\Big)}{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)} \quad \textit{a.s. on } \{\xi B = 0\},$$

(iii) $$\mathcal{L}\Big(1_B \xi \,\Big|\, 1_{B^c}\xi\Big) = \Gamma(\cdot \,|\, \mu B^c = 0) \quad \textit{a.s. on } \{\xi B = 0\}.$$

In subsequent proofs, we shall often use the fact that $P(A \,|\mathcal{F}) > 0$ a.s. on A for any A and $\mathcal{F}$, which follows from

$$P\Big\{P(A \,|\mathcal{F}) = 0;\, A\Big\} = E\Big\{P(A \,|\mathcal{F});\, P(A \,|\mathcal{F}) = 0\Big\} = 0.$$

Proof: Define for fixed $B \in \hat{\mathcal{S}}$

$$M_0 = \Big\{\mu;\, \mu B = 0,\, P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)_\mu > 0\Big\}.$$

Then for any measurable $M \subset M_0$,

$$\begin{aligned} &C\Big(M \times \{\mu B^c = 0\}\Big) \\ &\quad = E \sum_{\mu \le \xi} 1\Big\{\xi - \mu \in M,\, \mu B^c = 0\Big\} \\ &\quad = E \sum_{\mu \le \xi} 1\Big\{\xi - \mu \in M,\, P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0,\, \mu B^c = 0\Big\} \\ &\quad = P\Big\{1_{B^c}\xi \in M,\, P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0\Big\} \\ &\quad \ll E\Big\{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big);\, 1_{B^c}\xi \in M\Big\} \\ &\quad = P\Big\{\xi B = 0,\, 1_{B^c}\xi \in M\Big\} \\ &\quad \le P\{\xi \in M\}, \end{aligned}$$

and so $C(\cdot \times \{\mu B^c = 0\}) \ll \mathcal{L}(\xi)$ on M_0. Noting that $P(\xi B = 0 \,|\, 1_{B^c}\xi) > 0$ a.s. on $\{\xi B = 0\}$ and using the maximality of G, we get for any $B \in \hat{\mathcal{S}}$ and measurable $f \geq 0$ on $\mathcal{N}_S$

$$
\begin{aligned}
E \int & \Gamma(d\mu)\, f(\xi+\mu)\, 1\Big\{\xi B = \mu B^c = 0\Big\} \\
&= E \int \Gamma(d\mu)\, f(\xi+\mu)\, 1\Big\{\xi B = \mu B^c = 0,\ P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0\Big\} \\
&= \iint C(d\mu\, d\nu)\, f(\mu+\nu)\, 1\Big\{\nu B = \mu B^c = 0,\ P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)_\nu > 0\Big\} \\
&= E \sum_{\mu \leq \xi} f(\xi)\, 1\Big\{\xi B = \mu B,\ \mu B^c = 0,\ P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0\Big\} \\
&= E\Big\{f(\xi);\ P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0\Big\}.
\end{aligned}
$$

(ii) Replacing f in (i) by the function

$$
g(\mu) = f(1_B\mu)\, 1_M(1_{B^c}\mu), \quad \mu \in \mathcal{N}_S,
$$

for measurable $M \subset \mathcal{N}_S$ and $f \geq 0$ on $\hat{\mathcal{N}}_S$, we get

$$
\begin{aligned}
E\Big\{& f(1_B\xi)\, 1_M(1_{B^c}\xi);\ P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0\Big\} \\
&= E \int \Gamma(d\mu)\, f(\mu)\, 1_M(\xi)\, 1\Big\{\xi B = \mu B^c = 0\Big\} \\
&= E \int G\Big(1_{B^c}\xi,\, d\mu\Big)\, f(\mu)\, 1_M(1_{B^c}\xi)\, 1\Big\{\xi B = \mu B^c = 0\Big\}.
\end{aligned}
$$

Since M was arbitrary, we obtain a.s.

$$
\begin{aligned}
E\Big\{& f(1_B\xi) \,\Big|\, 1_{B^c}\xi\Big\}\ 1\Big\{P\Big(\xi B = 0 \,|\, 1_{B^c}\xi\Big) > 0\Big\} \\
&= \int G\Big(1_{B^c}\xi,\, d\mu\Big)\, f(\mu)\, 1\{\mu B^c = 0\}\, P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big),
\end{aligned}
$$

and since $P(\xi B = 0 \,|\, 1_{B^c}\xi) > 0$ a.s. on $\{\xi B = 0\}$, this gives

$$
1_{\mathcal{N}_B}\Gamma f = \frac{E\Big\{f(1_B\xi) \,\Big|\, 1_{B^c}\xi\Big\}}{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)} \ \text{ a.s. on } \{\xi B = 0\}.
$$

The assertion now follows since f was arbitrary.

(iii) By (ii), we have

$$
\Gamma\{\mu B^c = 0\} = \Big\{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)\Big\}^{-1} > 0 \ \text{ a.s. on } \{\xi B = 0\},
$$

and the assertion follows by division. □

Of special importance is the restriction of the Gibbs kernel to the class of single point masses δ_s. Identifying δ_s with the point $s \in S$ yields a random measure $\eta = g(\xi, \cdot)$ on S, known as the *Papangelou kernel* of ξ, which can also be defined directly as the maximal kernel satisfying the partial disintegration

$$
E \int \eta(ds)\, f(s, \xi) \leq E \int \xi(ds)\, f(s,\, \xi - \delta_s). \tag{10}
$$

Corollary 31.16 *(Papangelou kernel) The Papangelou kernel η of a point process ξ on S is a.s. locally finite, and for simple ξ we have $\eta(\operatorname{supp}\xi) = 0$ a.s. Furthermore, for any $B \in \hat{\mathcal{S}}$,*

$$\text{(i)}\qquad 1_B\,\eta = \frac{E\Big(1_B\,\xi;\ \xi B = 1 \,\Big|\, 1_{B^c}\xi\Big)}{P\Big(\xi B = 0 \,|\, 1_{B^c}\xi\Big)} \ \textit{ a.s. on } \{\xi B = 0\},$$

$$\text{(ii)}\qquad \mathcal{L}\Big(\tau_B \,\Big|\, 1_{B^c}\xi,\ \xi B = 1\Big) = \frac{1_B\,\eta}{\eta B} \ \textit{ a.s. on } \{\xi B = 0,\ \eta B > 0\}.$$

Proof: The local finiteness is clear, since by Lemma 31.14 (i),

$$\eta B = \Gamma\Big\{\mu B = \mu S = 1\Big\} < \infty \ \text{ a.s.,} \quad B \in \hat{\mathcal{S}}.$$

When ξ is simple, part (ii) of the same lemma yields a.s.

$$\begin{aligned}\eta(\operatorname{supp}\xi) &= \Gamma\Big\{\mu(\operatorname{supp}\xi) = \mu S = 1\Big\}\\ &= \int \Gamma(d\mu)\,\mu(\operatorname{supp}\xi) = 0.\end{aligned}$$

(i) By Theorem 31.15 (ii), we get for any $C \subset B$ in $\hat{\mathcal{S}}$

$$\begin{aligned}\eta C &= \Gamma\Big\{\mu C = \mu S = 1\Big\}\\ &= \frac{P\Big(\xi C = \xi B = 1 \,\Big|\, 1_{B^c}\xi\Big)}{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)} \ \text{ a.s. on } \{\xi B = 0\}.\end{aligned}$$

(ii) By (i), we have

$$\eta B = \frac{P\Big(\xi B = 1 \,\Big|\, 1_{B^c}\xi\Big)}{P\Big(\xi B = 0 \,|\, 1_{B^c}\xi\Big)} \ \text{ a.s. on } \{\xi B = 0\}.$$

When this is positive, the assertion follows by division. □

The theory simplifies under the classical regularity condition

$$P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big) > 0 \ \text{ a.s.,} \quad B \in \hat{\mathcal{S}},$$

traditionally labeled by (Σ).

Lemma 31.17 *(regularity) For a simple point process ξ on S satisfying* (Σ), *we have*

(i) *the right-hand side of Theorem 31.15* (i) *reduces to $Ef(\xi)$,*

(ii) *equality holds in* (9) *and* (10),

(iii) *when $A \in \sigma(1_{B^c}\xi)$ a.s. on $\{\xi B = 0\}$, we have $PA = 1$.*

Proof: (i) Obvious.

(ii) The equality in (9) holds by (i), and the equality in (10) follows.

(iii) The stated condition gives

$$E\Big\{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big);\, A^c\Big\} = P\Big(A^c \cap \{\xi B = 0\}\Big) = 0,$$

and (Σ) yields $PA^c = 0$. □

To illustrate the usefulness of the Papangelou kernel η, we show how invariance of η may imply that the underlying point process ξ is Cox.

Theorem 31.18 *(invariance and Cox property, Papangelou) Let ξ be a T-marked point process on* R *satisfying (Σ), let ν be a random measure on T, and define $\eta = \lambda \otimes \nu$. Then these conditions are equivalent:*

(i) *ξ is a Cox process directed by η,*

(ii) *ξ has Papangelou kernel η.*

Proof, (i) $\Rightarrow$ (ii): Let η be the Papangelou kernel of ξ. Assuming (i), and noting that $\zeta = \lambda \otimes \nu$ is $1_{B^c}\xi$-measurable for every bounded, measurable set $B \in \mathsf{R} \times T$, we see from Corollary 31.16 that, a.s. on $\{\xi B = 0\}$,

$$\eta B = \frac{P\Big(\xi B = 1 \,\Big|\, 1_{B^c}\xi\Big)}{P\Big(\xi B = 0 \,\Big|\, 1_{B^c}\xi\Big)} = \frac{e^{-\zeta B}\,\zeta B}{e^{-\zeta B}} = \zeta B.$$

Since also $\eta = 0$ a.s. on $\operatorname{supp}\xi$ by Lemma 31.14, we obtain $\eta = \zeta = \lambda \otimes \nu$ a.s.

(ii) $\Rightarrow$ (i): Under (ii), Corollary 31.16 shows that $\mathcal{L}(1_B\xi;\, \xi B = 1 \,|\, 1_{B^c}\xi)$ is a.s. λ-invariant on $\{\xi B = 0\}$ for every bounded, measurable rectangle $B \subset \mathsf{R} \times T$, which extends by Lemma 31.17 (iii) to all of Ω. By iteration it follows that the ordered points in $\mathcal{L}(1_B\xi;\, \xi B = n \,|\, 1_{B^c}\xi)$ have a shift-invariant joint distribution, and hence form a stationary binomial process on B. In other words, ξ is a mixed binomial process on every bounded, measurable set B, and Theorem 15.4 yields the desired Cox property. □

When ξ is a.s. bounded, the compound Campbell measure becomes symmetric in its two components, and the Palm and Gibbs kernels agree. This simple observation extends to a general relationship between the two kernels. Here we write $Q(\mu, \cdot)$ for a version of the reduced Palm kernel of ξ, where $\mu \in \hat{\mathcal{N}}_S$.

Proposition 31.19 *(Palm–Gibbs duality) Let ξ be a point process on S with Gibbs kernel G and reduced Palm kernel Q, where $\xi S < \infty$ a.s. Then $Q(\xi, \{0\}) > 0$ a.s., and*

$$\Gamma = G(\xi, \cdot) = \frac{Q(\xi, \cdot)}{Q(\xi, \{0\})} \quad \text{a.s.}$$

Proof: When $\xi S < \infty$ a.s., the compound Campbell measure is symmetric, in the sense that $C\tilde{f} = Cf$ with $\tilde{f}(s,t) = f(t,s)$. Since $\mathcal{L}(\xi) \ll C(\cdot \times \mathcal{N}_S)$, we may choose an $A \in \mathcal{B}_{\mathcal{N}_S}$ with $\xi \in A$ a.s. and $\mathcal{L}(\xi) \sim C(\cdot \times \mathcal{N}_S)$ on A. Then on $A \times \mathcal{N}_S$ we have the dual disintegrations

$$\begin{aligned} C &= \mathcal{L}(\xi) \otimes G \\ &\cong \nu \otimes Q \ \text{ on } \ A \times \mathcal{N}_S, \end{aligned}$$

where ν is the supporting measure on $\mathcal{N}_S$ associated with Q. The uniqueness in Theorem 3.4 yields

$$\Gamma \equiv G(\xi, \cdot) = p(\xi)\, Q(\xi, \cdot) \ \text{ a.s.}, \tag{11}$$

for a measurable function $p > 0$ on A. To determine p, we may apply Theorem 31.15 (ii) to the sets $M = \{0\}$ and $B = \emptyset$ to get a.s.

$$\Gamma\{0\} = \frac{P\big(1_\emptyset \xi = 0 \,\big|\, 1_S \xi\big)}{P\big(\xi\emptyset = 0 \,\big|\, 1_S \xi\big)} = \frac{P(\Omega \,|\, \xi)}{P(\Omega \,|\, \xi)} = 1.$$

Inserting this into (11) gives $p(\xi)\, Q(\xi, \{0\}) = 1$ a.s., and the assertion follows. □

We finally note how the duality between Palm and Gibbs kernels yields an inner conditioning property, similar to the outer conditioning in Theorem 31.15 (iii).

Corollary 31.20 *(inner conditioning) Let ξ be a point process on S with reduced Palm kernel Q. Then for any $B \in \hat{\mathcal{S}}$, we have $Q_{1_B \xi}\{\mu B = 0\} > 0$ and*

$$\mathcal{L}\big(1_{B^c}\xi \,\big|\, 1_B\, \xi\big) = Q_{1_B \xi}(\,\cdot \,|\, \mu B = 0\,) \quad a.s.$$

Proof: Applying Theorem 31.15 (ii) with ξ and B^c replaced by $1_B\, \xi$ and B, and using Proposition 31.19 with ξ replaced by $1_B\, \xi$, we get a.s.

$$\begin{aligned} \mathcal{L}\big(1_{B^c}\xi \,\big|\, 1_B\, \xi\big) &= \Gamma^B_{1_B \xi}\big(1_{B^c}\mu \,\big|\, \mu B = 0\big) \\ &= Q^B_{1_B \xi}\big(1_{B^c}\mu \in \cdot \,\big|\, \mu B = 0\big) \\ &= Q_{1_B \xi}\big(1_{B^c}\mu \in \cdot \,\big|\, \mu B = 0\big), \end{aligned}$$

where Γ^B and Q^B denote the Gibbs and Palm kernels of the restriction $1_B\, \xi$. □

Exercises

1. Put $\xi = \delta_\sigma$ for a random element σ in S. Show that Theorem 31.1 reduces to the disintegration theorem for regular conditional distributions $\mathcal{L}(\eta \,|\, \sigma)$. Also give the Campbell measure $C_{\xi,\eta}$ in this case.

2. Let G be a compact group with normalized Haar measure λ, acting measurably on S, T, consider a stationary random pair (σ, η) in $S \times T$, and put $\xi = \delta_\sigma$. Show that in this case the group actions are proper, and express the invariant Palm kernel in Theorem 31.2 in terms of conditional distributions. Also show that when $S = G$, (ii) gives an invariant representation of the conditional distributions $\mathcal{L}(\eta \,|\, \sigma)_r$.

3. Consider a stationary pair (ξ, η) on the d-dimensional torus T^d, where $\xi = \delta_\sigma$ for a random element σ in T^d, and extend by periodic continuation to a stationary pair $(\tilde{\xi}, \tilde{\eta})$ on R^d. Express the inversion formulas of Theorem 31.3 in terms of conditional distributions $\mathcal{L}(\eta \,|\, \sigma)_r$.

4. For (ξ, η) as in Lemma 31.3, define a random measure $\tilde{\xi}$ on the appropriate product space by $\tilde{\xi}(A \times B) = \int_B 1_A\{\theta_{-x}(\xi, \eta)\}\, \xi(ds)$. Show that $\tilde{\xi}$ is again stationary under shifts in R^d.

5. Let ξ be a stationary renewal process based on a distribution μ, and describe the associated Palm measure as the distribution of a random sequence $\tilde{\xi}$. Explain in what sense $\tilde{\xi}$ is cycle stationary. Conversely, given a stationary sequence $\tilde{\xi}$ of positive random variables, explain how $\mathcal{L}(\tilde{\xi})$ arises as the Palm distribution of a stationary, simple point process ξ on R. Also state conditions on μ and $\tilde{\xi}$ for the associated measures $\mathcal{L}(\tilde{\xi})$ and $\mathcal{L}(\xi)$ to be bounded.

6. Prove Theorem 31.5 (i) by an elementary argument, when the B_n are intervals in R^d. (*Hint:* If an interval I is divided, for each n, into subintervals I_{nj} with $\max_j |I_{nj}| \to 0$, then $\sum_j 1\{\xi I_{nj} = 1\} \to \xi I$ a.s. Now take expected values and use dominated convergence.)

7. In the context of Theorem 31.8, show that $Q_{\xi,\eta} = Q'_{\xi,\eta}$ iff $\bar{\xi} = E\bar{\xi} \in (0, \infty)$ a.s. Also give examples where $Q_{\xi,\eta}$ but not $Q'_{\xi,\eta}$ exists as a probability measure, and conversely.

8. Show that Theorems 31.6 and 31.8 fail for ordinary Palm measures, unless $\bar{\xi}$ is a.s. a constant.

9. Show by examples that the statements in Theorem 31.9 (i)$-$(ii) fail when the point process ξ is not simple.

10. Find the Palm distribution of a Bernoulli sequence, and write the result in the form of Corollary 31.11. In other words, to obtain the Palm distribution of (ξ_n), we simply replace ξ_0 by the value 1. Now apply Lemma 31.7. (Thus, a suitable random shift of (ξ_n) yields a sequence with the same distribution, except that the value at the origin in now 1.[5])

11. Show how Theorem 31.12 can be used to find the Palm distribution of a Cox process or a mixed binomial process.

12. Explain the statements of Theorem 31.13, in the special cases where ξ is a Cox process directed by η or a mixed binomial process with $\|\xi\| = \eta$.

13. Give an example of a simple point process that satisfies (Σ) and one that doesn't. Describe how the conditions in Theorem 31.15 and Corollary 31.16 simplify under condition (Σ).

14. Since a compound Campbell measure of a bounded point process ξ is symmetric in the two components, explain why the Gibbs kernel is based on a partial

[5] This is known as the paradox of *an extra head*.

disintegration only, whereas the Palm kernel is obtained through a full disintegration. Give an example of a bounded point process where the two kernels differ.

15. Since both the local approximation in Theorem 31.9 and the inner conditioning in Corollary 31.20 describe Palm measures in terms of elementary conditioning, explain how they differ and why they lead to different interpretations of the Palm distributions.

X. SDEs, Diffusions, and Potential Theory

Stochastic differential equations (SDEs) form a central topic of modern probability theory with a wealth of applications. In the autonomous case, the solutions provide probabilistic descriptions of continuous, strong Markov processes, known as diffusions. In particular, we explore the relations between weak and strong solutions, and characterize the solutions in terms of a martingale problem. In the one-dimensional case, diffusions may be described in terms of a scale function and a speed measure. Potential theoretic tools are used throughout Markov process theory, and some classical problems in the area have beautiful solutions in terms of Brownian motion. Our final chapter on stochastic differential geometry deals primarily with martingales and semi-martingales in a general differentiable manifold. For beginners we recommend a careful study of Chapter 32 and selected parts of Chapter 33, whereas the remaining material is more advanced and might be postponed.

— — —

32. Stochastic equations and martingale problems. Using a Picard iteration, we may construct pathwise solutions to suitable SDEs, combining into a stochastic flow on the underlying space. Next, we examine the relationship between weak solutions and the associated martingale problems, and explore the connections between weak and strong existence and uniqueness.

33. One-dimensional SDEs and diffusions. For a one-dimensional SDE without a drift term, we give precise conditions for existence and uniqueness. We proceed with a detailed study of one-dimensional diffusions, leading to a complete description in terms of a scale function and a speed measure. We further examine the recurrence and ergodic properties of diffusions on a natural scale, depending on the speed measure and nature of the endpoints.

34. PDE connections and potential theory. Here we explore some connections between Brownian motion and classical potential theory, including solutions to the three classical problems of electrostatics. Next, we explore the relationship between alternating capacities and random closed sets, as well as between excessive functions and super-martingales, leading to a probabilistic version of the classical Riesz decomposition.

35. Stochastic differential geometry. Here we consider semi-martingales and their covariation processes in a general differential manifold S, define martingales in terms of a connection, and explore criteria involving affine and convex functions. We further introduce the local characteristics of a semi-martingale in S, and justify their definitions by some natural embedding and projection properties. Finally, we specialize to the diffusion case, and characterize Brownian motion and related processes in a Riemannian manifold.

Chapter 32

Stochastic Equations and Martingale Problems

Progressive functions, drift and diffusion rates, Langevin equation, linear equations, weak, strong, and functional solutions, stochastic flows, Picard iteration, explosion, uniqueness, path-wise and in law, martingale problems, weak existence and continuity, measurability and mixtures, strong Markov and Feller properties, transformation of drift, scaling, transfer of solutions

Just as ordinary differential equations (ODEs) describe the evolution of suitably smooth dynamical systems, so the *stochastic differential equations* (SDEs) model the evolution of the random counterparts, the *diffusion processes*. The corresponding theories are in many ways analogous, except that the stochastic version involves an additional noise term, given by an Itô-type stochastic integral with respect to a Brownian motion. This leads to a close connection between the present SDE theory and the stochastic calculus developed in Chapters 18–19, which in turn depends in a crucial way on the martingale theory of Chapter 9.

The study of SDEs and associated diffusion processes leads far beyond the mentioned analogy with classical dynamical systems. In fact, the coefficients of such an equation determine a possibly time-dependent elliptic operator A, as in Theorem 17.24, which suggests the associated *martingale problem* of finding a process X, such that the processes M^f in Lemma 17.21 become martingales. It turns out to be essentially equivalent for X to be a weak solution to the given SDE, as will be seen from the fundamental Theorem 32.7.

The general theory of SDEs involves both weak and strong solutions. Here it is important to distinguish between the notions of *strong* and *weak existence*, and the associated notions of *pathwise uniqueness* and *uniqueness in law*. In this connection, we will establish the powerful result of Yamada and Watanabe, asserting that weak existence and pathwise uniqueness imply strong existence and uniqueness in law. Under the same conditions, we can in fact express the general solution in functional form as $X = F(X_0, B)$ a.s.

The SDEs studied in this chapter are typically of the form

$$dX_t^i = \sigma_j^i(t, X)\, dB_t^j + b^i(t, X)\, dt, \tag{1}$$

which may also be written in integrated form as

$$X_t^i = X_0^i + \textstyle\sum_j \displaystyle\int_0^t \sigma_j^i(s, X)\, dB_s^j + \int_0^t b^i(s, X)\, ds, \quad t \geq 0. \tag{2}$$

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_33

Here $B = (B^1, \dots, B^r)$ is a Brownian motion in R^r with respect to a filtration $\mathcal{F}$, and the solution $X = (X^1, \dots, X^d)$ is a continuous $\mathcal{F}$-semi-martingale in R^d. Furthermore, the coefficients σ and b are progressive[1] functions of suitable dimension, defined on the canonical path space $C_{\mathsf{R}_+,\mathsf{R}^d}$, equipped with the induced filtration $\mathcal{G}_t = \sigma\{x_s;\, s \le t\}$, $t \ge 0$. For convenience, we shall often refer to (1) as *equation* (σ, b).

For the integrals in (2) to exist, in the sense of Itô and Lebesgue integration, X must fulfill the integrability conditions

$$\int_0^t \left\{ \left| a^{ij}(s,X) \right| + \left| b^i(s,X) \right| \right\} ds < \infty \ \text{ a.s.}, \quad t \ge 0, \tag{3}$$

where $a^{ij} = \sigma^i_k \sigma^j_k$ or $a = \sigma\sigma'$, and the bars denote any norms in the spaces of $d \times d$-matrices and d-vectors, respectively. For the existence and adaptedness of the right-hand side, we also need the integrands in (2) be progressive, which is ensured by the following result.

Lemma 32.1 *(progressive functions) Let f be a progressive function on $\mathsf{R}_+ \times C_{\mathsf{R}_+,\mathsf{R}^d}$ for the induced filtration $\mathcal{G}$, and let X be a continuous, $\mathcal{F}$-adapted process in R^d. Then the process $Y_t = f(t, X)$ is $\mathcal{F}$-progressive.*

Proof: Fix any $t \ge 0$. Since X is adapted, we note that $\pi_s(X) = X_s$ is $\mathcal{F}_t$-measurable for every $s \le t$, where $\pi_s(x) = x_s$ on $C_{\mathsf{R}_+,\mathsf{R}^d}$. Since $\mathcal{G}_t = \sigma\{\pi_s;\, s \le t\}$, Lemma 1.4 shows that X is $\mathcal{F}_t/\mathcal{G}_t$-measurable. Hence, by Lemma 1.9, the mapping $\varphi(s,\omega) = \{s, X(\omega)\}$ is $(\mathcal{B}_t \otimes \mathcal{F}_t)/(\mathcal{B}_t \otimes \mathcal{G}_t)$-measurable from $[0,t] \times \Omega$ to $[0,t] \times C_{\mathsf{R}_+,\mathsf{R}^d}$, where $\mathcal{B}_t = \mathcal{B}[0,t]$. Also note that f is $(\mathcal{B}_t \otimes \mathcal{G}_t)$-measurable on $[0,t] \times C_{\mathsf{R}_+,\mathsf{R}^d}$ since f is progressive. Then Lemma 1.7 shows that $Y = f \circ \varphi$ is $(\mathcal{B}_t \otimes \mathcal{F}_t)/\mathcal{B}$-measurable on $[0,t] \times \Omega$. $\square$

Equation (2) exhibits the solution process X as a semi-martingale in R^d with drift components $b^i(X) \cdot \lambda$ and covariation processes $[X^i, X^j] = a^{ij}(X) \cdot \lambda$, where $a^{ij}(x) = a^{ij}(\cdot, x)$ and $b^i(x) = b^i(\cdot, x)$. The densities $a(t,X)$ and $b(t,X)$ may be regarded as *local characteristics* of X at time t. Of special interest is the *diffusion case*, where σ and b have the form

$$\begin{aligned} \sigma(t,x) &= \sigma(x_t), \\ b(t,x) &= b(x_t), \qquad t \ge 0,\ x \in C_{\mathsf{R}_+,\mathsf{R}^d}, \end{aligned}$$

for some measurable functions on R^d. Then the local characteristics at time t depend only on the current position X_t of the process, and the progressivity holds automatically.

We will distinguish between strong and weak solutions to an SDE (σ, b). For the former, we regard the filtered probability space $(\Omega, \mathcal{F}, P)$ as given, along with an $\mathcal{F}$-Brownian motion B and an $\mathcal{F}_0$-measurable random vector ξ. A *strong solution* is then defined as an adapted process X with $X_0 = \xi$ a.s.

[1]short for *progressively measurable*

satisfying (1). For a *weak solution*, only the initial distribution μ is given, and the solution consists of a triple $(\Omega, \mathcal{F}, P)$, along with an $\mathcal{F}$-Brownian motion B and an adapted process X with $\mathcal{L}(X_0) = \mu$ satisfying (1).

This leads to different notions of existence and uniqueness, for a given equation (σ, b). We say that *weak existence* holds for the initial distribution μ, if there exists a corresponding weak solution $(\Omega, \mathcal{F}, P, B, X)$. By contrast, *strong existence* for the given μ means that, for any basic triple $(\mathcal{F}, B, \xi)$ with $\mathcal{L}(\xi) = \mu$, there exists a strong solution X with $X_0 = \xi$ a.s. Next, we say that *uniqueness in law* holds for the initial distribution μ, if all weak solutions X with $\mathcal{L}(X_0) = \mu$ have the same distribution. Finally, *pathwise uniqueness* for a given distribution μ means that, whenever X and Y are two solutions with $X_0 = Y_0 = \xi$ a.s. and $\mathcal{L}(\xi) = \mu$, defined on a common filtered probability space with a given Brownian motion B, we have $X = Y$ a.s.

One of the simplest SDEs is the *Langevin equation*

$$dX_t = dB_t - X_t\, dt, \tag{4}$$

of importance for both theory and applications. Integrating by parts gives

$$\begin{aligned} d\Big(e^t X_t\Big) &= e^t\, dX_t + e^t X_t\, dt \\ &= e^t\, dB_t, \end{aligned}$$

which yields the explicit solution

$$X_t = e^{-t} X_0 + \int_0^t e^{-(t-s)}\, dB_s, \quad t \geq 0, \tag{5}$$

recognized as an *Ornstein–Uhlenbeck process*. Conversely, the latter process is easily seen to satisfy (4). We further note that $\theta_t X \xrightarrow{d} Y$ as $t \to \infty$, where Y denotes the stationary version of the process, encountered in Chapters 14 and 21. We can also get a stationary version[2] directly from (5) by choosing X_0 to be $N(0, \frac{1}{2})$ and independent of B.

We turn to a more general class of equations that can be solved explicitly. A further extension appears in Theorem 20.8.

Proposition 32.2 *(linear equations) Let U, V be continuous semi-martingales, and put $Z = \exp\Big(V - V_0 - \frac{1}{2}[V]\Big)$. Then the equation $dX = dU + X dV$ has the unique solution*

$$X = Z\Big\{X_0 + Z^{-1} \cdot \Big(U - [U, V]\Big)\Big\}. \tag{6}$$

Proof: Define $Y = X/Z$. Integrating by parts and noting that $dZ = Z\, dV$, we get

$$\begin{aligned} dU &= dX - X dV \\ &= Y dZ + Z\, dY + d\,[Y, Z] - X dV \\ &= Z\, dY + d\,[Y, Z]. \end{aligned} \tag{7}$$

[2]For full agreement with Chapter 14, we need to replace B by $B\sqrt{2}$ in (4) and (5).

In particular,

$$[U,V] = Z \cdot [Y,V] = [Y,Z].$$

Substituting this into (7) yields $Z\,dY = dU - d\,[U,V]$, which implies $dY = Z^{-1}d(U - [U,V])$. Now (6) follows as we integrate from 0 to t and note that $Y_0 = X_0$. Since all steps are reversible, the same argument shows that (6) is indeed a solution. □

Though most SDEs can not be solved explicitly, we may still derive some general conditions for strong existence, pathwise uniqueness, and continuous dependence on initial conditions, using the classical method of *Picard iteration*, originally devised for ordinary differential equations.

Theorem 32.3 *(strong solutions and stochastic flows, Itô) Let σ, b be bounded, progressive functions satisfying a Lipschitz condition*[3]

$$\left\{\sigma(w) - \sigma(w')\right\}_t^* + \left\{b(w) - b(w')\right\}_t^* \lesssim (w - w')_t^*, \quad t \geq 0, \tag{8}$$

and consider a Brownian motion B in R^r w.r.t. a complete filtration $\mathcal{F}$. Then there exists a jointly continuous process $X = (X_t^x)$ on $\mathsf{R}_+ \times \mathsf{R}^d$, such that for any $\mathcal{F}_0$-measurable random vector ξ in R^d, equation (σ, b) has the a.s. unique solution X^ξ starting at ξ.

For one-dimensional diffusion equations, a stronger result is established in Theorem 33.3. The solution process $X = (X_t^x)$ on $\mathsf{R}_+ \times \mathsf{R}^d$ is called the *stochastic flow* generated by B. Our proof is based on two lemmas, beginning with an elementary classical estimate.

Lemma 32.4 *(Gronwall) Let f be a continuous function on R_+, such that*

$$f(t) \leq a + b \int_0^t f(s)\,ds, \quad t \geq 0, \tag{9}$$

for some constants $a, b \geq 0$. Then

$$f(t) \leq a\,e^{bt}, \quad t \geq 0.$$

Proof: We may write (9) as

$$\frac{d}{dt}\left\{ e^{-bt} \int_0^t f(s)\,ds \right\} \leq a\,e^{-bt}, \quad t \geq 0.$$

Now integrate over $[0,t]$ and combine with (9). □

To state the next result, let S_X denote the process given by the right-hand side of (2).

[3] Recall that $\lesssim$ denotes inequality up to a constant factor.

Lemma 32.5 *(L^p-contraction) Let σ, b be bounded, progressive functions satisfying (8). Then for any $p \geq 2$, there exists a non-decreasing function $c \geq 0$ on R_+ such that, for any continuous, adapted processes X, Y in R^d,*

$$E\Big(S_X - S_Y\Big)_t^{*p} \leq 2\,E\Big|X_0 - Y_0\Big|^p + c_t \int_0^t E(X - Y)_s^{*p}\,ds, \quad t \geq 0.$$

Proof: By Theorem 18.7, condition (8), and Jensen's inequality,

$$\begin{aligned} & E\Big(S_X - S_Y\Big)_t^{*p} - 2E\Big|X_0 - Y_0\Big|^p \\ &\quad \lesssim\; E\Big\{(\sigma_X - \sigma_Y)\cdot B\Big\}_t^{*p} + E\Big\{(b_X - b_Y)\cdot\lambda\Big\}_t^{*p} \\ &\quad \lesssim\; E\Big(|\,\sigma_X - \sigma_Y|^2\cdot\lambda\Big)_t^{p/2} + E\Big(|\,b_X - b_Y|\cdot\lambda\Big)_t^{p} \\ &\quad \lesssim\; E\left|\int_0^t (X - Y)_s^{*2}\,ds\right|^{p/2} + E\left|\int_0^t (X - Y)_s^{*}\,ds\right|^{p} \\ &\quad \leq\; \Big(t^{p/2-1} + t^{p-1}\Big)\int_0^t E(X - Y)_s^{*p}\,ds. \end{aligned}$$

□

Proof of Theorem 32.3: To prove the existence, fix any $\mathcal{F}_0$-measurable random vector ξ in R^d, put $X_t^0 \equiv \xi$, and define recursively $X^n = S_{X^{n-1}}$ for $n \geq 1$. Since σ, b are bounded, we have $E(X^1 - X^0)_t^{*2} < \infty$, and Lemma 32.5 yields

$$E\Big(X^{n+1} - X^n\Big)_t^{*2} \leq c_t \int_0^t E\Big(X^n - X^{n-1}\Big)_s^{*2}\,ds, \quad t \geq 0,\ n \geq 1.$$

Hence, by induction,

$$E\Big(X^{n+1} - X^n\Big)_t^{*2} \leq \frac{c_t^n\,t^n}{n!}\,E\Big(X^1 - \xi\Big)_t^{*2} < \infty, \quad t, n \geq 0.$$

For any $k \in \mathsf{N}$, we get

$$\begin{aligned} \Big\|\sup_{n\geq k}\Big(X^n - X^k\Big)_t^*\Big\|_2 &\leq \sum_{n\geq k}\Big\|\Big(X^{n+1} - X^n\Big)_t^*\Big\|_2 \\ &\leq \Big\|\Big(X^1 - \xi\Big)_t^*\Big\|_2 \sum_{n\geq k}\Big(c_t^n\,t^n/n!\Big)^{1/2} < \infty. \end{aligned}$$

Thus, Lemma 5.6 yields a continuous, adapted process X with $X_0 = \xi$, such that $(X^n - X)_t^* \to 0$ a.s. and in L^2 for each $t \geq 0$. To see that X solves equation (σ, b), we conclude from Lemma 32.5 that

$$E\Big(X^n - S_X\Big)_t^{*2} \leq c_t \int_0^t E\Big(X^{n-1} - X\Big)_s^{*2}\,ds, \quad t \geq 0.$$

As $n \to \infty$, we get $E(X - S_X)_t^{*2} = 0$ for all t, which implies $X = S_X$ a.s.

Now consider two solutions X, Y with $|X_0 - Y_0| \leq \varepsilon$ a.s. By Lemma 32.5, we get for any $p \geq 2$

$$E(X - Y)_t^{*p} \leq 2\,\varepsilon^p + c_t \int_0^t E(X - Y)_s^{*p}\,ds, \quad t \geq 0,$$

and by Lemma 32.4 it follows that

$$E(X - Y)_t^{*p} \le 2\,\varepsilon^p\, e^{c_t t}, \quad t \ge 0. \tag{10}$$

If $X_0 = Y_0$ a.s., we may take $\varepsilon = 0$ and conclude that $X = Y$ a.s., which proves the asserted uniqueness. Letting X^x denote the solution X with $X_0 = x$ a.s., we get by (10)

$$E\Big|X^x - X^y\Big|_t^{*p} \le 2\,|x-y|^p\, e^{c_t t}, \quad t \ge 0.$$

Taking $p > d$ and applying Theorem 4.23 for each $T > 0$ with the metric $\rho_T(f,g) = (f-g)_T^*$, we conclude that the process (X_t^x) has a jointly continuous version on $\mathsf{R}_+ \times \mathsf{R}^d$.

The construction shows that if X, Y are solutions with $X_0 = \xi$ and $Y_0 = \eta$ a.s., then $X = Y$ a.s. on the set $\{\xi = \eta\}$. In particular, $X = X^\xi$ a.s. when ξ takes only countably many values. In general, ξ may be uniformly approximated by some random vectors $\xi_1, \xi_2, \dots$ in Q^d, and (10) yields $X_t^{\xi_n} \to X_t$ in L^2 for all $t \ge 0$. Since also $X_t^{\xi_n} \to X_t^\xi$ a.s. by the continuity of the flow, it follows that $X_t = X_t^\xi$ a.s. □

For many SDEs, the solutions may explode after a finite time. To deal with this possibility, we introduce as in Chapter 17 an extra absorbing state[4] Δ at infinity, so that the path space becomes $C_{\mathsf{R}_+,\overline{\mathsf{R}^d}}$ with $\overline{\mathsf{R}^d} = \mathsf{R}^d \cup \{\Delta\}$. Define $\zeta_n = \inf\{t;\, |X_t| \ge n\}$ for each n, put $\zeta = \sup_n \zeta_n$, and let $X_t = \Delta$ for $t \ge \zeta$. Given a Brownian motion B in R^r and an adapted process X in the extended path space, we say that X or the pair (X, B) solves equation (σ, b) on the interval $[0, \zeta)$, if

$$X_{t\wedge\zeta_n} = X_0 + \int_0^{t\wedge\zeta_n} \sigma(s, X)\, dB_s + \int_0^{t\wedge\zeta_n} b(s, X)\, ds, \quad t \ge 0,\ n \in \mathsf{N}. \tag{11}$$

When $\zeta < \infty$, we have $|X_{\zeta_n}| \to \infty$, and X is said to *explode* at time ζ.

Conditions for existence and uniqueness of possibly exploding solutions are obtainable from Theorem 32.3 by suitable localization. The following result can sometimes be used to exclude the possibility of explosion.

Proposition 32.6 *(explosion) For solutions to equation (σ, b), no explosion occurs a.s. when*

$$\sigma(x)_t^* + b(x)_t^* \lesssim 1 + x_t^*, \quad t \ge 0. \tag{12}$$

Proof: By Proposition 18.15 we may assume that X_0 is bounded. From (11) and (12), we get for suitable constants $c_t < \infty$

$$EX_{t\wedge\zeta_n}^{*2} \le 2\,E|X_0|^2 + c_t \int_0^t \Big(1 + EX_{s\wedge\zeta_n}^{*2}\Big)\, ds, \quad t \ge 0,\ n \in \mathsf{N},$$

and so by Lemma 32.4

$$1 + EX_{t\wedge\zeta_n}^{*2} \le \Big(1 + 2\,E|X_0|^2\Big) \exp(c_t\, t) < \infty, \quad t \ge 0,\ n \in \mathsf{N}.$$

[4] often referred to as the *coffin state* or *cemetery*

As $n \to \infty$ we obtain $EX^{*2}_{t\wedge\zeta} < \infty$, which implies $\zeta > t$ a.s. □

Our next aim is to characterize weak solutions to equation (σ, b) by a martingale property, involving only the solution process X. Then define

$$M^f_t = f(X_t) - f(X_0) - \int_0^t A_s f(X)\, ds, \quad t \ge 0, \;\; f \in \hat{C}^\infty, \tag{13}$$

where the operators A_s are given by

$$A_s f(x) = \tfrac{1}{2}\, a^{ij}(s,x)\, f''_{ij}(x_s) + b^i(s,x)\, f'_i(x_s), \quad s \ge 0, \;\; f \in \hat{C}^\infty. \tag{14}$$

In the diffusion case we may replace the integrand $A_s f(X)$ in (13) by the expression $Af(X_s)$, where A denotes the elliptic operator

$$Af(x) = \tfrac{1}{2}\, a^{ij}(x)\, f''_{ij}(x) + b^i(x)\, f'_i(x), \quad f \in \hat{C}^\infty, \;\; x \in \mathsf{R}^d. \tag{15}$$

A continuous process X in R^d, or its distribution P, is said to solve the *local martingale problem* for (a, b), if M^f is a local martingale for every $f \in \hat{C}^\infty$. For bounded a, b it is clearly equivalent that M^f be a true martingale, and the original problem turns into a *martingale problem*. The (local) martingale problem for (a, b) with initial distribution μ is said to be *well posed* if it has a unique solution P_μ. For degenerate initial distributions δ_x, we may write P_x instead of P_{δ_x}. We state the basic equivalence between weak solutions to an SDE and solutions to the associated local martingale problem. Here $\hat{\mathcal{M}}_S$ denotes the class of probability measures on S.

Theorem 32.7 *(weak solutions and martingale problems, Stroock & Varadhan) For progressive functions σ, b and distributions $P \in \hat{\mathcal{M}}(C_{\mathsf{R}_+,\mathsf{R}^d})$, these conditions are equivalent:*

(i) *equation (σ, b) has a weak solution with distribution P,*
(ii) *P solves the local martingale problem for $(\sigma\sigma', b)$.*

Proof: Write $a = \sigma\sigma'$. If (X, B) solves equation (σ, b), then

$$\begin{aligned}[X^i, X^j] &= \left[\sigma^i_k(X)\cdot B^k,\, \sigma^j_l(X)\cdot B^l\right]\\ &= \sigma^i_k\sigma^j_l(X)\cdot[B^k, B^l]\\ &= a^{ij}(X)\cdot\lambda.\end{aligned}$$

By Itô's formula, we get for any $f \in \hat{C}^\infty$

$$\begin{aligned}df(X_t) &= f'_i(X_t)\, dX^i_t + \tfrac{1}{2}\, f''_{ij}(X_t)\, d[X^i, X^j]_t\\ &= f'_i(X_t)\, \sigma^i_j(t, X)\, dB^j_t + A_t f(X)\, dt.\end{aligned}$$

Hence, $dM^f_t = f'_i(X_t)\, \sigma^i_j(t,X)\, dB^j_t$, and so M^f is a local martingale.

Conversely, let X solve the local martingale problem for (a, b). When $f^i_n \in \hat{C}^\infty$ with $f^i_n(x) = x^i$ for $|x| \le n$, a localization argument shows that the processes

$$M^i_t = X^i_t - X^i_0 - \int_0^t b^i(s, X)\, ds, \quad t \ge 0, \tag{16}$$

are continuous local martingales. Similarly, choosing $f_n^{ij} \in \hat{C}^\infty$ with $f_n^{ij}(x) = x^i x^j$ for $|x| \le n$, we may form the local martingales

$$M^{ij} = X^i X^j - X_0^i X_0^j - \left(X^i \beta^j + X^j \beta^i + \alpha^{ij}\right) \cdot \lambda,$$

where $\alpha^{ij} = a^{ij}(X)$ and $\beta^i = b^i(X)$. Integrating by parts and using (16), we get

$$\begin{aligned} M^{ij} &= X^i \cdot X^j + X^j \cdot X^i + [X^i, X^j] - \left(X^i \beta^j + X^j \beta^i + \alpha^{ij}\right) \cdot \lambda \\ &= X^i \cdot M^j + X^j \cdot M^i + [M^i, M^j] - \alpha^{ij} \cdot \lambda. \end{aligned}$$

Here the last two terms on the right form a local martingale, and so by Proposition 18.2,

$$[M^i, M^j]_t = \int_0^t a^{ij}(s, X)\, ds, \quad t \ge 0.$$

Hence, Theorem 19.13 yields a Brownian motion B with respect to a standard extension of the original filtration, such that

$$M_t^i = \int_0^t \sigma_k^i(s, X)\, dB_s^k, \quad t \ge 0.$$

Substituting this into (16) yields (2), which means that the pair (X, B) solves equation (σ, b). □

For subsequent needs, we note that the previous construction can be made measurable, in the following sense.

Lemma 32.8 *(functional representation) For any progressive functions σ, b, there exists a measurable map*

$$F : \hat{\mathcal{M}}(C_{\mathbb{R}_+, \mathbb{R}^d}) \times C_{\mathbb{R}_+, \mathbb{R}^d} \times [0, 1] \to C_{\mathbb{R}_+, \mathbb{R}^r},$$

such that when X solves the local martingale problem for $(\sigma\sigma', b)$ with distribution P, and $\vartheta \perp\!\!\!\perp X$ is $U(0, 1)$, we have

(i) *$B = F(P, X, \vartheta)$ is a Brownian motion in $\mathbb{R}^r$,*
(ii) *the pair (X, B) with induced filtration solves equation (σ, b).*

Proof: In the previous construction of B, the only non-elementary step is the stochastic integration with respect to (X, Y) in Theorem 19.13, where Y is an independent Brownian motion and the integrand is a progressive function of X, obtained by some elementary matrix algebra. Since the pair (X, Y) is again a solution to a local martingale problem, Proposition 18.26 yields the desired functional representation. □

Combining the martingale formulation with a compactness argument, we may deduce some general existence and continuity properties.

Theorem 32.9 *(weak existence and continuity, Skorohod) Let σ, b be bounded, progressive functions, such that $\sigma(t, \cdot)$ and $b(t, \cdot)$ are continuous on $C_{\mathbb{R}_+, \mathbb{R}^d}$ for every $t \ge 0$. Then*

(i) *the martingale problem for* $(\sigma\sigma', b)$ *has a solution* P_μ *for every initial distribution* μ,

(ii) *when the* P_μ *are unique, the mapping* $\mu \mapsto P_\mu$ *is weakly continuous.*

Proof: (i) For any $\varepsilon > 0$, $t \ge 0$, and $x \in C_{\mathsf{R}_+,\mathsf{R}^d}$, define

$$\sigma_\varepsilon(t,x) = \sigma\{(t-\varepsilon)_+,\, x\},$$
$$b_\varepsilon(t,x) = b\{(t-\varepsilon)_+,\, x\},$$

and let $a_\varepsilon = \sigma_\varepsilon \sigma_\varepsilon'$. Since σ, b are progressive, the processes $\sigma_\varepsilon(s,X)$ and $b_\varepsilon(s,X)$, $s \le t$, are measurable functions of X on $[0, (t-\varepsilon)_+]$. Hence, a strong solution X^ε to equation $(\sigma_\varepsilon, b_\varepsilon)$ may be constructed recursively on the intervals $[(n-1)\varepsilon,\, n\varepsilon]$, $n \in \mathsf{N}$, starting from an arbitrary random vector $\xi \perp\!\!\!\perp B$ in R^d with distribution μ. In particular, X^ε solves the martingale problem for the pair $(a_\varepsilon, b_\varepsilon)$.

Applying Theorem 18.7 to equation $(\sigma_\varepsilon, b_\varepsilon)$ and using the boundedness of σ, b, we get for any $p > 0$

$$E \sup_{0 \le r \le h} \left| X^\varepsilon_{t+r} - X^\varepsilon_t \right|^p \lesssim h^{p/2} + h^p \lesssim h^{p/2}, \quad t, \varepsilon \ge 0,\ \ h \in [0,1].$$

Taking $p > 2d$, we see from Corollary 23.7 that the family $\{X^\varepsilon\}$ is tight in $C_{\mathsf{R}_+,\mathsf{R}^d}$. By Theorem 23.2, we may then choose some $\varepsilon_n \to 0$, such that $X^{\varepsilon_n} \xrightarrow{d} X$ for a suitable X. To see that X solves the martingale problem for (a,b), let $f \in \hat{C}^\infty$ and $s < t$ be arbitrary, and consider any bounded, continuous function $g \colon C_{[0,s],\mathsf{R}^d} \to \mathsf{R}$. We need to show that

$$E\Big\{ f(X_t) - f(X_s) - \int_s^t A_r f(X) dr \Big\}\, g(X) = 0.$$

Then note that X^ε satisfies the corresponding equation for the operators A^ε_r, constructed from the pair $(a_\varepsilon, b_\varepsilon)$. Writing the two conditions as $E\varphi(X) = 0$ and $E\varphi_\varepsilon(X^\varepsilon) = 0$, respectively, it suffices by Theorem 5.27 to show that $\varphi_\varepsilon(x_\varepsilon) \to \varphi(x)$, whenever $x_\varepsilon \to x$ in $C_{\mathsf{R}_+,\mathsf{R}^d}$. This follows easily from the continuity conditions imposed on σ, b.

(ii) Assuming the solutions P_μ to be unique, let $\mu_n \xrightarrow{w} \mu$. Arguing as before, we see that (P_{μ_n}) is tight, and so by Theorem 23.2 it is also relatively compact. If $P_{\mu_n} \xrightarrow{w} Q$ along a sub-sequence, we see as before that Q solves the martingale problem for (a,b) with initial distribution μ. Hence $Q = P_\mu$, and the convergence extends to the original sequence. □

The well-posedness of the local martingale problem for (a,b) may now be extended from degenerate to arbitrary initial distributions. This requires a basic measurability property, which will also be needed later.

Theorem 32.10 *(measurability and mixtures, Stroock & Varadhan) Let* a, b *be progressive functions, such that for any* $x \in \mathsf{R}^d$ *the local martingale problem for* (a,b) *with initial distribution* δ_x *has a unique solution* P_x. *Then*

(i) *the* P_x *form a kernel from* R^d *to* $C_{\mathsf{R}_+,\mathsf{R}^d}$,

(ii) *for any initial distribution* μ, *the associated local martingale problem has the unique solution* $P_\mu = \int P_x\, \mu(dx)$.

Proof: By the proof of Theorem 32.7, it is enough to state the local martingale problem in terms of functions f belonging to some countable subclass $\mathcal{C} \subset \hat{C}^\infty$, consisting of suitably truncated versions of the coordinate functions x^i and their products $x^i x^j$. Now define $\mathcal{P} = \hat{\mathcal{M}}(C_{\mathsf{R}^d,\mathsf{R}^d})$ and $\mathcal{P}_M = \{P_x;\, x \in \mathsf{R}^d\}$, and write X for the canonical process in $C_{\mathsf{R}_+,\mathsf{R}^d}$. Let D be the class of measures $P \in \mathcal{P}$ with degenerate projections $P \circ X_0^{-1}$. Next let I consist of all measures $P \in \mathcal{P}$, such that X satisfies the integrability condition (3). Finally, put $\tau_n^f = \inf\{t;\, |M_t^f| \geq n\}$, and let L be the class of measures $P \in \mathcal{P}$, such that the processes $M_t^{f,n} = M^f(t \wedge \tau_n^f)$ exist and are martingales under P for all $f \in \mathcal{C}$ and $n \in \mathsf{N}$. Then clearly $\mathcal{P}_M = D \cap I \cap L$.

(i) It suffices to show that $\mathcal{P}_M$ is a measurable subset of $\mathcal{P}$, since the desired measurability will then follow by Theorem A1.1 and Lemma 3.2. The measurability of D is clear from Theorem 2.18. Even I is measurable, since the integrals on the left of (3) are measurable by Fubini's theorem. Finally, $L \cap I$ is a measurable subset of I, since the defining condition is equivalent to countably many relations of the form $E\Big(M_t^{f,n} - M_s^{f,n}; F\Big) = 0$ with $f \in \mathcal{C}$, $n \in \mathsf{N}$, $s < t$ in Q_+, and $F \in \mathcal{F}_s$.

(ii) For any probability measure μ on R^d, the measure $P_\mu = \int P_x\, \mu(dx)$ has clearly initial distribution μ, and the previous argument shows that even P_μ solves the local martingale problem for (a, b). To prove the uniqueness, let P be any measure with the stated properties. Then $E\Big(M_t^{f,n} - M_s^{f,n}; F \,\Big|\, X_0\Big) = 0$ a.s. for all f, n, $s < t$, and F as above, and so $P(\,\cdot\,|X_0)$ a.s. solves the local martingale problem with initial distribution δ_{X_0}. Thus, $P(\,\cdot\,|X_0) = P_{X_0}$ a.s., and we get $P = EP_{X_0} = \int P_x\, \mu(dx) = P_\mu$. This extends the well-posedness to arbitrary initial distributions. □

We return to the problem of constructing a Feller diffusion with generator A as in (15), solving a suitable SDE or the associated martingale problem. The following result may be regarded as a converse to Theorem 17.24.

Theorem 32.11 *(strong Markov and Feller properties, Stroock & Varadhan) Let* a, b *be measurable functions on* R^d, *such that for any* $x \in \mathsf{R}^d$ *the local martingale problem for* (a, b) *with initial distribution* δ_x *has a unique solution* P_x. *Then*

(i) *the family* (P_x) *satisfies the strong Markov property,*

(ii) *when* a, b *are bounded, continuous,* $T_t f(x) = E_x f(X_t)$ *defines a Feller semigroup on* C_0, *and the operator* A *in* (15) *extends uniquely to the associated generator.*

Proof: (i) By Theorem 32.10 it remains to prove that, for any state $x \in \mathsf{R}^d$ and bounded optional time τ,

$$\mathcal{L}_x\Big(X \circ \theta_\tau \,\Big|\, \mathcal{F}_\tau\Big) = P_{X_\tau} \quad \text{a.s.}$$

As in the previous proof, this is equivalent to countably many relations of the form

$$E_x\Big\{1_F\Big(M_t^{f,n}-M_s^{f,n}\Big)\circ\theta_\tau\Big|\,\mathcal{F}_\tau\Big\}=0\ \text{ a.s.}\tag{17}$$

with $s<t$ and $F\in\mathcal{F}_s$, where $M^{f,n}$ denotes the process M^f stopped at $\tau_n=\inf\{t;\,|M^f|\geq n\}$. Now $\theta_\tau^{-1}\mathcal{F}_s\subset\mathcal{F}_{\tau+s}$ by Lemma 9.5, and in the diffusion case,

$$\Big(M_t^{f,n}-M_s^{f,n}\Big)\circ\theta_\tau=M^f_{(\tau+t)\wedge\sigma_n}-M^f_{(\tau+s)\wedge\sigma_n},$$

where $\sigma_n=\tau+\tau_n\circ\theta_\tau$, which is again optional by Proposition 11.8. Thus, (17) follows by optional sampling from the local martingale property of M^f under P_x.

(ii) Assume that a,b are also bounded and continuous, and define $T_tf(x)=E_xf(X_t)$. By Theorem 32.9, the function T_tf is continuous for every $f\in C_0$ and $t>0$, and the path continuity implies that $T_tf(x)$ is continuous in t for every x. To see that $T_tf\in C_0$, it remains to show that $|X_t^x|\xrightarrow{P}\infty$ as $|x|\to\infty$, where X^x has distribution P_x. This follows from the SDE by the boundedness of σ,b, if for $0<r<|x|$ we write

$$\begin{aligned}P\Big\{|X_t^x|<r\Big\}&\leq P\Big\{\Big|X_t^x-x\Big|>|x|-r\Big\}\\&\leq\frac{E\Big|X_t^x-x\Big|^2}{(|x|-r)^2}\lesssim\frac{t+t^2}{(|x|-r)^2},\end{aligned}$$

and let $|x|\to\infty$ for fixed r,t. The last assertion is obvious from the uniqueness in law together with Theorem 17.23. □

Uniqueness in law is usually harder to prove than the weak existence. Some fairly general uniqueness criteria will be obtained in Theorems 33.1 and 34.2. For the moment, we will only exhibit some transformations that may simplify the problem. First we show how the drift may sometimes be eliminated by a change of probability measure.

Proposition 32.12 *(transformation of drift) Let σ, b, c be progressive functions of suitable dimension, where c is bounded. Then*
(i) *weak existence holds simultaneously for equations (σ,b) and $(\sigma,b+\sigma c)$,*
(ii) *when $c=\sigma' h$ for a progressive function h, uniqueness in law holds simultaneously for the two equations.*

Proof: (i) Let X be a weak solution to equation (σ,b), defined on the canonical space for (X,B) with induced filtration $\mathcal{F}$ and probability measure P. Put $V=c(X)$, and note that $(V^2\cdot\lambda)_t$ is bounded for each t. By Lemma 19.19 and Corollary 19.26, there exists a probability measure Q with $Q=\mathcal{E}(V'\cdot B)_t\cdot P$ on $\mathcal{F}_t$ for each $t\geq 0$, and we note that $\tilde{B}=B-V\cdot\lambda$ is a Q-Brownian motion. Under Q, we further get by Proposition 19.21

$$\begin{aligned}X-X_0&=\sigma(X)\cdot\Big(\tilde{B}+V\cdot\lambda\Big)+b(X)\cdot\lambda\\&=\sigma(X)\cdot\tilde{B}+(b+\sigma c)(X)\cdot\lambda,\end{aligned}$$

which shows that X is a weak solution to the SDE $(\sigma, b+\sigma c)$. Since the same argument applies to equation $(\sigma, b+\sigma c)$ with c replaced by $-c$, we conclude that weak existence holds simultaneously for the two equations.

(ii) Let $c = \sigma' h$, and suppose that uniqueness in law holds for equation $(\sigma, b + ah)$. Further, let (X, B) solve equation (σ, b) under both P and Q. Choosing V and $\tilde{B}$ as before, it follows that $(X, \tilde{B})$ solves equation $(\sigma, b+\sigma c)$, under the transformed distributions $\mathcal{E}(V'\cdot B)_t \cdot P$ and $\mathcal{E}(V'\cdot B)_t \cdot Q$ for (X, B). By hypothesis, the latter measures have then the same X-marginal, and the stated condition shows that $\mathcal{E}(V'\cdot B)$ is X-measurable. Thus, the X-marginals agree even for P, Q, which proves the uniqueness in law for equation (σ, b). Reversing the argument yields an implication in the other direction. □

Next we show how an SDE of diffusion type can be transformed by a random time change. The method will be used systematically in Chapter 33, to analyze the one-dimensional case.

Proposition 32.13 *(scaling) Let* σ, b, c *be measurable functions on* R^d*, where the range of* c *has compact closure in* $(0,\infty)$*. Then weak existence and uniqueness in law hold simultaneously for equations* (σ, b) *and* $(c\,\sigma, c^2 b)$.

Proof: Let X solve the local martingale problem for the pair (a, b), and introduce the process $V = c^2(X)\cdot\lambda$ with inverse (τ_s). By optional sampling, we note that $M^f_{\tau_s}$, $s \geq 0$, is again a local martingale, and the process $Y_s = X_{\tau_s}$ satisfies

$$M^f_{\tau_s} = f(Y_s) - f(Y_0) - \int_0^s c^2 Af(Y_r)\,dr.$$

Thus, Y solves the local martingale problem for (c^2a, c^2b).

Now let T be the map on $C_{\mathsf{R}_+,\mathsf{R}^d}$ transforming X into Y, and write T' for the corresponding map based on c^{-1}, so that T and T' are mutual inverses. Applying the previous argument to both maps, we conclude that a measure $P \in \hat{\mathcal{M}}(C_{\mathsf{R}_+,\mathsf{R}^d})$ solves the local martingale problem for (a, b) iff $P \circ T^{-1}$ solves the corresponding problem for (c^2a, c^2b). Thus, both existence and uniqueness hold simultaneously for the two problems. By Theorem 32.7, the last statement translates immediately into a corresponding assertion for the SDEs. □

We finally explore the relationship between weak and strong solutions. Under suitable conditions, we further prove the existence of a universal functional solution. To explain the terminology, let $\mathcal{G}$ be the filtration induced by the identity map (ξ, B) on the canonical space $\Omega = \mathsf{R}^d \times C_{\mathsf{R}_+,\mathsf{R}^r}$, so that $\mathcal{G}_t = \sigma\{\xi, B^t\}$, $t \geq 0$, where $B^t_s = B_{s\wedge t}$. Writing W^r for the r-dimensional Wiener measure, we introduce for every $\mu \in \hat{\mathcal{M}}_{\mathsf{R}^d}$ the $(\mu \otimes W^r)$-completion $\mathcal{G}^\mu_t$ of $\mathcal{G}_t$. The *universal completion* $\bar{\mathcal{G}}_t$ is defined as $\bigcap_\mu \mathcal{G}^\mu_t$, and we say that a function

$$F\colon\ \mathsf{R}^d \times C_{\mathsf{R}_+,\mathsf{R}^r} \to C_{\mathsf{R}_+,\mathsf{R}^d} \tag{18}$$

is *universally adapted* if it is adapted to the filtration $\bar{\mathcal{G}} = (\bar{\mathcal{G}}_t)$.

Theorem 32.14 *(existence, uniqueness, and functional solution, Yamada & Watanabe, OK) Let the functions σ, b be progressive and such that weak existence and pathwise uniqueness hold for solutions to equation (σ, b) starting at fixed points. Then*

(i) *strong existence and uniqueness in law hold for any initial distribution,*

(ii) *there exists a measurable, universally adapted function F as in* (18)*, such that every solution (X, B) to equation (σ, b) satisfies $X = F(X_0, B)$ a.s.*

Note that the function F in (ii) can be chosen to be independent of initial distribution μ. A couple of lemmas are needed for the proof, beginning with a statement clarifying the relationship between adaptedness, strong existence, and functional solutions.

Lemma 32.15 *(transfer of strong solution) Let (X, B) solve equation (σ, b), where X is adapted to the complete filtration induced by X_0 and B. Then*

(i) *$X = F(X_0, B)$ a.s. for a Borel measurable function F as in* (18)*,*

(ii) *for any basic triple $(\mathcal{F}, \tilde{B}, \xi)$ with $\xi \stackrel{d}{=} X_0$, the process $\tilde{X} = F(\xi, \tilde{B})$ is $\mathcal{F}$-adapted, and the pair $(\tilde{X}, \tilde{B})$ solves equation (σ, b).*

Proof: By Lemma 1.14, we have $X = F(X_0, B)$ a.s. for some Borel measurable function F as stated. The same result yields for every $t \geq 0$ a representation $X_t = G_t(X_0, B^t)$ a.s., and so $F(X_0, B)_t = G_t(X_0, B^t)$ a.s. Hence, $\tilde{X}_t = G_t(\xi, \tilde{B}^t)$ a.s., and so $\tilde{X}$ is $\mathcal{F}$-adapted. Since also $(\tilde{X}, \tilde{B}) \stackrel{d}{=} (X, B)$, Proposition 18.26 shows that even the former pair solves equation (σ, b). $\square$

Even weak solutions can be transferred to any probability space with a specified Brownian motion:

Lemma 32.16 *(transfer of weak solution) Let (X, B) solve equation (σ, b), and let $(\mathcal{F}, \tilde{B}, \xi)$ be a basic triple with $\xi \stackrel{d}{=} X_0$. Then*

(i) *there exists a process $\tilde{X}$ with*

$$\tilde{X} \underset{\xi, \tilde{B}}{\perp\!\!\!\perp} \mathcal{F}, \qquad \tilde{X}_0 = \xi \text{ a.s.}, \qquad (\tilde{X}, \tilde{B}) \stackrel{d}{=} (X, B),$$

(ii) *the filtration $\mathcal{G}$ induced by $(\tilde{X}, \mathcal{F})$ is a standard extension of $\mathcal{F}$,*

(iii) *the pair $(\tilde{X}, \tilde{B})$ with filtration $\mathcal{G}$ solves equation (σ, b).*

Proof: (i) By Theorem 8.17 and Proposition 8.20, there exists a process $\tilde{X} \perp\!\!\!\perp_{\xi, \tilde{B}} \mathcal{F}$ with $(\tilde{X}, \xi, \tilde{B}) \stackrel{d}{=} (X, X_0, B)$, and in particular $\tilde{X}_0 = \xi$ a.s.

(ii) Fix any $t \geq 0$, and put $\tilde{B}' = \tilde{B} - \tilde{B}^t$. Then $(\tilde{X}^t, \tilde{B}^t) \perp\!\!\!\perp \tilde{B}'$, since the corresponding relation holds for (X, B), and so $\tilde{X}^t \perp\!\!\!\perp_{\xi, \tilde{B}^t} \tilde{B}'$. Since also $\tilde{X}^t \perp\!\!\!\perp_{\xi, \tilde{B}} \mathcal{F}$, Proposition 8.12 yields $\tilde{X}^t \perp\!\!\!\perp_{\xi, \tilde{B}^t} (\tilde{B}', \mathcal{F})$, and hence $\tilde{X}^t \perp\!\!\!\perp_{\mathcal{F}_t} \mathcal{F}$. Then also $(\tilde{X}^t, \mathcal{F}_t) \perp\!\!\!\perp_{\mathcal{F}_t} \mathcal{F}$ by Corollary 8.11, which means that $\mathcal{G}_t \perp\!\!\!\perp_{\mathcal{F}_t} \mathcal{F}$.

(iii) Since standard extensions preserve martingales, Theorem 19.3 shows that $\tilde{B}$ remains a Brownian motion with respect to $\mathcal{G}$. As in Proposition 18.26, we conclude that the pair $(\tilde{X}, \tilde{B})$ solves equation (σ, b). □

Proof of Theorem 32.14: For clarity, we begin with a special case.

1. Fixed initial distribution. Let (X, B) solve equation (σ, b) with initial distribution μ and filtration $\mathcal{F}$. Then Lemma 32.16 yields a process $Y \perp\!\!\!\perp_{X_0,B} \mathcal{F}$ with $Y_0 = X_0$ a.s., such that (Y, B) solves equation (σ, b) for the filtration $\mathcal{G}$ induced by $(Y, \mathcal{F})$. Since $\mathcal{G}$ is a standard extension of $\mathcal{F}$, the pair (X, B) remains a solution for $\mathcal{G}$, and the pathwise uniqueness yields $X = Y$ a.s.

For every $t \geq 0$, we have $X^t \perp\!\!\!\perp_{X_0,B} X^t$ and $(X^t, B^t) \perp\!\!\!\perp (B - B^t)$, and so $X^t \perp\!\!\!\perp_{X_0,B^t} X^t$ a.s. by Proposition 8.12. Thus, Corollary 8.11 (ii) shows that X is adapted to the complete filtration induced by (X_0, B). Hence, Lemma 32.15 yields a measurable function F_μ with $X = F_\mu(X_0, B)$ a.s., such that for any basic triple $(\tilde{\mathcal{F}}, \tilde{B}, \xi)$ with $\xi \overset{d}{=} X_0$, the process $\tilde{X} = F_\mu(\xi, \tilde{B})$ is $\tilde{\mathcal{F}}$-adapted and solves equation (σ, b) along with $\tilde{B}$. In particular, $\tilde{X} \overset{d}{=} X$ since $(\xi, \tilde{B}) \overset{d}{=} (X_0, B)$, and the pathwise uniqueness shows that $\tilde{X}$ is the a.s. unique solution for the given triple $(\tilde{\mathcal{F}}, \tilde{B}, \xi)$. This proves the uniqueness in law.

2. General case. The previous case yields uniqueness in law for solutions starting at fixed points, and Theorem 32.10 shows that the corresponding distributions P_x form a kernel from R^d to $C_{\mathsf{R}_+,\mathsf{R}^d}$. Now Lemma 32.8 yields a measurable mapping G, such that for a process X with distribution P_x and a $U(0,1)$ random variable $\vartheta \perp\!\!\!\perp X$, the process $B = G(P_x, X, \vartheta)$ is a Brownian motion in R^r, and the pair (X, B) solves equation (σ, b). Writing Q_x for the distribution of (X, B), we see from Theorem 2.19 (v) and Lemma 3.2 (ii) that the mapping $x \mapsto Q_x$ is a kernel from R^d to $C_{\mathsf{R}_+,\mathsf{R}^{d+r}}$.

Changing the notation, write (X, B) for the canonical process in $C_{\mathsf{R}_+,\mathsf{R}^{d+r}}$. The special case yields $X = F_x(x, B) = F_x(B)$ a.s. Q_x, and so

$$Q_x(X \in \cdot \mid B) = \delta_{F_x(B)} \ \text{ a.s.,} \quad x \in \mathsf{R}^d. \tag{19}$$

By Proposition 9.27, we may choose versions $\nu_{x,w} = Q_x(X \in \cdot \mid B \in dw)$, combining into a probability kernel $\nu : \mathsf{R}^d \times C_{\mathsf{R}_+,\mathsf{R}^r} \to C_{\mathsf{R}_+,\mathsf{R}^d}$. Here (19) shows that $\nu_{x,w}$ is a.s. degenerate for each x, and since the set D of degenerate measures is measurable by Lemma 2.18, we can modify ν to get $\nu_{x,w} D \equiv 1$. Then

$$\nu_{x,w} = \delta_{F(x,w)}, \quad x \in \mathsf{R}^d, \ w \in C_{\mathsf{R}_+,\mathsf{R}^r}, \tag{20}$$

for some function F as in (18), which is product-measurable, by the kernel property of ν. Comparing (19) and (20) gives $F(x, B) = F_x(B)$ a.s. for all x.

Now fix any probability measure μ on R^d, and conclude as in Theorem 32.10 that $P_\mu = \int P_x\, \mu(dx)$ solves the local martingale problem for (a, b) with initial distribution μ. Hence, equation (σ, b) has a solution (X, B) with distribution μ for X_0. Since conditioning on $\mathcal{F}_0$ preserves martingales, the equation remains conditionally valid, given X_0. The pathwise uniqueness in the degenerate case yields $P\{X = F(X_0, B) \mid X_0\} = 1$ a.s., and so $X = F(X_0, B)$ a.s. In particular, the pathwise uniqueness extends to any initial distribution μ.

