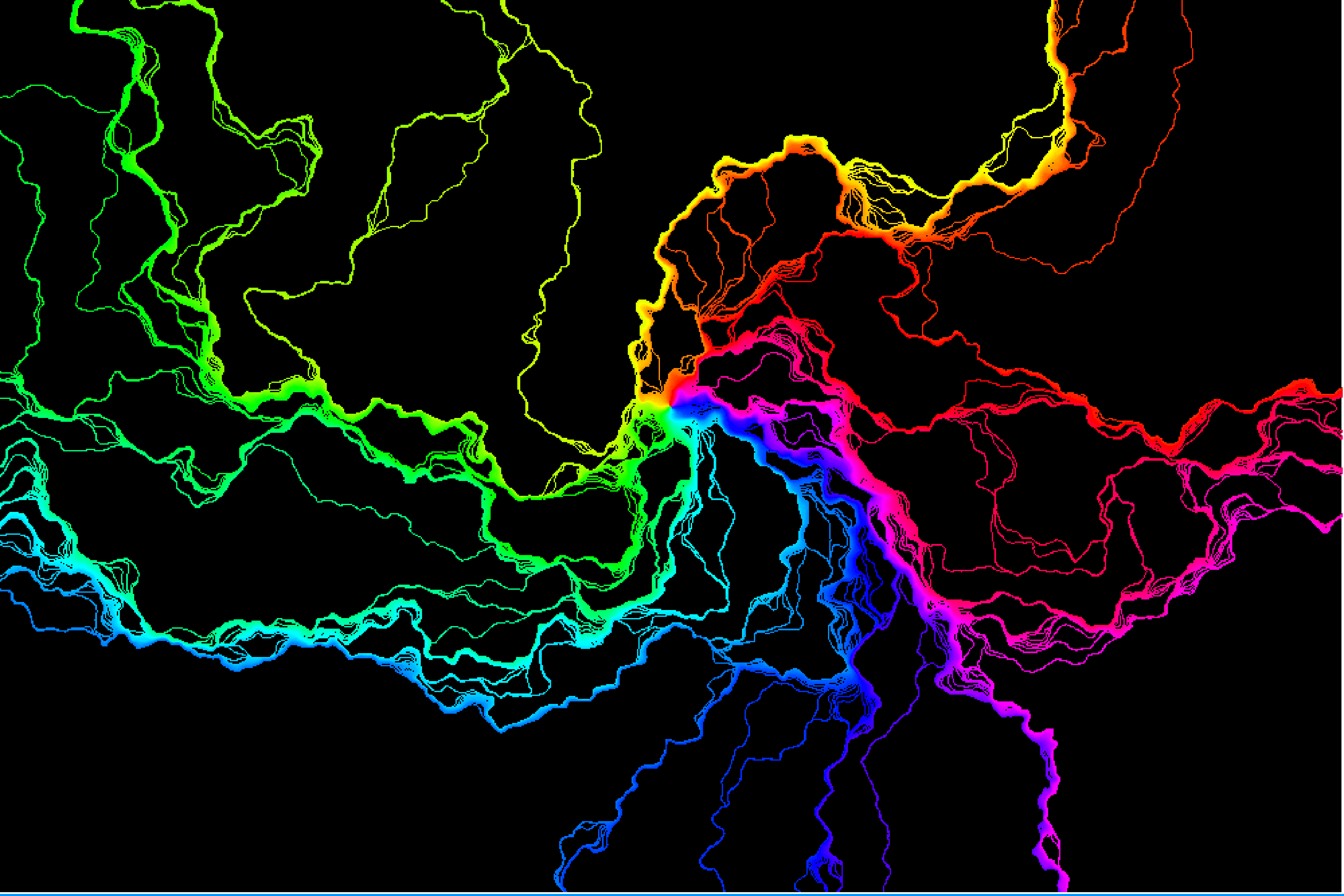


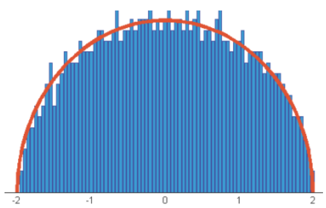
USTC 概率论习题课讲义 (试用版)

2022 秋、2023 春-刘党政班

作者：宗语轩

时间：2023 年 8 月

版本：0.6



Probability Theory is Measure Theory with a Soul. ——Mark Kac

前 言

概率论, 是研究随机现象并揭示随机现象中的结构和规律的数学分支, 同时也为统计理论与方法提供理论基础. 概率论起源于 17 世纪帕斯卡与费马有关机会性游戏的研究, 后经伯努利, 拉普拉斯, 棣莫弗, 高斯与李雅普诺夫等人的发展, 出现了一般形式的大数定律与中心极限定理. 1933 年 Kolmogorov 在其专著《概率论基础》中提出的基于测度论的概率公理化体系, 标志着现代概率论的建立. 之后受数理金融与统计力学发展的驱动, Lévy，Ito，Doob 等人发展了随机过程理论, 形成了如今完善的鞅论与随机分析. 其后概率论及其应用一直蓬勃发展, 同时也一直与其他数学分支及其他学科领域相互交叉渗透. 直至现在, 概率论已成为一门能够 "顶天立地" 的学科.

在 2022 年秋季学期, 笔者有幸担任刘党政老师班级的概率论课程助教, 也是笔者首次担任小班教学课程的助教. 在笔者担任助教期间, 发现不少同学尤其在期中之后的概率论学习出现一定困难. 一方面, 概率论本身内容比较杂, 很多概念初学起来并不容易理解, 以及大部分同学初学概率论都会有 "琐碎" 之感, 尤其体现在处理概率问题中运用的各种技巧和技术上, 因此, 初学概率论时出现瓶颈也是非常正常的现象. 另一方面, 很多同学自始至终对概率空间的理解不够深刻，以及对分析中最常见最基本的 technique 和最基本的测度论知识不熟悉, 从部分同学的作业和考试中也体现出概念使用混乱, 概率语言不会表述或表述不当等问题. 基于笔者担任助教期间的积累, 同时为了让后人更好地适应课程体系, 加深概念理解, 总结及延伸一些重要的 technique 及 idea, 并适当开拓视野, 笔者在原有的习题课讲义基础上编写了此份《USTC 概率论习题课讲义》合集.

本课程对应的教材内容为: Grimmett, Stirzaker: Probability and Random Process, Chapter 1-3, 4.1-4.10, 5.1, 5.6-5.10, 7.1-7.6. 本讲义的内容编排基本参照教材内容的顺序, 但也补充了一部分拓展知识. 另外, 由于科大数院的本科培养方案进行了改革, 至 2022 年春季学期起, 本课程从原有的 4 学分调整至现在的 3 学分, 因此现在的 3 学分课程在原有的 4 学分课程的基础上主要在随机游走、正态分布下的样本均值与样本方差、矩方法与组合计数、用 Linderberg 替换术证明 Linderberg-Feller 中心极限定理、熵、随机矩阵初步等内容以及一些例子上作了删减 (其中一部分内容会放在 1 学分的课程概率论进阶中讲授). 但笔者认为这些删去的大部分内容都是概率和统计甚至和其相关方向中比较重要的内容及 idea, 因此这些删去的内容也补充至本讲义中, 特此说明.

本讲义以 "专题选讲" 的形式为大家呈现. 由于本讲义是习题课讲义, 从定位上可以看作课堂内容的补充、延伸和拓展, 同时也结合各位读者的需求, 笔者在讲义的编排上主要分为 "基本内容"、"进阶内容" 两部分:

- **基础内容**主要侧重于课堂内容的整理或补充, 包括课堂内容的整合加工, 以及对已学过的概念进行适度延伸等, 有助于对课堂内容进一步理解以及知识框架的建立; 除此之外, 还补充了一些基本的工具及从所学内容延伸出的一些方法、technique 或 idea, 这也是笔者认为大家需要了解的部分.
- **进阶内容**主要侧重于课堂内容的拓展. 一部分是一些既与课内相关度较大, 又和其他领域有一定关联的 "趣味" 问题, 这部分所涉及到的知识点和证明的技术使用往往会比较综合, 旨在让读者了解概率论在其他领域上的一些 "渗透" 现象. 还有一部分是从课内内容出发拓展一些实实在在的知识或方法, 不仅对这些内容有助于更深刻的理解, 同时也跟概统的后续课程起到了过渡和衔接的作用.

此外, 在一些章节最后补充了适量的练习题, 大家要做到合理选用. 部分练习题提示与解答附在本讲义第三部分, 以供大家参考.

特别感谢刘党政老师对笔者助教工作的支持和鼓励, 特别感谢班级里的所有同学提供的问题、解法及各种资料, 同时特别感谢本科 19 级数院何家志同学提供的本课程往年笔记, 为本讲义的撰写提供了极大的帮助.

由于笔者水平有限, 加之时间紧迫, 部分内容未免片面或有所疏漏, 笔误亦在所难免, 恳请广大读者批评指正, 感激不尽!

中国科学技术大学 本科 2019 级数学科学学院 宗语轩

2022 年 12 月于合肥

参考书目:

[1] Grimmett, Stirzaker: Probability and Random Process, 2001.
[2] Rick Durrett: Probability: Theory and Examples, fifth edition, 2019.
[3] 李贤平: 《概率论基础》第三版, 高等教育出版社, 2010.
[4] 李贤平, 陈子毅: 《概率论基础学习指导书》, 高等教育出版社, 2011.
[5] 吴昊: 清华大学《概率论 1》课程讲义 (Introduction to Probability Theory).
[6] 苏淳, 冯群强: 《概率论》第三版, 科学出版社, 2020.
[7] E. Kowalski: An introduction to probabilistic number theory, 2021.
[8] Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers (6th ed.), Oxford University Press, 2008.
[9] Noga Alon, Joel H. Spencer: The Probabilistic Method, 4th Edition, 2016.

[0] 欢迎访问个人主页: http://home.ustc.edu.cn/~zyx240014/
闻道有先后, 内容有疏漏. 发现错误欢迎联系我: zyx240014@mail.ustc.edu.cn

目录

第 1 章 基本内容

1.1 组合恒等式

引理 1.1

我们有

(1) $\binom{n}{k} = \binom{n}{n-k}\ (0 \leqslant k \leqslant n)$.

(2) $k\binom{n}{k} = n\binom{n-1}{k-1}$.

(3) $\binom{n}{m} + \binom{n}{m+1} = \binom{n+1}{m+1}$

提示 (3) 即为 $n+1$ 个盒子选出 $m+1$ 个的组合总数, 考虑最后一个是否被选出, 两种情况的组合总数相加即可.

定理 1.1 (二项式定理)

对 $\forall x \in \mathbb{R}$ 及 $n \in \mathbb{N}^*$, 有

$$(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i.$$

提示 归纳. 利用**引理** 1.1 **(3)** , 我们有

$$(1+x)^n = (1+x)\cdot(1+x)^{n-1} = (1+x)\sum_{i=0}^{n-1}\binom{n-1}{i}x^i = \sum_{i=1}^{n-1}\left(\binom{n-1}{i}+\binom{n-1}{i-1}\right)x^i + 1 + x^n = \sum_{i=0}^{n}\binom{n}{i}x^i.$$

引理 1.2

利用上述引理和定理, 我们得到

(1) $\sum_{k=0}^{n}\binom{n}{k} = 2^n$.

(2) $\sum_{\substack{0\leqslant k\leqslant n\\ k\text{为偶数}}}\binom{n}{k} = \sum_{\substack{0\leqslant k\leqslant n\\ k\text{为奇数}}}\binom{n}{k} = 2^{n-1}$.

(3) $\binom{m+n}{k} = \sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}$. 特别地, $\binom{2n}{n} = \sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i} = \sum_{i=0}^{n}\binom{n}{i}^2$.

(4) $\sum_{k=0}^{n} k\binom{n}{k} = n2^{n-1}$.

(5) $\sum_{k=m}^{n}\binom{k}{m} = \binom{n+1}{m+1}$.

(6) $\sum_{k=0}^{n}\binom{m+k}{k} = \binom{m+n+1}{n}$.

提示 (1) (2) 对 $(1+x)^n$ 运用二项式定理, 分别取 $x=1,-1$ 即可. (3) 注意到 $(1+x)^{m+n} = (1+x)^m(1+x)^n$,

分别运用二项式定理, 考察等式两边 x^k 项的系数. (4)$(1+x)^n=\sum_{i=0}^{n}\binom{n}{i}x^i$ 两侧对 x 求导, 并取 $x=1$ 即可. (5) (6) 利用**引理** 1.1 **(3)** 归纳即可.

例题 1.1 设 $n\ge m$. 证明:

$$\sum_{k=0}^{m}\binom{m}{k}\binom{n+k}{m}=\sum_{k=0}^{m}\binom{m}{k}\binom{n}{k}2^k.$$

证明

$$\begin{aligned}\sum_{k=0}^{m}\binom{m}{k}\binom{n+k}{m}&=[x^m](1+x)^n\sum_{k=0}^{m}\binom{m}{k}(1+x)^k=[x^m](1+x)^n(2+x)^m\\&=[x^m](1+x)^n\sum_{k=0}^{m}\binom{m}{k}2^kx^{m-k}\\&=\sum_{k=0}^{m}\binom{m}{k}\binom{n}{k}2^k.\end{aligned}$$

其中 "$[x^m]f(x)$" 表示 $f(x)$ 中项 x^m 的系数.

注 也可以用组合方法 (~~实质是通过带有二项式定理的代数方法来编故事 (bushi~~):

已知 m 个男孩和 n 个女孩, 并满足条件:

(1) 给一些 (数量未知) 男孩吃苹果;

(2) 从所有人中选出 m 个人吃梨;

(3) 吃了梨的男孩必须也吃了苹果.

假设每个苹果 (梨) 不作区分, 每个人至多拿 1 个苹果及 1 个梨. 问一共有多少种符合条件的分法?

一方面, 设 (1) 中给 k 个男孩苹果, 其分法有 $\binom{m}{k}$ 种, 由 (3) 知, 另外没吃苹果的 $m-k$ 个男孩不可能吃梨, 结合 (2), 剩下梨的分法有 $\binom{n+k}{m}$ 种. 对 k 累加得

$$\sum_{k=0}^{m}\binom{m}{k}\binom{n+k}{m}.$$

另一方面, 设 (2) 中给 k 个女孩梨, 则给 $m-k$ 个男孩梨. 分法分别为 $\binom{n}{k}$ 和 $\binom{m}{m-k}=\binom{m}{k}$ 种. 由 (3) 知吃了梨的男孩一定吃了苹果. 剩下 k 个男孩可以吃苹果也可以不吃苹果, 分法共 2^k 种. 对 k 累加得

$$\sum_{k=0}^{m}\binom{m}{k}\binom{n}{k}2^k.$$

1.2 随机事件 (集合) 的运算

方法论:

$$\textbf{事件运算}\longleftrightarrow\textbf{集合运算}$$

对于事件 A,B, 有

$$\text{事件交, 并, 余, 差}\longleftrightarrow A\cap B, A\cup B, A^c(\text{对立事件}), A\backslash B$$

$$\omega\in A\cap B\Longleftrightarrow\omega\in A\text{且}\omega\in B\Longleftrightarrow A,B\text{同时发生}$$

$$\omega \in A \cup B \Longleftrightarrow \omega \in A\text{或}\omega \in B \Longleftrightarrow A\text{发生或}B\text{发生}$$

$$\omega \in A^c \Longleftrightarrow \omega \notin A \Longleftrightarrow A^c\text{发生} \Longleftrightarrow A\text{不发生}$$

$$\omega \in A\backslash B \Longleftrightarrow \omega \in A\text{且}\omega \notin B \Longleftrightarrow A\text{发生但同时}B\text{不发生}$$

设 $\{A_i, i \in I\}$ 为一列事件族, 类比数列上下确界的定义, 我们定义记号

$$\sup_{i\in I} A_i := \bigcup_{i\in I} A_i = \{x : \exists i \in I, x \in A_i\}, \qquad \inf_{i\in I} A_i := \bigcap_{i\in I} A_i = \{x : \forall i \in I, x \in A_i\}.$$

集合运算的三个基本性质: **交换律**, **结合律**, **分配律**. 分配律可表示为:

$$B \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I}(B \cap A_i), \qquad B \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I}(B \cup A_i)$$

定理 1.2 (De.Morgan 法则)

设 $\{A_i, i \in I\}$ 为一列事件族, 则有

(1) $\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c$;

(2) $\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c$.

♡

我们可以将 $\bigcup_{i\in I} A_i$ 转化成事件的无交并 (互不相容), 以利用概率测度的可列可加性进行计算:

命题 1.1

设 $\{A_i, i = 1, 2, \cdots, n\}$ 为一列事件族 $(n \leqslant +\infty)$, 则有

$$\bigcup_{i=1}^{n} A_i = \bigsqcup_{i=1}^{n}\left(A_i \backslash \bigcup_{k=1}^{i-1} A_k\right),$$

其中 $\bigsqcup$ 表示无交并.

♠

下面我们考虑事件序列 $\{A_n\}$ 的极限:

(1) **上限事件**: $\limsup\limits_{n\to+\infty} A_n := \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m = \{\omega : \text{有无穷多个}A_k\text{包含}\omega\}$, 即 $\{A_n\}$ 发生无穷多次;

(2) **下限事件**: $\liminf\limits_{n\to+\infty} A_n := \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m = \{\omega : \text{有限个}A_k\text{不包含}\omega\}$, 即 $\{A_n\}$ 不发生有限次.

注 利用 $\left\{\bigcup_{m=n}^{\infty} A_m\right\}_{n=1}^{\infty}$ 的递减性和 $\left\{\bigcap_{m=n}^{\infty} A_m\right\}_{n=1}^{\infty}$ 的递增性, 我们亦有:

$$\limsup_{n\to+\infty} A_n = \lim_{n\to\infty}\bigcup_{m=n}^{\infty} A_m, \qquad \liminf_{n\to+\infty} A_n = \lim_{n\to\infty}\bigcap_{m=n}^{\infty} A_m.$$

定义 1.1

我们称事件序列 $\{A_n\}$ **收敛**, 若 $\limsup\limits_{n\to+\infty} A_n = \liminf\limits_{n\to+\infty} A_n$. 记收敛的极限为 $A := \lim\limits_{n\to+\infty} A_n$.

♣

接上, 注意到

$$\bigcap_{m=n}^{\infty} A_m \subseteq A_n \subseteq \bigcup_{m=n}^{\infty} A_m,$$

不难验证 $A \in \mathcal{F}$, 且满足概率连续性: $\mathbb{P}(A) = \lim_{n\to+\infty} \mathbb{P}(A_n)$.

例题 1.2 设 $\{f_n(x)\}$ 及 $f(x)$ 是定义在 $\mathbb{R}$ 上的实值函数, 则使 $f_n(x)$ 不收敛于 $f(x)$ 的一切点 x 所形成的集合 D 可表示为

$$D = \bigcup_{k=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{ x : |f_n(x) - f(x)| \geqslant \frac{1}{k} \right\}.$$

引理 1.3

概率测度的一些性质:

(1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;

(2) 若 $A \subset B$, 则 $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \geq \mathbb{P}(A)$;

(3) $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}\left(A_i \backslash \bigcup_{k=1}^{i-1} A_k \right)$, $n \leqslant +\infty$;

(4) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;

(5) 次 (σ) 可加性:

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant \sum_{i=1}^{n} \mathbb{P}(A_i),\ n \leqslant +\infty;$$

(6) **Jordan 公式**:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i \right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{i_1 < \cdots < i_k} \mathbb{P}(A_{i_1} \cdots A_{i_k}).$$

其中 (3) 由**命题** 1.1得到, (5) 可由 (2), (3) 得到, (6) 为作业.

例题 1.3 设事件 $A_1, A_2, \cdots, A_n$ 相互独立, $\mathbb{P}(A_i) = p_i, i = 1, 2, \cdots, n$. 求:

(1) 这 n 个事件中至少发生一个的概率 P_1;

(2) 这 n 个事件中恰好发生一个的概率 P_2.

解 注意到 $\{$这 n 个事件中至少发生一个$\} = A_1 \cup A_2 \cup \cdots \cup A_n$, 故有

$$P_1 = \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = 1 - \mathbb{P}(A_1^c A_2^c \cdots A_n^c) \xlongequal{\text{独立}} 1 - \prod_{i=1}^{n} \mathbb{P}(A_i^c) = 1 - \prod_{i=1}^{n} (1 - p_i).$$

注意到 $\{$这 n 个事件中恰好发生一个$\} = (A_1^c A_2^c \cdots A_n) \sqcup (A_1^c A_2^c \cdots A_{n-1} A_n^c) \sqcup \cdots \sqcup (A_1 A_2^c \cdots A_n^c)$, 故有

$$\begin{aligned} P_2 &= \mathbb{P}\left(\bigsqcup_{i=1}^{n} (A_1^c A_2^c \cdots A_{i-1}^c A_i A_{i+1}^c \cdots A_n^c) \right) \\ &= \sum_{i=1}^{n} \mathbb{P}(A_1^c A_2^c \cdots A_{i-1}^c A_i A_{i+1}^c \cdots A_n^c) \\ &\xlongequal{\text{独立}} \sum_{k=1}^{n} p_k \prod_{i=1, i\neq k}^{n} (1 - p_i). \end{aligned}$$

例题 1.4 (赌徒破产问题) Player 财富为 k, Dealer 财富为 $N - k$, 掷一枚均匀硬币, 出现正面 H 时 Player 赢 1 Dealer 输 1, 否则 Dealer 赢 1 Player 输 1. 双方赌到一方输光时游戏结束. 问存在某个状态游戏结束的概率?

解 假设游戏结束后继续投掷硬币并保持两人财富保持不变. 记
$A_i=\{第(i-1)N+1次至第iN次投掷硬币均出现正面\ \mathrm{H}\}$, $B=\{存在某个状态游戏结束\}$. 则
$B^c=\{游戏无限进行下去\}$, 且 A_i 相互独立. 注意到对 $\forall i\in\mathbb{N}^*$, 有 $A_i\subseteq B$, 故对 $\forall n\in N^*$, 有 $B^c\subseteq\bigcap_{i=1}^{n}A_i^c$. 所以

$$\mathbb{P}(B^c)\leqslant\mathbb{P}\left(\bigcap_{i=1}^{n}A_i^c\right)\xlongequal{独立}\prod_{i=1}^{n}\mathbb{P}(A_i^c)=(1-2^{-N})^n.$$

由 n 的任意性, 令 $n\to+\infty$, 得 $\mathbb{P}(B^c)=0$, 即 $\mathbb{P}(B)=1$. 故存在某个状态游戏结束的概率为 1.

我们已经知道, 重复独立试验中, 小概率事件必然发生. 上例是其中的一个应用.

例题 1.5(Warming's theorem) 已知事件 $A_1,A_2,\cdots,A_n$, 记事件 N_k 表示这 n 个事件中恰好有 k 个发生, 证明:

$$\mathbb{P}(N_k)=\sum_{i=0}^{n-k}(-1)^i\binom{k+i}{k}S_{k+i},\quad S_j=\sum_{i_1<i_2<\cdots<i_j}\mathbb{P}(A_{i_1}A_{i_2}\cdots A_{i_j})$$

解 记 $A_S=\bigcap_{i\in S}A_i$, 则

$$\mathbb{P}(N_k)=\sum_{\substack{S\subseteq\{1,\cdots n\}\\|S|=k}}\mathbb{P}\left(A_S\bigcap_{j\notin S}A_j^c\right)$$

结合 Jordan 公式, 有

$$\begin{aligned}\mathbb{P}\left(A_S\bigcap_{j\notin S}A_j^c\right)&=\mathbb{P}(A_S)-\mathbb{P}\left(A_S\cap\left(\bigcup_{j\notin S}A_j\right)\right)=\mathbb{P}(A_S)-\mathbb{P}\left(\bigcup_{j\notin S}(A_S\cap A_j)\right)\\&=\mathbb{P}(A_S)-\sum_{j\notin S}\mathbb{P}(A_{S\cup\{j\}})+\sum_{\substack{j<k\\j,k\notin S}}\mathbb{P}(A_{S\cup\{j,k\}})-\cdots\end{aligned}$$

再代入上述 $\mathbb{P}(N_k)$ 的表达式即得.

注 Jordan 公式和 Warming's theorem 在一些排列组合下的概率问题中给出了精确的计算公式, 如下例:

例题 1.6(配对问题) n 对夫妇面对面随机占座, 求恰好有 k 对夫妇面对面坐的概率 $P_k^{(n)}$.$(0\leqslant k\leqslant n)$.

解 把 n 位男士和 n 位女士按序标号, 且同对夫妇标号相同. 记 A_i 表示标号 i 的夫妇面对面坐. 先看 $k=0$ 时的情形. 对 $\forall i_1<i_2<\cdots<i_k$, 有

$$\mathbb{P}(A_{i_1}A_{i_2}\cdots A_{i_k})=\frac{(n-k)!}{n!}$$

利用 Jordan 公式, 有

$$\begin{aligned}P_0^{(n)}&=\mathbb{P}\left(\bigcap_{k=1}^{n}A_k^c\right)=1-\left(\bigcup_{k=1}^{n}A_k\right)=1-\sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<\cdots<i_k}\mathbb{P}(A_{i_1}\cdots A_{i_k})\\&=1-\sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\frac{(n-k)!}{n!}\\&=\sum_{k=0}^{n}\frac{(-1)^k}{k!}.\end{aligned}$$

再看 $1\leqslant k\leqslant n$ 时的情形. 可利用 Warming's theorem 直接得到, 或考虑仅标号为第 $i_1,i_2,\cdots i_k$ 的夫妇面

对面坐的概率:

$$\begin{aligned}\mathbb{P}(A_{i_1}\cdots A_{i_k}A_{i_{k+1}}^c\cdots A_{i_n}^c) &= \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}|A_{i_1})\cdots\mathbb{P}(A_{i_k}|A_{i_1}\cdots A_{i_{k-1}})\mathbb{P}(A_{i_{k+1}^c}\cdots A_{i_n}^c|A_{i_1}\cdots A_{i_k})\\ &= \frac{1}{n}\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_0^{(n-k)}\\ &= \frac{(n-k)!}{n!}P_0^{(n-k)}\end{aligned}$$

故有

$$P_k^{(n)} = \sum_{i_1<\cdots<i_k}\mathbb{P}(A_{i_1}\cdots A_{i_k}A_{i_{k+1}}^c\cdots A_{i_n}^c) = \binom{n}{k}\frac{(n-k)!}{n!}P_0^{(n-k)} = \frac{1}{k!}\sum_{i=0}^{n-k}\frac{(-1)^i}{i!}.$$

注 我们发现, $\lim\limits_{n\to+\infty}P_k^{(n)} = \frac{\mathrm{e}^{-1}}{k!}$, 该值即为参数为 1 的泊松分布的随机变量取 k 值的概率 (后续章节会讲到).