Returning to the canonical setting, let (ξ, B) be the identity map on the canonical space $\mathsf{R}^d \times C_{\mathsf{R}_+,\mathsf{R}^r}$, endowed with the probability measure $\mu \otimes W^r$ and induced complete filtration $\mathcal{G}^\mu$. By the result in the special case, equation (σ, b) has a $\mathcal{G}^\mu$-adapted solution $X = F_\mu(\xi, B)$ with $X_0 = \xi$ a.s., and the previous discussion yields even $X = F(\xi, B)$ a.s. Hence, F is adapted to $\mathcal{G}^\mu$, and μ being arbitrary, the adaptedness extends to the universal completion $\bar{\mathcal{G}}_t = \bigcap_\mu \mathcal{G}_t^\mu$, $t \geq 0$. $\square$

Exercises

1. Show that for any $c \in (0, 1)$, the stochastic flow X_t^x in Theorem 32.3 is a.s. Hölder continuous in x with exponent c, uniformly for bounded x and t. (*Hint:* Apply Theorem 4.23 to the estimate in the proof of Theorem 32.3.)

2. Show that a process X in R^d is a Brownian motion iff the process $f(X_t) - \frac{1}{2}\int_0^t \Delta f(X_s)\, ds$ is a martingale for every $f \in \hat{C}^\infty$. Compare with Theorem 19.3 and Lemma 17.21.

3. Show that a Brownian bridge in R^d satisfies the SDE $dX_t = dB_t - (1-t)^{-1}X_t\, dt$ on $[0, 1)$ with initial condition $X_0 = 0$. Further show that if X^x is the solution starting at x, then $Y_t^x = X_t^x - (1-t)x$ is again a Brownian bridge. (*Hint:* Note that $M_t = X_t/(1-t)$ is a martingale on $[0, 1)$, and that Y^x satisfies the same SDE as X.)

4. Solve the preceding SDE, using Proposition 32.2 to express the Brownian bridge in terms of a Brownian motion. Compare with previously known formulas.

5. Given two continuous semi-martingales U, V, show that the Fisk–Stratonovich SDE $dX = dU + X \circ dV$ has the unique solution $X = Z(X_0 + Z^{-1} \circ U)$, where $Z = \exp(V - V_0)$. (*Hint:* Use Corollary 18.21 and the chain rule for FS-integrals, or derive the result from Proposition 32.2.)

6. Under suitable conditions, show how a Fisk–Stratonovich SDE can be converted into an Itô equation, and conversely. Also give a sufficient condition for the existence of a strong solution to an FS-equation.

7. Show that weak existence and uniqueness in law hold for the SDE $dX_t = \operatorname{sgn}(X_t+)\, dB_t$ with initial condition $X_0 = 0$, whereas strong existence and pathwise uniqueness fail. (*Hint:* Show that any solution X is a Brownian motion, and define $B = \operatorname{sgn}(X+) \cdot X$. Note that both X and $-X$ satisfy the given SDE.)

8. Show that weak existence holds for the SDE $dX_t = \operatorname{sgn}(X_t)\, dB_t$ with initial condition $X_0 = 0$, whereas strong existence and uniqueness in law fail. (*Hint:* We may take X to be a Brownian motion or put $X \equiv 0$.)

9. Show that strong existence holds for the SDE $dX_t = 1\{X_t \neq 0\}\, dB_t$ with initial condition $X_0 = 0$, whereas uniqueness in law fails. (*Hint:* Here $X = B$ and $X = 0$ are both solutions.)

10. Show that a given process X may satisfy SDEs with different pairs $(\sigma\sigma', b)$. (*Hint:* For a trivial example, take $X = 0$, $b = 0$, and $\sigma = 0$ or $\sigma(x) = \operatorname{sgn} x$.)

11. Construct a non-Markovian solution X to the SDE $dX_t = \mathrm{sgn}(X_t)\, dB_t$. (*Hint:* Let X be a Brownian motion, stopped at the first visit to 0 after time 1. We might also choose X to be 0 on $[0,1]$ and a Brownian motion on $[1,\infty)$.)

12. For X as in Theorem 32.3, construct an SDE in R^{md} satisfied by the process $(X_t^{x_1},\dots,X_t^{x_m})$ for arbitrary $x_1,\dots,x_m \in \mathsf{R}^d$. Conclude that $\mathcal{L}(X)$ is determined by $\mathcal{L}(X^x, X^y)$ for arbitrary $x,y \in \mathsf{R}^d$. (*Hint:* Note that $\mathcal{L}(X^x)$ is determined by $(\sigma\sigma', b)$ and x, and apply this result to the m-point motion.)

13. Find two SDEs as in Theorem 32.3 with solutions X, Y, such that $X^x \overset{d}{=} Y^x$ for all x but $X \overset{d}{\neq} Y$. (*Hint:* We may choose $dX = dB$ and $dY = \mathrm{sgn}(Y+)\, dB$.)

14. For a diffusion equation (σ, b) as in Theorem 32.3, show that the distribution of the associated flow X determines $\sum_j \sigma_j^i(x)\, \sigma_j^k(y)$ for arbitrary pairs $i,k \in \{1,\dots,d\}$ and $x,y \in \mathsf{R}^d$.

15. Show that if weak existence holds for the SDE (σ, b), then the pathwise uniqueness can be strengthened to the corresponding property for solutions X, Y with respect to possibly different filtrations.

16. Assume that weak existence and the stronger version of pathwise uniqueness hold for the SDE (σ, b). Use Theorem 8.17 and Lemma 32.15 to prove existence for every μ of an a.s. unique functional solution $F_\mu(X_0, B)$ with $\mathcal{L}(X_0) = \mu$.

Chapter 33

One-Dimensional SDEs and Diffusions

Removal of drift, weak existence and uniqueness, path-wise uniqueness, weak and strict comparison, hitting times, scale function, speed measure, time-change reduction, Green function, accessible and reflecting boundaries, entrance and exit boundaries, entrance laws and Feller properties, ratio ergodic theorem, recurrence and ergodicity, strong ergodicity, positive recurrence and invariant distribution

We have already recognized martingales and Markov processes as the most important dependence structures of modern probability. In both cases, Brownian motion is the prototype and most fundamental example. In fact, we proved in Theorem 19.4 that a real, continuous local martingale is nothing but a suitably time-changed Brownian motion. Similarly, we will show in Theorem 33.9 below that even a real and sufficiently regular continuous Markov process is essentially a Brownian motion, up to a natural transformation in both space and time.

Much more can be said in the one-dimensional case of continuous strong Markov processes—also known as *diffusion processes*—which motivates that we spend a whole chapter to explore their basic properties. We begin with a study of the basic *diffusion equation*

$$dX_t = \sigma(X_t)\,dB_t + b(X_t)\,dt, \qquad (1)$$

referred to below as equation (σ, b). Here we will see how the drift term can often be removed by a simple spatial transformation. This reduces (1) to the simple equation $dX_t = \sigma(X_t)\,dB_t$, where precise criteria can be given for weak existence and uniqueness in law, in terms of the function σ. For more general σ and b, we may give conditions for pathwise uniqueness and strict or weak comparison. From Theorem 32.11 we know that, if weak existence and uniqueness in law hold for (1), the solution process X is a continuous strong Markov process. It is clearly also a semi-martingale.

We proceed with a study of general, one-dimensional diffusion processes X. Our first aim is then to show how a spatial transformation, based on a suitable *scale function*, reduces X to a continuous local martingale, which is then said to be on a *natural scale*. In the latter case, we may next identify a suitable *speed measure*, determining the time-change reduction of X into a Brownian motion. The tools for this construction are purely probabilistic, as they are

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_34

based on an analysis of the hitting times τ_a and $\tau_{a,b}$ of states a or interval endpoints $a<b$. The ergodic behavior of X also depends in a crucial way on the properties at the boundary points, which may be absorbing or reflecting, as well as of entrance or exit type.

Preparing for a more detailed discussion of equation (σ,b), we recall from Proposition 32.12 how the drift term can sometimes be eliminated through a change of the underlying probability measure. Under suitable regularity conditions on the coefficients, we may use the alternative approach of transforming the state space. Then assume that X solves (1), and put $Y_t=p(X_t)$ for some function $p\in C^1$, possessing an absolutely continuous derivative p' with density p''. By the generalized Itô formula of Theorem 29.5, we have

$$\begin{aligned} dY_t &= p'(X_t)\,dX_t+\tfrac{1}{2}\,p''(X_t)\,d[X]_t \\ &= (\sigma p')(X_t)\,dB_t+\left(\tfrac{1}{2}\,\sigma^2p''+b\,p'\right)(X_t)\,dt. \end{aligned}$$

Here the drift term vanishes iff p solves the ordinary differential equation

$$\tfrac{1}{2}\,\sigma^2p''+b\,p'=0. \tag{2}$$

When b/σ^2 is locally integrable, (2) has the explicit solutions

$$p'(x)=c\,\exp\Big\{-2\int_0^x(b/\sigma^2)(u)\,du\Big\},\quad x\in\mathsf{R},$$

for an arbitrary constant c. The desired scale function p is then determined up to an affine transformation, and for $c>0$ it is strictly increasing with a unique inverse p^{-1}. A mapping by p reduces (1) to the form $dY_t=\tilde\sigma(Y_t)\,dB_t$, where $\tilde\sigma=(\sigma p')\circ p^{-1}$. Since the new equation is equivalent, it is clear that weak or strong existence or uniqueness hold simultaneously for the two equations.

Once the drift is removed, we are left with an equation of the form

$$dX_t=\sigma(X_t)\,dB_t. \tag{3}$$

Here exact criteria for weak existence and uniqueness in law may be given in terms of the singularity sets

$$\begin{aligned} S_\sigma &= \Big\{x\in\mathsf{R};\ \int_{x-}^{x+}\sigma^{-2}(y)\,dy=\infty\Big\}, \\ N_\sigma &= \Big\{x\in\mathsf{R};\ \sigma(x)=0\Big\}. \end{aligned}$$

Theorem 33.1 *(existence and uniqueness, Engelbert & Schmidt) For equation* (3) *with initial distribution* μ*,*

(i) *weak existence holds for all* $\mu\ \Leftrightarrow\ S_\sigma\subset N_\sigma$*,*
(ii) *uniqueness in law then holds for all* $\mu\ \Leftrightarrow\ S_\sigma=N_\sigma$*.*

We begin with a lemma, which will also be useful later on. Given any measure ν on R, we may introduce the associated singularity set

$$S_\nu=\Big\{x\in\mathsf{R};\ \nu(x-,\,x+)=\infty\Big\}.$$

Letting B be a one-dimensional Brownian motion with associated local time L, we may also introduce the additive functional

$$A_s = \int L_s^x \, \nu(dx), \quad s \geq 0. \tag{4}$$

Lemma 33.2 *(singularity set) Let B be a Brownian motion with local time L and initial distribution μ, and define A by* (4) *for a measure ν on* R. *Then a.s. P_μ,*

$$\inf\Big\{s \geq 0;\, A_s = \infty\Big\} = \inf\Big\{s \geq 0;\, B_s \in S_\nu\Big\}.$$

Proof: Fix any $t > 0$, and let R be the event where $B_s \notin S_\nu$ on $[0, t]$. Noting that $L_t^x = 0$ a.s. for x outside the range $B[0, t]$, we get a.s. on R

$$\begin{aligned} A_t &= \int_{-\infty}^{\infty} L_t^x \, \nu(dx) \\ &\leq \nu\Big(B[0,t]\Big) \sup\nolimits_x L_t^x < \infty, \end{aligned}$$

since $B[0, t]$ is compact and L_t^x is a.s. continuous and hence bounded.

Conversely, let $B_s \in S_\nu$ for some $s < t$. To show that $A_t = \infty$ a.s. on this event, we may use the strong Markov property to reduce to the case where $B_0 = a$ is non-random in S_ν. But then $L_t^a > 0$ a.s. by Tanaka's formula, and so the continuity of L yields for small enough $\varepsilon > 0$

$$\begin{aligned} A_t &= \int_{-\infty}^{\infty} L_t^x \, \nu(dx) \\ &\geq \nu\Big(a - \varepsilon, a + \varepsilon\Big) \inf_{|x-a|<\varepsilon} L_t^x = \infty. \end{aligned} \qquad \square$$

Proof of Theorem 33.1: (i) Let $S_\sigma \subset N_\sigma$. To prove the asserted weak existence, let Y be a Brownian motion with arbitrary initial distribution μ, and define $\zeta = \inf\{s \geq 0;\, Y_s \in S_\sigma\}$. By Lemma 33.2, the additive functional

$$A_s = \int_0^s \sigma^{-2}(Y_r) \, dr, \quad s \geq 0, \tag{5}$$

is continuous and strictly increasing on $[0, \zeta)$, and $A_t = \infty$ for $t > \zeta$. Further note that $A_\zeta = \infty$ when $\zeta = \infty$, whereas A_ζ may be finite when $\zeta < \infty$. In the latter case, A jumps from A_ζ to ∞ at time ζ.

Now introduce the inverse

$$\tau_t = \inf\Big\{s > 0;\, A_s > t\Big\}, \quad t \geq 0. \tag{6}$$

The process τ is clearly continuous and strictly increasing on $[0, A_\zeta]$, and for $t \geq A_\zeta$ we have $\tau_t = \zeta$. Further note that $X_t = Y_{\tau_t}$ is a continuous local martingale, and

$$\begin{aligned} t = A_{\tau_t} &= \int_0^{\tau_t} \sigma^{-2}(Y_r) \, dr \\ &= \int_0^t \sigma^{-2}(X_s) \, d\tau_s, \quad t < A_\zeta. \end{aligned}$$

Hence, for $t \leq A_\zeta$,

$$[X]_t = \tau_t = \int_0^t \sigma^2(X_s) \, ds. \tag{7}$$

Here both sides remain constant after time A_ζ, since $S_\sigma \subset N_\sigma$, and so (7) remains true for all $t \geq 0$. Hence, Theorem 19.13 yields a Brownian motion B satisfying (3), which means that X is a weak solution with initial distribution μ.

Conversely, assume weak existence for any initial distribution. To show that $S_\sigma \subset N_\sigma$, fix any $x \in S_\sigma$, and choose a solution X with $X_0 = x$. Since X is a continuous local martingale, Theorem 19.4 yields $X_t = Y_{\tau_t}$ for a Brownian motion Y starting at x and a random time-change τ satisfying (7). For A as in (5) and $t \geq 0$, we have

$$\begin{aligned} A_{\tau_t} &= \int_0^{\tau_t} \sigma^{-2}(Y_r)\, dr \\ &= \int_0^t \sigma^{-2}(X_s)\, d\tau_s \\ &= \int_0^t 1\Big\{\sigma(X_s) > 0\Big\}\, ds \leq t. \end{aligned} \tag{8}$$

Since $A_s = \infty$ for $s > 0$ by Lemma 33.2, we get $\tau_t = 0$ a.s., and so $X_t \equiv x$ a.s. But then $x \in N_\sigma$ by (7).

(ii) Let $N_\sigma \subset S_\sigma$, and consider a solution X with initial distribution μ. As before, we may write $X_t = Y_{\tau_t}$ a.s., where Y is a Brownian motion with initial distribution μ, and τ is a random time-change satisfying (7). Define A as in (5), put $\chi = \inf\{t \geq 0;\, X_t \in S_\sigma\}$, and note that $\tau_\chi = \zeta \equiv \inf\{s \geq 0;\, Y_s \in S_\sigma\}$. Since $N_\sigma \subset S_\sigma$, we get as in (8)

$$A_{\tau_t} = \int_0^{\tau_t} \sigma^{-2}(Y_s)\, ds = t, \quad t \leq \chi.$$

Furthermore, $A_s = \infty$ for $s > \zeta$ by Lemma 33.2, and so (8) yields $\tau_t \leq \zeta$ a.s. for all t, which means that τ remains constant after time χ. Thus, τ and A are related by (6), which shows that τ and then also X are measurable functions of Y. Since the distribution of Y depends only on μ, the same thing is true for X, which proves the asserted uniqueness in law.

Conversely, let S_σ be a proper subset of N_σ, and fix any $x \in N_\sigma \setminus S_\sigma$. As before, we may form a solution $X_t = Y_{\tau_t}$ starting at x, where Y is a Brownian motion starting at x, and τ is defined as in (6) from the process A in (5). Since $x \notin S_\sigma$, Lemma 33.2 gives $A_{0+} < \infty$ a.s., and so $\tau_t > 0$ a.s. for $t > 0$, which shows that X is a.s. non-constant. Since $x \in N_\sigma$, (3) has also the trivial solution $X_t \equiv x$. Thus, uniqueness in law fails for solutions starting at x. □

Proceeding with a study of pathwise uniqueness, we return to equation (1), writing $w(\sigma, \cdot)$ for the modulus of continuity of σ.

Theorem 33.3 *(pathwise uniqueness, Skorohod, Yamada & Watanabe) Let σ, b be bounded, measurable functions on* R*, such that*

$$\int_0^\varepsilon \Big\{w(\sigma, h)\Big\}^{-2} dh = \infty, \quad \varepsilon > 0, \tag{9}$$

and either b is Lipschitz continuous or $\sigma \neq 0$. Then pathwise uniqueness holds for equation (σ, b).

The significance of (9) is clear from the following lemma, where for any semi-martingale Y we write $L_t^x(Y)$ for the associated local time.

Lemma 33.4 *(local time) Let X^i solve equation (σ, b_i) for $i = 1, 2$, where σ satisfies* (9)*. Then*

$$L^0\Big(X^1 - X^2\Big) = 0 \;\; a.s.$$

Proof: Write $Y = X^1 - X^2$, $L_t^x = L_t^x(Y)$, and $w(x) = w(\sigma, |x|)$. Using (1) and Theorem 29.5, we get for any $t > 0$

$$\begin{aligned}\int_{-\infty}^{\infty} w_x^{-2} L_t^x \, dx &= \int_0^t \Big| w(Y_s) \Big|^{-2} d[Y]_s \\ &= \int_0^t \left| \frac{\sigma(X_s^1) - \sigma(X_s^2)}{w\Big(X_s^1 - X_s^2\Big)} \right|^2 ds \le t < \infty.\end{aligned}$$

By (1) and the right-continuity of L it follows that $L_t^0 = 0$ a.s. □

Proof of Theorem 33.3 *for $\sigma \neq 0$:* By Propositions 32.12 and 32.13 along with a simple localization argument, uniqueness in law holds for equation (σ, b) when $\sigma \neq 0$. To prove the pathwise uniqueness, consider any two solutions X and Y with $X_0 = Y_0$ a.s. Using Tanaka's formula, Lemma 33.4, and equation (σ, b), we get

$$\begin{aligned}d(X_t \vee Y_t) &= dX_t + d(Y_t - X_t)^+ \\ &= dX_t + 1\{Y_t > X_t\} \, d(Y_t - X_t) \\ &= 1\{Y_t \le X_t\} \, dX_t + 1\{Y_t > X_t\} \, dY_t \\ &= \sigma(X_t \vee Y_t) \, dB_t + b(X_t \vee Y_t) \, dt,\end{aligned}$$

which shows that $X \vee Y$ is again a solution. The uniqueness in law gives $X \overset{d}{=} X \vee Y$. Since $X \le X \vee Y$, it follows that $X = X \vee Y$ a.s., which implies $Y \le X$ a.s. The symmetric argument yields $X \le Y$ a.s. □

The assertion for Lipschitz continuous b is a special case of the following comparison result.

Theorem 33.5 *(weak comparison, Skorohod, Yamada) For $i = 1, 2$, let X^i solve equation (σ, b_i), where σ satisfies* (9) *and b_1 or b_2 is Lipschitz continuous. Then*

$$\left.\begin{aligned} b_1 \ge b_2 \\ X_0^1 \ge X_0^2 \;\; a.s. \end{aligned}\right\} \;\Rightarrow\; X^1 \ge X^2 \;\; a.s. \; on \; \mathsf{R}_+.$$

Proof: By symmetry we may take b_1 to be Lipschitz continuous. Since $X_0^2 \le X_0^1$ a.s., we get by Tanaka's formula and Lemma 33.4

$$\begin{aligned}\Big(X_t^2 - X_t^1\Big)^+ = \int_0^t 1\Big\{X_s^2 > X_s^1\Big\} \Big\{\sigma(X_t^2) - \sigma(X_t^1)\Big\} \, dB_t \\ + \int_0^t 1\Big\{X_s^2 > X_s^1\Big\} \Big\{b_2(X_s^2) - b_1(X_s^1)\Big\} \, ds.\end{aligned}$$

Using the martingale property of the first term, the Lipschitz continuity of b_1, and the condition $b_2 \le b_1$, we conclude that

$$\begin{aligned} E\Big(X_t^2 - X_t^1\Big)^+ &\le E\int_0^t 1\Big\{X_s^2 > X_s^1\Big\}\Big\{b_1(X_s^2) - b_1(X_s^1)\Big\}\,ds \\ &\lesssim E\int_0^t 1\Big\{X_s^2 > X_s^1\Big\}\,\Big|X_s^2 - X_s^1\Big|\,ds \\ &= \int_0^t E\Big(X_s^2 - X_s^1\Big)^+ ds. \end{aligned}$$

Then $E(X_t^2 - X_t^1)^+ = 0$ by Gronwall's lemma, and so $X_t^2 \le X_t^1$ a.s. □

By imposing stronger restrictions on the coefficients, we may strengthen the last conclusion to a strict inequality.

Theorem 33.6 *(strict comparison) For $i = 1, 2$, let X^i solve equation (σ, b_i), where σ is Lipschitz continuous and b_1, b_2 are continuous. Then*

$$\left.\begin{array}{r} b_1 > b_2 \\ X_0^1 \ge X_0^2 \ \ a.s. \end{array}\right\} \ \Rightarrow \ X^1 > X^2 \ \ a.s.\ on\ (0,\infty).$$

Proof: Since the b_i are continuous with $b_1 > b_2$, there exists a locally Lipschitz continuous function b on R with $b_1 > b > b_2$. By Theorem 32.3, equation (σ, b) has a solution X with $X_0 = X_0^1 \ge X_0^2$ a.s., and it suffices to show that $X^1 > X > X^2$ a.s. on $(0, \infty)$. This yields a reduction to the case where one of the functions b_i is locally Lipschitz. By symmetry we may take that function to be b_1.

By the Lipschitz continuity of σ and b_1, we may introduce the continuous semi-martingales

$$\begin{aligned} U_t &= \int_0^t \Big\{b_1(X_s^2) - b_2(X_s^2)\Big\}\,ds, \\ V_t &= \int_0^t \frac{\sigma(X_s^1) - \sigma(X_s^2)}{X_s^1 - X_s^2}\,dB_s + \int_0^t \frac{b_1(X_s^1) - b_1(X_s^2)}{X_s^1 - X_s^2}\,ds, \end{aligned}$$

subject to the convention $0/0 = 0$, and we note that

$$d\Big(X_t^1 - X_t^2\Big) = dU_t + \Big(X_t^1 - X_t^2\Big)\,dV_t.$$

Letting $Z = \exp\Big(V - \frac{1}{2}[V]\Big) > 0$, we get by Proposition 32.2

$$X_t^1 - X_t^2 = Z_t\Big(X_0^1 - X_0^2\Big) + Z_t\int_0^t Z_s^{-1}\Big\{b_1(X_s^2) - b_2(X_s^2)\Big\}\,ds,$$

and the assertion follows since $X_0^1 \ge X_0^2$ a.s. and $b_1 > b_2$. □

We turn to a systematic study of one-dimensional diffusions. By a *diffusion* in an interval $I \subset \mathsf{R}$ we mean a continuous strong Markov process taking values in I. Termination is only permitted at open end-points of I. Define

$\tau_y = \inf\{t \ge 0;\, X_t = y\}$, and say that X is *regular* if $P_x\{\tau_y < \infty\} > 0$ for any $x \in I^o$ and $y \in I$. Further write $\tau_{a,b} = \tau_a \wedge \tau_b$.

Our first step is to transform the general diffusion process into a continuous local martingale, using a suitable change of scale. This corresponds to a removal of drift in the SDE (1).

Theorem 33.7 *(scale function, Feller, Dynkin) Let X be a regular diffusion in an interval $I \subset \mathsf{R}$. Then*

(i) *there exists a continuous, strictly increasing function p on I, such that $p(X^{\tau_{a,b}})$ is a P_x-martingale for every $x \in [a,b] \subset I$,*

(ii) *an increasing function p is such as in* (i) *iff*

$$P_x\{\tau_b < \tau_a\} = \frac{p_x - p_a}{p_b - p_a}, \quad x \in [a,b].$$

A function p with the stated property is called a *scale function* for X, and X is said to be defined on a *natural scale* if the scale function can be chosen to be linear. In general, the process $Y = p(X)$ is a regular diffusion on a natural scale. We begin with a study of the functions

$$\begin{aligned} p_{a,b}(x) &= P_x\{\tau_b < \tau_a\}, \\ h_{a,b}(x) &= E_x \tau_{a,b}, \qquad a \le x \le b, \end{aligned}$$

which play a basic role in the sequel.

Lemma 33.8 *(hitting times) For any regular diffusion and constants $a < b$ in I, we have*

(i) *$p_{a,b}$ is continuous and strictly increasing on $[a,b]$,*

(ii) *$h_{a,b}$ is bounded on $[a,b]$.*

In particular, (ii) yields $\tau_{a,b} < \infty$ a.s. under P_x for any $a \le x \le b$.

Proof: (i) First we show that $P_x\{\tau_b < \tau_a\} > 0$ for any $a < x < b$. Then introduce the optional time $\sigma_1 = \tau_a + \tau_x \circ \theta_{\tau_a}$, and define recursively $\sigma_{n+1} = \sigma_n + \sigma_1 \circ \theta_{\sigma_n}$. By the strong Markov property, the σ_n form a random walk in $[0,\infty]$ under each P_x. If $P_x\{\tau_b < \tau_a\} = 0$, we get $\tau_b \ge \sigma_n \to \infty$ a.s. P_x, and so $P_x\{\tau_b = \infty\} = 1$, which contradicts the regularity of X.

Using the strong Markov property at τ_y, we next obtain

$$P_x\{\tau_b < \tau_a\} = P_x\{\tau_y < \tau_a\}\, P_y\{\tau_b < \tau_a\}, \quad a < x < y < b. \tag{10}$$

Since $P_x\{\tau_a < \tau_y\} > 0$, we have $P_x\{\tau_y < \tau_a\} < 1$, which shows that $P_x\{\tau_b < \tau_a\}$ is strictly increasing.

By symmetry it remains to prove that $P_y\{\tau_b < \tau_a\}$ is left-continuous on $(a,b]$. By (10) it is equivalent to show that, for any $x \in (a,b)$, the mapping $y \mapsto P_x\{\tau_y < \tau_a\}$ is left-continuous on $(x,b]$. Then let $y_n \uparrow y$, and note that $\tau_{y_n} \uparrow \tau_y$ a.s. P_x by the continuity of X. Hence, $\{\tau_{y_n} < \tau_a\} \downarrow \{\tau_y < \tau_a\}$, which implies convergence of the corresponding probabilities.

(ii) Fix any $c \in (a,b)$. By the regularity of X, we may choose $h > 0$ so large that

$$P_c\{\tau_a \le h\} \wedge P_c\{\tau_b \le h\} = \delta > 0.$$

When $x \in (a,c)$, the strong Markov property at τ_x yields

$$\begin{aligned}\delta &\le P_c\{\tau_a \le h\}\\ &\le P_c\{\tau_x \le h\}\, P_x\{\tau_a \le h\}\\ &\le P_x\{\tau_a \le h\}\\ &\le P_x\Big\{\tau_{a,b} \le h\Big\},\end{aligned}$$

and similarly for $x \in (c,b)$. By the Markov property at h and induction on n, we obtain

$$P_x\Big\{\tau_{a,b} > nh\Big\} \le (1-\delta)^n, \quad x \in [a,b],\ n \in \mathsf{Z}_+,$$

and Lemma 4.4 yields

$$\begin{aligned}E_x\tau_{a,b} &= \int_0^\infty P_x\Big\{\tau_{a,b} > t\Big\}\, dt\\ &\le h \sum_{n\ge 0} (1-\delta)^n < \infty.\end{aligned}$$ □

Proof of Theorem 33.7: Let p be a locally bounded and measurable function on I, such that $M = p(X^{\tau_{a,b}})$ is a martingale under P_x for any $a < x < b$. Then

$$\begin{aligned}p_x &= E_x M_0 = E_x M_\infty\\ &= E_x\, p(X_{\tau_{a,b}})\\ &= p_a P_x\{\tau_a < \tau_b\} + p_b P_x\{\tau_b < \tau_a\}\\ &= p_a + (p_b - p_a)\, P_x\{\tau_b < \tau_a\},\end{aligned}$$

and (ii) follows provided that $p_a \ne p_b$.

To construct a function p with the stated properties, fix any points $u < v$ in I, and define for arbitrary $a \le u$ and $b \ge v$ in I

$$p(x) = \frac{p_{a,b}(x) - p_{a,b}(u)}{p_{a,b}(v) - p_{a,b}(u)}, \quad x \in [a,b]. \tag{11}$$

To see that p is independent of a and b, consider a larger interval $[a',b'] \subset I$, and conclude from the strong Markov property at $\tau_{a,b}$ that, for any $x \in [a,b]$,

$$\begin{aligned}P_x\{\tau_{b'} < \tau_{a'}\} = {}& P_x\{\tau_a < \tau_b\}\, P_a\{\tau_{b'} < \tau_{a'}\}\\ &+ P_x\{\tau_b < \tau_a\}\, P_b\{\tau_{b'} < \tau_{a'}\},\end{aligned}$$

or

$$p_{a',b'}(x) = p_{a,b}(x)\Big\{\, p_{a',b'}(b) - p_{a',b'}(a)\Big\} + p_{a'b'}(a).$$

Thus, $p_{a,b}$ and $p_{a',b'}$ agree on $[a,b]$ up to an affine mapping, and so they yield the same value in (11).

By Lemma 33.8, the constructed function is continuous and strictly increasing, and it remains to show that $p(X^{\tau_{a,b}})$ is a martingale under P_x for any $a < b$

in I. Since the martingale property is preserved by affine transformations, it is equivalent to show that $p_{a,b}(X^{\tau_{a,b}})$ is a P_x-martingale. Then fix any optional time σ, and write $\tau = \sigma \wedge \tau_{a,b}$. By the strong Markov property at τ,

$$\begin{aligned} E_x\, p_{a,b}(X_\tau) &= E_x P_{X_\tau}\{\tau_b < \tau_a\} \\ &= P_x \theta_\tau^{-1}\{\tau_b < \tau_a\} \\ &= P_x\{\tau_b < \tau_a\} \\ &= p_{a,b}(x), \end{aligned}$$

and the desired martingale property follows by Lemma 9.14. □

To prepare for the next result, consider a Brownian motion B in R with jointly continuous local time L. For any measure ν on R, we introduce as in (4) the additive functional $A = \int L^x\, \nu(dx)$ with right-continuous inverse

$$\sigma_t = \inf\Big\{s > 0;\ A_s > t\Big\}, \quad t \ge 0.$$

If $\nu \neq 0$, the recurrence of B shows that A is a.s. unbounded. Hence, $\sigma_t < \infty$ a.s. for all t, and we may define $X_t = B_{\sigma_t}$, $t \ge 0$. We refer to $\sigma = (\sigma_t)$ as the *random time-change based on* ν, and to the process $X = B \circ \sigma$ as a correspondingly *time-changed Brownian motion.*

Theorem 33.9 *(speed measure and time change, Feller, Volkonsky, Itô & McKean)*

(i) *For a regular diffusion X on a natural scale in I, there exists a unique measure ν on I with $\nu[a,b] \in (0,\infty)$ for all $a < b$ in I^o, such that X is a time-changed Brownian motion based on an extension of ν to $\bar I$.*

(ii) *Any time-changed Brownian motion as above is a regular diffusion in I.*

The extended version of ν is called the *speed measure* of the diffusion. Contrary to what the term suggests, the process moves slowly through regions where ν is large. For Brownian motion itself, the speed measure is clearly Lebesgue measure. More generally, the speed measure of a regular diffusion solving equation (3) has density σ^{-2}.

To prove the uniqueness of ν, we need the following lemma, which is also useful for the subsequent classification of boundary behavior. Here we write $\sigma_{a,b} = \inf\{s > 0;\ B_s \notin (a,b)\}$.

Lemma 33.10 *(Green function) Let X be a time-changed Brownian motion based on ν. Then for any measurable function $f \ge 0$ on I and points $a < b$ in $\bar I$, we have*

$$E_x \int_0^{\tau_{a,b}} f(X_t)\, dt = \int_a^b g_{a,b}(x,y)\, f(y)\, \nu(dy), \quad x \in [a,b], \tag{12}$$

where

$$g_{a,b}(x,y) = E_x L^y_{\sigma_{a,b}} = \frac{2\,(x \wedge y - a)(b - x \vee y)}{b-a}, \quad x,y \in [a,b]. \tag{13}$$

When X is recurrent, this remains true for $a = -\infty$ or $b = \infty$.

Taking $f \equiv 1$ in (12), we get in particular

$$h_{a,b}(x) = E_x \tau_{a,b} = \int_a^b g_{a,b}(x,y)\,\nu(dy), \quad x \in [a,b], \tag{14}$$

which will be useful later on.

Proof: Clearly $\tau_{a,b} = A(\sigma_{a,b})$ for any $a, b \in \bar{I}$, and also for $a = -\infty$ or $b = \infty$ when X is recurrent. Since L^y is supported by $\{y\}$, we see from (4) that

$$\begin{aligned}\int_0^{\tau_{a,b}} f(X_t)\,dt &= \int_0^{\sigma_{a,b}} f(B_s)\,dA_s \\ &= \int_a^b f(y)\,L^y_{\sigma_{a,b}}\,\nu(dy).\end{aligned}$$

Taking expectations gives (12) with $g_{a,b}(x,y) = E_x L^y_{\sigma_{a,b}}$. To prove (13), we get by Tanaka's formula and optional sampling

$$E_x L^y_{\sigma_{a,b}\wedge s} = E_x \Big| B_{\sigma_{a,b}\wedge s} - y \Big| - |x-y|, \quad s \ge 0.$$

If a, b are finite, we may let $s \to \infty$ and conclude by monotone and dominated convergence that

$$g_{a,b}(x,y) = \frac{(y-a)\,(b-x)}{b-a} + \frac{(b-y)\,(x-a)}{b-a} - |x-y|,$$

which simplifies to (13). The result for infinite a or b follows immediately by monotone convergence. □

The next lemma will enable us to construct the speed measure ν from the functions $h_{a,b}$ in Lemma 33.8.

Lemma 33.11 *(consistence) For a regular diffusion on a natural scale in I, there exists a strictly concave function h on I^o, such that for any $a < b$ in I,*

$$h_{a,b}(x) = h(x) - \frac{x-a}{b-a}\,h(b) - \frac{b-x}{b-a}\,h(a), \quad x \in [a,b]. \tag{15}$$

Proof: Fix any $u < v$ in I, and define for any $a \le u$ and $b \ge v$ in I

$$h(x) = h_{a,b}(x) - \frac{x-u}{v-u}\,h_{a,b}(v) - \frac{v-x}{v-u}\,h_{a,b}(u), \quad x \in [a,b]. \tag{16}$$

To see that h is independent of a and b, consider any larger interval $[a', b'] \subset I$, and conclude from the strong Markov property at $\tau_{a,b}$ that, for any $x \in [a,b]$,

$$E_x \tau_{a',b'} = E_x \tau_{a,b} + P_x\{\tau_a < \tau_b\}\,E_a \tau_{a',b'} + P_x\{\tau_b < \tau_a\}\,E_b \tau_{a',b'},$$

or

$$h_{a',b'}(x) = h_{a,b}(x) + \frac{b-x}{b-a}\,h_{a',b'}(a) + \frac{x-a}{b-a}\,h_{a',b'}(b). \tag{17}$$

Thus, $h_{a,b}$ and $h_{a',b'}$ agree on $[a,b]$ up to an affine function and therefore yield the same value in (16).

If $a \le u$ and $b \ge v$, then (16) shows that h and $h_{a,b}$ agree on $[a,b]$ up to an affine function, and (15) follows since $h_{a,b}(a) = h_{a,b}(b) = 0$. The formula extends by (17) to arbitrary $a < b$ in I. $\square$

Since h is strictly concave, its left derivative h'_- is strictly decreasing and left-continuous, and hence determines a measure ν on I^o satisfying

$$2\,\nu[a,b) = h'_-(a) - h'_-(b), \quad a < b \text{ in } I^o. \tag{18}$$

For motivation, we note that this expression is consistent with (14).

To prove Theorem 33.9, we first need to clarify the behavior of X at the endpoints of I. If $b \neq I$ for an endpoint b, then by hypothesis the motion terminates when X reaches b, and we may attach b to I as an absorbing state. For convenience, we may then assume that I is a compact interval of the form $[a,b]$, where either endpoint may be *inaccessible*, in the sense that a.s. it can not be reached in finite time from a point in I^o.

For either endpoint b, the set $\Xi_b = \{t \ge 0;\, X_t = b\}$ is regenerative under P_b in the sense of Chapter 29. In particular, Theorem 29.7 shows that b is either *absorbing*, in the sense that $\Xi_b = \mathsf{R}_+$ a.s., or *reflecting*, in the sense that $\Xi_b^o = \emptyset$ a.s. In the latter case, the reflection is said to be *fast* if $\lambda\,\Xi_b = 0$ and *slow* if $\lambda\,\Xi_b > 0$. We give a more detailed discussion of the boundary behavior after the proof of the main theorem.

First we establish Theorem 33.9 in a special case. The general result will then be deduced by a pathwise comparison.

Proof of Theorem 33.9, *absorbing endpoints (Méléard):* Let X have distribution P_x, where $x \in I^o$, and put $\zeta = \inf\{t > 0;\, X_t \neq I^o\}$. For any $a < b$ in I^o with $x \in [a,b]$, the process $X^{\tau_{a,b}}$ is a continuous martingale, and so by Theorem 29.5

$$h(X_t) = h(x) + \int_0^t h'_-(X)\,dX - \int_I \tilde{L}_t^x\,\nu(dx), \quad t \in [0,\zeta), \tag{19}$$

where $\tilde{L}$ is the local time of X.

Next conclude from Theorem 19.4 that $X = B \circ [X]$ a.s. for a Brownian motion B starting at x. Using Theorem 29.5 twice, we get in particular, for any non-negative measurable function f,

$$\begin{aligned}\int_I f(x)\,\tilde{L}_t^x\,dx &= \int_0^t f(X_s)\,d[X]_s \\ &= \int_0^{[X]_t} f(B_s)\,ds \\ &= \int_I f(x)\,L_{[X]_t}^x\,dt,\end{aligned}$$

where L is the local time of B. Hence, $\tilde{L}_t^x = L_{[X]_t}^x$ a.s. for $t < \zeta$, and so the last term in (19) equals $A_{[X]_t}$ a.s.

For any optional time σ, put $\tau = \sigma \wedge \tau_{a,b}$, and conclude from the strong Markov property that

$$\begin{aligned} E_x\Big\{\tau + h_{a,b}(X_\tau)\Big\} &= E_x\Big(\tau + E_{X_\tau}\tau_{a,b}\Big) \\ &= E_x\Big(\tau + \tau_{a,b} \circ \theta_\tau\Big) \\ &= E_x\tau_{a,b} = h_{a,b}(x). \end{aligned}$$

Writing $M_t = h(X_t) + t$, it follows by Lemma 9.14 that $M^{\tau_{a,b}}$ is a P_x-martingale whenever $x \in [a,b] \subset I^o$. Comparing with (19) and using Proposition 18.2, we obtain $A_{[X]_t} = t$ a.s. for all $t \in [0,\zeta)$. Since A is continuous and strictly increasing on $[0,\zeta)$ with inverse σ, we get $[X]_t = \sigma_t$ a.s. for $t < \zeta$. The last relation extends to $[\zeta,\infty)$, provided we give infinite mass to ν at each endpoint. Then $X = B \circ \sigma$ a.s. on R_+.

Conversely, we note that $B \circ \sigma$ is a regular diffusion on I, whenever σ is a random time-change based on a measure ν with the stated properties. To prove the uniqueness of ν, fix any $a < x < b$ in I^o, and apply Lemma 33.10 with $f(y) = \{g_{a,b}(x,y)\}^{-1}$ to see that $\nu(a,b)$ is determined by P_x. □

Proof of Theorem 33.9, *general case:* Define ν on I^o as in (18), and extend the definition to $\bar I$, by assigning infinite mass to any absorbing endpoint. To reflecting endpoints we attach finite mass, to be specified later. Given a Brownian motion B, we see as before that the correspondingly time-changed process $\tilde X = B \circ \sigma$ is a regular diffusion in I. Letting $\zeta = \sup\{t;\, X_t \in I^o\}$ and $\tilde\zeta = \sup\{t;\, \tilde X_t \in I^o\}$, we further see from the previous case that X^ζ and $\tilde X^{\tilde\zeta}$ have the same distribution for any starting position $x \in I^o$.

Now fix any $a < b$ in I^o, and define recursively

$$\begin{aligned} \chi_1 &= \zeta + \tau_{a,b} \circ \theta_\zeta, \\ \chi_{n+1} &= \chi_n + \chi_1 \circ \theta_{\chi_n}, \quad n \in \mathsf{N}. \end{aligned}$$

Then the processes $Y_n^{a,b} = X^\zeta \circ \theta_{\chi_n}$ form a Markov chain in the path space. A similar construction for $\tilde X$ yields some processes $\tilde Y_n^{a,b}$, and we note that $(Y_n^{a,b}) \overset{d}{=} (\tilde Y_n^{a,b})$ for fixed a and b. Since the process $Y_n^{a',b'}$, obtained for a smaller interval $[a',b']$, can be measurably recovered from that for $[a,b]$, and similarly for $\tilde Y_n^{a',b'}$, the entire collections $(Y_n^{a,b})$ and $(\tilde Y_n^{a,b})$ have the same distribution. By Theorem 8.17, we may then assume that the two families agree a.s.

Now let $I = [a,b]$, where a is reflecting. By the nature of Brownian motion, the level sets Ξ_a and $\tilde\Xi_a$ for X and $\tilde X$ are a.s. perfect, and so we may form the corresponding excursion point processes ξ and $\tilde\xi$, local times L and $\tilde L$, and inverse local times T and $\tilde T$. Since excursions within $[a,b)$ agree a.s. for X and $\tilde X$, we can normalize the excursion laws of the two processes, using the law of large numbers, such that the corresponding parts of ξ and $\tilde\xi$ agree a.s. Then even T and $\tilde T$ agree, possibly apart from the lengths of excursions reaching b and the drift coefficient c in Theorem 29.13. For $\tilde X$ the latter is proportional to the mass $\nu\{a\}$, which can now be chosen such that c becomes the same as for X. This choice of $\nu\{a\}$ is clearly independent of starting position x for X and $\tilde{X}$.

If the other endpoint b is absorbing, then clearly $X = \tilde{X}$ a.s., and the proof is complete. If b is instead reflecting, the excursions from b agree a.s. for X and $\tilde{X}$. Repeating the argument with the roles of a and b interchanged, we get $X = \tilde{X}$ a.s., after a suitable adjustment of the mass $\nu\{b\}$. □

We proceed to classify the boundary behavior of a regular diffusion on a natural scale, in terms of the speed measure ν. A right endpoint b is called an *entrance boundary* for X if it is inaccessible, and yet

$$\lim_{r\to\infty} \inf_{y>x} P_y\{\tau_x \le r\} > 0, \quad x \in I^o. \tag{20}$$

By the Markov property at times nr, $n \in \mathsf{N}$, the limit in (20) then equals 1, and in particular $P_y\{\tau_x < \infty\} = 1$ for all $x < y$ in I^o. In Theorem 33.13 below, we show that an entrance boundary is an endpoint where X may enter but not exit.

The opposite behavior occurs at an *exit boundary*, defined as an endpoint b that is accessible and *naturally absorbing*, in the sense of remaining absorbing when the charge $\nu\{b\}$ is reduced to zero. If b is accessible but not naturally absorbing, we have already seen how the boundary behavior of X depends on the value of $\nu\{b\}$. Thus, in this case, b is absorbing when $\nu\{b\} = \infty$, slowly reflecting when $\nu\{b\} \in (0, \infty)$, and fast reflecting when $\nu\{b\} = 0$. For reflecting b, we further see from Theorem 33.9 that the set $\Xi_b = \{t \ge 0;\, X_t = b\}$ is a.s. perfect.

Theorem 33.12 *(boundary behavior, Feller) Consider a regular diffusion on a natural scale in $I = [a, b]$ with speed measure ν. Then for fixed $u \in I^o$, we have*

(i) *b is accessible iff*
$$b < \infty, \qquad \int_u^b (b - x)\, \nu(dx) < \infty,$$
(ii) *b is accessible, reflecting iff*
$$b < \infty, \qquad \nu(u, b] < \infty,$$
(iii) *b is an entrance boundary iff*
$$b = \infty, \qquad \int_u^b x\, \nu(dx) < \infty.$$

The stated conditions can be translated into similar criteria for any regular diffusion. In general, exit and other accessible boundaries may clearly be infinite, whereas entrance boundaries may be finite. *Explosion* is said to occur when X reaches an infinite boundary in finite time. A basic example of a regular diffusion on $(0, \infty)$ with an entrance boundary at 0 is given by the Bessel process $X_t = |B_t|$, where B is a Brownian motion in R^d with $d \ge 2$.

Proof of Theorem 33.12: (i) Since $\limsup_s(\pm B_s) = \infty$ a.s., Theorem 33.9 shows that X cannot explode, so any accessible endpoint is finite. Now assume that $a < c < u < b < \infty$. Then Lemma 33.8 shows that b is accessible iff $h_{c,b}(u) < \infty$, which holds by (14) iff $\int_u^b (b-x)\,\nu(dx) < \infty$.

(ii) Here $b < \infty$ by (i), in which case Lemma 33.2 shows that b is absorbing iff $\nu(u, b] = \infty$.

(iii) An entrance boundary b is inaccessible by definition, and therefore $\tau_u = \tau_{u,b}$ a.s. when $a < u < b$. Arguing as in the proof of Lemma 33.8, we also note that $E_y \tau_u$ is bounded for $y > u$. If $b < \infty$, we obtain the contradiction $E_y \tau_u = h_{u,b}(y) = \infty$, and so b must be infinite. From (14), we get by monotone convergence as $y \to \infty$

$$\begin{aligned} E_y \tau_u \;&=\; h_{u,\infty}(y) = 2\int_u^\infty \Big(x \wedge y - u\Big)\,\nu(dx) \\ &\to 2\int_u^\infty (x-u)\,\nu(dx), \end{aligned}$$

which is finite iff $\int_u^\infty x\,\nu(dx) < \infty$. □

We proceed to establish an important regularity property, which also clarifies the nature of entrance boundaries.

Theorem 33.13 *(entrance laws and Feller properties) For a regular diffusion X on I, form $\bar{I} \supset I$ by attaching the entrance boundaries of X. Then X extends to a continuous Feller process on $\bar{I}$.*

Proof: For any $f \in C_b$, $a, x \in I$, and $r, t \ge 0$, the strong Markov property at $\tau_x \wedge r$ yields

$$\begin{aligned} E_a f(X_{\tau_x \wedge r + t}) &= E_a T_t f\Big(X_{\tau_x \wedge r}\Big) \\ &= T_t f(x) P_a\{\tau_x \le r\} + E_a\Big\{T_t f(X_r);\, \tau_x > r\Big\}. \end{aligned} \tag{21}$$

To show that $T_t f$ is left-continuous at any $y \in I$, fix an $a < y$ in I^o, and choose $r > 0$ so large that $P_a\{\tau_y \le r\} > 0$. As $x \uparrow y$ we have $\tau_x \uparrow \tau_y$, and so $\{\tau_x \le r\} \downarrow \{\tau_y \le r\}$. Thus, the probabilities and expectations in (21) converge to the corresponding expressions for τ_y, and we get $T_t f(x) \to T_t f(y)$. The proof of the right-continuity is similar.

If an endpoint b is inaccessible but not of entrance type, and if $f(x) \to 0$ as $x \to b$, then clearly even $T_t f(x) \to 0$ at b for each $t > 0$. Now let ∞ be an entrance boundary, and consider a function f with a finite limit at ∞. We need to show that even $T_t f(x)$ converges, as $x \to \infty$ for fixed t. Then conclude from Lemma 33.10 that, as $a \to \infty$,

$$\begin{aligned} \sup_{x \ge a} E_x \tau_a &= 2 \sup_{x \ge a} \int_a^\infty \Big(x \wedge r - a\Big)\,\nu(dr) \\ &= 2\int_a^\infty (r-a)\,\nu(dr) \to 0. \end{aligned} \tag{22}$$

Next we note that, for any $a < x < y$ and $r \ge 0$,

$$\begin{aligned} P_y\{\tau_a \le r\} &\le P_y\Big\{\tau_x \le r,\, \tau_a - \tau_x \le r\Big\} \\ &= P_y\{\tau_x \le r\}\, P_x\{\tau_a \le r\} \\ &\le P_x\{\tau_a \le r\}. \end{aligned}$$

Thus, $P_x \circ \tau_a^{-1}$ converges vaguely as $x \to \infty$ for fixed a, and (22) shows that the convergence holds even in the weak sense.

Now fix any t and f, and introduce for every a the continuous function $g_a(s) = E_a f\{X_{(t-s)+}\}$. By the strong Markov property at $\tau_a \wedge t$ and Theorem 8.5, we get for any $x, y \ge a$

$$\begin{aligned} \Big|T_t f(x) - T_t f(y)\Big| \le \Big|&E_x\, g_a(\tau_a) - E_y\, g_a(\tau_a)\Big| \\ &+ 2\, \|f\| \Big(P_x + P_y\Big)\{\tau_a > t\}. \end{aligned}$$

Here the right-hand side tends to 0 as $x, y \to \infty$ and then $a \to \infty$, by (22) and the weak convergence of $P_x \circ \tau_a^{-1}$. Thus, $T_t f(x)$ is Cauchy convergent as $x \to \infty$, and we may denote the limit by $T_t f(\infty)$.

It is now easy to check that the extended operators T_t form a Feller semigroup on $C_0(\bar{I})$. Finally, Theorem 17.15 shows that the associated process, starting at a possible entrance boundary, has again a continuous version in the topology of $\bar{I}$. □

We may next establish a ratio ergodic theorem, for elementary additive functionals of a recurrent diffusion. It is instructive to compare with the general ratio limit theorems in Chapter 26.

Theorem 33.14 *(ratio ergodic theorem, Derman, Motoo & Watanabe) Let X be a regular, recurrent diffusion on a natural scale in I with speed measure ν. Then for any measurable functions $f, g \ge 0$ on I with $\nu f < \infty$ and $\nu g > 0$, we have as $t \to \infty$*

$$\frac{\int_0^t f(X_s)\, ds}{\int_0^t g(X_s)\, ds} \to \frac{\nu f}{\nu g} \quad \text{a.s. } P_x, \quad x \in I.$$

Proof: Fix any $a < b$ in I, put $\tau_a^b = \tau_b + \tau_a \circ \theta_{\tau_b}$, and define recursively some optional times $\sigma_0, \sigma_1, \dots$ by

$$\sigma_{n+1} = \sigma_n + \tau_a^b \circ \theta_{\sigma_n}, \quad n \ge 0,$$

starting with $\sigma_0 = \tau_a$. Write

$$\int_0^{\sigma_n} f(X_s)\, ds = \int_0^{\sigma_0} f(X_s)\, ds + \sum_{k=1}^{n} \int_{\sigma_{k-1}}^{\sigma_k} f(X_s)\, ds, \tag{23}$$

and note that the last sum has i.i.d. terms. By the strong Markov property and Lemma 33.10, we get for any $x \in I$

$$\begin{aligned} E_x \int_{\sigma_{k-1}}^{\sigma_k} f(X_s)\, ds &= E_a \int_0^{\tau_b} f(X_s)\, ds + E_b \int_0^{\tau_a} f(X_s)\, ds \\ &= \int f(y) \left\{ g_{-\infty,b}(y,a) + g_{a,\infty}(y,b) \right\} \nu(dy) \\ &= 2 \int f(y) \left\{ \left(b - y \vee a\right)_+ + \left(y \wedge b - a\right)_+ \right\} \nu(dy) \\ &= 2\,(b-a)\,\nu f. \end{aligned}$$

The same lemma shows that the first term in (23) is a.s. finite. Hence, the law of large numbers yields

$$\lim_{n\to\infty} n^{-1} \int_0^{\sigma_n} f(X_s)\, ds = 2\,(b-a)\,\nu f \quad \text{a.s. } P_x, \quad x \in I.$$

Writing $\kappa_t = \sup\{n \geq 0;\, \sigma_n \leq t\}$, we get by monotone interpolation

$$\lim_{t\to\infty} \kappa_t^{-1} \int_0^{t} f(X_s)\, ds = 2\,(b-a)\,\nu f \quad \text{a.s. } P_x, \quad x \in I. \tag{24}$$

This remains true when $\nu f = \infty$, since we can then apply (24) to some approximating functions $f_n \uparrow f$ with $\nu f_n < \infty$, and let $n \to \infty$. The assertion now follows as we apply (24) to both f and g. □

We may finally classify the asymptotic behavior of the process, depending on the boundedness of the speed measure ν and the nature of the endpoints. Applying a suitable affine mapping, we may transforms I^o into one of the intervals $(0,1)$, $(0,\infty)$, or $(-\infty,\infty)$. We need to distinguish between the three possibilities of boundary behavior at finite endpoints, codified as follows:

(— inaccessible boundary,
[— absorbing boundary,
[[— reflecting boundary.

Thus, we end up with totally ten different cases.

A diffusion is said to be ν*-ergodic* if it is recurrent and such that $\mathcal{L}_x(X_t) \xrightarrow{w} \nu/\nu I$ for all x. A recurrent diffusion may be either *null-recurrent* or *positive recurrent*, depending on whether $|X_t| \xrightarrow{P} \infty$ or not. We further say that *absorption* occurs at an endpoint b, when $X_t = b$ for all sufficiently large t, so that $1\{X_t = b\} \to 1$.

Theorem 33.15 *(recurrence and ergodicity, Feller, Maruyama & Tanaka) Let X be a regular diffusion on a natural scale with speed measure ν. Then the ergodic behavior of X depends as follows on the initial position x and the nature of the boundaries:*

- $(-\infty,\infty)$*: X is ν-ergodic when ν is bounded, otherwise null-recurrent,*

- $(0,\infty)$*:* $X_t \to 0$ *a.s.,*
- $[0,\infty)$*:* $1\{X_t = 0\} \to 1$ *a.s.,*
- $[[0,\infty)$*:* X *is* ν*-ergodic when* ν *is bounded, otherwise null-recurrent,*
- $(0,1)$*:* $X_t \to 0$ *or* 1*, with probabilities* $1-x$ *and* x*, respectively,*
- $[0,1)$*:* $1\{X_t = 0\} \to 1$ *or* $X_t \to 1$*, with probabilities* $1-x$ *and* x*,*
- $[0,1]$*:* $1\{X_t = 0\} \to 1$ *or* $1\{X_t = 1\} \to 1$*, with probabilities* $1-x$ *and* x*,*
- $[[0,1)$*:* $X_t \to 1$ *a.s.,*
- $[[0,1]$*:* $1\{X_t = 1\} \to 1$ *a.s.,*
- $[[0,1]]$*:* X *is* ν*-ergodic.*

First we prove the relatively elementary recurrence properties, distinguishing between the cases of absorption, convergence, and recurrence.

Proof of recurrence properties:

$[0,1]$: Theorem 33.7 (ii) yields $P_x\{\tau_0 < \infty\} = 1 - x$ and $P_x\{\tau_1 < \infty\} = x$.

$[0,\infty)$: The same theorem yields for any $b > x$

$$P_x\{\tau_0 < \infty\} \ge P_x\{\tau_0 < \tau_b\} = \frac{b-x}{b},$$

which tends to 1 as $b \to \infty$.

$(-\infty,\infty)$: The recurrence follows from the previous case.

$[[0,\infty)$: Since 0 is reflecting, $P_0\{\tau_y < \infty\} > 0$ for some $y > 0$, which extends to arbitrary y by the strong Markov property and the regularity of X. Arguing as in the proof of Lemma 33.8, we conclude that $P_0\{\tau_y < \infty\} = 1$ for all $y > 0$. The asserted recurrence now follows, as we combine with the statement for $[0,\infty)$.

$(0,\infty)$: Here $X = B \circ [X]$ a.s. for a Brownian motion B. Since $X > 0$, we have $[X]_\infty < \infty$ a.s., and so X converges a.s. Now $P_y\{\tau_{a,b} < \infty\} = 1$ for any $0 < a \le y \le b$. Applying the Markov property at an arbitrary time $t > 0$, we conclude that a.s. either $\liminf_t X_t \le a$ or $\limsup_t X_t \ge b$. Since a and b are arbitrary, X_∞ is then an endpoint of $(0,\infty)$, and hence equals 0.

$(0,1)$: As in the previous case, we have a.s. convergence to either 0 or 1. To find the corresponding probabilities, we see from Theorem 33.7 that

$$P_x\{\tau_a < \infty\} \ge P_x\{\tau_a < \tau_b\} = \frac{b-x}{b-a}, \quad 0 < a < x < b < 1.$$

Letting $b \to 1$ and then $a \to 0$, we obtain $P_x\{X_\infty = 0\} \ge 1 - x$. Similarly, $P_x\{X_\infty = 1\} \ge x$, and so equality holds in both relations.

$[0,1)$: Once again, $X \to 0$ or 1 with probabilities $1-x$ and x, respectively. Further note that

$$P_x\{\tau_0 < \infty\} \ge P_x\{\tau_0 < \tau_b\} = \frac{b-x}{b}, \quad 0 \le x < b < 1,$$

which tends to $1-x$ as $b \to 1$. Thus, absorption occurs when $X \to 0$.

$[[0,1]]$: As in the previous case, we get $P_0\{\tau_1 < \infty\} = 1$, and by symmetry also $P_1\{\tau_0 < \infty\} = 1$.

$[[0,1]$: As before $P_0\{\tau_1 < \infty\} = 1$, and so the same relation holds for P_x.

$[[0,1)$: Again $P_0\{\tau_b < \infty\} = 1$ for all $b \in (0,1)$. By the strong Markov property at τ_b and the result for $[0,1)$, it follows that $P_0\{X_t \to 1\} \geq b$. Letting $b \to 1$ gives $X_t \to 1$, a.s. under P_0. The result for P_x now follows by the strong Markov property at τ_x, applied under P_0. □

The ergodic properties will be proved along the lines of Theorem 11.22, which requires some additional lemmas.

Lemma 33.16 *(coupling) For any Feller processes $X \perp\!\!\!\perp Y$, the pair (X,Y) is again Feller.*

Proof: Use Theorem 5.30 and Lemma 17.3. □

We proceed with a continuous-time counterpart of Lemma 11.24.

Lemma 33.17 *(strong ergodicity) For a regular, recurrent diffusion with initial distribution μ_1 or μ_2, we have*

$$\lim_{t\to\infty} \left\| P_{\mu_1} \circ \theta_t^{-1} - P_{\mu_2} \circ \theta_t^{-1} \right\| = 0.$$

Proof: Let X and Y be independent with distributions P_{μ_1} and P_{μ_2}, respectively. By Theorem 33.13 and Lemma 33.16, the pair (X,Y) can be extended to a Feller diffusion, and so by Theorem 17.17 it is again strong Markov with respect to the induced filtration $\mathcal{G}$. Define $\tau = \inf\{t \geq 0;\, X_t = Y_t\}$, and note that τ is $\mathcal{G}$-optional by Lemma 9.6. The assertion now follows as for Lemma 11.24, provided we can show that $\tau < \infty$ a.s.

Here we first assume that $I = \mathsf{R}$. The processes X and Y are then continuous local martingales. By independence they remain local martingales for the extended filtration $\mathcal{G}$, and so even $X - Y$ is a local $\mathcal{G}$-martingale. Using the independence and recurrence of X and Y, we get $[X-Y]_\infty = [X]_\infty + [Y]_\infty = \infty$ a.s., which shows that even $X - Y$ is recurrent. In particular, $\tau < \infty$ a.s.

Next let $I = [[0,\infty)$ or $[[0,1]]$, and define $\tau_1 = \inf\{t \geq 0;\, X_t = 0\}$ and $\tau_2 = \inf\{t \geq 0;\, Y_t = 0\}$. By the continuity and recurrence of X and Y, we get $\tau \leq \tau_1 \vee \tau_2 < \infty$ a.s. □

Our next result is similar to the discrete-time version in Lemma 11.25.

Lemma 33.18 *(existence) Every regular, positive-recurrent diffusion has an invariant distribution.*

Proof: By Theorem 33.13 we may regard the transition kernels μ_t with associated operators T_t as defined on $\bar{I}$—the interval I with possible entrance boundaries attached. Since X is not null-recurrent, we may choose a bounded

Borel set B and some $x_0 \in I$ and $t_n \to \infty$, such that $\inf_n \mu_{t_n}(x_0, B) > 0$. Then Theorem 6.20 yields a measure μ on $\bar{I}$ with $\mu I > 0$, such that $\mu_{t_n}(x_0, \cdot) \xrightarrow{v} \mu$ along a sub-sequence, in the topology of $\bar{I}$. The convergence extends by Lemma 33.17 to arbitrary $x \in I$, and so

$$T_{t_n} f(x) \to \mu f, \quad f \in C_0(\bar{I}), \; x \in I. \tag{25}$$

Now fix any $h \geq 0$ and $f \in C_0(\bar{I})$, and note that even $T_h f \in C_0(\bar{I})$ by Theorem 33.13. Using (25), the semi-group property, and dominated convergence, we get for any $x \in I$

$$\begin{aligned} \mu(T_h f) &\leftarrow T_{t_n}\big(T_h f\big)(x) \\ &= T_h\big(T_{t_n} f\big)(x) \to \mu f. \end{aligned}$$

Thus, $\mu\mu_h = \mu$ for all h, which means that μ is invariant on $\bar{I}$. In particular, $\mu(\bar{I} \setminus I) = 0$ by the nature of entrance boundaries, and so $\mu/\mu I$ is an invariant distribution on I. □

Our final lemma provides the crucial connection between speed measure and invariant distributions.

Lemma 33.19 *(positive recurrence) For a regular, recurrent diffusion X on a natural scale in I with speed measure ν, these conditions are equivalent:*

(i) *$\nu I < \infty$,*

(ii) *X is positive-recurrent,*

(iii) *X has an invariant distribution.*

The invariant distribution is then unique and given by $\nu/\nu I$.

Proof: If the process is null-recurrent, then clearly no invariant distribution exists. The converse is also true by Lemma 33.18, and so (ii) ⇔ (iii). Now fix any bounded, measurable function $f : I \to \mathsf{R}_+$ with bounded support. By Theorem 33.14, Fubini's theorem, and dominated convergence, we have for any distribution μ on I

$$t^{-1} \int_0^t E_\mu f(X_s)\, ds = E_\mu\, t^{-1} \int_0^t f(X_s)\, ds \to \frac{\nu f}{\nu I}.$$

If μ is invariant, then $\mu f = \nu f/\nu I$, and so $\nu I < \infty$. If instead X is null-recurrent, then $E_\mu f(X_s) \to 0$ as $s \to \infty$, and we get $\nu f/\nu I = 0$, which implies $\nu I = \infty$. □

End of proof of Theorem 33.15: It remains to consider the cases where I equals (∞, ∞), $[[0, \infty)$, or $[[0, 1]]$, since otherwise we have convergence or absorption at some endpoint. In case of $[[0, 1]]$, Theorem 33.12 (ii) shows that ν is bounded. In the remaining cases, ν may be unbounded, so that X is null-recurrent by Lemma 33.19. If ν is bounded, then $\mu = \nu/\nu I$ is invariant by the same lemma, and the asserted ν-ergodicity follows from Lemma 33.17 with $\mu_1 = \mu$. □

Exercises

1. Prove pathwise uniqueness for the SDE $dX_t = (X_t^+)^{1/2}\,dB_t + c\,dt$ with $c > 0$. Also show that the solutions X^x with $X_0^x = x$ satisfy $X_t^x < X_t^y$ a.s. for $x < y$, up to the time when X^x reaches 0.

2. Show that solutions to equation $dX_t = \sigma(X_t)\,dB_t$ cannot explode. (*Hint:* If X explodes at time $\zeta < \infty$, then $[X]_\zeta = \infty$, and the local time of X tends to ∞ as $t \to \zeta$, uniformly on compacts. Now use Theorem 29.5 to see that $\zeta = \infty$ a.s.)

3. Assume in Theorem 33.1 that $S_\sigma = N_\sigma$. Show that the solutions X to (3) form a regular diffusion on a natural scale, on every connected component I of S_σ. Also note that the endpoints of I are either absorbing or exit boundaries for X. (*Hint:* Use Theorems 32.11, 29.4, and 29.5, and show that the exit time from any compact interval $J \subset I$ is finite.)

4. Assume in Theorem 33.1 that $S_\sigma \subset N_\sigma$, and form $\tilde\sigma$ from σ by taking $\tilde\sigma(x) = 1$ on $A = N_\sigma \setminus S_\sigma$. Show that solutions X to equation $(\tilde\sigma, 0)$ also solve equation $(\sigma, 0)$, but not conversely unless $A = \emptyset$. (*Hint:* Since $\lambda A = 0$, we have $\int 1_A(X_t)\,dt = \int 1_A(X_t)\,d[X]_t = 0$ a.s. by Theorem 29.5.)

5. Assume in Theorem 33.1 that $S_\sigma \subset N_\sigma$. Show that equation $(\sigma, 0)$ has solutions forming a regular diffusion on every connected component of S_σ^c. Prove the corresponding statement for the connected components of N_σ^c when N_σ is closed. (*Hint:* For S_σ^c, use the preceding result. For N_σ^c, choose X to be absorbed when it first reaches N_σ.)

6. For a regular diffusion in I, show that any two scale functions p_1, p_2 on I are related by an affine transformation.

7. For a regular diffusion X on a natural scale in I and with speed measure ν, define a new process $Y_t = X \circ A_t$, where $A_t = \int_0^t h(X_s)\,ds$ for a bounded, continuous function $h > 0$ on R_+. Show that Y is again a regular diffusion on a natural scale in I, and determine its speed measure.

8. In the setting of Theorem 33.14, show that the stated relation implies the convergence in Corollary 26.8 (i). Further use the result to prove a law of large numbers for regular, recurrent diffusions with bounded speed measure ν. (*Hint:* Note that $\nu g > 0$ implies $\int g(X_s)\,ds > 0$ a.s.)

9. Give examples of speed measures ν for diffusions on $[0, 1)$, $[[0, 1)$, and $(0, 1)$. In any of these cases, can we change the character of the boundary at 0 by changing ν? Will it make any difference if the right endpoint is ∞?

10. Identify the cases in Theorem 33.15 where X is positive-recurrent, null-recurrent, and transient. In the first of these cases, give the invariant distribution.

11. Let X be a Brownian motion in R^d absorbed at 0. Show that $Y = |X|^2$ is a regular diffusion in $(0, \infty)$, describe its boundary behavior for different d, and identify the corresponding case of Theorem 33.15. Verify the conclusion by computing the associated scale function and speed measure.

12. Explain how the ratio ergodic Theorem 33.14 is related to the ergodic properties in Theorems 26.4 and 26.21. Name a context where two or all three results apply.

Chapter 34

PDE Connections and Potential Theory

Cauchy problem, Feynman–Kac formula, PDE existence and SDE uniqueness, harmonic functions, regularity, Dirichlet's problem, transition and occupation densities, Greenian domain, Green function and potential, hitting and quitting kernels, sweeping measure and hitting, equilibrium measure and quitting, dependence on conductor and domain, time reversal, capacities and random sets, super-harmonic and excessive functions, additive functional as compensator, Riesz decomposition, measure-induced additive functional

In Chapters 17 and 32 we saw how elliptic differential operators arise naturally as the generators of nice diffusion processes. This is the ultimate cause of some profound connections between probability theory and partial differential equations (PDEs). In particular, a suitable extension of the operator $\frac{1}{2}\Delta$ appears as the generator of Brownian motion in R^d, which leads to a close relationship between classical potential theory and the theory of Brownian motion. More specifically, many basic problems in potential theory can be solved by probabilistic methods. Conversely, various hitting distributions for Brownian motion can be given a potential-theoretic interpretation.

This chapter explores some of the mentioned connections. First we derive the celebrated Feynman–Kac formula, and show how the *existence* of solutions to a given Cauchy problem implies *uniqueness* of solutions to the associated SDE. We then proceed with a probabilistic construction of Green functions and potentials, and solve the Dirichlet, sweeping, and equilibrium problems of classical potential theory in terms of Brownian motion. We further show how Green capacities and alternating set functions can be represented in a natural way in terms of random sets.

We conclude with a discussion of excessive and super-harmonic functions, along with their relations to modern martingale theory. To indicate the main ideas, let f be an excessive function of Brownian motion X on R^d. Then $f(X)$ is a continuous super-martingale under P_x for every x, and so it has a Doob–Meyer decomposition $M - A$. Here A can be chosen to be a continuous additive functional (CAF) of X, and we obtain an associated Riesz decomposition $f = U_A + h$, where U_A is the potential of A and h is the greatest harmonic minorant of f.

Some stochastic calculus from Chapters 18 and 32 is used at the beginning of the chapter, and we also rely on the theory of Feller processes from Chapter

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_35

17. As for Brownian motion, the present discussion is essentially self-contained, apart from some elementary facts from Chapters 14 and 19. Occasionally, we refer to Chapters 5 and 23 for some basic weak convergence theory. Finally, the results at the end of the chapter require the existence of Poisson processes from Proposition 15.6, as well as some basic facts about the Fell topology listed in Theorem A6.1. Potential-theoretic ideas are used in several other chapters, and some additional, essentially unrelated results appear in especially Chapters 26 and 29.

To begin with the general PDE connections, we consider an arbitrary Feller diffusion in R^d with associated semi-group operators T_t and generator $(A, \mathcal{D})$. Recall from Theorem 17.6 that, for any $f \in \mathcal{D}$, the function

$$\begin{aligned} u(t,x) &= T_t f(x) \\ &= E_x f(X_t), \quad t \geq 0, \ x \in \mathsf{R}^d, \end{aligned}$$

satisfies *Kolmogorov's backward equation* $\dot u = Au$, where $\dot u = \partial u/\partial t$. Thus, u provides a probabilistic solution to the *Cauchy problem*

$$\dot u = Au, \qquad u(0,x) = f(x). \tag{1}$$

Adding a *potential* term vu to (1), where $v \colon \mathsf{R}^d \to \mathsf{R}_+$, we obtain the more general problem

$$\dot u = Au - vu, \qquad u(0,x) = f(x). \tag{2}$$

Here the solution may be expressed in terms of the elementary *multiplicative functional* e^{-V}, where

$$V_t = \int_0^t v(X_s)\, ds, \quad t \geq 0.$$

Let $C^{1,2}$ be the class of functions $f \colon \mathsf{R}_+ \times \mathsf{R}^d$ of class C^1 in the time variable and class C^2 in the space variables. Write $C_b(\mathsf{R}^d)$ and $C_b^+(\mathsf{R}^d)$ for the classes of bounded, continuous functions from R^d to R and R_+, respectively.

Theorem 34.1 *(Cauchy problem, Feynman, Kac) Let $(A, \mathcal{D})$ be the generator of a Feller diffusion in R^d, and fix any $f \in C_b(\mathsf{R}^d)$ and $v \in C_b^+(\mathsf{R}^d)$. Then*

(i) *any bounded solution $u \in C^{1,2}$ to* (2) *is given by*

$$u(t,x) = E_x\, e^{-V_t} f(X_t), \quad t \geq 0, \ x \in \mathsf{R}^d.$$

(ii) *the function in* (i) *solves* (2) *whenever $f \in \mathcal{D}$.*

Formula (i) can be interpreted in terms of *killing*. Then consider an exponential random variable $\gamma \perp\!\!\!\perp X$ with mean 1, and define $\zeta = \inf\{t \geq 0;\, V_t > \gamma\}$. Letting $\tilde X$ be the process X killed at time ζ, we may express the right-hand side in (i) as $E_x f(\tilde X_t)$, with the understanding that $f(\tilde X_t) = 0$ when $t \geq \zeta$. Thus, $u(t,x) = \tilde T_t f(x)$, where $\tilde T_t$ is the transition operator of the killed process. It is easy to verify, directly from (i), that the family $(\tilde T_t)$ is again a Feller semigroup.

Proof of Theorem 34.1: Let $u \in C^{1,2}$ be a bounded solution to (2), fix any $t > 0$, and define

$$M_s = e^{-V_s}\, u\Big(t-s, X_s\Big), \quad s \in [0,t].$$

Writing $\overset{m}{=}$ for equality apart from a continuous local martingale or its differential, we see from Lemma 17.21, Itô's formula, and (2) that, for any $s < t$,

$$\begin{aligned} dM_s &= e^{-V_s}\Big\{du(t-s,X_s) - u(t-s,X_s)\,v(X_s)\,ds\Big\} \\ &\overset{m}{=} e^{-V_s}\{Au(t-s,X_s) - \dot u(t-s,X_s) - u(t-s,X_s)\,v(X_s)\}\,ds = 0. \end{aligned}$$

Thus, M is a continuous local martingale on $[0,t)$. Since M is bounded, the martingale property extends to t, and we get

$$\begin{aligned} u(t,x) &= E_x M_0 = E_x M_t \\ &= E_x u(0, X_t) \\ &= E_x\, e^{-V_t} f(X_t). \end{aligned}$$

Next let u be given by (i) for some $f \in \mathcal{D}$. Integrating by parts and using Lemma 17.21, we obtain

$$\begin{aligned} d\Big\{e^{-V_t} f(X_t)\Big\} &= e^{-V_t}\Big\{df(X_t) - (vf)(X_t)\,dt\Big\} \\ &\overset{m}{=} e^{-V_t}\Big(Af - vf\Big)(X_t)\,dt. \end{aligned}$$

Taking expectations and differentiating at $t = 0$, we conclude that the semigroup $\tilde T_t f(x) = E_x f(\tilde X_t) = u(t,x)$ has generator $\tilde A = A - v$ on $\mathcal{D}$. Now (2) follows by the last assertion in Theorem 17.6. □

Part (ii) of Theorem 34.1 can often be improved in special cases. In particular, if $v = 0$ and $A = \frac12\Delta = \frac12\sum_i \partial^2/\partial x_i^2$, so that X is a Brownian motion and (2) reduces to the standard *heat equation*, then $u(t,x) = E_x f(X_t)$ solves (2) for any bounded, continuous function f on R^d. To see this, we note that $u \in C^{1,2}$ on $(0,\infty)\times\mathsf{R}^d$, by the smoothness of the Brownian transition density. We may then obtain (2) by applying the backward equation to the function $T_h f(x)$ for a fixed $h \in (0,t)$.

Now consider the SDE in R^d

$$dX_t^i = \sigma_j^i(X_t)\,dB_t^j + b^i(X_t)\,dt, \tag{3}$$

with associated elliptic operator[1]

$$Av(x) = \tfrac12\, a^{ij}(x)\,\partial_{ij}^2 v(x) + b^i(x)\,\partial_i v(x), \quad x \in \mathsf{R}^d,\ v \in C^2,$$

where $a^{ij} = \sigma_k^i \sigma_k^j$. We show how *uniqueness* in law for solutions to (3) may be inferred from the *existence* of solutions to the associated Cauchy problem (1).

Theorem 34.2 *(existence and uniqueness, Stroock & Varadhan) For any* $f \in C_0^\infty(\mathsf{R}^d)$ *we have* (i) $\Rightarrow$ (ii)*, where*

[1]Summation over repeated indices is always understood.

(i) *equation* (1) *has a bounded solution on a set* $[0,\varepsilon]\times\mathsf{R}^d$,

(ii) *uniqueness in law holds for the SDE* (3).

Proof: Fix any $f\in C_0^\infty$ and $t\in(0,\varepsilon]$, and let u be a bounded solution to (1) on $[0,t]\times\mathsf{R}^d$. If X solves (3), we see as before that $M_s=u(t-s,X_s)$ is a martingale on $[0,t]$, and so

$$\begin{aligned} Ef(X_t) &= Eu(0,X_t)\\ &= EM_t = EM_0\\ &= Eu(t,X_0). \end{aligned}$$

Thus, the 1-dimensional distributions of X on $[0,\varepsilon]$ are uniquely determined by the initial distribution.

Now let X and Y be solutions with the same initial distribution. To prove that $X\overset{fd}{=}Y$, it is enough to consider times $0=t_0<t_1<\cdots<t_n$ with $t_k-t_{k-1}\le\varepsilon$ for all k. Assume that the distributions agree at $t_0,\dots,t_{n-1}=t$, and fix any set $C=\pi^{-1}_{t_0,\dots,t_{n-1}}B$ with $B\in\mathcal{B}^{nd}$. By Theorem 32.7, both $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ solve the local martingale problem for (a,b). If $P\{X\in C\}=P\{Y\in C\}>0$, Theorem 32.11 yields the same property for the conditional measures $\mathcal{L}(\theta_tX\,|\,X\in C)$ and $\mathcal{L}(\theta_tY\,|\,Y\in C)$. Since the corresponding initial distributions agree by hypothesis, the 1-dimensional result yields the extension

$$\mathcal{L}\Big(X_{t+h};\,X\in C\Big)=\mathcal{L}\Big(Y_{t+h};\,Y\in C\Big),\quad h\in(0,\varepsilon].$$

In particular, the distributions agree at times $t_0,\dots,t_n$. The general result now follows by induction. □

We now specialize to the case of a Brownian motion X in R^d. For any closed set $B\subset\mathsf{R}^d$, we introduce the *hitting time* $\tau_B=\inf\{t>0;X_t\in B\}$ and associated *hitting kernel*

$$H_B(x,dy)=P_x\Big\{\tau_B<\infty,\,X_{\tau_B}\in dy\Big\},\quad x\in\mathsf{R}^d.$$

For suitable functions f, we write $H_Bf(x)=\int f(y)\,H_B(x,dy)$.

By a *domain* in R^d we mean an open, connected subset $D\subset\mathsf{R}^d$. A function $u:D\to\mathsf{R}$ is said to be *harmonic*, if it belongs to $C^2(D)$ and satisfies the *Laplace equation* $\Delta u=0$. We also say that u has the *mean-value property*, if it is locally bounded and measurable, and such that for any ball $B\subset D$ with center x, the average of u over the boundary ∂B equals $u(x)$. The following analytic result is crucial for the probabilistic developments.

Lemma 34.3 *(harmonic function, Gauss, Koebe) For functions u on a domain $D\subset\mathsf{R}^d$, these conditions are equivalent and imply $u\in C_D^\infty$:*

(i) *u is harmonic,*

(ii) *u has the mean-value property.*

Proof: First let $u \in C^2(D)$, and fix a ball $B \subset D$ with center x. Writing $\tau = \tau_{\partial B}$ and noting that $E_x \tau < \infty$, we get by Itô's formula

$$E_x u(X_\tau) - u(x) = \tfrac{1}{2}\, E_x \int_0^\tau \Delta u(X_s)\, ds.$$

Here the first term on the left equals the average of u over ∂B, by the spherical symmetry of Brownian motion. If u is harmonic, then the right-hand side vanishes, and the mean-value property follows. If instead u is not harmonic, we may choose B such that $\Delta u \neq 0$ on B. Then the right-hand side is non-zero, and the mean-value property fails.

It remains to show that every function u with the mean-value property is infinitely differentiable. Then fix any infinitely differentiable, spherically symmetric probability density φ, supported by a ball of radius $\varepsilon > 0$ around the origin. Here the mean-value property yields $u = u * \varphi$, on the set where the right-hand side is defined, and by dominated convergence the infinite differentiability of φ carries over to $u * \varphi = u$. □

Before proceeding to the potential-theoretic developments, we need to introduce a regularity condition on the domain D. Writing $\zeta = \zeta_D = \tau_{D^c}$, we note that $P_x\{\zeta = 0\} = 0$ or 1 for every $x \in \partial D$ by Corollary 17.18. When this probability is 1, we say that x is *regular for* D^c or simply *regular*. If this holds for every $x \in \partial D$, the boundary ∂D is said to be regular, and we refer to D as a *regular domain*.

Regularity is a fairly weak condition. In particular, any domain with a smooth boundary is regular, and we shall see that even various edges and corners are allowed, provided they are not too sharp and directed inward. By a *spherical cone* in R^d with *vertex* v and *axis* $a \neq 0$ we mean a set of the form $C = \big\{x;\ \langle x - v, a\rangle \geq c|x - v|\big\}$, where $c \in (0, |a|]$.

Lemma 34.4 *(cone condition, Zaremba) For a domain $D \subset \mathsf{R}^d$, let $x \in \partial D$ be such that $C \cap G \subset D^c$ for a spherical cone C with vertex x and a neighborhood G of x. Then x is regular for D^c.*

Proof: By compactness of the unit sphere in R^d, we may cover R^d by $C_1 = C$ along with finitely many congruent cones $C_2, \dots, C_n$ with vertex x. By rotational symmetry,

$$\begin{aligned} 1 &= P_x\Big\{\min_{k \leq n} \tau_{C_k} = 0\Big\} \\ &\leq \sum_{k \leq n} P_x\Big\{\tau_{C_k} = 0\Big\} \\ &= n\, P_x\{\tau_C = 0\}, \end{aligned}$$

and so $P_x\{\tau_C = 0\} > 0$. Hence, Corollary 17.18 yields $P\{\tau_C = 0\} = 1$, and we get $\zeta_D \leq \tau_{C \cap G} = 0$ a.s. P_x. □

Now fix a domain $D \subset \mathsf{R}^d$ and a continuous function $f : \partial D \to \mathsf{R}$. A function u on $\bar{D}$ is said to solve the *Dirichlet problem* (D, f), if u is harmonic on D and continuous on $\bar{D}$ with $u = f$ on ∂D. The solution may be interpreted as the electrostatic potential in D, when the potential on the boundary is given by f.

Theorem 34.5 *(Dirichlet problem, Kakutani, Doob) For any regular domain $D \subset \mathsf{R}^d$ and function $f \in C_b(\partial D)$, we have*

(i) *the Dirichlet problem (D, f) has a solution*

$$u(x) = E_x\Big\{f(X_{\zeta_D});\, \zeta_D < \infty\Big\} = H_{D^c} f(x), \quad x \in \bar{D}, \tag{4}$$

(ii) *when $\zeta_D < \infty$ a.s., this is the only bounded solution,*

(iii) *when $d \geq 3$ and $f \in C_0(\partial D)$, it is the unique solution in $C_0(\bar{D})$.*

Thus, H_{D^c} agrees with the *sweeping (balayage) kernel* in Newtonian potential theory, determining the *harmonic measure* on ∂D. The following result clarifies the role of the regularity condition on ∂D.

Lemma 34.6 *(regularity, Doob) For any domain D and point $a \in \partial D$, these conditions are equivalent:*

(i) *a is regular for D^c,*

(ii) *for $f \in C_b(\partial D)$ and u as in (4), we have $u(x) \to f(a)$ as $x \to a$ in D.*

Proof: First let a be regular. For any $t > h > 0$ and $x \in D$, the Markov property yields

$$\begin{aligned} P_x\{\zeta > t\} &\leq P_x\Big\{\zeta \circ \theta_h > t - h\Big\} \\ &= E_x P_{X_h}\Big\{\zeta > t - h\Big\}. \end{aligned}$$

Here the right-hand side is continuous in x, by the continuity of the Gaussian kernel and dominated convergence, and so

$$\begin{aligned} \limsup_{x \to a} P_x\{\zeta > t\} &\leq E_a\, P_{X_h}\Big\{\zeta > t - h\Big\} \\ &= P_a\Big\{\zeta \circ \theta_h > t - h\Big\}. \end{aligned}$$

As $h \to 0$, the probability on the right tends to $P_a\{\zeta > t\} = 0$, and so $P_x\{\zeta > t\} \to 0$ as $x \to a$, which means that $\mathcal{L}_x(\zeta) \xrightarrow{w} \delta_0$. Since also $P_x \xrightarrow{w} P_a$ in $C_{\mathsf{R}_+,\mathsf{R}^d}$, Theorem 5.29 yields

$$\mathcal{L}_x(X, \zeta) \xrightarrow{w} \mathcal{L}_a(X, 0) \text{ in } C_{\mathsf{R}_+,\mathsf{R}^d} \times [0, \infty], \quad x \to a.$$

The continuity of the mapping $(x, t) \mapsto x_t$ implies $\mathcal{L}_x(X_\zeta) \xrightarrow{w} \mathcal{L}_a(X_0) = \delta_a$, and so $u(x) \to f(a)$, by the continuity of f.

Next assume (ii). If $d = 1$, then D is an interval, which is clearly regular. Next let $d \geq 2$. Then the Markov property yields for any $f \in C_b(\partial D)$

$$u(a) = E_a\Big\{f(X_\zeta);\, \zeta \leq h\Big\} + E_a\Big\{u(X_h);\, \zeta > h\Big\}, \quad h > 0.$$

As $h \to 0$, we get $u(a) = f(a)$ by dominated convergence, and for $f(x) = e^{-|x-a|}$ we obtain $P_a\{X_\zeta = a,\, \zeta < \infty\} = 1$. Since a.s. $X_t \neq a$ for all $t > 0$ by Theorem 19.6 (i), we obtain $P_a\{\zeta = 0\} = 1$, and so a is regular. □

Proof of Theorem 34.5: For u as in (4), fix any closed ball in D with center x and boundary S, and use the strong Markov property at $\tau = \tau_S$ to obtain

$$\begin{aligned} u(x) &= E_x\Big\{f(X_\zeta);\, \zeta < \infty\Big\} \\ &= E_x E_{X_\tau}\Big\{f(X_\zeta);\, \zeta < \infty\Big\} \\ &= E_x u(X_\tau). \end{aligned}$$

Hence, u has the mean-value property, and so it is harmonic by Lemma 34.3. Furthermore, Lemma 34.6 shows that u is continuous on $\bar{D}$ with $u = f$ on ∂D. Thus, u solves the Dirichlet problem (D, f).

Now let $d \geq 3$ and $f \in C_0(\partial D)$. For any $\varepsilon > 0$,

$$|u(x)| \leq \varepsilon + \|f\|\, P_x\Big\{\Big|f(X_\zeta)\Big| > \varepsilon,\, \zeta < \infty\Big\}. \tag{5}$$

Since X is transient by Theorem 19.6 (ii) and the set $\Big\{y \in \partial D;\, |f(y)| > \varepsilon\Big\}$ is bounded, the right-hand side of (5) tends to 0 as $|x| \to \infty$ and then $\varepsilon \to 0$, which shows that $u \in C_0(\bar{D})$.

To prove the asserted uniqueness, we may choose $f = 0$, and show that any solution u with the stated properties is identically 0. When $d \geq 3$ and $u \in C_0(\bar{D})$, this is clear by Lemma 34.3, which shows that harmonic functions have no local maxima or minima. Next let $\zeta < \infty$ a.s. and $u \in C_b(\bar{D})$. Then Corollary 18.19 yields $E_x u(X_{\zeta \wedge n}) = u(x)$ for any $x \in D$ and $n \in \mathsf{N}$, and so $u(x) = E_x u(X_\zeta) = 0$, by continuity and dominated convergence as $n \to \infty$. □

To prepare for the probabilistic construction of the Green function in an arbitrary domain $D \subset \mathsf{R}^d$, we need to study the transition densities of a Brownian motion, killed upon hitting the boundary ∂D. Recall that the unrestricted Brownian motion in R^d has transition densities

$$p_t(x, y) = (2\pi t)^{-d/2} e^{-|x-y|^2/2t}, \quad x, y \in \mathsf{R}^d,\ t > 0. \tag{6}$$

By the strong Markov property and Theorem 8.5, we get for any $t > 0$, $x \in D$, and $B \subset \mathcal{B}_D$

$$P_x\{X_t \in B\} = P_x\Big\{X_t \in B,\, t \leq \zeta\Big\} + E_x\Big\{T_{t-\zeta} 1_B(X_\zeta);\, t > \zeta\Big\}.$$

Thus, the killed process has transition densities

$$p_t^D(x, y) = p_t(x, y) - E_x\Big\{p_{t-\zeta}(X_\zeta, y);\, t > \zeta\Big\}, \quad x, y \in D,\ t > 0. \tag{7}$$

The following symmetry and continuity properties of p_t^D play a crucial role in the sequel.

Theorem 34.7 *(transition density, Hunt) For any domain D in R^d and time $t > 0$, we have*

(i) *p_t^D is symmetric and continuous on D^2,*

(ii) *for regular $a \in \partial D$,*

$$\lim_{x \to a} p_t^D(x, y) = 0, \quad y \in D.$$

Proof: (i) From (6) we note that $p_t(x, y)$ is uniformly continuous in (x, y) for fixed $t > 0$, as well as in (x, y, t) for $|x - y| > \varepsilon > 0$ and $t > 0$. By (7) it follows that $p_t^D(x, y)$ is equi-continuous in $y \in D$ for fixed $t > 0$. To prove the continuity in $x \in D$ for fixed $t > 0$ and $y \in D$, it is then enough to show that $P_x\{X_t \in B,\, t \le \zeta\}$ is continuous in x for fixed $t > 0$ and $B \in \mathcal{B}_D$. For any $h \in (0, t)$, the Markov property gives

$$P_x\Big\{X_t \in B,\, \zeta \ge t\Big\} = E_x\Big(P_{X_h}\{X_{t-h} \in B,\, \zeta \ge t - h\};\, \zeta > h\Big).$$

Thus, for any $x, y \in D$,

$$\begin{aligned} &\Big|(P_x - P_y)\{X_t \in B,\, t \le \zeta\}\Big| \\ &\qquad \le (P_x + P_y)\{\zeta \le h\} + \Big\|\mathcal{L}_x(X_h) - \mathcal{L}_y(X_h)\Big\|, \end{aligned}$$

which tends to 0 as $y \to x$ and then $h \to 0$. Combining the continuity in x with the equi-continuity in y, we conclude that $p_t^D(x, y)$ is continuous in $(x, y) \in D^2$ for fixed $t > 0$.

To prove the symmetry in x and y, it is now enough to establish the integrated version

$$\int_C P_x\{X_t \in B,\, \zeta > t\}\, dx = \int_B P_x\{X_t \in C,\, \zeta > t\}\, dx, \tag{8}$$

for bounded $B, C \in \mathcal{B}_D$. Then fix any compact set $F \subset D$. Letting $n \in \mathsf{N}$, and writing $h = 2^{-n}t$ and $t_k = kh$, we get by Proposition 11.2

$$\begin{aligned} &\int_C P_x\Big\{X_{t_k} \in F,\, k \le 2^n;\, X_t \in B\Big\}\, dx \\ &\qquad = \int_F \cdots \int_F 1_C(x_0)\, 1_B(x_{2^n}) \prod_{k \le 2^n} p_h(x_{k-1}, x_k)\, dx_0 \cdots dx_{2^n}. \end{aligned}$$

Here the right-hand side is symmetric in the pair (B, C), by the symmetry of $p_h(x, y)$. By dominated convergence as $n \to \infty$, we obtain (8) with F in place of D, and the stated version follows by monotone convergence as $F \uparrow D$.

(ii) From the proof of Lemma 34.6, recall that $\mathcal{L}_x(\zeta, X) \xrightarrow{w} \mathcal{L}_a(0, X)$ as $x \to a$ with a regular $a \in \partial D$. In particular, $\mathcal{L}_x(\zeta, X_\zeta) \xrightarrow{w} \delta_{0,a}$. By the boundedness and continuity of $p_t(x, y)$ for $|x - y| > \varepsilon > 0$, we see from (7) that $p_t^D(x, y) \to 0$. □

A domain $D \subset \mathsf{R}^d$ is said to be *Greenian*, if either $d \ge 3$, or if $d \le 2$ and $P_x\{\zeta_D < \infty\} = 1$ for all $x \in D$. Since the latter probability is harmonic in

x, it suffices by Lemma 34.3 to verify the stated property for a single $x \in D$. Given a Greenian domain D, we introduce the *Green function*

$$g^D(x,y) = \int_0^\infty p_t^D(x,y)\,dt, \quad x,y \in D.$$

For any measure μ on D, we further introduce the associated *Green potential*

$$G^D\mu(x) = \int g^D(x,y)\,\mu(dy), \quad x \in D.$$

Writing $G^D\mu = G^D f$ when $\mu(dy) = f(y)\,dy$, we get by Fubini's theorem

$$\begin{aligned} E_x \int_0^\zeta f(X_t)\,dt &= \int g^D(x,y)\,f(y)\,dy \\ &= G^D f(x), \quad x \in D, \end{aligned}$$

which identifies g^D as an *occupation density* of the killed process.

The next result shows that g^D and G^D agree with the Green function and Green potential of classical potential theory. Thus, $G^D\mu(x)$ may be interpreted as the electrostatic potential at x, arising from a charge distribution μ in D, when the boundary ∂D is grounded.

Theorem 34.8 *(Green function) For a Greenian domain $D \subset \mathsf{R}^d$,*

(i) *the function g^D is symmetric on D^2,*

(ii) *$g^D(x,y)$ is harmonic in $x \in D \setminus \{y\}$ for fixed $y \in D$,*

(iii) *for regular $b \in \partial D$,*

$$\lim_{x \to b} g^D(x,y) = 0, \quad y \in D.$$

The proof is straightforward when $d \geq 3$, but for $d \leq 2$ we need two technical lemmas. We begin with a uniform estimate for large t.

Lemma 34.9 *(uniform integrability) For a domain $D \subset \mathsf{R}^d$, bounded when $d \leq 2$, we have*

$$\lim_{t \to \infty} \sup_{x,y \in D} \int_t^\infty p_s^D(x,y)\,ds = 0.$$

Proof: For $d \geq 3$ we may take $D = \mathsf{R}^d$, in which case the result is obvious from (6). Now let $d = 2$. By obvious domination and scaling arguments, we may then assume that $|x| \leq 1$, $y = 0$, $D = \{z;\, |z| \leq 2\}$, and $t > 1$. Writing $p_t(x) = p_t(x,0)$, we get by (7)

$$\begin{aligned} p_t^D(x,0) &\leq p_t(x) - E_0\Big\{p_{t-\zeta}(1);\, \zeta \leq t/2\Big\} \\ &\leq p_t(0) - p_t(1)\,P_0\{\zeta \leq t/2\} \\ &\leq p_t(0)\,P_0\{\zeta > t/2\} + p_t(0) - p_t(1) \\ &\lesssim t^{-1}P_0\{\zeta > t/2\} + t^{-2}. \end{aligned}$$

As in Lemma 33.8 (ii), we have $E_0\zeta < \infty$, and so by Lemma 4.4 the right-hand side is integrable in $t \in [1,\infty)$. The proof for $d = 1$ is similar. □

We also need to show that bounded sets have bounded Green potential.

Lemma 34.10 *(boundedness) For a Greenian domain $D \subset \mathsf{R}^d$ and bounded set $B \in \mathcal{B}_D$, the function $G^D 1_B$ is bounded.*

Proof: By domination and scaling, along with the strong Markov property, we may take $B = \{x;\, |x| \le 1\}$, and show that $G^D 1_B(0) < \infty$. For $d \ge 3$ we may further choose $D = \mathsf{R}^d$, in which case the result follows by a simple computation. For $d = 2$, we may assume that $D \supset C \equiv \{x;\, |x| < 2\}$. Write $\sigma = \zeta_C + \tau_B \circ \theta_{\zeta_C}$ and $\tau_0 = 0$, and define recursively $\tau_{k+1} = \tau_k + \sigma \circ \theta_{\tau_k}$, $k \ge 0$. Putting $b = (1, 0)$ and using the strong Markov property at the times τ_k, we get

$$G^D 1_B(0) = G^C 1_B(0) + G^C 1_B(b) \sum_{k \ge 1} P_0\{\tau_k < \zeta\}.$$

Here $G^C 1_B(0) \vee G^C 1_B(b) < \infty$ by Lemma 34.9. Furthermore, the strong Markov property yields $P_0\{\tau_k < \zeta\} \le p^k$, where $p = \sup_{x \in B} P_x\{\sigma < \zeta\}$. Finally, we note that $p < 1$, since $P_x\{\sigma < \zeta\}$ is harmonic and hence continuous on B. The proof for $d = 1$ is similar. □

Proof of Theorem 34.8: (i) This is clear by Theorem 34.7.

(ii) If $d \ge 3$, or if $d = 2$ and D is bounded, we see from Theorem 34.7, Lemma 34.9, and dominated convergence that $g^D(x, y)$ is continuous in $x \in D \setminus \{y\}$ for fixed $y \in D$. Moreover, $G^D 1_B$ has the mean-value property in $D \setminus \bar{B}$ for bounded $B \in \mathcal{B}_D$. The latter property extends by continuity to the density $g^D(x, y)$, which is then harmonic in $x \in D \setminus \{y\}$ for fixed $y \in D$, by Lemma 34.3.

For $d = 2$ and unbounded D, define $D_n = \{x \in D;\, |x| < n\}$, and note as before that $g^{D_n}(x, y)$ has the mean-value property in $x \in D_n \setminus \{y\}$ for fixed $y \in D_n$. Since $p_t^{D_n} \uparrow p_t^D$ by dominated convergence, we have $g^{D_n} \uparrow g^D$, and so the mean-value property extends to the limit. For any $x \ne y$ in D, choose a circular disk B around y with radius $\varepsilon > 0$ small enough that $x \notin \bar{B} \subset D$. Then $\pi\, \varepsilon^2 g^D(x, y) = G^D 1_B(x) < \infty$ by Lemma 34.10. Thus, Lemma 34.3 shows that even $g^D(x, y)$ is harmonic in $x \in D \setminus \{y\}$.

(iii) Fix any $y \in D$, and assume that $x \to b \in \partial D$. Choose a Greenian domain $D' \supset D$ with $b \in D'$. Since $p_t^D \le p_t^{D'}$, and both $p_t^{D'}(\cdot, y)$ and $g^{D'}(\cdot, y)$ are continuous at b, whereas $p_t^D(x, y) \to 0$ by Theorem 34.7, we get $g^D(x, y) \to 0$ by Theorem 1.23. □

We proceed to show that a measure is determined by its Green potential, whenever the latter is finite. An extension appears as part of Theorem 34.12. For convenience, we write

$$P_t^D \mu(x) = \int p_t^D(x, y)\, \mu(dy), \quad x \in D,\ t > 0.$$

Theorem 34.11 *(uniqueness) For measures μ, ν on a Greenian domain $D \subset \mathsf{R}^d$, we have*

$$G^D \mu = G^D \nu < \infty \quad \Rightarrow \quad \mu = \nu.$$

Proof: For any $t>0$, we have

$$\begin{aligned}\int_0^t (P_s^D\mu)\,ds &= G^D\mu - P_t^D G^D\mu\\ &= G^D\nu - P_t^D G^D\nu\\ &= \int_0^t (P_s^D\nu)\,ds. \end{aligned}\tag{9}$$

By the symmetry of p^D, we further get for any measurable function $f\colon D\to \mathsf{R}_+$

$$\begin{aligned}\int f(x)\,P_s^D\mu(x)\,dx &= \int f(x)\,dx\int p_s^D(x,y)\,\mu(dy)\\ &= \int \mu(dy)\int f(x)\,p_s^D(x,y)\,dx\\ &= \int P_s^D f(y)\,\mu(dy).\end{aligned}$$

Hence,

$$\begin{aligned}\int f(x)\,dx\int_0^t P_s^D\mu(x)\,ds &= \int_0^t ds\int P_s^D f(y)\,\mu(dy)\\ &= \int \mu(dy)\int_0^t P_s^D f(y)\,ds,\end{aligned}$$

and similarly for ν. By (9), we obtain

$$\int \mu(dy)\int_0^t P_s^D f(y)\,ds = \int \nu(dy)\int_0^t P_s^D f(y)\,ds. \tag{10}$$

Assuming $f\in C_K^+(D)$, we get $P_s^D f\to f$ as $s\to 0$, and so $t^{-1}\int_0^t P_s^D f\,ds\to f$. If we can take limits in the outer integrations of (10), we obtain $\mu f=\nu f$, which implies $\mu=\nu$ since f is arbitrary.

To justify the argument, it is enough to show that $\sup_s P_s^D f$ is μ- and ν-integrable. Then conclude from Theorem 34.7 that $f\lesssim p_s^D(\cdot,y)$ for fixed $s>0$ and $y\in D$, and from Theorem 34.8 that $f\lesssim G^D f$. Here the latter property yields $P_s^D f\lesssim P_s^D G^D f\le G^D f$, whereas the former property yields for any $y\in D$ and $s>0$

$$\begin{aligned}\mu(G^D f) &= \int G^D\mu(x) f(x)\,dx\\ &\lesssim P_s^D G^D\mu(y)\\ &\le G^D\mu(y)<\infty,\end{aligned}$$

and similarly for ν. □

Now let $\mathcal{F}_D$ and $\mathcal{K}_D$ be the classes of closed and compact subsets of D, and write $\mathcal{F}_D^r$ and $\mathcal{K}_D^r$ for the subclasses of sets with regular boundary. For any $B\in\mathcal{F}_D$, we introduce the associated *hitting kernel*

$$H_B^D(x,dy) = P_x\Big\{\tau_B<\zeta_D,\ X_{\tau_B}\in dy\Big\},\quad x\in D.$$

Note that when X has initial distribution μ, the hitting distribution of X^ζ in B equals $\mu H_B^D=\int\mu(dx)\,H_B^D(x,\cdot)$.

The next result solves the *sweeping problem* of classical potential theory. To avoid some technical distractions, here and below, we consider only subsets with regular boundary. In general, the irregular part of the boundary can be shown to be *polar*, in the sense of being a.s. avoided by a Brownian motion. Given this result, one can easily remove all regularity restrictions.

Theorem 34.12 *(sweeping and hitting) For a Greenian domain $D \subset \mathsf{R}^d$ with subset $B \in \mathcal{F}_D^r$, let μ be a bounded measure on D with $G^D\mu < \infty$ on B. Then μH_B^D is the unique measure ν on B with*

$$G^D\mu = G^D\nu \text{ on } B.$$

For an electrostatic interpretation, suppose we insert a grounded conductor B into a domain D with grounded boundary and charge distribution μ. Then a charge distribution $-\mu H_B^D$ arises on B.

A lemma is needed for the proof. Here we define $g^{D\setminus B}(x,y) = 0$ when either x or y lies in B.

Lemma 34.13 *(fundamental identity) For a Greenian domain $D \subset \mathsf{R}^d$ with subset $B \in \mathcal{F}_D^r$,*

$$g^D(x,y) = g^{D\setminus B}(x,y) + \int_B H_B^D(x,dz)\, g^D(z,y), \quad x,y \in D.$$

Proof: Write $\zeta = \zeta_D$ and $\tau = \tau_B$. Subtracting relations (7) for the domains D and $D\setminus B$, and using the strong Markov property at τ together with Theorem 8.5, we get

$$\begin{aligned}
&p_t^D(x,y) - p_t^{D\setminus B}(x,y)\\
&\quad = E_x\Big\{p_{t-\tau}(X_\tau,y);\, \tau < \zeta \wedge t\Big\} - E_x\Big\{p_{t-\zeta}(X_\zeta,y);\, \tau < \zeta < t\Big\}\\
&\quad = E_x\Big\{p_{t-\tau}(X_\tau,y);\, \tau < \zeta \wedge t\Big\}\\
&\qquad\qquad - E_x\Big(E_{X_\tau}\Big\{p_{t-\tau-\zeta}(X_\zeta,y);\, \zeta < t-\tau\Big\};\, \tau < \zeta \wedge t\Big)\\
&\quad = E_x\Big\{p_{t-\tau}^D(X_\tau,y);\, \tau < \zeta \wedge t\Big\}.
\end{aligned}$$

Integrating with respect to t yields

$$\begin{aligned}
g^D(x,y) - g^{D\setminus B}(x,y) &= E_x\Big\{g^D(X_\tau,y);\, \tau < \zeta\Big\}\\
&= \int H_B^D(x,dz)\, g^D(z,y).
\end{aligned}$$

□

Proof of Theorem 34.12: Since ∂B is regular, we have $H_B^D(x,\cdot) = \delta_x$ for all $x \in B$, and so by Lemma 34.13 we get for all $x \in B$ and $z \in D$

$$\begin{aligned}
\int g^D(x,y)\, H_B^D(z,dy) &= \int g^D(z,y)\, H_B^D(x,dy)\\
&= g^D(z,x).
\end{aligned}$$

Integrating with respect to $\mu(dz)$ gives $G^D(\mu H_B^D)(x) = G^D\mu(x)$, which shows that $\nu = \mu H_B^D$ has the stated property.

Now consider any measure ν on B with $G^D\mu = G^D\nu$ on B. Noting that $g^{D\setminus B}(x,\cdot) = 0$ on B, whereas $H_B^D(x,\cdot)$ is supported by B, we get by Lemma 34.13 for any $x \in D$

$$\begin{aligned} G^D\nu(x) &= \int \nu(dz)\, g^D(z,x) \\ &= \int \nu(dz) \int g^D(z,y)\, H_B^D(x,dy) \\ &= \int H_B^D(x,dy)\, G^D\nu(y) \\ &= \int H_B^D(x,dy)\, G^D\mu(y). \end{aligned}$$

Thus, μ determines $G^D\nu$ on D, and so ν is unique by Theorem 34.11. □

Turning to the classical *equilibrium problem*, we introduce for any $K \in \mathcal{K}_D$ the *last exit* or *quitting time*

$$\gamma_K^D = \sup\Big\{t < \zeta_D;\ X_t \in K\Big\},$$

along with the associated *quitting kernel*

$$L_K^D(x,dy) = P_x\Big\{\gamma_K^D > 0;\ X(\gamma_K^D) \in dy\Big\}.$$

Theorem 34.14 *(equilibrium measure and quitting, Chung) For a Greenian domain $D \in \mathsf{R}^d$ with subset $K \in \mathcal{K}_D$,*

(i) *there exists a measure μ_K^D on ∂K with*

$$L_K^D(x,dy) = g^D(x,y)\,\mu_K^D(dy), \quad x \in D,$$

(ii) *μ_K^D is diffuse when $d \geq 2$,*

(iii) *when $K \in \mathcal{K}_D^r$, μ_K^D is the unique measure μ on K with $G^D\mu = 1$ on K.*

Here μ_K^D is called the *equilibrium measure* of K relative to D, and its total mass C_K^D is called the *capacity* of K in D. For an electrostatic interpretation, suppose we insert a conductor K with potential 1 into a domain D with grounded boundary. Then a charge distribution μ_K^D arises on the boundary of K.

Proof: Write $\gamma = \gamma_K^D$, and define

$$l_\varepsilon(x) = \varepsilon^{-1} P_x\Big\{0 < \gamma \leq \varepsilon\Big\}, \quad \varepsilon > 0.$$

Using Fubini's theorem, the simple Markov property, and dominated convergence as $\varepsilon \to 0$, we get for any $f \in C_b(D)$ and $x \in D$

$$\begin{aligned}
G^D(fl_\varepsilon)(x) &= E_x \int_0^\zeta f(X_t)\, l_\varepsilon(X_t)\, dt \\
&= \varepsilon^{-1} \int_0^\infty E_x\Big\{ f(X_t)\, P_{X_t}\{0 < \gamma \le \varepsilon\};\ t < \zeta \Big\}\, dt \\
&= \varepsilon^{-1} \int_0^\infty E_x\Big\{ f(X_t);\ t < \gamma \le t + \varepsilon \Big\}\, dt \\
&= \varepsilon^{-1} E_x \int_{(\gamma-\varepsilon)_+}^\gamma f(X_t)\, dt \\
&\to E_x\Big\{ f(X_\gamma);\ \gamma > 0 \Big\} \\
&= L_K^D f(x).
\end{aligned}$$

If f has compact support, then for every x we may replace f by the bounded, continuous function $f/g^D(x,\cdot)$, to get as $\varepsilon \to 0$

$$\int f(y)\, l_\varepsilon(y)\, dy \to \int \frac{L_K^D(x,dy)\, f(y)}{g^D(x,y)}. \tag{11}$$

Since the left-hand side is independent of x, the same thing is true for the measure

$$\mu_K^D(dy) = \frac{L_K^D(x,dy)}{g^D(x,y)}. \tag{12}$$

When $d = 1$ we have $g^D(x,x) < \infty$, and (12) is trivially equivalent to (i). If instead $d \ge 2$, then singletons are polar, and so the measure $L_K^D(x,\cdot)$ is diffuse, which implies the same property for μ_K^D. Thus, (12) is again equivalent to (i). Furthermore, the continuity of X implies that $L_K^D(x,\cdot)$, and then also μ_K^D is supported by ∂K.

Integrating equation (i) over D yields

$$P_x\Big\{\tau_K < \zeta_D\Big\} = G^D \mu_K^D(x), \quad x \in D,$$

and so for $K \in \mathcal{K}_D^r$ we get $G^D \mu_K^D = 1$ on K. If ν is another measure on K with $G^D \nu = 1$ on K, then $\nu = \mu_K^D$ by the uniqueness in Theorem 34.12. □

Now we will see how the equilibrium measures and capacities depend on the sets $K \in \mathcal{K}_D^r$.

Proposition 34.15 *(consistency) For a Greenian domain $D \subset \mathsf{R}^d$ with subsets $K \subset B$ in $\mathcal{K}_D^r$,*

(i) $\mu_K^D = \mu_B^D H_K^D = \mu_B^D L_K^D$,

(ii) $C_K^D = \int_B P_x\Big\{\tau_K < \zeta_D\Big\}\, \mu_B^D(dx)$.

Proof: By Theorem 34.12 and the defining properties of μ_B^D and μ_K^D, we have on K

$$\begin{aligned}
G^D\Big(\mu_B^D H_K^D\Big) &= G^D \mu_B^D \\
&= 1 = G^D \mu_K^D,
\end{aligned}$$

and so $\mu_B^D H_K^D = \mu_K^D$ by the same result. To prove the second equality in (i), we see from Theorem 34.14 that, for any $A \in \mathcal{B}_K$,

$$\begin{aligned} \mu_B^D L_K^D(A) &= \int \mu_B^D(dx) \int_A g^D(x,y)\, \mu_K^D(dy) \\ &= \int_A G^D \mu_B^D(y)\, \mu_K^D(dy) = \mu_K^D(A), \end{aligned}$$

since $G^D \mu_B^D = 1$ on $A \subset B$. Finally (i) $\Rightarrow$ (ii), since $H_K^D(x, K) = P_x\{\tau_K < \zeta_D\}$. □

Some basic properties of capacities and equilibrium measures follow immediately from Proposition 34.15. To explain the terminology, fix a space S, along with a class $\mathcal{U}$ of subsets, closed under finite unions. For any function $h \colon \mathcal{U} \to \mathsf{R}$ and sets $U, U_1, U_2, \ldots \in \mathcal{U}$, we define recursively the differences

$$\begin{aligned} \Delta_{U_1} h(U) &= h(U \cup U_1) - h(U), \\ \Delta_{U_1,\ldots,U_n} h(U) &= \Delta_{U_n}\Big\{\Delta_{U_1,\ldots,U_{n-1}} h(U)\Big\}, \quad n > 1, \end{aligned}$$

where the difference Δ_{U_n} in the last formula is taken with respect to U. Note that the higher-order differences $\Delta_{U_1,\ldots,U_n}$ are invariant under permutations of $U_1, \ldots, U_n$. Say that h is *alternating* or *completely monotone* if

$$(-1)^{n+1} \Delta_{U_1,\ldots,U_n} h(U) \geq 0, \quad n \in \mathsf{N}, \ \ U, U_1, U_2, \ldots \in \mathcal{U}.$$

Corollary 34.16 *(dependence on conductor, Choquet) For a Greenian domain $D \subset \mathsf{R}^d$,*

(i) *the capacity C_K^D is an alternating function of $K \in \mathcal{K}_D^r$,*

(ii) *$\mu_{K_n}^D \xrightarrow{w} \mu_K^D$ as $K_n \downarrow K$ or $K_n \uparrow K$ in $\mathcal{K}_D^r$.*

Proof: (i) Let ψ denote the path of X^ζ, regarded as a random closed set in D. Writing

$$\begin{aligned} h_x(K) &= P_x\Big\{\psi K \neq \emptyset\Big\} \\ &= P_x\{\tau_K < \zeta\}, \quad x \in D \setminus K, \end{aligned}$$

we get by induction

$$\begin{aligned} &(-1)^{n+1} \Delta_{K_1,\ldots,K_n} h_x(K) \\ &\qquad = P_x\Big\{\psi K = \emptyset,\, \psi K_1 \neq \emptyset,\, \ldots,\, \psi K_n \neq \emptyset\Big\} \geq 0, \end{aligned}$$

and the assertion follows by Proposition 34.15 with $K \subset B^o$.

(ii) Trivially, $\tau_{K_n} \downarrow \tau_K$ when $K_n \uparrow K$, and $\tau_{K_n} \uparrow \tau_K$ when $K_n \downarrow K$ since the K_n are closed. In the latter case, we have also $\bigcap_n \{\tau_{K_n} < \zeta\} = \{\tau_K < \zeta\}$ by compactness. Thus, in both cases $H_{K_n}^D(x, \cdot) \xrightarrow{w} H_K^D(x, \cdot)$ for all $x \in D \setminus \bigcup_n K_n$, and by dominated convergence in Proposition 34.15 with $B^o \supset \bigcup_n K_n$, we get $\mu_{K_n}^D \xrightarrow{w} \mu_K^D$. □

The next result solves the equilibrium problem for two conductors.

Corollary 34.17 *(condenser theorem)* *For any disjoint sets $B \in \mathcal{F}_D^r$ and $K \in \mathcal{K}_D^r$, there exists a unique signed measure ν on $B \cup K$ with $G^D\nu = 0$ on B and $G^D\nu = 1$ on K, given by*

$$\nu = \mu_K^{D\setminus B} - \mu_K^{D\setminus B} H_B^D.$$

Proof: Applying Theorem 34.14 to the domain $D \setminus B$ with subset K, we get $\nu = \mu_K^{D\setminus B}$ on K, and then $\nu = -\mu_K^{D\setminus B} H_B^D$ on B by Theorem 34.12. □

The symmetry between hitting and quitting kernels in Proposition 34.15 can be extended to an invariance under *time reversal* of the whole process. More precisely, putting $\gamma = \gamma_K^D$, we may relate the stopped process $X_t^\zeta = X_{\gamma\wedge t}$ to its reversal $\tilde{X}_t^\gamma = X_{(\gamma-t)+}$. For convenience, write $P_\mu = \int P_x\,\mu(dx)$, and refer to the induced measures as *distributions*, even when μ is not normalized.

Theorem 34.18 *(time reversal)* *For a Greenian domain $D \in \mathsf{R}^d$ with subset $K \in \mathcal{K}_D^r$, put $\gamma = \gamma_K^D$ and $\mu = \mu_K^D$. Then $X^\gamma \overset{d}{=} \tilde{X}^\gamma$ under P_μ.*

Proof: Let P_x and E_x refer to the process X^ζ. Fix any times $0 = t_0 < t_1 < \cdots < t_n$, and write $s_k = t_n - t_k$ and $h_k = t_k - t_{k-1}$. For any continuous functions $f_0, \ldots, f_n$ with compact supports in D, define

$$\begin{aligned} f^\varepsilon(x) &= E_x \prod_{k\geq 0} f_k(X_{s_k})\, l_\varepsilon(X_{t_n}) \\ &= E_x \prod_{k\geq 1} f_k(X_{s_k})\, E_{X_{s_1}}(f_0\, l_\varepsilon)(X_{t_1}), \end{aligned}$$

where the last equality holds by the Markov property at s_1. Proceeding as in the proof of Theorem 34.14, we get

$$\begin{aligned} \int \Big(f^\varepsilon G^D\mu\Big)(x)\, dx &= \int G^D f^\varepsilon(y)\, \mu(dy) \\ &\to E_\mu \textstyle\prod_k f_k(\tilde{X}_{t_k}^\gamma)\, 1\{\gamma > t_n\}. \end{aligned} \tag{13}$$

On the other hand, (11) shows that the measure $l_\varepsilon(x)\,dx$ tends vaguely to μ, and so by Theorem 34.7

$$\begin{aligned} E_x(f_0\, l_\varepsilon)(X_{t_1}) &= \int p_{t_1}^D(x,y)\,(f_0\, l_\varepsilon)(y)\, dy \\ &\to \int p_{t_1}^D(x,y)\, f_0(y)\, \mu(dy). \end{aligned}$$

Using dominated convergence, Fubini's theorem, Proposition 11.2, Theorem 34.7, and the relation $G^D\mu(x) = P_x\{\gamma > 0\}$, we obtain

$$\begin{aligned} &\int \Big(f^\varepsilon G^D\mu\Big)(x)\, dx \\ &\quad\to \int G^D\mu(x)\, dx \int f_0(y)\, \mu(dy)\, E_x \prod_{k>0} f_k(X_{s_k})\, p_{t_1}^D(X_{s_1}, y) \\ &\quad= \int f_0(x_0)\, \mu(dx_0) \int\cdots\int G^D\mu(x_n) \prod_{k>0} p_{h_k}^D(x_{k-1}, x_k) f_k(x_k)\, dx_k \\ &\quad= E_\mu \textstyle\prod_k f_k(X_{t_k})\, G^D\mu(X_{t_n}) \\ &\quad= E_\mu \textstyle\prod_k f_k(X_{t_k})\, 1\{\gamma > t_n\}. \end{aligned}$$

Comparing with (13), we see that X^γ and $\tilde{X}^\gamma$ have the same finite-dimensional distributions. □

We now extend Proposition 34.15 to exhibit the dependence on Greenian domains $D \subset D'$ as well. Then for fixed $K \in \mathcal{K}_D$, we define recursively the optional times

$$\begin{aligned} \tau_j &= \gamma_{j-1} + \tau_K^{D'} \circ \theta_{\gamma_{j-1}}, \\ \gamma_j &= \tau_j + \gamma_K^D \circ \theta_{\tau_j}, \qquad j \geq 1, \end{aligned}$$

starting with $\gamma_0 = 0$. Thus, τ_k and γ_k are the hitting or quitting times of K, during the k-th excursion in D reaching K, prior to the exit time $\zeta_{D'}$. The *extended hitting and quitting kernels* are given by

$$\begin{aligned} H_K^{D,D'}(x,\cdot) &= E_x \textstyle\sum_k \delta_{X(\tau_k)}, \\ L_K^{D,D'}(x,\cdot) &= E_x \textstyle\sum_k \delta_{X(\gamma_k)}, \end{aligned}$$

where the summations extend over all $k \in \mathsf{N}$ with $\tau_k < \infty$.

Theorem 34.19 *(dependence on conductor and domain) For a Greenian domains $D \subset D'$ in R^d with regular, compact subsets $K \subset K'$,*

$$\mu_K^D = \mu_{K'}^{D'} H_K^{D,D'} = \mu_{K'}^{D'} L_K^{D,D'}.$$

Proof: Define

$$l_\varepsilon = \varepsilon^{-1} P_x\{\gamma_K^D \in (0,\varepsilon]\}, \quad \varepsilon > 0.$$

Proceeding as in the proof of Theorem 34.14, we get for any $x \in D'$ and $f \in C_b(D')$

$$\begin{aligned} G^{D'}(f\, l_\varepsilon)(x) &= \varepsilon^{-1} E_x \int_0^{\zeta_{D'}} f(X_t)\, 1\Big\{\gamma_K^D \circ \theta_t \in (0,\varepsilon]\Big\}\, dt \\ &\to L_K^{D,D'} f(x). \end{aligned}$$

If f has compact support in D, we may conclude as before that

$$\int f(y)\, \mu_K^D(dy) \leftarrow \int (f\, l_\varepsilon)(y)\, dy \to \int \frac{L_K^{D,D'}(x,dy)\, f(y)}{g^{D'}(x,y)},$$

and so

$$L_K^{D,D'}(x,dy) = g^{D'}(x,y)\, \mu_K^D(dy).$$

Integrating with respect to $\mu_{K'}^{D'}$ and noting that $G^{D'}\mu_{K'}^{D'} = 1$ on $K' \supset K$, we obtain the second expression for μ_K^D.

To deduce the first expression for μ_K^D, note that $H_K^{D'} H_K^{D,D'} = H_K^{D,D'}$, by the strong Markov property at τ_K. Combining with the second expression for μ_K^D and using Theorem 34.18 and Proposition 34.15, we get

$$\begin{aligned} \mu_K^D &= \mu_K^{D'} L_K^{D,D'} = \mu_K^{D'} H_K^{D,D'} \\ &= \mu_{K'}^{D'} H_K^{D'} H_K^{D,D'} \\ &= \mu_{K'}^{D'} H_K^{D,D'}. \end{aligned}$$

□

The last result enables us to study the equilibrium measure μ_K^D and capacity C_K^D as functions of both D and K. In particular, we obtain the following continuity and monotonicity properties.

Corollary 34.20 *(dependence on domain) For a fixed regular, compact set $K \subset \mathsf{R}^d$, the equilibrium measure μ_K^D is a decreasing, upper continuous function of the Greenian domain $D \supset K$.*

Proof: The monotonicity is clear from Theorem 34.19 with $K = K'$, since

$$H_K^{D,D'}(x,\cdot) \geq \delta_x, \quad x \in K \subset D \subset D'.$$

It remains to prove that C_K^D is continuous from above and below in D for fixed K. By dominated convergence, it is then enough to show that $\kappa_K^{D_n} \to \kappa_K^D$, where $\kappa_K^D = \sup\{j;\, \tau_j < \infty\}$ is the number of D-excursions hitting K.

Assuming $D_n \uparrow D$, we need to show that if $X_s, X_t \in K$ and $X \in D$ on $[s,t]$, then $X \in D_n$ on $[s,t]$ for sufficiently large n. But this is clear, since the path is compact on the interval $[s,t]$. If instead $D_n \downarrow D$, we need to show that, for any $r < s < t$ with $X_r, X_t \in K$ and $X_s \notin D$, we have $X_s \notin D_n$ for sufficiently large n. But this is obvious. □

Next we show how Green capacities can be expressed in terms of random sets. Let χ denote the identity mapping on $\mathcal{F}_D$. Given any measure ν on $\mathcal{F}_D \setminus \{\emptyset\}$ with $\nu\{\chi K \neq \emptyset\} < \infty$ for all $K \in \mathcal{K}_D$, we may introduce a Poisson process η on $\mathcal{F}_D \setminus \{\emptyset\}$ with intensity measure ν, and form the associated random closed set $\varphi = \bigcup\{F;\, \eta\{F\} > 0\}$ in D. Letting π_ν be the distribution of φ, we note that

$$\begin{aligned} \pi_\nu\{\chi K = \emptyset\} &= P\Big\{\eta\{\chi K \neq \emptyset\} = 0\Big\} \\ &= \exp\Big(-\nu\{\chi K \neq \emptyset\}\Big), \quad K \in \mathcal{K}_D. \end{aligned}$$

Theorem 34.21 *(Green capacities and random sets, Choquet) For any Greenian domain $D \subset \mathsf{R}^d$, there exists a unique measure ν on $\mathcal{F}_D \setminus \{\emptyset\}$ such that*

$$\begin{aligned} C_K^D &= \nu\{\chi K \neq \emptyset\} \\ &= -\log \pi_\nu\{\chi K = \emptyset\}, \quad K \subset \mathcal{K}_D^r. \end{aligned}$$

Proof: Let ψ denote the path of X^ζ in D. Choose sets $K_n \uparrow D$ in $\mathcal{K}_D^r$ with $K_n \subset K_{n+1}^o$ for all n, and put $\mu_n = \mu_{K_n}^D$, $\psi_n = \psi K_n$, and $\chi_n = \chi K_n$. Define

$$\nu_n^p = \int P_x\Big\{\psi_p \in \cdot\,,\, \psi_n \neq \emptyset\Big\}\, \mu_p(dx), \quad n \leq p, \tag{14}$$

and conclude by the strong Markov property and Proposition 34.15 that

$$\nu_n^q\Big\{\chi_p \in \cdot\,,\, \chi_m \neq \emptyset\Big\} = \nu_m^p, \quad m \leq n \leq p \leq q. \tag{15}$$

By Corollary 8.22, there exist some measures ν_n on $\mathcal{F}_D$, $n \in \mathsf{N}$, with

$$\nu_n\{\chi_p \in \cdot\} = \nu_n^p, \quad n \le p, \tag{16}$$

and from (15) we note that

$$\nu_n\Big\{\cdot,\, \chi_m \ne \emptyset\Big\} = \nu_m, \quad m \le n. \tag{17}$$

Hence, the measures ν_n agree on $\{\chi_m \ne \emptyset\}$ for $n \ge m$, and so we may define $\nu = \sup_n \nu_n$. By (17) we have $\nu\{\cdot, \chi_n \ne \emptyset\} = \nu_n$ for all n. Assuming $K \in \mathcal{K}_D^r$ with $K \subset K_n^o$, we conclude from (14), (16), and Proposition 34.15 that

$$\begin{aligned} \nu\{\chi K \ne \emptyset\} &= \nu_n\{\chi K \ne \emptyset\} \\ &= \nu_n^n\{\chi K \ne \emptyset\} \\ &= \int P_x\Big\{\psi_n K \ne \emptyset\Big\}\, \mu_n(dx) \\ &= \int P_x\Big\{\tau_K < \zeta\Big\}\, \mu_n(dx) = C_K^D. \end{aligned}$$

The uniqueness of ν is clear by a monotone-class argument. □

We next explore the connection between alternating set functions and random closed sets. As in Chapter 23, we then fix an lcscH space S with Borel σ-field $\mathcal{S}$ and classes $\mathcal{G}$, $\mathcal{F}$, $\mathcal{K}$ of open, closed, and compact sets. Write $\hat{\mathcal{S}} = \{B \in \mathcal{S};\, \bar{B} \in \mathcal{K}\}$, and say that a class $\mathcal{U} \subset \hat{\mathcal{S}}$ is *separating*, if for any $K \in \mathcal{K}$ and $G \in \mathcal{G}$ with $K \subset G$, there exists a set $U \in \mathcal{U}$ with $K \subset U \subset G$.

For any non-decreasing function h on a separating class $\mathcal{U} \subset \hat{\mathcal{S}}$, we define the associated *inner* and *outer capacities* h^o and $\bar{h}$ by

$$\begin{aligned} h^o(G) &= \sup\Big\{h(U);\, U \in \mathcal{U},\, \bar{U} \subset G\Big\}, \quad G \in \mathcal{G}, \\ \bar{h}(K) &= \inf\Big\{h(U);\, U \in \mathcal{U},\, U^o \supset K\Big\}, \quad K \in \mathcal{K}. \end{aligned}$$

Those formulas clearly remain valid with $\mathcal{U}$ replaced by any separating subclass. For any random closed set φ in S, we introduce the associated *hitting function*

$$h(B) = P\{\varphi B \ne \emptyset\} \quad B \in \hat{\mathcal{S}}.$$

Theorem 34.22 *(alternating functions and random sets, Choquet) Let S be an lcscH space with classes $\mathcal{G}$, $\mathcal{F}$, $\mathcal{K}$ of open, closed, and compact sets, and let $\mathcal{U} \subset \hat{\mathcal{S}}$ be separating and closed under finite unions. Then*

(i) *the hitting function h of a random closed set in S is alternating with $h = \bar{h}$ on $\mathcal{K}$ and $h = h^o$ on $\mathcal{G}$,*

(ii) *for any alternating function $p : \mathcal{U} \to [0,1]$ with $p(\emptyset) = 0$, there exists a random closed set with hitting function h, such that $h = \bar{p}$ on $\mathcal{K}$ and $h = p^o$ on $\mathcal{G}$.*

First we clarify the algebraic part of the construction:

Lemma 34.23 *(discrete case) Let the class $\mathcal{U} \subset \hat{\mathcal{S}}$ be finite and closed under unions, and consider an alternating function $h : \mathcal{U} \to [0,1]$ with $h(\emptyset) = 0$. Then there exists a point process ξ on S with*

$$P\{\xi U > 0\} = h(U), \quad U \in \mathcal{U}.$$

Proof: This is obvious when $\mathcal{U} = \{\emptyset\}$. Proceeding by induction, assume the statement to be true when $\mathcal{U}$ is generated by up to $n-1$ sets, and consider a class $\mathcal{U}$ generated by n non-empty sets $B_1, \ldots, B_n$. By scaling we may assume that $h(B_1 \cup \cdots \cup B_n) = 1$.

For any $j \in \{1, \ldots, n\}$, let $\mathcal{U}_j$ be the class of unions formed by the sets $B_i \setminus B_j$, $i \neq j$, and define

$$\begin{aligned} h_j(U) &= \Delta_U h(B_j) \\ &= h(B_j \cup U) - h(B_j), \quad U \in \mathcal{U}_j. \end{aligned}$$

Then each h_j is again alternating with $h_j(\emptyset) = 0$, and so the induction hypothesis yields a point process ξ_j on $\bigcup_i B_i \setminus B_j$ with hitting function h_j. Note that h_j remains the hitting function of ξ_j on all of $\mathcal{U}$. We further introduce a point process ξ_{n+1} with

$$P \bigcap\nolimits_i \Big\{ \xi_{n+1} B_i > 0 \Big\} = (-1)^{n+1} \Delta_{B_1,\ldots,B_n} h(\emptyset).$$

For $1 \leq j \leq n+1$, write ν_j for the restriction of $\mathcal{L}(\xi_j)$ to the set $A_j = \bigcap_{i<j} \{\mu B_i > 0\}$, and put $\nu = \sum_j \nu_j$. Choose ξ to be the canonical point process on S with distribution ν.

To see that ξ has hitting function h, we note that for any $U \in \mathcal{U}$ and $j \leq n$,

$$\begin{aligned} \nu_j\{\mu U > 0\} &= P\Big\{ \xi_j B_1 > 0, \, \ldots, \, \xi_j B_{j-1} > 0, \, \xi_j U > 0 \Big\} \\ &= (-1)^{j+1} \Delta_{B_1,\ldots,B_{j-1},\,U}\, h_j(\emptyset) \\ &= (-1)^{j+1} \Delta_{B_1,\ldots,B_{j-1},\,U}\, h(B_j). \end{aligned}$$

It remains to show that, for any $U \in \mathcal{U} \setminus \{\emptyset\}$,

$$\sum_{j \leq n} (-1)^{j+1} \Delta_{B_1,\ldots,B_{j-1},\,U}\, h(B_j) + (-1)^{n+1} \Delta_{B_1,\ldots,B_n} h(\emptyset) = h(U),$$

which is clear from the fact that

$$\Delta_{B_1,\ldots,B_{j-1},\,U}\, h(B_j) = \Delta_{B_1,\ldots,B_j,\,U}\, h(\emptyset) + \Delta_{B_1,\ldots,B_{j-1},\,U}\, h(\emptyset). \qquad \square$$

Proof of Theorem 34.22: The direct assertion can be proved in the same way as Corollary 34.16. Conversely, let $\mathcal{U}$ and p be such as stated. By Lemma A6.3 we may take $\mathcal{U}$ to be countable, say $\mathcal{U} = \{U_1, U_2, \ldots\}$. For every n, let $\mathcal{U}_n$ be the class of unions of sets from $U_1, \ldots, U_n$. By Lemma 34.23 there exist some point processes $\xi_1, \xi_2, \ldots$ on S, such that

$$P\{\xi_n U > 0\} = p(U), \quad U \in \mathcal{U}_n, \; n \in \mathbb{N}.$$

The space $\mathcal{F}$ is compact by Theorem A6.1, and so Theorem 23.2 yields a random closed set φ in S with $\operatorname{supp}\xi_n \xrightarrow{d} \varphi$ along a sub-sequence $N' \subset \mathbb{N}$. Writing h_n and h for the associated hitting functions, we get

$$\begin{aligned} h(B^o) &\le \liminf_{n\in N'} h_n(B) \\ &\le \limsup_{n\in N'} h_n(B) = h(\bar{B}), \quad B \in \hat{\mathcal{S}}, \end{aligned}$$

and in particular,

$$h(U^o) \le p(U) \le h(\bar{U}), \quad U \in \mathcal{U}.$$

Using the stronger separation property $K \subset U^o \subset \bar{U} \subset G$, we may easily conclude that $h = p^o$ on $\mathcal{G}$ and $h = \bar{p}$ on $\mathcal{K}$. □

Turning to a discussion of super-harmonic and excessive functions, fix a domain $D \subset \mathbb{R}^d$, and let $T_t = T_t^D$ be the transition operators of a Brownian motion X in D, killed on the boundary ∂D. A function $f \ge 0$ on D is said to be *excessive*, if $T_t f \le f$ for all $t > 0$ and $T_t f \to f$ as $t \to 0$. Then clearly $T_t f \uparrow f$. Note that if f is excessive, then $f(X)$ is a super-martingale under P_x for every $x \in D$. The basic example of an excessive function is the Green potential $G^D\nu$ of a measure ν on a Greenian domain D, provided this potential is finite.

Though excessivity is defined globally in terms of the operators T_t^D, it is in fact a local property. For a precise statement, say that a measurable function $f \ge 0$ on D is *super-harmonic*, if for any ball B in D with center x, the average of f over the sphere ∂B is bounded by $f(x)$. As we shall see, it is enough to consider balls in D of radius less than an arbitrary $\varepsilon > 0$. Recall that f is *lower semi-continuous* if $x_n \to x$ implies $\liminf_n f(x_n) \ge f(x)$.

Theorem 34.24 *(super-harmonic and excessive functions, Doob) For any measurable function $f \ge 0$ on a domain $D \subset \mathbb{R}^d$, these conditions are equivalent:*

(i) *f is excessive,*

(ii) *f is super-harmonic and lower semi-continuous.*

Our proof is based on two lemmas, beginning with a relationship between the two continuity properties.

Lemma 34.25 *(semi-continuity) Let $f \ge 0$ be a measurable function on a domain $D \subset \mathbb{R}^d$, such that $T_t f \le f$ for all $t > 0$. Then these conditions are equivalent:*

(i) *f is excessive,*

(ii) *f is lower semi-continuous.*

Proof: Assume (i), and let $x_n \to x$ in D. By Theorem 34.7 and Fatou's lemma,

$$\begin{aligned} T_t f(x) &= \int p_t^D(x,y)\, f(y)\, dy \\ &\le \liminf_{n\to\infty} \int p_t^D(x_n,y)\, f(y)\, dy \\ &= \liminf_{n\to\infty} T_t f(x_n) \\ &\le \liminf_{n\to\infty} f(x_n), \end{aligned}$$

and as $t \to 0$ we get $f(x) \le \liminf_n f(x_n)$, which implies (ii).

Next assume (ii). Using the continuity of X and Fatou's lemma, we get as $t \to 0$ along an arbitrary sequence

$$\begin{aligned} f(x) &= E_x f(X_0) \\ &\le E_x \liminf_{t\to 0} f(X_t) \\ &\le \liminf_{t\to 0} E_x f(X_t) \\ &= \liminf_{t\to 0} T_t f(x) \\ &\le \limsup_{t\to 0} T_t f(x) \le f(x). \end{aligned}$$

Thus, $T_t f \to f$, proving (i). □

For smooth functions, the super-harmonic property is easily described.

Lemma 34.26 *(smooth functions) For any function $f \ge 0$ in C_D^2,*

$$\begin{aligned} f \textit{ is super-harmonic } &\Leftrightarrow \Delta f \le 0 \\ &\Rightarrow f \textit{ is excessive.} \end{aligned}$$

Proof: By Itô's formula, the process

$$M_t = f(X_t) - \tfrac{1}{2} \int_0^t \Delta f(X_s)\, ds, \quad t \in [0,\zeta), \tag{18}$$

is a continuous local martingale. Now fix any closed ball $B \subset D$ with center x, and write $\tau = \tau_{\partial B}$. Since $E_x \tau < \infty$, we get by dominated convergence

$$f(x) = E_x f(X_\tau) - \tfrac{1}{2}\, E_x \int_0^\tau \Delta f(X_s)\, ds.$$

Thus, f is super-harmonic iff the last expectation is ≤ 0, and the first assertion follows.

To prove the last implication, we note that the exit time $\zeta = \tau_{\partial D}$ is predictable, say with announcing sequence (τ_n). If $\Delta f \le 0$, we get from (18) by optional sampling

$$E_x\Big\{ f(X_{t\wedge\tau_n});\, t < \zeta \Big\} \le E_x f(X_{t\wedge\tau_n}) \le f(x).$$

Hence, Fatou's lemma yields $E_x\{f(X_t);\, t < \zeta\} = T_t f(x)$, and so f is excessive by Lemma 34.25. □

Proof of Theorem 34.24: If f is excessive or super-harmonic, then so is $f \wedge n$ for every $n > 0$ by Lemma 34.25. The converse statement is also true, by monotone convergence, and since the lower semi-continuity is preserved by increasing limits. Thus, we may henceforth assume that f is bounded.

Now assume (i). By Lemma 34.25 it is then lower semi-continuous, and it remains to prove that f is super-harmonic. Since the property $T_t f \leq f$ is preserved as we pass to a sub-domain, we may take D to be bounded. For each $h > 0$, we define $q_h = h^{-1}(f - T_h f)$ and $f_h = G^D q_h$. Since f and D are bounded, we have $G^D f < \infty$, and so $f_h = h^{-1} \int_0^h T_s f ds \uparrow f$. By the strong Markov property, we further see that for any optional time $\tau < \zeta$,

$$\begin{aligned} E_x f_h(X_\tau) &= E_x E_{X_\tau} \int_0^\infty q_h(X_s)\, ds \\ &= E_x \int_0^\infty q_h(X_{s+\tau})\, ds \\ &= E_x \int_\tau^\infty q_h(X_s)\, ds \leq f_h(x). \end{aligned}$$

In particular, f_h is super-harmonic for each h, and so by monotone convergence the same property holds for f.