练习 1.1 证明 **Bonferroni** 不等式:

$$\mathbb{P}\left(\bigcup_{r=1}^{n}A_r\right) \geqslant \sum_{r=1}^{n}\mathbb{P}(A_r) - \sum_{1\leqslant r<k\leqslant n}\mathbb{P}(A_r\cap A_k).$$

1.3 条件概率的应用

条件概率的应用, 关键在于如何利用已知的" 在给定某些事件下另一些事件发生概率的" 这个条件 (放大说就是利用已知的信息之间的关系) 去搭建桥梁, 进而解决概率问题. 相对常用的工具有:

引理 1.4 (全概率公式)

设 $B_1,\cdots,B_n$ 为 Ω 的一个划分, 且 $\forall i, \mathbb{P}(B_i)>0$, 则

$$\mathbb{P}(A) = \sum_{i=1}^{n}\mathbb{P}(B_i)\mathbb{P}(A\mid B_i)$$

引理 1.5 (贝叶斯公式)

设 $A_1,A_2,\cdots,A_n$ 为 Ω 的一个划分, $\mathbb{P}(A_i)>0, \forall i$, 则当 $\mathbb{P}(B)>0$ 时有:

$$\mathbb{P}(A_i\mid B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B\mid A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i)\mathbb{P}(B\mid A_i)}{\sum_{j=1}^{n}\mathbb{P}(A_j)\mathbb{P}(B\mid A_j)}$$

引理 1.6

设 $\mathbb{P}(A_1A_2\cdots A_{n-1})>0$, 则有

$$\mathbb{P}(A_1A_2\cdots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2)\cdots\mathbb{P}(A_n|A_1A_2\cdots A_{n-1})$$

注 若涉及到未知的含参整变量, 常常利用条件概率得到递推关系, 转化成数列递推方法求解.

例题 1.7 A 与 B 两个小孩从自己装有红与黄两种颜色积木的袋子里各摸出一块, A 摸出红色与黄色的概率分别是 0.8 与 0.2, B 摸出红色与黄色的概率分别是 0.9 与 0.1. 现有一块红色积木, 求其为 A 取出的概率.

解 记事件 X 表示积木是由 A 摸出的, 事件 Y 表示摸出的积木是红色的. 则有

$$\mathbb{P}(X|Y) = \frac{\mathbb{P}(Y|X)\mathbb{P}(X)}{\mathbb{P}(Y|X)\mathbb{P}(X)+\mathbb{P}(Y|X^c)\mathbb{P}(X^c)} = \frac{0.8\cdot0.5}{0.8\cdot0.5+0.9\cdot0.5} = \frac{8}{17}.$$

例题 1.8 设有甲和乙两个罐子, 甲罐中有 m 个红球和 n 个黑球, 乙罐中有 n 个红球和 m 个黑球, 且 $m > n$. 随机选取一个罐子再从中随机抽取一球, 发现为红球, 将其放回后并摇匀. 若再次在该罐中随机抽取一球, 问该球仍为红色的概率是否比 $\frac{1}{2}$ 大?

解 记事件 X 表示球是由甲取出的, 事件 R_1 表示第一次取出的球是红球, 事件 R_2 表示第二次取出的球是红球. 则有

$$\mathbb{P}(R_2|R_1) = \frac{\mathbb{P}(R_1R_2)}{\mathbb{P}(R_1)} = \frac{\mathbb{P}(R_1R_2|X)\mathbb{P}(X) + \mathbb{P}(R_1R_2|X^c)\mathbb{P}(X^c)}{\mathbb{P}(R_1|X)\mathbb{P}(X) + \mathbb{P}(R_1|X^c)\mathbb{P}(X^c)} = \frac{\frac{1}{2}((\frac{m}{m+n})^2 + (\frac{n}{m+n})^2)}{\frac{1}{2}(\frac{m}{m+n} + \frac{n}{m+n})} = \frac{m^2+n^2}{(m+n)^2} > \frac{1}{2}.$$

这里不能取等是因为 $m > n$.

例题 1.9 平面上有 n 个不同的点, 编号分别为 $1, 2, \cdots, n$. 现有一个质点在这些点上做随机游动, 假设每次它在某点上停留片刻之后就会在其余所有点中等概率选择一个并移动到该点上. 设其初始位置为点 1 上, 求它在第一次返回此点之前访问过点 2 的概率.

解 设所求概率对应的事件 C_1, 事件 A_i 表示在初始位置为点 i 下, 第一次返回点 1 之前访问过点 2 (访问过的点包括初始位置), 事件 B_{ij} 表示初始位置为点 i 下首次到 j. 则有

$$\begin{aligned}\mathbb{P}(C_1) &= \mathbb{P}(B_{12})\mathbb{P}(C_1|B_{12}) + \sum_{k=3}^{n}\mathbb{P}(B_{1k})\mathbb{P}(C_1|B_{1k}) = \mathbb{P}(B_{12})\mathbb{P}(A_2) + \sum_{k=3}^{n}\mathbb{P}(B_{1k})\mathbb{P}(A_k)\\ &= \frac{1}{n-1}\mathbb{P}(A_2) + \frac{n-2}{n-1}\mathbb{P}(A_k)\\ &= \frac{1}{n-1} + \frac{n-2}{n-1}\mathbb{P}(A_3).\end{aligned}$$

由对称性知, $\mathbb{P}(A_3) = \mathbb{P}(A_3^c) = \frac{1}{2}$. 所以 $\mathbb{P}(A_1) = \frac{n}{2(n-1)}$.

例题 1.10 甲乙两坛子中各装一只白球和一只黑球, 从两坛中各取出一球交换后放入另一坛中. 记事件 A_n, B_n, C_n 分别表示第 n 次交换后甲坛的白球数是 2, 1, 0, 记 p_n, q_n, r_n 表示其对应的概率, 求 p_n, q_n, r_n 的关系式, 并讨论当 $n \to +\infty$ 时的情形.

解 我们有

$$p_{n+1} = \mathbb{P}(A_{n+1}) = \mathbb{P}(A_{n+1}|A_n)\mathbb{P}(A_n) + \mathbb{P}(A_{n+1}|B_n)\mathbb{P}(B_n) + \mathbb{P}(A_{n+1}|C_n)\mathbb{P}(C_n) = \frac{1}{4}\mathbb{P}(B_n) = \frac{1}{4}q_n$$

类似地, 有

$$\begin{aligned}q_{n+1} &= \mathbb{P}(B_{n+1}) = \mathbb{P}(B_{n+1}|A_n)\mathbb{P}(A_n) + \mathbb{P}(B_{n+1}|B_n)\mathbb{P}(B_n) + \mathbb{P}(B_{n+1}|C_n)\mathbb{P}(C_n)\\ &= \mathbb{P}(A_n) + \frac{1}{2}\mathbb{P}(B_n) + \mathbb{P}(C_n)\\ &= p_n + \frac{1}{2}q_n + r_n.\end{aligned}$$

$$r_{n+1} = \mathbb{P}(A_{n+1}) = \mathbb{P}(C_{n+1}|A_n)\mathbb{P}(A_n) + \mathbb{P}(C_{n+1}|B_n)\mathbb{P}(B_n) + \mathbb{P}(C_{n+1}|C_n)\mathbb{P}(C_n) = \frac{1}{4}\mathbb{P}(B_n) = \frac{1}{4}q_n$$

将上述三式联立, 利用 $q_0 = 1, q_1 = \frac{1}{2}$, 消去 p_n, r_n, 得

$$q_{n+1} - q_n = -\frac{1}{2}(q_n - q_{n-1}) \implies q_n = \frac{2}{3}\left(1 - \left(-\frac{1}{2}\right)^{n+1}\right)$$

进而

$$p_n = r_n = \frac{1}{6}\left(1 - \left(-\frac{1}{2}\right)^{n}\right)$$

令 $n \to +\infty$, 可得 $p := \lim_{n\to+\infty} p_n = \frac{1}{6}, r := \lim_{n\to+\infty} r_n = \frac{1}{6}, q := \lim_{n\to+\infty} q_n = \frac{2}{3}$.

✍ **练习 1.2** 甲、乙两人轮流抛掷一枚均匀的骰子. 甲先掷, 一直到掷出了 1 点, 交给乙掷, 而到乙掷出了 1 点, 再交给甲掷, 并如此一直下去. 求第 n 次抛掷时由甲掷的概率.

✍ **练习 1.3** 100 名乘客登上一架正好有 100 个座位的飞机, 每名乘客对应一个座位. 第一位乘客先随机选择一个座位坐. 第二位乘客如果自己的座位空着就坐自己的座位, 否则就在其他空余的座位中随机选择一个座位坐. 第三位乘客如果自己的座位空着就坐自己的座位, 否则就在其他空余的座位中随机选择一个座位坐. 这个过程一直持续到所有的 100 名乘客都登机为止. 求最后一名乘客坐自己的座位的概率.

1.4 随机变量与分布函数

定义 1.2

$(\Omega, \mathcal{F}, \mathbb{P}), X : \Omega \to \mathbb{R}$, 若 $\forall x \in \mathbb{R}$ 有 $\{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}$, 则称 X 为 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的一个随机变量.

♣

注 一般用 $\{X \le x\}$ 表示 $\{\omega \in \Omega : X(\omega) \le x\}$, 但是我们不能 “忘记” 样本空间.

定义 1.3

设 X 为 $(\Omega, \mathcal{F}, \mathbb{P})$ 上随机变量, 则称函数 $F_X(x) = \mathbb{P}(\{X \le x\})$ 为 X 的(概率) 分布函数.

♣

区分随机变量 & 分布函数: 若已知随机变量, 我们必然能得到其分布函数. 反之存在性也满足 (见后), 但不能唯一确定! **分布函数会 “忘记” 样本空间!** 下面一个简单的例子即可说明:

例题 1.11 掷一枚均匀硬币.$\Omega = \{H, T\}$, 随机变量 X, Y 满足 $X(H) = 1, X(T) = -1, Y(H) = -1, Y(T) = 1$. 我们有

$$F_X(x) = F_Y(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{2}, & -1 \leqslant x < 1, \\ 1, & x \geqslant 1, \end{cases}$$

但是 $X \ne Y$.

回顾一下分布函数的性质:

引理 1.7

设 X 是随机变量, $F(x)$ 为其分布函数, 则有

(1) 单调递增: 若 $x < y$, 则 $F(x) \le F(y)$,

(2) $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to+\infty} F(x) = 1$,

(3) 右连续: $F(x+0) = F(x)$.

♡

事实上, 满足上述三条引理的一元函数称为分布函数 $F(x)$, 可以找到一个概率空间及随机变量 X, 使得 X 的分布函数是 $F(x)$.

$X \sim U(0,1)$((0,1) 上均匀分布), 当分布函数 F 严格递增时, $Y = F^{-1}(X)$ 有分布函数 F.

$$\mathbb{P}(Y \le y) = \mathbb{P}(F^{-1}(X) \le y) = \mathbb{P}(X \le F(y)) \xlongequal{F(\mathbb{R})\subseteq(0,1)} F(y)$$

更一般地, 对分布函数 $F(x)$, 定义

$$F^{-1}(y) = \sup\{x \mid F(x) < y\},\ y \in (0,1)$$

显然 $F^{-1}(y)$ 单调递增, 且我们有如下事实: 设 F 为分布函数, $X \sim U(0,1)$, 则 $Y = F^{-1}(X)$ 的分布函数为 F.

问题: $\mathbb{P}(F^{-1}(X) \le y) \overset{?}{=\!=} \mathbb{P}(X \le F(y))$

提示 只需验证 $F^{-1}(y) > x \iff y > F(x)$.

$\Longrightarrow$: $\exists x_0$, 使得 $x < x_0 < F^{-1}(y)$. 利用 $F^{-1}(y)$ 上确界的定义知,

$$x_0 \in \{x \mid F(x) < y\}.$$

再利用 F 的单调性知, $F(x) \leqslant F(x_0) < y$.

$\Longleftarrow$: 利用 F 的右连续性知, $\exists\delta > 0$, 使得

$$y > F(x+\delta).$$

再利用 F^{-1} 的定义知, $F^{-1}(y) \geqslant x+\delta > x$.

注 此结论表明均匀分布可用来产生其他分布, 在随机模拟中相当重要.

蒙特卡洛模拟: $g : [0,1] \to [0,1]$ 连续函数, 要计算 $I = \int_0^1 g(x)\mathrm{d}x$, 其中 (X,Y) 为 $[0,1]\times[0,1]$ 上均匀分布, $A = \{(x,y) : 0 \le y \le g(x), x \in [0,1]\}$. 向 $[0,1]^2$ 上随机投点 N 次, 落入 A 的次数记为 N_A, 即 $Y \leqslant g(x)$ 时 (X,Y) "成功". 频率稳定性建议:

$$\frac{N_A}{N} \to I = \mathbb{P}(A).$$

例题 1.12 (Buffon 问题) 平面上有间距为 2 平行线, 投长为 1 的针, 试求针与线相交的概率?

解 x 表示针的中点与最近一条平行线的距离, θ 表示针与线夹角, 明显

$$0 \leqslant x \leqslant 1, \quad 0 \leqslant \theta \leqslant \pi, \quad G = \{(\theta,x) : \theta \in [0,\pi], x \in [0,1]\}$$

A 表示针与线相交: $A := \left\{(\theta,x) \in G : x \le \frac{1}{2}\sin\theta\right\}$

$$\mathbb{P}(A) = \frac{|A|}{|G|} = \frac{1}{\pi}\int_0^\pi \int_0^{\frac{1}{2}\sin\theta} 1\mathrm{d}x\mathrm{d}\theta = \frac{1}{\pi}$$

定义 1.4

以连续型为例. 设随机向量 (X,Y) 联合密度 $f(x,y)$, 则 X 的边缘密度:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y,$$

边缘分布:

$$F_X(x) = \lim_{y\to+\infty} F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(x,v)\mathrm{d}v.$$

给定 $X = x(f_X(x) > 0)$ 下 Y 的条件密度:

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}(\text{关于}y\text{构成密度函数})$$

条件分布:

$$F_{Y|X}(y \mid x) = \int_{-\infty}^{y} \frac{f(x,v)}{f_X(x)}\mathrm{d}v$$

♣

离散型下定义类似. 由此可知, 对随机变量 X,Y, 若已知 $F_{X,Y}(x,y)$, 我们可以求得 $F_X(x),F_Y(y)$, $F_{X|Y}(x|y),F_{Y|X}(y|x)$. 反之,

$$F_X(x),F_Y(y)\nRightarrow F_{X,Y}(x,y),\quad F_X(x),F_{Y|X}(y|x)\Rightarrow F_{X,Y}(x,y).$$

注 $\mathbb{P}(A),\mathbb{P}(B)\nRightarrow\mathbb{P}(AB)=\mathbb{P}(B)\mathbb{P}(A|B)$.

例题 1.13 二元函数

$$F(x,y)=\begin{cases}1-\mathrm{e}^{-x}-x\mathrm{e}^{-y}, & 0\leqslant x\leqslant y,\\ 1-\mathrm{e}^{-y}-y\mathrm{e}^{-y}, & 0\leqslant y\leqslant x,\\ 0, & \text{else}\end{cases}$$

是否为某随机向量 (X,Y) 的联合分布函数? 若是, 分别求出 X 和 Y 的分布函数; 若不是, 请说明理由.

解 $F(x,y)$ 连续且 $\lim\limits_{x,y\to-\infty}F(x,y)=0$, $\lim\limits_{x,y\to+\infty}F(x,y)=1$. 而

$$\frac{\partial^2F}{\partial x\partial y}=\begin{cases}\mathrm{e}^{-y} & 0\leqslant x\leqslant y,\\ 0 & 0\leqslant y\leqslant x.\end{cases}$$

$\dfrac{\partial^2F}{\partial x\partial y}\geqslant 0$. 因此 $F(x,y)$ 是一个联合分布, 并有

$$F_X(x)=\lim_{y\to+\infty}F(x,y)=\begin{cases}1-\mathrm{e}^{-x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}\quad F_Y(y)=\lim_{x\to+\infty}F(x,y)=\begin{cases}1-\mathrm{e}^{-y}-y\mathrm{e}^{-y}, & y\geqslant 0,\\ 0, & y<0.\end{cases}$$

✍ **练习 1.4** 定义实数 m 成为分布函数 F 的中位数, 若

$$F(m-0)\leqslant\frac{1}{2}\leqslant F(m).$$

(1) 给出 $[0,1]$ 上均匀分布和二项分布 $B\left(n,\dfrac{1}{2}\right)$ 的中位数;

(2) 证明分布函数至少有一个中位数, 且中位数集合构成一个闭区间.

1.5 离散型及期望方差

Bernoulli 分布

$\mathbb{P}(X=1)=p,\mathbb{P}(X=0)=1-p$, 则

$$\begin{aligned}&\mathbb{E}[X]=p\cdot 1+(1-p)\cdot 0=p,\\ &\mathbb{E}[X^2]=p\cdot 1^2+(1-p)\cdot 0^2=p,\quad\Longrightarrow\ \mathrm{Var}(X)=p(1-p).\end{aligned}$$

二项分布

$X\sim B(n,p), \mathbb{P}(X=k)=\binom{n}{k}p^kq^{n-k}, k=0,1,\cdots,n$ 这里 $p\in(0,1), p+q=1$. X 服从参数为 (n,p) 的二项分布. 其中 $B(1,p)$ 即为 Bernoulli 分布.

$$\begin{aligned}
\mathbb{E}[X]&=\sum_{k=0}^{n}k\binom{n}{k}p^kq^{n-k}=\sum_{k=0}^{n}k\frac{n!}{k!(n-k)!}p^kq^{n-k}=\sum_{k=0}^{n-1}\frac{n!}{k!(n-1-k)!}p^{k+1}q^{n-1-k}\\
&=np\sum_{k=0}^{n-1}\binom{n-1}{k}p^kq^{n-1-k}=np,\\
\mathbb{E}[X(X-1)]&=\sum_{k=0}^{n}k(k-1)\binom{n}{k}p^kq^{n-k}=n(n-1)p^2,\\
\mathbb{E}[X^2]&=np(np+q), \mathrm{Var}(X)=npq=np(1-p).
\end{aligned}$$

注

1. 二项分布具有**可加性**: 设 $X\sim B(M,p), Y\sim B(N,p)$ 且 X,Y 相互独立，则有 $X+Y\sim B(M+N,p)$. (对任意有限个均成立).

$$\begin{aligned}
\mathbb{P}(X+Y=n)&=\sum_{k=0}^{n}\mathbb{P}(X=k,Y=n-k)=\sum_{k=0}^{n}\binom{M}{k}p^kq^{M-k}\binom{N}{n-k}p^{n-k}q^{N-n+k}\\
&=p^nq^{M+N-n}\sum_{k=0}^{n}\binom{M}{k}\binom{N}{n-k}\\
&=\binom{M+N}{n}p^nq^{M+N-n}.
\end{aligned}$$

2. 二项分布**可分解**: $X\sim B(n,p)$ 可以分解成 n 个相互独立且参数为 p 的 Bernoulli 分布之和, 即

$$X=X_1+X_2+\cdots+X_n,\qquad X_r \text{ i.i.d } \sim B(1,p),\ r=1,\cdots,n.$$

因此, 我们有

$$\begin{aligned}
\mathbb{E}[X]&=\mathbb{E}[X_1+X_2+\cdots+X_n]=\sum_{k=1}^{n}\mathbb{E}[X_k]=n\mathbb{E}[X_1]=np,\\
\mathrm{Var}(X)&=\mathrm{Var}(X_1+X_2+\cdots+X_n)\xlongequal{\text{独立}}\sum_{k=1}^{n}\mathrm{Var}(X_k)=n\mathrm{Var}(X_1)=np(1-p).
\end{aligned}$$

上例可知, 期望的线性性非常重要. 对于随机变量 X,Y, 有时 $X+Y$ 的分布较为复杂, 而 X 和 Y 的分布相对好入手. 转化成如下形式即可:

$$\int_{-\infty}^{+\infty}t\mathrm{d}F_{X+Y}(t)=\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]=\int_{-\infty}^{+\infty}t\mathrm{d}F_X(t)+\int_{-\infty}^{+\infty}t\mathrm{d}F_Y(t)$$

同时注意到, 期望只与分布有关. 若 X,Y 同分布, 则两者期望相同.

几何分布

$$
\begin{aligned}
\mathbb{P}(X=k)&=pq^{k-1}, k=1,2,\cdots,p\in(0,1),p+q=1\\
\mathbb{E}[X]&=\sum_{k=1}^{\infty}kpq^{k-1}\xlongequal{\text{绝对收敛}}p\left(\sum_{k=1}^{\infty}q^k\right)'=p\left(\frac{q}{1-q}\right)'=\frac{p}{(1-q)^2}=\frac{1}{p},\\
\mathbb{E}[X(X+1)]&=\sum_{k=1}^{\infty}k(k+1)pq^{k-1}\xlongequal{\text{绝对收敛}}p\left(\sum_{k=1}^{\infty}q^{k+1}\right)''=p\left(\frac{q^2}{1-q}\right)''=\frac{2p}{(1-q)^3}=\frac{2}{p^2},\\
\mathrm{Var}(X)&=\frac{1-p}{p^2}.
\end{aligned}
$$

注 离散情况下, 几何分布 $\Longleftrightarrow$ 无记忆性.

$\Rightarrow$: 无记忆性:

$$
\mathbb{P}(X=m+k\mid X>m)=\frac{\mathbb{P}(X=m+k)}{\mathbb{P}(X>m)}=\frac{pq^{m+k-1}}{q^m}=pq^{k-1}=\mathbb{P}(X=k).
$$

$\Leftarrow$: 记 $p=\mathbb{P}(X=k+1\mid X>k)$ 与 k 无关, 令 $r_k=\mathbb{P}(X>k)(r_0=1)$, 则

$$
p=\frac{\mathbb{P}(X=k+1)}{\mathbb{P}(X>k)}=\frac{r_k-r_{k+1}}{r_k}.
$$

解得 $r_k=(1-p)^k$. 因此 $\mathbb{P}(X=k)=r_{k-1}-r_k=p(1-p)^{k-1}=pq^{k-1}$.

泊松分布

$$
\begin{aligned}
X\sim P(\lambda),\mathbb{P}(X=k)&=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots,\lambda>0.\\
\mathbb{E}[X]&=\sum_{k=1}^{+\infty}k\mathbb{P}(X=k)=\lambda\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}=\lambda,\\
\mathbb{E}[X(X-1)]&=\sum_{k=1}^{+\infty}k(k-1)\mathbb{P}(X=k)=\lambda^2\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}=\lambda^2\qquad\Longrightarrow\ \mathrm{Var}(X)=\lambda.
\end{aligned}
$$

注

1. 泊松分布具有**可加性**: 设 $X\sim P(\lambda_1),Y\sim P(\lambda_2)$ 且 X,Y 相互独立，则有 $X+Y\sim P(\lambda_1+\lambda_2)$. (对任意有限个均成立).

$$
\begin{aligned}
\mathbb{P}(X+Y=n)&=\sum_{k=0}^{n}\mathbb{P}(X=k,Y=n-k)=\sum_{k=0}^{n}\frac{(\lambda_1)^k}{k!}e^{-\lambda_1}\cdot\frac{(\lambda_2)^{n-k}}{(n-k)!}e^{-\lambda_2}\\
&=\frac{e^{-\lambda_1-\lambda_2}}{n!}\sum_{k=0}^{n}\binom{n}{k}(\lambda_1)^k(\lambda_2)^{n-k}\\
&=\frac{e^{-\lambda_1-\lambda_2}}{n!}(\lambda_1+\lambda_2)^n.
\end{aligned}
$$

2. 泊松分布亦可分解, 泊松分布在随机选择下具有不变性. 如下例:

泊松翻转: 掷硬币 N 次, $N\sim P(\lambda)$,$\mathbb{P}(H)=p$, 记 X,Y 为 H,T 出现次数. 则 X 与 Y 独立, 因为

$$
\begin{aligned}
\mathbb{P}(X=x,Y=y)&=\mathbb{P}(X=x,Y=y,N=x+y)\\
&=\mathbb{P}(X=x,Y=y\mid N=x+y)\mathbb{P}(N=x+y)=\binom{x+y}{x}p^xq^y\frac{\lambda^{x+y}}{(x+y)!}e^{-\lambda}\\
&=\frac{(\lambda p)^x}{x!}e^{-\lambda p}\cdot\frac{(\lambda q)^y}{y!}e^{-\lambda q}.
\end{aligned}
$$

因此

$$\mathbb{P}(X=x)=\sum_{y=0}^{\infty}\mathbb{P}(X=x,Y=y)=\frac{(\lambda p)^x}{x!}e^{-\lambda p},\quad \mathbb{P}(Y=y)=\frac{(\lambda q)^y}{y!}e^{-\lambda q}.$$

X 与 Y 独立, $N=X+Y$ 且有 $X\sim P(\lambda p), Y\sim P(\lambda q)$.