Conversely, assume (ii). To prove (i), it suffices by Lemma 34.25 to show that $T_t f \leq f$ for all t. Then fix a spherically symmetric probability density $\psi \in C^\infty(\mathsf{R}^d)$ supported by the unit ball, and put $\psi_h(x) = h^{-d}\, \psi(x/h)$ for each $h > 0$. Writing ρ for the Euclidean metric in R^d, we define $f_h = \psi_h * f$ on the set $D_h = \{x \in D;\ \rho(x, D^c) > h\}$. Note that $f_h \in C^\infty(D_h)$ for all h, that f_h is super-harmonic on D_h, and that $f_h \uparrow f$. By Lemma 34.26 and monotone convergence, we conclude that f is excessive on each set D_h. Letting ζ_h be the first exit time from D_h, we obtain

$$E_x\Big\{f(X_t);\, t < \zeta_h\Big\} \leq f(x), \quad h > 0.$$

As $h \to 0$ we have $\zeta_h \uparrow \zeta$, and hence $\{t < \zeta_h\} \uparrow \{t < \zeta\}$. By monotone convergence, we get $T_t f(x) \leq f(x)$. □

Now we prove the remarkable fact that, although an excessive function f may not be continuous, the super-martingale $f(X)$ is a.s. continuous under P_x for every x.

Theorem 34.27 *(continuity, Doob) Let f be an excessive function on a domain $D \subset \mathsf{R}^d$, and let X be a Brownian motion, killed upon hitting ∂D. Then the process $f(X_t)$ is a.s. continuous on $[0, \zeta)$.*

Our proof is based on the following invariance under time reversal, for a stationary version of Brownian motion. Though no such process exists in the usual probabilistic sense, we do have the following:

Lemma 34.28 *(time reversal, Doob) Let P_x be the distribution of Brownian motion X in R^d starting at x, and put $\bar{P} = \int P_x\, dx$. Then for any $c > 0$, these processes have the same distribution under $\bar{P}$:*

$$Y_t^c = X_t, \quad \tilde{Y}_t^c = X_{c-t}, \qquad t \in [0, c].$$

Proof: Introduce the processes

$$\begin{aligned} B_t &= X_t - X_0, \\ \tilde{B}_t &= X_{c-t} - X_c, \qquad t \in [0, c], \end{aligned}$$

and note that B and $\tilde{B}$ are Brownian motions on $[0, c]$ under each P_x. Fix any measurable function $f \geq 0$ on $C_{[0,c],\mathbb{R}^d}$. By Fubini's theorem and the invariance of Lebesgue measure, we get

$$\begin{aligned} \bar{E} f(\tilde{Y}^c) &= \bar{E} f\Big(X_0 - \tilde{B}_c + \tilde{B}\Big) \\ &= \int E_x f\Big(x - \tilde{B}_c + \tilde{B}\Big)\, dx \\ &= \int E_0 f\Big(x - \tilde{B}_c + \tilde{B}\Big)\, dx \\ &= E_0 \int f\Big(x - \tilde{B}_c + \tilde{B}\Big)\, dx \\ &= E_0 \int f(x + \tilde{B})\, dx \\ &= \int E_x f(Y^c)\, dx \\ &= \bar{E} f(Y^c). \end{aligned}$$

□

Proof of Theorem 34.27: Since $f \wedge n$ is again excessive for each $n > 0$ by Theorem 34.24, we may assume that f is bounded. As in the proof of the same theorem, we may then approximate f by smooth, excessive functions $f_h \uparrow f$ on suitable subdomains $D_h \uparrow D$. Since $f_h(X)$ is a continuous super-martingale, up to the exit time ζ_h from D_h, Theorem 9.33 shows that $f(X)$ is a.s. right-continuous on $[0, \zeta)$, under any initial distribution μ. Using the Markov property at rational times, we may extend the a.s. right-continuity to the random time set $T = \{t \geq 0;\, X_t \in D\}$.

To strengthen the result to a.s. continuity on T, we note that $f(X)$ is right-continuous on T, a.e. $\bar{P}$. By Lemma 34.28, it follows that $f(X)$ is also left-continuous on T, a.e. $\bar{P}$. Thus, $f(X)$ is continuous on T, a.s. P_μ for arbitrary $\mu \ll \lambda^d$. Since $P_\mu \circ X_h^{-1} \ll \lambda^d$ for any μ and $h > 0$, we conclude that $f(X)$ is a.s. continuous on $T \cap [h, \infty)$ for any $h > 0$. Combining with the right-continuity at 0, we obtain the asserted continuity on $[0, \zeta)$. □

If f is excessive, then $f(X)$ is a super-martingale under P_x for every x, and hence has a Doob–Meyer decomposition $f(X) = M - A$. It is remarkable that we can choose A to be a continuous additive functional (CAF) of X independent of x. A similar situation was encountered in connection with Theorem 29.23.

Theorem 34.29 *(excessive functions and additive functionals, Meyer) Let f be an excessive function on a domain $D \subset \mathbb{R}^d$, and let P_x be the distribution of Brownian motion X in D, killed upon hitting ∂D. Then*

$$M = f(X) + A \ \text{ a.s. } P_x, \quad x \in D,$$

for some a.s. unique processes A and M, where

(i) *A is a CAF of X,*

(ii) *M is a continuous local P_x-martingale on $[0,\zeta)$, for every $x \in D$.*

The main difficulty of the proof is to construct a version of the process A that compensates $-f(X)$ under *every* measure P_μ. Here the following lemma is helpful.

Lemma 34.30 *(universal compensation) Consider an excessive function f on a domain $D \subset \mathsf{R}^d$, a distribution $m \sim \lambda^d$ on D, and a P_m-compensator A of $-f(X)$ on $[0,\zeta)$. Then for any distribution μ and constant $h > 0$, the process $A \circ \theta_h$ is a P_μ-compensator of $-f(X \circ \theta_h)$ on $[0, \zeta \circ \theta_h)$.*

In other words, the process $M_t = f(X_t) + A_{t-h} \circ \theta_h$ is a local P_μ-martingale on $[h,\zeta)$ for every μ and h.

Proof: For any bounded P_m-martingale M and initial distribution $\mu \ll m$, we note that M is also a P_μ-martingale. To see this, write $k = d\mu/dm$, and note that $P_\mu = k(X_0) \cdot P_m$. It is equivalent to show that $N_t = k(X_0)M_t$ is a P_m-martingale, which is clear since $k(X_0)$ is $\mathcal{F}_0$-measurable with mean 1.

Now fix any distribution μ and constant $h > 0$. To prove the stated property of A, it is enough to show that, for any bounded P_m-martingale M, the process $N_t = M_{t-h} \circ \theta_h$ is a P_μ-martingale on $[h,\infty)$. Then fix any times $s < t$ and sets $F \in \mathcal{F}_h$ and $G \in \mathcal{F}_s$. Using the Markov property at h, and noting that $P_\mu \circ X_h^{-1} \ll m$, we get

$$\begin{aligned} E_\mu\Big(M_t \circ \theta_h;\ F \cap \theta_h^{-1} G\Big) &= E_\mu\Big\{E_{X_h}(M_t; G); F\Big\} \\ &= E_\mu\Big\{E_{X_h}(M_s; G); F\Big\} \\ &= E_\mu\Big(M_s \circ \theta_h;\ F \cap \theta_h^{-1} G\Big). \end{aligned}$$

Hence, a monotone-class argument yields $E_\mu\Big(M_t \circ \theta_h \,\Big|\, \mathcal{F}_{h+s}\Big) = M_s \circ \theta_h$ a.s. $\square$

Proof of Theorem 34.29: Let A^μ be the P_μ-compensator of $-f(X)$ on $[0,\zeta)$, and note that A^μ is a.s. continuous, e.g. by Theorem 19.11. Fix any distribution $m \sim \lambda^d$ on D, and conclude from Lemma 34.30 that $A^m \circ \theta_h$ is a P_μ-compensator of $-f(X \circ \theta_h)$ on $[0, \zeta \circ \theta_h)$, for any μ and $h > 0$. Since this remains true for the process $A^\mu_{t+h} - A^\mu_h$, we get for any μ and $h > 0$

$$A^\mu_t = A^\mu_h + A^m_{t-h} \circ \theta_h, \quad t \geq h, \ \text{a.s. } P_\mu. \tag{19}$$

Restricting h to the positive rationals, we define

$$A_t = \lim_{h \to 0} A^m_{t-h} \circ \theta_h, \quad t > 0,$$

whenever the limit exists and is continuous and non-decreasing with $A_0 = 0$, and put $A = 0$ otherwise. By (19) we have $A = A^\mu$ a.s. P_μ for every μ, and so A is a P_μ-compensator of $-f(X)$ on $[0,\zeta)$ for every μ. For each $h > 0$, it follows by Lemma 34.30 that $A \circ \theta_h$ is a P_μ-compensator of $-f(X \circ \theta_h)$ on $[0, \zeta \circ \theta_h)$, and since this is also true for the process $A_{t+h} - A_h$, we get $A_{t+h} = A_h + A_t \circ \theta_h$ a.s. P_μ. Thus, A is a CAF. □

We can now establish a probabilistic version of the classical Riesz decomposition. To avoid some obscuring technicalities, we restrict our attention to locally bounded functions f. By the *greatest harmonic minorant* of f we mean a harmonic function $h \le f$, dominating all other such functions. Recall that the *potential* U_A of a CAF A of X is given by $U_A(x) = E_x A_\infty$.

Theorem 34.31 *(Riesz decomposition) Let $f \ge 0$ be a locally bounded function on a domain $D \subset \mathsf{R}^d$, and let X be a Brownian motion in D, killed upon hitting ∂D. Then these conditions are equivalent:*

(i) *f is excessive,*

(ii) *$f = U_A + h$ for a CAF A of X and a harmonic function $h \ge 0$.*

In that case,

(iii) *A is the compensator of $-f(X)$,*

(iv) *h is the greatest harmonic minorant of f.*

A similar result for uniformly α-excessive functions of an arbitrary Feller process was obtained in Theorem 29.23. From the classical Riesz representation on Greenian domains, we know that U_A may also be written as the Green potential of a unique measure ν_A, so that $f = G^D \nu_A + h$. In the special case where $D = \mathsf{R}^d$ with $d \ge 3$, we recall from Theorem 29.21 that $\nu_A B = \bar{E}(1_B \cdot A)_1$. A similar representation holds in the general case.

Proof of Theorem 34.31: First let A be a CAF with $U_A < \infty$. By the additivity of A and the Markov property of X, we get for any $t > 0$

$$\begin{aligned} U_A(x) &= E_x A_\infty = E_x\big(A_t + A_\infty \circ \theta_t\big) \\ &= E_x A_t + E_x E_{X_t} A_\infty \\ &= E_x A_t + T_t U_A(x). \end{aligned}$$

By dominated convergence, $E_x A_t \downarrow 0$ as $t \to 0$, and so U_A is excessive. Even $U_A + h$ is then excessive for any harmonic function $h \ge 0$.

Conversely, let f be excessive and locally bounded. By Theorem 34.29 there exists a CAF A, such that $M = f(X) + A$ is a continuous local martingale on $[0, \zeta)$. For any localizing and announcing sequence $\tau_n \uparrow \zeta$, we get

$$\begin{aligned} f(x) &= E_x M_0 = E_x M_{\tau_n} \\ &= E_x f(X_{\tau_n}) + E_x A_{\tau_n} \\ &\ge E_x A_{\tau_n}. \end{aligned}$$

As $n \to \infty$, we obtain $U_A \le f$ by monotone convergence.

By the additivity of A and the Markov property of X, we have

$$\begin{aligned} E_x\Big(A_\infty \,\Big|\, \mathcal{F}_t\Big) &= A_t + E_x\Big(A_\infty \circ \theta_t \,\Big|\, \mathcal{F}_t\Big) \\ &= A_t + E_{X_t} A_\infty \\ &= M_t - f(X_t) + U_A(X_t). \end{aligned} \tag{20}$$

Writing $h = f - U_A$, it follows that $h(X)$ is a continuous local martingale. Since h is locally bounded, we conclude by optional sampling and dominated convergence that h has the mean-value property. Thus, h is harmonic by Lemma 34.3.

To prove the uniqueness of A, suppose that f has also a representation $U_B + k$ for some CAF B and harmonic function $k \ge 0$. Proceeding as in (20), we get

$$A_t - B_t = E_x\Big(A_\infty - B_\infty \,\Big|\, \mathcal{F}_t\Big) + h(X_t) - k(X_t), \quad t \ge 0,$$

which shows that $A - B$ is a continuous local martingale. Hence, Proposition 18.2 yields $A = B$ a.s.

To prove (iv), consider any harmonic minorant $k \ge 0$. Since $f - k$ is again excessive and locally bounded, it has a representation $U_B + l$ for some CAF B and harmonic function l. But then $f = U_B + k + l$, and so $A = B$ a.s. and $h = k + l \ge k$. □

For a sufficiently regular measure ν on R^d, we may construct an associated CAF A of Brownian motion X, such that A increases only when X visits the support of ν. This clearly extends the notion of local time. For convenience we may write $G^D(1_D\nu) = G^D\nu$.

Proposition 34.32 *(additive functional induced by measure) Let the measure ν on R^d be such that $U(1_D\nu)$ is bounded for every bounded domain D. Then*

(i) *there exists a CAF A of Brownian motion X, such that for any D*

$$E_x A_{\zeta_D} = G^D\nu(x), \quad x \in D,$$

(ii) *ν and A determine each other a.s. uniquely,*

(iii) $\operatorname{supp} A \subset \Big\{t \ge 0;\ X_t \in \operatorname{supp} \nu\Big\}$ *a.s.*

The proof is straightforward, given the classical Riesz decomposition, and we indicate only the main steps.

Proof: (i)−(ii) A simple calculation shows that $G^D\nu$ is excessive for any bounded domain D. Since $G^D\nu \le U(1_D\nu)$, it is also bounded. Hence, Theorem 34.31 yields a CAF A_D of X on $[0, \zeta_D)$ and a harmonic function $h_D \ge 0$, such that $G^D\nu = U_{A_D} + h_D$. In fact, $h_D = 0$ by Riesz' theorem.

Now consider another bounded domain $D' \supset D$. We claim that $G^{D'}\nu - G^D\nu$ is harmonic on D. This is clear from the analytic definitions, and it also follows, under a regularity condition, from Lemma 34.13. Since A_D and $A_{D'}$ are compensators of $-G^D\nu(X)$ and $-G^{D'}\nu(X)$, respectively, $A_D - A_{D'}$ is a martingale on $[0, \zeta_D)$, and so $A_D = A_{D'}$ a.s. up to time ζ_D. Now choose some bounded domains $D_n \uparrow \mathsf{R}^d$, and define $A = \sup_n A_{D_n}$, so that $A = A_D$ a.s. on $[0, \zeta_D)$ for all D. It is easy to see that A is a CAF of X, and that (i) holds for any bounded domain D. The uniqueness of ν is clear from the uniqueness in the classical Riesz decomposition.

(iii) Note that $G^D\nu$ is harmonic on $D \setminus \operatorname{supp} \nu$ for every D, so that $G^D\nu(X)$ is a local martingale on the predictable set $\big\{t < \zeta_D;\ X_t \notin \operatorname{supp} \nu\big\}$. $\square$

Exercises

1. For any domain $D \subset \mathsf{R}^2$ and point $x \in \partial D$, let $x \in I \subset D^c$ for a line segment I. Show that x is regular for D^c. (*Hint:* Consider the windings around x of a Brownian motion starting at x, using the strong Markov property and Brownian scaling.)

2. Compute the Newtonian potential kernel $g = g^D$ when $D = \mathsf{R}^d$ with $d \geq 3$, and check by direct computation that $g(x, y)$ is harmonic in $x \neq y$ for fixed y.

3. For any domain $D \subset \mathsf{R}^d$, show that $p_t(x, y) - p_t^D(x, y) \to 0$ as $t \to 0$, uniformly for $x \neq y$ in any compact set $K \subset D$. Also prove the same convergence as $\inf\{|x|;\ x \notin D\} \to \infty$, uniformly for bounded $t > 0$ and $x \neq y$. (*Hint:* Note that $p_t(x, y)$ is uniformly bounded for $|x - y| > \varepsilon > 0$, and use (7).)

4. For a domain $D \subset \mathsf{R}^d$ with $d \geq 3$, show that $g(x, y) - g^D(x, y)$ is uniformly bounded for $x \neq y$ in a compact set $K \subset D$. Also show that the difference tends to 0 as $\inf\{|x|;\ x \notin D\} \to \infty$, uniformly for $x \neq y$ in K. (*Hint:* Use Lemma 34.13.)

5. Show that the equilibrium measure μ_K^D is restricted to the outer boundary of K and agrees for all sets K with the same outer boundary. (Here the *outer boundary* of K consists of all points $x \in \partial K$ that can be connected to D^c or ∞ by a path through K^c.) Prove a corresponding statement for the sweeping measure ν in Theorem 34.12.

6. For any Greenian domain $D \subset \mathsf{R}^d$, disjoint sets $K_1, \ldots, K_n \in \mathcal{K}_D^r$, and constants $p_1, \ldots, p_d \in \mathsf{R}$, prove the existence of a unique signed measure ν on $\bigcup_j K_j$ with $G^D\nu = p_j$ on K_j for all j. (*Hint:* Use Corollary 34.17 recursively.)

7. Let $\varphi_1 \perp\!\!\!\perp \varphi_2$ be random sets with distributions π_{ν_1}, π_{ν_2}. Show that $\varphi_1 \cup \varphi_2$ has distribution $\pi_{\nu_1+\nu_2}$.

8. Extend Theorem 34.22 to unbounded functions p. (*Hint:* Consider the restrictions to compact sets, and proceed as in Theorem 34.21.)

9. For a second degree polynomial $f(x_1, \ldots, x_d)$, find necessary and sufficient conditions for f to be super-harmonic. (*Hint:* First discard the constant and first order terms. Then diagonalize to reduce to a linear combination in the squares x_k^2.)

10. Say that f is *sub-harmonic* if $-f$ is super-harmonic. Show that every smooth, convex function is sub-harmonic. Then show that every positive linear combination of sub-harmonic functions is again sub-harmonic. Finally, give an example of a sub-harmonic function that is not convex. (*Hint:* Use the result of the previous exercise.)

11. Show that there is no distribution ν on R^d such that, under P_ν, a Brownian motion X_t and its time reversal $Y_t = X_{1-t}$ have the same distribution on $[0, 1]$. (*Hint:*

Note that X_1 has distribution $\nu * \mu$, where μ is the standard normal distribution. To see that the equation $\nu * \mu = \nu$ has no solution, extend by iteration to $\nu * \mu^{*n} = \nu$, and let $n \to \infty$.)

12. Show that a Gaussian process is time-reversible iff it is stationary. Then extend this result to any mixture of Gaussian processes. (*Hint:* Use Lemma 14.1.)

Chapter 35

Stochastic Differential Geometry

Semi-martingales, bilinear forms, pull-back, covariation process, semi-martingale integral, connection, chain rule, differential operators and Christoffel symbols, martingale criteria, induced connections, affine and convex maps, geodesics and martingale criteria, local drift and diffusion rates, sub-manifolds and projection, diffusions, Riemannian metric and connection, isotropic martingales, Brownian motion, harmonic maps, Lévy processes in Lie groups

Just as differential geometry deals with the geometric properties of smooth curves and other objects in a differential manifold S, its stochastic counterpart involves a study of sufficiently regular processes in S, which may include the development of a relevant stochastic calculus. We may think of the manifold as a smooth surface of finite dimension n, embedded in a higher-dimensional Euclidean space. Though such an embedding always exists, it is far from unique and may even allow some smooth bending and stretching. Thus, we insist that all objects on S are described *intrinsically* in terms of the local properties of the surface *itself*. Indeed, though the differential structure of S can be described in terms of some local coordinates in a small neighborhood of an arbitrary point, the actual choice of coordinates is fairly arbitrary, and no properties are allowed to depend on any specific choice.

When developing stochastic calculus in a Euclidean space, we may start with Brownian motion, then proceed to continuous martingales, via a stochastic integration or random time change, and eventually reach the general semi-martingales via Itô's formula. In a general differential manifold S we must proceed in the opposite direction, since there is no natural way to define a Brownian motion or some more general martingales in S. Indeed, to define martingales we need to endow S with a *connection* ∇, and for the notion of Brownian motion in S we need even a *Riemannian metric* ρ. Following such a program leads to an amazingly rich and beautiful theory, which has the further advantage of providing some deeper insight into the classical notions of martingales and stochastic calculus.

In this chapter, we consider first the class of general *semi-martingales* in S, along with their *covariation processes*. We then introduce *connections* on S, and study the associated classes of *martingales* and *geodesics*. Finally, we consider a general *Riemannian metric* with associated *canonical connection*, which allow us to define a *Brownian motion* and related processes in S. Though no semi-martingale decomposition is possible in an abstract manifold S, owing to the lack of any additive structure, we may still define some *local characteristics*

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1_36

of a semi-martingale in S, describing the intrinsic *drift* and *diffusion rates* of the process. In particular, a martingale is then a process with drift 0.

Though the mentioned notions are all intrinsic, they can be justified by some natural embedding and projection properties. Then recall that every differential manifold S can be embedded as a smooth sub-manifold of a suitable Euclidean space. For a Riemannian manifold S, we can even choose the embedding to preserve the Riemannian structure.[1] In both cases, the relevant martingale and rate properties agree with the corresponding extrinsic notions. The same thing is true for the appropriate projection properties.

Though the reader is assumed to be at least a vaguely familiar with differential geometry, some basic notions and results are reviewed here and in Appendix 7. Write $\mathcal{S}$ for the class of smooth functions $f, g, \dots$ on S, and let T_S be the space of smooth (tangent) vector fields $u, v, \dots$ on S. A *form* is a smooth function α on the dual space of cotangent vectors T_S^*, and a *bilinear form* is a smooth function b on the dual $(T_S^*)^{\otimes 2}$ of the product space T_S^2. Any simple or bilinear forms can be written as finite sums[2]

$$\alpha = a_i\, df^{\,i}, \qquad b = b_{ij}\,(df^{\,i} \otimes dg^j),$$

for suitable $a_i, b_{ij}, f^{\,i}, g^j \in \mathcal{S}$, where $df(u) = uf$, and we note that in local coordinates $df = (\partial_i f)(x)\, dx^i$. Further note that $(\alpha \otimes \beta)(u, v) = (\alpha u)(\beta v)$.

For any smooth mapping $\varphi\colon S \to S'$, the *push-forward* $\varphi \circ u$ is a mapping from T_S to $T_{S'}$, given by[3]

$$(\varphi \circ u) f = u(f \circ \varphi), \quad f \in \mathcal{S}',$$

and the dual *pull-back* φ^* or φ^{*2} of a simple or bilinear form on S' into a corresponding form on S is defined by

$$\begin{aligned} (\varphi^* \alpha) u &= \alpha(\varphi \circ u), \\ (\varphi^{*2} b)(u, v) &= b(\varphi \circ u,\, \varphi \circ v). \end{aligned} \tag{1}$$

In particular, we note that

$$\begin{aligned} \varphi^{*2}(\alpha \otimes \beta) &= (\varphi^* \alpha)(\varphi^* \beta), \\ (\varphi^* df) u &= (\varphi \circ u) f = u(f \circ \varphi) \\ &= d(f \circ \varphi) u. \end{aligned}$$

A continuous[4] process X in a manifold S is called a *semi-martingale*, if $f(X)$ is a real semi-martingale for every smooth function f on S. A fundamental role is played by the *covariation process*[5] $b\,[X]$, defined as follows for every bilinear form b on S.

[1] Those are the celebrated embedding theorems of Whitney and Nash.

[2] Note that the coefficients are functions, not constants.

[3] To ease the access for probabilists, we often use a simplified notation. Traditional versions of the stated formulas are sometimes inserted, here or in Appendix 7.

[4] usually omitted, since all processes in this chapter are assumed to be continuous

[5] $b\,[X]$ is often written as $\int b(\,dX, dX)$

Theorem 35.1 *(covariation process) Let X be a semi-martingale in a smooth manifold S. Then for any bilinear form b there exists a finite-variation process $b[X]$, depending linearly on b, such that a.s. for smooth functions f, g and maps $\varphi\colon S\to S'$,*

(i) $(df\otimes dg)[X]=[f(X),\,g(X)]$,

(ii) $(fb)[X]=f(X)\cdot b[X]$,

(iii) $(\varphi^{*2}b)[X]=b[\varphi(X)]$,

(iv) *b is symmetric, non-negative definite $\Rightarrow$ $b[X]$ is non-decreasing.*

The map $b\mapsto b[X]$ is a.s. uniquely determined by (i)–(ii).

(In traditional notation, (ii) and (iii) may be written as

$$\int (fb)\,(dX,dX) \;=\; \int (f\circ X)\,d\Big\{\int b(dX,dX)\Big\},$$
$$\int_S (T^*\varphi\otimes T^*\varphi)\,b(dX,dX) \;=\; \int_{S'} b\,\Big\{d(\varphi\circ X),\,d(\varphi\circ X)\Big\}.)$$

Proof: Fix a local chart with coordinates $x_1,\dots,x_n$, and consider any bilinear form b, represented as a finite sum $b_{ij}(d\varphi^i\otimes d\varphi^j)$ in terms of some smooth functions b_{ij} and φ^i. If $b[X]$ exist with the stated properties, then by (i)–(ii), Itô's formula, and the chain rule for Stieltjes integrals,

$$\begin{aligned} b[X] &= b_{ij}(X)\cdot(d\varphi^i\otimes d\varphi^j)[X]\\ &= b_{ij}(X)\cdot[\varphi^i(X),\varphi^j(X)]\\ &= b_{ij}(X)\cdot\Big\{\partial_h\varphi^i(X)\,\partial_k\varphi^j(X)\cdot[X^h,X^k]\Big\}\\ &= \Big\{b_{ij}(\partial_h\varphi^i)(\partial_k\varphi^j)\Big\}(X)\cdot[X^h,X^k]. \end{aligned} \tag{2}$$

To see that $b[X]$ is well-defined by this expression, we need to show that the right-hand side vanishes a.s. when $b=0$. Using the coordinate representation $d\varphi=\partial_i\varphi(x)\,dx^i$, we obtain

$$\begin{aligned} b &= b_{ij}(d\varphi^i\otimes d\varphi^j)\\ &= b_{ij}(x)\Big\{\partial_h\varphi^i(x)\,dx^h\Big\}\otimes\Big\{\partial_k\varphi^j(x)\,dx^k\Big\}\\ &= \Big\{b_{ij}(\partial_h\varphi^i)(\partial_k\varphi^j)\Big\}(x)\,\Big(dx^h\otimes dx^k\Big). \end{aligned}$$

If $b=0$, we get $b_{ij}(\partial_h\varphi^i)(\partial_k\varphi^j)=0$ for all h and k, and (2) yields $b[X]=0$ a.s.

With $b[X]$ as in (2), property (i) becomes a special case, whereas (ii) follows from (2) if we use the same chain rule to write

$$\begin{aligned} (fb)[X] &= \Big\{fb_{ij}\,(d\varphi^i\otimes d\varphi^j)\Big\}\,[X]\\ &= (f\,b_{ij})(X)\cdot[\varphi^i(X),\,\varphi^j(X)]\\ &= f(X)\cdot\Big\{b_{ij}(X)\cdot[\varphi^i(X),\,\varphi^j(X)]\Big\}\\ &= f(X)\cdot b[X]. \end{aligned}$$

(iii) By the linearity of both sides, it is enough to take $b = f(dg \otimes dh)$ for some smooth functions f, g, h. Noting that

$$\begin{aligned}\varphi^{*2}\Big\{f(dg \otimes dh)\Big\} &= (f \circ \varphi)(\varphi^* dg \otimes \varphi^* dh)\\ &= (f \circ \varphi)\Big\{d(g \circ \varphi) \otimes d(h \circ \varphi)\Big\},\end{aligned}$$

we get by (i)−(ii)

$$\begin{aligned}(\varphi^{*2} b)[X] &= \varphi^{*2}\Big\{f(dg \otimes dh)\Big\}[X]\\ &= (f \circ \varphi)\Big\{d(g \circ \varphi) \otimes d(h \circ \varphi)\Big\}[X]\\ &= (f \circ \varphi)(X) \cdot \Big[g \circ \varphi(X),\ h \circ \varphi(X)\Big]\\ &= f(\varphi \circ X) \cdot \Big[g(\varphi \circ X),\ h(\varphi \circ X)\Big]\\ &= \Big\{f(dg \otimes dh)\Big\}[\varphi \circ X]\\ &= b[\varphi \circ X].\end{aligned}$$

(iv) In local coordinates, we may write b as a finite sum $b_{ij}(x)(dx^i \otimes dx^j)$. Letting $A = (a_{ij})$ be the positive square root of the matrix $B = (b_{ij})$, so that $b_{ij} = \sum_k a_{ik} a_{jk}$, we get by (i)−(ii)

$$\begin{aligned}b\,[X] &= b_{ij}(X) \cdot [X^i, X^j]\\ &= \textstyle\sum_k \Big[a_{ik}(X) \cdot X^i,\ a_{jk}(X) \cdot X^j\Big] \geq 0.\end{aligned} \qquad \square$$

Given a semi-martingale X in a differentiable manifold S, along with a first order form $\alpha = f_i\, dh^i$ on S, we define the integral $\langle \alpha, X\rangle$ of α along X as the real semi-martingale[6]

$$\langle\, \alpha, X\rangle = \int f_i(X) \circ d\,(h^i \circ X), \tag{3}$$

where $\int Y \circ dX = Y \circ X$ denotes the Fisk–Stratonovich integral of Y with respect to X. Note that we must use FS rather than Itô integrals to ensure independence of the choice of local coordinates[7]. This is shown in statement (i) below, along with some further basic properties.

Theorem 35.2 *(semi-martingale integral) For a semi-martingale X and a first order form α in a manifold S, define $\langle\, \alpha, X\rangle$ by* (3). *Then a.s.*

(i) $\langle\, \alpha, X\rangle$ *is independent of the representation of* α,

(ii) $\langle\, df, X\rangle_t = f(X_t) - f(X_0)$,

(iii) $\langle f\alpha, X\rangle = f(X) \circ \langle\, \alpha, X\rangle$,

(iv) $\Big[\langle\, \alpha, X\rangle, \langle\, \beta, X\rangle\Big] = (\alpha \otimes \beta)[X]$.

[6] Also written as $\int\langle\, d\alpha, \mathbf{d}X\rangle = \int\langle\, \alpha, \delta X\rangle$. Summation over repeated indices is understood. Our circle notation must not be confused with the composition of functions.

[7] Even in the extrinsic case we must use FS-integrals, but then for a different reason: to ensure that the resulting process will stay in the manifold.

Proof: (i) By linearity, it is enough to show that $f_i\,dh^i = 0$ implies $\langle f_i\,dh^i, X\rangle = 0$ a.s. In local coordinates, we may write $X = (X^1, \dots, X^n)$ and note that $f_i\,\partial_k h^i = 0$ for all k. Using Theorem 18.21 (i)−(ii), we get

$$\begin{aligned}\int f_i(X) \circ d\,(h^i(X)) &= \int f_i(X) \circ \int \Big(\partial_k h^i(X)\Big) \circ dX^k \\ &= \int f_i(X)\,\partial_k h^i(X) \circ dX^k \\ &= \int \Big(f_i\,\partial_k h^i\Big)(X) \circ dX^k = 0.\end{aligned}$$

(ii) Take $\alpha = 1 \cdot df$ in (3).

(iii) Using (3) and Theorem 18.21 (ii), we get for $\alpha = g_i\,dh^i$

$$\begin{aligned}\langle f\alpha, X\rangle &= \int f g_i(X) \circ d\,(h^i(X)) \\ &= \int f(X) \circ \int g_i(X) \circ d\,(h^i(X)) \\ &= \int f(X) \circ d\,\langle\,\alpha, X\rangle.\end{aligned}$$

(iv) Assuming $\alpha = f_i\,dh^i$ and $\beta = g_j\,dh^j$, and using (3) and Theorem 35.1 (i)−(ii), along with some basic properties of the covariation, we get

$$\begin{aligned}\Big[\langle\,\alpha, X\rangle, \langle\,\beta, X\rangle\Big] &= \Big[f_i(X)\cdot h^i(X),\; g_j(X)\cdot h^j(X)\Big] \\ &= f_i\,g_j(X)\cdot \Big[h^i(X),\; h^j(X)\Big] \\ &= f_i\,g_j(X)\cdot \Big(dh^i \otimes dh^j\Big)[X] \\ &= f_i\,g_j\Big(dh^i \otimes dh^j\Big)\,[X] \\ &= \Big(f_i\,dh^i \otimes g_j\,dh^j\Big)[X] \\ &= (\alpha \otimes \beta)[X].\end{aligned}$$

□

− − −

Before proceeding to the intrinsic theory of martingales in a differentiable manifold, we insert a somewhat informal discussion[8] of semi-martingales in a smooth sub-manifold S of a Euclidean space R^n, motivating the general results in Theorems 35.8, 35.16−35.17, and 35.20. We may then define a *martingale* in S as an S-valued semi-martingale X in R^n with canonical decomposition $M + A$, where locally for $X_t = x \in S$, the martingale increment dM_t is restricted to the tangent space T_x, whereas the compensating part dA_t is perpendicular to T_x. The role of the drift term A is to ensure that the process X will stay in the manifold.

Given a semi-martingale X in R^n, we define the *projection* $\hat{X} = \pi_S(X)$ onto S starting at $x \in S$ as the unique strong solution to the Fisk–Stratonovich SDE

$$\hat{X}^i_t = x^i + \int_0^t P^i_j(\hat{X}_s) \circ dX^j_s, \quad t \geq 0,\;\; i \leq n,$$

[8]This extrinsic discussion is included mainly for motivation; it can be skipped by purist or hurried readers. The intrinsic theory is resumed after the statement of Lemma 35.3.

where $P_j^i(x)$ denotes the projection[9] of the j-th unit vector in R^n onto the tangent space T_xS. Note that we must use FS rather than Itô integrals to ensure that the solution $\hat{X}$ will stay in the manifold. It is suggestive to use the shorthand notation

$$\begin{aligned} d\hat{X}_t &= P_T(\hat{X}_t) \circ dX_t \\ &= P_T(\hat{X}_t)\, dX_t - \tfrac{1}{2}\, d\Big[Q_T(\hat{X}), X\Big]_t, \end{aligned}$$

where $P_T(x)$ and $Q_T(x)$ denote the complementary projections onto the tangent space T_x and its orthogonal complement $T_x^\perp$. Here the second expression is the Itô form of the same projection, which shows that if X is a local martingale in R^n, then $\hat{X}$ is a martingale in S. The two covariance processes are related by

$$\begin{aligned} d[\hat{X}, \hat{X}]_t &= d\Big[P_T(\hat{X}) \cdot X,\, P_T(\hat{X}) \cdot X\Big]_t \\ &= P_T(\hat{X}_t)\, d[X, X]_t\, P_T(\hat{X}_t) \\ &= P_T^2(\hat{X}_t)\, d[X, X]_t, \end{aligned}$$

which exhibits the covariance process $\hat{C}_t = [\hat{X}, \hat{X}]_t$ as the tangential projection of the process $C_t = [X, X]_t$. It is suggestive to write the latter relation in the form $[\hat{X}, \hat{X}] = [X, X]_S$, or simply as $\hat{C} = C_S$.

When X is a Brownian motion in R^n, we have $C_t = tI$ for the identity matrix I in R^n, and so

$$[\hat{X}, \hat{X}]_t = P_T^2(\hat{X}) I \cdot \lambda,$$

or simply $d\hat{C}_t = I_S\, dt$, where the projection I_S may be regarded as the identity matrix in S. Here $\hat{X}$ is called a *Brownian motion* in S. More generally, if X is an isotropic local martingale in R^n, so that $d[X, X]_t = d[X]_t I$ for a non-decreasing *rate process* $[X]$ in R_+, the projection $\hat{X}$ is again *isotropic*, in the sense that $d\hat{C} = d[X]_t I_S$. The latter process may clearly be time-changed into a Brownian motion, just as in the Euclidean case of Theorem 19.4.

Our somewhat informal, extrinsic discussion is summarized below. The stated results, of some independent interest, will justify our intrinsic definitions of drift and diffusion rates in Theorem 35.15.

Lemma 35.3 *(projection onto sub-manifold) Let X be a semi-martingale in R^n, with projection $\hat{X}$ onto a smooth sub-manifold $S \subset \mathsf{R}^n$. Then*

(i) *$\hat{X}$ is a semi-martingale in S,*
(ii) *X is a local martingale in R^n $\Rightarrow$ $\hat{X}$ is a martingale in S,*
(iii) *X is an isotropic local martingale in R^n $\Rightarrow$ $\hat{X}$ is isotropic in S,*
(iv) *X is a Brownian motion in R^n $\Rightarrow$ $\hat{X}$ is a Brownian motion S.*

— — —

Returning to the intrinsic theory, we define martingales in a smooth manifold S in terms of a *connection*, defined as a linear map ∇ of smooth functions f on S into symmetric, bilinear forms ∇f, satisfying[10]

$$\nabla f^2 = 2\, f\, \nabla f + 2\, (df)^{\otimes 2}, \quad f \in \mathcal{S}. \tag{4}$$

[9]To be precise, we need to extend $P(x)$ to a neighborhood of S, since the equation must first be stated, before we can argue that the solution process $\hat{X}$ stays in S. This is similar to the extension needed in Theorem 35.8 below.

[10]Connections are sometimes called *Hessians* and denoted by $\nabla = \text{Hess}$.

Given a connection ∇, we define an associated *martingale* in S as a semi-martingale X satisfying

$$f(X) \overset{m}{=} \tfrac{1}{2}\nabla f[X], \quad f \in \mathcal{S}, \tag{5}$$

where $X \overset{m}{=} Y$ in R means that $X - Y$ is a local martingale.

For motivation, we state a couple of elementary results. Further justifications are given by Theorems 35.8 and 35.16−35.17 below, which show how the intrinsic martingales and projection properties agree with the corresponding extrinsic notions and results when S is a smooth sub-manifold of R^n.

Lemma 35.4 *(connections and martingales)*

(i) *A continuous local martingale X in R^n satisfies* (4)−(5) *with*

$$(\nabla f)_{ij} = \partial^2_{ij} f, \quad i, j \le n, \;\; f \in \mathcal{S}.$$

(ii) *For a smooth manifold S, let ∇ be a linear map from $\mathcal{S}$ to the class of symmetric, bilinear maps on S. Then for any semi-martingale X in S satisfying* (5)*,*

$$\nabla f^2[X] = 2\, f\, \nabla f[X] + 2\,(df)^{\otimes 2}[X] \;\; a.s., \quad f \in \mathcal{S}.$$

Proof: (i) Equation (4) is obtained by repeated use of the chain rule for differentiation, whereas (5) holds by Itô's formula.

(ii) Applying (5) to both f and f^2 and using Theorem 35.1 (i)−(ii), along with Theorems 18.11 and 18.16 for real semi-martingales, we get

$$\begin{aligned}\tfrac{1}{2}\nabla f^2[X] &\overset{m}{=} (f \circ X)^2 \\ &= 2\, f(X) \cdot (f \circ X) + [f \circ X] \\ &\overset{m}{=} f(X) \cdot \nabla f[X] + (df)^{\otimes 2}[X] \\ &= f\, \nabla f[X] + (df)^{\otimes 2}[X].\end{aligned}$$

Since both sides have locally finite variation and start at 0, the assertion follows by Proposition 18.2. □

The defining property (4) of connections extends to a general chain rule:

Theorem 35.5 *(chain rule) Let S be a smooth manifold with connection ∇. Then for any smooth functions $f: S \to \mathsf{R}^n$ and $\varphi: \mathsf{R}^n \to \mathsf{R}$,*

$$\nabla(\varphi \circ f) = \big(\partial_i \varphi \circ f\big)\nabla f^i + \big(\partial^2_{ij}\varphi \circ f\big)\big(df^i \otimes df^j\big).$$

Proof: For any vector fields u, v, define a linear operator w on $\mathcal{S}$ by

$$wf = uvf - \nabla f(u, v), \quad f \in \mathcal{S}. \tag{6}$$

Using (4), (6), and the product rule for u, v, we get

$$\begin{aligned}\tfrac{1}{2}\,wf^2 &= \tfrac{1}{2}\,uvf^2 - \tfrac{1}{2}\,\nabla f^2(u,v)\\ &= u(f\,vf) - f\,\nabla f(u,v) - (df)^{\otimes 2}(u,v)\\ &= f\,uvf + uf\,vf - f\,\nabla f(u,v) - uf\,vf\\ &= f\big(uvf - \nabla f(u,v)\big) = f\,wf,\end{aligned}$$

and so Lemma A7.1 shows that w is a smooth vector field. Using (6) and the chain rule for u, v, w, we obtain

$$\begin{aligned}\nabla(\varphi\circ f)(u,v) &= uv(\varphi\circ f) - w(\varphi\circ f)\\ &= u\Big\{(\partial_i\varphi\circ f)\,vf^i\Big\} - (\partial_i\varphi\circ f)\,wf^i\\ &= \Big(\partial^2_{ij}\varphi\circ f\Big)(uf^j)(vf^i) + (\partial_i\varphi\circ f)\Big(uvf^i - wf^i\Big)\\ &= \Big(\partial^2_{ij}\varphi\circ f\Big)\Big(df^i\otimes df^j\Big)(u,v) - (\partial_i\varphi\circ f)\,\nabla f^i(u,v),\end{aligned}$$

and the assertion follows since u, v were arbitrary. □

In local coordinates, a connection on S is simply a special second order differential operator.

Theorem 35.6 *(connections as differential operators) For a smooth manifold S, let ∇ be a linear map of any $f\in\mathcal{S}$ into a symmetric, bilinear form ∇f. Then these conditions are equivalent:*

(i) *∇ is a connection on S,*

(ii) *in local coordinates there exist some smooth functions $\Gamma^k_{ij}=\Gamma^k_{ji}$, $i,j,k\le n$, such that*
$$(\nabla f)_{ij} = \Big(\partial^2_{ij} - \Gamma^k_{ij}\,\partial_k\Big)f,\quad f\in\mathcal{S}.$$

The functions Γ^k_{ij}, known as the *Christoffel symbols* of ∇, depend in a subtle way[11] on the choice of local coordinates. When $S=\mathsf{R}^n$ with Cartesian coordinates, we may choose $\Gamma^k_{ij}\equiv 0$ to form the *flat connection* ∇_n, first encountered in Lemma 35.4 (i).

Proof, (i) ⇒ (ii): Since the operator ∇ is local by (6), it is enough to consider the representation of ∇f on a local chart $U\subset S$. Here any smooth function $f: U\to\mathsf{R}$ has a representation $f=\tilde f\circ\varphi$, for a smooth function $\tilde f:\mathsf{R}^n\to\mathsf{R}$ and some local coordinates $\varphi^1,\dots,\varphi^n$ on U. The bilinear forms $\nabla\varphi^k$ may be written uniquely[12] as

$$\nabla\varphi^k = -\Gamma^k_{ij}\Big(d\varphi^i\otimes d\varphi^j\Big),\quad i,j,k\le n, \tag{7}$$

[11] In particular, they do not transform as tensors.

[12] The minus sign is traditional in geometry and has no special significance for us.

for some smooth functions Γ^k_{ij}. By Theorem 35.5,

$$\begin{aligned}\nabla f &= \nabla(\tilde f\circ\varphi) = \big(\partial_k \tilde f\circ\varphi\big)\,\nabla\varphi^k + \big(\partial^2_{ij}\tilde f\circ\varphi\big)\big(d\varphi^i\otimes d\varphi^j\big)\\ &= \Big\{\partial^2_{ij}\tilde f\circ\varphi - \Gamma^k_{ij}\big(\partial_k\tilde f\circ\varphi\big)\Big\}\big(d\varphi^i\otimes d\varphi^j\big)\\ &= \Big\{\big(\partial^2_{ij}\tilde f - \Gamma^k_{ij}\partial_k\tilde f\big)\circ\varphi\Big\}\big(d\varphi^i\otimes d\varphi^j\big),\end{aligned}$$

which proves the asserted relation in the more precise form

$$(\nabla f)_{ij} = \big(\partial^2_{ij} - \Gamma^k_{ij}\partial_k\big)\tilde f\circ\varphi.$$

(ii) $\Rightarrow$ (i): In Lemma 35.4 (i) we saw that the operator $(\nabla_n f)_{ij} = \partial^2_{ij} f$ on $\mathcal{S}$ satisfies (4). It remains to note that $\partial_k f^2 = 2f\,\partial_k f$ for all $k\le n$, by the chain rule for differentiation. □

This yields a simple description of martingales in local coordinates.

Theorem 35.7 *(martingale criteria) Let X be a semi-martingale in a smooth manifold S with connection ∇, fix some local coordinates $f^1,\dots,f^n$ in S, and put $X^i = f^i(X)$. Then these conditions are equivalent:*

(i) *X is a ∇-martingale,*

(ii) $f^i(X) \overset{m}{=} \frac12\,\nabla f^i[X],\ i\le n,$

(iii) $X^k \overset{m}{=} -\frac12\,\Gamma^k_{ij}(X)\cdot[X^i,X^j],\ k\le n.$

Proof, (i) $\Rightarrow$ (ii): Use (5).

(ii) $\Leftrightarrow$ (iii): By Theorems 35.1 (i) and 35.6, we have

$$\begin{aligned}\nabla f^k[X] &= -\Gamma^k_{ij}(X)\cdot\big(df^i\otimes df^j\big)[X]\\ &= -\Gamma^k_{ij}(X)\cdot\big[f^i\circ X,\ f^j\circ X\big]\\ &= -\Gamma^k_{ij}(X)\cdot[X^i,X^j].\end{aligned}$$

(ii) $\Rightarrow$ (i): By Itô's formula, condition (ii), and Theorems 35.1 (i)−(ii) and 35.5, we get for any smooth function φ on R^n

$$\begin{aligned}2\,\varphi\circ f(X) &\overset{m}{=} 2\,(\partial_i\varphi\circ f)(X)\cdot X^i + \big(\partial^2_{ij}\varphi\circ f\big)(X)\cdot[X^i,X^j]\\ &\overset{m}{=} (\partial_i\varphi\circ f)(X)\cdot\nabla f^i[X] + \big(\partial^2_{ij}\varphi\circ f\big)(X)\cdot\big(df^i\otimes df^j\big)[X]\\ &= (\partial_i\varphi\circ f)\nabla f^i[X] + \big(\partial^2_{ij}\varphi\circ f\big)\big(df^i\otimes df^j\big)[X]\\ &= \nabla(\varphi\circ f)[X].\end{aligned}$$

Hence, (5) remains true for $\varphi\circ f$, and (i) follows. □

For any local martingales M^i in R^n, Theorem 35.7 (iii) defines an SDE in X, possessing a unique solution in U up to an explosion time. This yields a large class of martingales associated with every connection.

The flat connection ∇_n on R^n induces a connection ∇_S on every smooth sub-manifold S. We proceed to describe ∇_S and characterize the associated martingales. Note the consistency with the extrinsic results in Lemma 35.3.

Theorem 35.8 *(induced connection on sub-manifold) For a smooth sub-manifold $S \subset \mathsf{R}^n$, let U be a neighborhood of S with a unique, smooth orthogonal projection π_S onto S, and write φ for the embedding $S \to \mathsf{R}^n$. Then*

(i) *∇_n induces a connection ∇_S on S, given by*

$$\nabla_S f = \varphi^{*2}\, \nabla_n (f \circ \pi_S), \quad f \in \mathcal{S}_S,$$

(ii) *a semi-martingale X in S is a ∇_S-martingale, iff $X \overset{m}{=} H \cdot C$ in R^n for a non-decreasing, continuous process C and a bounded, measurable process H in R^n satisfying*

$$H_t \perp T_{X_t}, \quad t \geq 0.$$

Proof: (i) To see that ∇_S is a connection, fix any $f \in \mathcal{S}_S$ with an orthogonal extension $\tilde{f}$ to U satisfying $\tilde{f} = f \circ \pi_S$, and put $\bar{u} = \varphi \circ u$ and $\bar{v} = \varphi \circ v$. Then

$$\begin{aligned} \tfrac{1}{2}\, \nabla_S f^2(u,v) &= \tfrac{1}{2}\, \nabla_n \tilde{f}^2(\bar{u},\bar{v}) \\ &= \tilde{f}\, \nabla_n \tilde{f}(\bar{u},\bar{v}) + (\bar{u}\tilde{f})(\bar{v}\tilde{f}) \\ &= f\, \nabla_S f(u,v) + (uf)(vf). \end{aligned}$$

(ii) For an S-valued semi-martingale $X \overset{m}{=} A$ in R^n, we get by Itô's formula, Theorems 35.1 (iii) and 35.6, and part (i)

$$\begin{aligned} f(X) = \tilde{f}(X) &\overset{m}{=} \partial_i \tilde{f}(X) \cdot X^i + \tfrac{1}{2}\, \partial^2_{ij} \tilde{f}(X) \cdot [X^i, X^j] \\ &\overset{m}{=} \operatorname{grad} \tilde{f}(X) \cdot A + \tfrac{1}{2}\, \varphi^{*2}\, \nabla_n \tilde{f}[X] \\ &= \operatorname{grad} \tilde{f}(X) \cdot A + \tfrac{1}{2}\, \nabla_S f[X]. \end{aligned}$$

Hence, X is a ∇_S-martingale iff $\operatorname{grad} \tilde{f} \cdot A = 0$ a.s. for all f, which clearly holds under the stated condition on A.

Conversely, let $\operatorname{grad} \tilde{f}(X) \cdot A = 0$ a.s. for all $f \in \mathcal{S}_S$. Writing $C = \sum_i \int |dA^i|$, we get $A = \sum_i H^i \cdot C$ for some bounded, predictable processes H^i, and f being arbitrary, we obtain $H^i_t \perp T_{X_t} S$ for C-a.e. t. Since modifying the H^i on a C-null set will not affect the sum $\sum_i H^i \cdot C$, we may assume that the stated orthogonality holds identically. □

A smooth mapping φ between two manifolds S, S' with connections ∇_S, $\nabla_{S'}$ is said to be *affine* if it commutes with the two connections, in the sense that

$$\nabla_S (f \circ \varphi) = \varphi^{*2} (\nabla_{S'} f), \quad f \in \mathcal{S}_{S'}. \tag{8}$$

In particular, a function $f \in \mathcal{S}$ is affine if it is so as a mapping from (S, ∇) into R with the flat connection, so that $\nabla f = 0$.

A *geodesic* in (S, ∇) is defined as an affine mapping $\gamma\colon I \to S$, for a real interval I with the flat connection. It describes the *motion* of a particle along a path in S, not to be confused with the path itself. We may characterize geodesics by a differential equation, here given both globally and in local coordinates.

Lemma 35.9 *(geodesics) For a manifold S with connection ∇ and an interval $I \subset \mathsf{R}$, consider a smooth map $\gamma: I \to S$. Then γ is a geodesic, iff for every $t \in I$,*

(i) $\partial^2(f \circ \gamma_t) = \nabla f(\dot\gamma_t, \dot\gamma_t), \quad f \in \mathcal{S}$,

(ii) $\ddot\gamma_t^k = -\left(\Gamma_{ij}^k \circ \gamma_t\right) \dot\gamma_t^i \dot\gamma_t^j, \quad i, j, k \le n.$

Proof: (i) Apply (8) to pairs of tangent vectors $\partial = d/dt$ in $T_t I$.

(ii) Apply (i) to the coordinate functions on a local chart. □

Affine maps can be characterized in terms of both geodesics and martingales. Note the striking similarity between the two cases.

Theorem 35.10 *(affine maps) Let φ be a smooth map between two manifolds S, S' with connections $\nabla_S, \nabla_{S'}$. Then these conditions are equivalent:*

(i) *φ is affine,*

(ii) *γ is a geodesic in S $\Rightarrow$ $\varphi \circ \gamma$ is a geodesic in S',*

(iii) *M is a martingale in S $\Rightarrow$ $\varphi \circ M$ is a martingale in S'.*

Proof, (i) ⇒ (iii): Assume (i), and let M be a martingale in S. For any smooth function f on S', we get by (1), (5), (8), and Theorem 35.1 (iv)

$$\begin{aligned} 2\, f \circ \varphi \circ M &\overset{m}{=} \nabla_S(f \circ \varphi)[M] \\ &= \varphi^{*2}(\nabla_{S'} f)[M] \\ &= \nabla_{S'} f[\varphi \circ M], \end{aligned}$$

which shows that $\varphi(M)$ is a martingale in S', proving (iii).

(iii) ⇒ (ii): Assume (iii), let $\gamma: I \to S$ be a geodesic in S, and put $\psi = \varphi \circ \gamma$. Consider a Brownian motion B in I, starting at $t = 0$ and stopped at some points $r \pm \varepsilon \in I$. Since γ is affine, we see as before that $\varphi \circ B$ is a martingale in S, and so by (iii) the process $\psi \circ B$ is a martingale in S'. Applying (5) to the martingales B and $\psi \circ B$ and using Theorem 35.1 (iii), we get for any f on S'

$$\begin{aligned} \nabla_I(f \circ \psi)[B] &\overset{m}{=} f \circ \psi(B) \\ &\overset{m}{=} \nabla_{S'}[\psi \circ B] \\ &= \psi^{*2}\, \nabla_{S'}[B]. \end{aligned}$$

Since both sides have locally finite variation and vanish at 0, they agree a.s., and so up to an optional time $\tau > 0$,

$$\int_0^{t \wedge \tau} (f \circ \psi)''(B_s)\, ds = \int_0^{t \wedge \tau} \nabla_{S'} f\left[\dot\psi \circ B_s\right] ds.$$

By the continuity of the two integrands, we get at $s = 0$

$$(f \circ \psi)''(r) = \nabla_{S'} f(\dot\psi_r, \dot\psi_r),$$

and since $r \in I$ was arbitrary, we conclude from Lemma 35.9 that ψ is a geodesic, proving (ii).

(ii) $\Rightarrow$ (i): For any $u \in T_S$, Lemma 35.9 yields a geodesic γ in S with $\dot\gamma(0) = u$, and so by (ii) the map $\psi = \varphi \circ \gamma$ is a geodesic in S'. Since γ has speed u at 0, we further note that ψ has speed $\varphi(u)$ at 0. Using (1) and applying Lemma 35.9 to both φ and ψ, we get for any smooth function f on S'

$$\begin{aligned}\nabla_S(f\circ\varphi)(u,u) &= \nabla_S(f\circ\varphi)(\dot\gamma_0,\dot\gamma_0)\\ &= (f\circ\varphi\circ\gamma)''(0)\\ &= \nabla_{S'}f(\dot\psi_0,\dot\psi_0)\\ &= \nabla_{S'}f(\varphi\circ u,\varphi\circ u)\\ &= \varphi^{*2}(\nabla_{S'}f)(u,u).\end{aligned}$$

Then by symmetry the two bilinear forms agree, and (8) follows. □

Corollary 35.11 *(uniqueness) A connection ∇ on a smooth manifold S is determined by each of the following:*

(i) *the class of all ∇-geodesics,*

(ii) *the class of all ∇-martingales.*

Proof: Let ∇, ∇' be connections on S with the same set of associated geodesics or martingales. Then Theorem 35.10 shows that the identity map $\varphi\colon (S,\nabla)\to(S,\nabla')$ is affine, and so for any smooth function f on S,

$$\begin{aligned}\nabla f &= \nabla(f\circ\varphi)\\ &= \varphi^{*2}(\nabla' f) = \nabla' f,\end{aligned}$$

which means that $\nabla = \nabla'$. □

A smooth function f on S is said to be *convex* if $\nabla f(u,u) \ge 0$ for all $u \in T_S$. In local coordinates, this means that the symmetric matrix $(\nabla f)_{ij}$ is non-negative definite. We begin with some elementary properties.

Lemma 35.12 *(affine and convex maps) Let S, S' be smooth manifolds with connections $\nabla_S, \nabla_{S'}$. Then for smooth functions f and φ,*

(i) *f is affine on S $\Leftrightarrow$ $f, -f$ are convex,*

(ii) $\left.\begin{array}{c}\varphi\colon S\to S' \textit{ is affine}\\ f \textit{ is convex on } S'\end{array}\right\} \Rightarrow f\circ\varphi$ *is convex on* S,

(iii) $\left.\begin{array}{c}f \textit{ is convex on } S\\ h \textit{ is convex, increasing on } \mathsf{R}\end{array}\right\} \Rightarrow h\circ f$ *is convex on* S.

Proof: (i) If both f and $-f$ are convex, then $\nabla f(u,u) = 0$ for all $u \in T_S$, and so by symmetry $\nabla f = 0$.

(ii) Using the definitions of affinity and convexity along with (1), we get for any $u \in T_S$

$$\begin{aligned}\nabla_S (f \circ \varphi)(u,u) &= \left(\varphi^{*2} \nabla_{S'} f\right)(u,u) \\ &= (\nabla_{S'} f)(\varphi \circ u, \varphi \circ u) \geq 0,\end{aligned}$$

which shows that $f \circ \varphi$ is convex on S.

(iii) Using Lemma 35.5 and the conditions on f and h, we get for any $u \in T_S$

$$\nabla_S (h \circ f)(u,u) = (h' \circ f) \nabla_S f(u,u) + (h'' \circ f)(uf)^2 \geq 0,$$

which shows that $h \circ f$ is convex on S. □

We may next characterize convexity in terms of geodesics or martingales.

Theorem 35.13 *(convex maps) Let f be a smooth function on a manifold S with connection ∇. Then these conditions are equivalent:*

(i) *f is convex on S,*

(ii) *γ is a geodesic in S $\Rightarrow$ $f \circ \gamma$ is convex on R,*

(iii) *M is a martingale in S $\Rightarrow$ $f \circ M$ is a local sub-martingale in R.*

Proof, (i) $\Rightarrow$ (iii): If f be convex on S and M be a martingale in S, then by (5) and Theorem 35.1 (iv),

$$f(M) \overset{m}{=} \tfrac{1}{2} \nabla f[M] \geq 0,$$

which shows that $f \circ M$ is a local sub-martingale, proving (iii).

(iii) $\Rightarrow$ (ii): Fix a geodesic $\gamma \colon I \to S$, and let M be a martingale on I. Then $\gamma \circ M$ is a martingale in S by Theorem 35.10, and so by (iii) $f \circ \gamma(M)$ is a local sub-martingale on I. Since M was arbitrary, a converse of Jensen's inequality in Lemma 4.5 shows that $f \circ \gamma$ is convex, proving (ii).

(ii) $\Rightarrow$ (i): For any $u \in T_S$ there exists a geodesic γ in S with $\dot\gamma_0 = u$. Then $f \circ \gamma$ is convex by (ii), and Lemma 35.9 (i) yields

$$\begin{aligned}\nabla f(u,u) &= \nabla f(\dot\gamma_0, \dot\gamma_0) \\ &= (f \circ \gamma)''(0) \geq 0,\end{aligned}$$

which shows that f is convex, proving (i). □

Conversely, geodesics and martingales in S can be characterized in terms of convex functions.[13] Here we state a simplified local version.[14]

[13] Since there may not be enough affine functions, we need to use the convex functions as convenient substitutes.

[14] The corresponding global statement is false in general, since there may not be enough *globally* convex functions. Fortunately, there are plenty of *locally* convex functions.

Theorem 35.14 *(convexity criteria, Darling) Let S be a manifold with connection ∇. Then every point $x \in S$ has an open neighborhood U, such that*

(i) *a smooth curve $\gamma : I \to U$ is a geodesic iff*

$$f \text{ is convex on } U \;\Rightarrow\; f \circ \gamma \text{ is convex on } I,$$

(ii) *a semi-martingale X in U is a martingale iff*

$$f \text{ is convex on } U \;\Rightarrow\; f \circ X \text{ is a local sub-martingale.}$$

Proof: The necessity in parts (i) and (ii) is clear from Theorem 35.13. We omit the more difficult proof of the sufficiency.[15] □

We return to a more detailed study of semi-martingales X in S. Though X has no semi-martingale decomposition in S, due to the absence of any linear structure, its local characteristics[16] can still be defined intrinsically. By their local nature, it is then enough to consider semi-martingales in any local chart U. For any smooth function f and bilinear form b on S, we define the finite-variation processes A_f and C_b by

$$f(X) \overset{m}{=} \tfrac{1}{2}\, \nabla f[X] + A_f, \qquad b\,[X] = C_b. \tag{9}$$

To avoid some distracting technicalities, we assume[17] that $A_f \ll \lambda$ and $C_b \ll \lambda$ a.s., where λ denotes Lebesgue measure on R_+.

Theorem 35.15 *(local drift and diffusion rates) Let S be a smooth manifold with connection ∇, and consider a semi-martingale X in S, such that a.s. $A_f \ll \lambda$ and $C_b \ll \lambda$ for all f and b. Then*

(i) *there exists an a.s. unique random tangent vector[18] A on S, such that*

$$f(X) \overset{m}{=} \tfrac{1}{2}\, \nabla f[X] + \langle A, f\rangle \cdot \lambda, \quad f \in \mathcal{S},$$

(ii) *there exists an a.s. unique symmetric, second order, contra-variant random tensor[19] C on S, such that for any bilinear form b on S,*

$$b\,[X] = \langle C, b\,\rangle \cdot \lambda \;\text{ a.s.},$$

(iii) *X solves the local martingale problem*

$$f(X) \overset{m}{=} \left\{ \tfrac{1}{2}\, \langle C, \nabla f\rangle + \langle A, f\rangle \right\} \cdot \lambda, \quad f \in \mathcal{S}, \tag{10}$$

which in local coordinates becomes

$$f(X) \overset{m}{=} \left\{ \tfrac{1}{2}\, C^{ij} \left(\partial^2_{ij} - \Gamma^k_{ij}\, \partial_k \right) f + A^i\, \partial_i f \right\} \cdot \lambda.$$

---

[15] Detailed discussions appear in Emery (1989), pp. 42–46, and (1998/2013), pp. 55–59.

[16] Our local characteristics are totally different from those of Meyer (1982).

[17] More generally, if $A_f \ll \xi$ and $C_b \ll \xi$ a.s. for a diffuse random measure ξ on R_+, we may reduce to the case $\xi = \lambda$ by a random time change.

[18] In other words, A_t is a random element of T_{X_t} for every $t \ge 0$.

[19] Thus, C_t is a symmetric random element of $T^{\otimes 2}_{X_t}$ — the dual of the space of bilinear forms on S. Schwartz suggests we think of such objects as *second order tangent vectors*.

Proof: (i) Fix any $f^1, \ldots, f^n \in \mathcal{S}$ and a smooth function φ on R^n, write $Y^i = f^i(X)$ and $A_f = Af \cdot \lambda$, and put $f = (f^i)$ and $Y = (Y^i)$. Using Itô's formula, equation (9), Lemma 18.14, and Theorems 35.1 (i)−(ii) and 35.5, we get

$$\begin{aligned}
\varphi(Y) &\overset{m}{=} \partial_i \varphi(Y) \cdot Y^i + \tfrac{1}{2}\, \partial^2_{ij}\varphi(Y) \cdot [Y^i, Y^j] \\
&\overset{m}{=} \tfrac{1}{2}\, \partial_i \varphi(Y) \cdot \nabla f^i[X] + \partial_i\varphi(Y) \cdot (Af^i \cdot \lambda) \\
&\qquad + \tfrac{1}{2}\, \partial^2_{ij}\varphi(Y) \cdot \big(df^i \otimes df^j\big)[X] \\
&= \tfrac{1}{2}\, (\partial_i \varphi \circ f) \nabla f^i[X] + \partial_i \varphi(Y) A f^i \cdot \lambda \\
&\qquad + \tfrac{1}{2}\, \big(\partial^2_{ij}\varphi \circ f\big)\big(df^i \otimes df^j\big)[X] \\
&= \tfrac{1}{2}\, \nabla(\varphi \circ f)[X] + \partial_i \varphi(Y) A f^i \cdot \lambda.
\end{aligned}$$

We may also apply (9) directly to the function $\varphi \circ f$ to get

$$(\varphi \circ f)(X) \overset{m}{=} \tfrac{1}{2}\, \nabla(\varphi \circ f)[X] + A(\varphi \circ f) \cdot \lambda,$$

and so by comparison,

$$A(\varphi \circ f) = (\partial_i \varphi \circ f)(X) A f^i \;\text{ a.e.}$$

Hence, Lemma A7.1 shows that A is indeed a random tangent vector.

(ii) For any local coordinates $\varphi^1, \ldots, \varphi^n$, we define

$$\begin{aligned}
C^{ij} \cdot \lambda &= \big(d\varphi^i \otimes d\varphi^j\big)[X] \\
&= [\varphi^i \circ X,\, \varphi^j \circ X].
\end{aligned}$$

Letting b be a bilinear form with coordinate representation $b_{ij}(d\varphi^i \otimes d\varphi^j)$, we obtain the duality relation

$$\begin{aligned}
\langle C, b \rangle \cdot \lambda &= b_{ij}\, C^{ij} \cdot \lambda \\
&= b_{ij}\big(d\varphi^i \otimes d\varphi^j\big)[X] = b[X],
\end{aligned}$$

where C is the random operator $S \to T_S^{\otimes 2}$ with coordinate representation (C^{ij}). Thus, C is a.e. unique and independent of any choice of local coordinates.

(iii) Combine (i)−(ii), and use Theorem 35.5, along with the coordinate representation of A. □

We refer to the processes A and C above as the *intrinsic drift* and *diffusion rates* of X. Note in particular that X is a martingale iff $A = 0$ a.e. Claim (i) may be surprising, since $\nabla f[X]$ is linear in f while $f(X)$ is not.

In normal coordinates around a point $x \in S$, the random differential operator in (iii) reduces at x to the form

$$L_t = \tfrac{1}{2}\, C_t^{ij}\, \partial^2_{ij} + A_t^i\, \partial_i,$$

which may be thought of as the *characteristic differential operator*[20] of X. This suggests that we regard the pair (A, C) as the *random generator* of X.

[20]This relates our local characteristics to the Schwartz–Meyer theory of second order stochastic calculus.

Our definitions of intrinsic drift and diffusion rates are justified by the following results for semi-martingales in Euclidean sub-manifolds, which formally agree with our findings in the extrinsic case.[21]

Theorem 35.16 *(semi-martingales in sub-manifold) Let X be an S-valued semi-martingale in R^n, where S is a smooth sub-manifold with induced connection ∇. Then*

(i) *the local drift of X in S is the tangential projection of the drift in R^n,*

(ii) *the local diffusion rates of X in S agree with those in R^n.*

Proof: (i) Using the notation of Theorem 35.8, we may write condition (i) of Theorem 35.15 in the form

$$f(X) \overset{m}{=} \tfrac{1}{2}\,\nabla f[X] + \Big\langle \operatorname{grad} \tilde f(X) \,\Big|\, A \Big\rangle \cdot \lambda.$$

Now let $X \overset{m}{=} H \cdot \lambda + V \cdot \lambda$, where H and V denote the tangential and normal components of the Euclidean drift rate. Proceeding as in the proof of Theorem 35.8, we get

$$f(X) \overset{m}{=} \tfrac{1}{2}\,\nabla f[X] + \Big\langle \operatorname{grad} \tilde f(X) \,\Big|\, H \Big\rangle \cdot \lambda + \Big\langle \operatorname{grad} \tilde f(X) \,\Big|\, V \Big\rangle \cdot \lambda.$$

Here the last term vanishes, since $\operatorname{grad} \tilde f(x) \in T_x \perp V$, and so by comparison

$$\Big\langle \operatorname{grad} \tilde f(X) \,\Big|\, A \Big\rangle \cdot \lambda = \Big\langle \operatorname{grad} \tilde f(X) \,\Big|\, H \Big\rangle \cdot \lambda.$$

Since f was arbitrary, it follows that $A = H$ a.e., as asserted.

(ii) Writing φ for the embedding $S \to \mathsf{R}^n$, we get for any bilinear form $b = h(df \otimes dg)$ with $f, g, h \in \mathcal{S}$

$$\begin{aligned} \varphi^{*2} b &= (h \circ \varphi)\Big\{(\varphi^* df) \otimes (\varphi^* dg)\Big\} \\ &= (h \circ \varphi)\Big\{d(f \circ \varphi) \otimes d(g \circ \varphi)\Big\}, \end{aligned}$$

and so

$$\begin{aligned} \varphi^{*2} b(u,v) &= (h \circ \varphi)\, u(f \circ \varphi)\, v(g \circ \varphi) \\ &= h(uf)(vg) \\ &= b(u,v), \end{aligned}$$

which extends to $\varphi^{*2} b = b$ for general b. Hence, Theorem 35.1 (iii) gives

$$b\,[\varphi \circ X] = \varphi^{*2} b\,[X] = b\,[X],$$

as desired. □

For a further justification, we show how the local characteristics of a semi-martingale in R^n are transformed by projection. This again agrees formally with the extrinsic results in Lemma 35.3.

[21] This may be remarkable, since the definitions of drift and diffusion rates are totally different in the two cases. A similar remark applies to Corollary 35.17.

Corollary 35.17 *(projection onto sub-manifold) Let X be a semi-martingale in R^n with projection $\hat{X}$ onto a smooth sub-manifold S with induced connection ∇. Then*

(i) *$\hat{X}$ is a semi-martingale in S,*

(ii) *the local drift of $\hat{X}$ equals the tangential projection of the drift of X,*

(iii) *the local diffusion rates of $\hat{X}$ are tangential projections of the diffusion rates of X.*

Proof: Proceed as in Lemma 35.3 and use Theorem 35.16. □

The local rate processes A and C are said to be *autonomous*, if $A_t = a(X_t)$ and $C_t = c(X_t)$ for a deterministic vector field a and a second order contravariant tensor field c on S. In that case, the local martingale problem of Theorem 35.15 (iii) yields some more precise information about X. As before, we define a *diffusion* in S as a continuous strong Markov process.