例题 1.14 已知随机变量 X,Y, 且 X,Y 相互独立. 求如下两情形的条件概率 $\mathbb{P}(X=k|X+Y=j)$:

(1) $X\sim B(m,p), Y\sim B(n,p)$;

(2) $X\sim P(\lambda_1), Y\sim P(\lambda_2)$.

解 (1) 利用 $X+Y\sim B(m+n,p)$, 我们有

$$\begin{aligned}\mathbb{P}(X=k|X+Y=j)&=\frac{\mathbb{P}(X=k,X+Y=j)}{\mathbb{P}(X+Y=j)}=\frac{\mathbb{P}(X=k)\mathbb{P}(Y=j-k)}{\mathbb{P}(X+Y=j)}\\&=\frac{\binom{m}{k}p^k(1-p)^{m-k}\binom{n}{j-k}p^{j-k}(1-p)^{n-j+k}}{\binom{m+n}{j}p^j(1-p)^{m+n-j}}\\&=\frac{\binom{m}{k}\binom{n}{j-k}}{\binom{m+n}{j}}.\end{aligned}$$

(2) 利用 $X+Y\sim P(\lambda_1+\lambda_2,p)$, 我们有

$$\begin{aligned}\mathbb{P}(X=k|X+Y=j)&=\frac{\mathbb{P}(X=k,X+Y=j)}{\mathbb{P}(X+Y=j)}=\frac{\mathbb{P}(X=k)\mathbb{P}(Y=j-k)}{\mathbb{P}(X+Y=j)}\\&=\frac{\frac{\lambda_1^k}{k!}\mathrm{e}^{-\lambda_1}\frac{\lambda_2^{j-k}}{(j-k)!}\mathrm{e}^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^j}{j!}\mathrm{e}^{-\lambda_1+\lambda_2}}\\&=\binom{j}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{j-k}.\end{aligned}$$

思考: 求上例两情形的条件期望 $\mathbb{E}[X|X+Y]$.

✍ **练习 1.5** 独立重复伯努利试验中, p 为每次试验成功的概率, S_n 表示第 n 次试验成功时试验总次数. 求:

(1) 条件概率 $\mathbb{P}(S_{n+1}=k|S_n=j)$;

(2) 条件概率 $\mathbb{P}(S_n=k|S_{n+1}=j)$.

✍ **练习 1.6** 有足够多套同类型的卡片组, 每套卡片组共 n 张各不相同的卡片, 每花 1 张券就可以在完整的一套卡片组中随机抽取 1 张卡片. 某人想集齐一套完整的卡片组, 设他恰好集齐卡片组时花费的总券数为 X_n, 求 $\lim\limits_{n\to+\infty}\frac{\mathbb{E}[X_n]}{n\ln n}$.

✍ **练习 1.7 Grimmett 3.5.4, 3.11.15.**

✍ **练习 1.8** 给定 $b>a>0$, 离散随机变量 X 取值于区间 $[a,b]$.

(1) 证明: $\mathrm{Var}(X)\leqslant\frac{(b-a)^2}{4}$.

(2) 求 $\mathbb{E}[X]\mathbb{E}[\frac{1}{X}]$ 的取值范围.

1.6 示性函数与概率论

示性函数: $I_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A.\end{cases}$, 是一种随机变量. 例子: $B(1,p)$. 先从集合的观点看起:

引理 1.8

对集合 $A, B, A_j (j\in J)$, 有

(1) $I_A^k=I_A (k\in N^*)$.

(2) $I_{A\cap B}=I_AI_B$, $I_{A^c}=1-I_A$, $I_{A\cup B}=I_A+I_B-I_AI_B$, $I_{A\backslash B}=I_A(1-I_B)$.

(3) $I_{A\Delta B}=|I_A-I_B|$.

(4) $I_{\bigcup_j A_j}=1-\prod\limits_{j\in J}(1-I_{A_j})$.

提示 (4) 利用 De.Morgan 法则: $\bigcup\limits_{i\in I}A_i=\left(\bigcap\limits_{i\in I}A_i^c\right)^c$.

转化方式:

(1) 对事件 A, $\mathbb{P}(A)=\mathbb{E}[I_A]$, 且有 $\mathbb{E}[I_A]=\mathbb{E}[I_A^k](k\in N^*)$, 算方差时取 $k=2$.

(2) 分解成其他的随机变量之和. 例如: 对离散型随机变量 X, 有 $X=\sum\limits_x xI_{\{X=x\}}$ 及

$$\mathbb{E}[X]=\sum_x x\mathbb{P}(X=x)=\sum_x x\mathbb{E}[I_{\{X=x\}}]=\mathbb{E}[\sum_x xI_{\{X=x\}}].$$

也可以分解成若干示性函数的和. 由此可用来计算方差或其他 k 阶矩 (参考作业题 **3.3.3(a)**).

动机: 利用期望的性质, 尤其是线性性.

引理 1.9

对非负整值随机变量 X, 有

$$\mathbb{E}[X]=\sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

证明 利用 Fubini 定理, 我们有

$$\mathbb{E}[X]=\mathbb{E}[\sum_{n=0}^{X-1}1]=\mathbb{E}[\sum_{n=0}^{\infty}I_{\{X>n\}}]\xlongequal{\text{Fubini}}\sum_{n=0}^{\infty}\mathbb{E}[I_{\{X>n\}}]=\sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

注 求分布列的常见转化方法:

$$\mathbb{P}(X=n)=\mathbb{P}(X\geqslant n)-\mathbb{P}(X\geqslant n+1)=\mathbb{P}(X\leqslant n)-\mathbb{P}(X\leqslant n-1)$$

$$\begin{aligned}\mathbb{P}(X=m,Y=n)&=\mathbb{P}(X\geqslant m,Y\geqslant n)-\mathbb{P}(X\geqslant m+1,Y\geqslant n)-\mathbb{P}(X\geqslant m,Y\geqslant n+1)\\&\quad+\mathbb{P}(X\geqslant m+1,Y\geqslant n+1)\\&=\mathbb{P}(X\geqslant m,Y\leqslant n)-\mathbb{P}(X\geqslant m+1,Y\leqslant n)-\mathbb{P}(X\geqslant m,Y\leqslant n-1)\\&\quad+\mathbb{P}(X\geqslant m+1,Y\leqslant n-1).\end{aligned}$$

例题 1.15 利用引理 1.8 **(4)**, 证明**Jordan 公式**:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right)=\sum_{k=1}^n(-1)^{k-1}\sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\mathbb{P}(A_{i_1}\cdots A_{i_k}).$$

证明 利用

$$I_{\bigcup_j A_j} = 1 - \prod_{j\in J}(1 - I_{A_j}) = \sum_{k=1}^{n}(-1)^{k-1}\sum_{1\leqslant i_1<\cdots<i_k\leqslant n} I_{A_{i_1}}\cdots I_{A_{i_k}} = \sum_{k=1}^{n}(-1)^{k-1}\sum_{1\leqslant i_1<\cdots<i_k\leqslant n} I_{A_{i_1}\cap\cdots\cap A_{i_k}},$$

两边同时取期望并利用期望的线性性即得.

例题 1.16 一座楼共 n 层, 有一个载有 m 人的升降电梯在这 n 层楼里停靠若干次, 每个人随机选择一层楼停靠. 求电梯平均停靠的总次数.

解 记 T 为电梯停靠的次数, 直接求 T 的分布非常困难, 但我们可以将其分解成 n 个示性函数之和:

$$T = I_{A_1} + \cdots + I_{A_n},$$

其中事件 A_i 表示第 i 层楼有人停. 故有

$$\mathbb{E}[T] = \sum_{i=1}^{n}\mathbb{E}[I_{A_i}] = \sum_{i=1}^{n}\mathbb{P}(A_i) = \sum_{i=1}^{n}\left(1-\left(\frac{n-1}{n}\right)^m\right) = n\left(1-\left(1-\frac{1}{n}\right)^m\right).$$

练习 1.9 **Grimmett 3.11.36.**

练习 1.10 用示性函数的方法证明 **Waring 公式**(详见 **Grimmett 1.8.13**).

1.7 条件期望

定义 1.5

设

$$\psi(x) := \mathbb{E}[Y|X=x] = \sum_y y f_{Y|X}(y|x).$$

称 $\psi(X)$ 为 Y 关于 X 的条件期望, 记 $\mathbb{E}[Y|X]$.

注 $\mathbb{E}[Y|X]$ 为一个与 X 有关的**随机变量**. 由定义知, 求条件期望 $\mathbb{E}[Y|X]$ 时往往先求 $\mathbb{E}[Y|X=x]$ 的表达式, 再用 X 代替表达式里的 x, 得到的新表达式即为 $\mathbb{E}[Y|X]$.

定理 1.3

设 X, Y 为离散型随机变量, 则有

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

例题 1.17 (多项分布) $X = (X_1,\cdots,X_r)$. $\mathbb{P}(X_1 = k_1,\cdots,X_r = k_r) = \dfrac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r}$, $\sum_{i=1}^{r} k_i = n$, $\sum_{i=1}^{r} p_i = 1$ 且 $p_i > 0$. 现对 $i \neq j$,

(1) 求 $\mathrm{Cov}(X_i, X_j)$, $\rho(X_i, X_j)$.

(2) 求 $\mathbb{E}[X_j|X_i > 0]$.

解 (1) 由题意得, $X_i \sim B(n, p_i)$, $X_i + X_j \sim B(n, p_i + p_j)$. 故有

$$\mathbb{E}[X_i] = np_i,\quad \mathbb{E}[X_i + X_j] = n(p_i+p_j),\quad \mathrm{Var}(X_i) = np_i(1-p_i),\quad \mathrm{Var}(X_i+X_j) = n(p_i+p_j)(1-p_i-p_j).$$

再利用 $\mathrm{Var}(X_i + X_j) = \mathrm{Var}(X_i) + \mathrm{Var}(X_j) + 2\mathrm{Cov}(X_i, X_j)$, 得

$$\mathrm{Cov}(X_i, X_j) = -np_ip_j,\quad \rho(X_i, X_j) = \frac{-np_i\cdot np_j}{\sqrt{np_i(1-p_i)np_j(1-p_j)}} = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}.$$

(2) 利用条件期望的性质, 我们有

$$\mathbb{E}[X_j] = \mathbb{E}[\mathbb{E}[X_j|X_i]] = \mathbb{E}[X_j|X_i > 0]\mathbb{P}(X_i > 0) + \mathbb{E}[X_j|X_i = 0]\mathbb{P}(X_i = 0).$$

而

$$\begin{aligned}\mathbb{P}(X_j = k|X_i = 0) &= \frac{\mathbb{P}(X_j = k, X_i = 0)}{\mathbb{P}(X_i = 0)} = \frac{n!}{k!0!(n-k)!}p_j^k p_i^0(1-p_i-p_j)^{n-k} \cdot \frac{1}{(1-p_i)^n} \\ &= \binom{n}{k}\left(\frac{p_j}{1-p_i}\right)^k\left(1-\frac{p_j}{1-p_i}\right)^{n-k}.\end{aligned}$$

因此,

$$\mathbb{E}[X_j|X_i = 0] = \frac{np_j}{1-p_i}.$$

利用 $\mathbb{E}[X_j] = np_j, \mathbb{P}(X_i = 0) = (1-p_i)^n, \mathbb{P}(X_i > 0) = 1-(1-p_i)^n$, 有

$$\mathbb{E}[X_j|X_i > 0] = \frac{np_j(1-(1-p_i)^{n-1})}{1-(1-p_i)^n}.$$

注 补充协方差中两个很有用的结论:

1. $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i, X_j).$
2. 双线性性: $\mathrm{Cov}(aX+bY, Z) = a\mathrm{Cov}(X, Z)+b\mathrm{Cov}(Y, Z), \mathrm{Cov}(X, aY+bZ) = a\mathrm{Cov}(X, Y)+b\mathrm{Cov}(X, Z).$

我们可以从 “投影” 的角度理解条件期望, 下面事实说明 $\mathbb{E}[Y|X] = \arg\min_{g(x)}\mathbb{E}[(Y-g(x))^2]$. 换言之, 假设 “$X$” 是你已知的信息, 此时 “$Y$” 可以做最好的估计:

例题 1.18 设 (X, Y) 是联合离散随机向量, X 与 Y 的二阶矩存在, 记 $\psi(X) := \mathbb{E}[Y|X]$. 若 g 为可测函数且 $g(X)$ 的二阶矩存在, 证明:

$$\mathbb{E}[(Y-\psi(X))^2] \leqslant \mathbb{E}[(Y-g(X))^2]$$

解 我们有

$$\begin{aligned}\mathbb{E}[(Y-g(X))^2] &= \mathbb{E}[(Y-\psi(X)+\psi(X)-g(X))^2] \\ &= \mathbb{E}[(Y-\psi(X))^2] + 2\mathbb{E}[(Y-\psi(X))(\psi(X)-g(X))] + \mathbb{E}[(\psi(X)-g(X))^2].\end{aligned}$$

利用 $\mathbb{E}[(\psi(X)-g(X))^2] \geqslant 0$ 以及

$$\begin{aligned}\mathbb{E}[(Y-\psi(X))(\psi(X)-g(X))] &= \mathbb{E}[\mathbb{E}[(Y-\psi(X))(\psi(X)-g(X))|X]] \\ &= \mathbb{E}[(\psi(X)-g(X))\mathbb{E}[(Y-\psi(X))|X]] \\ &= \mathbb{E}[(\psi(X)-g(X))(\mathbb{E}[Y|X]-\psi(X))] \\ &= 0.\end{aligned}$$

注 类似地, 我们亦可以从 “投影” 的角度理解方差: $\mathrm{Var}(X) = \min_{a\in\mathbb{R}}\mathbb{E}[(X-a)^2]$. 因为

$$\mathbb{E}[(X-a)^2] = a^2 - 2a\mathbb{E}[X] + \mathbb{E}[X^2] = (a-\mathbb{E}[X])^2 + \mathrm{Var}(X) \geqslant \mathrm{Var}(X),$$

等号当且仅当 $a = \mathbb{E}[X]$ 时取等, 此时即为 $\mathrm{Var}(X)$ 的形式.

例题 1.19 独立重复地掷一个均匀骰子, 按序记录投掷的点数. 记 τ_{ij} 表示首次出现 “ij” 时骰子已投掷的总次数, 其中 $1 \leqslant i, j \leqslant 6$. 求 $\mathbb{E}[\tau_{11}], \mathbb{E}[\tau_{12}]$.

解 记 N_1, N_2 分别表示第 1, 2 次骰子投掷的点数, τ_i 表示首次出现 "i" 时骰子已投掷的总次数, 则有

$$\begin{aligned}
\mathbb{E}[\tau_{11}] &= \mathbb{E}[\mathbb{E}[\tau_{11}|N_1]] = \mathbb{E}[\tau_{11}|N_1 = 1]\mathbb{P}(N_1 = 1) + \mathbb{E}[\tau_{11}|N_1 \neq 1]\mathbb{P}(N_1 \neq 1) \\
&= \mathbb{E}[\mathbb{E}[\tau_{11}|N_1 = 1, N_2]] \cdot \frac{1}{6} + (1 + \mathbb{E}[\tau_{11}]) \cdot \frac{5}{6} \\
&= (\mathbb{E}[\tau_{11}|N_1 = 1, N_2 = 1]\mathbb{P}(N_2 = 1) + \mathbb{E}[\tau_{11}|N_1 = 1, N_2 \neq 1]\mathbb{P}(N_2 \neq 1)) \cdot \frac{1}{6} + (1 + \mathbb{E}[\tau_{11}]) \cdot \frac{5}{6} \\
&= \left(2 \cdot \frac{1}{6} + (2 + \mathbb{E}[\tau_{11}]) \cdot \frac{5}{6}\right) \cdot \frac{1}{6} + (1 + \mathbb{E}[\tau_{11}]) \cdot \frac{5}{6} \\
&= \frac{35}{36}\mathbb{E}[\tau_{11}] + \frac{7}{6}
\end{aligned}$$

因此

$$\mathbb{E}[\tau_{11}] = 42.$$

类似地, 有

$$\begin{aligned}
\mathbb{E}[\tau_{12}] &= \mathbb{E}[\mathbb{E}[\tau_{12}|N_1]] = \mathbb{E}[\tau_{12}|N_1 = 1]\mathbb{P}(N_1 = 1) + \mathbb{E}[\tau_{12}|N_1 \neq 1]\mathbb{P}(N_1 \neq 1) \\
&= \mathbb{E}[\tau_{12}|N_1 = 1] \cdot \frac{1}{6} + (1 + \mathbb{E}[\tau_{12}]) \cdot \frac{5}{6}.
\end{aligned}$$

同时, 我们有

$$\begin{aligned}
\mathbb{E}[\tau_{12}|N_1 = 1] &= \mathbb{E}[\mathbb{E}[\tau_{12}|N_1 = 1, N_2]] \\
&= \mathbb{E}[\tau_{12}|N_1 = 1, N_2 = 1]\mathbb{P}(N_2 = 1) + \mathbb{E}[\tau_{12}|N_1 = 1, N_2 = 2]\mathbb{P}(N_2 = 2) \\
&+ \mathbb{E}[\tau_{12}|N_1 = 1, N_2 \neq 1, 2]\mathbb{P}(N_2 \neq 1, 2) \\
&= (1 + \mathbb{E}[\tau_{12}|N_1 = 1]) \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + (2 + \mathbb{E}[\tau_{12}]) \cdot \frac{4}{6} \\
&= \frac{1}{6}\mathbb{E}[\tau_{12}|N_1 = 1] + \frac{2}{3}\mathbb{E}[\tau_{12}] + \frac{11}{6}.
\end{aligned}$$

将上述两式联立可得, $\mathbb{E}[\tau_{12}] = 36$.

练习 1.11 小朋友 A 有 N 块积木, N 服从参数为 λ 的泊松分布, 每块积木由小朋友 B 独立地以 $\frac{1}{2}$ 的概率拿走. 记小朋友 B 拿走的积木块数为 K, 求 $\mathbb{E}[K], \mathbb{E}[N|K]$.

练习 1.12 一个瓮包括了 b 个蓝球和 r 个红球.

(1) 在瓮中随机取球, 直到第一个蓝球被取出后停止. 证明取出的总球数的期望值是 $\frac{b+r+1}{b+1}$.

(2) 在瓮中随机取球, 直到某一种颜色的球都被取出后停止. 求瓮中取出的总球数的期望值.

练习 1.13 证明:

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X])$$

注 这里定义条件方差:

$$\mathrm{Var}(Y|X) = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2.$$

1.8 概率母函数

定义 1.6

非负整数值随机变量 X 的(概率) 母函数: $G_X(s) := \mathbb{E}[s^X] = \sum_{i=0}^{\infty} s^i \mathbb{P}(X=i)$.

♣

定义 1.7

(X,Y) 非负整数值向量, 联合概率母函数

$$G(s,t) := \mathbb{E}[s^X t^Y] = \sum_{i,j=0}^{\infty} s^i t^j \mathbb{P}(X=i, Y=j).$$

♣

$G(s)$ 为概率母函数的判定条件: $G(s)$ 系数非负, $G(1)=1$.

概率母函数的优点是它简单, 在离散型随机过程中会用到相关结果. 后续的一一对应说明, 对这类随机变量概率分布的研究可转化为母函数的研究, 加之母函数是一类幂级数, 有许多好的性质以便于处理, 尤其用于求概率分布的数字特征 (矩) 上. 缺点一方面是其局限性大 (仅适用于非负整数值随机变量), 另一方面是不能用其完全确定分布函数. 以上问题会通过引入特征函数来解决, 以后上课会讲.

概率母函数的**性质**:

(1) 唯一性: 对非负整值随机变量 X, 母函数 $\Longleftrightarrow$ 分布列.

$\Longleftarrow$: 定义;

$\Longrightarrow$: $G_X(s)$ 是在 $(-R,R)$ 内任意阶可微的幂级数 $(R \geq 1)$, 并有

$$\mathbb{P}(X=i) = \frac{G^{(i)}(0)}{i!}.$$

(2) 算矩: $\mathbb{E}[X] = G'(1)$, $\mathbb{E}[X(X-1)\cdots(X-k+1)] = G^{(k)}(1)$, $\mathrm{Var}(X) = G''(1) + G'(1)(1-G'(1))$.

(3) 卷积: 设 $X_1, \cdots, X_n$ 相互独立, $G_k(s) = \mathbb{E}[s^{X_k}]$. 则 $Y = \sum_{k=1}^{n} X_k$ 母函数为

$$G_Y(s) = G_1(s)\cdots G_n(s).$$

(4) 套娃: 设 $\{X_k : k \geq 1\}$i.i.d, 母函数为 $G_X(s)$, 随机变量 N 与 $\{X_k : k \geq 1\}$ 独立, 其母函数为 $G_N(s)$. 则 $Y = \sum_{k=1}^{N} X_k$ 有母函数

$$G_Y(s) = G_N\left(G_X(s)\right).$$

(5) 独立: X 与 Y 独立 $\Leftrightarrow G(s,t) = G_X(s)G_Y(t)$.

例题 1.20 计算以下常见离散型的母函数及矩:

(1) 二项分布: $X \sim B(n,p)$, $G_X(s) = \sum_{i=0}^{n} \mathrm{C}_n^i p^i q^{n-i} s^i = (ps+q)^n$. 利用

$$G_X'(1) = np(ps+q)\Big|_{s=1} = np, \qquad G_X''(1) = n(n-1)p^2(ps+q)\Big|_{s=1} = n(n-1)p^2$$

可求得

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}(X) = np(1-p).$$

(2) **几何分布**: X 服从参数为 p 的几何分布,$G_X(s)=\sum_{i=1}^{\infty}pq^{i-1}s^i=\frac{ps}{1-qs}$. 利用

$$G'_X(1)=\frac{p}{(1-qs)^2}\Big|_{s=1}=\frac{1}{p},\qquad G''_X(1)=\frac{2pq}{(1-qs)^3}\Big|_{s=1}=\frac{2(1-p)}{p^2}$$

可求得

$$\mathbb{E}[X]=\frac{1}{p},\qquad \mathrm{Var}(X)=\frac{1-p}{p^2}.$$

(3) **泊松分布**: $X\sim P(\lambda)$, $G_X(s)=\sum_{i=0}^{\infty}s^i\frac{\lambda^i}{i!}e^{-\lambda}=\mathrm{e}^{\lambda(s-1)}$. 利用

$$G'_X(1)=\lambda\mathrm{e}^{\lambda(s-1)}\Big|_{s=1}=\lambda,\qquad G''_X(1)=\lambda^2\mathrm{e}^{\lambda(s-1)}\Big|_{s=1}=\lambda^2$$

可求得

$$\mathbb{E}[X]=\mathrm{Var}(X)=\lambda.$$

注 可以通过母函数来验证一些离散型随机变量的**可加性**:

二项分布: $X_1\sim B(n_1,p), X_2\sim B(n_2,p)$, X_1 与 X_2 独立, 则

$$G_{X_1+X_2}(s)=(ps+q)^{n_1+n_2}$$

即 $X_1+X_2\sim B(n_1+n_2,p)$.

泊松分布: $X_1\sim P(\lambda_1), X_2\sim P(\lambda_2)$, X_1 与 X_2 独立, 则

$$G_{X_1+X_2}(s)=e^{(\lambda_1+\lambda_2)(s-1)}$$

即 $X_1+X_2\sim P(\lambda_1+\lambda_2)$.

引理 1.10

设 $\{X_k:k\geq 1\}$ i.i.d X, 随机变量 N 与 $\{X_k:k\geq 1\}$ 独立, $Y=\sum_{k=1}^{N}X_k$. 用母函数的方法证明:

$$\mathbb{E}[Y]=\mathbb{E}[N]\cdot\mathbb{E}[X],\qquad \mathrm{Var}(Y)=\mathbb{E}[N]\cdot\mathrm{Var}[X]+\mathrm{Var}[N]\cdot(\mathbb{E}[X])^2.$$

注 在本课程的作业和考试中若使用该引理必须给出证明.

证明 记 X 的母函数为 $F(s)$, N 的母函数为 $G(s)$. 则 Y 的母函数为 $Y(s)=G(F(s))$. 而

$$Y'(s)=G'(F(s))F'(s),\qquad Y''(s)=G''(s)(F(s))(F'(s))^2+G'(F(s))F''(s).$$

因此有

$$\mathbb{E}[Y]=Y'(1)=G'(F(1))F'(1)=G'(1)F'(1)=\mathbb{E}[N]\cdot\mathbb{E}[X]$$

及

$$\begin{aligned}\mathrm{Var}(Y)&=Y''(1)+Y'(1)-(Y'(1))^2=G''(1)(F'(1))^2+G'(1)F''(1)+G'(1)F'(1)-(G'(1)F'(1))^2\\&=G'(1)(F''(1)+F'(1)-(F'(1))^2)+(F'(1))^2(G''(1)+G'(1)-(G'(1))^2)\\&=\mathbb{E}[N]\cdot\mathrm{Var}[X]+\mathrm{Var}[N]\cdot(\mathbb{E}[X])^2.\end{aligned}$$

例题 1.21 设 $S_N=X_1+X_2+\cdots+X_N$ 为 N 个相互独立的随机变量之和, 其中每个随机变量等概率地取值 $1,2,\cdots,m$. 求:

(1) S_N 的概率母函数 $G(s)$.

(2) 关于 k 的序列 $\mathbb{P}(S_N \leqslant k)$ 的母函数 $G_1(s)$.

(3) 设 N 服从参数为 $p \in (0,1)$ 的几何分布, 且 N 与 $\{X_k : k \geq 1\}$ 独立, 试回答 (2) 中的问题.