Corollary 35.18 *(diffusions in manifold) Let X be a semi-martingale in a manifold S with connection ∇. Then*

(i) *if* (10) *has a unique solution X with autonomous rates $A_t = a(X_t)$ and $C_t = c(X_t)$, it defines a diffusion in S,*

(ii) *if the functions a, c in* (i) *are smooth, then* (10) *has a locally unique solution X, defining a Feller diffusion in S.*

Proof: Use Theorems 32.3 and 32.11 (i)−(ii). □

Part (iii) of Theorem 35.15 suggests that, if the function c is non-singular, we might remove the drift a by a suitable change of connection ∇, so that X becomes a martingale. We turn to a separate discussion of such drift-free diffusions.

Theorem 35.19 *(diffusive martingales) Let X be a semi-martingale in a manifold S with connection ∇, and let L be a non-negative, symmetric, linear*[22] *map of bilinear forms b on S into smooth functions Lb on S. Then these conditions are equivalent:*

(i) $f(X) \stackrel{m}{=} \frac{1}{2} L(\nabla f)(X) \cdot \lambda, \quad f \in \mathcal{S}$,

(ii) *X is a martingale with $b[X] = Lb\,(X) \cdot \lambda$,*

(iii) *X is a diffusion process in S with generator $\frac{1}{2} L\nabla$.*

Proof: For clarity and importance, we give a direct proof:

(ii) ⇒ (i): The two properties in (ii) yield for any $f \in \mathcal{S}$

$$\begin{aligned} 2\, f(X) &\stackrel{m}{=} \nabla f[X] \\ &= L(\nabla f)(X) \cdot \lambda, \end{aligned}$$

[22] so that in particular $L(fb) = f Lb$

proving (i).

(i) ⇒ (ii): By (4) and the linearity of L,

$$\begin{aligned} L(\nabla f^2) &= 2\,L\Big\{f\nabla f + (df)^{\otimes 2}\Big\} \\ &= 2f\,L(\nabla f) + 2\,L(df)^{\otimes 2}, \end{aligned}$$

and so by (i),

$$\begin{aligned} f^2 \circ X &\overset{m}{=} \tfrac{1}{2}\int L(\nabla f^2)(X_t)\,dt \\ &= \int f\,L(\nabla f)(X_t)\,dt + \int L(df)^{\otimes 2}(X_t)\,dt. \end{aligned}$$

On the other hand, we get by Itô's formula and (i)

$$\begin{aligned} f^2 \circ X &= 2\,f(X)\cdot(f\circ X) + [f\circ X] \\ &\overset{m}{=} \int f(X_t)\,L(\nabla f)(X_t)\,dt + [f\circ X]. \end{aligned}$$

Comparing the two expressions and using Theorem 35.1 (i), we obtain

$$\begin{aligned} (df)^{\otimes 2}[X] &= [f\circ X] \\ &= L(df)^{\otimes 2}(X)\cdot\lambda, \end{aligned}$$

which extends by polarization and Theorem 35.1 (ii) to

$$\begin{aligned} f(dg\otimes dh)[X] &= f(X)\cdot\Big\{L(dg\otimes dh)(X)\cdot\lambda\Big\} \\ &= fL(dg\otimes dh)(X)\cdot\lambda \\ &= L\Big\{f(dg\otimes dh)\Big\}(X)\cdot\lambda, \end{aligned}$$

and finally to $b[X] = Lb(X)\cdot\lambda$. Applying the latter formula to ∇f and using (i), we get

$$\begin{aligned} \nabla f[X] &= L(\nabla f)(X)\cdot\lambda \\ &\overset{m}{=} 2\,f(X), \end{aligned}$$

which proves (5), showing that X is a martingale.

(i)–(ii) ⇔ (iii): Use Corollary 35.18 and Lemma 17.21. □

A differentiable manifold S is said to be *Riemannian*, if it is endowed with a positive definite, bilinear form ρ, called the *metric tensor* on S. For any vectors $u, v \in T_S$, ρ determines an inner product $\rho(u,v) = \langle u\,|\,v\rangle$ with associated norm $\|\cdot\|$. For any smooth curve $\gamma: I \to S$, the associated *length* and *energy* are defined as integrals of $\|\dot\gamma_t\|$ and $\frac{1}{2}\|\dot\gamma_t\|^2$, respectively. Any semi-martingale X in S has the *Riemannian quadratic variation*[23] $\rho[X]$.

The Riemannian structure on S induces a Riemannian metric on any sub-manifold S', preserving the length and energy of any smooth curve, as well as the quadratic variation of any semi-martingale. This again agrees with the situation in the extrinsic case.

[23]The process $\rho[X]$ is sometimes written as $\int(\,dX, dX)$ or $\int\langle\,dX\,|\,dX\rangle$.

Lemma 35.20 *(induced metric on sub-manifold) For a Riemannian manifold (S, ρ_S), let $S' \subset S$ be a smooth sub-manifold with associated embedding $\varphi : S' \to S$. Then*

(i) *$\rho_{S'} = \varphi^{*2} \rho_S$ is a Riemannian metric on S',*

(ii) *for any smooth curve γ in S', the length $\|\dot\gamma\|$ and energy $\frac{1}{2}\|\dot\gamma\|^2$ agree under ρ_S and $\rho_{S'}$,*

(iii) *for any semi-martingale X in S', the quadratic variations $\rho[X]$ agree under ρ_S and $\rho_{S'}$.*

Proof: (i) For any $u \in T_x S'$, we have $(u \circ \varphi) f = u(f \circ \varphi) = uf$, and so $u \circ \varphi = u$, which implies $\rho_{S'}(u, v) = \rho_S(u, v)$.

(ii) From (i) we get $\|\dot\gamma_t\|_{S'} = \|\dot\gamma_t\|_S$ for any curve γ in S.

(iii) For any semi-martingale X in S', Theorem 35.1 yields

$$\begin{aligned} \rho_{S'}[X] &= \varphi^{*2} \rho_S[X] \\ &= \rho_S[\varphi \circ X] = \rho_S[X]. \end{aligned} \qquad \square$$

We have seen how a smooth sub-manifold S of R^n can be endowed in a natural way with both a connection ∇ and a Riemannian metric ρ. It is remarkable that ∇, and hence also the associated class of martingales, depend *intrinsically* only on the metric ρ, regardless of the way (S, ρ) is embedded[24] into R^n. For that reason, ∇ is often called the *canonical* or *Riemannian connection.*[25]

For a precise statement, note that in local coordinates, the Riemannian metric ρ on S is given by a symmetric, positive definite matrix ρ_{ij}, depending smoothly on $x \in S$. The associated inverse matrix is denoted by ρ^{ij}. The notions of *Lie derivative* $\mathcal{L}_u b$ and *gradient* $\operatorname{grad} f = \hat f$ are explained in Appendix 7. A coordinate neighborhood of a point $x \in S$ is said to be *normal*, if the local coordinates are ortho-normal at x, and every geodesic through x describes a uniform motion along a straight line.

Theorem 35.21 *(Riemannian connection) On a Riemannian manifold (S, ρ), conditions* (i)–(iii), (v) *are equivalent and determine a unique connection ∇:*

(i) *for every $x \in S$, we have $\Gamma_{ij}^k(x) = 0$ in normal coordinates around x,*

(ii) *in terms of Lie derivatives, ∇ is given by*

$$\nabla f = \tfrac{1}{2} \mathcal{L}_{\operatorname{grad} f}(\rho), \quad f \in \mathcal{S},$$

(iii) *in local coordinates, ∇ has Christoffel symbols*

$$\Gamma_{ij}^k = \tfrac{1}{2} \rho^{kl} \Big(\partial_i \rho_{lj} + \partial_j \rho_{il} - \partial_l \rho_{ij} \Big),$$

[24]By a deep theorem in differential geometry, every smooth Riemannian manifold can be isometrically embedded into a Euclidean space of suitable dimension. Though the embedding is far from unique, the induced connection is always the same.

[25]also known as the *Levi-Civita connection*

(iv) *for $S = \mathsf{R}^n$ with Euclidean metric ρ, ∇ reduces to the flat connection,*

(v) *when (S, ρ) is isometrically embedded into R^n, ∇ agrees with the induced connection ∇_S of Theorem 35.8.*

Proof: (i) The uniqueness is clear from Theorem 35.6. The existence follows from parts (ii)−(iii), if we can show that the constructed connection satisfies the condition in (i). This holds by (v) and Theorems 35.8 and 35.20, along with the fact that any local neighborhood $U \subset S$ can be isometrically embedded into R^n for a suitable n.

(ii) To show that ∇ is a connection, we note that $\langle \hat f \,|\, u\rangle = df(u) = uf$, and hence for any $u \in T_S$,

$$\begin{aligned}\langle\, \mathrm{grad}\, f^2 | \, u\rangle &= uf^2 = 2f(uf)\\ &= 2f\langle \hat f \,|\, u\rangle\\ &= \langle\, 2f(\hat f)\,|\, u\rangle,\end{aligned}$$

which shows that $\mathrm{grad}\, f^2 = 2f(\hat f)$. Using Corollary A7.5, we get

$$\begin{aligned}\nabla f^2 &= \tfrac12 \mathcal{L}_{\mathrm{grad}\, f^2}(\rho)\\ &= \tfrac12 \mathcal{L}_{2f(\hat f)}(\rho) = \mathcal{L}_{f(\hat f)}(\rho)\\ &= f\,\mathcal{L}_{\mathrm{grad}\, f}(\rho) + df \otimes \langle \hat f \,|\, \cdot\rangle + \langle \cdot \,|\, \hat f\,\rangle \otimes df\\ &= 2f\,\nabla f + 2\,(df)^{\otimes 2}.\end{aligned}$$

(iii) Writing $\hat f = \mathrm{grad}\, f$ and $\hat g = \mathrm{grad}\, g$, and using Theorem A7.4 (ii) and (iv), we get from (ii) for any vector field u

$$\begin{aligned}2\,\nabla f(u, \hat g) &= (\mathcal{L}_{\mathrm{grad}\, f}\,\rho)(u, \hat g)\\ &= \hat f\langle u \,|\, \hat g\rangle - \langle \mathcal{L}_{\mathrm{grad}\, f}\, u \,|\, \hat g\rangle - \langle u \,|\, \mathcal{L}_{\mathrm{grad}\, f}\, \hat g\rangle\\ &= \hat f(ug) - (\mathcal{L}_{\mathrm{grad}\, f}\, u)g - \langle \mathcal{L}_{\mathrm{grad}\, f}\, \hat g \,|\, u\rangle\\ &= u(\hat f g) - \Big\langle u \,\Big|\, [\hat f, \hat g]\Big\rangle.\end{aligned}$$

Interchanging f and g and noting that $[\hat f, \hat g] + [\hat g, \hat f\,] = 0$ and $\hat f g = \hat g f = \langle \hat f \,|\, \hat g\rangle$, we obtain

$$\nabla f(u, \hat g) + \nabla g(u, \hat f) = u\langle \hat f \,|\, \hat g\rangle.$$

Choosing $u = \hat h = \mathrm{grad}\, h$ and permuting f, g, and h, we get

$$2\,\nabla f(\hat g, \hat h) = \hat g\langle \hat f \,|\, \hat h\rangle + \hat h\langle \hat f \,|\, \hat g\rangle - \hat f\langle \hat g \,|\, \hat h\rangle.$$

Noting that every coordinate vector ∂_i in T_S is locally a gradient field, and inverting the mapping $\hat f = \mathrm{grad}\, f$ into $\partial_i f = \rho_{ij} \hat f^j$, we have by (7)

$$2\,\Gamma^i_{jk}\,\rho_{il} = \partial_j\,\rho_{lk} + \partial_k\,\rho_{jl} - \partial_l\,\rho_{jk}.$$

Now solve for Γ^i_{jk} by applying the inverse matrix (ρ^{il}) to both sides.

(iv) Here $\rho = I$ is constant, and (iii) yields $\Gamma^i_{jk} = 0$ for all i, j, k.

(v) Write $\varphi\colon S \to \mathsf{R}^n$ for the embedding map, let U be a neighborhood of S with a unique, smooth, orthogonal projection $\pi_S\colon U \to S$, and introduce the orthogonal extension[26] $f' = f \circ \pi_S$ of any smooth function f on S. Using the

[26] in Theorem 35.8 denoted by $\tilde f$

definitions of $\hat{f} = \text{grad}\, f$ in S and R^n, we get

$$\begin{aligned}\Big\langle \varphi\circ\hat{f}\,\Big|\,\varphi\circ u\Big\rangle_n &= \langle\,\hat{f}\,|\,u\rangle_S = df(u)\\ &= uf = (\varphi\circ u)f'\\ &= \langle \text{grad}\, f'\circ\varphi\,|\,\varphi\circ u\rangle_n,\end{aligned}$$

which implies $\text{grad}\, f'\circ\varphi = \varphi\circ\hat{f}$ on S. By the definitions of ρ, ∇, and ∇_S, we obtain

$$\begin{aligned}\nabla f &= \tfrac{1}{2}\,\mathcal{L}_{\text{grad}\, f}(\rho)\\ &= \tfrac{1}{2}\,\mathcal{L}_{\text{grad}\, f'\circ\varphi}(\varphi^{*2}I)\\ &= \varphi^{*2}\nabla_n f' = \nabla_S f.\end{aligned}$$

□

In a Riemannian manifold (S,ρ), we always choose ∇ to be the canonical connection. Given a bilinear form b on (S,ρ), we define

$$\text{tr}\, b = \rho^{ij}\, b_{ij} = \textstyle\sum_i b(e_i, e_i), \tag{11}$$

for any ortho-normal basis (e_i) in T_x. A martingale M in a Riemannian manifold (S,ρ) is said to be *isotropic* if

$$\|\hat{f}\| = \|\hat{g}\| \quad\Rightarrow\quad [f\circ M] = [\,g\circ M]\ \text{ a.s.}$$

We may then introduce the *rate process* $n^{-1}\rho[M]$ of M, where $n = \dim S$. If it is absolutely continuous, its density is called the (local) *rate* of M. An isotropic martingale with constant rate 1 is called a *Brownian motion*.

Theorem 35.22 *(isotropic martingales) Let M be a martingale in a Riemannian manifold (S,ρ) with* $\dim S = n$. *Then these conditions are equivalent:*

(i) *M is isotropic,*

(ii) $b\,[M] = \text{tr}\, b(M)\cdot T$ *a.s. for an increasing process T.*

In that case, we have a.s.

(iii) *$T = n^{-1}\rho[M]$, and $M = B\circ T$ for a Brownian motion B in S.*

Proof, (i) ⇒ (ii)–(iii): Let $h^1,\dots,h^n$ be normal coordinates around a point $x\in S$. Polarizing (i) and alternating the signs of the h^i, we get by Theorem 35.1 (i)

$$\begin{aligned}\big(dh^i\otimes dh^j\big)[M] &= \big[h^i\circ M,\, h^j\circ M\big]\\ &= \delta^{ij}\,[h^i\circ M],\end{aligned}$$

and so by Theorem 35.1 (ii),

$$\begin{aligned}b\,[M] &= b_{ij}(M)\,\delta^{ij}\cdot[h^1\circ M]\\ &= \textstyle\sum_i b_{ii}(M)\cdot[h^1\circ M]\\ &= n^{-1}\text{tr}\, b(M)\cdot\rho[M],\end{aligned}$$

proving (ii) with $T = n^{-1}\rho[M]$. The time-change representation now follows as in Theorem 19.4 by means of Theorem 18.24.

(ii) $\Rightarrow$ (i): Assuming (ii), we get for any $f \in \mathcal{S}$

$$\begin{aligned}[f \circ M] &= (df)^{\otimes 2}[M] \\ &= \mathrm{tr}(df)^{\otimes 2}(M) \cdot T \\ &= \|\hat{f}\|^2 \big(\mathrm{tr}\, e^{\otimes 2}\big)(M) \cdot T,\end{aligned}$$

proving (i). □

For the canonical connection ∇ on a Riemannian manifold (S, ρ), we define the *Laplacian*[27] Δ on S by

$$\Delta f(x) = \mathrm{tr}(\nabla f)(x), \quad f \in \mathcal{S}, \ x \in S. \tag{12}$$

We list some characterizations of Brownian motion:

Corollary 35.23 *(Brownian motion) For a semi-martingale X in a Riemannian manifold (S, ρ), these conditions are equivalent:*

(i) $f(X) \overset{m}{=} \frac{1}{2}\Delta f(X) \cdot \lambda$ *for all* $f \in \mathcal{S}$,

(ii) X *is a martingale with* $b[X] = \mathrm{tr}\, b(X) \cdot \lambda$,

(iii) X *is an isotropic martingale with constant rate* 1,

(iv) X *is a diffusion process in* S *with generator* $\frac{1}{2}\Delta$.

Proof, (ii) $\Leftrightarrow$ (iii): Use Theorem 35.22.

(i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iv): Use Theorem 35.19. □

If the function c in Corollary 35.18 is non-singular, we can change the Riemannian metric ρ to get the Brownian condition $b[X] = \mathrm{tr}\, b(X) \cdot \lambda$ fulfilled. Unfortunately, this may change the canonical connection ∇ as well, and so the resulting process might have a non-zero drift. The latter can often be removed by a suitable Girsanov-type transformation.

A mapping φ from a Riemannian manifold (S, ρ) to a manifold S' with connection ∇ is said to be *harmonic*, if

$$\Delta(f \circ \varphi) = \mathrm{tr}\Big(\varphi^{*2}\, \nabla_{S'} f\Big), \quad f \in \mathcal{S}'. \tag{13}$$

Theorem 35.24 *(harmonic maps) Let φ be a smooth map from a Riemannian manifold (S, ρ) to a manifold S' with connection ∇. Then these conditions are equivalent:*

(i) φ *is harmonic,*

(ii) B *is a Brownian motion in* $S \Rightarrow \varphi \circ B$ *is a martingale in* S',

(iii) M *is an isotropic martingale in* $S \Rightarrow \varphi \circ M$ *is a martingale in* S'.

[27]also called the *Laplace–Beltrami operator*

Proof, (i) ⇒ (iii): Let φ be harmonic, and let M be an isotropic martingale in S. Using (13) and Theorems 35.1 (iii) and 35.23 (i)−(ii), we get for any $f \in \mathcal{S}'$

$$\begin{aligned} 2\, f \circ \varphi(M) &\overset{m}{=} \Delta(f \circ \varphi)(M) \cdot \rho[M] \\ &= \mathrm{tr}\Big(\varphi^{*2} \nabla_{S'} f\Big)(M) \cdot \rho[M] \\ &= \Big(\varphi^{*2} \nabla_{S'} f\Big)[M] \\ &= \nabla_{S'} f[\varphi \circ M], \end{aligned}$$

which shows that $\varphi \circ M$ is a martingale in S'.

(iii) ⇒ (ii): Special case.

(ii) ⇒ (i): Assuming (ii), we get as before for any $f \in \mathcal{S}'$

$$\begin{aligned} \mathrm{tr}\Big(\varphi^{*2} \nabla_{S'} f\Big)(B) \cdot \lambda &= \Big(\varphi^{*2} \nabla_{S'} f\Big)[B] \\ &= \nabla_{S'} f[\varphi \circ B] \\ &\overset{m}{=} 2\, f \circ \varphi(B) \\ &\overset{m}{=} \Delta(f \circ \varphi)(B) \cdot \lambda. \end{aligned}$$

Since the extreme sides have locally finite variation, they agree a.s., and (13) follows by continuity, proving (i). □

A *Lie group* is defined as a differential manifold endowed with a smooth group structure. On any Lie group S, we may introduce a left-invariant Riemannian metric ρ, generated by an arbitrary inner product on the basic tangent space T_ι, where ι is the identity element of S. Since the group operation may not be commutative, we need to distinguish between left and right increments of a process X over an interval $[s, t]$, given by $X_s^{-1} X_t$ or $X_t\, X_s^{-1}$, respectively. In the Abelian case, both versions reduce to $X_t - X_s$.

As before, we define a *Lévy process* in S as a process with stationary, independent left increments. We proceed to characterize Lévy processes in terms of the local characteristics in Theorem 35.15. The result is a version for Lie groups of the elementary Theorem 11.5.

Theorem 35.25 *(Lévy processes in Lie groups) In a Lie group S with a left-invariant Riemannian metric ρ, let X be a semi-martingale with local rates A and C. Then these conditions are equivalent:*

(i) *X has stationary, independent left increments,*

(ii) *$A = a(X)$ and $C = c(X)$ for some smooth, left-invariant functions a, c on S.*

Proof, (ii) ⇒ (i): By Corollary 35.18 (ii), X is a Feller diffusion, whose distribution is uniquely determined by a and c. If the latter are left-invariant, Theorem 11.5 shows that X has stationary, independent left increments.

(i) ⇒ (ii): For any $s < t$, Proposition 18.17 and Theorem 35.1 show that the increment $b[X]_t - b[X]_s$ depends measurably on the values of X in the interval $[s, t]$. By stationarity of the increments, the corresponding mapping is left-invariant. Taking right derivatives in Theorem 2.15 and using the stationarity again, we conclude that C_t depends measurably on X in any right neighborhood $[t, t+\varepsilon)$, where the affecting map is again invariant. By the $0-1$ law in Corollary 9.26, C_t then depends measurably on X_t for every $t > 0$, and so $C_t = c(X_t)$ by Lemma 1.14 for a measurable and invariant function c.

To deal with A, we need to show that if $f(X)$ has a semi-martingale decomposition $M + V$, then $V_t - V_s$ depends measurably on X in the interval $[s, t]$. The required measurability then follows by approximation with sums of conditional expectations, as in the proof in Lemma 10.7, since the conditioning σ-field $\mathcal{F}_s$ may be replaced by conditioning on X_s, by independence of the increments. Since such a measurability is already known for the term $\frac{1}{2}\nabla f[X]$ in (9), it holds for Af as well. We may now proceed as before to see that $A = a(X_t)$ for some measurable and invariant function a.

Since a is a left-invariant vector field on S, it is automatically smooth[28]. A similar argument may yield the desired smoothness of the second order invariant vector field c. □

Every continuous process in S with stationary, independent left increments can be shown to be a semi-martingale of the stated form. The intrinsic drift rate a clearly depends on the connection ∇, which in turn depends on the metric ρ, via Theorem 35.21. However, ρ is not unique since it can be generated by any inner product on T_ι, and so there is no natural definition of Brownian motion in S. In fact, martingales and drift rates are not well-defined either, since even the canonical connection may fail to be unique[29].

Exercises

1. Show that when $S = \mathsf{R}^n$, a semi-martingale in S is just a continuous semi-martingale $X = (X^1, \ldots, X^n)$ in the usual sense. Further show how the covariance matrix $[X^i, X^j]$, $i, j \le n$, can be recovered from $b[X]$.

2. For any semi-martingales X, Y and bilinear form b on S, construct a process $b\,[X, Y]$ of locally finite variation, depending linearly on b, such that $(df \otimes dg)[X, Y] = [f(X), g(Y)]$ for smooth functions f, g, so that in particular $b\,[X, X] = b\,[X]$. (*Hint:* Regard (X, Y) as a semi-martingale in S^2, and define $b\,[X, Y] = (\pi_1 \otimes \pi_2)^* b\,[(X, Y)]$, where π_1, π_2 are the coordinate projections $S^2 \to S$, and $(\pi_1 \otimes \pi_2)^* b$ denotes the pull-back of b by $\pi_1 \otimes \pi_2$ into a bilinear form on S^2. Now check the desired formula.)

3. Explain where in the proof of Theorem 35.2 it is important that we use the FS rather than the Itô integral. Specify which of the four parts of the theorem depend on that convention.

---

[28] Cf. Proposition 3.7 in Warner (1983), p. 85.

[29] This is far from obvious; thanks to Ming Liao for answering my question.

4. Same requests as in the previous exercise for the extrinsic Lemma 35.3 with accompanying discussion. (*Hint:* Note that the reasons are very different.)

5. Show that the class of connections on S is *not* closed under addition or multiplication by scalars. (*Hint:* This is clear from (4), and even more obvious from Theorem 35.6.)

6. Show that for any connection ∇ with Christoffel symbols Γ_{ij}^k, the latter can be changed arbitrarily in a neighborhood of a point $x \in S$, such that the new functions $\tilde{\Gamma}_{ij}^k$ are Christoffel symbols of a connection $\tilde{\nabla}$ on S. (*Hint:* Extend the $\tilde{\Gamma}_{ij}^k$ to smooth functions agreeing with Γ_{ij}^k outside a larger neighborhood.)

7. Show that the class of martingales in S depends on the choice of connection. Thus, letting X be a (non-trivial) martingale for a connection ∇, show that we can choose another connection ∇', such that X fails to be a ∇'-martingale. (*Hint:* Use Theorem 35.7 together with the previous exercise.)

8. Compare the martingale notions in Lemma 35.3 and Theorem 35.8. Thus, check that, for a given embedding, the two notions agree. On the other hand, explain in what sense one notion is extrinsic while the other is intrinsic.

9. Show that the notion of geodesic depends on the choice of connection. Thus, given a geodesic γ for a connection ∇, show that we can choose another connection ∇' such that γ fails to be a ∇'-geodesic. (*Hint:* This might be seen most easily from Corollary 35.11.)

10. For a given connection ∇ on S, show that the classes of geodesics and martingales are invariant under a scaling of time. Thus, if γ is a geodesic in S, then so is the function $\gamma'_t = \gamma_{ct}$, $t \geq 0$, for any constant $c > 0$. Similarly, if M is a martingale in S, then so is the process $M'_t = M_{ct}$. (*Hint:* For geodesics, use Lemma 35.9. For martingales, we can use Theorem 35.7.)

11. When $S = \mathsf{R}^n$ with the flat connection, show that the notions of affine and convex functions agree with the elementary notions for Euclidean spaces. Further show that, in this case, a geodesic represents motion with constant speed along a straight line. For Euclidean spaces S, S' with flat connections, verify the properties and equivalences in Lemma 35.12 and Theorems 35.10 and 35.13.

12. Show that the intrinsic notion of drift depends on the choice of connection, whereas the intrinsic diffusion rate doesn't. Thus, show that if a semi-martingale X has drift A for connection ∇, we can choose another connection ∇' such that the associated drift is different. (*Hint:* Note that X is a martingale for ∇ iff $A = 0$. For the second assertion, note that $b\,[X]$ is independent of ∇.)

13. For $S = \mathsf{R}^n$ with the flat connection, explain how our intrinsic local characteristics are related to the classical notions for ordinary semi-martingales.

14. Show that in Theorem 35.15, we can replace Lebesgue measure λ by any diffuse measure on R, and even by a suitable random measure.

15. Clarify the relationship between the intrinsic results of Theorem 35.16 and Corollary 35.17 and the extrinsic ones in Lemma 35.3. (*Hint:* Since the definitions of local characteristics are totally different in the two cases, the formal agreement merely justifies our intrinsic definitions. Further note that the intrinsic properties hold for *any* smooth embedding of S, whereas the extrinsic ones refer to a specific choice of embedding.)

16. Show that for $S = \mathsf{R}^d$ with the Euclidean metric, the intrinsic notion of isotropic martingales agrees with the notion for continuous local martingales in Chapter 19. Prove a corresponding statement for Brownian motion.

17. Show that if X is a martingale in a Riemannian manifold (S, ρ), it may fail to be so for a different choice of metric ρ. Further show that if X is an isotropic martingale in (S, ρ), we can change ρ so that it remains a martingale but is no longer isotropic. Finally, if X is a Brownian motion in (S, ρ), we can change ρ so that it remains an isotropic martingale, though with a rate different from 1.

Appendices

1. Measurable maps, 2. General topology, 3. Linear spaces, 4. Linear operators, 5. Function and measure spaces, 6. Classes and spaces of sets, 7. Differential geometry

Here we review some basic topology, functional analysis, and differential geometry, needed at various places throughout the book. We further consider some special results from advanced measure theory, and discuss the topological properties of various spaces of functions, measures, and sets, of crucial importance for especially the weak convergence theory in Chapter 23. Proofs are included only when they are not easily accessible in the standard literature.

1. Measurable maps

The basic measure theory required for the development of modern probability was surveyed in Chapters 1–3. Here we add some less elementary facts, needed for special purposes.

If a measurable mapping is invertible, the measurability of its inverse can sometimes be inferred from the measurability of the range.

Theorem A1.1 *(range and inverse, Kuratowski) For any measurable bijection f between two Borel spaces S and T, the inverse $f^{-1}: T \to S$ is again measurable.*

Proof: See Parthasarathy (1967), Section I.3. □

We turn to the basic projection and section theorems, which play important roles in some more advanced contexts. Given a measurable space $(\Omega, \mathcal{F})$, we define the *universal completion* of $\mathcal{F}$ as the σ-field $\bar{\mathcal{F}} = \bigcap_\mu \mathcal{F}^\mu$, where $\mathcal{F}^\mu$ denotes the completion of $\mathcal{F}$ with respect to μ, and the intersection extends over all probability measures μ on $\mathcal{F}$. For any spaces Ω and S, we define the *projection* πA of a set $A \subset \Omega \times S$ onto Ω as the union $\bigcup_s A_s$, where

$$A_s = \big\{\omega \in \Omega;\ (\omega, s) \in A\big\}, \quad s \in S.$$

Theorem A1.2 *(projection and sections, Lusin, Choquet, Meyer) For any measurable space $(\Omega, \mathcal{F})$ and Borel space $(S, \mathcal{S})$, consider a set $A \in \mathcal{F} \otimes \mathcal{S}$ with projection πA onto Ω. Then*

(i) $\qquad \pi A \in \bar{\mathcal{F}} = \bigcap_\mu \mathcal{F}^\mu,$

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1

(ii) *for any probability measure P on $\mathcal{F}$, there exists a random element ξ in S, such that*

$$\{\omega, \xi(\omega)\} \in A, \quad \omega \in \pi A \ \text{ a.s. } P.$$

Proof: See Dellacherie & Meyer (1975), Section III.44. □

We note an application to functional representations needed in Chapter 28:

Corollary A1.3 *(conditioning representation) Consider a probability space $(\Omega, \mathcal{F}, P)$ with sub-σ-field $\mathcal{G} \subset \mathcal{F}$ and a measurable function $f: T \times U \to S$, where S, T, U are Borel. Let ξ, η be random elements in S and U, such that*

$$\mathcal{L}(\xi \,|\, \mathcal{G}) \in \Big\{\mathcal{L}\{f(t,\eta)\};\, t \in T\Big\} \ \text{ a.s.}$$

Then there exist a $\mathcal{G}$-measurable random element τ in T and a random element $\tilde{\eta}$ in U, such that

$$\eta \overset{d}{=} \tilde{\eta} \perp\!\!\!\perp \tau, \qquad \xi = f(\tau, \tilde{\eta}) \ \text{ a.s.}$$

Proof: Fix a version of the conditional distribution $\mathcal{L}(\xi \,|\, \mathcal{G})$. Applying Theorem A1.2 (ii) to the product-measurable set

$$A = \Big\{(\omega, t) \in \Omega \times T;\, \mathcal{L}(\xi \,|\, \mathcal{G})(\omega) = \mathcal{L}\{f(t,\eta)\}\Big\},$$

we obtain a $\mathcal{G}$-measurable random element τ in T satisfying

$$\begin{aligned} \mathcal{L}(\xi \,|\, \mathcal{G}) &= \mathcal{L}\Big\{f(t,\eta)\Big\}_{t=\tau} \\ &= \mathcal{L}(\xi \,|\, \tau) \ \text{ a.s.,} \end{aligned}$$

where the last equality holds by the chain rule for conditioning. Letting $\zeta \overset{d}{=} \eta$ in U with $\zeta \perp\!\!\!\perp \tau$ and using Lemma 4.11, we get a.s.

$$\begin{aligned} \mathcal{L}\Big\{f(\tau,\zeta)\Big|\, \tau\Big\} &= \mathcal{L}\Big\{f(t,\eta)\Big\}_{t=\tau} \\ &= \mathcal{L}(\xi \,|\, \tau), \end{aligned}$$

and so $\{f(\tau,\zeta), \tau\} \overset{d}{=} (\xi, \tau)$. Then Theorem 8.17 yields a random pair $(\tilde{\eta}, \tilde{\tau}) \overset{d}{=} (\zeta, \tau)$ in $U \times T$ with

$$\Big\{f(\tilde{\tau}, \tilde{\eta}), \tilde{\tau}\Big\} = (\xi, \tau) \ \text{ a.s.}$$

In particular $\tilde{\tau} = \tau$ a.s., and so $\xi = f(\tau, \tilde{\eta})$ a.s. Further note that $\tilde{\eta} \overset{d}{=} \zeta \overset{d}{=} \eta$, and that $\tilde{\eta} \perp\!\!\!\perp \tilde{\tau} = \tau$ since $\zeta \perp\!\!\!\perp \tau$ by construction. □

2. General topology

Elementary metric topology, including properties of open and closed sets, continuity, convergence, and compactness, are used throughout this book. For special purposes, we also need the slightly more subtle notions of separability and completeness, dense or nowhere dense sets, uniform and equi-continuity, and approximation. We assume the reader to be well familiar with most of this material. Here we summarize some basic ideas of general, non-metric topology, of special importance in Sections 5–6 below and Chapter 23. Proofs may be found in most textbooks on real analysis.

A *topology* on a space S is defined as a class $\mathcal{T}$ of subsets of S, called the *open sets*, such that $\mathcal{T}$ is closed under finite intersections and arbitrary unions and contains S and $\emptyset$. The complements of open sets are said to be *closed*. For any $A \subset S$, the *interior* A^o is the union of all open set contained in A. Similarly, the *closure* $\bar{A}$ is the intersection of all closed sets containing A, and the boundary ∂A is the difference $\bar{A} \setminus A^o$. A set A is *dense* in S if $\bar{A} = S$, and *nowhere dense* if $(\bar{A})^c$ is dense in S. The space S is *separable* if it has a countable, dense subset. It is *connected* if S and $\emptyset$ are the only sets that are both open and closed. For any $A \subset S$, the *relative topology* on A consists of all sets $A \cap G$ with $G \in \mathcal{T}$.

The family of topologies on S is partially ordered under set inclusion. The smallest topology on S has only the two sets S and $\emptyset$, whereas the largest one, the *discrete topology* 2^S, contains all subsets of S. For any class $\mathcal{C}$ of subsets of S, the *generated topology* is the smallest topology $\mathcal{T} \supset \mathcal{C}$. A topology $\mathcal{T}$ is said to be *metrizable*, if there exists a metric ρ on S, such that $\mathcal{T}$ is generated by all open ρ-balls. In particular, it is *Polish* if ρ can be chosen to be separable and complete.

An open set containing a point $x \in S$ is called a *neighborhood* of x. A class $\mathcal{B}_x$ of neighborhoods of x is called a *neighborhood base* at x, if every neighborhood of x contains a set in $\mathcal{B}_x$. A *base* $\mathcal{B}$ of $\mathcal{T}$ is a class containing a neighborhood base at every point. Every base clearly generates the topology. The space S is said to be *first countable*, if it has a countable neighborhood base at every point, and *second countable*, if it has a countable base. Every metric space is clearly first countable, and it is second countable iff it is separable. A topological space S is said to be *Hausdorff*, if any two distinct points have disjoint neighborhoods, and *normal* if all singleton sets are closed, and any two disjoint closed sets are contained in disjoint open sets. In particular, all metric spaces are normal.

A mapping f between two topological spaces S, S' is said to be *continuous*, if $f^{-1}B$ is open in S for every open set $B \subset S'$. The spaces S, S' are *homeomorphic*, if there exists a bijection (1−1 correspondence) $f\colon S \to S'$ such that f and f^{-1} are both continuous. For any family $\mathcal{F}$ of functions f_i mapping S into some topological spaces S_i, the *generated topology* on S is the smallest one making all the f_i continuous. In particular, if S is a Cartesian product of some topological spaces S_i, $i \in I$, so that the elements of S are functions $x = (x_i;\, i \in I)$, then the *product topology* on S is generated by all coordinate projections $\pi_i\colon x \mapsto x_i$. A set $A \subset S$ is said to be *compact*, if every cover of A by open sets contains a finite subcover, and *relatively compact* if its closure $\bar{A}$ is compact. By *Tychonov*'s theorem, any Cartesian product of compact spaces is again compact in the product topology. A topological space S is said to be *locally compact*, if it has a base consisting of sets with compact closure.

A set I is said to be *partially ordered* under a relation $\prec$, if

- $i \prec j \prec i \;\Rightarrow\; i = j$,
- $i \prec i$ for all $i \in I$,
- $i \prec j \prec k \;\Rightarrow\; i \prec k$.

It is said to be *directed* if also

- $i, j \in I \;\Rightarrow\; i \prec k$ and $j \prec k$ for some $k \in I$.

A *net* in a topological space S is a mapping of a directed set I into S. In particular, sequences are nets on $I = \mathsf{N}$ with the usual order. A net (x_i) is said to *converge* to a *limit* $x \in S$, written as $x_i \to x$, if for every neighborhood A of x there exists an $i \in I$, such that $x_j \in A$ for all $j \succ i$. Possible limits are unique when S is Hausdorff, but not in general. We further say that (x_i) *clusters* at x, if for any neighborhood A of x and $i \in I$, there exists a $j \succ i$ with $x_j \in A$. For sequences (x_n), this means that $x_n \to x$ along a sub-sequence.

Net convergence and clustering are constantly used[1] throughout analysis and probability, such as in continuous time and higher dimensions, or in the definitions of total variation and Riemann integrals. They also serve to characterize the topology itself:

Lemma A2.1 *(closed sets) For sets A in a topological space S, these conditions are equivalent:*

(i) *A is closed,*
(ii) *for every convergent net in A, even the limit lies in A,*
(iii) *for every net in A, all cluster points also lie in A.*

When S is a metric space, it suffices in (ii)−(iii) *to consider sequences.*

We can also use nets to characterize compactness:

Lemma A2.2 *(compact sets) For sets A in a topological space S, these conditions are equivalent:*

(i) *A is compact,*
(ii) *every net in A has at least one cluster point in A.*

If the net in (ii) *has exactly one cluster point x, it converges to x. When S is a metric space, it is enough in* (ii) *to consider sequences.*

For sequences, property (ii) above amounts to *sequential compactness*—the existence of a convergent sub-sequence with limit in A. For the notion of *relative sequential compactness*, the limit is not required to lie in A.

Given a compact topological space X, let C_X be the class of continuous functions $f : X \to \mathsf{R}$ equipped with the supremum norm $\|f\| = \sup_x |f(x)|$. An *algebra* $\mathcal{A} \subset C_X$ is defined as a linear subspace closed under multiplication. It is said to *separate points*, if for any $x \neq y$ in X there exists an $f \in \mathcal{A}$ with $f(x) \neq f(y)$.

[1] only the terminology may be less familiar

Theorem A2.3 *(Stone–Weierstrass) For a compact topological space X, let $\mathcal{A} \subset C_X$ be an algebra separating points and containing the constant function 1. Then $\bar{\mathcal{A}} = C_X$.*

Like many other results in this and subsequent sections, this statement is also true in a complex version.

Throughout this book we are making use of spaces that are locally compact, second countable, and Hausdorff, here referred to as *lcscH spaces*. We record some standard facts about such spaces:

Lemma A2.4 *(lcscH spaces) Let the space S be lcscH. Then*

(i) *S is Polish and σ-compact,*

(ii) *S has a separable, complete metrization, such that a set is relatively compact iff it is bounded,*

(iii) *if S is not already compact, it can be compactified by the attachment of a single point $\Delta \neq S$.*

3. Linear spaces

In this and the next section, we summarize some basic ideas of functional analysis, such as Hilbert and Banach spaces, and operators between linear spaces. Aspects of this material are needed in several chapters throughout the book. Proofs may be found in textbooks on real analysis.

A *linear space*[2] over R is a set X with two operations, addition $x + y$ and multiplication by scalars[3] cx, such that for any $x, y, z \in X$ and $a, b \in \mathsf{R}$,

- $x + (y + z) = (x + y) + z$,
- $x + y = y + x$,
- $x + 0 = x$ for some $0 \in X$,
- $x + (-x) = 0$ for some $-x \in X$,
- $a(bx) = (ab)x$,
- $(a + b)x = ax + bx$,
- $a(x + y) = ax + ay$,
- $1x = x$.

A *subspace* of X is a subset closed under the mentioned operations, hence a linear space in its own right. A function $\|x\| \geq 0$ on X is called a *norm*, if for any $x, y \in X$ and $c \in \mathsf{R}$,

[2] also called a *vector space*

[3] Here and below, we can also use the set C of complex numbers as our scalar field. Usually this requires only trivial modifications.

- $\|x\| = 0 \Leftrightarrow x = 0$,
- $\|cx\| = |c| \, \|x\|$,
- $\|x + y\| \le \|x\| + \|y\|$.

The norm generates a metric ρ on X, given by $\rho(x, y) = \|x - y\|$. If ρ is *complete*, in the sense that every Cauchy sequence converges, then X is called a *Banach space*. In that case, any closed subspace is again a Banach space.

Given a linear space X, we define a *linear functional* on X as a function $f: X \to \mathsf{R}$, such that

$$f(ax + by) = a\, f(x) + b\, f(y), \quad x, y \in X, \ a, b \in \mathsf{R}.$$

Similarly, a *convex functional* is defined as a map $p: X \to \mathsf{R}_+$ satisfying

- $p(cx) = c\, p(x), \quad x \in X, \, c \ge 0$,
- $p(x + y) \le p(x) + p(y), \quad x, y \in X$.

Lemma A3.1 *(Hahn–Banach) Let X be a linear space with subspace Y, fix a convex functional p on X, and let f be a linear functional on Y with $f \le p$ on Y. Then f extends to a linear functional on X, such that $f \le p$ on X.*

A set M in a linear space X is said to be *convex* if

$$c\, x + (1 - c)\, y \in M, \quad x, y \in M, \ c \in [0, 1].$$

A *hyper-plane* in X is determined by an equation $f(x) = a$, for a linear functional $f \ne 0$ and a constant a. We say that M is *supported* by the plane $f(x) = a$ if $f(x) \le a$, or $f(x) \ge a$, for all $x \in M$. The sets M, N are *separated* by the plane $f(x) = a$ if $f(x) \le a$ on M and $f(x) \ge a$ on N, or vice versa. For clarity, we state the following support and separation result for convex sets in a Euclidean space.

Corollary A3.2 *(convex sets)*

(i) *For any convex set $M \subset \mathsf{R}^d$ and point $x \in \partial M$, there exists a hyper-plane through x supporting M.*

(ii) *For any disjoint convex sets $M, N \subset \mathsf{R}^d$, there exists a hyper-plane separating M and N.*

When X is a normed linear space, we say that a linear functional f on X is *bounded* if $|f(x)| \le c\|x\|$ for some constant $c < \infty$. Then there is a smallest value of c, denoted by $\|f\|$, so that $|f(x)| \le \|f\| \, \|x\|$. The space of bounded linear functionals f on X is again a normed linear space (in fact a Banach space), called the *dual* of X and denoted by X^*. Continuing recursively, we may form the *second dual* $X^{**} = (X^*)^*$, etc.

Theorem A3.3 *(Hahn–Banach) Let X be a normed linear space with a subspace Y. Then for any $g \in Y^*$, there exists an $f \in X^*$ with $\|f\| = \|g\|$ and $f = g$ on Y.*

Corollary A3.4 *(natural embedding) For a normed linear space X, we may define a linear isometry $T: X \to X^{**}$ by*

$$(Tx)f = f(x), \quad x \in X, \; f \in X^*.$$

The space X is said to be *reflexive* if $T(X) = X^{**}$, so that X and X^{**} are isomorphic, written as $X \cong X^{**}$. The reflexive property extends to any closed linear subspace of X. Familiar examples of reflexive spaces include l^p and L^p for $p > 1$ (but not for $p = 1$), where the duals equal l^q and L^q, respectively, with $p^{-1} + q^{-1} = 1$. For $p = 2$, we get a Hilbert space H with the unique property $H \cong H^*$.

On a normed linear space X, we may consider not only the *norm topology* induced by $\|x\|$, but also the *weak topology* generated by the dual X^*. Thus, the weak topology on X^* is generated by the second dual X^{**}. Even more important is the *weak* topology* on X^*, generated by the *coordinate maps* $\pi_x: X^* \to \mathsf{R}$ given by

$$\pi_x f = f(x), \quad x \in X, \; f \in X^*.$$

Lemma A3.5 *(topologies on X^*) For a normed linear space X, write $\mathcal{T}_w^*$, $\mathcal{T}_w$, $\mathcal{T}_n$ for the weak*, weak, and norm topologies on X^*. Then*

(i) $\mathcal{T}_w^* \subset \mathcal{T}_w \subset \mathcal{T}_n$,

(ii) $\mathcal{T}_w^* = \mathcal{T}_w = \mathcal{T}_n$ *when X is reflexive.*

An *inner product* on a linear space X is a function $\langle x, y\rangle$ on X^2, such that for any $x, y, z \in X$ and $a, b \in \mathsf{R}$,

- $\langle x, x\rangle \geq 0$, with equality iff $x = 0$,
- $\langle x, y\rangle = \langle y, x\rangle$,
- $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$.

The function $\|x\| = \langle x, x\rangle^{1/2}$ is then a norm on X, satisfying the *Cauchy inequality*[4]

- $|\langle x, y\rangle| \leq \|x\| \, \|y\|, \quad x, y \in X.$

If the associated metric is complete, then X is called a *Hilbert space*[5]. If $\langle x, y\rangle = 0$, we say that x and y are *orthogonal* and write $x \perp y$. For any subset $A \subset H$, we write $x \perp A$ if $x \perp y$ for all $y \in A$. The *orthogonal complement* $A^\perp$, consisting of all $x \in H$ with $x \perp A$, is a closed linear subspace of H.

A linear functional f on a Hilbert space H is clearly continuous iff it is *bounded*, in the sense that

$$|f(x)| \leq c\|x\|, \quad x \in H,$$

for some constant $c \geq 0$, in which case we define $\|f\|$ as the smallest constant c with this property. The bounded linear functionals form a linear space H^* with norm $\|f\|$, which is again a Hilbert space isomorphic to H.

---

[4]also known as the *Schwarz* or *Buniakovsky* inequality

[5]Hilbert spaces appear at many places throughout the book, especially in Chapters 8, 14, 18–19, and 21.

Lemma A3.6 *(Riesz representation[6]) For functions f on a Hilbert space H, these conditions are equivalent:*

(i) *f is a continuous linear functional on H,*

(ii) *there exists a unique element $y \in H$ satisfying*

$$f(x) = \langle x, y \rangle, \quad x \in H.$$

In this case, $\|f\| = \|y\|$.

An *ortho-normal system* (ONS) in H is a family of orthogonal elements $x_i \in H$ with norm 1. It is said to be *complete* or to form an *ortho-normal basis* (ONB) if $y \perp (x_i)$ implies $y = 0$.

Lemma A3.7 *(ortho-normal bases) For any Hilbert space H, we have[7]*

(i) *H has an ONB (x_i),*

(ii) *the cardinality of (x_i) is unique,*

(iii) *every $x \in H$ has a unique representation*

$$x = \textstyle\sum_i b_i x_i, \text{ where } b_i = \langle x, x_i \rangle,$$

(iv) *if $x = \sum_i b_i x_i$ and $y = \sum_i c_i x_i$, then*

$$\langle x, y \rangle = \textstyle\sum_i b_i c_i, \qquad \|x\|^2 = \textstyle\sum_i b_i^2.$$

The cardinality in (ii) is called the *dimension* of H. Since any finite-dimensional Hilbert space is isomorphic[8] to a Euclidean space R^n, we may henceforth assume the dimension to be infinite. The simplest infinite-dimensional Hilbert space is l^2, consisting of all infinite sequences $x = (x_n)$ in R with $\|x\|^2 = \sum_n x_n^2 < \infty$. Here the inner product is given by

$$\langle x, y \rangle = \sum_{n \geq 1} x_n y_n, \qquad x = (x_n), \ y = (y_n),$$

and we may choose an ONB consisting of the elements $(0, \dots, 0, 1, 0, \dots)$, with 1 in the n-th position.

Another standard example is $L^2(\mu)$, for any σ-finite measure μ with infinite support on a measurable space S. Here the elements are (equivalence classes of) measurable functions f on S with $\|f\|^2 = \mu f^2 < \infty$, and the inner product is given by $\langle f, g \rangle = \mu(fg)$. Familiar ONB's include the Haar functions when μ equals Lebesgue measure λ on $[0, 1]$, and suitably normalized Hermite polynomials when $\mu = N(0, 1)$ on R.

All the basic Hilbert spaces are essentially equivalent:

[6] also known as the *Riesz–Fischer theorem*

[7] This requires the axiom of choice, assumed throughout this book.

[8] In other words, there exists a 1−1 correspondence between the two spaces preserving the algebraic structure.

Lemma A3.8 *(separable Hilbert spaces) For an infinite-dimensional Hilbert space H, these conditions are equivalent:*

(i) *H is separable,*

(ii) *H has countably infinite dimension,*

(iii) *$H \cong l^2$,*

(iv) *$H \cong L^2(\mu)$ for any σ-finite measure μ with infinite support.*

Most Hilbert spaces H arising in applications are separable, and for simplicity we may often take $H = l^2$, or let $H = L^2(\mu)$ for a suitable measure μ. For any Hilbert spaces H_1, H_2, we may form a new Hilbert space $H_1 \otimes H_2$, called the *tensor product* of H_1, H_2. When $H_1 = L^2(\mu_1)$ and $H_2 = L^2(\mu_2)$, it may be thought of as the space $L^2(\mu_1 \otimes \mu_2)$, where $\mu_1 \otimes \mu_2$ is the product measure of μ_1, μ_2. If $f_1, f_2, \ldots$ and $g_1, g_2, \ldots$ are ONB's in H_1 and H_2, the functions $h_{ij} = f_i \otimes g_j$ form an ONB in $H_1 \otimes H_2$, where $(f \otimes g)(x, y) = f(x)g(y)$. Iterating the construction, we may form the n-fold tensor powers $H^{\otimes n} = H \otimes \cdots \otimes H$, for every $n \in \mathsf{N}$.

4. Linear operators

A *linear operator* between two normed linear spaces X, Y is a mapping $T : X \to Y$ satisfying

$$T(ax + by) = a\,Tx + b\,Ty, \quad x, y \in X, \;\; a, b \in \mathsf{R}.$$

It is said to be *bounded*[9] if $\|Tx\| \le c\,\|x\|$ for some constant $c < \infty$, in which case there is a smallest number $\|T\|$ with this property, so that

$$\|Tx\| \le \|T\|\,\|x\|, \quad x \in X.$$

When $\|T\| \le 1$ we call T a *contraction operator*. The space of bounded linear operators $X \to Y$ is denoted by $\mathcal{B}_{X,Y}$, and when $X = Y$ we write $\mathcal{B}_{X,X} = \mathcal{B}_X$. Those are again linear spaces, under the operations

$$\begin{aligned} (T_1 + T_2)\,x &= T_1 x + T_2 x, \\ (c\,T)x &= c\,(Tx), \qquad x \in X, \;\; c \in \mathsf{R}, \end{aligned}$$

and the function $\|T\|$ above is a norm on $\mathcal{B}_{X,Y}$. Furthermore, $\mathcal{B}_{X,Y}$ is complete, hence a Banach space, whenever this is true for Y. In particular, we note that $\mathcal{B}_{X,\mathsf{R}}$ equals the dual space X^*.

Theorem A4.1 *(Banach–Steinhaus) Let $\mathcal{T} \subset \mathcal{B}_{X,Y}$, for a Banach space X and a normed linear space Y. Then*

$$\sup_{T \in \mathcal{T}} \|Tx\| < \infty, \;\; x \in X \quad \Leftrightarrow \quad \sup_{T \in \mathcal{T}} \|T\| < \infty.$$

[9]Bounded operators are especially important in Chapter 26.

If $T \in \mathcal{B}_{X,Y}$ and $S \in \mathcal{B}_{Y,Z}$, the product $ST \in \mathcal{B}_{X,Z}$ is given by $(ST)x = S(Tx)$, and we note that $\|ST\| \leq \|S\|\,\|T\|$. Two operators $S, T \in \mathcal{B}_X$ are said to *commute* if $ST = TS$. The *identity operator* $I \in \mathcal{B}_X$ is defined by $Ix = x$ for all $x \in X$. Furthermore, $T \in \mathcal{B}_X$ is said to be *idempotent* if $T^2 = T$. For any operators $T, T_1, T_2, \ldots \in \mathcal{B}_{X,Y}$, we say that T_n converges *strongly* to T if

$$\|T_n x - Tx\| \to 0 \text{ in } Y, \quad x \in X.$$

We also consider convergence in the *norm topology*, defined by $\|T_n - T\| \to 0$ in $\mathcal{B}_{X,Y}$.

For any $A \in \mathcal{B}_{X,Y}$, the *adjoint* A^* of A is a linear map on Y^* given by

$$(A^* f)(x) = f(Ax), \quad x \in X, \ f \in Y^*.$$

If also $B \in \mathcal{B}_{Y,Z}$, we note that $(BA)^* = A^* B^*$. Further note that $I^* = I$, where the latter I is given by $If = f$.

Lemma A4.2 *(adjoint operator) For any normed linear spaces X, Y, the map $A \to A^*$ is a linear isometry from $\mathcal{B}_{X,Y}$ to $\mathcal{B}_{Y^*,X^*}$.*

When $A \in \mathcal{B}_{X,Y}$ is bijective, its inverse $A^{-1}\colon Y \to X$ is again linear, though it need not be bounded.

Theorem A4.3 *(inverses and adjoints) For a Banach space X and a normed linear space Y, let $A \in \mathcal{B}_{X,Y}$. Then these conditions are equivalent:*

(i) *A has a bounded inverse $A^{-1} \in \mathcal{B}_{Y,X}$,*

(ii) *A^* has a bounded inverse $(A^*)^{-1} \in \mathcal{B}_{X^*,Y^*}$.*

In that case, $(A^*)^{-1} = (A^{-1})^*$.

For a Hilbert space H we may identify H^* with H, which leads to some simplifications. Then for any $A \in \mathcal{B}_H$ there exists a unique operator $A^* \in \mathcal{B}_H$ satisfying

$$\langle Ax, y\rangle = \langle x, A^* y\rangle, \quad x, y \in H,$$

which again is called the *adjoint* of A. The mapping $A \mapsto A^*$ is a linear isometry on H satisfying $(A^*)^* = A$. We say that A is *self-adjoint* if $A^* = A$, so that $\langle Ax, y\rangle = \langle x, Ay\rangle$. In that case,

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x\rangle|.$$

For any $A \in \mathcal{B}_H$, the *null space* N_A and *range* R_A are given by

$$N_A = \{x \in H;\, Ax = 0\},$$
$$R_A = \{Ax;\, x \in H\},$$

and we note that $(R_A)^\perp = N_{A^*}$ and $(N_A)^\perp = \overline{R}_{A^*}$

Lemma A4.4 *(unitary operators*[10]*) For any $U \in \mathcal{B}_H$, these conditions are equivalent:*

(i) $UU^* = U^*U = I$,

(ii) *U is an isometry on H,*

(iii) *U maps any ONB in H into an ONB.*

We state an abstract version of Theorem 1.35:

Lemma A4.5 *(orthogonal decomposition) Let M be a closed subspace of a Hilbert space H. Then for any $x \in H$, we have*

(i) *x has a unique decomposition $x = y + z$ with $y \in M$ and $z = M^\perp$,*

(ii) *y is the unique element of M minimizing $\|x - y\|$,*

(iii) *writing $y = \pi_M x$, we have $\pi_M \in \mathcal{B}_H$ with $\|\pi_M\| \le 1$, and $\pi_M + \pi_{M^\perp} = I$.*

Lemma A4.6 *(projections*[11]*) When $A \in \mathcal{B}_H$, these conditions are equivalent:*

(i) *$A = \pi_M$ for a closed subspace $M \subset H$,*

(ii) *A is self-adjoint and idempotent.*

In that case, $M = \overline{R}_A$.

For any sub-spaces $M, N \subset H$, we introduce the *orthogonal complement*

$$M \ominus N = M \cap N^\perp = \{x \in M;\ x \perp N\}.$$

When $M \perp N$, we further consider the *direct sum*

$$M \oplus N = \{x + y;\ x \in M,\ y \in N\}.$$

The *conditional orthogonality* $M \perp_R N$ is defined by[12]

$$M \underset{R}{\perp} N \quad \Leftrightarrow \quad \pi_{R^\perp} M \perp \pi_{R^\perp} N.$$

We list some classical propositions of probabilistic relevance.

Lemma A4.7 *(orthogonality) For any closed sub-spaces $M, N \subset H$, these conditions are equivalent:*

(i) $M \perp N$,

(ii) $\pi_M \pi_N = 0$,

(iii) *$\pi_M + \pi_N = \pi_R$ for some R,*

In that case, $R = M \oplus N$.

Lemma A4.8 *(commutativity) For any closed sub-spaces $M, N \subset H$, these conditions are equivalent:*

(i) $M \underset{M \cap N}{\perp} N$,

(ii) $\pi_M \pi_N = \pi_N \pi_M$,

(iii) *$\pi_M \pi_N = \pi_R$ for some R,*

In that case, $R = M \cap N$.

---

[10]Unitary operators are especially important in Chapters 14 and 27–28.

[11]Projection operators play a fundamental role in especially Chapter 8, where most of the quoted results have important probabilistic counterparts.

[12]Here our terminology and notation are motivated by Theorem 8.13.

Lemma A4.9 *(tower property) For any closed sub-spaces $M, N \subset H$, these conditions are equivalent:*

(i) $N \subset M$,

(ii) $\pi_M \pi_N = \pi_N$,

(iii) $\pi_N \pi_M = \pi_N$,

(iv) $\|\pi_N x\| \le \|\pi_M x\|, \quad x \in H$.

The next result may be compared with the probabilistic versions in Corollary 25.18 and Theorem 30.11.

Theorem A4.10 *(mean ergodic theorem) For a bi-measurable, bijection*[13] *T on (S, μ) with $\mu \circ f^{-1} = \mu$, define $Uf = f \circ T$ for $f \in L^2(\mu)$, and let M be the null space of $U - I$. Then*

(i) *U is a unitary operator on $L^2(\mu)$,*

(ii) $$n^{-1} \sum_{k<n} f \circ T^k = n^{-1} \sum_{k<n} U^k f \to \pi_M f, \quad f \in L^2(\mu).$$

We also need to consider possibly unbounded operators[14] A between two Banach spaces X, Y, defined on a linear subspace $\mathcal{D} \subset X$ called the *domain* of A. We say that A is *closed* if its *graph*

$$G = \big\{(x, Ax);\ x \in \mathcal{D}\big\}$$

is a closed subset of $X \times Y$, in the product topology for the norm topologies on X, Y. More generally, an operator $(A, \mathcal{D})$ is said to be *closable*, if the closure $\bar{G}$ in $X \times Y$ is the graph of a single-valued operator $\bar{A}$, called the *closure* of A.

Lemma A4.11 *(closed graph theorem) For a linear operator A between two Banach spaces X, Y, we have*

$$A \text{ is bounded} \quad \Leftrightarrow \quad A \text{ is closed.}$$

Lemma A4.12 *(closable operators) Let $(A, \mathcal{D})$ be a linear operator between the Banach spaces X, Y. Then these conditions are equivalent:*

(i) *A is closable,*

(ii) *for any $x_1, x_2, \ldots \in \mathcal{D}$,*
$$\left.\begin{array}{r} x_n \to 0 \\ Ax_n \to y \end{array}\right\} \Rightarrow y = 0.$$

---

[13]Though measure-preserving maps are not invertible in general, they can be made so by a suitable randomization. Cf. Lemmas 25.2 and 27.4.

[14]Unbounded operators play fundamental roles in especially Chapters 17 and 21.

The domain of an operator may be hard to identify and is often too large for technical convenience. When $(A, \mathcal{D})$ is a closed operator from X to Y, a linear subspace $D \subset \mathcal{D}$ is called a *core* of A, if the restriction of A to D is closable with closure $(A, \mathcal{D})$.

For a suitable space $\mathcal{C}$ of functions on a topological space S, we say that an operator A on $\mathcal{C}$ is *local*, if for any $s \in S$ with a neighborhood G, we have

$$f = 0 \text{ on } G \quad \Rightarrow \quad Af(s) = 0.$$

Then by linearity, $Af(s)$ depends only on values of f in a neighborhood of s. When $S = \mathsf{R}^d$, this holds for any differential operator A, and the converse assertion is often true under additional assumptions. Local operators play important roles in Chapters 17, 21, and 35.

5. Function and measure spaces

Here we first collect some basic facts about the function spaces $C_{T,S}$ or $D_{\mathsf{R}_+,S}$, of special importance in probability theory. Though processes with paths in those spaces are considered throughout the book, most topological results mentioned here are not needed until Chapter 23, where they are fundamental for the theory of convergence in distribution.

Let $C_{T,S}$ be the space of continuous, S-valued functions on T, where S, T are separable, complete metric spaces and T is locally compact. For any functions $x, x_1, x_2, \ldots \in C_{T,S}$, we define the *locally uniform convergence*[15] $x_n \xrightarrow{ul} x$ by

$$\sup_{t \in K} \rho(x_t^n, x_t) \to 0, \quad K \in \mathcal{K}_T, \tag{1}$$

where ρ is the metric on S and $\mathcal{K}_T$ denotes the class of compact sets $K \subset T$. The distance in (1) defines a separable and complete metric on $C_{K,S}$, and we may construct a similar metric $\hat{\rho}$ on $C_{T,S}$ by applying (1) to a sequence of compact sets[16] $K_n \in \mathcal{K}_T$ with $K_n^o \uparrow T$. The *evaluation maps* on $C_{T,S}$ are given by $\pi_t : x \mapsto x_t$ for all $t \in T$.

Lemma A5.1 *(locally uniform topology on $C_{T,S}$) Let S, T be separable, complete metric spaces, with T locally compact. Then there exists a topology $\mathcal{T}$ on $C_{T,S}$, such that*

(i) *$\mathcal{T}$ induces the convergence $x_n \xrightarrow{ul} x$ in* (1),
(ii) *$C_{T,S}$ is Polish under $\mathcal{T}$,*
(iii) *$\mathcal{T}$ generates the Borel σ-field $\sigma\{\pi_t;\, t \in T\}$.*

Proof: We need to prove only (iii), the remaining claims being obvious. The maps π_t are continuous, hence Borel measurable, and so the generated

[15] Here and below, the associated net convergence is given by the same formulas.
[16] which exist since T is locally compact

σ-field $\mathcal{C}$ is contained in $\mathcal{B}_{C_{T,S}}$. For the converse, we need to show that any open subset $G \subset C_{T,S}$ lies in $\mathcal{C}$. Then choose a separable and complete metric $\hat{\rho}$ in $C_{T,S}$, and note that G is a countable union of open balls

$$B_x^r = \left\{ y \in C_{T,S};\ \hat{\rho}(x,y) < r \right\}.$$

The latter lie in $\mathcal{C}$, since for any countable dense set $T' \subset T$,

$$\bar{B}_x^r = \bigcap_{t \in T'} \left\{ y \in C_{T,S};\ \rho(x_t, y_t) \le r \right\}.$$

□

We state a version of the classical *Arzelà–Ascoli criterion* for compactness in $C_{T,S}$. For any metrics d on T and ρ on S, we define the associated *moduli of continuity* by

$$w_K(x,h) = \sup\left\{ \rho(x_s, x_t);\ s,t \in K,\ d(s,t) \le h \right\}, \quad h > 0,\ K \in \mathcal{K}_T.$$

Theorem A5.2 *(compactness in $C_{T,S}$, Arzelà, Ascoli) Let $A \subset C_{T,S}$, where S, T are separable, complete metric spaces with T locally compact, and fix a dense set $T' \subset T$. Then A is relatively compact in the locally uniform topology iff*

(i) *$\pi_t A$ is relatively compact in S for every $t \in T'$,*

(ii) $\lim_{h \to 0} \sup_{x \in A} w_K(x,h) = 0, \quad K \in \mathcal{K}_T.$

In that case,

(iii) *$\bigcup_{t \in K} \pi_t A$ is relatively compact in S for all $K \in \mathcal{K}_T$.*

Proof: This is essentially a special case of Theorem A5.4 below. Various versions are proved in textbooks on real analysis. Probabilists may consult Dudley (1989), Section 2.4. □

Turning to spaces of functions with possible jump discontinuities, let $D_{\mathsf{R}_+,S}$ be the space of rcll[17] functions on R_+ with values in a separable, complete metric space (S,ρ). Note that such functions x can have only jump discontinuities, and that only finitely many jumps of size $\rho(x_t, x_{t-}) \ge \varepsilon$ may occur in every finite interval. Here the topology of locally uniform convergence $x_n \xrightarrow{ul} x$ is less useful and technically a bit awkward, since the separability of S may not be preserved in $D_{\mathsf{R}_+,S}$.

For a more useful and convenient topology, we allow the path of each x to be shifted in time by an increasing bijection λ on R_+, so that λ is continuous and strictly increasing with $\lambda_0 = 0$. For the associated *Skorohod* convergence $x_n \xrightarrow{s} x$ in $D_{\mathsf{R}_+,S}$, we require

$$\lambda_n \xrightarrow{ul} \iota, \qquad x_n \circ \lambda_n \xrightarrow{ul} x, \tag{2}$$

for some bijections λ_n as above, where ι denotes the identity map on R_+. Though clearly $x_n \xrightarrow{ul} x$ implies $x_n \xrightarrow{s} x$, the converse implication is false. We state the basic properties of the corresponding *Skorohod J_1-topology*:

[17]right-continuous with left-hand limits, often written as the French acronym *càdlàg*

Lemma A5.3 *(Skorohod topology on $D_{\mathsf{R}_+,S}$, Skorohod, Prohorov, Kolmogorov) For any separable, complete metric space S, there exists a topology $\mathcal{T}$ on $D_{\mathsf{R}_+,S}$, such that*

(i) *$\mathcal{T}$ induces the convergence $x_n \xrightarrow{s} x$ in* (2),

(ii) *$D_{\mathsf{R}_+,S}$ is Polish under $\mathcal{T}$,*

(iii) *$\mathcal{T}$ generates the Borel σ-field $\sigma\{\pi_t;\ t \geq 0\}$.*

Proof: Detailed proofs of various versions appear in Billingsley (1968), pp. 111–116, Ethier & Kurtz (1986), pp. 116–122, and Jacod & Shiryaev (1987), pp. 291–301. □

To state an associated compactness criterion, we define some *modified moduli of continuity* on $D_{\mathsf{R}_+,S}$ by

$$\tilde{w}_t(x,h) = \inf_{(I_k)} \max_k \sup_{r,s\in I_k} \rho(x_r,x_s), \quad x \in D_{\mathsf{R}_+,S},\ \ t,h > 0, \tag{3}$$

where the infimum extends over all partitions of the interval $[0,t)$ into subintervals $I_k = [u,v)$ with $v - u \geq h$ when $v < t$. Note that $\tilde{w}_t(x,h) \to 0$ as $h \to 0$ for fixed $x \in D_{\mathsf{R}_+,S}$ and $t > 0$.

Theorem A5.4 *(compactness in $D_{\mathsf{R}_+,S}$, Prohorov) Let $A \subset D_{\mathsf{R}_+,S}$ for a separable, complete metric space S, and fix a dense set $T \subset \mathsf{R}_+$. Then A is relatively compact in the Skorohod topology iff*

(i) *$\pi_t A$ is relatively compact in S for all $t \in T$,*

(ii) $$\lim_{h\to 0} \sup_{x\in A} \tilde{w}_t(x,h) = 0, \quad t > 0. \tag{4}$$

In that case,

(iii) *$\bigcup_{s\leq t} \pi_s A$ is relatively compact in S for all $t \geq 0$.*

Proof: Detailed proofs are implicit in Ethier & Kurtz (1986), pp. 122–127. Related versions appear in Billingsley (1968), pp. 116–120, and Jacod & Shiryaev (1987), pp. 292–298. □

We turn to some measure spaces. Given a separable, complete metric space S, let $\mathcal{M}_S$ be the class of locally finite measures μ on S, so that $\mu B < \infty$ for all bounded Borel sets B. For any $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_S$, we define the *vague convergence* $\mu_n \xrightarrow{v} \mu$ by

$$\mu_n f \to \mu f, \quad f \in \hat{C}_S, \tag{5}$$

where $\hat{C}_S$ is the class of bounded, continuous functions $f \geq 0$ on S with bounded support. The maps $\pi_f\colon \mu \mapsto \mu f$ induce the *vague topology*[18] on $\mathcal{M}_S$:

[18]When S is locally compact, it is equivalent to require the supports of f to be compact. In general, that would give a smaller and less useful topology.