解 (1) 我们有

$$G(s) = (\mathbb{E}[S^X])^N = \left(\frac{\sum_{k=1}^m s^k}{m}\right)^N = \left(\frac{s(1-s^m)}{m(1-s)}\right)^N.$$

(2) 注意到 $\mathbb{P}(S_N \leqslant 0) = 0$, 我们有

$$G(s) = \sum_{k=1}^{\infty} \mathbb{P}(S_N = k)s^k = \sum_{k=1}^{\infty} (\mathbb{P}(S_N \leqslant k) - \mathbb{P}(S_N \leqslant k-1))s^k = \sum_{k=1}^{\infty} \mathbb{P}(S_N \leqslant k)s^k - s\left(\sum_{k=1}^{\infty} \mathbb{P}(S_N \leqslant k)s^k\right)$$

因此

$$G_1(s) = \sum_{k=1}^{\infty} \mathbb{P}(S_N \leqslant k)s^k = \frac{G(s)}{1-s} = \frac{s^N(1-s^m)^N}{m^N(1-s)^{N+1}}.$$

(3) 记 N 的母函数是 $G_N(s)$, 例 1.20 已求得 $G_N(s) = \dfrac{ps}{1-(1-p)s}$, 故 (2) 中母函数 $G_0(s)$ 为

$$G_0(s) = G_N(G_1(s)) = \frac{pG_1(s)}{1-(1-p)G_1(s)} = \frac{ps(1-s^m)}{m(1-s)^2 - s(1-s^m)(1-s)(1-p)}.$$

练习 1.14 **Grimmett 5.12.8, 5.12.15**.

1.9 一维简单随机游走

$S_0 = a, S_n = S_{n-1} + X_n = S_0 + \sum_{i=1}^{n} X_i$ $X_1, \cdots, X_n$ 相互独立, 且 $\mathbb{P}(X_i = 1) = p, \mathbb{P}(X_i = -1) = q$, 这里 $p + q = 1$.

注 $p = \frac{1}{2}$ 时, 对称简单随机游走.

课堂上我们已经讲了双吸收壁模型、时齐性 + 空齐性 +Markov 性、计数 $N_n(a,b)$、反射原理及其应用—投票定理, 接下来我们再补充一些内容.

我们先回顾投票定理:

定理 1.4 (投票定理)

若 $b > 0$, 则

$$\#\{\text{从}(0,0)\text{到}(n,b)\text{且不再过}x\text{轴轨道}\} = \frac{b}{n}N_n(0,b).$$

由此推出:

定理 1.5 (不返回出发点)

$S_0 = 0, n \geq 1$, 则

$$\mathbb{P}(S_1S_2\cdots S_n \neq 0, S_n = b) = \frac{|b|}{n}\mathbb{P}(S_n = b).$$

因此有

$$\mathbb{P}(S_1S_2\cdots S_n \neq 0) = \frac{1}{n}\mathbb{E}[|S_n|].$$

证明 不妨设 $b > 0$, 利用投票定理得

$$\begin{aligned}\mathbb{P}(S_1S_2\cdots S_n \neq 0, S_n = b) &= \frac{b}{n}N_n(0,b)p^{\frac{n+b}{2}}q^{\frac{n-b}{2}}\\ &= \frac{b}{n}\mathbb{P}(S_n = b).\end{aligned}$$

设 $M_n = \max\{S_i : 0 \leq i \leq n\}$, 当 $S_0 = 0$ 时, $M_n \geq 0$.

定理 1.6

$S_0 = 0, r \geq 1$, 则有

$$\mathbb{P}(M_n \geq r, S_n = b) = \begin{cases}\mathbb{P}(S_n = b), & b \geq r,\\ \left(\frac{q}{p}\right)^{r-b}\mathbb{P}(S_n = 2r-b), & b < r.\end{cases}$$

证明 仅考虑 $b < r$, 记

$$Q = \{(0,0) \to (n,b)\text{且过某点}(i,r)\}$$

$\forall \pi \in Q$, 记 (i_π, r) 为 π 与 $y = r$ 的第一个交点, 然后反射 (i_π, r) 后面部分得到新轨道 π'(连接 $(0,0)$ 与 $(n, 2r-b)$), 两者有 $1-1$ 对应. 而

$$\frac{\mathbb{P}(\pi)}{\mathbb{P}(\pi')} = \frac{p^{\frac{n+b}{2}}q^{\frac{n-b}{2}}}{p^{\frac{n+2r-b}{2}}q^{\frac{n-2r+b}{2}}} = \left(\frac{q}{p}\right)^{r-b},$$

因此

$$\mathbb{P}(M_n \geq r, S_n = b) = \sum_{\pi\in Q}\mathbb{P}(\pi) = \left(\frac{q}{p}\right)^{r-b}\sum_{\pi'}\mathbb{P}(\pi') = \left(\frac{q}{p}\right)^{r-b}\mathbb{P}(S_n = 2r-b).$$

注 当 $p = \dfrac{1}{2}$ 时,

$$\begin{aligned}\mathbb{P}(M_n \geq r) = \mathbb{P}(S_n \geq r) + \sum_{b=-\infty}^{r-1}\left(\frac{q}{p}\right)^{r-b}\mathbb{P}(S_n = 2r-b) &\xlongequal{c=2r-b} \mathbb{P}(S_n \geq r) + \sum_{c=r+1}^{+\infty}\left(\frac{q}{p}\right)^{c-r}\mathbb{P}(S_n = c)\\ &= \mathbb{P}(S_n = r) + \sum_{c=r+1}^{+\infty}\left(1 + \left(\frac{q}{p}\right)^{c-r}\right)\mathbb{P}(S_n = c)\\ &\xlongequal{p=\frac{1}{2}} \mathbb{P}(S_n = r) + 2\mathbb{P}(S_n \geq r+1).\end{aligned}$$

定理 1.7 (首中时定理)

$S_0 = 0$, 在时刻 n 首次击中点 b 概率为

$$f_b(n) = \frac{|b|}{n}\mathbb{P}(S_n = b),\ n \geq 1.$$

证明 不妨设 $b>0$, 注意此时 $t=n$ 时达到新的最大值点, 因此

$$\begin{aligned}
f_b(n) &= \mathbb{P}(S_n=b, M_{n-1}=b-1, S_{n-1}=b-1)=\mathbb{P}(M_{n-1}=S_{n-1}=b-1, X_n=1)\\
&= p\mathbb{P}(M_{n-1}=b-1, S_{n-1}=b-1)\\
&= p\left(\mathbb{P}(M_{n-1}\geq b-1, S_{n-1}=b-1)-\mathbb{P}(M_{n-1}\geq b, S_{n-1}=b-1)\right)\\
&\xlongequal{\text{定理 1.6}} p\mathbb{P}(S_{n-1}=b-1)-p\cdot\frac{q}{p}\mathbb{P}(S_{n-1}=b+1)=\frac{n+b}{2n}\mathbb{P}(S_n=b)-\frac{n-b}{2n}\mathbb{P}(S_n=b)\\
&= \frac{b}{n}\mathbb{P}(S_n=b),
\end{aligned}$$

注 对比**定理** 1.4和**定理** 1.5, 思考能否类同投票定理计数观点来说明上述定理?

定理 1.8

$p=\dfrac{1}{2}, S_0=0$, 记最后一次访问原点: $T_{2n}=\max\{0\leq i\leq 2n: S_i=0\}$, 则有

$$\mathbb{P}(T_{2n}=2k)=\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2n-2k}=0).$$

证明

$$\begin{aligned}
\mathbb{P}(T_{2n}=2k) &= \mathbb{P}(S_{2k}=0, S_{2k+1}\cdots S_{2n}\neq 0)=\mathbb{P}(S_{2k}=0)\mathbb{P}(S_{2k+1}\cdots S_{2n}\neq 0\mid S_{2k}=0)\\
&\xlongequal{\text{时齐性}} \mathbb{P}(S_{2k}=0)\mathbb{P}(S_1S_2\cdots S_{2n-2k}\neq 0).
\end{aligned}$$

令 $m=n-k$,

$$\begin{aligned}
\mathbb{P}(S_1\cdots S_{2m}\neq 0) &\xlongequal{\text{定理 1.5}} \frac{1}{2m}\mathbb{E}[|S_{2m}|]=\frac{2}{2m}\sum_{i=1}^m 2i\cdot\mathbb{P}(S_{2m}=2i)=2\left(\frac{1}{2}\right)^{2m}\sum_{i=1}^m\frac{m+i-(m-i)}{2m}\binom{2m}{m+i}\\
&= 2\left(\frac{1}{2}\right)^{2m}\sum_{i=1}^m\left(\binom{2m-1}{m+i-1}-\binom{2m-1}{m+i}\right)=\left(\frac{1}{2}\right)^{2m}\binom{2m}{m}\\
&= \mathbb{P}(S_{2m}=0).
\end{aligned}$$

注 $\mathbb{P}(S_1S_2\cdots S_{2m}\neq 0)=\mathbb{P}(S_{2m}=0)$.

Stirling 公式: $n!\sim\left(\dfrac{n}{e}\right)^n\sqrt{2\pi n}, n\to\infty$. 因此

$$\begin{aligned}
\mathbb{P}(S_{2k}=0) &= \left(\frac{1}{2}\right)^{2k}\binom{2k}{k}\sim\frac{1}{\sqrt{\pi k}}, k\to\infty,\\
\mathbb{P}(S_{2n-2k}=0) &\sim \frac{1}{\sqrt{\pi(n-k)}}, n-k\to\infty.
\end{aligned}$$

例题 1.22 (反正弦律) 接**定理** 1.8, $n\to+\infty$ 时, 我们有

$$\begin{aligned}
\mathbb{P}(T_{2n}\leq 2xn) &= \sum_{k:k\leq xn}\mathbb{P}(T_{2n}=2k)\sim\sum_{\frac{k}{n}\leq x}\frac{1}{\pi\sqrt{k(n-k)}}=\sum_{\frac{k}{n}\leq x}\frac{1}{\pi}\frac{1}{\sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}}\cdot\frac{1}{n}\\
&\sim \int_0^x\frac{\mathrm{d}u}{\pi\sqrt{u(1-u)}}=\frac{2}{\pi}\arcsin\sqrt{x},
\end{aligned}$$

即 $\dfrac{T_{2n}}{2n}$ 的渐近分布为 $\dfrac{2}{\pi}\arcsin\sqrt{x}$.

例题 1.23 直线上简单随机游动 $S_n=\sum_{k=1}^n X_k, S_0=0$, 这里 X_i i.i.d 且 $P(X_i=1)=p, P(X_i=1)=1-p, 0<p<1$.

(1) 求 $\mathbb{E}(S_n), \mathrm{Var}(S_n), \mathrm{Cov}(S_m, S_n)$.

(2) Y 服从参数为 p 的几何分布且与 $\{X_k\}$ 独立时, 求 $\mathrm{Var}(S_Y)$.

(3) 对正整数 k, 求 S_{n+k} 关于 S_n 的条件分布列 $f_{S_{n+k}|S_n}$ 与条件期望 $\mathbb{E}[S_{n+k}|S_n]$.

(4) 求 $\mathbb{P}(S_1 \geqslant -1, S_2 \geqslant -1, \cdots S_{2n-2} \geqslant -1, S_{2n-1} = -1)$.

解 (1) 我们有

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n(2p-1),$$

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n(\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2) = n(1-(2p-1)^2) = 4np(1-p),$$

$$\mathrm{Cov}(S_m, S_n) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{\min\{m,n\}} \mathrm{Var}(X_i) = 4\min\{m,n\}\, p(1-p).$$

(2) 可直接取条件期望求解, 也可以通过引理 1.10(在本课程的作业和考试中若使用该引理必须给出证明) 求得

$$\mathrm{Var}(S_Y) = \mathbb{E}[Y]\cdot \mathrm{Var}[X_1] + \mathrm{Var}[Y]\cdot(\mathbb{E}[X_1])^2 = \frac{1}{p}\cdot 4p(1-p) + \frac{1-p}{p^2}\cdot(2p-1)^2 = \frac{(1-p)(8p^2-4p+1)}{p^2}.$$

(3) 利用空齐性和时齐性, 我们有

$$\mathbb{P}(S_{n+k} = a+b|S_n = a) = \mathbb{P}(S_k = b|S_0 = 0) = \begin{cases} \binom{k}{\frac{k+b}{2}} p^{\frac{k+b}{2}}(1-p)^{\frac{k-b}{2}}, & 2 \mid (k+b) \\ 0, & 2 \nmid (k+b) \end{cases}$$

且有

$$\mathbb{E}[S_{n+k}|S_n] = \mathbb{E}[S_n + X_{n+1} + \cdots + X_{n+k}|S_n] = S_n + \sum_{k=n+1}^{n+k} \mathbb{E}[X_k|S_n] \xlongequal{\text{独立}} S_n + \sum_{k=n+1}^{n+k} \mathbb{E}[X_k] = S_n + k(2p-1).$$

(4) 利用首中时原理, 我们有

$$\begin{aligned}
\mathbb{P}(S_1 \geqslant -1, \cdots S_{2n-2} \geqslant -1, S_{2n-1} = -1) &= \frac{1}{1-p}\mathbb{P}(S_1 \geqslant -1, \cdots S_{2n-2} \geqslant -1, S_{2n-1} = -1, S_{2n} = -2) \\
&= \frac{1}{1-p}\mathbb{P}(S_1 \geqslant -1, \cdots S_{2n-2} \geqslant -1, S_{2n-1} \geqslant -1, S_{2n} = -2) \\
&= \frac{1}{1-p}\frac{2}{2n}\mathbb{P}(S_{2n} = -2) \\
&= \frac{1}{n(1-p)}\binom{2n}{n+1} p^{n-1}(1-p)^{n+1}.
\end{aligned}$$

✍ **练习 1.15 Grimmett 3.11.27, 3.11.28**.

✍ **练习 1.16** 平面上一粒子 “向右向上” 随机游走, $S_0 = 0, S_n = S_{n-1} + X_n, n = 1, 2, \cdots$, 这里 X_i i.i.d $\mathbb{P}(X_i = (1,0)) = \mathbb{P}(X_i = (0,1)) = \frac{1}{2}$. 记 $C_{n,m}$ 为粒子从 $(0,0)$ 到 (mn, n) 且始终在直线 $y = \frac{x}{m}$ 及其上方运动的概率. 求

(1) 当 $m = 1$ 时计算 $C_{n,1}$;

(2) 当 $m = 2$ 时计算 $C_{3,2}$, 并验证序列 $\{C_{n,2}\}$ 的母函数 $G(s)$ 满足方程

$$G(s) = 1 + \frac{s}{8}G^3(s).$$

1.10 连续型随机变量

随机向量间的变换: 以二元为例, $(X_1,X_2)\to(Y_1,Y_2)=T(X_1,X_2)$, $y_i=y_i(x_1,x_2), i=1,2, T: D\subset\mathbb{R}^2\to R\subset\mathbb{R}^2$, $1-1$ 映射, $(x_1,x_2)\to(y_1,y_2)$, 其逆 $x_i=x_i(y_1,y_2)$ 有连续偏导数.

定理 1.9

(X_1,X_2) 有密度 $f(x_1,x_2)$, 则

$$f_{Y_1,Y_2}(y_1,y_2)=f(x_1(y_1,y_2),x_2(y_1,y_2))\,|J(y_1,y_2)|\,I_R$$

这里 $J=\begin{vmatrix}\dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_2}{\partial y_1}\\ \dfrac{\partial x_1}{\partial y_2} & \dfrac{\partial x_2}{\partial y_2}\end{vmatrix}$

证明 本质变量替换, 设 $A\subset D, B=T(A)\subset T(D)=R$, 则 $(X_1,X_2)\in A\Leftrightarrow(Y_1,Y_2)\in B$, 则

$$\begin{aligned}\mathbb{P}((Y_1,Y_2)\in B)&=\mathbb{P}((X_1,X_2)\in A)\\&=\iint_A f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\&=\iint_B f(x_1(y_1,y_2),x_2(y_1,y_2))\,|J|\,\mathrm{d}y_1\mathrm{d}y_2\end{aligned}$$

取 $B=T(D)\cap(-\infty,y_1]\times(-\infty,y_2]$

注

1. 若 $D_0\subset D,\mathbb{P}((X_1,X_2)\in D_0)=1, T$ 在 D_0 上单射, 结论亦对.
2. 这里二元情形可以一般化地推广到 n 元情形.

例题 1.24 设 X,Y 独立同 $N(0,1)$, 令

$$X=R\cos\Theta, Y=R\sin\Theta, R\geq 0, 0\leq\Theta\leq 2\pi$$

求 (R,Θ) 的联合密度.

解 $f(x,y)=\dfrac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$, $|J|=r$, 则

$$f_{R,\Theta}(r,\theta)=\frac{1}{2\pi}re^{-\frac{1}{2}r^2}, r>0,\theta\in[0,2\pi)$$

表明 R 与 Θ 独立, 且 $\Theta\sim U[0,2\pi)$, $f_R(r)=re^{-\frac{1}{2}r^2}, r>0$

副产品: 产生独立正态随机数, 设 U_1,U_2 独立同 $U[0,1]$, 令

$$\begin{aligned}X&=\sqrt{-2\log U_1}\cos(2\pi U_2)\\Y&=\sqrt{-2\log U_1}\sin(2\pi U_2)\end{aligned}$$

则 X 与 Y 独立同 $N(0,1)$.

例题 1.25 X,Y 独立同 $N(0,1)$, 令 $U=X, V=\rho X+\sqrt{1-\rho^2}Y, \rho\in(-1,1)$, 则 (U,V) 服从二元标准正态分布. 即

$$f_{U,V}(u,v)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\}.$$

例题 1.26 (次序统计量) 随机变量 $X_1,\cdots,X_n$ i.i.d $\sim F(x)$, 其密度函数为 $f(x)$. 将 $X_1,\cdots,X_n$ 从小到大排列为 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$.

(1) 证明:

$$\mathbb{P}(X_{(n)} < x) = (F(x))^n, \quad \mathbb{P}(X_{(1)} > x) = (1 - F(x))^n.$$

由此可以分别求出 $X_{(1)}, X_{(n)}$ 的密度函数.

(2) 证明: 对 $y_i = x_{(i)}$, n 个次序统计量 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 的联合分布

$$\mathbb{P}(X_{(1)} < y_1, \cdots, X_{(n)} < y_n) = \begin{cases} n!\mathbb{P}(X_1 < y_1, \cdots X_n < y_n, X_1 < X_2 < \cdots < X_n), & y_1 < \cdots < y_n, \\ 0, & \text{else}. \end{cases}$$

由此导出其联合密度

$$g(y_1, \cdots, y_n) = \begin{cases} n! f(y_1) \cdots f(y_n), & y_1 < \cdots < y_n, \\ 0, & \text{else}, \end{cases}$$

注 这里 $X_{(i)} = x_{(i)}$ 已给定次序, 故 else 情况下为 0.

(3) 对 $\forall 1 \leqslant i < j \leqslant n$, 求单个统计量 $X_{(i)}$ 的密度函数 $g_i(x)$ 和两个次序统计量 $(X_{(i)}, X_{(j)})$ 的联合密度函数 $g_{ij}(x, y)$ (仅考虑 $x < y$).

(4) (留作习题) 若 X_i i.i.d. $\sim U(0,1)$, 定义极差: $R = X_{(n)} - X_{(1)}$, 求 R 的密度函数.

解 (1)

$$\mathbb{P}(X_{(n)} < x) = \mathbb{P}(X_1 < x, \cdots, X_n < x) = \prod_{k=1}^{n} \mathbb{P}(X_k < x) = (F(x))^n,$$
$$\mathbb{P}(X_{(1)} > x) = \mathbb{P}(X_1 > x, \cdots, X_n > x) = \prod_{k=1}^{n} \mathbb{P}(X_k > x) = (1 - F(x))^n.$$

(2) 仅考虑 $x_1 < \cdots < x_n$. 注意到

$$\mathbb{P}(X_{(1)} < x_1, \cdots, X_{(n)} < x_n) = \sum_{\pi} \mathbb{P}(X_{\pi_1} < x_1, \cdots, X_{\pi_n} < x_n, X_{\pi_1} < \cdots < X_{\pi_n}),$$

即可. 其中 $(\pi_1, \cdots \pi_n)$ 为 $(1, 2, \cdots, n)$ 的排列.

(3) 注意到

$$\int \cdots \int_{a < x_1 < \cdots < x_n < b} f(x_1) \cdots f(x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n = \frac{1}{n!}(F(b) - F(a))^n.$$

我们有

$$\begin{aligned} g_i(x) = & \int \cdots \int_{-\infty < x_1 < \cdots < x_n < +\infty} n! f(x_1) \cdots f(x_{i-1}) f(x) f(x_{i+1}) \cdots f(x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \mathrm{d}x_n \\ = & n! \int \cdots \int_{-\infty < x_1 < \cdots < x_{i-1} < x} f(x_1) \cdots f(x_{i-1}) f(x) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \\ & \times \int \cdots \int_{x < x_{i+1} < \cdots < x_n < +\infty} f(x_{i+1}) \cdots f(x_n) \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \\ = & \frac{n!}{(i-1)!(n-i)!} (F(x))^{i-1} (1 - F(x))^{n-i} f(x). \end{aligned}$$

及

$$\begin{aligned}
g_{i,j}(x,y) =& n! \int \cdots \int_{-\infty<x_1<\cdots<x_{i-1}<x} f(x_1)\cdots f(x_{i-1}) f(x) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \\
& \times \int \cdots \int_{x<x_{i+1}<\cdots<x_{j-1}<y} f(x_{i+1})\cdots f(x_{j-1}) f(y) \mathrm{d}x_{i+1} \cdots \mathrm{d}x_{j-1} \\
& \times \int \cdots \int_{y<x_{j+1}<\cdots<x_n<+\infty} f(x_{j+1})\cdots f(x_n) \mathrm{d}x_{j+1} \cdots \mathrm{d}x_n \\
=& \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1}(F(y)-F(x))^{j-i-1}(1-F(y))^{n-j} f(x) f(y).
\end{aligned}$$

(4) 提示: 作变量代换: $Z = X_{(1)}, R = X_{(n)} - X_{(1)}$ 即可, 这里 $|J| = 1$, 代换后求 R 的边缘密度即可. 答案:

$$g_R(r) = \begin{cases} n(n-1)r^{n-2}(1-r), & 0 < r < 1, \\ 0, & \text{else}. \end{cases}$$

例题 1.27 (指数分布) $X \sim \mathrm{Exp}(\lambda), \lambda > 0$, 密度函数

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

分布函数 $F(x) = 1 - e^{-\lambda x}, x \geq 0$. 自行验证: $\mathbb{E}[X] = \dfrac{1}{\lambda}, \mathrm{Var}(X) = \dfrac{1}{\lambda^2}$.

(1) 若 X 是非负连续性随机变量, 则 X 是指数分布 $\Longleftrightarrow$ X 无记忆性.

(2) 设 $X_1, \cdots X_n$ 相互独立且分别服从参数为 $\lambda_1, \cdots \lambda_n$ 指数分布, 求 $Y = \min\{X_1, \cdots, X_n\}$ 的分布.

(3) 设 $X_1, \cdots X_n$ i.i.d $\sim \mathrm{Exp}(\lambda)$, 定义 $Y = X_1 + \cdots + X_n$, 则 Y 的密度函数是

$$g(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} \mathrm{e}^{-\lambda x}, x \geqslant 0$$

其中 $\Gamma(n) = (n-1)!$. 我们记 Y 服从 Γ 分布 $\Gamma(n, \lambda)$. 请读者自行验证 Γ 分布具有可加性: 若 $Z_1 \sim \Gamma(n, \lambda), Z_2 \sim \Gamma(m, \lambda), Z_1, Z_2$ 独立, 则有 $Z_1 + Z_2 \sim \Gamma(n+m, \lambda)$ (直接计算或通过特征函数说明).

解 (1) $\Longrightarrow$: 若 $X \sim \mathrm{Exp}(\lambda)$, 则对 $s, t > 0$ 有

$$\mathbb{P}(X > t+s \mid X > t) = \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > t)} = \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s);$$

$\Longleftarrow$: 记 $G(x) = \mathbb{P}(X > x)$, 则有

$$G(s+t) = G(s)G(t).$$

其中 $s, t \geqslant 0$. 对 $\forall n \in \mathbb{N}^*$ 及 $x \in \mathbb{R}^*$, 有 $G(nx) = (G(x))^n$. 取 $x = \dfrac{1}{n}$, 并记 $a = G(1) \geqslant 0$, 则有

$$G\left(\frac{1}{n}\right) = a^{\frac{1}{n}}.$$

因此, 对任意正整数 m, n, 有

$$G\left(\frac{m}{n}\right) = \left(G\left(\frac{1}{n}\right)\right)^m = a^{\frac{m}{n}}.$$

因此 $G(x) = a^x$ 对所有正有理数成立, 利用 $G(x)$ 的连续性 (或单调性), 则 $G(x) = a^x$ 对 $\forall x \geqslant 0$ 均成立. 因为 $G(x) \in [0,1]$, 且 $\lim\limits_{x \to +\infty} G(x) = 0$, 故 $a \in (0,1)$, 可写为 $a = \mathrm{e}^{-\lambda}$, 其中 $\lambda > 0$. 因此

$$F(x) = 1 - G(x) = 1 - \mathrm{e}^{-\lambda x}.$$

(2) X_i 的分布函数 $F_i(x)=1-\mathrm{e}^{-\lambda_i x}$, 类似例 1.26(**1**), 可得 Y 的分布函数

$$F_Y(x)=1-\prod_{i=1}^{n}(1-F_i(x))=1-\mathrm{e}^{-(\lambda_1+\cdots+\lambda_n)x},\ x>0.$$

因此 $Y\sim\mathrm{Exp}(\lambda_1+\cdots+\lambda_n)$.

(3) 可通过归纳计算得到 (自行验证), 或作变量代换 $Y_k=X_1+\cdots+X_k$, $k=1,\cdots,n$, 则有

$$\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\begin{pmatrix}1&&&\\1&1&&\\\vdots&\vdots&\ddots&\\1&1&\cdots&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}:=A\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}.$$

$$|J|=\det\left(\frac{\partial x_i}{\partial y_j}\right)_{ij}=\det(A^{-1})=1.$$

而 $(X_1,\cdots X_n)$ 的联合密度是

$$f(x_1,\cdots x_n)=\lambda^n\mathrm{e}^{-\lambda(x_1+\cdots+x_n)},\ x_1,\cdots,x_n>0.$$

故 $(Y_1,\cdots Y_n)$ 的联合密度为

$$g(y_1,\cdots y_n)=\lambda^n\mathrm{e}^{-\lambda y_n},\ y_n\geqslant y_{n-1}\geqslant\cdots\geqslant y_1\geqslant 0.$$

所以 Y_n 的联合密度

$$\begin{aligned}g_n(y_n)&=\int_0^{y_n}\int_0^{y_{n-1}}\cdots\int_0^{y_2}g(y_1,\cdots y_n)\mathrm{d}y_1\cdots\mathrm{d}y_{n-1}=\lambda^n\mathrm{e}^{-\lambda y_n}\int_0^{y_{n-1}}\cdots\int_0^{y_2}1\,\mathrm{d}y_1\cdots\mathrm{d}y_{n-1}\\&=\frac{\lambda^n}{(n-1)!}{y_n}^{n-1}\mathrm{e}^{-\lambda y_n}.\end{aligned}$$

例题 1.28 (正态分布 (高斯分布)) $X\sim N(\mu,\sigma^2)$, 分布函数记为 $F(x)$, 密度函数

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2},\ -\infty<x<+\infty,\ \mu\in\mathbb{R},\ \sigma>0.$$

自行验证: $\mathbb{E}[X]=\mu$, $\mathrm{Var}(X)=\sigma^2$.

(1) 请读者自行验证正态分布具有**可加性**: 若 $Z_1\sim N(\mu_1,\sigma_1^2)$, $Z_2\sim N(\mu_2,\sigma_2^2)$, Z_1,Z_2 独立, 则有 $Z_1+Z_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$ (直接计算或通过特征函数说明).

(2) $\mu=0,\sigma=1$. 记 $\gamma_n=\mathbb{E}[X^n]$, 则有

$$\gamma_n=\begin{cases}0, & n=2k-1,\\(2k-1)!!, & n=2k.\end{cases}$$

由此可知, 正态分布的任意阶矩均有限, 且奇阶矩为零.

(3) **Mill's ratio** $\mu=0,\sigma=1$, 利用 $f'(x)+xf(x)=0$, 证明:

$$\frac{1}{x}-\frac{1}{x^3}<\frac{1-F(x)}{f(x)}<\frac{1}{x}.$$

更一般地, 请读者自证:

$$\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}-\frac{15}{x^7}<\frac{1-F(x)}{f(x)}<\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}.$$

(4) $\mu=0,\sigma=1,a>0$, 证明: 当 $x\to+\infty$ 时, 有

$$\mathbb{P}\left(X>x+\frac{a}{x}\middle|X>x\right)\to\mathrm{e}^{-a}.$$

解 (2) 奇阶矩为零显然, 因为函数 $x^{2k-1}f(x)$ 是奇函数. 对于偶阶矩, 令 $t=\frac{x^2}{2}$, 则有

$$\mathbb{E}[X^{2k}]=\int_{-\infty}^{\infty}\frac{x^{2k}}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int_0^{\infty}\frac{2^{k-1}t^{k-\frac{1}{2}}}{\sqrt{\pi}}e^{-t}dt=\frac{2^{k-1}}{\sqrt{\pi}}\Gamma(k+\frac{1}{2})=(2k-1)!!.$$

(3) 反复利用分部积分, 并利用 $f'(x)+xf(x)=0$, 有

$$1-F(x)=\int_x^{+\infty}f(u)du=-\int_x^{+\infty}\frac{f'(u)}{u}du=\frac{f(x)}{x}+\int_x^{+\infty}\frac{f'(u)}{u^3}du=\frac{f(x)}{x}-\frac{f(x)}{x^3}-\int_x^{+\infty}\frac{3f'(u)}{u^5}du.$$

(4) 利用 (3) 可知,

$$\mathbb{P}\left(X>x+\frac{a}{x}\Big|X>x\right)=\frac{1-F\left(x+\frac{a}{x}\right)}{1-F(x)}=(1+o(1))\frac{e^{-\frac{1}{2}(x+\frac{a}{x})^2}}{e^{-\frac{1}{2}x^2}}\xrightarrow{x\to+\infty}e^{-a}.$$

✍ **练习 1.17** **Grimmett 4.6.7, 4.6.8, 4.7.2, 4.7.7(4.14.18), 4.14.14, 4.14.19, 4.14.27, 4.14.28, 4.14.29, 4.14.33, 4.14.39, 4.14.46**.

✍ **练习 1.18** 若 X,Y 独立且服从 $N(0,1)$, 求 $U=X^2+Y^2$ 与 $V=\frac{X}{Y}$ 的密度函数, 并证明它们是独立的.

✍ **练习 1.19** 若气体分子的速度是随机向量 $V=(X,Y,Z)$, 各分量相互独立且服从 $N(0,\sigma^2)$. 证明: $S=\sqrt{X^2+Y^2+Z^2}$ 服从 **Maxwell 分布律**:

$$p(s)=\sqrt{\frac{2}{\pi}}\frac{s^2}{\sigma^3}\exp\left(-\frac{s^2}{2\sigma^2}\right),s>0.$$

✍ **练习 1.20** 对连续型随机变量 X, 定义实数 m 为 X 的分布函数 F 的**中位数**, 若

$$F(m-0)\leqslant\frac{1}{2}\leqslant F(m)\Longleftrightarrow\mathbb{P}(X\leqslant m)\geqslant\frac{1}{2}\text{ 且 }\mathbb{P}(X\geqslant m)\geqslant\frac{1}{2}.$$

(1) **(本小问练习 1.4 已证)** 证明分布函数至少有一个中位数, 且中位数集合构成一个闭区间.

(2) 若 $\mathbb{E}[|X|]<\infty$, 证明:

$$\mathbb{E}[|X-m|]=\inf_{x\in\mathbb{R}}\mathbb{E}[|X-x|]$$

当且仅当 m 是 X 的中位数.

(3) 若 $\mathbb{E}[X^2]<\infty$, m 是 X 的中位数. 证明:

$$|\mathbb{E}[X]-m|\leqslant\sqrt{\mathrm{Var}(X)}.$$

✍ **练习 1.21** 已知一列标准高斯 (正态) 随机变量 $Z_1,\cdots,Z_n$ 独立同分布. 定义

$$A_n=\max_{1\le i\le n}|Z_i|,B_n=\max_{1\le i\le n}Z_i.$$

证明: 当 $n\to+\infty$ 时, 有

$$\mathbb{E}[A_n]=\sqrt{2\log n}+o(1),\quad\mathbb{E}[B_n]=\sqrt{2\log n}+o(1).$$

1.11 多元正态分布

定义 1.8

$\vec{X}=(X_1,\cdots,X_n)$ 服从 n 维正态分布, 若其密度

$$f(\vec{x})=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})'\right]$$

这里 $\vec{\mu}\in\mathbb{R}^n,\Sigma=(\sigma_{ij})_{i,j=1}^n$ 为正定矩阵, 记 $\vec{X}\sim N(\vec{\mu},\Sigma)$

♣

通过课堂上已证明的定理得到一些 **Facts**, 由此说明多元正态分布具有很好的" 结构":

(1) 多元正态分布可以通过均值向量和协方差矩阵唯一确定;
(2) 多元正态分布做任何线性变换后仍是多元正态分布.
(3) 随机向量服从多维正态分布，那么这个随机向量的某一部分所满足的多维正态分布可以直接由均值向量和协方差矩阵中的对应部分决定 (打洞);
(4) 多元正态分布独立性与不相关性等价;

推论 1.1

设 $\vec{X}$ 服从 n 维正态分布 $N(\vec{\mu},\Sigma)$, 则存在 n 维正交矩阵 A, 使得 $\vec{Y}=A\vec{X}$ 的各个分量相互独立.

提示 对任意 n 维正定矩阵 Σ, 存在 n 维正交矩阵 A, 满足 $A\Sigma A^T=I_n$.

推论 1.2

设 $\vec{X}$ 服从 n 维正态分布 $N(\vec{\mu},\Sigma)$, 且 $\vec{X}$ 的各个分量独立同方差, 则对任意 n 维正交矩阵 A,$\vec{Y}=A\vec{X}$ 的各个分量独立同方差.

提示 $\Sigma=\sigma^2 I_n, A\Sigma A^T=\sigma^2 I_n$.

定理 1.10

$\vec{X}\sim N(\vec{\mu},\Sigma)$, 则 $Y=\vec{X}\cdot\vec{t}^T$

$$\phi(\vec{t})=\mathbb{E}\left[e^{i\vec{x}\cdot\vec{t}^T}\right]=\mathbb{E}[e^{isY}]\Big|_{s=1}$$

$Y\sim N(\vec{\mu}\cdot\vec{t}^T,\vec{t}\Sigma\vec{t}^T)$, 当 $\vec{t}\neq 0$ 时

$$\phi(\vec{t})=e^{i\vec{\mu}\cdot\vec{t}^T s-\frac{1}{2}\vec{t}\Sigma\vec{t}^T s^2}\Big|_{s=1}=e^{i\vec{\mu}\cdot\vec{t}'-\frac{1}{2}\vec{t}\Sigma\vec{t}^T}$$

由此我们得到

定理 1.11

$\vec{X}$ 服从 n 维正态分布 $N(\vec{\mu},\Sigma)$ 当且仅当对任意 n 维实向量 $\vec{t}$, $\vec{Y}=\vec{t}^T\vec{X}$ 均服从一维正态分布 $N(\vec{t}^T\vec{\mu},\vec{t}^T\Sigma\vec{t})$.

注

1. 只满足 n 维随机向量 $\vec{X}$ 的各分量均服从一维正态分布不能推出 $\vec{X}$ 服从 n 维正态分布, 反例如下: 设 X,Y 独立且服从 $N(0,1)$, 令

$$Z=\begin{cases}|Y|, & X\geqslant 0,\\ -|Y|, & X<0.\end{cases}$$

则 $Z\sim N(0,1)$, 但 (Y,Z) 不服从二维正态分布 (可计算出 $\mathbb{P}(Y+Z=0)=\frac{1}{2}$, 因此 $Y+Z$ 不服从一维正态分布, 再由**定理** 1.11即得).

2. 我们知道, 多元正态分布独立性与不相关性等价. 但对于多个服从一维正态分布的随机变量, 该等价性不再成立. 例如 1 中的随机变量 X,Y, $\mathrm{Cov}(Y,Z)=0$ 即不相关, 但 Y,Z 不独立.

我们亦可用特征函数的方法来验证课堂上多元正态分布中已讲过的定理及命题, 这里不再赘述.

例题 1.29 $\vec{X} \sim N(\vec{\mu}, \Sigma), \vec{X}^T = (\vec{X}^{(1)}, \vec{X}^{(2)}), \vec{\mu}^T = (\vec{\mu}^{(1)}, \vec{\mu}^{(2)})$ $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, 求 $\vec{X}^{(2)}|\vec{X}^{(1)}$ 的条件分布及 $\mathbb{E}[\vec{X}^{(2)}|\vec{X}^{(1)}]$, $\mathrm{Var}[\vec{X}^{(2)}|\vec{X}^{(1)}]$.

证明 打洞. 利用

$$\begin{pmatrix} I_{n_1} & O \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n_2} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} I_{n_1} & -\Sigma_{11}^{-1}\Sigma_{12} \\ O & I_{n_2} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix},$$

我们构造

$$\vec{Y} = \begin{pmatrix} \vec{Y}^{(1)} \\ \vec{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} I_{n_1} & O \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n_2} \end{pmatrix}\begin{pmatrix} \vec{X}^{(1)} \\ \vec{X}^{(2)} \end{pmatrix} = \begin{pmatrix} \vec{X}^{(1)} \\ -\Sigma_{21}\Sigma_{11}^{-1}\vec{X}^{(1)} + \vec{X}^{(2)} \end{pmatrix},$$

则有

$$\vec{Y} \sim N\left(\begin{pmatrix} \vec{\mu}^{(1)} \\ -\Sigma_{21}\Sigma_{11}^{-1}\vec{\mu}^{(1)} + \vec{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}\right).$$

因此 $\vec{Y}^{(1)}$ 与 $\vec{Y}^{(2)}$ 独立,

$$\vec{Y}^{(2)}|\vec{Y}^{(1)} \sim N(-\Sigma_{21}\Sigma_{11}^{-1}\vec{\mu}^{(1)} + \vec{\mu}^{(2)}, \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

而 $\vec{X}^{(1)} = \vec{Y}^{(1)}$, 所以

$$\vec{X}^{(2)}|\vec{X}^{(1)} \sim N(\Sigma_{21}\Sigma_{11}^{-1}(\vec{X}^{(1)} - \vec{\mu}^{(1)}) + \vec{\mu}^{(2)}, \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}),$$

因此 $\mathbb{E}[\vec{X}^{(2)}|\vec{X}^{(1)}] = \Sigma_{21}\Sigma_{11}^{-1}(\vec{X}^{(1)} - \vec{\mu}^{(1)}) + \vec{\mu}^{(2)}$, $\mathrm{Var}[\vec{X}^{(2)}|\vec{X}^{(1)}] = \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

1.12 正态分布下的样本均值与样本方差

两个统计量: 从总体中抽取样本 $X_1, \cdots, X_n$

样本均值: $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

样本方差: $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$

例题 1.30 $X_1, \cdots, X_n$ i.i.d X, $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, 则 $\mathbb{E}[\overline{X}] = \mu$, $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$, $\mathbb{E}[S^2] = \sigma^2$(用来估计 μ 和 σ^2)

证明 只证明 $\mathbb{E}[S^2] = \sigma^2$.

$$\begin{aligned}
\mathbb{E}[S^2] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}[(x_i - \mu - (\overline{X} - \mu))^2] = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{Var}(X_i) + \mathrm{Var}(\overline{X}) - 2\mathbb{E}[(X_i - \mu)(\overline{X} - \mu)]) \\
&= \frac{1}{n-1}\left(n\sigma^2 + n\cdot\frac{1}{n}\sigma^2 - 2n\cdot\frac{1}{n}\sigma^2\right) \\
&= \sigma^2.
\end{aligned}$$

定义 1.9

当 X 有密度函数

$$f(x) = \frac{1}{2^{\frac{d}{2}}\Gamma\left(\frac{d}{2}\right)} x^{\frac{d}{2}-1} e^{-\frac{1}{2}x},\ x > 0.$$

则 X 服从 d 个自由度的卡方分布，记 $X \sim \chi^2(d)$.

注 $\chi^2(d) = \Gamma\left(\frac{d}{2}, \frac{1}{2}\right)$.

引理 1.11

$Y_1, \cdots, Y_d$ 独立同 $N(0,1)$, $X = \sum_{i=1}^{d} Y_i^2$, 则 $X \sim \chi^2(d)$.

证明 $(Y_1, \cdots, Y_d)$ 联合密度 $f(y_1, \cdots, y_d) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\sum_{j=1}^{d} y_j^2}$, 因此

$$F_X(x) = \mathbb{P}\left(\sum_{j=1}^{d} Y_i^2 \le x\right) = \int_{\sum y_i^2 \le x} f(y_1, \cdots, y_d)\mathrm{d}y_1 \cdots \mathrm{d}y_d.$$

极坐标换元后有:

$$\begin{aligned} F_X(x) &= \mathrm{C}_d \int_0^{\sqrt{x}} r^{d-1} e^{-\frac{1}{2}r^2} \mathrm{d}r \\ &= \frac{\mathrm{C}_d}{2} \int_0^{x} r^{\frac{d}{2}-1} e^{-\frac{1}{2}r} \mathrm{d}r. \end{aligned}$$

利用 $F_X(\infty) = 1$, 得: $\frac{\mathrm{C}_d}{2} = \frac{1}{2^{\frac{d}{2}}\Gamma\left(\frac{d}{2}\right)}$.

注 卡方分布具有可加性, 其证明已留为作业.

定理 1.12

$X_1, \cdots, X_n$ 独立同 $N(\mu, \sigma^2)$, 则

(1) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$.

(2) $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$.

(3) $\overline{X}$ 与 S^2 独立.

证明 令 $Y_i = \frac{1}{\sigma}(X_i - \mu)$, 则 $Y_i \sim N(0,1)$, 且 $Y_1, \cdots, Y_n$ 相互独立, 显见

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{\sigma}(\overline{X} - \mu).$$

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n-1}{\sigma^2} S^2.$$

$(Y_1,\cdots,Y_n)$ 密度 $f(y)=\dfrac{1}{(\sqrt{2\pi})^n}e^{-\frac{1}{2}\sum_{i=1}^n y_i^2}$, 取正交阵

$$A=\begin{pmatrix} \frac{1}{\sqrt{n}} & * & * & * \\ \vdots & * & * & * \\ \vdots & * & * & * \\ \frac{1}{\sqrt{n}} & * & * & * \end{pmatrix},$$

令 $(Z_1,\cdots,Z_n)=(Y_1,\cdots,Y_n)A$, 则 $Z_1=\sqrt{n}\overline{Y}$, 且 $Z_1,\cdots,Z_n$ 独立同 $N(0,1)$, 因为 $\vec{Y}\sim N(0,I_n)$ 知 $\vec{Z}\sim N(0,A^TI_nA)=N(0,I_n)$. 又可知

$$Z_1^2+\cdots+Z_n^2=Y_1^2+\cdots+Y_n^2,$$

有

$$\begin{aligned}\sum_{i=2}^n Z_i^2&=\sum_{i=1}^n Y_i^2-n\overline{Y}^2=\sum_{i=1}^n(Y_i-\overline{Y}+\overline{Y})^2-n\overline{Y}^2\\&=\sum_{i=1}^n(Y_i-\overline{Y})^2=\frac{n-1}{\sigma^2}S^2.\end{aligned}$$

因此 $\overline{Y}\sim N(0,\dfrac{1}{n})$, $\dfrac{n-1}{\sigma^2}S^2\sim\chi^2(n-1)$, 且 $\overline{X}$ 与 S^2 独立.

1.13 期望浅谈

在概率论中, 关于分布函数的积分都是指 Lebesgue 积分, 而不是 Riemann 积分. 这两者有根本区别, 后者是对函数的定义域 “分蛋糕”, 而前者是对函数的值域 “分蛋糕”(参考可测函数 (随机变量) 的定义). 同时注意到, 可测函数对函数的连续性根本不作要求. 从动机讲, 为了把积分对象扩充到更大的一类函数——可测函数类上, 而 Riemann 积分就不够用了, 因此必须要在此基础上进一步推广, 同时注意到可测函数在极限运算下封闭, 为保持可测函数的结构和运算, 于是就据此建立了 Lebesgue 积分这一套理论. 具体内容可以参考实分析教材.

我们回顾一下课堂中已讲过的内容: 定义一般随机变量 X 的期望 $\mathbb{E}[X]$ 的过程——三步走战略 (若第一步是示性函数则为四步走).

非负简单随机变量 $\longrightarrow$ 非负随机变量 $\longrightarrow$ 一般随机变量

$$X=\sum_{i=1}^n x_iI_{A_i},\ X_n=nI_{A_n}+\sum_{j=1}^{n\cdot 2^n}\frac{j-1}{2^n}I_{A_{n,j}}\xrightarrow{n\to+\infty}X,\ X=X^+-X^-.$$

其中对 $X=\sum_{i=1}^n x_iI_{A_i}$,

$$\mathbb{E}[X]=\sum_{i=1}^n x_i\mathbb{P}(A_i).$$

X 可积指期望 $\mathbb{E}[|X|]$ 存在 $\Longleftrightarrow\mathbb{E}[X^+],\mathbb{E}[X^-]$ 均存在.

以上过程可以借助下面几个 **Facts** 来理解:

1. 随机变量 $X:\Omega\to\mathbb{R}$ 是实值 Borel 可测函数: $\forall x\in\mathbb{R},\{X\leqslant x\}\in\mathcal{F}$. **动机**: 对于随机变量, 相比过程 (Ω), 更在乎输出结果.
2. 随机变量在极限运算下封闭: $\forall n\in\mathbb{N}^*$, X_n 是随机变量. 若 $\liminf\limits_{n\to\infty}X_n=\limsup\limits_{n\to\infty}X_n$, 则 $X:=\lim\limits_{n\to+\infty}X_n$ 也是随机变量 ($\liminf\limits_{n\to\infty}X_n,\limsup\limits_{n\to\infty}X_n$ 是随机变量).
3. X 未必是连续型的.

期望的一些**性质**: 设 X,Y,X_n 是随机变量, a,b,c 是常数, 则有

(1) **非负性**: 当 $X\geqslant 0$ 时, $\mathbb{E}[X]\geqslant 0$.

(2) **归一性**: $\mathbb{E}[c]=c,c\in\mathbb{R}$.

(3) **线性性**: $\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$.

(4) **单调性**: 当 $X\leqslant Y$ 时, $\mathbb{E}[X]\leqslant\mathbb{E}[Y]$.

(5) **绝对值不等式**: $|\mathbb{E}[X]|\leqslant\mathbb{E}[|X|]$.

(6) **单调收敛定理**: $X_{n+1}\geqslant X_n\geqslant 0,\forall n\in\mathbb{N}^*$, 且 $X_n\to X,n\to+\infty$, 则

$$\mathbb{E}[X_n]\to\mathbb{E}[X],n\to\infty.$$

这里 $\mathbb{E}[X]\leqslant+\infty$.

(7) **控制收敛定理**: $X_n\to X,|X_n|\leqslant Y,\forall n\in\mathbb{N}^*$, 当 $\mathbb{E}[Y]<\infty$ 时

$$\mathbb{E}[X_n]\to\mathbb{E}[X],n\to\infty,$$

其中 Y 为常数 c 时又称**有界收敛定理**.

(8) **Fatou 引理**: $X_n\geqslant 0$, 则

$$\mathbb{E}[\liminf_{n\to\infty}X_n]\leqslant\liminf_{n\to\infty}\mathbb{E}[X_n].$$

注

1. 由上述期望定义知, 若 $\mathbb{P}(A)=0,\forall A\in\mathcal{F}$, 则有 $\displaystyle\int_A X\mathrm{d}F=0$.
2. 由**注 1** 可知, 在原集合基础上挖去概率测度为 0 的集合并不会改变积分值, 因此很多性质我们只需考虑几乎处处 (a.s.) 条件下满足即可: $\exists\Omega_0,\mathbb{P}(\Omega_0)=0,\forall\omega\in\Omega\backslash\Omega_0$.

定理 1.13 (逐项积分定理)

设随机变量 $X_n\geqslant 0$, 则有

$$\mathbb{E}[\sum_{k=1}^{\infty}X_k]=\sum_{k=1}^{\infty}\mathbb{E}[X_k].$$

$S_n=\displaystyle\sum_{k=1}^{n}X_k\uparrow\sum_{k=1}^{\infty}X_k$, 再利用单调收敛定理即可.

利用**引理** 1.9或上述定理可得: 对于非负整值随机变量 X, 有 $\mathbb{E}[X]=\displaystyle\sum_{n=1}^{\infty}\mathbb{P}(X\geqslant n)$. 课堂上我们已经证明: 对于连续型随机变量 X, 我们亦有

$$\mathbb{E}[X]=\int_0^{\infty}(1-F(x))\mathrm{d}x-\int_{-\infty}^{0}F(x)\mathrm{d}x.$$

其中 F 是 X 的分布函数. 对于一般随机变量, 我们可以进行估计:

引理 1.12

设 X 是随机变量, 则有

$$\sum_{n=1}^{\infty}\mathbb{P}(|X|\geqslant n)\leqslant\mathbb{E}(|X|)\leqslant\sum_{n=1}^{\infty}\mathbb{P}(|X|\geqslant n)+1$$

证明 注意到

$$\sum_{n=1}^{\infty}\mathbb{P}(|X|\geqslant n)=\sum_{n=1}^{\infty}\mathbb{P}([|X|]\geqslant n)=\mathbb{E}([|X|])\leqslant\mathbb{E}(|X|)\leqslant\mathbb{E}([|X|]+1)=\sum_{n=1}^{\infty}\mathbb{P}(|X|\geqslant n)+1.$$

即可.

例题 1.31 设 $X\geqslant 0$, 证明

$$\mathbb{E}[X]=0\iff X\xlongequal{\text{a.s.}}0.$$

证明 $\Longleftarrow$: 注意到上述注 **1** 即可;

$\Longrightarrow$: 即证 $\mathbb{P}(X>0)=0$. 而 $\{X>0\}=\bigcup_{n=1}^{\infty}\left\{X>\frac{1}{n}\right\}$. 故有

$$\mathbb{P}\left(X>\frac{1}{n}\right)=\mathbb{E}[I_{\{X>\frac{1}{n}\}}]=n\mathbb{E}\left[\frac{1}{n}I_{\{X>\frac{1}{n}\}}\right]\leqslant n\mathbb{E}[XI_{\{X>\frac{1}{n}\}}]\leqslant n\mathbb{E}[X]=0.$$

由概率测度的次可加性知,

$$\mathbb{P}(X>0)\leqslant\sum_{n=1}^{\infty}\mathbb{P}\left(X>\frac{1}{n}\right)=0.$$

注 类似上述方法可以证明: 若 X 在 Ω 上非负可积, 则 X 在 Ω 上 a.s. 有限. (注意到 $\{X=+\infty\}=\bigcap_{k=1}^{\infty}\{X>k\}$ 即可)

例题 1.32 设随机变量 X 可积, $\mathbb{P}(A_n)\to 0$, 证明: $\int_{A_n}X\mathrm{d}F\to 0$.