Lemma A5.5 *(vague topology on $\mathcal{M}_S$) For a separable, complete metric space S, there exists a topology $\mathcal{T}$ on $\mathcal{M}_S$, such that*

(i) *$\mathcal{T}$ induces the convergence $\mu_n \xrightarrow{v} \mu$ in* (5),
(ii) *$\mathcal{M}_S$ is Polish under $\mathcal{T}$,*
(iii) *$\mathcal{T}$ generates the Borel σ-field $\sigma\{\pi_f;\ f \in \hat{C}_S\}$.*

Proof: See K(17), pp. 111–117. □

On the sub-class $\mathcal{M}_S^b$ of bounded measures on S, we also consider the *weak topology* and associated *weak convergence* $\mu_n \xrightarrow{w} \mu$, given by (5) with $\hat{C}_S$ replaced by the class C_S of bounded, continuous functions on S. No further analysis is required, since this agrees with the vague topology and convergence, when the metric ρ on S is replaced by the equivalent metric $\rho \wedge 1$, so that all sets in S become bounded.

Theorem A5.6 *(vague compactness in $\mathcal{M}_S$, Prohorov, Matthes et al.) Let $A \subset \mathcal{M}_S$ for a separable, complete metric space S. Then A is vaguely relatively compact iff*

(i) $\sup_{\mu \in A} \mu B < \infty, \quad B \in \hat{\mathcal{S}},$
(ii) $\inf_{K \in \mathcal{K}} \sup_{\mu \in A} \mu(B \setminus K) = 0, \quad B \in \hat{\mathcal{S}}.$

In particular, a set $A \subset \mathcal{M}_S^b$ is weakly relatively compact iff (i)−(ii) *hold with $B = S$.*

Proof: See K(17), pp. 114, 116. □

We finally consider spaces of measure-valued rcll functions. Here we may characterize compactness in terms of countably many one-dimensional projections. Here we write $D_{\mathsf{R}_+,S} = D_S$ for simplicity.

Theorem A5.7 *(measure-valued functions) Given an lcscH space S, there exist some $f_1, f_2, \ldots \in \hat{C}_S$, such that these conditions are equivalent, for any $A \subset D_{\mathcal{M}_S}$:*

(i) *A is relatively compact in $D_{\mathcal{M}_S}$,*
(ii) *$Af_j = \{xf_j;\ x \in A\}$ is relatively compact in D_{R_+} for every $j \in \mathsf{N}$.*

Proof: If A is relatively compact, then so is Af for every $f \in \hat{C}_S$, since the map $x \mapsto xf$ is continuous from $D_{\mathcal{M}_S}$ to D_{R_+}. To prove the converse, we may choose a dense collection $f_1, f_2, \ldots \in \hat{C}_S$, closed under addition, such that Af_j is relatively compact for every j. In particular, $\sup_{x \in A} x_t f_j < \infty$ for all $t \geq 0$ and $j \in \mathsf{N}$, and so by Theorem A5.6 the set $\{x_t;\ x \in A\}$ is relatively compact in $\mathcal{M}_S$ for every $t \geq 0$. By Theorem A5.4 it remains to verify (4), where $\tilde{w}$ may be defined in terms of the complete metric

$$\rho(\mu,\nu) = \sum_{k\geq 1} 2^{-k}\Big\{\Big|\mu f_k - \nu f_k\Big| \wedge 1\Big\}, \quad \mu,\nu \in \mathcal{M}_S. \tag{6}$$

If (4) fails, then either we may choose some $x^n \in A$ and $t_n \to 0$ with $\limsup_n \rho(x^n_{t_n}, x^n_0) > 0$, or else there exist some $x^n \in A$ and bounded $s_t < t_n < u_n$ with $u_n - s_n \to 0$ such that

$$\limsup_{n\to\infty}\Big\{\rho(x^n_{s_n}, x^n_{t_n}) \wedge \rho(x^n_{t_n}, x^n_{u_n})\Big\} > 0. \tag{7}$$

In the former case, (6) yields $\limsup_n |x^n_{t_n} f_j - x^n_0 f_j| > 0$ for some $j \in \mathsf{N}$, which contradicts the relative compactness of Af_j. Assuming (7) instead, we have by (6) some $i, j \in \mathsf{N}$ with

$$\limsup_{n\to\infty}\Big\{\Big|x^n_{s_n} f_i - x^n_{t_n} f_i\Big| \wedge \Big|x^n_{t_n} f_j - x^n_{u_n} f_j\Big|\Big\} > 0. \tag{8}$$

Now for any $a, a', b, b' \in \mathsf{R}$,

$$\tfrac{1}{2}\Big(|a| \wedge |b'|\Big) \leq \Big(|a| \wedge |a'|\Big) \vee \Big(|b| \wedge |b'|\Big) \vee \Big(|a+a'| \wedge |b+b'|\Big).$$

Since the set $\{f_k\}$ is closed under addition, (8) yields the same relation with a common $i = j$. But then (4) fails for Af_i, which contradicts the relative compactness of Af_i by Theorem A5.4. Thus, (4) fails for A, and so A is relatively compact. □

6. Classes and spaces of sets

Given an lcscH space S, we introduce the classes $\mathcal{G}$, $\mathcal{F}$, $\mathcal{K}$ of open, closed, and compact subsets, respectively. Here we may regard $\mathcal{F}$ as a space in its own right, endowed with the *Fell topology* and associated convergence[19]. To introduce the latter, choose a metric ρ on S making every closed ρ-ball compact, and define

$$\rho(s, F) = \inf_{x\in F} \rho(s,x), \quad s \in S, \;\; F \in \mathcal{F}.$$

Then for any $F, F_1, F_2, \ldots \in \mathcal{F}$, we define $F_n \xrightarrow{f} F$ by the condition

$$\rho(s, F_n) \to \rho(s, F), \quad s \in S. \tag{9}$$

We show that (9) is independent of the choice of ρ, and is induced by the topology generated by the sets

$$\begin{aligned} &\Big\{F \in \mathcal{F};\, F \cap G \neq \emptyset\Big\}, \qquad G \in \mathcal{G}, \\ &\Big\{F \in \mathcal{F};\, F \cap K = \emptyset\Big\}, \qquad K \in \mathcal{K}. \end{aligned} \tag{10}$$

Theorem A6.1 *(Fell topology on $\mathcal{F}$) Let $\mathcal{F}$ be the class of closed sets in an lcscH space S. Then there exists a topology $\mathcal{T}$ on $\mathcal{F}$, such that*

(i) *$\mathcal{T}$ generates the convergence $F_n \xrightarrow{f} F$ in* (9),

(ii) *$\mathcal{F}$ is compact and metrizable under $\mathcal{T}$,*

(iii) *$\mathcal{T}$ is generated by the sets* (10),

(iv) *$\Big\{F \in \mathcal{F};\, F \cap B \neq \emptyset\Big\}$ is universally Borel measurable for every $B \in \mathcal{S}$.*

---

[19]This is important for our discussion of random sets in Chapters 23 and 34.

Proof, (i)−(iii): We show that the Fell topology in (iii) is generated by the maps $F \mapsto \rho(s, F)$, $s \in S$. The convergence criterion in (i) then follows, once the metrization property is established. To see that the stated maps are continuous, put $B_s^r = \{t \in S;\ \rho(s,t) < r\}$, and note that

$$\{F;\ \rho(s,F) < r\} = \{F;\ F \cap B_s^r \neq \emptyset\},$$
$$\{F;\ \rho(s,F) > r\} = \{F;\ F \cap \bar{B}_s^r = \emptyset\}.$$

Here the sets on the right are open, by the definition of the Fell topology and the choice of ρ. Thus, the Fell topology contains the ρ-topology.

To prove the converse, fix any $F \in \mathcal{F}$ and a net $\{F_i\} \subset \mathcal{F}$ with directed index set $(I, \prec)$, such that $F_i \to F$ in the ρ-topology. To see that convergence holds even in the Fell topology, let $G \in \mathcal{G}$ be arbitrary with $F \cap G \notin \emptyset$. Fix any $s \in F \cap G$. Since $\rho(s, F_i) \to \rho(s, F) = 0$, we may further choose some $s_i \in F_i$ with $\rho(s, s_i) \to 0$. Since G is open, there exists an $i \in I$ with $s_j \in G$ for all $j \succ i$. Then also $F_j \cap G \notin \emptyset$ for all $j \succ i$.

Next let $K \in \mathcal{K}$ with $F \cap K = \emptyset$. Define $r_s = \frac{1}{2}\,\rho(s, F)$ for each $s \in K$, and put $G_s = B_{s,r_s}$. Since K is compact, it is covered by finitely many balls G_{s_k}. For each k we have $\rho(s_k, F_i) \to \rho(s_k, F)$, and so there exists an $i_k \in I$ with $F_j \cap G_{s_k} = \emptyset$ for all $j \succ i_k$. Letting $i \in I$ with $i \succ i_k$ for all k, it is clear that $F_j \cap K = \emptyset$ for all $j \succ i$.

(ii) Fix a countable, dense set $D \subset S$, and assume that $\rho(s, F_i) \to \rho(s, F)$ for all $s \in D$. For any $s, s' \in S$,

$$\Big|\rho(s, F_j) - \rho(s, F)\Big| \le \Big|\rho(s', F_j) - \rho(s', F)\Big| + 2\rho(s, s').$$

Given any s and $\varepsilon > 0$, we can make the left-hand side $< \varepsilon$ by choosing $s' \in D$ with $\rho(s, s') < \varepsilon/3$, and then letting $i \in I$ be such that $|\rho(s', F_j) - \rho(s', F)| < \varepsilon/3$ for all $j \succ i$. Hence, the Fell topology is also generated by the maps $F \mapsto \rho(s, F)$ for all $s \in D$. But then $\mathcal{F}$ is homeomorphic to a subset of $\bar{\mathsf{R}}_+^\infty$, which is second-countable and metrizable.

To see that $\mathcal{F}$ is compact, it suffices to show that every sequence $(F_n) \subset \mathcal{F}$ contains a convergent sub-sequence. Then choose a sub-sequence with $\rho(s, F_n)$ converging in $\bar{\mathsf{R}}_+$ for all $s \in D$, and hence also for all $s \in S$. Since the functions $\rho(s, F_n)$ are equi-continuous, the limit f is again continuous, and so the set $F = \{s \in S;\ f(s) = 0\}$ is closed.

To obtain $F_n \to F$, we need to show that, whenever $F \cap G \neq \emptyset$ or $F \cap K = \emptyset$ for some $G \in \mathcal{G}$ or $K \in \mathcal{K}$, the same relation holds eventually even for F_n. In the former case, fix any $s \in F \cap G$, and note that $\rho(s, F_n) \to f(s) = 0$. Hence, we may choose some $s_n \in F_n$ with $s_n \to s$, and since $s_n \in G$ for large n, we get $F_n \cap G \neq \emptyset$. In the latter case, assume that instead $F_n \cap K \neq \emptyset$ along a sub-sequence. Then there exist some $s_n \in F_n \cap K$, and so $s_n \to s \in K$ along a further sub-sequence. Here $0 = \rho(s_n, F_n) \to \rho(s, F)$, which yields the contradiction $s \in F \cap K$.

(iv) The mapping $(s,F) \mapsto \rho(s,F)$ is jointly continuous, hence Borel measurable. Now S and $\mathcal{F}$ are both separable, and so the Borel σ-field in $S \times \mathcal{F}$ agrees with the product σ-field $\mathcal{S} \otimes \mathcal{B}_{\mathcal{F}}$. Since $s \in F$ iff $\rho(s,F) = 0$, it follows that $\{(s,F);\, s \in F\}$ belongs to $\mathcal{S} \otimes \mathcal{B}_{\mathcal{F}}$. Hence, so does $\{(s,F);\, s \in F \cap B\}$ for arbitrary $B \in \mathcal{S}$. The assertion now follows by Theorem A1.2. $\square$

Say that a class $\mathcal{U} \subset \hat{\mathcal{S}}$ is *separating*, if for any $K \subset G$ with $K \in \mathcal{K}$ and $G \in \mathcal{G}$, there exists a $U \in \mathcal{U}$ with $K \subset U \subset G$. A class $\mathcal{I} \subset \hat{\mathcal{S}}$ is *pre-separating* if the finite unions of $\mathcal{I}$-sets form a separating class. When S is Euclidean, we typically choose $\mathcal{I}$ to be a class of intervals or rectangles and $\mathcal{U}$ as the corresponding class of finite unions.

Lemma A6.2 *(separation) For any monotone function $h\colon \hat{\mathcal{S}} \to \mathsf{R}$, we have the separating class*

$$\hat{\mathcal{S}}_h = \left\{B \in \hat{\mathcal{S}};\, h(B^o) = h(\bar{B})\right\}.$$

Proof: Fix a metric ρ in S such that every closed ρ-ball is compact, and let $K \in \mathcal{K}$ and $G \in \mathcal{G}$ with $K \subset G$. For any $\varepsilon > 0$, define $K_\varepsilon = \{s \in S;\, d(s,K) < \varepsilon\}$ and note that $\bar{K}_\varepsilon = \{s \in S;\, \rho(s,K) \le \varepsilon\}$. Since K is compact, we have $\rho(K, G^c) > 0$, and so $K \subset K_\varepsilon \subset G$ for sufficiently small $\varepsilon > 0$. Furthermore, the monotonicity of h yields $K_\varepsilon \in \hat{\mathcal{S}}_h$ for almost every $\varepsilon > 0$. $\square$

We often need the separating class to be countable.

Lemma A6.3 *(countable separation) Every separating class $\mathcal{U} \subset \hat{\mathcal{S}}$ contains a countable separating subclass.*

Proof: Fix a countable topological base $\mathcal{B} \subset \hat{\mathcal{S}}$, closed under finite unions. For every $B \in \mathcal{B}$, choose some compact sets $K_{B,n} \downarrow \bar{B}$ with $K^o_{B,n} \supset \bar{B}$. Then for every pair $(B,n) \in \mathcal{B} \times \mathsf{N}$, choose $U_{B,n} \in \mathcal{U}$ with $\bar{B} \subset U_{B,n} \subset K^o_{B,n}$. The family $\{U_{B,n}\}$ is clearly separating. $\square$

To state the next result, fix any metric spaces $S_1, S_2, \ldots$, and introduce the product spaces $S^n = S_1 \times \cdots \times S_n$ and $S = S_1 \times S_2 \times \cdots$, endowed with their product topologies. For any $m < n < \infty$, let π_m and π_{mn} be the natural projections of S or S^n onto S^m. Say that the sets $A_n \subset S^n$, $n \in \mathsf{N}$, form a *projective sequence* if $\pi_{mn} A_n \subset A_m$ for all $m \le n$. Their *projective limit* in S is then defined as the set $A = \bigcap_n \pi_n^{-1} A_n$.

Lemma A6.4 *(projective limits) For any metric spaces $S_1, S_2, \ldots$, consider a projective sequence of non-empty, compact sets $K_n \subset S_1 \times \cdots \times S_n$, $n \in \mathsf{N}$. Then the projective limit $K = \bigcap_n \pi_n^{-1} K_n$ is again non-empty and compact.*

Proof: Since the K_n are non-empty, we may choose some sequences $x^n = (x^n_m) \in \pi_n^{-1} K_n$, $n \in \mathsf{N}$. By the projective property of the sets K_m, we have $\pi_m x^n \in K_m$ for all $m \le n$. In particular, the sequence $x^1_m, x^2_m, \ldots$ is relatively compact in S_m for each $m \in \mathsf{N}$, and so a diagonal argument yields convergence $x^n \to x = (x_m) \in S$ along a sub-sequence $N' \subset \mathsf{N}$, which implies $\pi_m x^n \to \pi_m x$ along N' for each $m \in \mathsf{N}$. Since the K_m are closed, we obtain $\pi_m x \in K_m$ for all m, and so $x \in K$, which shows that K is non-empty. The compactness of K may be proved by a similar argument, where we assume that $x^1, x^2, \ldots \in K$. □

We also include a couple of estimates for convex sets, needed in Chapter 25. Here $\partial_\varepsilon B$ denotes the ε-neighborhood of the boundary ∂B.

Lemma A6.5 *(convex sets) If $B \subset \mathsf{R}^d$ is convex and $\varepsilon > 0$, then*

(i) $\lambda^d(B - B) \le \binom{2d}{d} \lambda^d B$,

(ii) $\lambda^d(\partial_\varepsilon B) \le 2\left\{(1+\varepsilon/r_B)^d - 1\right\}\lambda^d B$.

Proof: For (i), see Rogers & Shephard (1958). Part (ii) is elementary. □

Throughout this book, we assume the *Zermelo-Fraenkel axioms* (ZF) of set theory, amended with the *Axiom of Choice* (AC), stated below. It is known that, if the ZF axioms are *consistent*[20] (non-contradictory), then so is the ZF system amended with AC.

Axiom of Choice, AC: *For any class of sets $S_t \ne \emptyset$, $t \in T$, there exists a function f on T with*

$$f(t) \in S_t, \quad t \in T.$$

In other words, for any collection of non-empty sets S_t, the product set $\mathsf{X}_t S_t$ is again non-empty. We often need the following equivalent statement involving partially ordered sets $(S, \prec)$. When $A \subset S$ is linearly ordered under $\prec$, we say that $b \in S$ is an *upper bound* of A if $s \prec b$ for all $s \in A$. Furthermore, an element $m \in S$ is said to be *maximal* if $m \prec s \in S \;\Rightarrow\; s = m$.

Lemma A6.6 *(Zorn's lemma, Kuratowski) Let $(S, \prec)$ be a partially ordered set. Then* (i) ⇒ (ii), *where*

(i) *every linearly ordered subset $A \subset S$ has an upper bound $b \in S$,*

(ii) *S has a maximal element $m \in S$.*

Proof: See Dudley (1989) for a comprehensive discussion of AC and its equivalents. □

[20] The consistency of ZF must be taken on faith, since no proof is possible. It had better be true, or else the entire edifice of modern mathematics would collapse. Most logicians seem to agree that even questions of truth are meaningless.

7. Differential geometry

Here we state some basic facts from *intrinsic*[21] differential geometry, needed in Chapter 35. A more detailed discussion of the required results is given by Emery (1989).

Lemma A7.1 *(tangent vectors) Let M be a smooth manifold with class $\mathcal{S}$ of smooth functions $M \to \mathsf{R}$, and consider a point-wise smooth map $u: \mathcal{S} \to \mathcal{S}$. Then these conditions are equivalent*[22]*:*

(i) *u is a differential operator*[23] *of the form*

$$uf = u^i \, \partial_i f, \quad f \in \mathcal{S},$$

(ii) *u is linear and satisfies*

$$uf^2 = 2f(uf), \quad f \in \mathcal{S},$$

(iii) *for any smooth functions $f = (f^1, \ldots, f^n): M \to \mathsf{R}^n$ and $\varphi: \mathsf{R}^n \to \mathsf{R}$,*

$$u(\varphi \circ f) = (\partial_i \varphi \circ f) \, uf^i.$$

Proof, (i) $\Rightarrow$ (iii): Consider the restriction to a local chart U, and write $f^i = \tilde{f}^i \circ \psi$ in terms of some local coordinates $\psi^1, \ldots, \psi^n$. Using (i) (twice) along with the elementary chain rule, we get

$$\begin{aligned} u(\varphi \circ f) &= (u\psi^i) \Big(\partial_i (\varphi \circ \tilde{f}) \circ \psi \Big) \\ &= (u\psi^i) \Big(\partial_j \varphi \circ \tilde{f} \circ \psi \Big) \Big(\partial_i \tilde{f}^j \circ \psi \Big) \\ &= \Big(\partial_j \varphi \circ f \Big) (u\psi^i) \Big(\partial_i \tilde{f}^j \circ \psi \Big) \\ &= \Big(\partial_j \varphi \circ f \Big) (uf^j). \end{aligned}$$

(iii) $\Rightarrow$ (ii): Taking $n = 2$ and $\varphi(x, y) = ax + by$ yields the linearity of u. Then take $n = 1$ and $\varphi(x) = x^2$ to get the relation in (ii).

(ii) $\Rightarrow$ (i): Condition (ii) gives $u1 = 0$, and by polarization $u(fg) = f(ug) + g(uf)$, which extends by iteration to

$$u(fgh) = fg(uh) + gh(uf) + hf(ug), \quad f, g, h \in \mathcal{S}. \tag{11}$$

If $f \in \mathcal{S}$ with $f = 0$ in a neighborhood of a point $x \in M$, we may choose $g \in \mathcal{S}$ with $g(a) = 0$ and $fg = f$, where the former formula yields $uf(x) = 0$, showing that the operator u is local. Thus, we may henceforth take f to be supported by a single chart U, and write $f = \tilde{f} \circ \varphi$ for some smooth functions $\varphi: U \to \mathsf{R}^n$ and $\tilde{f}: \mathsf{R}^n \to \mathsf{R}$, where the φ^i form local coordinates on U. Fixing any $a \in U$, we may assume that $\varphi(a) = 0$. By Taylor's formula,

$$\tilde{f}(r) = \tilde{f}(0) + r^i \partial_i \tilde{f}(0) + r^i r^j h_{ij}(r), \quad r \in \mathsf{R}^n,$$

---

[21] meaning that no notion or result depends on the choice of local coordinates

[22] Here and below, summation over repeated indices is understood.

[23] In local coordinates, we can identify u with the geometric vector $u = (u^1, \ldots, u^n)$.

for some smooth functions h_{ij} on R^n. Inserting $r = \varphi(x)$, applying u to both sides, and noting that $uc = 0$ for constants c, we get

$$uf(x) = u\varphi^i(x)\,\partial_i\tilde{f}(0) + u\Big\{\varphi^i\varphi^j\,(h_{ij}\circ\varphi)\Big\}(x).$$

Taking $x = a$ and using (11), we obtain

$$uf(a) = \Big\{u\varphi^i(a)\Big\}\,\partial_i\tilde{f}(0),$$

and the assertion follows with $u^i = u\varphi^i$. □

The space of tangent vectors at x is denoted by T_xS, and a smooth function of tangent vectors is called a (tangent) *vector field* on S. The space of vector fields $u, v, \ldots$ on S will be denoted by T_S. A *cotangent* at x is an element of the dual space T_x^*, and smooth functions $\alpha, \beta, \ldots$ of cotangent vectors are called (simple) *forms*. Similarly, smooth functions $b, c, \ldots$ on the dual T_S^{*2} of the product space T_S^2 are called *bilinear forms*.

The dual coupling of simple or bilinear forms α or b with vector fields u, v is often written as

$$\langle\alpha, u\rangle = \alpha(u), \qquad \langle b, (u,v)\rangle = b(u,v).$$

A special role is played by the forms df, given by

$$\langle df, u\rangle = df(u) = uf, \quad f \in \mathcal{S}.$$

For any simple forms α, β, we define a bilinear form $\alpha\otimes\beta$ by

$$\begin{aligned}(\alpha\otimes\beta)(u,v) &= \langle\alpha, u\rangle\langle\beta, v\rangle\\ &= \alpha(u)\beta(v),\end{aligned}$$

so that in particular

$$(df\otimes dg)(u,v) = (uf)(vg).$$

In general, we have the following representations[24] in terms of simple or bilinear forms df and $df\otimes dg$.

Lemma A7.2 *(simple and bilinear forms) For a smooth manifold S,*

(i) *every form on S is a finite sum*

$$\alpha = a_i\,df^i \text{ with } a_i, f^i \in \mathcal{S},$$

(ii) *every bilinear form on S is a finite sum*

$$b = b_{ij}\,(df^i\otimes dg^j) \text{ with } b_{ij}, f^i, g^j \in \mathcal{S}.$$

Such simple and bilinear forms can be represented as follows in terms of local coordinates:

[24]Note that the coefficients are functions, not constants.

Lemma A7.3 *(coordinate representations) Let S be a smooth manifold with local coordinates $x^1, \dots, x^n$. Then for any $f, g, h \in \mathcal{S}$,*

(i) $f\,dg = (f\partial_i g)(x)\, dx^i$,

(ii) $f(dg \otimes dh) = \big\{ f(\partial_i g)(\partial_j h) \big\}(x)\, (dx^i \otimes dx^j)$.

For a smooth mapping φ between two manifolds S, S', the *push-forward* of a tangent vector $u \in T_x S$ into a tangent vector $\varphi \circ u \in T_{\varphi_x} S'$ is given by[25]

$$\begin{aligned} (\varphi \circ u) f &= u(f \circ \varphi), \\ df(\varphi \circ u) &= d(f \circ \varphi) u. \end{aligned}$$

The dual operations of *pull-back* of a simple or bilinear form α or b on S, here denoted by φ^* or φ^{*2}, respectively, are given, for tangent vectors u, v at suitable points, by

$$\begin{aligned} (\varphi^* \alpha) u &= \alpha(\varphi \circ u), \\ (\varphi^{*2} b)(u, v) &= b(\varphi \circ u,\, \varphi \circ v). \end{aligned}$$

(In traditional notation, the latter formulas may be written as

$$\begin{aligned} \big\langle (T_x^* \varphi)\, \alpha, A \big\rangle_x \;&=\; \big\langle \alpha, (T_x \varphi) A \big\rangle_{\varphi(x)}, \\ (T^*\varphi \otimes T^*\varphi)\, b(x)(A, B) \;&=\; b(\varphi(x))\, \big\{ (T_x\varphi) A, (T_x \varphi) B \big\}.) \end{aligned}$$

Here we list some useful pull-back formulas, first for simple forms,

$$\begin{aligned} \varphi^*(f\alpha) &= (f \circ \varphi)(\varphi^* \alpha), \\ \varphi^*(df) &= d(f \circ \varphi), \end{aligned}$$

and then for bilinear forms,

$$\begin{aligned} \varphi^{*2}(fb) &= (f \circ \varphi)(\varphi^{*2} b), \\ \varphi^{*2}(\alpha \otimes \beta) &= \varphi^* \alpha \otimes \varphi^* \beta, \\ (\psi \circ \varphi)^{*2} &= \varphi^{*2} \circ \psi^{*2}. \end{aligned}$$

A *Riemannian metric* on S is a symmetric, positive-definite, bilinear form ρ on S, determining an inner product $\langle u \,|\, v \rangle = \rho(u, v)$ with associated norm $\|u\|$ on S. For any $f \in \mathcal{S}$, we define the *gradient* $\operatorname{grad} f = \hat{f}$ as the unique vector field on S satisfying

$$\begin{aligned} \rho(\hat{f}, u) &= \langle \hat{f} \,|\, u \rangle \\ &= df(u) = uf, \end{aligned}$$

for any vector field u on S. The inner product on T_S may be transferred to the forms df through the relation $\langle df \,|\, dg \rangle = \langle \hat{f} \,|\, \hat{g} \rangle$, so that[26]

---

[25] To facilitate the access for probabilists, we often use a simplified notation. Traditional versions of the stated formulas are sometimes inserted for reference and comparison.

[26] The mapping $df \mapsto \hat{f}$ extends immediately to a linear map $\alpha \mapsto \alpha^\sharp$ between T_S^* and T_S with inverse $u \mapsto u^\flat$, sometimes called the *musical isomorphisms*. Unfortunately, the nice interpretation of $\hat{f}$ as a gradient vector seems to be lost in general.

$$\hat{f}g = \hat{g}f = \langle \hat{f} | \, \hat{g} \rangle$$
$$= \langle df | \, dg \rangle.$$

Letting $\rho = \rho_{ij}(dx^i \otimes dx^j)$ in local coordinates and writing (ρ^{ij}) for the inverse of the matrix (ρ_{ij}), we note that

$$\langle u \, | \, v \rangle = \rho_{ij} \, u^i v^j, \qquad \langle \hat{f} \, | \, \hat{g} \rangle = \rho^{ij} \hat{f}_i \, \hat{g}_j,$$
$$\hat{f}^{\,i} = \rho^{ij} \, \partial_j f, \qquad \partial_i f = \rho_{ij} \hat{f}^{\,j}.$$

For special purposes, we also need some Lie derivatives. Since the formal definitions may be confusing for the novice[27], we list only some basic formulas, which can be used recursively to calculate the Lie derivatives along a given vector field u of a smooth function f, another vector field v, a form α, or a bilinear form b.

Theorem A7.4 *(Lie derivatives)*

(i) $\mathcal{L}_u f = uf$,

(ii) $\mathcal{L}_u v = uv - vu = [u, v]$,

(iii) $\langle \mathcal{L}_u \alpha, v \rangle = \mathcal{L}_u \langle \alpha, v \rangle - \langle \alpha, \mathcal{L}_u v \rangle$,

(iv) $(\mathcal{L}_u b)(v, w) = u\{b(v, w)\} - b\,(\mathcal{L}_u v, w) - b\,(v, \mathcal{L}_u w)$,

Proof: See Emery (1989), pp. 16–19. □

Thus, $\mathcal{L}_u f$ is simply the vector field u itself, applied to the function f. The u-derivative of a vector field v equals the commutator or *Lie bracket* $[u, v]$. Since the u-derivative of a simple or bilinear form α or b can be calculated recursively from (i)$-$(iv), we never need to go back to the underlying definitions in terms of derivatives. From the stated formulas, it is straightforward to derive the following useful relation:

Corollary A7.5 *(transformation rule) For any smooth function f, bilinear form b, and vector field u, we have*

$$\mathcal{L}_{fu} b = f \mathcal{L}_u b + df \otimes b(u, \cdot) + b(\cdot, u) \otimes df.$$

Proof: See Emery (1989), p. 19. □

Exercise

1. Explain in what sense we can think of ξ in Theorem A1.2 as a measurable selection. Further state the result in terms of random sets.

2. A partially ordered set I is said to be *linearly ordered* if, for any two elements $i, j \in I$, we have either $i \prec j$ or $j \prec i$. Give an example of a partially ordered set that is directed by not linearly ordered.

[27]and are presumably well-known to the experts

3. For a topological space S, give an example of a net in S that is not indexed by a linearly ordered set I.

4. Give an example of a Polish space that is not locally compact. (*Hint:* This is true for for many function spaces.)

5. Give an example of a convex functional on R^d that is not linear. (*Hint:* If M is a convex body in R^d, then the functional $p(x) = \inf\{r > 0;\ (x/r) \in M\}$ is convex. When is it linear?)

6. Give examples of two different norms in R^d. Then give the geometric meaning of a linear functional on R^d. Finally, for each of the norms, explain the meaning of the Hahn–Banach theorem, and show that the claimed extension of f may not be unique. (*Hint:* The l^p norms are different for $1 \le p \le \infty$. We may think of the dual as a space of vectors.)

7. For R^d with the usual 'dot' product, give examples of two different ONBs, and explain the geometric meaning of the statements in Lemma A3.7.

8. In the Euclidean space R^d, give the geometric and/or algebraic meaning of adjoints, unitary operators, and projections. Further explain in what sense geometric projections are self-adjoint and idempotent, and give the geometric meaning of the various propositions involving projection operators.

9. Prove Lemmas A4.7, A4.8, and A4.9. (*Hint:* Here Lemma A4.6 may be helpful.)

10. Show that the operator U in Theorem A4.10 is unitary, and prove the result. Then explain how the statement is related to Theorems 25.18 and 30.11.

11. Prove Lemma A6.5 (ii). (*Hint* (*Day*): First show that if $S_r \subset B$, then $B + S_\varepsilon \subset (1 + \varepsilon/r)B$, where S_r denotes an r-ball around 0.)

Notes and References

Here my modest aims are to trace the origins of some basic ideas in each chapter[1]*, to give references for the main results cited in the text, and to suggest some literature for further reading. No completeness is claimed, and knowledgeable readers will notice many errors and omissions, for which I apologize in advance. I also inserted some footnotes with short comments about the lives and careers of a few notable people*[2]*. Here my selection is very subjective, and many others could have been included.*

1. Sets and functions, measures and integration

The existence of non-trivial, countably additive measures was first discovered by Borel (1895/98), who constructed Lebesgue measure on the Borel σ-field in $\mathbb{R}$. The corresponding integral was introduced by Lebesgue[3] (1902/04), who also established the dominated convergence theorem. The monotone convergence theorem and Fatou's lemma were later obtained by Levi (1906a) and Fatou (1906), respectively. Lebesgue also introduced the higher-dimensional Lebesgue measure and proved a first version of Fubini's theorem, subsequently generalized by Fubini (1907) and Tonelli (1909). The integration theory was extended to general measures and abstract spaces by many authors, including Radon (1913) and Fréchet (1928).

The norm inequalities in Lemma 1.31 were first noted for finite sums by Hölder (1889) and Minkowski (1907), respectively, and were later extended to integrals by Riesz (1910). Part (i) for $p = 2$ goes back to Cauchy (1821) for finite sums and to Buniakowsky (1859) for integrals. The Hilbert space projection theorem can be traced back to Levi (1906b).

Monotone-class theorems were first established by Sierpiński (1928). Their use in probability theory goes back to Halmos (1950), Loève (1955), and Dynkin (1961). Mazurkiewicz (1916) proved that a topological space is Polish iff it is homeomorphic to a countable intersection of open sets in $[0, 1]^\infty$. The fact that every uncountable Borel set in a Polish space is Borel isomorphic to 2^∞ was proved independently by Alexandrov (1916) and Hausdorff (1916). Careful proofs of both results appear in Dudley (1989). Localized Borel spaces were used systematically in K(17).

Most results in this chapter are well known and can be found in any textbook on real analysis. Many graduate-level probability texts, such as Billingsley (1995) and Çinlar (2011), contain introductions to measure

[1]A comprehensive history of modern probability theory still remains to be written.

[2]I included only people who are no longer alive.

[3]Henri Lebesgue (1875–1941), French mathematician, along with Borel a founder of measure theory, which made modern probability possible.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1

theory. More advanced or comprehensive accounts are given, with numerous remarks and references, by Dudley (1989) and Bogachev (2007).

We are tacitly adopting ZFC or ZF+AC—the Zermelo–Fraenkel system amended with the Axiom of Choice—as the axiomatic foundation of modern analysis and probability. See, e.g., Dudley (1989) for a detailed discussion.

2. Measure extension and decomposition

As already mentioned, Borel (1895/98) was the first to prove the existence of one-dimensional Lebesgue measure. The modern construction via outer measures in due to Carathéodory (1918). Functions of bounded variation were introduced by Jordan (1881), who proved that any such function is the difference of two non-decreasing functions. The corresponding decomposition of signed measures was obtained by Hahn (1921). Integrals with respect to non-decreasing functions were defined by Stieltjes (1894), but their importance was not recognized until Riesz[4] (1909b) proved his representation theorem for linear functionals on $C[0,1]$. The a.e. differentiability of a function of bounded variation was first proved by Lebesgue (1904).

Vitali (1905) noted the connection between absolute continuity and the existence of a density. The Radon–Nikodym theorem was then proved in increasing generality by Radon (1913), Daniell (1920), and Nikodym (1930). A combined proof that also yields the Lebesgue decomposition was devised by von Neumann. Atomic decompositions and factorial measures appear in K(75/76) and K(83/86), respectively, and the present discussion is adapted from K(17).

Constructions of Lebesgue measure and proofs of the Radon–Nikodym theorem and Lebesgue's differentiation theorem are given in most textbooks on real analysis. Detailed accounts adapted to the needs of probabilists appear in, e.g., Loève (1977), Billingsley (1995), and Dudley (1989).

3. Kernels, disintegration, and invariance

Though kernels have long been used routinely in many areas of probability theory, they are hardly mentioned in most real analysis or early probability texts. Their basic properties were listed in K(97/02), where the basic kernel operations are also mentioned. Though the disintegration of measures on a product space is well known and due to Doob (1938), the associated disintegration formula is rarely stated and often used without proof, an early exception being K(75/76). Partial disintegrations go back to K(83/86). The theory of iterated disintegration was developed in K(10/11b), for applications to Palm measures.

Invariant measures on special groups were identified by explicit computation by many authors, including Hurwitz (1897). Haar (1933) proved the

[4]Riesz, Frigyes (1880–1956) and Marcel (1886–1969), Hungarian mathematicians and brothers, making fundamental contributions to functional analysis, potential theory, and other areas. Among their many students, we note Rényi (Frigyes) and Cramér (Marcel).

existence of invariant measures on a general lcscH group. Their uniqueness was then established by WEIL (1936/40), who also extended the theory to homogeneous spaces. The present representation of invariant measures, under a non-topological, proper group action, was established in K(11a). The usefulness of s-finite measures was recognized by SHARPE (1988) and GETOOR (1990).

Invariant disintegrations were first used to construct Palm distributions of stationary point processes. Here the regularization approach goes back to RYLL-NARDZEWSKI (1961) and PAPANGELOU (1974a), whereas the simple skew factorization was first noted by MATTHES (1963) and further developed by MECKE (1967) and TORTRAT (1969). The general factorization and regularization results for invariant measures on a product space were obtained in K(07).

The notions of strict and weak asymptotic invariance, going back to DOBRUSHIN (1956), were further studied by DEBES et al. (1970/71) and MATTHES et al. (1974/78). STONE (1968) noted that absolute continuity implies local invariance, and the latter notion was studied more systematically in K(78b).

The existence and uniqueness of Haar measures appears, along with much related material, in many textbooks on real or harmonic analysis, such as in HEWITT & ROSS (1979). The existence of regular conditional distributions is proved in most graduate-level probability texts, though often without a statement of the associated disintegration formula. For a more detailed and comprehensive account of the remaining material, we refer to K(17).

4. Processes, distributions, and independence

Countably additive probability measures were first used by BOREL[5] (1909/24), who constructed random variables as measurable functions on the Lebesgue unit interval and proved Theorem 4.18 for independent events. CANTELLI (1917) noticed that the 'easy' part remains true without the independence assumption. Lemma 4.5 was proved by JENSEN (1906), after a special case had been noted by HÖLDER.

The modern framework—with random variables as measurable functions on an abstract probability space $(\Omega, \mathcal{A}, P)$ and with expected values as P-integrals over Ω—was used implicitly by KOLMOGOROV from (1928) on. It was later formalized in the monograph of KOLMOGOROV[6] (1933c), which also contains the author's $0-1$ law, discovered long before the $0-1$ law of HEWITT

[5]ÉMILE BOREL (1871–1956), French mathematician, was the first to base probability theory on countably additive measures, 24 years before Kolmogorov's axiomatic foundations, which makes him the true founder of the subject.

[6]ANDREY KOLMOGOROV (1903–87), Russian mathematician, making profound contributions to many areas of mathematics, including probability. His 1933 paper was revolutionary, not only for the axiomatization of probability theory, which was essentially known since BOREL (1909), but also for his definition of conditional expectations, his $0-1$ law, and his existence theorem for processes with given finite-dimensional distributions. As a leader of the Moscow probability school, he had countless of students.

& Savage (1955).

Early work in probability theory deals mostly with properties depending only on the finite-dimensional distributions. Wiener (1923) was the first to construct the distribution of a process as a measure on a function space. The general continuity criterion in Theorem 4.23, essentially due to Kolmogorov, was first published by Slutsky (1937), with minor extensions added by Loève (1955) and Chentsov (1956). The search for general regularity properties was initiated by Doob (1937/47). It soon became clear, through the work of Lévy (1934/37), Doob (1951/53), and Kinney (1953), that most processes of interest have right-continuous versions with left-hand limits.

Introductions to measure-theoretic probability are given by countless authors, including Breiman (1968), Chung (1974), Billingsley (1995), and Çinlar (2011). Some more specific regularity properties are discussed in Loève (1977) and Cramér & Leadbetter (1967). Earlier texts tend to give more weight to distribution functions and their densities, less to measures and σ-fields.

5. Random sequences, series, and averages

The weak law of large numbers was first obtained by Bernoulli (1713), for the sequences named after him. More general versions were then established with increasing rigor by Bienaymé (1853), Chebyshev (1867), and Markov (1899). A necessary and sufficient condition for the weak law of large numbers was finally obtained by Kolmogorov (1928/29).

Khinchin & Kolmogorov (1925) studied series of independent, discrete random variables, and showed that convergence holds under the condition in Lemma 5.16. Kolmogorov (1928/29) then obtained his maximum inequality and showed that the three conditions in Theorem 5.18 are necessary and sufficient for a.s. convergence. The equivalence with convergence in distribution was later noted by Lévy (1937).

A strong law of large numbers for Bernoulli sequences was stated by Borel (1909a), but the first rigorous proof is due to Faber (1910). The simple criterion in Corollary 5.22 was obtained in Kolmogorov (1930). In (1933a) Kolmogorov showed that existence of the mean is necessary and sufficient for the strong law of large numbers, for general i.i.d. sequences. The extension to exponents $p \neq 1$ is due to Marcinkiewicz & Zygmund (1937). Proposition 5.24 was proved in stages by Glivenko (1933) and Cantelli (1933).

Riesz (1909a) introduced the notion of convergence in measure, for probability measures equivalent to convergence in probability, and showed that it implies a.e. convergence along a subsequence. The weak compactness criterion in Lemma 5.13 is due to Dunford (1939). The functional representation of Proposition 5.32 appeared in K(96a), and Corollary 5.33 was given by Stricker & Yor (1978).

The theory of weak convergence was founded by Alexandrov (1940/43), who proved in particular the 'portmanteau' Theorem 5.25. The continuous mapping Theorem 5.27 was obtained for a single function $f_n \equiv f$ by MANN & WALD (1943), and then in general by PROHOROV (1956) and RUBIN. The coupling Theorem 5.31 is due for complete S to SKOROHOD (1956) and in general to DUDLEY (1968).

More detailed accounts of the material in this chapter may be found in many textbooks, such as in LOÈVE (1977) and CHOW & TEICHER (1997). Additional results on random series and a.s. convergence appear in STOUT (1974) and KWAPIEŃ & WOYCZYŃSKI (1992).

6. Gaussian and Poisson convergence

The central limit theorem—so first called by PÓLYA (1920)—has a long and glorious history, beginning with the work of DE MOIVRE[7] (1738), who obtained the familiar approximation of binomial probabilities in terms of the normal density. LAPLACE (1774, 1812/20) stated a general version of the central limit theorem, but his proof was incomplete, as was the proof of CHEBYSHEV (1867/90). The normal laws were used by GAUSS (1809/16) to model the distribution of errors in astronomical observations.

The first rigorous proof was given by LIAPOUNOV (1901), though under an extra moment condition. Then LINDEBERG (1922a) proved his fundamental Theorem 6.13, which led in turn to the basic Proposition 6.10, in a series of papers by LINDEBERG (1922b) and LÉVY (1922a–c). BERNSTEIN (1927) obtained the first extension to higher dimensions. The general problem of normal convergence, regarded for two centuries as the central (indeed the only nontrivial) problem in probability, was eventually solved in the form of Theorem 6.16, independently by FELLER (1935) and LÉVY (1935a).

Many authors have proved local versions of the central limit theorem, the present version being due to STONE (1965). The domain of attraction to the normal law was identified by LÉVY, FELLER, and KHINCHIN. General laws of attraction were studied by DOEBLIN (1947, posth.), who noticed the connection with KARAMATA's (1930) theory of regular variation.

Even the simple Poisson approximation of binomial probabilities was noted by DE MOIVRE (1711). It was rediscovered by POISSON (1837), who also noted that the limits form a probability distribution in its own right. Applications of this 'law of small numbers' were compiled by COURNOT (1843) and BORTKIEWICZ[8] (1898).

Though characteristic functions were used in probability already by LAPLACE (1812/20), their first *rigorous* use to prove a limit theorem is credited to LIAPOUNOV (1901). The first general continuity theorem was established by LÉVY (1922c), who assumed the characteristic functions to converge uniformly in a neighborhood of the origin. The simpler criterion in Theorem 6.23 is due to BOCHNER (1933). Our direct approach to Theorem 6.3 may be new, in

---

[7]ABRAHAM DE MOIVRE (1667–1754), French mathematician. To escape the persecution of huguenots he fled to England, where he befriended Newton. He discovered both the normal and the Poisson approximation to the binomial distribution.

[8]Known for the example of the number of deaths by horse kicks in the Prussian army.

avoiding the relatively deep selection theorem of Helly (1911/12). The basic Corollary 6.5 was noted by Cramér & Wold (1936).

Introductions to characteristic functions and classical limit theorems may be found in many textbooks, notably Loève (1977). Feller (1971) is a rich source of further information on Laplace transforms, characteristic functions, and classical limit theorems. For more detailed or advanced results on characteristic functions, see Lukacs (1970).

7. Infinite divisibility and general null arrays

The scope of the classical central limit problem was broadened by Lévy (1925) to encompass the study of suitably normalized partial sums, leading to the classes of stable and self-decomposable limiting distributions. To include the case of the classical Poisson approximation, Kolmogorov proposed a further extension to general triangular arrays, subject to the sole condition of uniform asymptotic negligibility, allowing convergence to general infinitely divisible laws. de Finetti (1929) saw how the latter distributions are related to processes with stationary, independent increments, and posed the problem of finding their general form.

Here the characteristic functions were first characterized by Kolmogorov (1932), though only under a moment condition that was later removed by Lévy (1934/35). Khinchin (1937/38) noted the potential for significant simplifications in the one-dimensional case, leading to the celebrated *Lévy–Khinchin formula*. The probabilistic interpretation in terms of Poisson processes is essentially due to Itô (1942b). The simple form for non-negative random variables was noted by Jiřina (1964).

For the mentioned limit problem, Feller (1937) and Khinchin (1937) proved independently that all limiting distributions are infinitely divisible. It remained to characterize the convergence to specific limits, continuing the work by Feller (1935) and Lévy (1935a) for Gaussian limits. The general criteria were found independently by Doeblin (1939a) and Gnedenko (1939). Convergence criteria for infinitely divisible distributions on R_+^d were given by Jiřina (1966) and Nawrotzki (1968). Corresponding criteria for null-arrays were obtained in K(75/76), after the special case of Z_+-valued variables had been settled by Kerstan & Matthes (1964). Extreme value theory is surveyed by Leadbetter et al. (1983). Lemma 7.16 appears in Doeblin (1939a).

The limit theory for null arrays of random variables was once regarded as the central topic of probability theory, and comprehensive treatments were given by Gnedenko & Kolmogorov (1954/68), Loève (1977), and Petrov (1995). The classical approach soon got a bad reputation, because of the numerous technical estimates involving characteristic and distribution functions, and the subject is often omitted in modern textbooks.[9] Eventually, Feller

[9]This is a huge mistake, not only since characteristic functions provide the easiest approach to the classical limit theorems, but also for their crucial role in connection with

(1966/71) provided a more readable, even enjoyable account. The present concise probabilistic approach, based on compound Poisson approximations and methods of modern measure theory, was pioneered in K(97). It has the further advantage of extending easily to the multi-variate case.

8. Conditioning and disintegration

Though conditional densities have been computed by statisticians ever since Laplace (1774), a general approach to conditioning was not devised until Kolmogorov (1933c), who defined conditional probabilities and expectations as random variables on the basic probability space, using the Radon–Nikodym theorem, which had recently become available. His original notion of conditioning with respect to a random vector was extended by Halmos (1950) to arbitrary random elements, and then by Doob (1953) to general sub-σ-fields.

Our present Hilbert space approach to conditioning, essentially due to von Neumann (1940), is simpler and more suggestive, in avoiding the use of the less intuitive Radon–Nikodym theorem. It has the further advantage of leading to the attractive interpretation of martingales as projective families of random variables.

The existence of regular conditional distributions was studied by several authors, beginning with Doob (1938). It leads immediately to the often used but rarely stated disintegration Theorem 8.5, extending Fubini's theorem to the case of dependent random variables. The fundamental importance of conditional distributions is not always appreciated.[10]

The interpretation of the simple Markov property in terms of conditional independence was indicated already by Markov (1906), and the formal statement of Proposition 8.9 appears in Doob (1953). Further properties of conditional independence have been listed by Döhler (1980) and others. Our statement of conditional independence in terms of Hilbert-space projections is new. The subtle device of iterated conditioning was developed in K(10/11b).

The fundamental transfer Theorem 8.17, noted by many authors, may have been stated explicitly for the first time by Aldous (1981). It leads in particular to a short proof of Daniell's (1918/19/20) existence theorem for sequences of random elements in a Borel space, which in turn implies Kolmogorov's (1933c) celebrated extension to continuous time and arbitrary index sets. Łomnicki & Ulam (1934) noted that no topological assumptions are needed for independent random elements, a result that was later extended by Ionescu Tulcea (1949/50) to measures specified by a sequence of conditional distributions.

The traditional Radon–Nikodym approach to conditional expectations appears in most textbooks, such as in Billingsley (1995). Our projection approach is essentially the one used by Dellacherie & Meyer (1975) and

---

martingales, Lévy processes, random measures, and potentials.

[10]Dellacherie & Meyer (1975) write: "[They] have a bad reputation, and probabilists rarely employ them without saying 'we are obliged' to use conditional distributions . . . ." We may hope that times have changed, and their importance is now universally recognized.

Elliott (1982).

9. Optional times and martingales

Martingales were first introduced by Bernstein[11] (1927/37), in his efforts to relax the independence assumption in the classical limit theorems. Both Bernstein and Lévy (1935a-b/37) extended Kolmogorov's maximum inequality and the central limit theorem to a general martingale context. The *term* martingale (originally denoting part of a horse's harness and later used for a special gambling system) was introduced in the probabilistic context by Ville (1939).

The first martingale convergence theorem was obtained by Jessen[12] (1934) and Lévy (1935b), both of whom proved Theorem 9.24 for filtrations generated by a sequence of independent random variables. A sub-martingale version of the same result appears in Sparre-Andersen & Jessen (1948). The independence assumption was removed by Lévy (1937/54), who also noted the simple martingale proof of Kolmogorov's $0-1$ law and proved his conditional version of the Borel–Cantelli lemma.

The general convergence theorem for discrete-time martingales was proved by Doob[13] (1940), and the basic regularity theorems for continuous-time martingales first appeared in Doob (1951). The theory was extended to submartingales by Snell (1952) and Doob (1953). The latter book is also the original source of such fundamental results as the martingale closure theorem, the optional sampling theorem, and the L^p-inequality.

Though hitting times have long been used informally, general optional times (often called stopping times) seem to appear for the first time in Doob (1936). Abstract filtrations were not introduced until Doob (1953). Progressive processes were introduced by Dynkin (1961), and the modern definition of the σ-fields $\mathcal{F}_\tau$ is due to Yushkevich.

The first general exposition of martingale theory, truly revolutionary for its time, was given by Doob (1953). Most graduate-level texts, such as Çinlar (2011), contain introductions to martingales, filtrations, and optional times. More extensive coverage of the discrete-time theory appears in Neveu (1975), Chow & Teicher (1997), and Rogers & Williams (1994). The encyclopedic treatises of Dellacherie & Meyer (1975/80) remain standard references for the continuous-time theory.

10. Predictability and compensation

Predictable and totally inaccessible times appear implicitly, along with

[11]Sergei Bernstein (1880–1968), Ukrainian and Sovjet mathematician working in analysis and probability, invented martingales and proved martingale versions of some classical results for independent variables.

[12]Børge Jessen (1907–93), Danish mathematician working in analysis and geometry, proved the first martingale convergence theorem, long before Doob.

[13]Joseph Doob (1910–2004), American mathematician and a founder of modern probability, known for path-breaking work on martingales and probabilistic potential theory.

quasi-left continuous processes, in the work of Blumenthal (1957) and Hunt (1957/58). A systematic study of optional times and their associated σ-fields was initiated by Chung & Doob (1965), Meyer (1966/68), and Doléans (1967a). Their ideas were further developed by Dellacherie (1972), Dellacherie & Meyer (1975/80), and others into a 'general theory of processes', which has in many ways revolutionized modern probability.

The connection between excessive functions and super-martingales, noted by Doob (1954), suggested a continuous-time extension of the elementary Lemma 9.10. Such a result was eventually proved by Meyer[14] (1962/63) in the form of Lemma 10.7, after special decompositions in the Markovian context had been obtained by Volkonsky (1960) and Shur (1961). Doléans (1967) proved the equivalence of natural and predictable processes, leading to the ultimate version of the Doob–Meyer decomposition in the form of Theorem 10.5.

The original proof, appearing in Dellacherie (1972) and Dellacherie & Meyer (1975/80), is based on some deep results in capacity theory. The present more elementary approach combines Rao's (1969) simple proof of Lemma 10.7, based on Dunford's (1939) weak compactness criterion, with Doob's (1984) ingenious approximation of totally inaccessible times. Itô & Watanabe (1965) noted the extension to general sub-martingales, by suitable localization. The present proof of Theorem 10.14 is taken from Chung & Walsh (1974). The moment inequality in Proposition 10.20 was proved independently by Garsia (1973) and Neveu (1975), after a special case had been obtained by Burkholder et al. (1972).

Compensators of optional times first arose as *hazard functions* in reliability theory. More general compensators were later studied in the Markovian context by S. Watanabe (1964) under the name of *Lévy systems*. Grigelionis (1971) and Jacod (1975) constructed the compensators of adapted random measures on a product space $\mathsf{R}_+ \times S$. The elementary formula for the induced compensator of a random pair (τ, χ) is classical and appears in, e.g., Dellacherie (1972) and Jacod (1975). The equation $Z = 1 - Z_- \cdot \bar\eta$ was studied in greater generality by Doléans (1970). Discounted compensators were introduced in K(90), which contains the fundamental martingale and mapping results like Theorem 10.27.

Induced compensators of point processes have been studied by many authors, including Last & Brandt (1995) and Jacobsen (2006). More advanced and comprehensive accounts of the 'general theory' are given by Dellacherie (1972), Elliott (1982), Rogers & Williams (1987/2000), and Dellacherie & Meyer (1975/80).

---

[14] Paul-André Meyer (1934–2003), French mathematician, making profound contributions to martingales, stochastic integration, Markov processes, potential theory, Malliavin calculus, quantum probability, and stochastic differential geometry. As a leader of the famous Strasbourg school of probability, his work inspired probabilists all over the world.

11. Markov properties and discrete-time chains

Discrete-time Markov chains in a finite state space were introduced by MARKOV[15] (1906), who proved the first ergodic theorem and introduced the notion of irreducible chains. KOLMOGOROV (1936a/b) extended the theory to a countable state space and general transition probabilities. In particular, he obtained a decomposition of the state space into irreducible sets, classified the states with respect to recurrence and periodicity, and described the asymptotic behavior of the n-step transition probabilities. Kolmogorov's original proofs were analytic. The more intuitive coupling approach was introduced by DOEBLIN (1938a/b), long before the strong Markov property was formalized.

BACHELIER had noted the connection between random walks and diffusions, which inspired KOLMOGOROV (1931a) to give a precise definition of Markov processes in continuous time. His treatment is purely analytic, with the distribution specified by a family of transition kernels satisfying the Chapman–Kolmogorov relation, previously noted in special cases by CHAPMAN (1928) and SMOLUCHOVSKY. KOLMOGOROV makes no reference to sample paths. The transition to probabilistic methods began with the work of LÉVY (1934/35) and DOEBLIN (1938a/b).

Though the strong Markov property had been used informally since BACHELIER (1900/01), it was first proved rigorously in a special case by DOOB (1945). The elementary inequality of OTTAVIANI is from (1939). General filtrations were introduced in Markov process theory by BLUMENTHAL (1957). The modern setup, with a canonical process X defined on the path space Ω, equipped with a filtration $\mathcal{F}$, a family of shift operators θ_t, and a collection of probability measures P_x, was developed systematically by DYNKIN (1961/65). A weaker form of Theorem 11.14 appears in BLUMENTHAL & GETOOR (1968), and the present version is from K(87).

Most graduate-level texts contain elementary introductions to Markov chains in discrete and continuous time. More extensive discussions appear in FELLER (1968/71) and FREEDMAN (1971a). Most literature on general Markov processes is fairly advanced, leading quickly into discussions of semi-group and potential theory, here treated in Chapters 17 and 34. Standard references include DYNKIN[16] (1965), BLUMENTHAL & GETOOR (1968), SHARPE (1988), and DELLACHERIE & MEYER (1987/92). The coupling method, soon forgotten after Doeblin's tragic death[17], was revived by LINDVALL (1992) and

[15] ANDREI MARKOV (1856–1922), Russian mathematician, inventor of Markov processes.

[16] EUGENE DYNKIN (1924–2014), Russian-American mathematician, worked on Lie groups before turning to probability. Laid the foundations of the modern theory of Markov processes, and made seminal contributions to the theory of super-processes.

[17] WOLFGANG DOEBLIN (1915–40), born in Berlin to the famous author Alfred Döblin. When Hitler came to power, his Jewish family fled to Paris. After a short career inspired by Lévy and Fréchet, he joined the French army to fight the Nazis. When his unit got cornered, he shot himself, to avoid getting caught and sent to an extermination camp. Aged 25, he had already made revolutionary contributions to many areas of probability, including construction of diffusions with given drift and diffusion rates, time-change reduction of martingales to Brownian motion, solution of the general limit problem for null arrays, and recurrence

THORISSON (2000).

12. Random walks and renewal processes

Random walks arose early in a wide range of applications, such as in gambling, queuing, storage, and insurance; their history can be traced back to the origins of probability. Already BACHELIER (1900/01) used random walks to approximate diffusion processes. Such approximations were used in potential theory in the 1920s, to yield probabilistic interpretations in terms of a simple symmetric random walk. Finally, random walks played an important role in the sequential analysis developed by WALD (1947).

The modern developments began with PÓLYA's (1921) discovery that a simple symmetric random walk in Z^d is recurrent for $d \leq 2$, otherwise transient. His result was later extended to Brownian motion by LÉVY (1940) and KAKUTANI (1944a). The general recurrence criterion in Theorem 12.4 was obtained by CHUNG & FUCHS (1951), and the probabilistic approach to Theorem 12.2 was found by CHUNG & ORNSTEIN (1962). The first condition in Corollary 12.7 is even necessary for recurrence, as noted independently by ORNSTEIN (1969) and STONE (1969). WALD (1944) discovered the equations named after him.

The reflection principle was first used by ANDRÉ (1887) in the context of the ballot problem. The systematic study of fluctuation and absorption problems for random walks began with the work of POLLACZEK (1930). Ladder times and heights, first introduced by BLACKWELL, were explored in an influential paper by FELLER (1949). The factorizations in Theorem 12.16 were originally derived by the Wiener–Hopf technique, developed by PALEY & WIENER (1934) as a general tool in Fourier analysis. Theorem 12.17 is due for $u = 0$ to SPARRE-ANDERSEN (1953/54) and in general to BAXTER (1961). The intricate combinatorial methods used by the former author were later simplified by FELLER and others.

Though renewals in Markov chains are implicit already in some early work of KOLMOGOROV and LÉVY, the general renewal process may have been first introduced by PALM (1943). The first renewal theorem was obtained by ERDÖS et al. (1949) for random walks in Z_+, though the result is then an easy consequence of KOLMOGOROV's (1936a/b) ergodic theorem for Markov chains on a countable state space, as noted by CHUNG. BLACKWELL (1948/53) extended the result to random walks in R_+. The ultimate version for transient random walks in R is due to FELLER & OREY (1961). The first coupling proof of Blackwell's theorem was given by LINDVALL (1977). Our proof is a modification of an argument in ATHREYA et al. (1978), which originally did not cover all cases. The method seems to require the existence of a possibly infinite mean. An analytic approach to the general case appears in FELLER (1971).

Elementary introductions to random walks are given by many authors, including CHUNG (1974) and FELLER (1971). A detailed treatment of random

---

and ergodic properties of Markov processes.

walks in Z^d is given by Spitzer (1976). Further aspects of renewal theory are discussed in Smith (1954).

13. Jump-type chains and branching processes

Continuous-time Markov chains have been studied by many authors, beginning with Kolmogorov (1931a). The transition functions of general pure jump-type Markov processes were explored by Pospišil (1935/36) and Feller (1936/40), and the corresponding path properties were analysed by Doeblin (1939b) and Doob (1942b). The first continuous-time version of the strong Markov property was obtained by Doob (1945). The equivalence of the counting and renewal descriptions of Poisson processes was established by Bateman (1910). The significance of pseudo-Poisson processes was recognized by Feller.

The classical, discrete-time branching process was first studied by Bienaymé[18] (1845), who found in particular the formula for the extinction probability. His results were partially rediscovered by Watson & Galton[19] (1874), after whom the process has traditionally been named. In the critical case, the asymptotic survival probability and associated distribution were found by Kolmogorov (1938) and Yaglom (1947), and the present comparison argument is due to Spitzer. The transition probabilities of a birth & death process with rates $n\mu$ and $n\lambda$ were obtained by Palm (1943) and Kendall (1948). The associated ancestral process was identified in K(17).

The diffusion limit of a Bienaymé process was obtained by Feller (1951). The associated space-time process, known as a *super-process*, was discovered by Watanabe (1968) and later studied extensively by Dawson (1977/93) and others. The corresponding ancestral structure was uncovered in profound work by Dynkin (1991), Dawson & Perkins (1991), and Le Gall (1991).

Comprehensive accounts of pure-jump type Markov processes have been given by many authors, beginning with Doob 1953), Chung (1960), and Kemeny et al. (1966). The underlying regenerative structure was analyzed by Kingman (1972). Countless authors have discussed applications to queuing[20] theory and other areas, including Takács (1962). Detailed accounts of Bienaymé and related processes include those by Harris (1963), Athreya & Ney (1972), and Jagers (1975). Etheridge (2000) gives an accessible introduction to super-processes.

14. Gaussian processes and Brownian motion

As we have seen, the Gaussian distribution first arose in the work of de Moivre (1738) and Laplace (1774, 1812/20), and was popularized through

[18] Irénée-Jules Bienaymé (1796–1878), French probabilist and statistician.

[19] The latter became infamous as a founder of *eugenics*—the 'science' of breeding a human master race—later put in action by Hitler and others. Since his results were either wrong or already known, I would be happy to drop his name from the record.

[20] I prefer the traditional spelling, as in *ensuing* and *pursuing*.

GAUSS' (1809) theory of errors. MAXWELL[21] (1875/78) derived the Gaussian law as the velocity distribution for the molecules in a gas, assuming the hypotheses of Proposition 14.2. FREEDMAN (1962/63) used de Finetti's theorem to characterize sequences and processes with rotatable symmetries. His results were later recognized as equivalent to a relationship between positive definite and completely monotone functions proved by SCHOENBERG (1938). Isonormal Gaussian processes were introduced by SEGAL (1954).

The Brownian motion *process*, first introduced by THIELE (1880), was rediscovered and used by BACHELIER[22] (1900) to model fluctuations on the stock market. Bachelier also noted several basic properties of the process, such as the distribution of the maximum and the connection with the heat equation. His work, long neglected and forgotten, was eventually revived by Kolmogorov and others. He is now recognized as the founder of mathematical finance.

The mathematical notion must not be confused with the *physical phenomenon* of Brownian motion—the irregular motion of pollen particles suspended in water. Such a motion, first noted by VAN LEEUWENHOEK[23], was originally thought to be a biological phenomenon, motivating some careful studies during different seasons, meticulously recorded by the botanist BROWN (1828). Eventually, some evidence from physics and chemistry suggested the existence of atoms[24], explaining Brownian motion as caused by the constant bombardment by water molecules. The atomic hypothesis was still controversial, when EINSTEIN (1905/06), ignorant of Bachelier's work, modeled the motion by the mentioned process, and used it to estimate the size of the molecules[25]. A more refined model for the physical Brownian motion was proposed by LANGEVIN (1908) and ORNSTEIN & UHLENBECK (1930).

A rigorous study of the Brownian motion process was initiated by WIENER[26] (1923), who constructed its distribution as a measure on the space of continuous paths. The significance of Wiener's revolutionary paper was not appreciated until after the pioneering work of KOLMOGOROV (1931a/33b), LÉVY (1934/35), and FELLER (1936). The fact that the Brownian paths are nowhere differentiable was noted in a celebrated paper by PALEY et al. (1933).

Wiener also introduced stochastic integrals of deterministic L^2-functions, later studied in further detail by PALEY et al. (1933). The spectral repre-

---

[21]JAMES CLERK MAXWELL (1831–79), Scottish physicist, known for his equations describing the electric and magnetic fields, and a co-founder of statistical mechanics.

[22]LOUIS BACHELIER (1870–1946), French mathematician, the founder of mathematical finance and discoverer of the Brownian motion process.

[23]ANTONIE VAN LEEUWENHOEK (1632–1723), Dutch scientist, who was the first to construct a microscope powerful enough to observe the micro-organisms in 'plain' water.

[24]A similar claim by Democritus can be dismissed, as it was based on a logical fallacy.

[25]This is one of the three revolutionary achievements of ALBERT EINSTEIN (1879–1955) during his 'miracle year' of 1905, the others being his discovery of special relativity, including the celebrated formula $E = mc^2$, and his explanation of the photo-electric effect, marking the beginning of quantum mechanics and earning him a Nobel prize in 1921.

[26]NORBERT WIENER (1894–1964), famous American mathematician, making profound contributions to numerous areas of analysis and probability.

sentation of stationary processes, originally deduced from Bochner's (1932) theorem by Cramér (1942), was later recognized as equivalent to a general Hilbert space result of Stone (1932). The chaos expansion of Brownian functionals was discovered by Wiener (1938), and the theory of multiple integrals with respect to Brownian motion was developed in a seminal paper of Itô (1951c).

The law of the iterated logarithm was discovered by Khinchin[27], first (1923/24) for Bernoulli sequences, and then (1933b) for Brownian motion. A more detailed study of Brownian paths was pursued by Lévy (1948), who proved the existence of quadratic variation in (1940) and the arcsine laws in (1948). Though many proofs of the latter have since been given, the present deduction from basic symmetry properties may be new. Though the strong Markov property was used implicitly by Lévy and others, the result was not carefully stated and proved until Hunt (1956).

Most modern probability texts contain introductions to Brownian motion. The books by Itô & McKean (1965), Freedman (1971b), Karatzas & Shreve (1991), and Revuz & Yor (1999) give more detailed information on the subject. Further discussion of multiple Wiener–Itô integrals appears in Kallianpur (1980), Dellacherie et al. (1992), and Nualart (2006). The advanced theory of Gaussian processes is surveyed by Adler (1990).

15. Poisson and related processes

As we have seen, the Poisson distribution was recognized by de Moivre (1711) and Poisson (1837) as an approximation to the binomial distribution. In a study of Bible chronology, Ellis (1844) obtains Poisson processes as approximations of more general renewal processes. Point processes with stationary, independent increments were later considered by Lundberg (1903) to model a stream of insurance claims, by Rutherford & Geiger (1908) to describe the process of radioactive decay, and by Erlang (1909) to model the incoming calls to a telephone exchange. Here the first rigorous proof of the Poisson distribution for the increments was given by Bateman (1910), who also showed that the holding times are i.i.d. and exponentially distributed. The Poisson property in the non-homogeneous case is implicit in Lévy (1934/35), but it was not rigorously proven until Copeland & Regan (1936).

The mixed binomial representation of Poisson processes is used implicitly by Wiener (1938), in his construction of higher-dimensional Poisson processes. It was stated more explicitly by Moyal (1962), leading to the elementary construction of general Poisson processes, noted by both Kingman (1967) and Mecke (1967). Itô (1942b) noted that a marked point process with stationary, independent increments is equivalent to a Poisson process in the plane. The corresponding result in the non-homogeneous case eluded Itô, but

[27] Alexandr Khinchin (1894–1959), Russian mathematician, along with Kolmogorov a leader of the Moscow school of analysis and probability. He wrote monographs on statistical mechanics, information theory, and queuing theory.

follows from Kingman's (1967) general description of random measures with independent increments.

The mapping and marking properties of a Poisson process were established by Prékopa (1958), after partial results and special cases had been noted by Bartlett (1949) and Doob (1953). Cox processes, first introduced by Cox (1955), were studied extensively by Kingman (1964), Krickeberg (1972), and Grandell (1976). Thinnings were first considered by Rényi[28] (1956). Laplace transforms of random measures were used systematically by von Waldenfels (1968) and Mecke (1968); their use to construct Cox processes and transforms relies on a lemma of Jiřina (1964). One-dimensional uniqueness criteria were obtained in the Poisson case by Rényi (1967), and then in general independently by Mönch (1971) and in K(73a/83), with further extensions by Grandell (1976). The symmetry characterization of mixed Poisson and binomial processes was noted independently by Matthes et al. (1974), Davidson (1974b), and K(73a).

S. Watanabe (1964) proved that a ql-continuous simple point process is Poisson iff its compensator is deterministic, a result later extended to general marked point processes by Jacod (1975). A corresponding time-change reduction to Poisson was obtained independently by Meyer (1971) and Papangelou (1972); a related Poisson approximation was obtained by Brown (1978/79). The extended time-change result in Theorem 15.15 was obtained in K(90).

Multiple Poisson integrals were first mentioned by Itô (1956), though some underlying ideas can be traced back to papers by Wiener (1938) and Wiener & Wintner (1943). Many sophisticated methods have been applied through the years to the study of such integrals. The present convergence criterion for double Poisson integrals is a special case of some general results for multiple Poisson and Lévy integrals, established in K & Szulga (1989).

Introductions to Poisson and related processes are given by Kingman (1993) and Last & Penrose (2017). More advanced and comprehensive accounts of random measure theory appear in Matthes et al. (1978), Daley & Vere-Jones (2003/08), and K(17).

16. Independent-increment and Lévy processes

Up to the 1920s, Brownian motion and the Poisson process were essentially the only known processes with independent increments. In (1924/25) Lévy discovered the stable distributions and noted that they, too, could be associated with suitable 'decomposable' processes. de Finetti (1929) saw the connection between processes with independent increments and infinitely divisible distributions, and posed the problem of characterizing the latter. As noted in Chapter 7, the problem was solved by Kolmogorov (1932) and Lévy (1934/35).