证明 只需证: $\int_{A_n}|X|\mathrm{d}F\to 0$. 做截断: 定义随机变量

$$X_k=|X|I_{\{|X|<k\}}+kI_{\{|X|\geqslant k\}}.$$

利用单调收敛定理, 有

$$0\leqslant X_k\uparrow|X|\implies\mathbb{E}[X_k]\uparrow\mathbb{E}[|X|].$$

因此

$$\int_{A_n}|X|\mathrm{d}F=\int_{A_n}X_k\mathrm{d}F+\int_{A_n}(|X|-X_k)\mathrm{d}F\leqslant k\mathbb{P}(A_n)+\mathbb{E}[|X|-X_k].$$

对 $\forall\varepsilon>0$, 取 $k=k_0$, 使得 $\mathbb{E}[|X|-X_{k_0}]<\frac{\varepsilon}{2}$. 且 $\exists N\in\mathbb{N}^*$, 当 $n>N$ 时, 有

$$\mathbb{P}(A_n)<\frac{\varepsilon}{2k_0}.$$

回到上式, 则有

$$\int_{A_n}|X|\mathrm{d}F<k_0\cdot\frac{\varepsilon}{2k_0}+\frac{\varepsilon}{2}=\varepsilon.$$

注 体会一下这题用的**截断法**, 在后续讲到随机变量的收敛中这是一个常用的 technique.

练习 1.22 **Grimmett 4.3.3** .

1.14 特征函数

X 的**特征函数**: $\phi_X(t) = \mathbb{E}[e^{itX}]$. 多元情形: $\vec{X} = (X_1, \cdots, X_n)$ 特征函数, $\phi_{\vec{X}}(\boldsymbol{t}) = \mathbb{E}[e^{i\sum_j t_j X_j}]$.

注 对于连续型随机变量, 有

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx.$$

此时 $\phi(x)$ 是 $f(x)$ 的 **Fourier 变换**.

定理 1.14

对特征函数: $\phi(t) = \mathbb{E}[e^{itX}]$

(1) $\phi(0) = 1, |\phi(t)| \le 1, \overline{\phi(t)} = \phi(-t)$.

(2) ϕ 在 $\mathbb{R}$ 上一致连续.

(3) ϕ 非负定: $\forall t_1, \cdots, t_n \in \mathbb{R}, z_1, \cdots, z_n \in \mathbb{C}$, 有

$$\sum_{j,k=1}^{n} \phi(t_j - t_k) z_j \bar{z}_k \ge 0$$

注 **定理**1.14完全划分特征函数, Bochner 定理表明反之亦成立.

定理 1.15 (独立性)

我们有

(1) 当 X 与 Y 独立时, $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

(2) X 与 Y 独立 $\Leftrightarrow \phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t)$.

注

1. 对于独立性证明或求有关独立随机变量的和的问题, 用特征函数处理常常能够简化计算.
2. (1) 的逆命题并不成立, 我们有如下反例: 设 (X_1, X_2) 的联合密度是

$$f(x,y) = \frac{1 + xy(x^2 + y^2)}{4}, \quad |x|, |y| \leqslant 1,$$

则 X_1, X_2 不相互独立, 但有 $\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t)$.

利用我们已学过的反转公式和唯一性定理, 可以得到

$$X\text{与}Y\text{同分布} \iff \phi_X(t) = \phi_Y(t).$$

即特征函数与分布函数是相互唯一确定的. 同时对连续型随机变量, 我们有如下定理 (Fourier 逆变换):

定理 1.16

设 X 是连续型随机变量, 其密度函数为 $f(x)$, 特征函数为 $\phi(x)$, 若 $\int_{-\infty}^{+\infty} |\phi(t)| dt < +\infty$, 则

$$f(x) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{-itx} \phi(t) dt$$

证明 利用反转公式 + 控制收敛定理, 具体见 **Grimmett 5.12.20**.

注 连续型随机变量的密度函数和特征函数可用 Fourier 变换联系.

定理 1.17 (Parseval 等式)

对随机变量 X,Y, 有

$$\int_{-\infty}^{+\infty}\phi_X(t)\mathrm{d}F_Y(t)=\int_{-\infty}^{+\infty}\phi_Y(t)\mathrm{d}F_X(t).$$

证明 不妨设 X,Y 独立, 则有

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}XY}]=\mathbb{E}[\mathbb{E}[\mathrm{e}^{\mathrm{i}XY}|Y]]=\mathbb{E}[\phi_X(Y)]=\int_{-\infty}^{+\infty}\phi_X(t)\mathrm{d}F_Y(t).$$

注 Parseval 等式可以把分布函数的性质转化为特征函数的性质, 即 Fourier 变换是自伴变换.

例题 1.33 常见离散型及连续型随机变量的特征函数:

(1) 二项分布: $B(n,p),\phi(t)=(p\mathrm{e}^{\mathrm{i}t}+q)^n$;

(2) 几何分布: $G(p),\phi(t)=\dfrac{p\mathrm{e}^{\mathrm{i}t}}{1-q\mathrm{e}^{\mathrm{i}t}}$;

(3) 泊松分布: $P(\lambda),\phi(t)=\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}$;

(4) 指数分布: $\mathrm{Exp}(\lambda),\phi(t)=\dfrac{\lambda}{\lambda-\mathrm{i}t}$;

(5) Γ 分布: $\Gamma(n,\lambda),\phi(t)=\left(\dfrac{\lambda}{\lambda-\mathrm{i}t}\right)^n$ (利用指数分布特征函数即可导出);

(6) 均匀分布: $U(a,b),\phi(t)=\dfrac{\mathrm{e}^{\mathrm{i}bt}-\mathrm{e}^{\mathrm{i}at}}{\mathrm{i}t(b-a)}$. 特别地, 当 $a=-u,b=u$ 时, $\phi(t)=\dfrac{\sin ut}{ut}$.

(7) 正态分布: $N(\mu,\sigma^2),\phi(t)=\mathrm{e}^{\mathrm{i}\mu t-\frac{\sigma^2t^2}{2}}$;

注

1. 我们可以通过其特征函数【先猜】其分布, 再利用唯一性定理【后证】.
2. 利用特征函数的方法, 我们可以直接得到二项分布、泊松分布、Γ 分布和正态分布的可加性. 如

$$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2),X,Y\text{ 独立 }\implies X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2).$$

例题 1.34 设 $\phi(t)$ 是特征函数, 证明 $\overline{\phi}(t),\phi^2(t),|\phi(t)|^2,\mathrm{Re}(\phi(t))$ 是特征函数.

解 (1) 设 $\phi(t)$ 是 X 的特征函数, 则有

$$\overline{\phi}(t)=\overline{\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}]}=\mathbb{E}[\mathrm{e}^{-\mathrm{i}tX}]=\phi_{-X}(t).$$

(2)(3) 设 $\phi(t)$ 是 X_1,X_2 的特征函数, 且 X_1,X_2 相互独立, 则有

$$\phi_{X_1+X_2}(t)=\phi_{X_1}(t)\phi_{X_2}(t)=\phi^2(t),\quad \phi_{X_1-X_2}(t)=\phi_{X_1}(t)\phi_{-X_2}(t)=\phi(t)\overline{\phi}(t)=|\phi(t)|^2.$$

(4) 设 $\phi(t)$ 是 X 的特征函数, Z 满足 $\mathbb{P}(Z=X)=\mathbb{P}(Z=-X)=\dfrac{1}{2}$, 则有

$$\phi_Z(t)=\frac{1}{2}(\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}]+\mathbb{E}[\mathrm{e}^{-\mathrm{i}tX}])=\frac{1}{2}(\phi(t)+\overline{\phi}(t))=\mathrm{Re}(\phi(t)).$$

注 $X\overset{d}{=\!=}-X\iff\phi(t)$ 是实值函数.

练习 1.23 **Grimmett 5.7.6, 5.7.9, 5.8.1, 5.8.6, 5.9.4, 5.12.14, 5.12.23, 5.12.26(a)-(d), 5.12.27(b)-(d), 5.12.28**

.

练习 1.24 设 X 和 Y 为独立同分布的随机变量, 且 $\mathrm{Var}(X)=1$. 证明:

$$\forall a,b\in\mathbb{R},aX+bY\overset{D}{=\!=}\sqrt{a^2+b^2}X\iff X\sim N(0,1).$$

1.15 矩方法与组合计数

$$\gamma_k = \int_{\mathbb{R}} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathrm{d}x,\ \gamma_{2m-1} = 0,\ \gamma_{2m} = (2m-1)!!.$$

组合诠释: 将 $1, 2, \cdots, 2m$ 配成 m 对, 共有 $(2m-1)!!$ 种.

定理 1.18

$\{X_k\}$ 相互独立, 且满足

(1) $\mathbb{E}[X_k] = 0, \mathrm{Var}(X_k) = 1, \forall k$

(2) (一致有界高阶矩) $\forall m \geq 3$, $\mathrm{C}_m = \sup_k \mathbb{E}[|X_k|^m] < \infty$

则对 $S_n = \sum_{i=1}^{n} X_i$ 有

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right] \to \gamma_k, n \to \infty$$

进而,

$$\frac{S_n}{\sqrt{n}} \xrightarrow{D} N(0,1)$$

证明

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right] = n^{-\frac{k}{2}} \sum_{1 \leq i_1, \cdots, i_k \leq n} \mathbb{E}[X_{i_1} X_{i_2} \cdots X_{i_k}]$$

(1) $k = 0, 1$ 显然

(2) $k = 2$

$$\mathrm{LHS} = \frac{1}{n} \sum_{i_1} \mathbb{E}[X_{i_1}^2] + \frac{1}{n} \sum_{i_1 \neq i_2} \mathbb{E}[X_{i_1} X_{i_2}] = 1$$

(3) $k = 3$

$$\begin{aligned} \mathrm{LHS} &= n^{-\frac{3}{2}} \sum_{i_1} \mathbb{E}[X_{i_1}^3] + n^{-\frac{3}{2}} \sum_{i_1 \neq i_2} 3\mathbb{E}[X_{i_1}^2 X_{i_2}] \\ &+ n^{-\frac{3}{2}} \sum_{i_1 \neq i_2 \neq i_3} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3}] = O\left(n^{-\frac{1}{2}}\right) \to 0 \end{aligned}$$

(4) $k = 4$

$$\begin{aligned} \mathrm{LHS} &= \frac{1}{n^2} \sum_{i_1} \mathbb{E}[X_{i_1}^4] + \frac{1}{n^2} \sum_{i_1 \neq i_2} 4\mathbb{E}[X_{i_1}^3 X_{i_2}] + \frac{1}{n^2} \sum_{i_1 \neq i_2} 3\mathbb{E}[X_{i_1}^2 X_{i_2}^2] \\ &+ \frac{1}{n^2} \sum_{i_1 \neq i_2 \neq i_3} 6\mathbb{E}[X_{i_1}^2 X_{i_2} X_{i_3}] + \frac{1}{n^2} \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}] \\ &= O\left(\frac{1}{n}\right) + 0 + \frac{1}{n^2} \cdot 3n(n-1) + 0 + 0 \to 3 \end{aligned}$$

(5) 一般情形, LHS 中非零项 $\mathbb{E}[X_{i_1} \cdots X_{i_k}]$ 必可写为形如 $\mathbb{E}[X_{i_1}^{a_1} \cdots X_{i_m}^{a_m}], i_1 \neq \cdots \neq i_m$, 且 $a_1, \cdots, a_m \geq 2$. 因 $a_1 + \cdots + a_m = k$, 则 $m \leq \left[\frac{k}{2}\right]$, 表明 k 为奇数时,

$$\mathrm{LHS} = n^{-\frac{k}{2}} O(n^m) \to 0.$$

当 k 为偶数时, 主项必在 $m = \frac{k}{2}$ 时取到, 这时 $a_1 = a_2 = \cdots = a_m = 2$, 从 $i_1, i_2, \cdots, i_{2m-1}, i_{2m}$ 两两配对, 配对总数是 $\gamma_k = (k-1)!!$. 对每一种给定的配对方式, 对应极限

$$n^{-\frac{k}{2}} \sum_{\substack{i_1,\cdots,i_k \\ \text{一种配对}}} 1 = \frac{n(n-1)\cdots(n-m+1)}{n^m} \to 1.$$

即证.

详细分类: $\{1,2,\cdots,k\}$ 划分为 m 组, 每组至少 2 个数字, 记 $\Gamma_k^{(m)}$ 为所有划分方式, 每种方式形如 $V = \{V_1, V_2, \cdots, V_m\}$, 这里 $V_i \cap V_j = \emptyset (\forall i \neq j)$, $\bigcup_j V_j = \{1,\cdots,k\}$, $\#V_j \geq 2$. 可定义等价类: $\vec{i} = (i_1, i_2, \cdots, i_k) \in V$, 若

$$i_p = i_q \Leftrightarrow p, q \in V_j$$

定理 1.18中依分布收敛可由矩收敛定理证出.

定理 1.19 (矩收敛定理)

假设

(1) $\forall k \in \mathbb{N}, \gamma_{k,n} = \int x^k \mathrm{d}F_n$ 存在

(2) $\forall k, \lim_{n\to\infty} \gamma_{k,n} = \gamma_k$

(3) $\gamma_k = \int x^k \mathrm{d}F$, 且满足 Carleman 条件

$$\sum_{k=1}^{\infty} (\gamma_{2k})^{-\frac{1}{2k}} = \infty$$

则 $F_n \xrightarrow{w} F$.

注

1. 自行验证标准正态分布的 k 阶矩 γ_k 满足 Carleman 条件, 从而**定理** 1.18中
$$\frac{S_n}{\sqrt{n}} \xrightarrow{D} N(0,1).$$
成立
2. 如果高阶矩不存在, 我们可以尝试用截断法解决.

定理 1.20 (Wick 公式)

$$\mathbb{E}[X^{2m}] = \sum_{\sigma\in P(2m)} \mathbb{E}[X^2]\cdots\mathbb{E}[X^2]$$

假设 $(X_1, \cdots, X_{2n}) \sim N(0, \Sigma), \Sigma \geq 0$(非负定), 则

$$\mathbb{E}[X_1 X_2 \cdots X_{2n}] = \sum_{\sigma\in P(2m)} \prod_{i,j\in\sigma} \mathbb{E}[X_i X_j].$$

证明 令 $Y = \sum_{j=1}^{2n} \lambda_j X_j$, 则 $Y \sim N(0, \sigma^2)$

$$\sigma^2 = \sum_{i,j} \lambda_i \lambda_j \mathbb{E}[X_i X_j]$$

令 $g(\lambda)=\mathbb{E}[e^Y]$, 则

$$\mathrm{LHS}=\left(\frac{\partial^{2n}}{\partial\lambda_1\cdots\partial\lambda_{2n}}g(\lambda)\right)\bigg|_{\lambda=0}$$

另外

$$g(\lambda)=\sum_{k=0}^{\infty}\frac{\sigma^{2k}(2k-1)!!}{(2k)!}$$

只须看 $k=n$ 时,

$$\mathrm{LHS}=\frac{(2n-1)!!}{(2n)!}\frac{\partial^{2n}}{\partial\lambda_1\cdots\partial\lambda_{2n}}\left(\sum_{i,j}\lambda_i\lambda_j\mathbb{E}[X_iX_j]\right)^n=\sum_{\sigma\in P(2m)}\prod_{i,j\in\sigma}\mathbb{E}[X_iX_j]$$

$X\sim N(0,1),\mathbb{E}[X^{2n}]=(2n-1)!!$

Wick 乘积公式可以参考

https : //en.wikipedia.org/wiki/Isserlis%27_theorem

. 证明可在 http://www.math.utah.edu/ davar/math7880/F18/GaussianAnalysis.pdf 当中的一节里面找到.

1.16 中心极限定理 (CLT)

定理 1.21 (Lévy-Cramér 连续性定理)

F_n 为分布函数, $\phi_n(t)=\int e^{itx}\mathrm{d}F_n$

(1) 若 F 为分布函数, 且 $F_n\xrightarrow{w}F$, 则 $\phi_n(t)\to\phi(t),\forall t\in\mathbb{R}$(且内闭一致收敛, 这里暂且不证), 这里 $\phi(t)=\int e^{itx}\mathrm{d}F$

(2) 若 $\phi(t):=\lim_{n\to\infty}\phi_n(t)$ 存在, 且 $\phi(t)$ 在 $t=0$ 处连续, 则 ϕ 为某分布函数 F 的特征函数, 且 $F_n\xrightarrow{w}F$.

定义 1.10

称随机变量 X_n依分布收敛至 X, 弱

$$F_{X_n}\xrightarrow{w}F_X$$

记为 $X_n\xrightarrow{D}X$.

课堂上我们已经用特征函数的方法证明了 i.i.d 形式的 CLT

定理 1.22 (i.i.d CLT)

设 $\{X_k\}$ 独立同分布, 且 $\mu=\mathbb{E}[X_1],\sigma^2=\mathrm{Var}(X_1),\sigma>0$, 则

$$\frac{S_n-n\mu}{\sqrt{n\sigma^2}}\xrightarrow{D}N(0,1).$$

我们进一步考虑更一般的情形.

问题: 如何利用特征函数证明依分布收敛?

命题 1.2

设 $c_n \to c \in \mathbb{C}$, 则有

$$\lim_{n\to\infty}\left(1+\frac{c_n}{n}\right)^n = \mathrm{e}^c$$

证明 先看下面两个引理:

引理 1.13

设 $b \in \mathbb{C}, |b| \leqslant 1$, 则有 $|\mathrm{e}^b - b - 1| \leqslant |b|^2$.

证明 利用 $\mathrm{e}^b = \sum\limits_{k=0}^{+\infty}\frac{b^k}{k!}$, 有

$$|\mathrm{e}^b - b - 1| \leqslant \frac{|b|^2}{2}\left(1+\frac{|b|}{3}+\frac{|b|^2}{3\cdot 4}+\cdots\right) \leqslant \frac{|b|^2}{2}\left(1+\frac{1}{3}+\frac{1}{3\cdot 4}+\cdots\right) \leqslant |b|^2.$$

引理 1.14

设 $z_1,\cdots z_n, w_1,\cdots,w_n \in \mathbb{C}$ 且模长均不超过 r, 则

$$\left|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right| \leqslant r^{n-1}\sum_{k=1}^n |z_k - w_k|.$$

证明 我们有

$$\begin{aligned}
\left|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right| &= \left|\prod_{k=1}^n z_k - \prod_{k=1}^{n-1} z_k\cdot w_n + \prod_{k=1}^{n-1} z_k\cdot w_n - \prod_{k=1}^{n-2} z_k\cdot w_{n-1}w_n + \cdots + z_1\prod_{k=2}^n w_k - \prod_{k=1}^n w_k\right| \\
&\leqslant \left|\prod_{k=1}^n z_k - \prod_{k=1}^{n-1} z_k\cdot w_n\right| + \left|\prod_{k=1}^{n-1} z_k\cdot w_n - \prod_{k=1}^{n-2} z_k\cdot w_{n-1}w_n\right| + \cdots + \left|z_1\prod_{k=2}^n w_k - \prod_{k=1}^n w_k\right| \\
&= \left|\prod_{k=1}^{n-1} z_k\right||z_n - w_n| + \left|\prod_{k=1}^{n-2} z_k\cdot z_n\right||z_{n-1} - w_{n-1} + \cdots + \left|\prod_{k=2}^n w_k\right||z_1 - w_1| \\
&\leqslant r^{n-1}\sum_{k=1}^n |z_k - w_k|.
\end{aligned}$$

回到原命题. 设 $|c_n| \leqslant r, r > 0$. 注意到

$$\left|1+\frac{c_n}{n}\right| \leqslant 1+\frac{|c_n|}{n} \leqslant 1+\frac{r}{n} \leqslant \mathrm{e}^{\frac{r}{n}}, \quad |\mathrm{e}^{\frac{c_n}{n}}| \leqslant \mathrm{e}^{\frac{|c_n|}{n}} \leqslant \mathrm{e}^{\frac{r}{n}}$$

利用 $\lim\limits_{n\to+\infty}\mathrm{e}^{c_n} = \mathrm{e}^c$ 及上述引理可得

$$\left|\left(1+\frac{c_n}{n}\right)^n - \mathrm{e}^{c_n}\right| \leqslant (\mathrm{e}^{\frac{r}{n}})^{n-1}\cdot n\cdot\left|\mathrm{e}^{\frac{c_n}{n}} - 1 - \frac{c_n}{n}\right| \leqslant \mathrm{e}^r\cdot n\cdot\frac{|c_n|^2}{n^2} \leqslant \mathrm{e}^r\cdot\frac{r^2}{n} \xrightarrow{n\to+\infty} 0.$$

注 用特征函数证明依分布收敛, 对 i.i.d 和的情形, 想办法写成 $\left(1+\frac{c_n}{n}\right)^n$ 的形式.

例题 1.35 设 $X_1,\cdots X_n$ 独立同分布, X_1 对称且满足 $\mathbb{P}(|X_1|>x) = x^{-2}$, $x \geqslant 1$. 记 $S_n = \sum\limits_{k=1}^n X_k$, 则有

$$\frac{S_n}{\sqrt{n\log n}} \xrightarrow{D} N(0,1).$$

注 X_1 无二阶矩, 因为 $\mathbb{E}[X^2] = \int_1^{+\infty} x^{-1}\mathrm{d}x = +\infty$.

证明 利用特征函数, 有

$$\mathbb{E}[e^{it\frac{S_n}{\sqrt{n\log n}}}] = (\mathbb{E}[e^{it\frac{X_1}{\sqrt{n\log n}}}])^n = \left(\int_{|x|>1} e^{it\frac{X_1}{\sqrt{n\log n}}}|x|^{-3}dx\right)^n = \left(2\int_1^{+\infty}\cos\left(t\frac{X_1}{\sqrt{n\log n}}\right)x^{-3}dx\right)^n$$
$$= \left(1+2\int_1^{+\infty}\left(\cos\left(t\frac{X_1}{\sqrt{n\log n}}\right)-1\right)x^{-3}dx\right)^n$$

为转化 $\left(1+\frac{c_n}{n}\right)^n$ 的形式, 考虑 $c_n := 2n\int_1^{+\infty}\left(\cos\left(t\frac{X_1}{\sqrt{n\log n}}\right)-1\right)x^{-3}dx$, 令 $y=\frac{x}{\sqrt{n\log n}}$, 则

$$c_n = 2n\int_{\frac{1}{\sqrt{n\log n}}}^{+\infty}(n\log n)^{-1}(\cos ty-1)y^{-3}dy = \frac{1}{\log n}2\int_{\frac{1}{\sqrt{n\log n}}}^{+\infty}(\cos ty-1)y^{-3}dy.$$

把 n 转化成 x, 对 x 取极限, 并利用 L'Hospital 法则, 有

$$\begin{aligned}\lim_{x\to+\infty}\frac{2\int_{\frac{1}{\sqrt{x\log x}}}^{+\infty}(\cos ty-1)y^{-3}dy}{\log x} &= \lim_{x\to+\infty}\frac{2(x\log x)^{\frac{3}{2}}(\frac{1}{\sqrt{x\log x}})'(\cos\frac{t}{\sqrt{x\log x}}-1)}{\frac{1}{x}}\\ &= \lim_{x\to+\infty}x(1+\log x)(\cos\frac{t}{\sqrt{x\log x}}-1)\\ &= -\frac{t^2}{2}.\end{aligned}$$

因此 $\lim_{n\to+\infty}c_n = -\frac{t^2}{2}$. 由命题 1.2 可知

$$\lim_{n\to+\infty}\mathbb{E}[e^{it\frac{S_n}{\sqrt{n\log n}}}] = e^{-\frac{t^2}{2}} \implies \frac{S_n}{\sqrt{n\log n}} \xrightarrow{D} N(0,1).$$

课堂上亦给出了更一般形式的 CLT:

Linderberg 条件: 对 $X_1,\cdots,X_n$, 记 $a_k=\mathbb{E}[X_k], b_k^2=\mathrm{Var}[X_k], B_n^2=\sum_{k=1}^n b_k^2, F_k$ 为 X_k 分布函数.

$$\forall\epsilon>0,\ \lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^n\int_{|x-a_k|>\epsilon B_n}(x-a_k)^2dF_k=0 \tag{L}$$

定理 1.23 (Linderberg-Feller CLT)

设 $X_1,\cdots,X_n$ 相互独立, 满足(L), 则

$$\frac{1}{B_n}\sum_{k=1}^n(X_k-a_k)\to N(0,1)$$

注 事实上, 我们可以把上述定理推广到随机变量组列 (双下标) 中, 结论依然成立 (具体参考 **Durrett Theorem 3.4.10**):

定理 1.24

对每个 $n\in\mathbb{N}^*$, 随机变量 $X_{n,m}, 1\leqslant m\leqslant n$ 相互独立, 且满足

(1) $\mathbb{E}[X_{n,m}]=0, \forall m,n.$

(2) $\sum_{m=1}^n X_{n,m}^2\to\sigma^2>0.$

(3) 对 $\forall\varepsilon>0$, 均有 $\lim\limits_{n\to+\infty}\sum\limits_{m=1}^{n}\mathbb{E}[X_{n,m}^2 I_{\{|X_{n,m}|\geqslant\varepsilon\}}]=0.$

则当 $n\to+\infty$ 时, 有

$$\sum_{m=1}^{n}X_{n,m}\xrightarrow{D}N(0,1).$$

注 对于 Linderberg-Feller CLT 一般有两种证明方法. 一种方法是从弱收敛的角度通过特征函数证明逐点收敛, 证明可以参考 **Durrett Theorem 3.4.10**; 另一种方法利用依分布收敛的各种等价刻画的性质, 运用 Lindeberg 替换术解决, 更加 technical, 证明见本讲义第 3 部分阅读材料.

下面看一个相关的例子:

例题 1.36 设 $\{X_n\}$ 为相互独立的随机变量列, 记 $S_n=\sum\limits_{k=1}^{n}X_k$, 在以下两个条件下试选择合适的数列 $\{a_n\},\{B_n\}$ 并证明

$$\frac{S_n-a_n}{B_n}\xrightarrow{\mathrm{D}}N(0,1).$$

(1) $\mathbb{P}(X_k=\sqrt{k})=\mathbb{P}(X_k=-\sqrt{k})=\dfrac{1}{2}.$

(2) $\mathbb{P}(X_k=1)=\dfrac{1}{k},\mathbb{P}(X_k=0)=1-\dfrac{1}{k}.$

证明 (1) $\mathbb{E}[X_k]=0,\mathrm{Var}(X_k)=k$. 可推测

$$a_n=0,\quad B_n=\sqrt{\sum_{k=1}^{n}\mathrm{Var}(X_k)}=\sqrt{\frac{n(n+1)}{2}}.$$

下面验证其满足 Linderberg 条件. 对 $\forall\varepsilon>0$, 取正整数 N, 使得 $N>\dfrac{2}{\varepsilon^2}$. 则当 $n>N$ 时, 均有

$$\frac{\varepsilon B_n}{\sqrt{n}}=\varepsilon\sqrt{\frac{n+1}{2}}>\varepsilon\sqrt{\frac{N}{2}}=1.$$

因此

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x|>\epsilon B_n}x^2\mathrm{d}F_k=0.$$

(2) $\mathbb{E}[X_k]=\dfrac{1}{k},\mathrm{Var}(X_k)=\dfrac{1}{k}-\dfrac{1}{k^2}$. 可推测

$$a_n=\sum_{k=1}^{n}\frac{1}{k},\quad B_n=\sqrt{\sum_{k=1}^{n}\mathrm{Var}(X_k)}=\sqrt{\sum_{k=1}^{n}\frac{1}{k}-\sum_{k=1}^{n}\frac{1}{k^2}}.$$

下面验证其满足 Linderberg 条件. 记 $C_n=\sum\limits_{k=1}^{n}\dfrac{1}{k^2}$, 则 $1\leqslant C_n\leqslant\dfrac{\pi^2}{6}$, 因此 B_n 递增至 $+\infty$. 因为 $|0-\mathbb{E}[X_k]|,|1-\mathbb{E}[X_k]|\leqslant 1$, 故对 $\forall\varepsilon>0,\exists N\in\mathbb{N}^*$, 使得 $B_N>\dfrac{1}{\varepsilon}$. 则当 $n>N$ 时, 均有

$$\varepsilon B_n>\varepsilon B_N>1.$$

因此

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x|>\epsilon B_n}x^2\mathrm{d}F_k=0.$$

注

1. 由 Slutsky's theorem 可知 $\{a_n\},\{B_n\}$ 不一定唯一. 如 (2) 中取 $a_n = \log n, B_n = \sqrt{\log n}$ 上述结论依然成立.
2. 这类题目也可以通过验证 Lyapunov 条件来解决 (并非万能方法).

✍ **练习 1.25** **Grimmett 5.10.3, 5.12.32, 5.12.34, 5.12.42, 5.12.50, 5.12.52** .

✍ **练习 1.26** 设随机变量列 $\{X_k\}$ 相互独立且服从 $[0,1]$ 上的均匀分布, 令 $Y_n = (\prod_{i=1}^{n} X_i)^{-\frac{1}{n}}$, 证明:

$$\sqrt{n}(Y_n - \mathrm{e}) \xrightarrow{D} N(0, \mathrm{e}^2).$$

✍ **练习 1.27** 设随机变量 X_n 服从参数为正整数 n 的泊松分布, 试选择合适的数列 $\{a_n\},\{b_n\}$ 并证明

$$\frac{X_n - a_n}{b_n} \xrightarrow{\mathrm{D}} N(0,1).$$

✍ **练习 1.28** 设 $\{X_k\}$ 相互独立且服从指数分布, 其中 $\mathbb{E}[X_k] = \mu_k$. 证明: 若

$$\lim_{n\to+\infty} \frac{\max\{\mu_1,\cdots,\mu_n\}}{\sqrt{\sum_{k=1}^{n}\mu_k^2}} = 0,$$

则

$$\frac{1}{\sqrt{\sum_{k=1}^{n}\mu_k^2}} \sum_{k=1}^{n}(X_k - \mu_k) \xrightarrow{D} N(0,1).$$

1.17 随机变量列的收敛与极限定理

1.17.1 基本工具与技术

命题 1.3 (矩不等式)

我们有

(1) **Hölder 不等式**: $p, q > 1, \dfrac{1}{p} + \dfrac{1}{q} = 1$,

$$\mathbb{E}[|XY|] \le (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}$$

(2) **Minkowski 不等式**:

$$(\mathbb{E}[|X+Y|^p])^{\frac{1}{p}} \le (\mathbb{E}[|X|^p])^{\frac{1}{p}} + (\mathbb{E}[|Y|^p])^{\frac{1}{p}}$$

(3) **Markov 不等式**:

$$\mathbb{P}(|X| \ge a) \le \frac{\mathbb{E}[|X|]}{a}, a > 0$$

(4) **Chebyshev 不等式**:

$$\mathbb{P}(|X - \mathbb{E}[X]| \ge a) \le \frac{1}{a^2}\mathrm{Var}(X)$$

(5) 设 $s > 0$, 若 $\mathbb{E}[|X|^s] < +\infty$, 则对 $r \in [0,s]$, 均有 $\mathbb{E}[|X|^r] < +\infty$. 因此 $X_n \xrightarrow{s} X \implies X_n \xrightarrow{r} X$.

(6) C_r 不等式:

$$\mathbb{E}|X_1 + \cdots + X_n|^r \leqslant C_{n,r}(\mathbb{E}[|X_1|^r] + \cdots \mathbb{E}[|X_n|^r]).$$

其中

$$C_{n,r}=\begin{cases}1, & 0<r<1,\\ n^{r-1}, & r\geqslant 1.\end{cases}$$

(7) **Lyapunov 不等式**: 对 $\forall 0<r<s$, 有

$$(\mathbb{E}[|X|^r])^{\frac{1}{r}}\leqslant(\mathbb{E}[|X|^s])^{\frac{1}{s}}.$$

证明 (5) 注意到

$$|X|^r=|X|^rI_{\{|X|\leqslant 1\}}+|X|^rI_{\{|X|>1\}}\leqslant 1+|X|^s,$$

两边取期望即得.

(6) 当 $0<r\leqslant 1$ 时, 利用 x^r 的凹性, 有

$$|X_1+\cdots+X_n|^r\leqslant|X_1|^r+\cdots+|X_n|^r.$$

当 $r\geqslant 1$ 时, 利用 x^r 的凸性, 有

$$\left(\frac{|X_1|+\cdots+|X_n|}{n}\right)^r\leqslant\frac{|X_1|^r+\cdots+|X_n|^r}{n}$$

再利用 $|X_1+\cdots+X_n|\leqslant|X_1|+\cdots+|X_n|$ 即可. 综上, 有 $|X_1+\cdots+X_n|^r\leqslant C_{n,r}(|X_1|^r+\cdots+|X_n|^r)$. 最后对两边同时取期望即得.

(7) 利用 Hölder 不等式, 并取 $Y=1, p=\dfrac{s}{r}$, 用 $|X|^r$ 代替 $|X|$ 即可.

注 Marvkov 不等式 (及其推广) 是在尾概率估计中一个重要的不等式, 为随机变量偏离某些值的概率给出了上界. 更一般地, 设 g 是 $\mathbb{R}$ 上的非负 Borel 可测函数, 我们有

(1) 若 g 为偶函数, 且在 $\mathbb{R}^+$ 上单调递增 (指不严格, 下同), 则对 $\forall a\geqslant 0$, 有

$$\mathbb{P}(|X|\geqslant a)=\mathbb{P}(g(|X|)\geqslant g(a))\leqslant\frac{\mathbb{E}[g(|X|)]}{g(a)}.$$

(2) 若 g 在 $\mathbb{R}$ 上单调递增, 则对 $\forall a\geqslant 0$, 有

$$\mathbb{P}(X\geqslant a)=\mathbb{P}(g(X)\geqslant g(a))\leqslant\frac{\mathbb{E}[g(X)]}{g(a)}.$$

下面两个简单的特例非常值得关注, 在尾概率估计中非常常见. 对 $a>0$, 有

$$\mathbb{P}(|X|\geqslant a)\leqslant\frac{\mathbb{E}[|X|^r]}{a^r}, r>0,\qquad \mathbb{P}(X>a)\leqslant \mathrm{e}^{-ta}\mathbb{E}[\mathrm{e}^{tX}].$$

对于前者, 越高阶矩得到有关矩的信息越多, 也在证明随机变量的收敛时放缩更易施展手脚, 减少限制. 后者估计便于利用矩母函数的性质, 由此导出一些集中不等式 (超出本课程范围).

例题 1.37 (单边 Chebyshev 不等式) 设 $\mathrm{Var}(X)=\sigma^2\in(0,+\infty)$, 则对 $\forall\lambda>0$, 有

$$\mathbb{P}(X-\mathbb{E}[X]\geqslant\lambda)\leqslant\frac{\sigma^2}{\sigma^2+\lambda^2}.$$

证明 记 $Y=X-\mathbb{E}[X]$, 设 $u>0$, 则有

$$\mathbb{P}(X-\mathbb{E}[X]\geqslant\lambda)=\mathbb{P}(Y+u\geqslant\lambda+u)\leqslant\mathbb{P}((Y+u)^2\geqslant(\lambda+u)^2)\leqslant\frac{\mathbb{E}[(Y+u)^2]}{(\lambda+u)^2}=\frac{\sigma^2+u^2}{(\lambda+u)^2}\xlongequal{u=\frac{\sigma^2}{\lambda}}\frac{\sigma^2}{\sigma^2+\lambda^2}.$$

其中最后一个不等式是一个简单的函数最值问题, 最小值当 $u=\dfrac{\sigma^2}{\lambda}$ 时取等.

注 对称地, 有

$$\mathbb{P}(X-\mathbb{E}[X]\leqslant-\lambda)\leqslant\frac{\sigma^2}{\sigma^2+\lambda^2}.$$

例题 1.38 设 $X_1,\cdots X_n$ 独立同分布, $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\dfrac{1}{2}$, 记 $S_n=\sum\limits_{k=1}^{n}X_i$, 证明: 对 $\forall a>0$, 均有

$$\mathbb{P}(S_n>a)\leqslant \mathrm{e}^{-\frac{a^2}{2n}}.$$

证明 我们有

$$\mathbb{P}(S_n>a)\leqslant \mathrm{e}^{-ta}\mathbb{E}[\mathrm{e}^{tS_n}]=\mathrm{e}^{-ta}(\mathbb{E}[\mathrm{e}^{tX_1}])^n=\mathrm{e}^{-ta}\left(\frac{\mathrm{e}^t+\mathrm{e}^{-t}}{2}\right)^n\leqslant \mathrm{e}^{-ta+\frac{n}{2}t^2}=\mathrm{e}^{\frac{n}{2}(t-\frac{a}{n})^2-\frac{a^2}{2n}}\xlongequal{t=\frac{a}{n}}\mathrm{e}^{-\frac{a^2}{2n}}.$$

其中

$$\frac{\mathrm{e}^t+\mathrm{e}^{-t}}{2}=\frac{1}{2}\left(\sum_{k=0}^{+\infty}\frac{t^k}{k!}+\sum_{k=0}^{+\infty}(-1)^k\frac{t^k}{k!}\right)=\sum_{k=0}^{+\infty}\frac{t^{2k}}{(2k)!}\leqslant\sum_{k=0}^{+\infty}\frac{t^{2k}}{2^k k!}=\mathrm{e}^{\frac{t^2}{2}}.$$

注 $\dfrac{\mathrm{e}^t+\mathrm{e}^{-t}}{2}\leqslant \mathrm{e}^{\frac{t^2}{2}}$ 是常用的不等式放缩, 值得一记. 类似地还有 $\dfrac{\mathrm{e}^t-\mathrm{e}^{-t}}{2t}\leqslant \mathrm{e}^{\frac{t^2}{2}}$.

下面是概率测度下证明随机变量列收敛及一些放缩中常用的拆项小技巧, 请读者自行验证:

命题 1.4

我们有

$$\mathbb{P}(|X+Y|\leqslant a+b)\geqslant \mathbb{P}(|X|\leqslant a,|Y|\leqslant b),\quad \mathbb{P}(|X+Y|\geqslant a+b)\leqslant \mathbb{P}(|X|\geqslant a)+\mathbb{P}(|Y|\geqslant b).$$

注 独立复制的副产品: 设 X' 是 X 的独立复制 (X,X' i.i.d), 对 $\forall x,a$, 均有

$$\mathbb{P}(|X-X'|>x)=\mathbb{P}(|(X-a)-(X'-a)|>x)\leqslant \mathbb{P}(|X-a|\geqslant\frac{x}{2})+\mathbb{P}(|X'-a|\geqslant\frac{x}{2})=2\mathbb{P}(|X-a|\geqslant\frac{x}{2}).$$

回顾一下 Borel-Cantelli 引理:

定理 1.25 (Borel-Cantelli 引理)

记 $\{A_n,\ \text{i.o.}\}=\limsup\limits_{n\to\infty}A_n=\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{m=n}^{\infty}A_m$(infinitely often)

(1) 当 $\sum\limits_{n=1}^{\infty}\mathbb{P}(A_n)<\infty$ 时有, $\mathbb{P}(A_n,\ \text{i.o.})=0$

(2) 假设 $\{A_n\}$ 相互独立, 当 $\sum\limits_{n=1}^{\infty}\mathbb{P}(A_n)=\infty$ 时有 $\mathbb{P}(A_n,\ \text{i.o.})=1$

随机变量的截尾术

对于随机变量 X, 主要有以下三类结尾方法:

$$X_1=XI_{\{|X|\leqslant M\}}=\begin{cases}X, & |X|\leqslant M,\\ 0, & |X|>M.\end{cases},\quad X_2=\begin{cases}X, & |X|\leqslant M,\\ M, & X>M,\\ -M, & X<-M.\end{cases},\quad X_3=\begin{cases}X, & X\leqslant M,\\ M, & X>M.\end{cases}.$$

注 截尾法常常用于证明有关随机变量列的收敛中, 运用了转化的思想. 对于随机变量列 $\{X_n\}$, 我们常常令 $M\uparrow+\infty$ 或者取 $M=n$ (更一般取 $M=k_n$, $\{k_n\}$ 递增) 再令 $n\to+\infty$ 来处理. 这样截尾的好处是截尾后的随机变量具有有界性 (可能还利用了随机变量列的单调性). 一方面, 有界性的保证大大可以施展手脚 (比如很多定理、命题的适用条件都有有界性的要求, 以及随机变量的有界性能保证其任意阶矩的存在性, 从而便于矩不等式的施展); 另一方面, 单调性或是尾巴 $X-X_i$ 部分保证了目标的转化条件.

我们下举一例有关截尾的例子, 这也是强大数律的证明中一个重要的转化步骤. 后续我们还会看到

一些用到截尾的例子, 可见其精妙之处.

例题 1.39 设非负随机变量 X_i 独立同分布且 $\mathbb{E}[X_1]<\infty$, 令 $Y_n=X_nI_{\{|X_n|\leqslant n\}}$, 则有

$$\frac{1}{n}\sum_{k=1}^{n}(X_k-Y_k)\xrightarrow{\text{a.s.}}0.$$

从而

$$\frac{1}{n}\sum_{k=1}^{n}X_k\xrightarrow{\text{a.s.}}a\Longleftrightarrow\frac{1}{n}\sum_{k=1}^{n}Y_k\xrightarrow{\text{a.s.}}a$$

证明 对 $X\geqslant 0$, 利用 (见引理 **1.12**)

$$\sum_{m=1}^{\infty}\mathbb{P}(X\geq m)\leq\mathbb{E}[X]\leq 1+\sum_{m=1}^{\infty}\mathbb{P}(X\geq m),$$

有

$$\sum_{n=1}^{+\infty}\mathbb{P}(Y_n\neq X_n)=\sum_{n=1}^{+\infty}\mathbb{P}(X_n\geq n)=\sum_{n=1}^{+\infty}\mathbb{P}(X_1\geq n)\leq\mathbb{E}[X_1]<\infty$$

利用 Borel-Cantelli 引理, $\mathbb{P}(X_n\neq Y_n\ \text{i.o.})=0$, 即几乎处处 $\{X_n\neq Y_n\}$ 只发生有限次, 从而

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-Y_i)\xrightarrow{\text{a.s.}}0.$$

注 通过截尾, 证明原随机变量列的收敛转化成证明截尾后的随机变量列的收敛.

补充: 独立复制与对称化方法

这个方法在概率中是非常重要的技术, 考虑到内容上限就不作过多补充, 之后的一些概率类课程还会专门用到这个方法, 现在只以两个例子给大家自行体会.

例题 1.40 设 $f,g:\mathbb{R}\to\mathbb{R}$ 且单调递增, 证明, 对任意随机变量 X, 均有

$$\mathbb{E}[f(X)g(X)]\geqslant\mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

证明 设 X' 是 X 的独立复制, 则有

$$(f(X)-f(X'))(g(X)-g(X'))\geqslant 0.$$

因此

$$0\leqslant\mathbb{E}[(f(X)-f(X'))(g(X)-g(X'))]=2\mathbb{E}[f(X)g(X)]-2\mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

回忆一下, 我们称随机变量 X 是对称的, 若 $X\overset{d}{=\!=}-X$.

例题 1.41 设随机变量 $e_1,\cdots e_n$ 独立同分布, 且 $\mathbb{E}[e_i]=0$, 证明: 对任意 $h_1,\cdots h_n\in\mathbb{R}$, 均有

$$\mathbb{E}\left|\sum_{i=1}^{n}h_ie_i\right|\leqslant 2\mathbb{E}\left|\sum_{i=1}^{n}|h_i|e_i\right|.$$

证明 若 X_i 对称, 则

$$h_iX_i\overset{d}{=\!=}|h_i|X_i\implies(h_1X_1,\cdots,h_nX_n)\overset{d}{=\!=}(|h_1|X_1,\cdots,|h_n|X_n)\implies\sum_{i=1}^{n}h_iX_i\overset{d}{=\!=}\sum_{i=1}^{n}|h_i|X_i.$$

设 $\{e_i'\}$ 是 $\{e_i\}$ 的一个独立复制, 则 $e_i - e_i'$ 对称. 故有

$$\begin{aligned}
\mathbb{E}\left|\sum_{i=1}^n h_i e_i\right| &= \mathbb{E}\left|\sum_{i=1}^n h_i(e_i - \mathbb{E}[e_i'|e_1,\cdots,e_n])\right| = \mathbb{E}\left|\mathbb{E}(\sum_{i=1}^n h_i[(e_i - e_i')|e_1,\cdots,e_n])\right| \leqslant \mathbb{E}\left|\sum_{i=1}^n h_i(e_i - e_i')\right| \\
&= \mathbb{E}\left|\sum_{i=1}^n h_i(e_i - e_i')\right| \leqslant \mathbb{E}\left|\sum_{i=1}^n |h_i|e_i\right| + \mathbb{E}\left|\sum_{i=1}^n |h_i|e_i'\right| \\
&= 2\mathbb{E}\left|\sum_{i=1}^n |h_i|e_i\right|.
\end{aligned}$$

1.17.2 浅谈四种收敛

回顾随机变量列四种收敛的定义:

定义 1.11

设 $X, X_1, X_2, \cdots, X_n$ 为 $(\Omega, \mathcal{F}, \mathbb{P})$ 上随机变量.

(1) 几乎处处 (以概率 **1** 收敛) :

$$\mathbb{P}\left(\{\omega \in \Omega : X_n(\omega) \to X(\omega), n \to \infty\}\right) = 1$$

记为 $X_n \xrightarrow{\text{a.s.}} X$.

(2) **r** 阶收敛 $(r > 0)$:$\mathbb{E}[|X_n|^r] < \infty, \forall n$ 且

$$\mathbb{E}[|X_n - X|^r] \to 0, n \to \infty$$

记为 $X_n \xrightarrow{r} X$, 特别 $r = 1$ 时为 平均收敛, $r = 2$ 时为 均方收敛.

(3) 依概率收敛:

$$\forall \epsilon > 0, P(|X_n - X| > \epsilon) \to 0, n \to \infty$$

记为 $X_n \xrightarrow{P} X$.

(4) 依分布收敛:

$$F_{X_n} \xrightarrow{w} F_X$$

记为 $X_n \xrightarrow{D} X$.

♣

注 只有依分布收敛会 “忘记” 样本空间 (因为分布函数会 “忘记” 样本空间), 其他类型的收敛都要考虑样本空间. 因而 $X_n \to X, Y_n \to Y \nRightarrow \mathbb{P}(X = Y) = 1$ 及 $X_n + Y_n \xrightarrow{D} X + Y$.

证明依概率收敛:

(1) 估计 $\mathbb{P}(|X_n - X| > \varepsilon)$, 可能要利用**命题**1.1中的矩不等式或其他方法.

(2) 弱大数律 (注意条件是 **i.i.d** 且一阶矩存在).

(3) 若收敛极限是常数 c, 利用 $X_n \xrightarrow{D} c \Longleftrightarrow X_n \xrightarrow{P} c$, 转化成证明依分布收敛.

⋮

⋮

例题 1.42 有一列零均值随机变量 $X_1, \cdots X_n$, $\exists c > 0$, 使得对 $\forall n \in \mathbb{N}^*$, $\mathrm{Var}(X_n) < c$. 记 $S_n = \sum_{k=1}^n X_k$, 若

$$\mathrm{Cov}(X_i, X_j) \to 0 \quad as\ |i - j| \to +\infty,$$

证明: $\dfrac{S_n}{n} \xrightarrow{P} 0$.

证明 对 $\forall\varepsilon>0$, 由 Chebyshev 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right|>\varepsilon\right)\leqslant\frac{\mathbb{E}[S_n^2]}{\varepsilon^2n^2}=\frac{1}{\varepsilon^2n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k)+\frac{2}{\varepsilon^2n^2}\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i,X_j)\leqslant\frac{c}{\varepsilon^2n}+\frac{2}{\varepsilon^2n^2}\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i,X_j).$$

对 $\forall\delta>0$, 由题意知, $\exists N\in\mathbb{N}^*$, 使得 $|\mathrm{Cov}(X_i,X_j)|\leqslant\delta,|i-j|>N$. 又因为 $\mathrm{Cov}(X_i,X_j)<c$, 因此

$$\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i,X_j)=\sum_{\substack{1\leqslant i<j\leqslant n\\|i-j|\leqslant N}}\mathrm{Cov}(X_i,X_j)+\sum_{\substack{1\leqslant i<j\leqslant n\\|i-j|>N}}\mathrm{Cov}(X_i,X_j)\leqslant nNc+n^2\delta.$$

故有

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right|>\varepsilon\right)\leqslant\frac{c}{\varepsilon^2n}+\frac{2Nc}{\varepsilon^2n}+2\frac{\delta}{\varepsilon^2}.$$

令 $n\to+\infty$, 再令 $\delta\to0$, 则有 $\mathbb{P}\left(\left|\frac{S_n}{n}\right|>\varepsilon\right)=0$, 即 $\frac{S_n}{n}\xrightarrow{\mathrm{P}}0$.

证明几乎处处收敛:

(1) 利用定义转化: 存在 $\Omega_0\subset\Omega$, 满足 $\mathbb{P}(\Omega_0)=1$, 且对 $\forall\omega\in\Omega_0$, 有

$$X_n(\omega)\xrightarrow{n\to+\infty}X(\omega),\ n\to+\infty.$$

固定了 ω 后 $X_n(\omega),X(\omega)$ 变为常数, 于是转化成常数列极限问题: $a_n\to a$. 例如:

- $X_n\xrightarrow{\text{a.s.}}X,Y_n\xrightarrow{\text{a.s.}}Y\implies X_n+Y_n\xrightarrow{\text{a.s.}}X+Y,X_nY_n\xrightarrow{\text{a.s.}}XY$.
- $X_n\xrightarrow{\text{a.s.}}X\implies\frac{1}{n}\sum_{k=1}^{n}X_n\xrightarrow{\text{a.s.}}X$. (思考: 换成依概率收敛是否成立?)
- **Cauchy 列**: $X_n\xrightarrow{\text{a.s.}}X\iff\sup_{k\geqslant n}|X_k-X|\xrightarrow{\text{a.s.}}0\iff\sup_{k\geqslant n}|X_k-X_n|\xrightarrow{\text{a.s.}}0\iff$ $\sup_{k,m\geqslant n}|X_k-X_m|\xrightarrow{\text{a.s.}}0$.
- $\frac{X_n}{n}\xrightarrow{\text{a.s.}}0\iff\frac{\max_{1\leqslant k\leqslant n}|X_k|}{n}\xrightarrow{\text{a.s.}}0$.

(2) 运用 Borel-Cantelli 引理解决 (可能会结合**子序列**, **截尾**, 独立复制与对称化等方法):

$$\forall\varepsilon>0,\sum_{n=1}^{+\infty}\mathbb{P}(|X_n-X|>\varepsilon)<\infty\implies\forall\varepsilon>0,\mathbb{P}(|X_n-X|>\varepsilon,\text{i.o.})=0\iff X_n\xrightarrow{\text{a.s.}}X.$$

若 $\{X_n-X\}$ 相互独立, 则为等价关系. 最后进行 $\mathbb{P}(|X_n-X|>\varepsilon)$ 的估计, 可能要借助各种矩不等式或其他方法解决, 用 Markov 不等式时对矩选取合适的阶数, 例子见作业题 **Grimmett7.11.6**.

(3) 强大数律 (注意条件是 i.i.d 且一阶矩存在), 例子见作业题 **Grimmett7.5.1**.

(4) * 见后, 例如转化成证明有关子列依概率收敛问题.