[28] Alfréd Rényi (1921–70), Hungarian mathematician. Through his passion for the subject, he inspired a whole generation of mathematicians.

The paper of Lévy[29] (1934/35) was truly revolutionary[30], in giving at the same time a complete description of processes with independent increments. Though his intuition was perfectly clear, his reasoning was sometimes a bit sketchy, and a more careful analysis was given by Itô (1942b). In the meantime, analytic proofs of the representation formula for the characteristic function were given by Lévy himself (1937), by Feller (1937), and by Khinchin (1937). The Lévy–Khinchin formula for the characteristic function was extended by Jacod (1983) to the case of possibly fixed discontinuities. Our probabilistic description in the general case is new and totally different.

The basic convergence Theorem 16.14 for Lévy processes and the associated approximation of random walks in Corollary 16.17 are essentially due to Skorohod (1957), though with rather different statements and proofs. The special case of stable processes has been studied by countless authors, ever since Lévy (1924/25). Versions of Proposition 16.19 were discovered independently by Rosiński & Woyczyński (1986) and in K(92).

The local characteristics of a semi-martingale were introduced independently by Grigelionis (1971) and Jacod (1975). Tangent processes, first mentioned by Itô (1942b), were used by Jacod (1984) to extend the notion of semi-martingales. The existence of tangential sequences with conditionally independent terms was noted by Kwapień & Woyczyński (1992), and a careful proof appears in de la Peña & Giné (1999). A discrete-time version of the tangential comparison was obtained by Zinn (1986) and Hitchenko (1988), and a detailed proof appears in Kwapień & Woyczyński (1992). Both results were extended to continuous time in K(17a). The simplified comparison criterion may be new.

Modern introductions to Lévy processes have been given by Bertoin (1996) and Sato (2013). The latter gives a detailed account of the Lévy–Itô representation in the stochastically continuous case. The analytic approach in the general case is carefully explained by Jacod & Shiryaev (1987), which also contains an exposition of the associated limit theory.

17. Feller processes and semi-groups

Semi-group ideas are implicit in Kolmogorov's pioneering (1931a) paper, whose central theme is to identify some local characteristics determining the transition probabilities through a system of differential equations, the so-called Kolmogorov forward and backward equations. Markov chains and diffusion processes were originally treated separately, but in (1935) Kolmogorov proposed a unified framework, with transition kernels regarded as operators (initially operating on measures rather than on functions), and with local char-

[29] Paul Lévy (1886–1971), French mathematician and a founder of modern probability, making profound contributions to the central limit theorem, stable processes, Brownian motion, Lévy processes, and local time.

[30] Loève (1978), who had been a student of Lévy, writes (alluding to Greek mythology): "Decomposability sprang forth fully armed from the forehead of P. Lévy ... His analysis ... was so complete that since then only improvements of detail have been added."

acteristics given by an associated generator.

Kolmogorov's ideas were taken up by FELLER[31] (1936), who derived existence and uniqueness criteria for the forward and backward equations. The general theory of contraction semi-groups on a Banach space was developed independently by HILLE (1948) and YOSIDA (1948), both of whom recognized its significance for the theory of Markov processes. The power of the semi-group approach became clear through the work of FELLER (1952/54), who gave a complete description of the generators of one-dimensional diffusions, in particular characterizing the boundary behavior of the process in terms of the domain of the generator.

A systematic study of Markov semi-groups was now initiated by DYNKIN (1955a). The usual approach is to postulate the strong continuity (F_3), instead of the weaker and more easily verified condition (F_2). The positive maximum principle first appeared in the work of ITÔ (1957), and the core condition in Proposition 17.9 is due to WATANABE (1968).

The first regularity theorem was obtained by DOEBLIN (1939b), who gave conditions for the paths to be step functions. A sufficient condition for continuity was then obtained by FORTET (1943). Finally, KINNEY (1953) showed that any Feller process has a version with rcll paths, after DYNKIN (1952) had derived the same property under a Hölder condition. The use of martingale methods in the study of Markov processes dates back to KINNEY (1953) and DOOB (1954).

The strong Markov property for Feller processes was proved independently by DYNKIN & YUSHKEVICH (1956) and BLUMENTHAL (1957), after special cases had been obtained by DOOB (1945), HUNT (1956), and RAY (1956). BLUMENTHAL's (1957) paper also contains his $0-1$ law. DYNKIN (1955a) introduced his *characteristic operator*, and a version of Theorem 17.24 appears in DYNKIN (1956).

There is a vast literature on the approximation of Markov chains and Markov processes, covering a wide range of applications. The use of semi-group methods to prove limit theorems can be traced back to LINDEBERG's (1922a) proof of the central limit theorem. The general results of Theorems 17.25 and 17.28 were developed in stages by TROTTER (1958a), SOVA (1967), KURTZ (1969/75), and MACKEVIČIUS (1974). Our proof of Theorem 17.25 uses ideas from GOLDSTEIN (1976).

A splendid introduction to semi-group theory is given by the relevant chapters in FELLER (1971). In particular, Feller shows how the one-dimensional Lévy–Khinchin formula and associated limit theorems can be derived by semi-group methods. More detailed and advanced accounts of the subject appear in DYNKIN (1965), ETHIER & KURTZ (1986), and DELLACHERIE & MEYER (1987/92).

---

[31] WILLIAM FELLER (1906–70), Croatian-American mathematician, making path-breaking contributions to many areas of analysis and probability, including classical limit theorems and diffusion processes. His introductory volumes on probability theory, arguably the best probability books ever written, became bestsellers and were translated to many languages.

18. Itô integration and quadratic variation

The first stochastic integral with a random integrand was defined by Itô[32] (1942a/44), using Brownian motion as the integrator and product measurable and adapted processes as integrands. Doob (1953) made the connection with martingale theory. A version of the fundamental substitution rule was proved by Itô (1951a), and later extended by many authors. The compensated integral in Corollary 18.21 was introduced independently by Fisk (1966) and Stratonovich (1966).

The quadratic variation process was originally obtained from the Doob–Meyer decomposition. Fisk (1966) noted how it can also be obtained directly from the process, as in Proposition 18.17. Our present construction was inspired by Rogers & Williams (2000). The BDG inequalities were originally proved for $p > 1$ and discrete time by Burkholder (1966). Millar (1968) noted the extension to continuous martingales, in which context the further extension to arbitrary $p > 0$ was obtained independently by Burkholder & Gundy (1970) and Novikov (1971). Kunita & Watanabe (1967) introduced the covariation of two martingales and used it to characterize the integral. They further established some general inequalities related to Proposition 18.9.

Itô's original integral was extended to square-integrable martingales by Courrège (1962/63) and Kunita & Watanabe (1967), and then to continuous semi-martingales by Doléans & Meyer (1970). The idea of localization is due to Itô & Watanabe (1965). Theorem 18.24 was obtained by Kazamaki (1972), as part of a general theory of random time change. Stochastic integrals depending on a parameter were studied by Doléans (1967b) and Stricker & Yor (1978), and the functional representation in Proposition 18.26 first appeared in K(96a).

Elementary introductions to Itô calculus and its applications appear in countless textbooks, including Chung & Williams (1990) and Klebaner (2012). For more advanced accounts and further information, we recommend especially Ikeda & Watanabe (2014), Rogers & Williams (2000), Karatzas & Shreve (1998), and Revuz & Yor (1999).

19. Continuous martingales and Brownian motion

The fundamental characterization of Brownian motion in Theorem 19.3 was proved by Lévy (1937), who also in (1940) noted the conformal invariance up to a time change of complex Brownian motion and stated the polarity of singletons. A rigorous proof of Theorem 19.6 was provided by Kakutani (1944a/b). Kunita & Watanabe (1967) gave the first modern proof of Lévy's characterization, based on Itô's formula for exponential martingales.

[32]Kiyosi Itô (1915–2008), Japanese mathematician, the creator of stochastic calculus, SDEs, multiple WI-integrals, and excursion theory. He is also notable for reworking and making rigorous Lévy's theory of processes with independent increments.

The corresponding time-change reduction of continuous local martingales to a Brownian motion was discovered by DOEBLIN (1940b), and used by him to solve KOLMOGOROV's fundamental problem of constructing diffusion processes with given drift and diffusion rates[33]. However, his revolutionary paper was deposited in a sealed envelope to the Academy of Sciences in Paris, and was not known to the world until 60 years later. In the meantime, the result was rediscovered independently by DAMBIS (1965) and DUBINS & SCHWARZ (1965). A multi-variate version of Doeblin's theorem was noted by KNIGHT (1971), and a simplified proof was given by MEYER (1971). A systematic study of isotropic martingales was initiated by GETOOR & SHARPE (1972).

The skew-product representation in Corollary 19.7 is due to GALMARINO (1963). The integral representation in Theorem 19.11 is essentially due to ITÔ (1951c), who noted its connection with multiple stochastic integrals and chaos expansions. A one-dimensional version of Theorem 19.13 appears in DOOB (1953).

The change of measure was first considered by CAMERON & MARTIN (1944), which is the source of Theorem 19.23, and in WALD's (1946/47) work on sequential analysis, containing the identity of Lemma 19.25 in a version for random walks. The Cameron–Martin theorem was gradually extended to more general settings by many authors, including MARUYAMA (1954/55), GIRSANOV (1960), and VAN SCHUPPEN & WONG (1974). The martingale criterion of Theorem 19.24 was obtained by NOVIKOV (1972).

The material in this chapter is covered by many texts, including the monographs by KARATZAS & SHREVE (1998) and REVUZ & YOR (1999). A more advanced and amazingly informative text is JACOD (1979).

20. Semi-martingales and stochastic integration

DOOB (1953) conceived the idea of a stochastic integration theory for general L^2-martingales, based on a suitable decomposition of continuous-time sub-martingales. MEYER's (1962) proof of such a result opened the door to the L^2-theory, which was then developed by COURRÈGE (1962/63) and KUNITA & WATANABE (1967). The latter paper contains in particular a version of the general substitution rule. The integration theory was later extended in a series of papers by MEYER (1967) and DOLÉANS-DADE & MEYER (1970), and reached maturity with the notes of MEYER (1976) and the books by JACOD (1979), MÉTIVIER & PELLAUMAIL (1980), MÉTIVIER (1982), and DELLACHERIE & MEYER (1975/80).

The basic role of predictable processes as integrands was recognized by MEYER (1967). By contrast, semi-martingales were originally introduced in an ad hoc manner by DOLÉANS-DADE & MEYER (1970), and their basic preservation laws were only gradually recognized. In particular, JACOD (1975)

[33] A totally different construction was given independently by ITÔ (1942a), through his invention of stochastic calculus and the theory of SDEs. Itô's approach was also developed independently by GIHMAN (1947/50/51).

used the general Girsanov theorem of van Schuppen & Wong (1974) to show that the semi-martingale property is preserved under absolutely continuous changes of the probability measure. The characterization of general stochastic integrators as semi-martingales was obtained independently by Bichteler (1979) and Dellacherie (1980), in both cases with support from analysts.

Quasi-martingales were originally introduced by Fisk (1965) and Orey (1966). The decomposition of Rao (1969b) extends a result by Krickeberg (1956) for L^1-bounded martingales. Yoeurp (1976) combined a notion of *stable subspaces*, due to Kunita & Watanabe (1967), with the Hilbert space structure of $\mathcal{M}^2$ to obtain an orthogonal decomposition of L^2-martingales, equivalent to the decompositions in Theorem 20.14 and Proposition 20.16. Elaborating on those ideas, Meyer (1976) showed that the purely discontinuous component admits a representation as a sum of compensated jumps.

SDEs driven by general Lévy processes were considered already by Itô (1951b). The study of SDEs driven by general semi-martingales was initiated by Doléans-Dade (1970), who obtained her exponential process as a solution to the equation in Theorem 20.8. The scope of the theory was later expanded by many authors, and a comprehensive account is given by Protter (1990).

The martingale inequalities in Theorems 20.12 and 20.17 have ancient origins. Thus, a version of the latter result for independent random variables was proved by Kolmogorov (1929) and, in a sharper form, by Prohorov (1959). Their result was extended to discrete-time martingales by Johnson et al. (1985) and Hitczenko (1990). The present statements appeared in K & Sztencel (1991).

Early versions of the inequalities in Theorem 20.12 were proved by Khinchin (1923/24) for symmetric random walks and by Paley (1932) for Walsh series. A version for independent random variables was obtained by Marcinkiewicz & Zygmund (1937/38). The extension to discrete-time martingales is due to Burkholder (1966) for $p > 1$ and to Davis (1970) for $p = 1$. The result was extended to continuous time by Burkholder et al. (1972), who also noted how the general result can be deduced from the statement for $p = 1$. The present proof is a continuous-time version of Davis' original argument.

Introductions to general semi-martingales and stochastic integration are given by Dellacherie & Meyer (1975/80), Jacod & Shiryaev (1987), and Rogers & Williams (2000). Protter (1990) offers an alternative approach, originally suggested by Dellacherie (1980) and Meyer. The book by Jacod (1979) remains a rich source of further information on the subject.

21. Malliavin calculus

Malliavin[34] (1978) sensationally gave a probabilistic proof of Hörmander's (1967) celebrated regularity theorem for solutions to certain elliptic

[34] Paul Malliavin (1925–2010), French mathematician, already famous for contributions to harmonic analysis and other areas, before his revolutionary work in probability.

partial differential equations. The underlying ideas[35], forming the core of a stochastic calculus of variations known as Malliavin calculus, were further developed by many authors, including Stroock (1981), Bismut (1981), and Watanabe (1987), into a powerful tool with many applications.

Ramer (1974) and Skorohod (1975) independently extended Itô's stochastic integral to the case of anticipating integrands. Their integral was recognized by Gaveau & Trauber (1982) as the adjoint of the Malliavin derivative D. The expression for the integrand in Itô's representation of Brownian functionals was obtained by Clark (1970/71), under some regularity conditions that were later removed by Ocone (1984). The differentiability theorem for Itô integrals was proved by Pardoux & Peng (1990).

Many introductions to Malliavin calculus exist, including an early account by Bell (1987). A standard reference is the book by Nualart (1995/2006), which also contains a wealth of applications to anticipating stochastic calculus, stochastic PDEs, mathematical finance, and other areas. I also found the notes of Kunze (2013) especially helpful.

22. Skorohod embedding and functional convergence

The first functional limit theorems were obtained by Kolmogorov (1931b/33a), who considered special functionals of a random walk. Erdös[36] & Kac (1946/47) conceived the idea of an 'invariance principle' that would allow the extension of functional limit theorems from special cases to a general setting. They also treated some interesting functionals of a random walk. The first functional limit theorems were obtained by Donsker[37] (1951/52) for random walks and empirical distribution functions, following an idea of Doob (1949). A general theory based on sophisticated compactness arguments was later developed by Prohorov (1956) and others.

Skorohod's[38] (1965) embedding theorem provided a new, probabilistic approach to Donsker's theorem. Extensions to the martingale context were obtained by many authors, beginning with Dubins (1968). Lemma 22.19 appears in Dvoretzky (1972). Donsker's weak invariance principle was supplemented by a strong version due to Strassen (1964), allowing extensions of many a.s. limit theorems for Brownian motion to suitable random walks. In particular, his result yields a simple proof of the Hartman & Wintner (1941) law of the iterated logarithm, originally deduced from some deep results of Kolmogorov (1929).

[35]Denis Bell tells me that much of this material was known before the work of Malliavin, including the notion of Malliavin derivative.

[36]Paul Erdös (1913–96), Hungarian, the most prolific and eccentric mathematician of the 20th century. Constantly on travel and staying with friends, his only possession was a small suitcase where he kept his laundry. His motto was 'a new roof—a new proof'. He also offered monetary awards, successively upgraded, for solutions to open problems.

[37]Monroe Donsker (1924–91), American mathematician, who proved the first functional limit theorem, and along with Varadhan developed the theory of large deviations.

[38]Anatoliy Skorohod (1930–2011), Ukrainian mathematician, making profound contributions to many areas of stochastic processes.

Komlós et al. (1975/76) showed how the approximation rate in the Skorohod embedding can be improved by a more delicate 'strong approximation'. For an exposition of their work and its numerous applications, see Csörgö & Révész (1981).

Many texts on different levels contain introductions to weak convergence and functional limit theorems. A standard reference remains Billingsley (1968), which contains a careful treatment of the general theory and its numerous applications. Books focusing on applications in special areas include Pollard (1984) for empirical distributions and Whitt (2002) for queuing theory. Versions of the Skorohod embedding, even for martingales, appear in Hall & Heyde (1980) and Durrett (2019).

23. Convergence in distribution

After Donsker (1951/52) had proved his functional limit theorems for random walks and empirical distribution functions, a general theory of weak convergence in function spaces was developed by the Russian school, in seminal papers by Prohorov (1956), Skorohod (1956/57), and Kolmogorov (1956). Thus, Prohorov[39] (1956) proved his fundamental compactness Theorem 23.2, in a setting for separable, complete metric spaces. The abstract theory was later extended in various directions by Le Cam (1957), Varadarajan (1958), and Dudley (1966/67).

Skorohod (1956) considered four different topologies on $D_{[0,1]}$, of which the J_1-topology considered here is the most important for applications. The theory was later extended to $D_{\mathbb{R}_+}$ by Stone (1963) and Lindvall (1973). Moment conditions for tightness were developed by Chentsov (1956) and Billingsley (1968), before the powerful criterion of Aldous (1978) became available. Kurtz (1975) and Mitoma (1983) noted that tightness in $D_{\mathbb{R}_+,S}$ can often be verified by conditions on the one-dimensional projections, as in Theorem 23.23.

Convergence criteria for random measures were established by Prohorov (1961) for compact spaces, by von Waldenfels (1968) for locally compact spaces, and by Harris (1971) for separable, complete metric spaces. In the latter setting, convergence and tightness criteria for point processes, first announced in Debes et al. (1970), were proved in Matthes et al. (1974). The relationship between weak and vague convergence in distribution was clarified, for locally compact spaces, in K(75/76). Strong convergence of point processes was studied extensively by Matthes et al. (1974), where it forms a foundation of their theory of infinitely divisible point processes.

One-dimensional convergence criteria, noted for locally compact spaces in K(75/83) and K(96b), were extended to Polish spaces by Peterson (2001). The power of the Cox transform in this context was explored by Grandell (1976). The one-dimensional criteria provide a link to random closed sets

[39] Yuri Prohorov (1929–2913), Russian probabilist, founder of the modern theory of weak convergence.

with the associated FELL (1962) topology. Random sets had been studied extensively by many authors, including CHOQUET (1953/54), KENDALL (1974), and MATHERON (1975), when associated convergence criteria were identified in NORBERG[40] (1984).

Detailed accounts of the theory and applications of weak convergence appear in many textbooks and monographs, including PARTHASARATHY (1967), BILLINGSLEY (1968), POLLARD (1984), ETHIER & KURTZ (1986), JACOD & SHIRYAEV (1987), and WHITT (2002). The convergence theory for random measures is discussed in MATTHES et al. (1978), DALEY & VERE-JONES (2008), and K(17b). An introduction to random sets appears in SCHNEIDER & WEIL (2008), and more comprehensive treatments are given by MATHERON (1975) and MOLCHANOV (2005).

24. Large deviations

The theory of large deviations originated with some refinements of the central limit theorem noted by many authors, beginning with KHINCHIN (1929). Here the object of study is the ratio of tail probabilities $r_n(x) = P\{\zeta_n > x\} / P\{\zeta > x\}$, where ζ is $N(0,1)$ and $\zeta_n = n^{-1/2} \sum_{k \le n} \xi_k$ for some i.i.d. random variables ξ_k with mean 0 and variance 1, so that $r_n(x) \to 1$ for each x. A precise asymptotic expansion was obtained by CRAMÉR[41] (1938), when x varies with n at a rate $x = o(n^{1/2})$; see PETROV (1995) for details.

In the same paper, CRAMÉR (1938) obtained the first true large deviation result, in the form of our Theorem 24.3, though under some technical assumptions that were later removed by CHERNOFF (1952) and BAHADUR (1971). VARADHAN (1966) extended the result to higher dimensions and rephrased it as a general large deviation principle. At about the same time, SCHILDER (1966) proved his large deviation result for Brownian motion, using the present change-of-measure approach. Similar methods were used by FREIDLIN & WENTZELL (1970/98) to study random perturbations of dynamical systems.

This was long after SANOV (1957) had obtained his large deviation result for empirical distributions of i.i.d. random variables. The relative entropy $H(\nu|\mu)$ appearing in the limit had already been introduced in statistics by KULLBACK & LEIBLER (1951). Its crucial link to the Legendre–Fenchel transform Λ^*, long anticipated by physicists, was formalized by DONSKER & VARADHAN (1975/83). The latter authors also developed some profound and far-reaching extensions of Sanov's theorem, in a series of revolutionary papers[42]. ELLIS

---

[40]After this work, based on an unpublished paper of mine, NORBERG went on to write some deep papers on random capacities, inspired by unpublished notes of W. VERVAAT.

[41]HARALD CRAMÉR (1893–1985), Swedish mathematician and a founder of risk, large deviation, and extreme-value theories. He also wrote the first rigorous treatise on statistical theory. At the end of a distinguished academic career, he became chancellor of the entire Swedish university system.

[42]This work earned Varadhan an Abel prize in 2007, long after the death of his main collaborator. Together with Stroock, Varadhan also made profound contributions to the theory of multi-variate diffusions.

(1985) gives a detailed exposition of those results, along with a discussion of their physical significance.

Some of the underlying principles and techniques were developed at a later stage. Thus, an abstract version of the projective limit approach was introduced by Dawson & Gärtner (1987). Bryc (1990) supplemented Varadhan's (1966) functional version of the LDP with a converse proposition. Similarly, Ioffe (1991) appended a powerful inverse to the classical 'contraction principle'. Finally, Pukhalsky (1991) established the equivalence, under suitable regularity conditions, of the exponential tightness and the goodness of the rate function.

Strassen (1964) established his formidable law of the iterated logarithm by direct estimates. Freedman (1971b) gives a detailed account of the original approach. Varadhan (1984) recognized the result as a corollary to Schilder's theorem, and a complete proof along the suggested lines, different from ours, appears in Deuschel & Stroock (1989).

Gentle introductions to large deviation theory and its applications are given by Varadhan (1984) and Dembo & Zeitouni (1998). The more demanding text of Deuschel & Stroock (1989) provides much additional insight for the persistent reader.

25. Stationary processes and ergodic theorems

The history of ergodic theory dates back to the work of Boltzmann[43] (1887) in statistical mechanics. His 'ergodic hypothesis'—the conjectural equality of time and ensemble averages—was long accepted as an heuristic principle. In probabilistic terms it amounts to the convergence $t^{-1}\int_0^t f(X_s)\,ds \to Ef(X_0)$, where X_t represents the state of the system—typically the configuration of all molecules in a gas—at time t, and the expected value is taken with respect to an invariant probability measure on a compact sub-manifold of the state space.

The ergodic hypothesis was sensationally proved as a mathematical theorem, first in an L^2-version by von Neumann (1932), after Koopman (1931) had noted the connection between measure-preserving transformations and unitary operators on a Hilbert space, and then in the pointwise form of Birkhoff[44] (1932). The proof of the latter version was simplified in stages: first by Yosida & Kakutani (1939), who showed that the statement is an easy consequence of the maximal ergodic lemma, and then by Garsia (1965), who gave a short proof of the latter result. Khinchin (1933a/34) pioneered a translation of the basic ergodic theory into the probabilistic setting of stationary sequences and processes.

The first multi-variate ergodic theorem was obtained by Wiener (1939), who proved his result in the special case of averages over concentric balls. More general versions were established by many authors, including Day (1942) and

[43] Ludwig Boltzmann (1844–1906), Austrian physicist, along with Maxwell and Gibbs a founder of statistical mechanics.

[44] George David Birkhoff (1884–1944), prominent American mathematician.

Pitt (1942). The classical methods were pushed to the limit in a notable paper by Tempel'man (1972). The first ergodic theorem for non-commuting transformations was obtained by Zygmund (1951), where the underlying maximum inequalities go back to Hardy & Littlewood (1930). Sucheston (1983) noted that the statement follows easily from Maker's (1940) lemma.

The ergodic theorem for random matrices was proved by Furstenberg & Kesten (1960), long before the sub-additive ergodic theorem became available. The latter was originally proved by Kingman (1968), under the stronger hypothesis of joint stationarity of the array $(X_{m,n})$. The present extension and shorter proof are due to Liggett (1985).

The ergodic decomposition of invariant measures dates back to Krylov & Bogolioubov (1937), though the basic role of the invariant σ-field was not recognized until the work of Farrell (1962) and Varadarajan (1963). The connection between ergodic decompositions and sufficient statistics is explored in an elegant paper of Dynkin (1978). The traditional approach to the subject is via Choquet theory, as surveyed by Dellacherie & Meyer (1983).

The coupling equivalences in Theorem 25.25 (i) were proved by Goldstein (1979), after Griffeath (1975) had obtained a related result for Markov chains. The shift coupling part of the same theorem was established by Berbee (1979) and Aldous & Thorisson (1993), and the version for abstract groups was then obtained by Thorisson (1996).

The original *ballot theorem*, due to Whitworth (1878) and later rediscovered by Bertrand (1887), is one of the earliest rigorous results of probability theory. It states that if two candidates A and B in an election get the proportions p and $1 - p$ of the votes, then A will lead throughout the ballot count with probability $(2p-1)_+$. Extensions and new approaches have been noted by many authors, beginning with André (1887) and Barbier (1887). A modern combinatorial proof is given by Feller (1968), and a simple martingale argument appears in Chow & Teicher (1997). A version for cyclically stationary sequences and processes was obtained by Takács (1967). All earlier versions are subsumed by the present statement from K(99a), whose proof relies heavily on Takács' ideas.

The first version of Theorem 25.29 was obtained by Shannon (1948), who proved convergence in probability for stationary and ergodic Markov chains in a finite state space. The Markovian restriction was lifted by McMillan (1953), who also strengthened the result to convergence in L^1. Carleson (1958) extended McMillan's result to a countable state space. The a.s. convergence is due to Breiman (1957/60) and Ionescu Tulcea (1960) for finite state spaces and to Chung (1961) in the countable case.

Elementary introductions to stationary processes have been given by many authors, beginning with Doob (1953) and Cramér & Leadbetter (1967). A modern and comprehensive survey of general ergodic theorems is given by Krengel (1985). A nice introduction to information theory is given by Billingsley (1965). Thorisson (2000) gives a detailed account of modern coupling principles.

26. Ergodic properties of Markov processes

The first ratio ergodic theorems were obtained by Doeblin (1938b), Doob (1938/48a), Kakutani (1940), and Hurewicz (1944). Hopf (1954) and Dunford & Schwartz (1956) extended the pointwise ergodic theorem to general $L^1 - L^\infty$-contractions, and the ratio ergodic theorem was extended to positive L^1-contractions by Chacon & Ornstein (1960). The present approach to their result in due to Akcoglu & Chacon (1970).

The notion of Harris recurrence goes back to Doeblin (1940a) and Harris[45] (1956). The latter author used the condition to ensure the existence, in discrete time, of a σ-finite invariant measure. A corresponding continuous-time result was obtained by H. Watanabe (1964). The total variation convergence of Markov transition probabilities was obtained for a countable state space by Orey (1959/62), and in general by Jamison & Orey (1967). Blackwell & Freedman (1964) noted the equivalence of mixing and tail triviality. The present coupling approach goes back to Griffeath (1975) and Goldstein (1979) for the case of strong ergodicity, and to Berbee (1979) and Aldous & Thorisson (1993) for the corresponding weak result.

There is an extensive literature on ergodic theorems for Markov processes, mostly dealing with the discrete-time case. General expositions have been given by many authors, beginning with Neveu (1971) and Orey (1971). Our treatment of Harris recurrent Feller processes is adapted from Kunita (1990), who in turn follows the discrete-time approach of Revuz (1984). Krengel (1985) gives a comprehensive survey of abstract ergodic theorems.

27. Symmetric distributions and predictable maps

de Finetti[46] (1930) proved that an infinite sequence of random events is exchangeable iff the joint distribution is mixed i.i.d. He later (1937) claimed the corresponding result for general random variables, and a careful proof for even more general state spaces was given by Hewitt & Savage[47] (1955). Ryll-Nardzewski[48] (1957) noted that the statement remains true under the weaker hypothesis of contractability, and Bühlmann (1960) extended the result to processes on R_+. The sampling approach to de Finetti's theorem was noted in K(99b).

Finite exchangeable sequences arise in the context of sampling from a finite population, in which context a variety of limit theorems were obtained by Chernoff & Teicher (1958), Hájek (1960), Rosén (1964), Billingsley (1968), and Hagberg (1973). The representation of exchangeable processes on $[0,1]$ was discovered independently in K(73b), along with the associated

[45] Ted Harris (1919–2005), American mathematician, making profound contributions to branching processes, Markov ergodic theory, and convergence of random measures.

[46] Bruno de Finetti (1906–85), Italian mathematician and philosopher, whose famous theorem became a cornerstone in his theory of subjective probability and Bayesian statistics.

[47] This paper is notable for containing the authors' famous $0-1$ law.

[48] Czesław Ryll-Nardzewski (1926–2015), Polish mathematician prominent in analysis and probability, making profound contributions to both exchangeability and Palm measures.

functional limit theorems. The sum of compensated jumps was shown in K(74) to converge uniformly a.s.

The optional skipping theorem—the special case of Theorem 27.7 for i.i.d. sequences and increasing sequences of predictable times—was first noted by Doob[49] (1936/53). It formalizes the intuitive idea that a gambler can't 'beat the odds' by using a clever gambling strategy, as explained in Billingsley (1986). The general predictable sampling theorem, proved in K(88) along with its continuous-time counterpart, yields a simple proof of Sparre-Andersen's (1953/54) notable identity[50] for random walks, which in turn leads to a short proof of Lévy's third arcsine law.

Comprehensive surveys of exchangeability theory have been given by Aldous (1985) and in K(05).

28. Multi-variate arrays and symmetries

The notion of *partial exchangeability*—the invariance in distribution of a sequence of random variables under a proper sub-group of permutations—goes back to de Finetti (1938). Separately exchangeable arrays on N^2 were first studied by Dawid (1972). Symmetric, jointly exchangeable arrays on N^2 arise naturally in the context of *U-statistics*, whose theory goes back to Hoeffding (1948). They were later studied by McGinley & Sibson (1975), Silverman (1976), and Eagleson & Weber (1978). More recently, exchangeable arrays have been used extensively to describe random graphs.

General representations of exchangeable arrays were established independently in profound papers by Aldous (1981) and Hoover (1979), using totally different methods. Thus, Aldous uses probabilistic methods to derive the representation of separately exchangeable arrays on N^2, whereas Hoover derives the representation of jointly exchangeable arrays of arbitrary dimension, using ideas from non-standard analysis and mathematical logic, going back to Gaifman (1961) and Krauss (1969). Hoover also identifies pairs of functions f, g that can be used to represent the same array. Probabilistic proofs of Hoover's results were found[51] in K(92), along with the extensions to contractable arrays. Hoover's uniqueness criteria were amended in K(89/92).

The notion of rotatable arrays was suggested by some limit theorems for U-statistics, going back to Hoeffding (1948). Rotatable arrays also arise naturally in quantum mechanics, where the asymptotic eigenvalue distribution of X was studied by Wigner (1955), leading in particular to his celebrated *semicircle law* for the 'Gaussian orthogonal ensemble', where the entries $X_{ij} = X_{ji}$ are i.i.d. $N(0,1)$ for $i < j$. Olson & Uppuluri (1972) noted that, under the stated independence assumption, the Gaussian distribution follows from the

---

[49]See also the recollections of Halmos (1985), pp. 74–76, who was a student of Doob.

[50]Feller (1966/71) writes that Sparre-Andersen's result 'was a sensation greeted with incredulity, and the original proof was of an extraordinary intricacy and complexity'. He goes on to give a simplified proof, which is still quite complicated.

[51]Hoover's unpublished paper requires expert knowledge in mathematical logic. Before developing my own proofs, I had to ask the author if my interpretations were correct.

joint rotatability suggested by physical considerations[52]. Scores of physicists and mathematicians have extended Wigner's results, but under independence assumptions without physical significance.

Separately and jointly rotatable arrays on $\mathbb{N}^2$ were first studied by Dawid (1977/78), who also conjectured the representation in the separate case. The statement was subsequently proved by Aldous (1981), under some simplifying assumptions that were later removed in K(88). The latter paper also gives the representation in the jointly rotatable case. The higher-dimensional versions, derived in K(95), are stated most conveniently in terms of multiple Wiener–Itô integrals on tensor products of Hilbert spaces. Those results lead in turn to related representations of exchangeable and contractable random sheets.

Motivated by problems in population genetics, Kingman (1978) established his 'paint-box' representation of exchangeable partitions, later extended to more general symmetries in K(05). Algebraic and combinatorial aspects of exchangeable partitions have been studied extensively by Pitman (1995/2002) and Gnedin (1997).

A comprehensive account of contractable, exchangeable, and rotatable arrays is given in K(05), where the proofs of some major results take up to a hundred pages. From the huge literature on the eigenvalue distribution of random matrices, I can especially recommend the books by Mehta (1991) for physicists and by Anderson et al. (2010) for mathematicians.

29. Local time, excursions, and additive functionals

Local time of Brownian motion at a fixed point was discovered and explored by Lévy (1939), who devised several explicit constructions, mostly of the type of Proposition 29.12. Much of Lévy's analysis is based on the observation in Corollary 29.3. The elementary Lemma 29.2 is due to Skorohod (1961/62). Tanaka's (1963) formula, first noted for Brownian motion, was taken by Meyer (1976) as a basis for a general semi-martingale approach. The resulting local time process was recognized as an occupation density, independently by Meyer (1976) and Wang (1977). Trotter (1958b) proved that Brownian local time has a jointly continuous version, and the regularization theorem for general continuous semi-martingales was obtained by Yor (1978).

Modern excursion theory originated with the seminal paper of Itô (1972), which was partly inspired by work of Lévy (1939). In particular, Itô proved a version of Theorem 29.11, assuming the existence of local time. Horowitz (1972) independently studied regenerative sets and noted their connection with subordinators, equivalent to the existence of a local time. A systematic theory of regenerative processes was developed by Maisonneuve (1974). The remarkable Theorem 29.17 was discovered independently by Ray (1963) and Knight (1963), and the present proof is essentially due to Walsh (1978). Our construction of the excursion process is close in spirit to Lévy's original ideas and those of Greenwood & Pitman (1980).

[52]An observer can't distinguish between arrays differing by a joint rotation.

Elementary additive functionals of integral type had been discussed extensively in the literature, when DYNKIN proposed a study of the general case. The existence Theorem 29.23 was obtained by VOLKONSKY (1960), and the construction of local time in Theorem 29.24 dates back to BLUMENTHAL & GETOOR (1964). The integral representation of CAFs in Theorem 29.25 was proved independently by VOLKONSKY (1958/60) and MCKEAN & TANAKA (1961). The characterization of an additive functional in terms of a suitable measure on the state space dates back to MEYER (1962), and an explicit representation of the associated measure was found by REVUZ (1970), after special cases had been considered by HUNT (1957/58).

An introduction to Brownian local time appears in KARATZAS & SHREVE (1998). The books by ITÔ & MCKEAN (1965) and REVUZ & YOR (1999) contain an abundance of further information on the subject. The latter text may also serve as an introduction to additive functionals and excursion theory. For more information on the latter topics, the reader may consult BLUMENTHAL & GETOOR (1968), BLUMENTHAL (1992), and DELLACHERIE et al. (1992).

30. Random measures, smoothing and scattering

The Poisson convergence of superpositions of many small, independent point processes has been noted in increasing generality by many authors, including PALM (1943), KHINCHIN (1955), OSOSKOV (1956), GRIGELIONIS (1963), GOLDMAN (1967), and JAGERS (1972). Limit theorems under simultaneous thinning and rescaling of a given point process were obtained by RÉNYI (1956), NAWROTZKI (1962), BELYAEV (1963), and GOLDMAN (1967). The Cox convergence of dissipative transforms, quoted from K(17b), extends the thinning theorem in K(75), which in turn implies MECKE's (1968) characterization of Cox processes. The partly analogous theory of strong convergence, developed by MATTHES et al. (1974/78), forms a basis for their theory of infinitely divisible point processes.

The limit theorem of DOBRUSHIN[53] (1956) and its various extensions played a crucial role in the development of the subject. Related results for evolutions generated by independent random walks go back to MARUYAMA (1955). The Poisson convergence of traffic flows was first noted by BREIMAN (1963). Precise version of those results, including the role of asymptotically invariant measures, were given by STONE (1968). For constant motion, the fact that independence alone implies the Poisson property was noted in K(78c). Wiener's multi-variate ergodic theorem was extended to random measures by NGUYEN & ZESSIN (1979). A general theory of averaging, smoothing, and scattering, with associated criteria for weak or strong Cox convergence, was developed in K(17b).

Line and flat processes have been studied extensively, first by MILES (1969/

[53]ROLAND DOBRUSHIN (1929–95), Russian mathematical physicist, famous for his work in statistical mechanics.

71/74) in the Poisson case, and then more generally by Davidson[54] (1974a/c), Krickeberg (1974a/b), Papangelou (1974a-b/76), and in K(76/78b-c/81). Spatial branching processes were studied extensively by Matthes et al. (1974/78). Here the stability dichotomy with associated truncation criteria go back to Debes et al. (1970/71). The Palm tree criterion for stability[55] was noted in K(77). Further developments in this area appear in Liemant et al. (1988).

A systematic development of point process theory was initiated by the East-German school, as documented by Matthes[56] et al. (1974/78/82). Daley & Vere-Jones (2003/08) give a detailed introduction to point processes and their applications. General random measure theory has been studied extensively since Jagers (1974) and K(75/76/83/86), and a modern, comprehensive account was attempted in K(17b).

31. Palm and Gibbs kernels, local approximation

In his study of intensity fluctuations in telephone traffic, now recognized as the beginning of modern point process theory, Palm[57] (1943) introduced some Palm probabilities associated with a simple, stationary point process on $\mathbb{R}$. An extended and more rigorous account of Palm's ideas was given by Khinchin (1955), in a monograph on queuing theory, and some basic inversion formulas became known as the *Palm–Khinchin equations*.

Kaplan (1955) established the fundamental time-cycle duality as an extension of results for renewal processes by Doob (1948b), after a partial discrete-time version had been noted by Kac (1947). Kaplan's result was rediscovered in the context of Palm distributions, independently by Ryll-Nardzewski (1961) and Slivnyak (1962). In the special case of intervals on the line, Theorem 31.5 (i) was first noted by Korolyuk, as cited by Khinchin (1955), and part (iii) of the same theorem was obtained by Ryll-Nardzewski (1961). The general versions are due to König & Matthes (1963) and Matthes (1963) for $d = 1$, and to Matthes et al. (1978) for $d > 1$.

Averaging properties of Palm and related measures have been explored by many authors, including Slivnyak (1962), Zähle (1980), and Thorisson (1995). Palm measures of stationary point processes on the line have been studied in detail by many authors, including Nieuwenhuis (1994) and Thorisson (2000). The duality between Palm measures and moment densities was noted in K(99c). Local approximations of general point processes and the associated Palm measures were studied in K(09). The iteration approach to

[54]Rollo Davidson (1944–1970), British math prodigy, making profound contributions to stochastic analysis and geometry, before dying in a climbing accident at age 25. A yearly prize was established to his memory.

[55]often referred to as 'Kallenberg's backward method'—there I got what I deserve!

[56]Klaus Matthes (1931–98), a leader of the East-German probability school, who along with countless students and collaborators developed the modern theory of point processes, with applications to queuing, branching processes, and statistical mechanics. Served for many years as head of the Weierstrass Institute of the Academy of Sciences in DDR.

[57]Conny Palm (1907–51), Swedish engineer and applied mathematician, and the founder of modern point process theory. Also known for designing the first Swedish computer.

conditional distributions and Palm measures goes back to K(10/11b).

Exterior conditioning plays a basic role in statistical mechanics, where the theory of Gibbs measures, inspired by the pioneering work of GIBBS[58] (1902), has been pursued by scores of physicists and mathematicians. In particular, Theorem 31.15 (i) is a version of the celebrated 'DLR equation', noted by DOBRUSHIN (1970) and LANFORD & RUELLE (1969). Connections with modern point process theory were recognized and explored by NGUYEN & ZESSIN (1979), MATTHES et al. (1979), and RAUCHENSCHWANDTNER (1980). Motivated by problems in stochastic geometry, PAPANGELOU (1974b) introduced the kernel named after him, under some simplifying assumptions that were later removed in K(78a). An elegant projection approach to Papangelou kernels was noted by VAN DER HOEVEN (1983). The general theory of Palm and Gibbs kernels was developed in K(83/86).

As with many subjects of the previous chapter, here too the systematic development was initiated by MATTHES et al. (1974/78/82). A comprehensive account of general Palm and Gibbs kernels is given in K(17b), which also contains a detailed exposition of the stationary case. Applications to queuing theory and related areas are discussed by many authors, including FRANKEN et al. (1981) and BACCELLI & BRÉMAUD (2000).

32. Stochastic equations and martingale problems

Long before the existence of any general theory for SDEs, LANGEVIN (1908) proposed his equation to model the *velocity* of a Brownian particle. The solution process was later studied by ORNSTEIN & UHLENBECK (1930), and was thus named after them. A more rigorous discussion appears in DOOB (1942a).

The general idea of a stochastic differential equation goes back to BERNSTEIN (1934/38), who proposed a pathwise construction of diffusion processes by a discrete approximation, leading in the limit to a formal differential equation driven by a Brownian motion. KOLMOGOROV posed the problem of constructing a diffusion process with given drift and diffusion rates[59], which motivated ITÔ (1942a/51b) to develop his theory of SDEs, including a precise definition of the stochastic integral, conditions for existence and uniqueness of solutions, the Markov property of the solution process, and the continuous dependence on initial state. Similar results were later obtained independently by GIHMAN (1947/50/51). A general theory of stochastic flows was developed by KUNITA (1990) and others.

The notion of a weak solution was introduced by GIRSANOV (1960), and criteria for weak existence appear in SKOROHOD (1965). The ideas behind the transformations in Propositions 32.12 and 32.13 date back to GIRSANOV (1960) and VOLKONSKY (1958), respectively. The notion of a martingale problem goes back to LÉVY's martingale characterization of Brownian motion and

[58]JOSIAH WILLARD GIBBS (1839–1903), American physicist, along with Maxwell and Boltzmann a founder of statistical mechanics.

[59]The problem was also solved independently by DOEBLIN (1940b), who used a totally different method based on a random time change; see the notes to Chapter 19.

Dynkin's theory of the characteristic operator. A comprehensive theory was developed by Stroock & Varadhan (1969), who established the equivalence with weak solutions to the associated SDEs, gave general criteria for uniqueness in law, and deduced conditions for the strong Markov and Feller properties. The measurability part of Theorem 32.10 extends an exercise in Stroock & Varadhan (1979).

Yamada & Watanabe (1971) proved that weak existence and pathwise uniqueness imply strong existence and uniqueness in law. Under the same conditions, they also established the existence of a functional solution, possibly depending on the initial distribution. The latter dependence was later removed in K(96a). Ikeda & Watanabe (1989) noted how the notions of pathwise uniqueness and uniqueness in law extend by conditioning from degenerate to arbitrary initial distributions.

The theory of SDEs is covered by countless textbooks on different levels, including Ikeda & Watanabe (2014), Rogers & Williams (2000), and Karatzas & Shreve (1998). More information on the martingale problem appears in Jacod (1979), Stroock & Varadhan (1979), and Ethier & Kurtz (1986). There is also an extensive literature on SPDEs—stochastic partial differential equations—where a standard reference is Walsh (1984).

33. One-dimensional SDEs and diffusions

The study of continuous Markov processes with associated parabolic differential equations, initiated by Kolmogorov (1931a) and Feller (1936), took a new turn with the seminal papers of Feller (1952/54), who studied the generators of one-dimensional diffusions, within the framework of the newly developed semi-group theory. In particular, Feller gave a complete description in terms of scale function and speed measure, classified the boundary behavior, and showed how the latter is determined by the domain of the generator. Finally, he identified the cases when explosion occurs, corresponding to the absorption cases in Theorem 33.15.

A more probabilistic approach to those results was developed by Dynkin (1955b/59), who along with Ray (1956) continued Feller's study of the relationship between analytic properties of the generator and sample path properties of the process. The idea of constructing diffusions on a natural scale through a time change of Brownian motion is due to Hunt (1958) and Volkonsky (1958), and the full description in Theorem 33.9 was completed by Volkonsky (1960) and Itô & McKean (1965). The present stochastic calculus approach is based on ideas in Méléard (1986).

The ratio ergodic Theorem 33.14 was first obtained for Brownian motion by Derman (1954), by a method originally devised for discrete-time chains by Doeblin (1938a). It was later extended to more general diffusions by Motoo & Watanabe (1958). The ergodic behavior of recurrent, one-dimensional diffusions was analyzed by Maruyama & Tanaka (1957).

For one-dimensional SDEs, Skorohod (1965) noticed that Itô's original

Lipschitz condition for pathwise uniqueness can be replaced by a weaker Hölder condition. He also obtained a corresponding comparison theorem. The improved conditions in Theorems 33.3 and 33.5 are due to Yamada & Watanabe (1971) and Yamada (1973), respectively. Perkins (1982) and Le Gall (1983) noted how the semi-martingale local time can be used to simplify and unify the proofs of those and related results. The fundamental weak existence and uniqueness criteria in Theorem 33.1 were discovered by Engelbert & Schmidt (1984/85), whose (1981) $0-1$ law is implicit in Lemma 33.2.

Elementary introductions to one-dimensional diffusions appear in Breiman (1968), Freedman (1971b), and Rogers & Williams (2000). More detailed and advanced accounts are given by Dynkin (1965) and Itô & McKean (1965). Further information on one-dimensional SDEs appears in Karatzas & Shreve (1998) and Revuz & Yor (1999).

34. PDE connections and potential theory

The fundamental solution to the heat equation in terms of the Gaussian kernel was obtained by Laplace (1809). A century later Bachelier (1900) noted the relationship between Brownian motion and the heat equation. The PDE connections were further explored by many authors, including Kolmogorov (1931a), Feller (1936), Kac (1951), and Doob (1955). A first version of Theorem 34.1 was obtained by Kac (1949), who was in turn inspired by Feynman's (1948) work on the Schrödinger equation. Theorem 34.2 is due to Stroock & Varadhan (1969).

In a discussion of the Dirichlet problem, Green (1828) introduced the functions named after him. The Dirichlet, sweeping, and equilibrium problems were all studied by Gauss (1840), in a pioneering paper on electrostatics. The rigorous developments in potential theory began with Poincaré (1890/99), who solved the Dirichlet problem for domains with a smooth boundary. The equilibrium measure was characterized by Gauss as the unique measure minimizing a certain energy functional, but the existence of the minimum was not rigorously established until Frostman (1935). The cone condition for regularity is due to Zaremba (1909).

The first probabilistic connections were made by Phillips & Wiener (1923) and Courant et al. (1928), who solved the Dirichlet problem in the plane by a method of discrete approximation, involving a version of Theorem 34.5 for a simple symmetric random walk. Kolmogorov & Leontovich (1933) evaluated a special hitting distribution for two-dimensional Brownian motion and noted that it satisfies the heat equation. Kakutani (1944b/45) showed how the harmonic measure and sweeping kernel can be expressed in terms of a Brownian motion. The probabilistic methods were extended and perfected by Doob (1954/55), who noted the profound connections with martingale theory. A general potential theory was later developed by Hunt (1957/58) for broad classes of Markov processes.

The interpretation of Green functions as occupation densities was known to Kac (1951), and a probabilistic approach to Green functions was developed by Hunt[60] (1956). The connection between equilibrium measures and quitting times, implicit already in Spitzer (1964) and Itô & McKean (1965), was exploited by Chung[61] (1973) to yield the representation in Theorem 34.14.

Time reversal of diffusion processes was first considered by Schrödinger (1931). Kolmogorov (1936b/37) computed the transition kernels of the reversed process and gave necessary and sufficient conditions for symmetry. The basic role of time reversal and duality in potential theory was recognized by Doob (1954) and Hunt (1958). Proposition 34.15 and the related construction in Theorem 34.21 go back to Hunt, but Theorem 34.19 may be new. The measure ν in Theorem 34.21 is related to the *Kuznetsov measures*, discussed extensively in Getoor (1990). The connection between random sets and alternating capacities was established by Choquet (1953/54), and a corresponding representation of infinitely divisible random sets was obtained by Matheron (1975).

The classical decomposition of subharmonic functions goes back to Riesz (1926/30). The basic connection between super-harmonic functions and super-martingales was established by Doob (1954), leading to a probabilistic counterpart of Riesz' theorem. This is where he recognized the need for a general decomposition theorem for super-martingales, generalizing the elementary Lemma 9.10. Doob also proved the continuity of the composition of an excessive function with Brownian motion.

Introductions to probabilistic potential theory and other PDE connections appear in Bass (1995/98), Chung (1995), and Karatzas & Shreve (1998). A detailed exposition of classical probabilistic potential theory is given by Port & Stone (1978). Doob (1984) provides a wealth of further information on both the analytic and the probabilistic aspects. An introduction to Hunt's work and subsequent developments is given by Chung (1982). More advanced and comprehensive treatments appear in Blumenthal & Getoor (1968), Dellacherie & Meyer (1987/92), and Sharpe (1988), which also cover the theory of additive functionals and their potentials. The connection between random sets and alternating capacities is discussed in Molchanov (2005).

35. Stochastic differential geometry

Brownian motion and stochastic calculus on Riemannian manifolds and Lie groups were considered already by Yosida (1949), Itô (1950), Dynkin (1961), and others, involving the sophisticated machinery of Levy-Civita connections and constructions by 'rolling without slipping'. Martingales in Riemannian

[60] Gilbert Hunt (1916–2008), American mathematician and a prominent tennis player, whose deep and revolutionary papers established a close connection between Markov processes and general potential theory.

[61] Kai Lai Chung (1917–2009), Chinese-American mathematician, making significant contributions to many areas of modern probability. He became a friend of mine and even spent a week as a guest in our house.

manifolds have been studied by DUNCAN (1983) and KENDALL (1988). A Lévy–Khinchin type representation of infinitely divisible distributions on Lie groups was obtained by HUNT (1956b), and a related theory of Lévy processes in Lie groups and homogeneous spaces has been developed by LIAO (2018).

SCHWARTZ[62] (1980/82) realized that much of the classical theory could be developed for semi-martingales in a differentiable manifold without any Riemannian structure. His ideas[63] were further pursued by MEYER (1981/82), arguably leading to a significant simplification[64] of the whole area.

The need of a *connection* to define martingales in a differential manifold was recognized by Meyer, and the present definition of martingales in connected manifolds goes back to BISMUT (1981). Characterizations of martingales in terms of convex functions were given by DARLING (1982). Our notion of local characteristics of a semi-martingale, believed to be new[65], arose from attempts to make probabilistic sense of the beautiful but somewhat obscure Schwartz–Meyer theory for general semi-martingales. Though the diffusion rates require no additional structure, the drift rate makes sense only in manifolds with a connection[66].

A survey of Brownian motion and SDEs in a Riemannian manifold appears in IKEDA & WATANABE (2014). For beginners we further recommend the 'overture' to the Riemannian theory in ROGERS & WILLIAMS (2000), and the introduction to stochastic differential geometry by KENDALL (1988). Detailed and comprehensive accounts of the Schwartz–Meyer theory are given by ÉMERY (1989/2000).

Appendices

A modern and comprehensive account of general measure theory is given by BOGACHEV (2007). Another useful reference is DUDLEY (1989), who also discusses the axiomatization of mathematics. The projection and section theorems rely on capacity theory, for which we refer to DELLACHERIE (1972) and DELLACHERIE & MEYER (1975/83). Most of the material on topology and functional analysis can be found in standard textbooks.

The J_1-topology was introduced by SKOROHOD (1956), and detailed expositions appear in BILLINGSLEY (1968), ETHIER & KURTZ (1986), and JACOD & SHIRYAEV (1987). Our discussion of topologies on $\mathcal{M}_S$ is adapted from K(17b). The topology on the space of closed sets was introduced in a more general setting by FELL (1962), and a detailed proof of Theorem A6.1, differ-

[62]LAURENT SCHWARTZ (1915–2002), French mathematician, whose creation of distribution theory earned him a Fields medal in 1950. Towards the end of a long and distinguished career, he got interested in stochastic differential geometry.

[63]Michel Emery tells me that much of this material was known before the work of Schwartz, whose aim was to develop a Bourbaki-style approach to stochastic calculus in manifolds.

[64]hence the phrase *sans larmes* (without tears) in the title of Meyer's 1981 paper

[65]A notion of local characteristics also appears in MEYER (1981), though with a totally different meaning.

[66]simply because a martingale is understood to be a semi-martingale without drift

ent from ours, appears in Matheron (1975). A modern and comprehensive account of the relevant set theory is given by Molchanov (2004).

The material on differential geometry is standard, and our account is adapted from Émery (1989), though with a simplified notation.

Bibliography

Here we list only publications that are explicitly mentioned in the text or notes or are directly related to results cited in the book. Knowledgeable readers will notice that many books or papers of historical significance are missing. The items are ordered alphabetically after the main part of the last name, as in de F̲inetti and von N̲eumann.

Adler, R.J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Proceses.* Inst. Math. Statist., Hayward, CA.

Akcoglu, M.A., Chacon, R.V. (1970). Ergodic properties of operators in Lebesgue space. *Adv. Appl. Probab.* **2**, 1–47.

Aldous, D.J. (1978). Stopping times and tightness. *Ann. Probab.* **6**, 335–340.

— (1981). Representations for partially exchangeable arrays of random variables. *J. Multivar. Anal.* **11**, 581–598.

— (1985). Exchangeability and related topics. *Lect. Notes in Math.* **1117**, 1–198. Springer, Berlin.

Aldous, D., Thorisson, H. (1993). Shift-coupling. *Stoch. Proc. Appl.* **44**, 1-14.

Alexandrov, A.D. (1940/43). Additive set-functions in abstract spaces. *Mat. Sb.* **8**, 307–348; **9**, 563–628; **13**, 169–238.

Alexandrov, P. (1916). Sur la puissance des ensembles mesurables B. *C.R. Acad. Sci. Paris* **162**, 323–325.

Anderson, G.W., Guionnet, A., Zeitouni, O. (2010). *Am Introduction to Random Matrices.* Cambridge Univ. Press.

André, D. (1887). Solution directe du problème résolu par M. Bertrand. *C.R. Acad. Sci. Paris* **105**, 436–437.

Athreya, K., McDonald, D., Ney, P. (1978). Coupling and the renewal theorem. *Amer. Math. Monthly* **85**, 809–814.

Athreya, K.B., Ney, P.E. (1972). *Branching Processes.* Springer, Berlin.

Baccelli, F., Brémaud, P. (2000). *Elements of Queueing [sic] Theory*, 2nd ed. Springer, Berlin.

Bachelier, L. (1900). Théorie de la spéculation. *Ann. Sci. École Norm. Sup.* **17**, 21–86.

— (1901). Théorie mathématique du jeu. *Ann. Sci. École Norm. Sup.* **18**, 143–210.

Bahadur, R.R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia.

Barbier, É. (1887). Généralisation du problème résolu par M. J. Bertrand. *C.R. Acad. Sci. Paris* **105**, 407, 440.

Bartlett, M.S. (1949). Some evolutionary stochastic processes. *J. Roy. Statist. Soc. B* **11**, 211–229.

Bass, R.F. (1995). *Probabilistic Techniques in Analysis.* Springer, NY.

— (1998). *Diffusions and Elliptic Operators.* Springer, NY.

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1

Bateman, H. (1910). Note on the probability distribution of α-particles. *Philos. Mag.* **20** (6), 704–707.

Baxter, G. (1961). An analytic approach to finite fluctuation problems in probability. *J. d'Analyse Math.* **9**, 31–70.

Bell, D.R. (1987). *The Malliavin Calculus.* Longman & Wiley.

Belyaev, Y.K. (1963). Limit theorems for dissipative flows. *Th. Probab. Appl.* **8**, 165–173.

Berbee, H.C.P. (1979). *Random Walks with Stationary Increments and Renewal Theory.* Mathematisch Centrum, Amsterdam.

Bernoulli, J. (1713). *Ars Conjectandi.* Thurnisiorum, Basel.

Bernstein, S.N. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.* **97**, 1–59.

— (1934). Principes de la théorie des équations différentielles stochastiques. *Trudy Fiz.-Mat., Steklov Inst., Akad. Nauk.* **5**, 95–124.

— (1937). On some variations of the Chebyshev inequality (in Russian). *Dokl. Acad. Nauk SSSR* **17**, 275–277.

— (1938). Équations différentielles stochastiques. *Act. Sci. Ind.* **738**, 5–31.

Bertoin, J. (1996). *Lévy Processes.* Cambridge Univ. Press.

Bertrand, J. (1887). Solution d'un problème. *C.R. Acad. Sci. Paris* **105**, 369.

Bichteler, K. (1979). Stochastic integrators. *Bull. Amer. Math. Soc.* **1**, 761–765.

Bienaymé, I.J. (1845). De la loi de multiplication et de la durée des familles. *Soc. Philomat. Paris Extraits, Sér.* **5**, 37–39.

— (1853). Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *C.R. Acad. Sci. Paris* **37**, 309–324.

Billingsley, P. (1965). *Ergodic Theory and Information.* Wiley, NY.

— (1968). *Convergence of Probability Measures.* Wiley, NY.

— (1986/95). *Probability and Measure*, 3rd ed. Wiley, NY.

Birkhoff, G.D. (1932). Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA* **17**, 656–660.

Bismut, J.M. (1981). Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's condition. *Z. Wahrsch. verw. Geb.* **56**, 469–505.

Blackwell, D. (1948). A renewal theorem. *Duke Math. J.* **15**, 145–150.

— (1953). Extension of a renewal theorem. *Pacific J. Math.* **3**, 315–320.

Blackwell, D., Freedman, D. (1964). The tail σ-field of a Markov chain and a theorem of Orey. *Ann. Math. Statist.* **35**, 1291–1295.

Blumenthal, R.M. (1957). An extended Markov property. *Trans. Amer. Math. Soc.* **82**, 52–72.

— (1992). *Excursions of Markov Processes.* Birkhäuser, Boston.

Blumenthal, R.M., Getoor, R.K. (1964). Local times for Markov processes. *Z. Wahrsch. verw. Geb.* **3**, 50–74.

— (1968). *Markov Processes and Potential Theory.* Academic Press, NY.

Bochner, S. (1932). *Vorlesungen über Fouriersche Integrale*, Akad. Verlagsges., Leipzig. Repr. Chelsea, NY 1948.

— (1933). Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. *Math. Ann.* **108**, 378–410.

Bogachev, V.I. (2007). *Measure Theory* **1–2**. Springer.

Boltzmann, L. (1887). Über die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *J. Reine Angew. Math.* **100**, 201–212.

Borel, E. (1895). Sur quelques points de la théorie des fonctions. *Ann. Sci. École Norm. Sup.* (3) **12**, 9–55.

— (1898). *Leçons sur la Théorie des Fonctions.* Gauthier-Villars, Paris.

— (1909a). Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo* **27**, 247–271.

— (1909b/24). *Elements of the Theory of Probability*, 3rd ed., Engl. transl. Prentice-Hall 1965.

Bortkiewicz, L.v. (1898). *Das Gesetz der kleinen Zahlen.* Teubner, Leipzig.

Breiman, L. (1957/60). The individual ergodic theorem of infomation theory. *Ann. Math. Statist.* **28**, 809–811; **31**, 809–810.

— (1963). The Poisson tendency in traffic distribution. *Ann. Math. Statist.* **34**, 308–311.

— (1968). *Probability.* Addison-Wesley, Reading, MA. Repr. SIAM, Philadelphia 1992.

Brown, R. (1828). A brief description of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Ann. Phys.* **14**, 294–313.

Brown, T.C. (1978). A martingale approach to the Poisson convergence of simple point processes. *Ann. Probab.* **6**, 615–628.

— (1979). Compensators and Cox convergence. *Math. Proc. Cambridge Phil. Soc.* **90**, 305–319.

Bryc, W. (1990). Large deviations by the asymptotic value method. In *Diffusion Processes and Related Problems in Analysis* (M. Pinsky, ed.), 447–472. Birkhäuser, Basel.

Bühlmann, H. (1960). Austauschbare stochastische Variabeln und ihre Grenzwertsätze. *Univ. Calif. Publ. Statist.* **3**, 1–35.

Buniakowsky, V.Y. (1859). Sur quelques inégalités concernant les intégrales ordinaires et les intégrales aux différences finies. *Mém. de l'Acad. St.-Petersbourg* **1**:9.

Burkholder, D.L. (1966). Martingale transforms. *Ann. Math. Statist.* **37**, 1494–1504.

Burkholder, D.L., Davis, B.J., Gundy, R.F. (1972). Integral inequalities for convex functions of operators on martingales. *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **2**, 223–240.

Burkholder, D.L., Gundy, R.F. (1970). Extrapolation and interpolation of quasi-linear operators on martingales. *Acta Math.* **124**, 249–304.

Cameron, R.H., Martin, W.T. (1944). Transformation of Wiener integrals under translations. *Ann. Math.* **45**, 386–396.

Cantelli, F.P. (1917). Su due applicazione di un teorema di G. Boole alla statistica matematica. *Rend. Accad. Naz. Lincei* **26**, 295–302.

— (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari* **4**, 421–424.

Carathéodory, C. (1918/27). *Vorlesungen über reelle Funktionen*, 2nd ed. Teubner, Leipzig (1st ed. 1918). Repr. Chelsea, NY 1946.

Carleson, L. (1958). Two remarks on the basic theorems of information theory. *Math. Scand.* **6**, 175–180.

Cauchy, A.L. (1821). *Cours d'analyse de l'École Royale Polytechnique*, Paris.

Chacon, R.V., Ornstein, D.S. (1960). A general ergodic theorem. *Illinois J. Math.* **4**, 153–160.

Chapman, S. (1928). On the Brownian displacements and thermal diffusion of grains suspended in a non-uniform fluid. *Proc. Roy. Soc. London (A)* **119**, 34–54.

Chebyshev, P.L. (1867). Des valeurs moyennes. *J. Math. Pures Appl.* **12**, 177–184.

— (1890). Sur deux théorèmes relatifs aux probabilités. *Acta Math.* **14**, 305–315.

Chentsov, N.N. (1956). Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the "heuristic" approach to the Kolmogorov–Smirnov tests. *Th. Probab. Appl.* **1**, 140–144.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23**, 493–507.

Chernoff, H., Teicher, H. (1958). A central limit theorem for sequences of exchangeable random variables. *Ann. Math. Statist.* **29**, 118–130.

Choquet, G. (1953/54). Theory of capacities. *Ann. Inst. Fourier Grenoble* **5**, 131–295.

Chow, Y.S., Teicher, H. (1978/97). *Probability Theory: Independence, Interchangeability, Martingales*, 3nd ed. Springer, NY.

Chung, K.L. (1960). *Markov Chains with Stationary Transition Probabilities.* Springer, Berlin.

— (1961). A note on the ergodic theorem of information theory. *Ann. Math. Statist.* **32**, 612–614.

— (1968/74). *A Course in Probability Theory*, 2nd ed. Academic Press, NY.

— (1973). Probabilistic approach to the equilibrium problem in potential theory. *Ann. Inst. Fourier Grenoble* **23**, 313–322.

— (1982). *Lectures from Markov Processes to Brownian Motion.* Springer, NY.

— (1995). *Green, Brown, and Probability.* World Scientific, Singapore.

Chung, K.L., Doob, J.L. (1965). Fields, optionality and measurability. *Amer. J. Math.* **87**, 397–424.

Chung, K.L., Fuchs, W.H.J. (1951). On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.* **6**.

Chung, K.L., Ornstein, D.S. (1962). On the recurrence of sums of random variables. *Bull. Amer. Math. Soc.* **68**, 30–32.

Chung, K.L., Walsh, J.B. (1974). Meyer's theorem on previsibility. *Z. Wahrsch. verw. Geb.* **29**, 253–256.

Chung, K.L., Williams, R.J. (1983/90). *Introduction to Stochastic Integration*, 2nd ed. Birkhäuser, Boston.

Çinlar, E. (2011). *Probability and Stochastics.* Springer.

Clark, J.M.C. (1970/71). The representation of functionals of Brownian motion by stochastic integrals. *Ann. Math. Statist.* **41**, 1282–95; **42**, 1778.

Copeland, A.H., Regan, F. (1936). A postulational treatment of the Poisson law. *Ann. Math,* **37**, 357–362.

Courant, R., Friedrichs, K., Lewy, H. (1928). Über die partiellen Differentialgleichungen der mathematischen Physik. *Math. Ann.* **100**, 32–74.

Cournot, A.A. (1843). *Exposition de la théorie des chances et des probabilités.* Hauchette, Paris.

Courrège, P. (1962/63). Intégrales stochastiques et martingales de carré intégrable. *Sem. Brelot–Choquet–Deny* **7**. Publ. Inst. H. Poincaré.

Cox, D.R. (1955). Some statistical methods connected with series of events. *J. R. Statist. Soc. Ser. B* **17**, 129–164.

Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actual. Sci. Indust.* **736**, 5–23.

— (1942). On harmonic analysis in certain functional spaces. *Ark. Mat. Astr. Fys.* **28B**:12 (17 pp.).

Cramér, H., Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes.* Wiley, NY.

Cramér, H., Wold, H. (1936). Some theorems on distribution functions. *J. London Math. Soc.* **11**, 290–295.

Csörgö, M., Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, NY.

Daley, D.J., Vere-Jones, D. (2003/08). *An Introduction to the Theory of Point Processes* **1-2**, 2nd ed. Springer, NY.

Dambis, K.E. (1965). On the decomposition of continuous submartingales. *Th. Probab. Appl.* **10**, 401–410.

Daniell, P.J. (1918/19). Integrals in an infinite number of dimensions. *Ann. Math.* (2) **20**, 281–288.

— (1919/20). Functions of limited variation in an infinite number of dimensions. *Ann. Math.* (2) **21**, 30–38.

— (1920). Stieltjes derivatives. *Bull. Amer. Math. Soc.* **26**, 444–448.

Darling, R.W.R. (1982). Martingales in manifolds—definition, examples and behaviour under maps. *Lect. Notes in Math.* **921**, Springer 1982.

Davidson, R. (1974a). Stochastic processes of flats and exchangeability. In: *Stochastic Geometry* (E.F. Harding & D.G. Kendall, eds.), pp. 13–45. Wiley, London.

— (1974b). Exchangeable point-processes. *Ibid.*, 46–51.

— (1974c). Construction of line-processes: second-order properties. *Ibid.*, 55–75.

Davis, B.J. (1970). On the integrability of the martingale square function. *Israel J. Math.* **8**, 187–190.

Dawid, A.P. (1977). Spherical matrix distributions and a multivariate model. *J. Roy. Statist. Soc.* (B) **39**, 254–261.

— (1978). Extendibility of spherical matrix distributions. *J. Multivar. Anal.* **8**, 559–566.

Dawson, D.A. (1977). The critical measure diffusion. *Z. Wahrsch. verw. Geb.* **40**, 125–145.

— (1993). Measure-valued Markov processes. In: *École d'été de probabilités de Saint-Flour XXI–1991. Lect. Notes in Math.* **1541**, 1–126. Springer, Berlin.

Dawson, D.A., Gärtner, J. (1987). Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. *Stochastics* **20**, 247–308.

Dawson, D.A., Perkins, E.A. (1991). Historical processes. *Mem. Amer. Math. Soc.* **93**, #454.

Day, M.M. (1942). Ergodic theorems for Abelian semigroups. *Trans. Amer. Math. Soc.* **51**, 399–412.

Debes, H., Kerstan, J., Liemant, A., Matthes, K. (1970/71). Verallgemeinerung eines Satzes von Dobrushin I, III. *Math. Nachr.* **47**, 183–244; **50**, 99–139.

Dellacherie, C. (1972). *Capacités et Processus Stochastiques.* Springer, Berlin.

— (1980). Un survol de la théorie de l'intégrale stochastique. *Stoch. Proc. Appl.* **10**, 115-144.

Dellacherie, C., Maisonneuve, B., Meyer, P.A. (1992). *Probabilités et Potentiel, V.* Hermann, Paris.

Dellacherie, C., Meyer, P.A. (1975/80/83/87). *Probabilités et Potentiel, I–IV.* Hermann, Paris. Engl. trans., North-Holland.

Dembo, A., Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, NY.

Derman, C. (1954). Ergodic property of the Brownian motion process. *Proc. Natl. Acad. Sci. USA* **40**, 1155–1158.

Deuschel, J.D., Stroock, D.W. (1989). *Large Deviations.* Academic Press, Boston.

Dobrushin, R.L. (1956). On Poisson laws for distributions of particles in space (in Russian). *Ukrain. Mat. Z.* **8**, 127–134.

— (1970). Gibbsian random fields without hard core conditions. *Theor. Math. Phys.* **4**, 705–719.

Doeblin, W. (1938a). Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkan.* **2**, 77–105.

— (1938b). Sur deux problèmes de M. Kolmogoroff concernant les chaînes dénombrables. *Bull. Soc. Math. France* **66**, 210–220.

— (1939a). Sur les sommes d'un grand nombre de variables aléatoires indépendantes. *Bull. Sci. Math.* **63**, 23–64.

— (1939b). Sur certains mouvements aléatoires discontinus. *Skand. Aktuarietidskr.* **22**, 211–222.

— (1940a). Eléments d'une théorie générale des chaînes simples constantes de Markoff. *Ann. Sci. Ecole Norm. Sup. 3*, **57**, 61–111.

— (1940b/2000). Sur l'équation de Kolmogoroff. *C.R. Acad. Sci.* **331**.

— (1947). Sur l'ensemble de puissances d'une loi de probabilité. *Ann. Ec. Norm. Sup.*

Döhler, R. (1980). On the conditional independence of random events. *Th. Probab. Appl.* **25**, 628–634.

Doléans(-Dade), C. (1967a). Processus croissants naturel et processus croissants très bien mesurable. *C.R. Acad. Sci. Paris* **264**, 874–876.

— (1967b). Intégrales stochastiques dépendant d'un paramètre. *Publ. Inst. Stat. Univ. Paris* **16**, 23–34.

— (1970). Quelques applications de la formule de changement de variables pour les semi-martingales. *Z. Wahrsch. verw. Geb.* **16**, 181–194.

Doléans-Dade, C., Meyer, P.A. (1970). Intégrales stochastiques par rapport aux martingales locales. *Lect. Notes in Math.* **124**, 77–107. Springer, Berlin.

Donsker, M.D. (1951/52). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6**.

— (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23**, 277–281.

Donsker, M.D., Varadhan, S.R.S. (1975/83). Asymptotic evaluation of certain Markov process expectations for large time, I–IV. *Comm. Pure Appl. Math.* **28**, 1–47, 279–301; **29**, 389–461; **36**, 183–212.

Doob, J.L. (1936). Note on probability. *Ann. Math.* (2) **37**, 363–367.

— (1937). Stochastic processes depending on a continuous parameter. *Trans. Amer. Math. Soc.* **42**, 107–140.

— (1938). Stochastic processes with an integral-valued parameter. *Trans. Amer. Math. Soc.* **44**, 87–150.

— (1940). Regularity properties of certain families of chance variables. *Trans. Amer. Math. Soc.* **47**, 455–486.

— (1942a). The Brownian movement and stochastic equations. *Ann. Math.* **43**, 351–369.

— (1942b). Topics in the theory of Markoff chains. *Trans. Amer. Math. Soc.* **52**, 37–64.

— (1945). Markoff chains—denumerable case. *Trans. Amer. Math. Soc.* **58**, 455–473.

— (1947). Probability in function space. *Bull. Amer. Math. Soc.* **53**, 15–30.

— (1948a). Asymptotic properties of Markov transition probabilities. *Trans. Amer. Math. Soc.* **63**, 393–421.

— (1948b). Renewal theory from the point of view of the theory of probability. *Trans. Amer. Math. Soc.* **63**, 422–438.

— (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20**, 393–403.

— (1951). Continuous parameter martingales. *Proc. 2nd Berkeley Symp. Math. Statist. Probab.*, 269–277.

— (1953). *Stochastic Processes.* Wiley, NY.

— (1954). Semimartingales and subharmonic functions. *Trans. Amer. Math. Soc.* **77**, 86–121.

— (1955). A probability approach to the heat equation. *Trans. Amer. Math. Soc.* **80**, 216–280.

— (1984). *Classical Potential Theory and its Probabilistic Counterpart.* Springer, NY.

— (1994). *Measure Theory.* Springer, NY.

Dubins, L.E. (1968). On a theorem of Skorohod. *Ann. Math. Statist.* **39**, 2094–2097.

Dubins, L.E., Schwarz, G. (1965). On continuous martingales. *Proc. Natl. Acad. Sci. USA* **53**, 913–916.

Dudley, R.M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10**, 109–126.

— (1967). Measures on non-separable metric spaces. *Illinois J. Math.* **11**, 449–453.

— (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39**, 1563–1572.

— (1989). *Real Analysis and Probability.* Wadsworth, Brooks & Cole, Pacific Grove, CA.

Duncan, T.E. (1983). Some geometric methods for stochastic integration in manifolds. *Geometry and Identification*, 63–72. *Lie groups: Hist., Frontiers and Appl. Ser. B 1*, Math. Sci. Press, Brooklyn 1983.

Dunford, N. (1939). A mean ergodic theorem. *Duke Math. J.* **5**, 635–646.

Dunford, N., Schwartz, J.T. (1956). Convergence almost everywhere of operator averages. *J. Rat. Mech. Anal.* **5**, 129–178.

Durrett, R. (1991/2019). *Probability Theory and Examples*, 2nd ed. Wadsworth, Brooks & Cole, Pacific Grove, CA.

Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **2**, 513–535.

Dynkin, E.B. (1952). Criteria of continuity and lack of discontinuities of the second kind for trajectories of a Markov stochastic process (Russian). *Izv. Akad. Nauk SSSR, Ser. Mat.* **16**, 563–572.

— (1955a). Infinitesimal operators of Markov stochastic processes (Russian). *Dokl. Akad. Nauk SSSR* **105**, 206–209.

— (1955b). Continuous one-dimensional Markov processes (Russian). *Dokl. Akad. Nauk SSSR* **105**, 405–408.

— (1956). Markov processes and semigroups of operators. Infinitesimal operators of Markov processes. *Th. Probab. Appl.* **1**, 25–60.

— (1959). One-dimensional continuous strong Markov processes. *Th. Probab. Appl.* **4**, 3–54.

— (1961a). *Theory of Markov Processes.* Engl. trans., Prentice-Hall and Pergamon Press, Englewood Cliffs, NJ, and Oxford. (Russian orig. 1959.)

— (1961b). Non-negative eigenfunctions of the Laplace–Beltrami operator and Brownian motion in certain symmetric spaces (in Russian). *Dokl. Akad. Nauk. SSSR*, **141**, 288–291.

— (1965). *Markov Processes, Vols. 1–2.* Engl. trans., Springer, Berlin. (Russian orig. 1963.)

— (1978). Sufficient statistics and extreme points. *Ann. Probab.* **6**, 705–730.

— (1991). Branching particle systems and superprocesses. *Ann. Probab.* **19**, 1157–1194.

Dynkin, E.B., Yushkevich, A.A. (1956). Strong Markov processes. *Th. Probab. Appl.* **1**, 134–139.

Eagleson, G.K., Weber, N.C. (1978). Limit theorems for weakly exchangeable arrays. *Math. Proc. Cambridge Phil. Soc.* **84**, 123–130.

Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. Engl. trans. in *Investigations on the Theory of the Brownian Movement.* Repr. Dover, NY 1956.

— (1906). On the theory of Brownian motion. *Ibid.*

Elliott, R.J. (1982). *Stochastic Calculus and Applications.* Springer, NY.

Ellis, R.L. (1844). On a question in the theory of probabilities. *Cambridge Math. J.* **4** (21), 127–133.

Ellis, R.S. (1985). *Entropy, Large Deviations, and Statistical Mechanics.* Springer, NY.

Émery, M. (1989). *Stochastic Calculus in Manifolds.* Springer.

— (2000/13). Martingales continues dans les variétés différentiables. *Lect. Notes in Math.* **1738**, 3–84. Reprinted in *Stochastic Differential Geometry at Saint-Flour*, pp. 263–346. Springer.

Engelbert, H.J., Schmidt, W. (1981). On the behaviour of certain functionals of the Wiener process and applications to stochastic differential equations. *Lect. Notes in Control and Inform. Sci.* **36**, 47–55.

— (1984). On one-dimensional stochastic differential equations with generalized drift. *Lect. Notes in Control and Inform. Sci.* **69**, 143–155. Springer, Berlin.

— (1985). On solutions of stochastic differential equations without drift. *Z. Wahrsch. verw. Geb.* **68**, 287–317.

Erdös, P., Feller, W., Pollard, H. (1949). A theorem on power series. *Bull. Amer. Math. Soc.* **55**, 201–204.

Erdös, P., Kac, M. (1946). On certain limit theorems in the theory of probability. *Bull. Amer. Math. Soc.* **52**, 292–302.

— (1947). On the number of positive sums of independent random variables. *Bull. Amer. Math. Soc.* **53**, 1011–1020.

Erlang, A.K. (1909). The theory of probabilities and telephone conversations. *Nyt. Tidskr. Mat. B* **20**, 33–41.

Etheridge, A.M. (2000). *An Introduction to Superprocesses.* Univ. Lect. Ser. **20**, Amer. Math. Soc., Providence, RI.

Ethier, S.N., Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, NY.

Faber, G. (1910). Über stetige Funktionen, II. *Math. Ann.* **69**, 372–443.

Farrell, R.H. (1962). Representation of invariant measures. *Illinois J. Math.* **6**, 447–467.

Fatou, P. (1906). Séries trigonométriques et séries de Taylor. *Acta Math.* **30**, 335–400.

Fell, J.M.G. (1962). A Hausdorff topology for the closed subsets of a locally compact non-Hausdorff space. *Proc. Amer. Math. Soc.* **13**, 472–476.

Feller, W. (1935/37). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie, I–II. *Math. Z.* **40**, 521–559; **42**, 301–312.

— (1936). Zur Theorie der stochastischen Prozesse (Existenz und Eindeutigkeitssätze). *Math. Ann.* **113**, 113–160.

— (1937). On the Kolmogoroff–P. Lévy formula for infinitely divisible distribution functions. *Proc. Yugoslav Acad. Sci.* **82**, 95–112.

— (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48**, 488–515; **58**, 474.

— (1949). Fluctuation theory of recurrent events. *Trans. Amer. Math. Soc.* **67**, 98–119.

— (1951). Diffusion processes in genetics. *Proc. 2nd Berkeley Symp. Math. Statist. Probab.*, 227–246. Univ. Calif. Press, Berkeley.

— (1952). The parabolic differential equations and the associated semi-groups of transformations. *Ann. Math.* **55**, 468–519.

— (1954). Diffusion processes in one dimension. *Trans. Amer. Math. Soc.* **77**, 1–31.

— (1950/57/68; 1966/71). *An Introduction to Probability Theory and its Applications*, **1** (3rd ed.); **2** (2nd ed.). Wiley, NY (1st eds. 1950, 1966).

Feller, W., Orey, S. (1961). A renewal theorem. *J. Math. Mech.* **10**, 619–624.

Feynman, R.P. (1948). Space-time approach to nonrelativistic quantum mechanics. *Rev. Mod. Phys.* **20**, 367–387.

Finetti, B. de (1929). Sulle funzioni ad incremento aleatorio. *Rend. Acc. Naz. Lincei* **10**, 163–168.

— (1930). Fuzione caratteristica di un fenomeno aleatorio. *Mem. R. Acc. Lincei (6)* **4**, 86–133.

— (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1–68.

— (1938). Sur la condition d'équivalence partielle. *Act. Sci. Ind.* **739**, 5–18.

Fisk, D.L. (1965). Quasimartingales. *Trans. Amer. Math. Soc.* **120**, 369–389.

— (1966). Sample quadratic variation of continuous, second-order martingales. *Z. Wahrsch. verw. Geb.* **6**, 273–278.

Fortet, R. (1943). Les fonctions aléatoires du type de Markoff associées à certaines équations linéaires aux dérivées partielles du type parabolique. *J. Math. Pures Appl.* **22**, 177–243.

Franken, P., König, D., Arndt, U., Schmidt, V. (1981). *Queues and Point Processes.* Akademie-Verlag, Berlin.

Fréchet, M. (1928). *Les Espaces Abstraits.* Gauthier-Villars, Paris.

Freedman, D. (1962/63). Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Statist.* **33**, 916–923; **34**, 1194–1216.

— (1971a). *Markov Chains.* Holden-Day, San Francisco. Repr. Springer, NY 1983.

— (1971b). *Brownian Motion and Diffusion.* Holden-Day, San Francisco. Repr. Springer, NY 1983.

Freidlin, M.I., Wentzell, A.D. (1970). On small random permutations of dynamical systems. *Russian Math. Surveys* **25**, 1–55.

— (1998). *Random Perturbations of Dynamical Systems.* Engl. trans., Springer, NY. (Russian orig. 1979.)

Frostman, O. (1935). Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions. *Medd. Lunds Univ. Mat. Sem.* **3**, 1-118.

Fubini, G. (1907). Sugli integrali multipli. *Rend. Acc. Naz. Lincei* **16**, 608–614.

Furstenberg, H., Kesten, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31**, 457–469.

GAIFMAN, H. (1961). Concerning measures in first order calculi. *Israel J. Math.* **2**, 1–18.

GALMARINO, A.R. (1963). Representation of an isotropic diffusion as a skew product. *Z. Wahrsch. verw. Geb.* **1**, 359–378.

GARSIA, A.M. (1965). A simple proof of E. Hopf's maximal ergodic theorem. *J. Math. Mech.* **14**, 381–382.

— (1973). *Martingale Inequalities: Seminar Notes on Recent Progress.* Math. Lect. Notes Ser. Benjamin, Reading, MA.

GAUSS, C.F. (1809/16). *Theory of Motion of the Heavenly Bodies.* Engl. trans., Dover, NY 1963.

— (1840). Allgemeine Lehrsätze in Beziehung auf die im vehrkehrten Verhältnisse des Quadrats der Entfernung wirkenden Anziehungs- und Abstossungs-Kräfte. *Gauss Werke* **5**, 197–242. Göttingen 1867.

GAVEAU, B., TRAUBER, P. (1982). L'intégrale stochastique comme opérateur de divergence dans l'espace fonctionnel. *J. Functional Anal.* **46**, 230–238.

GETOOR, R.K. (1990). *Excessive Measures.* Birkhäuser, Boston.

GETOOR, R.K., SHARPE, M.J. (1972). Conformal martingales. *Invent. Math.* **16**, 271–308.

GIBBS, J.W. (1902). *Elementary Principles in Statistical Mechanics.* Yale Univ. Press.

GIHMAN, I.I. (1947). On a method of constructing random processes (Russian). *Dokl. Akad. Nauk SSSR* **58**, 961–964.

— (1950/51). On the theory of differential equations for random processes, I–II (Russian). *Ukr. Mat. J.* **2**:4, 37–63; **3**:3, 317–339.

GIRSANOV, I.V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Th. Probab. Appl.* **5**, 285–301.

GLIVENKO, V.I. (1933). Sulla determinazione empirica della leggi di probabilità. *Giorn. Ist. Ital. Attuari* **4**, 92–99.

GNEDENKO, B.V. (1939). On the theory of limit theorems for sums of independent random variables (Russian). *Izv. Akad. Nauk SSSR Ser. Mat.* 181–232, 643–647.

GNEDENKO, B.V., KOLMOGOROV, A.N. (1954/68). *Limit Distributions for Sums of Independent Random Variables.* Engl. trans., 2nd ed., Addison-Wesley, Reading, MA. (Russian orig. 1949.)

GNEDIN, A.V. (1997). The representation of composition structures. *Ann. Probab.* **25**, 1437–1450.

GOLDMAN, J.R. (1967). Stochastic point processes: limit theorems. *Ann. Math. Statist.* **38**, 771–779.

GOLDSTEIN, J.A. (1976). Semigroup-theoretic proofs of the central limit theorem and other theorems of analysis. *Semigroup Forum* **12**, 189–206.

GOLDSTEIN, S. (1979). Maximal coupling. *Z. Wahrsch. verw. Geb.* **46**, 193–204.

GRANDELL, J. (1976). *Doubly Stochastic Poisson Processes. Lect. Notes in Math.* **529**. Springer, Berlin.

GREEN, G. (1828). An essay on the application of mathematical analysis to the theories of electricity and magnetism. Repr. in *Mathematical Papers*, Chelsea, NY 1970.

GREENWOOD, P., PITMAN, J. (1980). Construction of local time and Poisson point processes from nested arrays. *J. London Math. Soc.* (2) **22**, 182–192.

Griffeath, D. (1975). A maximal coupling for Markov chains. *Z. Wahrsch. verw. Geb.* **31**, 95–106.

Grigelionis, B. (1963). On the convergence of sums of random step processes to a Poisson process. *Th. Probab. Appl.* **8**, 172–182.

— (1971). On the representation of integer-valued measures by means of stochastic integrals with respect to Poisson measure. *Litovsk. Mat. Sb.* **11**, 93–108.

Haar, A. (1933). Der Maßbegriff in der Theorie der kontinuerlichen Gruppen. *Ann. Math.* **34**, 147–169.

Hagberg, J. (1973). Approximation of the summation process obtained by sampling from a finite population. *Th. Probab. Appl.* **18**, 790–803.

Hahn, H. (1921). *Theorie der reellen Funktionen.* Julius Springer, Berlin.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**, 361–374.

Hall, P., Heyde, C.C. (1980). *Martingale Limit Theory and its Application.* Academic Press, NY.

Halmos, P.R. (1950). *Measure Theory*, Van Nostrand, Princeton. Repr. Springer, NY 1974.

— (1985). *I want to be a Mathematician, an Automathography.* Springer, NY.

Hardy, G.H., Littlewood, J.E. (1930). A maximal theorem with function-theoretic applications. *Acta Math.* **54**, 81–116.

Harris, T.E. (1956). The existence of stationary measures for certain Markov processes. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* **2**, 113–124.

— (1963/89). *The Theory of Branching Processes.* Dover Publ., NY.

— (1971). Random measures and motions of point processes. *Z. Wahrsch. verw. Geb.* **18**, 85–115.

Hartman, P., Wintner, A. (1941). On the law of the iterated logarithm. *J. Math.* **63**, 169–176.

Hausdorff, F. (1916). Die Mächtigkeit der Borelschen Mengen. *Math. Ann.* **77**, 430–437.

Helly, E. (1911/12). Über lineare Funktionaloperatoren. *Sitzungsber. Nat. Kais. Akad. Wiss.* **121**, 265–297.

Hewitt, E., Ross, K.A. (1979). *Abstract Harmonic Analysis I*, 2nd ed. Springer NY.

Hewitt, E., Savage, L.J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470–501.

Hille, E. (1948). Functional analysis and semi-groups. *Amer. Math. Colloq. Publ.* **31**, NY.

Hitczenko, P. (1988). Comparison of moments for tangent sequences of random variables. *Probab. Th. Rel. Fields* **78**, 223–230.

— (1990). Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* **18**, 1656–1668.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19**, 293–325.

Hoeven, P.C.T. van der (1983). *On Point Processes.* Mathematisch Centrum, Amsterdam.

Hölder, O. (1889). Über einen Mittelwertsatz. *Nachr. Akad. Wiss. Göttingen, math.-phys. Kl.*, 38–47.

Hoover, D.N. (1979). Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton.

Hopf, E. (1954). The general temporally discrete Markov process. *J. Rat. Mech. Anal.* **3**, 13–45.

Hörmander, L. (1967). Hypoelliptic second order differential equations. *Acta Math.* **119**, 147–171.

Horowitz, J. (1972). Semilinear Markov processes, subordinators and renewal theory. *Z. Wahrsch. verw. Geb.* **24**, 167–193.

Hunt, G.A. (1956a). Some theorems concerning Brownian motion. *Trans. Amer. Math. Soc.* **81**, 294–319.

— (1956b). Semigroups of measures on Lie groups. *Trans. Amer. Math. Soc.* **81**, 264–293.

— (1957/58). Markoff processes and potentials, I–III. *Illinois J. Math.* **1**, 44–93, 316–369; **2**, 151–213.

Hurewicz, W. (1944). Ergodic theorem without invariant measure. *Ann. Math.* **45**, 192–206.

Hurwitz, A. (1897). Über die Erzeugung der Invarianten durch Integration. *Nachr. Ges. Göttingen, math.-phys. Kl.*, 71–90.

Ikeda, N., Watanabe, S. (1989/2014). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. Kodansha and Elsevier.

Ioffe, D. (1991). On some applicable versions of abstract large deviations theorems. *Ann. Probab.* **19**, 1629–1639.

Ionescu Tulcea, A. (1960). Contributions to information theory for abstract alphabets. *Ark. Mat.* **4**, 235–247.

Ionescu Tulcea, C.T. (1949/50). Mesures dans les espaces produits. *Atti Accad. Naz. Lincei Rend.* **7**, 208–211.

Itô, K. (1942a). Differential equations determining Markov processes (Japanese). *Zenkoku Shijō Sūgaku Danwakai* **244**:1077, 1352–1400.

— (1942b). On stochastic processes I: Infinitely divisible laws of probability. *Jap. J. Math.* **18**, 261–301.

— (1944). Stochastic integral. *Proc. Imp. Acad. Tokyo* **20**, 519–524.

— (1946). On a stochastic integral equation. *Proc. Imp. Acad. Tokyo* **22**, 32–35.

— (1950). Stochastic differential equations in a differentiable manifold. *Nagoya Math. J.* **1**, 35–47.

— (1951a). On a formula concerning stochastic differentials. *Nagoya Math. J.* **3**, 55–65.

— (1951b). On stochastic differential equations. *Mem. Amer. Math. Soc.* **4**, 1–51.

— (1951c). Multiple Wiener integral. *J. Math. Soc. Japan* **3**, 157–169.

— (1956). Spectral type of the shift transformation of differential processes with stationary increments. *Trans. Amer. Math. Soc.* **81**, 253–263.

— (1957). *Stochastic Processes* (Japanese). Iwanami Shoten, Tokyo.

— (1972). Poisson point processes attached to Markov processes. *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **3**, 225–239.

Itô, K., McKean, H.P. (1965). *Diffusion Processes and their Sample Paths.* Repr. Springer, Berlin 1996.

Itô, K., Watanabe, S. (1965). Transformation of Markov processes by multiplicative functionals. *Ann. Inst. Fourier* **15**, 15–30.

Jacobsen, M. (2006). *Point Process Theory and Applications.* Birkhäuser, Boston.

Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivative, representation of martingales. *Z. Wahrsch. verw. Geb.* **31**, 235–253.

— (1979). *Calcul Stochastique et Problèmes de Martingales. Lect. Notes in Math.* **714**. Springer, Berlin.

— (1983). Processus à accroissements indépendants: use condition nécessaire et suffisante de convergence en loi. *Z. Wahrsch. verw. Geb.* **63**, 109–136.

— (1984). Une généralisation des semimartingales: les processus admettant un processus à accroissements indépendants tangent. *Lect. Notes in Math.* **1059**, 91–118. Springer, Berlin.

Jacod, J., Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer, Berlin.

Jagers, P. (1972). On the weak convergence of superpositions of point processes. *Z. Wahrsch. verw. Geb.* **22**, 1–7.

— (1974). Aspects of random measures and point processes. *Adv. Probab. Rel. Topics* **3**, 179–239. Marcel Dekker, NY.

— (1975). *Branching Processes with Biological Applications.* Wiley.

Jamison, B., Orey, S. (1967). Markov chains recurrent in the sense of Harris. *Z. Wahrsch. verw. Geb.* **8**, 206–223.

Jensen, J.L.W.V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **30**, 175–193.

Jessen, B. (1934). The theory of integration in a space of an infinite number of dimensions. *Acta Math.* **63**, 249–323.

Jiřina, M. (1964). Branching processes with measure-valued states. In: *Trans. 3rd Prague Conf. Inf. Th. Statist. Dec. Func. Rand. Proc.*, pp. 333–357.

— (1966). Asymptotic behavior of measure-valued branching processes. *Rozpravy Československé Akad. Věd. Řada Mat. Přirod. Věd* **76**:3, Praha.

Johnson, W.B., Schechtman, G., Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13**, 234–253.

Jordan, C. (1881). Sur la série de Fourier. *C.R. Acad. Sci. Paris* **92**, 228–230.

Kac, M. (1947). On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math. Soc.* **53**, 1002–1010.

— (1949). On distributions of certain Wiener functionals. *Trans. Amer. Math. Soc.* **65**, 1–13.

— (1951). On some connections between probability theory and differential and integral equations. *Proc. 2nd Berkeley Symp. Math. Statist. Probab.*, 189–215. Univ. of California Press, Berkeley.

Kakutani, S. (1940). Ergodic theorems and the Markoff process with a stable distribution. *Proc. Imp. Acad. Tokyo* **16**, 49–54.

— (1944a). On Brownian motions in n-space. *Proc. Imp. Acad. Tokyo* **20**, 648–652.

— (1944b). Two-dimensional Brownian motion and harmonic functions. *Proc. Imp. Acad. Tokyo* **20**, 706–714.

— (1945). Markoff process and the Dirichlet problem. *Proc. Japan Acad.* **21**, 227–233.

Kallenberg, O. (1973a). Characterization and convergence of random measures and point processes. *Z. Wahrsch. verw. Geb.* **27**, 9–21.

— (1973b). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrsch. verw. Geb.* **27**, 23–36.

— (1974). Series of random processes without discontinuities of the second kind. *Ann. Probab.* **2**, 729–737.

— (1975a). Limits of compound and thinned point processes. *J. Appl. Probab.* **12**, 269–278.

— (1975/76, 1983/86). *Random Measures*, 1st–2nd eds, 3rd–4th eds. Akademie-Verlag & Academic Press, Berlin, London.

— (1976/80/81). On the structure of stationary flat processes, I–III. *Z. Wahrsch. verw. Geb.* **37**, 157–174; **52**, 127–147; **56**, 239–253.

— (1977). Stability of critical cluster fields. *Math. Nachr.* **77**, 1–43.

— (1978a). On conditional intensities of point processes. *Z. Wahrsch. verw. Geb.* **41**, 205–220.

— (1978b). On the asymptotic behavior of line processes and systems of non-interacting particles. *Z. Wahrsch. verw. Geb.* **43**, 65–95.

— (1978c). On the independence of velocities in a system of non-interacting particles. *Ann. Probab.* **6**, 885–890.

— (1987). Homogeneity and the strong Markov property. *Ann. Probab.* **15**, 213–240.

— (1988a). Spreading and predictable sampling in exchangeable sequences and processes. *Ann. Probab.* **16**, 508–534.

— (1988b). Some new representation in bivariate exchangeability. *Probab. Th. Rel. Fields* **77**, 415–455.

— (1989). On the representation theorem for exchangeable arrays. *J. Multivar. Anal.* **30**, 137–154.

— (1990). Random time change and an integral representation for marked stopping times. *Probab. Th. Rel. Fields* **86**, 167–202.

— (1992). Some time change representations of stable integrals, via predictable transformations of local martingales. *Stoch. Proc. Appl.* **40**, 199–223.

— (1992). Symmetries on random arrays and set-indexed processes. *J. Theor. Probab.* **5**, 727–765.

— (1995). Random arrays and functionals with multivariate rotational symmetries. *Probab. Th. Rel. Fields* **103**, 91–141.

— (1996a). On the existence of universal functional solutions to classical SDEs. *Ann. Probab.* **24**, 196–205.

— (1996b). Improved criteria for distributional convergence of point processes. *Stoch. Proc. Appl.* **64**, 93-102.

— (1997/2002). *Foundations of Modern Probability*, eds. 1–2. Springer, NY.

— (1999a). Ballot theorems and sojourn laws for stationary processes. *Ann. Probab.* **27**, 2011–2019.

— (1999b). Asymptotically invariant sampling and averaging from stationary-like processes. *Stoch. Proc. Appl.* **82**, 195–204.

— (1999c). Palm measure duality and conditioning in regenerative sets. *Ann. Probab.* **27**, 945–969.

— (2005). *Probabilistic Symmetries and Invariance Principles.* Springer.

— (2007). Invariant measures and disintegrations with applications to Palm and related kernels. *Probab. Th. Rel. Fields* **139**, 285–310, 311.

— (2009). Some local approximation properties of simple point processes. *Probab. Th. Rel. Fields* **143**, 73–96.

— (2010). Commutativity properties of conditional distributions and Palm measures. *Comm. Stoch. Anal.* **4**, 21–34.

— (2011a). Invariant Palm and related disintegrations via skew factorization. *Probab. Th. Rel. Fields* **149**, 279–301.

— (2011b). Iterated Palm conditioning and some Slivnyak-type theorems for Cox and cluster processes. *J. Theor. Probab.* **24**, 875–893.

— (2017a). Tangential existence and comparison, with applications to single and multiple integration. *Probab. Math. Statist.* **37**, 21–52.

— (2017b). *Random Measures, Theory and Applications.* Springer.

Kallenberg, O., Sztencel, R. (1991). Some dimension-free features of vector-valued martingales. *Probab. Th. Rel. Fields* **88**, 215–247.

Kallenberg, O., Szulga, J. (1989). Multiple integration with respect to Poisson and Lévy processes. *Probab. Th. Rel. Fields* **83**, 101–134.

Kallianpur, G. (1980). *Stochastic Filtering Theory.* Springer, NY.

Kaplan, E.L. (1955). Transformations of stationary random sequences. *Math. Scand.* **3**, 127–149.

Karamata, J. (1930). Sur une mode de croissance régulière des fonctions. *Mathematica (Cluj)* **4**, 38–53.

Karatzas, I., Shreve, S.E. (1988/91/98). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, NY.

Kazamaki, N. (1972). Change of time, stochastic integrals and weak martingales. *Z. Wahrsch. verw. Geb.* **22**, 25–32.

Kemeny, J.G., Snell, J.L., Knapp, A.W. (1966). *Denumerable Markov Chains.* Van Nostrand, Princeton.

Kendall, D.G. (1948). On the generalized "birth-and-death" process. *Ann. Math. Statist.* **19**, 1–15.

— (1974). Foundations of a theory of random sets. In *Stochastic Geometry* (eds. E.F. Harding, D.G. Kendall), pp. 322–376. Wiley, NY.

Kendall, W.S. (1988). Martingales on manifolds and harmonic maps. Geometry of random motion. *Contemp. Math.* **73**, 121–157.

Kerstan, J., Matthes, K. (1964). Stationäre zufällige Punktfolgen II. *Jahresber. Deutsch. Mat.-Verein.* **66**, 106–118.

Khinchin, A.Y. (1923). Über dyadische Brücke. *Math. Z.* **18**, 109–116.

— (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.* **6**, 9–20.

— (1929). Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Ann.* **101**, 745–752.

— (1933a). Zur mathematischen Begründing der statistischen Mechanik. *Z. Angew. Math. Mech.* **13**, 101–103.

— (1933b). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Springer, Berlin. Repr. Chelsea, NY 1948.

— (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* **109**, 604–615.

— (1937). Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. *Mat. Sb.* **2**, 79–119.

— (1938). *Limit Laws for Sums of Independent Random Variables* (Russian). Moscow.

— (1955). *Mathematical Methods in the Theory of Queuing* (Russian). Engl. trans., Griffin, London 1960.

Khinchin, A.Y., Kolmogorov, A.N. (1925). Über Konvergenz von Reihen deren Glieder durch den Zufall bestimmt werden. *Mat. Sb.* **32**, 668–676.

Kingman, J.F.C. (1964). On doubly stochastic Poisson processes. *Proc. Cambridge Phil. Soc.* **60**, 923–930.

— (1967). Completely random measures. *Pac.. J. Math.* **21**, 59–78.

— (1968). The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. (B)* **30**, 499–510.

— (1972). *Regenerative Phenomena.* Wiley, NY.

— (1978). The representation of partition structures. *J. London Math, Soc.* **18**, 374–380.

— (1993). *Poisson Processes.* Clarendon Press, Oxford.

Kinney, J.R. (1953). Continuity properties of Markov processes. *Trans. Amer. Math. Soc.* **74**, 280–302.

Klebaner, F.C. (2012). *Introduction to Stochastic Calculus with Applications*, 3rd ed. Imperial Coll. Press, London.

Knight, F.B. (1963). Random walks and a sojourn density process of Brownian motion. *Trans. Amer. Math. Soc.* **107**, 56–86.

— (1971). A reduction of continuous, square-integrable martingales to Brownian motion. *Lect. Notes in Math.* **190**, 19–31. Springer, Berlin.

Kolmogorov, A.N. (1928/29). Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Math. Ann.* **99**, 309–319; **102**, 484–488.

— (1929). Über das Gesatz des iterierten Logarithmus. *Math. Ann.* **101**, 126–135.

— (1930). Sur la loi forte des grandes nombres. *C.R. Acad. Sci. Paris* **191**, 910–912.

— (1931a). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* **104**, 415–458.

— (1931b). Eine Verallgemeinerung des Laplace–Liapounoffschen Satzes. *Izv. Akad. Nauk USSR, Otdel. Matem. Yestestv. Nauk* **1931**, 959–962.

— (1932). Sulla forma generale di un processo stocastico omogeneo (un problema di B. de Finetti). *Atti Accad. Naz. Lincei Rend.* (6) **15**, 805–808, 866–869.

— (1933a). Über die Grenzwertsätze der Wahrscheinlichkeitsrechnung. *Izv. Akad. Nauk USSR, Otdel. Matem. Yestestv. Nauk* **1933**, 363–372.

— (1933b). Zur Theorie der stetigen zufälligen Prozesse. *Math. Ann.* **108**, 149–160.

— (1933c). *Foundations of the Theory of Probability* (German), Springer, Berlin. Engl. trans., Chelsea, NY 1956.

— (1935). Some current developments in probability theory (in Russian). *Proc. 2nd All-Union Math. Congr.* **1**, 349–358. Akad. Nauk SSSR, Leningrad.

— (1936a). Anfangsgründe der Markoffschen Ketten mit unendlich vielen möglichen Zuständen. *Mat. Sb.* **1**, 607–610.

— (1936b). Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112**, 155–160.

— (1937). Zur Umkehrbarkeit der statistischen Naturgesetze. *Math. Ann.* **113**, 766–772.

— (1938). Zur Lösung einer biologischen Aufgabe. *Izv. Nauchno—Issled. Inst. Mat. Mech. Tomsk. Gosud. Univ.* **2**, 1–6.

— (1956). On Skorohod convergence. *Th. Probab. Appl.* **1**, 213–222.

Kolmogorov, A.N., Leontovich, M.A. (1933). Zur Berechnung der mittleren Brownschen Fläche. *Physik. Z. Sowjetunion* **4**, 1–13.

Komlós, J., Major, P., Tusnády, G. (1975/76). An approximation of partial sums of independent r.v.'s and the sample d.f., I–II. *Z. Wahrsch. verw. Geb.* **32**, 111–131; **34**, 33–58.

König, D., Matthes, K. (1963). Verallgemeinerung der Erlangschen Formeln, I. *Math. Nachr.* **26**, 45–56.

Koopman, B.O. (1931). Hamiltonian systems and transformations in Hilbert space. *Proc. Nat. Acad. Sci. USA* **17**, 315–318.

Krauss, P.H. (1969). Representations of symmetric probability models. *J. Symbolic Logic* **34**, 183–193.

Krengel, U. (1985). *Ergodic Theorems.* de Gruyter, Berlin.

Krickeberg, K. (1956). Convergence of martingales with a directed index set. *Trans. Amer. Math. Soc.* **83**, 313–357.

— (1972). The Cox process. *Symp. Math.* **9**, 151–167.

— (1974a). Invariance properties of the correlation measure of line-processes. In: *Stochastic Geometry* (E.F. Harding & D.G. Kendall, eds.), pp. 76–88. Wiley, London.

— (1974b). Moments of point processes. *Ibid.*, pp. 89–113.

Krylov, N., Bogolioubov, N. (1937). La théorie générale de la mesure dans son application à l'étude des systèmes de la mécanique non linéaires. *Ann. Math.* **38**, 65–113.

Kullback, S., Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22**, 79–86.

Kunita, H. (1990). *Stochastic Flows and Stochastic Differential Equations.* Cambridge Univ. Press, Cambridge.

Kunita, H., Watanabe, S. (1967). On square integrable martingales. *Nagoya Math. J.* **30**, 209–245.

Kunze, M. (2013). *An Introduction to Malliavin Calculus.* Lecture notes, Universität Ulm.

Kuratowski, K. (1933). *Topologie I.* Monografje Matem., Warsaw & Lwow.

Kurtz, T.G. (1969). Extensions of Trotter's operator semigroup approximation theorems. *J. Funct. Anal.* **3**, 354–375.

— (1975). Semigroups of conditioned shifts and approximation of Markov processes. *Ann. Probab.* **3**, 618–642.

Kwapień, S., Woyczyński, W.A. (1992). *Random Series and Stochastic Integrals: Single and Multiple.* Birkhäuser, Boston.

Lanford, O.E., Ruelle, D. (1969). Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.* **13**, 194–215.

Langevin, P. (1908). Sur la théorie du mouvement brownien. *C.R. Acad. Sci. Paris* **146**, 530–533.

Laplace, P.S. de (1774). Mémoire sur la probabilité des causes par les événemens. Engl. trans. in *Statistical Science* **1**, 359–378.

— (1809). Mémoire sur divers points d'analyse. Repr. in *Oeuvres Complètes de Laplace* **14**, 178–214. Gauthier-Villars, Paris 1886–1912.

— (1812/20). *Théorie Analytique des Probabilités*, 3rd ed. Repr. in *Oeuvres Complètes de Laplace* **7**. Gauthier-Villars, Paris 1886–1912.

Last, G., Brandt, A. (1995). *Marked Point Processes on the Real Line: The Dynamic Approach.* Springer, NY.

Last, G., Penrose, M.D. (2017). *Lectures on the Poisson Process.* Cambridge Univ. Press.

Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, NY.

Lebesgue, H. (1902). Intégrale, longeur, aire. *Ann. Mat. Pura Appl.* **7**, 231–359.

— (1904). *Leçons sur l'Intégration et la Recherche des Fonctions Primitives.* Paris.

Le Cam, L. (1957). Convergence in distribution of stochastic processes. *Univ. California Publ. Statist.* **2**, 207–236.

Le Gall, J.F. (1983). Applications des temps locaux aux équations différentielles stochastiques unidimensionelles. *Lect. Notes in Math.* **986**, 15–31.

— (1991). Brownian excursions, trees and measure-valued branching processes. *Ann. Probab.* **19**, 1399–1439.

Levi, B. (1906a). Sopra l'integrazione delle serie. *Rend. Ist. Lombardo Sci. Lett.* (2) **39**, 775–780.

— (1906b). Sul principio de Dirichlet. *Rend. Circ. Mat. Palermo* **22**, 293–360.

Lévy, P. (1922a). Sur le rôle de la loi de Gauss dans la théorie des erreurs. *C.R. Acad. Sci. Paris* **174**, 855–857.

— (1922b). Sur la loi de Gauss. *C.R. Acad. Sci. Paris* 1682–1684.

— (1922c). Sur la détermination des lois de probabilité par leurs fonctions caractéristiques. *C.R. Acad. Sci. Paris* **175**, 854–856.

— (1924). Théorie des erreurs. La loi de Gauss et les lois exceptionelles. *Bull. Soc. Math. France* **52**, 49–85.

— (1925). *Calcul des Probabilités.* Gauthier-Villars, Paris.

— (1934/35). Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Ann. Scuola Norm. Sup. Pisa* (2) **3**, 337–366; **4**, 217–218.

— (1935a). Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchainées. *J. Math. Pures Appl.* (8) **14**, 347–402.

— (1935b). Propriétés asymptotiques des sommes de variables aléatoires enchainées. *Bull. Sci. Math.* (2) **59**, 84–96, 109–128.

— (1937/54). *Théorie de l'Addition des Variables Aléatoires*, 2nd ed. Gauthier-Villars, Paris (1st ed. 1937).

— (1939). Sur certain processus stochastiques homogènes. *Comp. Math.* **7**, 283–339.

— (1940). Le mouvement brownien plan. *Amer. J. Math.* **62**, 487–550.

— (1948/65). *Processus Stochastiques et Mouvement Brownien*, 2nd ed. Gauthier-Villars, Paris (1st ed. 1948).

Liao, M. (2018). *Invariant Markov Processes under Lie Group Action.* Springer Intern. Publ.

Liapounov, A.M. (1901). Nouvelle forme du théorème sur la limite des probabilités. *Mem. Acad. Sci. St. Petersbourg* **12**, 1–24.

Liemant, A., Matthes, K., Wakolbinger, A. (1988). *Equilibrium Distributions of Branching Processes.* Akademie-Verlag, Berlin.

Liggett, T.M. (1985). An improved subadditive ergodic theorem. *Ann. Probab.* **13**, 1279–1285.

Lindeberg, J.W. (1922a). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Zeitschr.* **15**, 211–225.

— (1922b). Sur la loi de Gauss. *C.R. Acad. Sci. Paris* **174**, 1400–1402.

Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *J. Appl. Probab.* **10**, 109–121.

— (1977). A probabilistic proof of Blackwell's renewal theorem. *Ann. Probab.* **5**, 482–485.

— (1992). *Lectures on the Coupling Method.* Wiley, NY.

Loève, M. (1977/78). *Probability Theory* **1–2**, 4th ed. Springer, NY (eds. 1–3, 1955/63).

Łomnicki, Z., Ulam, S. (1934). Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités: I. Variables indépendantes. *Fund. Math.* **23**, 237–278.

Lukacs, E. (1960/70). *Characteristic Functions*, 2nd ed. Griffin, London.

Lundberg, F. (1903). *Approximate representation of the probability function. Reinsurance of collective risks* (in Swedish). Ph.D. thesis, Almqvist & Wiksell, Uppsala.

Mackevičius, V. (1974). On the question of the weak convergence of random processes in the space $D[0,\infty)$. *Lithuanian Math. Trans.* **14**, 620–623.

Maisonneuve, B. (1974). *Systèmes Régénératifs. Astérique* **15**. Soc. Math. de France.

Maker, P. (1940). The ergodic theorem for a sequence of functions. *Duke Math. J.* **6**, 27–30.

Malliavin, P. (1978). Stochastic calculus of variations and hypoelliptic operators. In: *Proc. Intern. Symp. on Stoch. Diff. Equations, Kyoto 1976*, pp. 195–263. Wiley 1978.

Mann, H.B., Wald, A. (1943). On stochastic limit and order relations. *Ann. Math. Statist.* **14**, 217–226.

Marcinkiewicz, J., Zygmund, A. (1937). Sur les fonctions indépendantes. *Fund. Math.* **29**, 60–90.

— (1938). Quelques théorèmes sur les fonctions indépendantes. *Studia Math.* **7**, 104–120.

Markov, A.A. (1899). The law of large numbers and the method of least squares (Russian). *Izv. Fiz.-Mat. Obshch. Kazan Univ.* (2) **8**, 110–128.

— (1906). Extension of the law of large numbers to dependent events (Russian). *Bull. Soc. Phys. Math. Kazan* (2) **15**, 135–156.

Maruyama, G. (1954). On the transition probability functions of the Markov process. *Natl. Sci. Rep. Ochanomizu Univ.* **5**, 10–20.

— (1955). Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo* **4**, 48–90.

Maruyama, G., Tanaka, H. (1957). Some properties of one-dimensional diffusion processes. *Mem. Fac. Sci. Kyushu Univ.* **11**, 117–141.

Matheron, G. (1975). *Random Sets and Integral Geometry.* Wiley, London.

Matthes, K. (1963). Stationäre zufällige Punktfolgen, I. *Jahresber. Deutsch. Math.-Verein.* **66**, 66–79.

Matthes, K., Kerstan, J., Mecke, J. (1974/78/82). *Infinitely Divisible Point Processes.* Wiley, Chichester (German/English/Russian eds.).

Matthes, K., Warmuth, W., Mecke, J. (1979). Bemerkungen zu einer Arbeit von Nguyen Xuan Xanh und Hans Zessin. *Math. Nachr.* **88**, 117–127.

Maxwell, J.C. (1875). *Theory of Heat*, 4th ed. Longmans, London.

— (1878). On Boltzmann's theorem on the average distribution of energy in a system of material points. *Trans. Cambridge Phil. Soc.* **12**, 547.

Mazurkiewicz, S. (1916). Über Borelsche Mengen. *Bull. Acad. Cracovie* **1916**, 490–494.

McGinley, W.G., Sibson, R. (1975). Dissociated random variables. *Math. Proc. Cambridge Phil. Soc.* **77**, 185–188.

McKean, H.P. (1969). *Stochastic Integrals.* Academic Press, NY.

McKean, H.P., Tanaka, H. (1961). Additive functionals of the Brownian path. *Mem. Coll. Sci. Univ. Kyoto, A* **33**, 479–506.

McMillan, B. (1953). The basic theorems of information theory. *Ann. Math. Statist.* **24**, 196–219.

Mecke, J. (1967). Stationäre zufällige Maße auf lokalkompakten Abelschen Gruppen. *Z. Wahrsch. verw. Geb.* **9**, 36–58.

— (1968). Eine characteristische Eigenschaft der doppelt stochastischen Poissonschen Prozesse. *Z. Wahrsch. verw. Geb.* **11**, 74–81.

Mehta, M.L. (1991). *Random Matrices.* Academic Press.

Méléard, S. (1986). Application du calcul stochastique à l'étude des processus de Markov réguliers sur $[0,1]$. *Stochastics* **19**, 41–82.

Métivier, M. (1982). *Semimartingales: A Course on Stochastic Processes.* de Gruyter, Berlin.

Métivier, M., Pellaumail, J. (1980). *Stochastic Integration.* Academic Press, NY.

Meyer, P.A. (1962). A decomposition theorem for supermartingales. *Illinois J. Math.* **6**, 193–205.

— (1963). Decomposition of supermartingales: the uniqueness theorem. *Illinois J. Math.* **7**, 1–17.

— (1966). *Probability and Potentials.* Engl. trans., Blaisdell, Waltham.

— (1967). Intégrales stochastiques, I–IV. *Lect. Notes in Math.* **39**, 72–162. Springer, Berlin.

— (1968). Guide détaillé de la théorie générale des processus. *Lect. Notes in Math.* **51**, 140–165. Springer.

— (1971). Démonstration simplifiée d'un théorème de Knight. *Lect. Notes in Math.* **191**, 191–195. Springer, Berlin.

— (1976). Un cours sur les intégrales stochastiques. *Lect. Notes in Math.* **511**, 245–398. Springer, Berlin.

— (1981). Géométrie stochastique sans larmes. *Lecture Notes in Math.* **850**, 44–102.

— (1982). Géométrie différentielle stochastique (bis). *Lecture Notes in Math.* **921**, 165–207.

Miles, R.E. (1969/71). Poisson flats in Euclidean spaces, I–II. *Adv. Appl. Probab.* **1**, 211–237; **3**, 1–43.

— (1974). A synopsis of 'Poisson flats in Euclidean spaces'. In: *Stochastic Geometry* (E.F. Harding & D.G. Kendall, eds), pp. 202–227. Wiley, London.

Millar, P.W. (1968). Martingale integrals. *Trans. Amer. Math. Soc.* **133**, 145–166.

Minkowski, H. (1907). *Diophantische Approximationen.* Teubner, Leipzig.

Mitoma, I. (1983). Tightness of probabilities on $C([0,1];\mathcal{S}')$ and $D([0,1];\mathcal{S}')$. *Ann. Probab.* **11**, 989–999.

Moivre, A. de (1711). *De Mensura Sortis* (On the measurement of chance). Engl. trans., *Int. Statist. Rev.* **52** (1984), 229–262.

— (1738). *The Doctrine of Chances*, 2nd ed. Repr. Case and Chelsea, London and NY 1967.

Molchanov, I. (2005). *Theory of Random Sets.* Springer, London.

Mönch, G. (1971). Verallgemeinerung eines Satzes von A. Rényi. *Studia Sci. Math. Hung.* **6**, 81–90.

Motoo, M., Watanabe, H. (1958). Ergodic property of recurrent diffusion process in one dimension. *J. Math. Soc. Japan* **10**, 272–286.

Moyal, J.E. (1962). The general theory of stochastic population processes. *Acta Math.* **108**, 1–31.

Nawrotzki, K. (1962). Ein Grenzwertsatz für homogene zufällige Punktfolgen (Verallgemeinerung eines Satzes von A. Rényi). *Math. Nachr.* **24**, 201–217.

— (1968). Mischungseigenschaften stationärer unbegrenzt teilbarer zufälliger Maße. *Math. Nachr.* **38**, 97–114.

Neumann, J. von (1932). Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA* **18**, 70–82.

— (1940). On rings of operators, III. *Ann. Math.* **41**, 94–161.

Neveu, J. (1971). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

— (1975). *Discrete-Parameter Martingales.* North-Holland, Amsterdam.

Nguyen, X.X., Zessin, H. (1979). Ergodic theorems for spatial processes. *Z. Wahrsch. verw. Geb.* **48**, 133–158.

NIEUWENHUIS, G. (1994). Bridging the gap between a stationary point process and its Palm distribution. *Statist. Nederl.* **48**, 37–62.

NIKODYM, O.M. (1930). Sur une généralisation des intégrales de M. J. Radon. *Fund. Math.* **15**, 131–179.

NORBERG, T. (1984). Convergence and existence of random set distributions. *Ann. Probab.* **12**, 726–732.

NOVIKOV, A.A. (1971). On moment inequalities for stochastic integrals. *Th. Probab. Appl.* **16**, 538–541.

— (1972). On an identity for stochastic integrals. *Th. Probab. Appl.* **17**, 717–720.

NUALART, D. (1995/2006). *The Malliavin Calculus and Related Topics.* Springer, NY.

OCONE, D. (1984). Malliavin calculus and stochastic integral representation of diffusion processes. *Stochastics* **12**, 161–185.

OLSON, W.H., UPPULURI, V.R.R. (1972). Asymptotic distribution of eigenvalues of random matrices. *Proc. 6th Berkely Symp. Math. Statist. Probab.* **3**, 615–644.

OREY, S. (1959). Recurrent Markov chains. *Pacific J. Math.* **9**, 805–827.

— (1962). An ergodic theorem for Markov chains. *Z. Wahrsch. verw. Geb.* **1**, 174–176.

— (1966). F-processes. *Proc. 5th Berkeley Symp. Math. Statist. Probab.* **2**:1, 301–313.

— (1971). *Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand, London.

ORNSTEIN, D.S. (1969). Random walks. *Trans. Amer. Math. Soc.* **138**, 1–60.

ORNSTEIN, L.S., UHLENBECK, G.E. (1930). On the theory of Brownian motion. *Phys. Review* **36**, 823–841.

OSOSKOV, G.A. (1956). A limit theorem for flows of homogeneous events. *Th. Probab. Appl.* **1**, 248–255.

OTTAVIANI, G. (1939). Sulla teoria astratta del calcolo delle probabilità proposita dal Cantelli. *Giorn. Ist. Ital. Attuari* **10**, 10–40.

PALEY, R.E.A.C. (1932). A remarkable series of orthogonal functions I. *Proc. London Math. Soc.* **34**, 241–264.

PALEY, R.E.A.C., WIENER, N. (1934). Fourier transforms in the complex domain. *Amer. Math. Soc. Coll. Publ.* **19**.

PALEY, R.E.A.C., WIENER, N., ZYGMUND, A. (1933). Notes on random functions. *Math. Z.* **37**, 647–668.

PALM, C. (1943). Intensity Variations in Telephone Traffic (German). *Ericsson Technics* **44**, 1–189. Engl. trans., *North-Holland Studies in Telecommunication* **10**, Elsevier 1988.

PAPANGELOU, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165**, 486–506.

— (1974a). On the Palm probabilities of processes of points and processes of lines. In: *Stochastic geometry* (E.F. Harding & D.G. Kendall, eds), pp. 114–147. Wiley, London.

— (1974b). The conditional intensity of general point processes and an application to line processes. *Z. Wahrsch. verw. Geb.* **28**, 207–226.

— (1976). Point processes on spaces of flats and other homogeneous spaces. *Math. Proc. Cambridge Phil. Soc.* **80**, 297–314.

Pardoux, E., Peng, S. (1990). Adapted solution of a backward stochastic differential equation. *Systems and Control Letters* **14**, 55–61.

Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces.* Academic Press, NY.

Peña, V.H. de la, Giné, E. (1999). *Decoupling: from Dependence to Independence.* Springer, NY.

Perkins, E. (1982). Local time and pathwise uniqueness for stochastic differential equations. *Lect. Notes in Math.* **920**, 201–208. Springer, Berlin.

Peterson, L.D. (2001). Convergence of random measures on Polish spaces. Dissertation, Auburn Univ.

Petrov, V.V. (1995). *Limit Theorems of Probability Theory.* Clarendon Press, Oxford.

Phillips, H.B., Wiener, N. (1923). Nets and Dirichlet problem. *J. Math. Phys.* **2**, 105–124.

Pitman, J.W. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Th. Rel. Fields* **102**, 145–158.

Pitt, H.R. (1942). Some generalizations of the ergodic theorem. *Proc. Camb. Phil. Soc.* **38**, 325–343.

Poincaré, H. (1890). Sur les équations aux dérivées partielles de la physique mathématique. *Amer. J. Math.* **12**, 211–294.

— (1899). *Théorie du Potentiel Newtonien.* Gauthier-Villars, Paris.

Poisson, S.D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités.* Bachelier, Paris.

Pollaczek, F. (1930). Über eine Aufgabe der Wahrscheinlichkeitstheorie, I–II. *Math. Z.* **32**, 64–100, 729–750.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, NY.

Pólya, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.* **8**, 171–181.

— (1921). Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz. *Math. Ann.* **84**, 149–160.

Port, S.C., Stone, C.J. (1978). *Brownian Motion and Classical Potential Theory.* Academic Press, NY.

Pospišil, B. (1935/36). Sur un problème de M.M.S. Bernstein et A. Kolmogoroff. *Časopis Pěst. Mat. Fys.* **65**, 64–76.

Prékopa, A. (1958). On secondary processes generated by a random point distribution of Poisson type. *Ann. Univ. Sci. Budapest Sect. Math.* **1**, 153–170.

Prohorov, Y.V. (1956). Convergence of random processes and limit theorems in probability theory. *Th. Probab. Appl.* **1**, 157–214.

— (1959). Some remarks on the strong law of large numbers. *Th. Probab. Appl.* **4**, 204–208.

— (1961). Random measures on a compactum. *Soviet Math. Dokl.* **2**, 539–541.

Protter, P. (1990). *Stochastic Integration and Differential Equations.* Springer, Berlin.

Pukhalsky, A.A. (1991). On functional principle of large deviations. In *New Trends in Probability and Statistics* (V. Sazonov and T. Shervashidze, eds.), 198–218. VSP Moks'las, Moscow.

Radon, J. (1913). Theorie und Anwendungen der absolut additiven Mengenfunktionen. *Wien Akad. Sitzungsber.* **122**, 1295–1438.

Ramer, R. (1974). On nonlinear transformations of Gaussian measures. *J. Functional Anal.* **15**, 166–187.

Rao, K.M. (1969a). On decomposition theorems of Meyer. *Math. Scand.* **24**, 66–78.

— (1969b). Quasimartingales. *Math. Scand.* **24**, 79–92.

Rauchenschwandtner, B. (1980). *Gibbsprozesse und Papangeloukerne.* Verb. wissensch. Gesellsch. Österreichs, Vienna.

Ray, D.B. (1956). Stationary Markov processes with continuous paths. *Trans. Amer. Math. Soc.* **82**, 452–493.

— (1963). Sojourn times of a diffusion process. *Illinois J. Math.* **7**, 615–630.

Rényi, A. (1956). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutato Int. Közl.* **1**, 519–527.

— (1967). Remarks on the Poisson process. *Studia Sci. Math. Hung.* **2**, 119–123.

Revuz, D. (1970). Mesures associées aux fonctionnelles additives de Markov, I–II. *Trans. Amer. Math. Soc.* **148**, 501–531; *Z. Wahrsch. verw. Geb.* **16**, 336–344.

— (1984). *Markov Chains*, 2nd ed. North-Holland, Amsterdam.

Revuz, D., Yor, M. (1991/94/99). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin.

Riesz, F. (1909a). Sur les suites de fonctions mesurables. *C.R. Acad. Sci. Paris* **148**, 1303–1305.

— (1909b). Sur les opérations fonctionelles linéaires. *C.R. Acad. Sci. Paris* **149**, 974–977.

— (1910). Untersuchungen über Systeme integrierbarer Funktionen. *Math. Ann.* **69**, 449–497.

— (1926/30). Sur les fonctions subharmoniques et leur rapport à la théorie du potentiel, I–II. *Acta Math.* **48**, 329–343; **54**, 321–360.

Rogers, C.A., Shephard, G.C. (1958). Some extremal problems for convex bodies. *Mathematica* **5**, 93–102.

Rogers, L.C.G., Williams, D. (1994/2000). *Diffusions, Markov Processes, and Martingales* **1–2**, 2nd ed. Cambridge Univ. Press.

Rosén, B. (1964). Limit theorems for sampling from a finite population. *Ark. Mat.* **5**, 383–424.

Rosiński, J., Woyczyński, W.A. (1986). On Itô stochastic integration with respect to p-stable motion: Inner clock, integrability of sample paths, double and multiple integrals. *Ann. Probab.* **14**, 271–286.

Rutherford, E., Geiger, H. (1908). An electrical method of counting the number of particles from radioactive substances. *Proc. Roy. Soc. A* **81**, 141–161.

Ryll-Nardzewski, C. (1957). On stationary sequences of random variables and the de Finetti's [sic] equivalence. *Colloq. Math.* **4**, 149–156.

— (1961). Remarks on processes of calls. *Proc. 4th Berkeley Symp. Math. Statist. Probab.* **2**, 455–465.

Sanov, I.N. (1957). On the probability of large deviations of random variables (Russian). Engl. trans.: *Sel. Trans. Math. Statist. Probab.* **1** (1961), 213–244.

Sato, K.-i. (1999/2013). *Lévy Processes and Infinitely Divisible Distributions*, 2nd ed. Cambridge Univ. Press.

Schilder, M. (1966). Some asymptotic formulae for Wiener integrals. *Trans. Amer. Math. Soc.* **125**, 63–85.

Schneider, R., Weil, W. (2008). *Stochastic and Integral Geometry.* Springer, Berlin.

Schoenberg, I.J. (1938). Metric spaces and completely monotone functions. *Ann. Math.* **39**, 811–841.

Schrödinger, E. (1931). Über die Umkehrung der Naturgesetze. *Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl.* 144–153.

Schuppen, J.H. van, Wong, E. (1974). Transformation of local martingales under a change of law. *Ann. Probab.* **2**, 879–888.

Schwartz, L. (1980). Semi-martingales sur des variétés et martingales conformes sur des variétés analytiques complexes. *Lect. Notes in Math.* **780**, Springer.

— (1982). Géométrie différentielle du 2^e ordre, semimartingales et équations différentielles stochastiques sur une variété différentielle. *Lect. Notes in Math.* **921**, 1–148. Springer

Segal, I.E. (1954). Abstract probability spaces and a theorem of Kolmogorov. *Amer. J. Math.* **76**, 721–732.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27**, 379–423, 623–656.

Sharpe, M. (1988). *General Theory of Markov Processes.* Academic Press, Boston.

Shiryaev, A.N. (1995). *Probability*, 2nd ed. Springer, NY.

Shur, M.G. (1961). Continuous additive functionals of a Markov process. *Dokl. Akad. Nauk SSSR* **137**, 800–803.

Sierpiński, W. (1928). Une théorème générale sur les familles d'ensemble. *Fund. Math.* **12**, 206–210.

Silverman, B.W. (1976). Limit theorems for dissociated random variables. *Adv. Appl. Probab.* **8**, 806–819.

Skorohod, A.V. (1956). Limit theorems for stochastic processes. *Th. Probab. Appl.* **1**, 261–290.

— (1957). Limit theorems for stochastic processes with independent increments. *Th. Probab. Appl.* **2**, 122–142.

— (1961/62). Stochastic equations for diffusion processes in a bounded region, I–II. *Th. Probab. Appl.* **6**, 264–274; **7**, 3–23.

— (1965). *Studies in the Theory of Random Processes.* Addison-Wesley, Reading, MA. (Russian orig. 1961.)

— (1975). On a generalization of a stochastic integral. *Theory Probab. Appl.* **20**, 219–233.

Slivnyak, I.M. (1962). Some properties of stationary flows of homogeneous random events. *Th. Probab. Appl.* **7**, 336–341.

Slutsky, E.E. (1937). Qualche proposizione relativa alla teoria delle funzioni aleatorie. *Giorn. Ist. Ital. Attuari* **8**, 183–199.

Smith, W.L. (1954). Asymptotic renewal theorems. *Proc. Roy. Soc. Edinb.* A **64**, 9–48.

Snell, J.L. (1952). Application of martingale system theorems. *Trans. Amer. Math. Soc.* **73**, 293–312.

Sova, M. (1967). Convergence d'opérations linéaires non bornées. *Rev. Roumaine Math. Pures Appl.* **12**, 373–389.

Sparre-Andersen, E. (1953/54). On the fluctuations of sums of random variables, I–II. *Math. Scand.* **1**, 263–285; **2**, 195–223.

Sparre-Andersen, E., Jessen, B. (1948). Some limit theorems on set-functions. *Danske Vid. Selsk. Mat.-Fys. Medd.* **25**:5 (8 pp.).

Spitzer, F. (1964). Electrostatic capacity, heat flow, and Brownian motion. *Z. Wahrsch. verw. Geb.* **3**, 110–121.

— (1976). *Principles of Random Walk*, 2nd ed. Springer, NY.

Stieltjes, T.J. (1894/95). Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse* **8**, 1-122; **9**, 1–47.

Stone, C.J. (1963). Weak convergence of stochastic processes defined on a semi-infinite time interval. *Proc. Amer. Math. Soc.* **14**, 694–696.

— (1965). A local limit theorem for multidimensional distribution functions. *Ann. Math. Statist.* **36**, 546-551.

— (1968). On a theorem by Dobrushin. *Ann. Math. Statist.* **39**, 1391–1401.

— (1969). On the potential operator for one-dimensional recurrent random walks. *Trans. Amer. Math. Soc.* **136**, 427–445.

Stone, M.H. (1932). Linear transformations in Hilbert space and their applications to analysis. *Amer. Math. Soc. Coll. Publ.* **15**.

Stout, W.F. (1974). *Almost Sure Convergence.* Academic Press, NY.

Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. verw. Geb.* **3**, 211–226.

Stratonovich, R.L. (1966). A new representation for stochastic integrals and equations. *SIAM J. Control* **4**, 362–371.

Stricker, C., Yor, M. (1978). Calcul stochastique dépendant d'un paramètre. *Z. Wahrsch. verw. Geb.* **45**, 109–133.

Stroock, D.W. (1981). The Malliavin calculus and its applications to second order parabolic differential equations, I–II. *Math. Systems Theory* **14**, 25–65, 141–171.

Stroock, D.W., Varadhan, S.R.S. (1969). Diffusion processes with continuous coefficients, I–II. *Comm. Pure Appl. Math.* **22**, 345–400, 479–530.

— (1979). *Multidimensional Diffusion Processes.* Springer, Berlin.

Sucheston, L. (1983). On one-parameter proofs of almost sure convergence of multiparameter processes. *Z. Wahrsch. verw. Geb.* **63**, 43–49.

Takács, L. (1962). *Introduction to the Theory of Queues.* Oxford Univ. Press, NY.

— (1967). *Combinatorial Methods in the Theory of Stochastic Processes.* Wiley, NY.

Tanaka, H. (1963). Note on continuous additive functionals of the 1-dimensional Brownian path. *Z. Wahrsch. verw. Geb.* **1**, 251–257.

Tempel'man, A.A. (1972). Ergodic theorems for general dynamical systems. *Trans. Moscow Math. Soc.* **26**, 94–132.

Thiele, T.N. (1880). On the compensation of quasi-systematic errors by the method of least squares (in Danish). *Vidensk. Selsk. Skr., Naturvid. Mat.* **12** (5), 381–408.

Thorisson, H. (1995). On time and cycle stationarity. *Stoch. Proc. Appl.* **55**, 183–209.

— (1996). Transforming random elements and shifting random fields. *Ann. Probab.* **24**, 2057–2064.

— (2000). *Coupling, Stationarity, and Regeneration.* Springer, NY.

Tonelli, L. (1909). Sull'integrazione per parti. *Rend. Acc. Naz. Lincei* (5) **18**, 246–253.

Tortrat, A. (1969). Sur les mesures aléatoires dans les groupes non abéliens. *Ann. Inst. H. Poincaré, Sec. B*, **5**, 31–47.

Trotter, H.F. (1958a). Approximation of semi-groups of operators. *Pacific J. Math.* **8**, 887–919.

— (1958b). A property of Brownian motion paths. *Illinois J. Math.* **2**, 425–433.

Varadarajan, V.S. (1958). Weak convergence of measures on separable metric spaces. On the convergence of probability distributions. *Sankhyā* **19**, 15–26.

— (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109**, 191–220.

Varadhan, S.R.S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19**, 261–286.

— (1984). *Large Deviations and Applications*. SIAM, Philadelphia.

Ville, J. (1939). *Étude Critique de la Notion du Collectif.* Gauthier-Villars, Paris.

Vitali, G. (1905). Sulle funzioni integrali. *Atti R. Accad. Sci. Torino* **40**, 753–766.

Volkonsky, V.A. (1958). Random time changes in strong Markov processes. *Th. Probab. Appl.* **3**, 310–326.

— (1960). Additive functionals of Markov processes. *Trudy Mosk. Mat. Obshc.* **9**, 143–189.

Wald, A. (1944). On cumulative sums of random variables. *Ann. Math. Statist.* **15**, 283–296.

— (1946). Differentiation under the integral sign in the fundamental identity of sequential analysis. *Ann. Math. Statist.* **17**, 493–497.

— (1947). *Sequential Analysis.* Wiley, NY.

Waldenfels, W.v. (1968). Characteristische Funktionale zufälliger Maße. *Z. Wahrsch. verw. Geb.* **10**, 279–283.

Walsh, J.B. (1978). Excursions and local time. *Astérisque* **52–53**, 159–192.

— (1984). An introduction to stochastic partial differential equations. *Lect. Notes in Math.* **1180**, 265–439.

Wang, A.T. (1977). Generalized Itô's formula and additive functionals of Brownian motion. *Z. Wahrsch. verw. Geb.* **41**, 153–159.

Warner, F.W. (1983). *Foundations of Differential Manifolds and Lie groups.* Springer, Berlin.

Watanabe, H. (1964). Potential operator of a recurrent strong Feller process in the strict sense and boundary value problem. *J. Math. Soc. Japan* **16**, 83–95.

Watanabe, S. (1964). On discontinuous additive functionals and Lévy measures of a Markov process. *Japan. J. Math.* **34**, 53–79.

— (1968). A limit theorem of branching processes and continuous state branching processes. *J. Math. Kyoto Univ.* **8**, 141–167.

— (1987). Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. *Ann. Probab.* **15**, 1–39.

Watson, H.W., Galton, F. (1874). On the probability of the extinction of families. *J. Anthropol. Inst. Gr. Britain, Ireland* **4**, 138–144.

Weil, A. (1936). La mesure invariante dans les espaces de groupes et les espaces homogènes. *Enseinement Math.* **35**, 241.

— (1940). *L'intégration dans les Groupes Topologiques et ses Applications.* Hermann et Cie, Paris.

Whitt, W. (2002). *Stochastic-Process Limits.* Springer, NY.

Whitworth, W.A. (1878). Arrangements of m things of one sort and n things of another sort under certain conditions of priority. *Messenger of Math.* **8**, 105–14.

Wiener, N. (1923). Differential space. *J. Math. Phys.* **2**, 131–174.

— (1938). The homogeneous chaos. *Amer. J. Math.* **60**, 897–936.

— (1939). The ergodic theorem. *Duke Math. J.* **5**, 1–18.

Wiener, N., Wintner, A. (1943). The discrete chaos. *Amer. J. Math.* **65**, 279–298.

Wigner, E.P. (1955). Characteristic vectors of bordered matrices with infinite dimension. *Ann. Math.* **62**, 548–564.

Williams, D. (1991). *Probability with Martingales.* Cambridge Univ. Press.

Yaglom, A.M. (1947). Certain limit theorems of the theory of branching processes. *Dokl. Acad. Nauk SSSR* **56**, 795–798.

Yamada, T. (1973). On a comparison theorem for solutions of stochastic differential equations and its applications. *J. Math. Kyoto Univ.* **13**, 497–512.

Yamada, T., Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11**, 155–167.

Yoeurp, C. (1976). Décompositions des martingales locales et formules exponentielles. *Lect. Notes in Math.* **511**, 432–480. Springer, Berlin.

Yor, M. (1978). Sur la continuité des temps locaux associée à certaines semimartingales. *Astérisque* **52–53**, 23–36.

Yosida, K. (1948). On the differentiability and the representation of one-parameter semigroups of linear operators. *J. Math. Soc. Japan* **1**, 15–21.

— (1949). Brownian motion on the surface of the 3-sphere. *Ann. Math. Statist.* **20**, 292–296.

Yosida, K., Kakutani, S. (1939). Birkhoff's ergodic theorem and the maximal ergodic theorem. *Proc. Imp. Acad.* **15**, 165–168.

Zähle, M. (1980). Ergodic properties of general Palm measures. *Math. Nachr.* **95**, 93–106.

Zaremba, S. (1909). Sur le principe du minimum. *Bull. Acad. Sci. Cracovie.*

Zinn, J. (1986). Comparison of martingale difference sequences. In: *Probability in Banach Spaces. Lect. Notes in Math.* **1153**, 453–457. Springer, Berlin.

Zygmund, A. (1951). An individual ergodic theorem for noncommutative transformations. *Acta Sci. Math. (Szeged)* **14**, 103–110.

Indices

Authors

O. Kallenberg, *Foundations of Modern Probability*, Probability Theory and Stochastic Modelling 99, https://doi.org/10.1007/978-3-030-61871-1

Topics

Symbols