⋮

⋮

例题 1.43 设 $X_1,\cdots X_n$ 独立同分布, 证明:

$$\frac{X_n}{n}\xrightarrow{\text{a.s.}}0\iff\mathbb{E}|X_1|<\infty.$$

证明 我们有

$$\frac{X_n}{n}\xrightarrow{\text{a.s.}}0\iff\forall\varepsilon>0,\mathbb{P}\left(\frac{|X_n|}{n}>\varepsilon,\text{i.o.}\right)=0\iff\forall\varepsilon>0,\sum_{n=1}^{+\infty}\mathbb{P}\left(\frac{|X_n|}{\varepsilon}>n\right)<\infty.$$

利用

$$\sum_{m=1}^{\infty}\mathbb{P}(|X|\geq m)\leq\mathbb{E}[|X|]\leq1+\sum_{m=1}^{\infty}\mathbb{P}(|X|\geq m)=\sum_{m=0}^{\infty}\mathbb{P}(|X|\geq m),$$

有

$$\frac{\mathbb{E}\left[|X_n|\right]}{\varepsilon}-1\leqslant\sum_{n=1}^{+\infty}\mathbb{P}\left(\frac{|X_n|}{\varepsilon}>n\right)\leqslant\frac{\mathbb{E}\left[|X_n|\right]}{\varepsilon}.$$

因此

$$\frac{X_n}{n}\xrightarrow{\text{a.s.}}0\Longleftrightarrow\forall\varepsilon>0,\sum_{n=1}^{+\infty}\mathbb{P}\left(\frac{|X_n|}{\varepsilon}>n\right)<\infty\Longleftrightarrow\mathbb{E}|X_1|<\infty.$$

注意到

$$X_n\xrightarrow{\text{a.s.}}X\Longleftrightarrow\forall\varepsilon>0,\mathbb{P}(|X_n-X|>\varepsilon,\text{i.o.})=0\Longleftrightarrow\mathbb{P}\left(\lim_{n\to\infty}\bigcup_{k=n}^{+\infty}\{|X_k-X|>\varepsilon\}\right)=0$$

$$\Longleftrightarrow\lim_{n\to\infty}\mathbb{P}\left(\bigcup_{k=n}^{+\infty}\{|X_k-X|>\varepsilon\}\right)=0\Longleftrightarrow\lim_{n\to\infty}\mathbb{P}\left(\sup_{k\geqslant n}|X_k-X|>\varepsilon\right)=0.$$

由此推得如下命题:

命题 1.5

我们有

(1) $X_n\xrightarrow{\text{a.s.}}X\Longleftrightarrow\sup\limits_{k\geqslant n}|X_k-X|\xrightarrow{\text{P}}0$. 特别地, 若 $|X_n|\downarrow 0$, 则

$$X_n\xrightarrow{\text{a.s.}}0\Longleftrightarrow\sup_{k\geqslant n}|X_k|\xrightarrow{\text{P}}0\Longleftrightarrow X_n\xrightarrow{\text{P}}0.$$

(2) 利用 Cauchy 列等价转换,

$$\begin{aligned}X_n\xrightarrow{\text{a.s.}}X&\Longleftrightarrow\sup_{k\geqslant n}|X_k-X|\xrightarrow{\text{P}}0\\&\Longleftrightarrow\sup_{k\geqslant n}|X_k-X_n|\xrightarrow{\text{a.s.}}0\quad\Longleftrightarrow\sup_{k\geqslant n}|X_k-X_n|\xrightarrow{\text{P}}0\\&\Longleftrightarrow\sup_{k,m\geqslant n}|X_k-X_m|\xrightarrow{\text{a.s.}}0\Longleftrightarrow\sup_{k,m\geqslant n}|X_k-X_m|\xrightarrow{\text{P}}0\end{aligned}$$

问题: 对于一般情形, 能否把几乎处处收敛转化成依概率收敛问题?

引理 1.15

我们有

$$X_n\xrightarrow{\text{P}}X\Longleftrightarrow\forall\varepsilon>0,\sup_{k\geqslant n}\mathbb{P}(|X_k-X_n|>\varepsilon)\xrightarrow{n\to+\infty}0,$$

注意: 区分 $\sup\limits_{k\geqslant n}\mathbb{P}(|X_k-X_n|>\varepsilon)$&$\mathbb{P}(\sup\limits_{k\geqslant n}|X_k-X_n|>\varepsilon)$, 两者不一样!

证明 $\Longrightarrow$: 对 $\forall\varepsilon>0$, 由 $X_n\xrightarrow{\text{P}}X$ 知, 对 $\forall\delta>0,\exists n_0\in\mathbb{N}^*$, 当 $n\geqslant n_0$ 时, 有 $\mathbb{P}(|X_n-X|>\frac{\varepsilon}{2})<\frac{\delta}{2}$. 故当 $k\geqslant n\geqslant n_0$ 时,

$$\mathbb{P}(|X_k-X_n|>\varepsilon)\leqslant\mathbb{P}(|X_k-X-(X_n-X)|>\varepsilon)\leqslant\mathbb{P}\left(|X_k-X|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(|X_n-X|>\frac{\varepsilon}{2}\right)<\delta.$$

$\Longleftarrow$: $\forall k\in\mathbb{N}^*,\exists n_k\uparrow+\infty$, 使得

$$\mathbb{P}(|X_{n_{k+1}}-X_{n_k}|>2^{-k})<2^{-k}\quad\forall k\in\mathbb{N}^*$$

所以

$$\sum_{k=1}^{+\infty}\mathbb{P}(|X_{n_{k+1}}-X_{n_k}|>2^{-k})<1 \implies \mathbb{P}(|X_{n_{k+1}}-X_{n_k}|>2^{-k},\text{i.o.})=0$$

记 $\Omega_0=\left\{|X_{n_{k+1}}-X_{n_k}|>2^{-k},\text{i.o.}\right\}$, 则当 $\omega\notin\Omega_0$ 时,

$$\sum_{k=1}^{+\infty}|X_{n_{k+1}}(\omega)-X_{n_k}(\omega)|<\infty$$

因此 $X_{n_k}(\omega)$ 为 Cauchy 列. 故存在极限 $X(\omega)$, 满足 $X_{n_k}(\omega)\to X(\omega)$. 所以

$$X_{n_k}\xrightarrow{\text{a.s.}}X.$$

因此当 $n>n_k$ 时, 利用

$$\mathbb{P}(|X_n-X|>\varepsilon)\leqslant\mathbb{P}\left(|X_{n_k}-X|>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(|X_n-X_{n_k}|>\frac{\varepsilon}{2}\right)$$

可知

$$X_n\xrightarrow{\text{P}}X.$$

命题 1.6

我们有

(1) $X_n\xrightarrow{\text{P}}X \implies$ 存在子列 $\{n_k\}$, $X_{n_k}\xrightarrow{\text{a.s.}}X$.

(2) $X_n\xrightarrow{\text{P}}X \iff$ 对任意子列 $\{n_k\}$, $\exists\left\{n_k'\right\}\subset\{n_k\}$, $X_{n_k'}\xrightarrow{\text{a.s.}}X$.

证明 (1) $\implies$: 参考引理 1.15的证明;

(2) $\implies$: 利用 (1) 及

$$a_n\to a\iff\forall\{n_k\},\exists\left\{n_k'\right\}\subset\{n_k\},a_{n_k'}\to a.$$

$\impliedby$: 反证. 若 $X_n\xrightarrow{\text{P}}X$ 不成立, 则 $\exists\varepsilon>0,\delta>0,\limsup\limits_{n\to\infty}\mathbb{P}(|X_n-X|>\varepsilon)>\delta$, 因此

$$\exists\{n_k\},\mathbb{P}(|X_{n_k}-X|>\varepsilon)>\delta \implies \forall\left\{n_k'\right\}\subset\{n_k\},\mathbb{P}(|X_{n_k'}-X|>\varepsilon)>\delta.$$

矛盾.

注 因此, 随机变量列的几乎处处收敛给出了另一个关于子列依概率收敛的刻画条件.

证明 r 阶收敛:

(1) 利用**命题** 1.3中的矩不等式, 如 **Hölder 不等式**、**Minkowski 不等式**、**C_r 不等式**等. 看一看笔记和作业题中的几个例子.

$$X_n\xrightarrow{\text{r}}X,Y_n\xrightarrow{\text{r}}Y \implies X_n+Y_n\xrightarrow{\text{r}}X+Y,X_nY_n\xrightarrow{\text{r}}XY.$$

(2) 借助矩不等式和截尾等技术, 运用单调收敛定理、控制收敛定理、Fatou 引理等证明.

⋮

⋮

已学过的定理: 若 $\exists k>0$, 使得 $\mathbb{P}(|X_n|\leqslant k)=1,\forall n\in\mathbb{N}^*$, 则

$$X_n\xrightarrow{\text{P}}X \implies X_n\xrightarrow{\text{r}}X.$$

更一般地情形留给读者自行验证: 若 $|X_n|\leqslant Y$, Y 绝对可积且 $X_n\xrightarrow{\text{P}}X$, 则 $X_n\xrightarrow{1}X$.

注 r 阶收敛不是本门课重点, 加之多数人没有系统学习实变的内容, 以及没有引入一致可积等概念, 因此对这门课而言一般来说不会专门在这方面出过难的题目, 后续高等概率论课程还会深入讲解这部分内

容.

例题 1.44 设随机变量 $X_1, \cdots, X_n$ 独立同分布, 且服从 $[0, a]$ 上的均匀分布, 其中 $a > 0$. 记

$$M_n = \max\{X_1, \cdots, X_n\},$$

分别在 a.s., r 阶收敛的意义下证明 $M_n \to a$.

解 显然 $M_n < a$, 对 $\forall 0 < \varepsilon < a$ 有

$$\mathbb{P}(M_n \leqslant a - \varepsilon) = \prod_{k=1}^{n} \mathbb{P}(X_k \leqslant a - \varepsilon) = \left(\frac{a-\varepsilon}{a}\right)^n.$$

因此

$$\sum_{n=1}^{+\infty} \mathbb{P}(M_n < a - \varepsilon) = \sum_{n=1}^{+\infty} \left(\frac{a-\varepsilon}{a}\right)^n < +\infty.$$

由 Borel-Cantelli 引理知 $\mathbb{P}(M_n \leqslant a - \varepsilon, \text{i.o.}) = 0$. 因此 $M_n \xrightarrow{\text{a.s.}} a$.

同时有

$$\begin{aligned}
\mathbb{E}[|M_n - a|^r] &= \mathbb{E}[|M_n - a|^r I_{\{|M_n - a| \leqslant \varepsilon\}}] + \mathbb{E}[|M_n - a|^r I_{\{|M_n - a| > \varepsilon\}}] \\
&\leqslant \varepsilon^r + a^r \mathbb{P}(|M_n - a| > \varepsilon) \\
&= \varepsilon^r + a^r \left(\frac{a-\varepsilon}{a}\right)^n.
\end{aligned}$$

令 $n \to +\infty$, 再令 $\varepsilon \to 0$, 则有

$$M_n \xrightarrow{r} a.$$

证明/利用依分布收敛:

(1) 定义: $F_{X_n} \xrightarrow{w} F_X$. 例子见作业题 **Grimmett5.12.32, 5.12.39** .

(2) 利用连续性定理, 用特征函数证明或者 CLT、Lindeberg-Feller CLT 等, 内容详见 **1.16 节**. 例子见作业题 **Grimmett5.12.41** .

(3) 弱大数律 (注意条件是 **i.i.d** 且一阶矩存在).

(4) 利用 $X_n \xrightarrow{\text{P}} X \Longrightarrow X_n \xrightarrow{\text{D}} X$.

(5) 利用依分布收敛的等价条件, 各种性质转化, 具体见后.

⋮

⋮

例题 1.45 随机变量 $X_1, X_2, \cdots X_n$ 独立同分布且 $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1$, $S_n = \sum_{k=1}^{n} X_k$, 则

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\text{D}} N(0, 1),$$

但不存在随机变量 Z, 使得

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\text{P}} Z.$$

证明 利用 i.i.d CLT 即可推出 $\frac{S_n}{\sqrt{n}} \xrightarrow{\text{D}} N(0,1)$, 现反证法证明后一个命题: 若存在随机变量 Z, 使得 $\frac{S_n}{\sqrt{n}} \xrightarrow{\text{P}} Z$ 成立, 则由极限分布的唯一性及前一个命题知 $Z \sim N(0,1)$. 一方面,

$$\frac{S_{2n} - S_n}{\sqrt{2n}} = \frac{S_{2n}}{\sqrt{2n}} - \frac{1}{\sqrt{2}} \frac{S_n}{\sqrt{n}} \xrightarrow{\text{P}} \left(1 - \frac{1}{\sqrt{2}}\right) Z.$$

另一方面,

$$\frac{S_{2n}-S_n}{\sqrt{2n}} \overset{\mathrm{D}}{=} \frac{S_n}{\sqrt{2n}} \xrightarrow{\mathrm{P}} \frac{1}{\sqrt{2}}Z,$$

矛盾. 因此不存在 Z 满足 $\dfrac{S_n}{\sqrt{n}} \xrightarrow{\mathrm{P}} Z$.

另外对于连续性随机变量, 也可以通过证明密度函数列的逐点收敛性推出依分布收敛:

命题 1.7

设 $\{X_n\}, X$ 均为连续型随机变量, 记 X_n, X 的密度函数是 $f_n(x), f(x)$, 若对 $\forall x \in \mathbb{R}$, 均有 $f_n(x) \to f(x)$, 则有 $X_n \xrightarrow{D} X$.

证明 对 $\forall x \in \mathbb{R}$, 利用 $|X| = X^+ + X^- = 2X^+ - X$, 有

$$\begin{aligned}|\mathbb{P}(X_n \leqslant x) - \mathbb{P}(X \leqslant x)| &= \left|\int_{-\infty}^{x} (f_n(y) - f(y))\mathrm{d}y\right| \leqslant \int_{-\infty}^{+\infty} |f_n(y) - f(y)|\mathrm{d}y = 2\int_{-\infty}^{+\infty} (f(y) - f_n(y))^+\mathrm{d}y \\ &\leqslant 2\int_{-\infty}^{+\infty} f(y)\mathrm{d}y < +\infty.\end{aligned}$$

利用控制收敛定理, 有

$$\lim_{n\to+\infty}\int_{-\infty}^{+\infty} |f_n(y) - f(y)|\mathrm{d}y = 2\lim_{n\to+\infty}\int_{-\infty}^{+\infty} (f(y) - f_n(y))^+\mathrm{d}y = 2\int_{-\infty}^{+\infty} \lim_{n\to+\infty}(f(y) - f_n(y))^+\mathrm{d}y = 0.$$

进而

$$\lim_{n\to+\infty} |\mathbb{P}(X_n \leqslant x) - \mathbb{P}(X \leqslant x)| = 0 \implies X_n \xrightarrow{D} X.$$

回顾已讲过的关于依分布收敛的内容:

引理 1.16

$X_n \xrightarrow{D} X \iff$ 对任意子列 $\{n_k\}$, 存在子子列 $\{n_k'\} \subset \{n_k\}$, 使得

$$X_{n_k'} \xrightarrow{D} X.$$

定理 1.26 (Skorokhod 表示定理)

设 $X_n \xrightarrow{D} X$, 则存在 $(\Omega, \mathcal{F}, \mathbb{P})$ 及其上的 Y_n, Y 满足:

(1) Y_n 与 Y 同分布, Y 与 X 同分布,

(2) $Y_n \xrightarrow{\text{a.s.}} Y$.

定理 1.27

$X_n \xrightarrow{D} X \iff \forall g \in C_b(\mathbb{R})$(有界连续函数), 有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)], n \to \infty$.

下面看一个和**定理** 1.27有关的例子:

例题 1.46 设 $X_n \xrightarrow{\mathrm{D}} X$, 且存在 $r, C > 0$, 使得 $\mathbb{E}[|X_n|^r] \leqslant C$, 证明: 对 $\forall 0 < s < r$, 均有

$$\lim_{n\to+\infty} \mathbb{E}[|X_n|^s] = \mathbb{E}[X^s].$$

证明 利用定理 1.27, 对 $\forall g \in C_b(\mathbb{R})$(有界连续函数), 有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)], n \to \infty$. 设 $M > 0$, 定义

$$g_M(x) = \begin{cases} |x|^s, & |x| < M, \\ M^s, & |x| \geqslant M. \end{cases}$$

则 $g_M(x)$ 是有界连续函数, 以及当 $M \to +\infty$ 时 $0 \leqslant g_M(x) \uparrow |x|^s$, 因此有

$$\lim_{n\to+\infty} \mathbb{E}[g_M(X_n)] = \mathbb{E}[g_M(X)].$$

同时我们有

$$0 \leqslant \mathbb{E}[|X_n|^s] - \mathbb{E}[g_M(X_n)] \leqslant \mathbb{E}[|X_n|^s I_{\{|X_n|\geqslant M\}}] \leqslant \frac{1}{M^{r-s}}\mathbb{E}[|X_n|^r I_{\{|X_n|\geqslant M\}}] \leqslant \frac{C}{M^{r-s}},$$

因此

$$\mathbb{E}[g_M(X_n)] \leqslant \mathbb{E}[|X_n|^s] \leqslant \mathbb{E}[g_M(X_n)] + \frac{C}{M^{r-s}}.$$

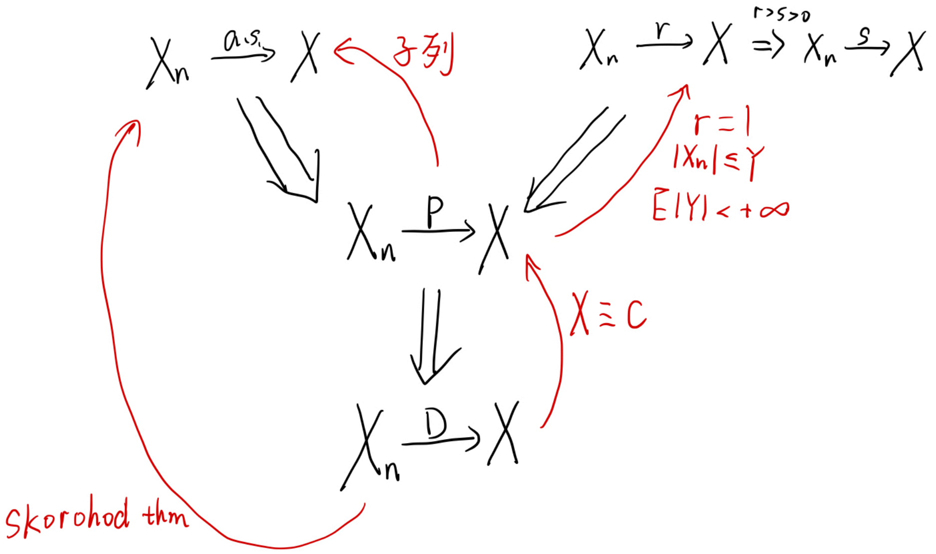
先令 $n \to +\infty$, 再令 $M \uparrow +\infty$, 结合 $\lim\limits_{M\to+\infty} \mathbb{E}[g_M(X)] = \mathbb{E}[|X|^s]$(利用单调收敛定理), 得

$$\lim_{n\to+\infty} \mathbb{E}[|X_n|^s] = \mathbb{E}[X^s].$$

注 这里的截尾处理非常精妙, 自行体会.

1.17.3 结论拾零及应用

本节主要以例子来呈现. 先回顾一下各收敛间的关系:



例题 1.47 设 $f:\mathbb{R}\to\mathbb{R}$ 是连续函数, 证明:

(1) $X_n \xrightarrow{\text{a.s.}} X \implies f(X_n) \xrightarrow{\text{a.s.}} f(X)$.

(2) $X_n \xrightarrow{\text{P}} X \implies f(X_n) \xrightarrow{\text{P}} f(X)$.

(3) $X_n \xrightarrow{\text{D}} X \implies f(X_n) \xrightarrow{\text{D}} f(X)$.

证明 (1) $\exists \Omega_0 \subset \Omega$, 使得 $\mathbb{P}(\Omega_0) = 1$ 且对 $\forall \omega \in \Omega_0$, 均有 $X_n(\omega) \to X(\omega)$, 由 f 的连续性知 $f(X_n(\omega)) \to f(X(\omega))$, 因此 $f(X_n) \xrightarrow{\text{a.s.}} f(X)$.

(2) 反证: 若命题不成立, 则存在子列 $\{n_k\}$, $\varepsilon, \delta > 0$, 满足

$$\mathbb{P}(|f(X_{n_k}) - f(X)| > \varepsilon) > \delta$$

因为 $X_{n_k} \xrightarrow{\text{P}} X$, 利用命题1.6, 存在子列 $\{n_k'\} \subset \{n_k\}$, 使得 $X_{n_k'} \xrightarrow{\text{a.s.}} X$. 由 (1) 知 $f(X_{n_k'}) \xrightarrow{\text{a.s.}} f(X)$, 从而 $f(X_{n_k'}) \xrightarrow{\text{P}} f(X)$, 矛盾.

(3) 利用 Skorokhod 表示定理, 在一个概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上存在 Y_n,Y, 使得 $Y_n \overset{D}{=\!=} X_n, Y \overset{D}{=\!=} X$ 且 $Y_n \xrightarrow{\text{a.s.}} Y$. 由 (1) 知 $f(Y_n) \xrightarrow{\text{a.s.}} f(Y)$, 因此 $f(Y_n) \xrightarrow{\text{D}} f(Y) \implies f(X_n) \xrightarrow{\text{D}} f(X)$.

例题 1.48 设 $X_1 \cdots X_n$ 是随机变量列, $N_1, N_2, \cdots$ 是取值为正整数值的随机变量, 证明:

(1) 若 $X_n \xrightarrow{\text{a.s.}} X$ 且 $N_k \xrightarrow{a.s.} \infty$, 则 $X_{N_k} \xrightarrow{\text{a.s.}} X$.

(2) 若 $X_n \xrightarrow{\text{a.s.}} X$ 且 $N_k \xrightarrow{P} \infty$, 则 $X_{N_k} \xrightarrow{\text{P}} X$.

(3) 若 $X_n \xrightarrow{\text{P}} X, N_k \xrightarrow{P} \infty$ 且 $\{X_n\}$ 与 $\{N_k\}$ 独立, 则 $X_{N_k} \xrightarrow{\text{P}} X$.

(4) 若 $X_n \xrightarrow{\text{D}} X, N_k \xrightarrow{P} \infty$ 且 $\{X_n\}$ 与 $\{N_k\}$ 独立, 则 $X_{N_k} \xrightarrow{\text{D}} X$.

证明 (1) $\exists \Omega_0 \subset \Omega$, 使得 $\mathbb{P}(\Omega_0) = 1$ 且对 $\forall \omega \in \Omega_0$, 均有 $X_n(\omega) \to X(\omega)$ 及 $N_k(\omega) \to +\infty$. 故对 $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*$, 当 $n > N$ 时, 有 $|X_n(\omega) - X(\omega)| < \varepsilon$ 且 $\exists K \in \mathbb{N}^*$, 当 $k > K$ 时, 有 $N_k(\omega) > n$, 从而 $|X_{N_k}(\omega) - X(\omega)| < \varepsilon$. 故 $X_{N_k}(\omega) \to X(\omega)$, 因此 $X_{N_k} \xrightarrow{\text{a.s.}} X$.

(2) 对 $\forall \varepsilon > 0$, 有

$$\mathbb{P}(|X_{N_k} - X| > \varepsilon) \leqslant \mathbb{P}(N_k \leqslant n) + \mathbb{P}(|X_{N_k} - X| > \varepsilon, N_k > n) \leqslant \mathbb{P}(N_k \leqslant n) + \mathbb{P}(\sup_{m \geqslant n} |X_m - X| > \varepsilon)$$

先令 $k \to +\infty$, 再令 $n \to +\infty$, 结合命题 1.5即得.

(3) 对 $\forall \varepsilon > 0, \delta > 0, \exists N \in \mathbb{N}^*$, 当 $n > N$ 时, 有 $\mathbb{P}(|X_n - X| > \varepsilon) < \delta$, $\exists K \in \mathbb{N}^*$, 当 $k > K$ 时, 有 $\mathbb{P}(N_k \leqslant N) < \delta$. 因此

$$\begin{aligned}
\mathbb{P}(|X_{N_k} - X| > \varepsilon) &\leqslant \mathbb{P}(N_k \leqslant N) + \sum_{n=N+1}^{+\infty} \mathbb{P}(|X_{N_k} - X| > \varepsilon, N_k = n) \\
&= \mathbb{P}(N_k \leqslant N) + \sum_{n=N+1}^{+\infty} \mathbb{P}(|X_n - X| > \varepsilon)\mathbb{P}(N_k = n) \\
&\leqslant 2\delta.
\end{aligned}$$

令 $\delta \to 0$ 即得.

(4) 设 $\phi_n(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}X_n t}], \phi(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}Xt}], \psi_k(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}X_{N_k} t}] = \mathbb{E}[\mathbb{E}[\mathrm{e}^{\mathrm{i}X_{N_k} t}|N_k]] = \sum_{n=1}^{+\infty} \mathbb{P}(N_k = n)\phi_n(t)$, 则

$$|\psi_k(t) - \phi(t)| \leqslant \sum_{n=1}^{+\infty} \mathbb{P}(N_k = n)|\phi_n(t) - \phi(t)|$$

对 $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*$, 则对 $n > N$, 有 $|\phi_n(t) - \phi(t)| < \varepsilon, \exists K \in \mathbb{N}^*$. 当 $k > K$ 时, 有 $\mathbb{P}(N_k \leqslant N) < \dfrac{\varepsilon}{M}$ (其中 $M = \max\limits_{1 \leqslant i \leqslant n} |\phi_i(t) - \phi(t)|$), 因此

$$|\psi_k(t) - \phi(t)| \leqslant \frac{\varepsilon}{M} \max_{1 \leqslant i \leqslant n} |\phi_i(t) - \phi(t)| + \mathbb{P}(N_k > N)\varepsilon \leqslant 2\varepsilon.$$

令 $\varepsilon \to 0$ 即得.

定理 1.28 (Slutsky 引理)

$X_n \xrightarrow{\text{D}} X, Y_n \xrightarrow{\text{P}} b, Z_n \xrightarrow{\text{P}} c$. 则

$$X_n Y_n + Z_n \xrightarrow{\text{D}} bX + c.$$

证明 见本文附录, 可以暂时忽略.

注

1. 若 $b \neq 0$, 则有 $Y_n^{-1} \xrightarrow{\text{P}} b^{-1}$. 另外可以配合**例 1.47(2)** 食用.
2. Slutsky 引理在研究参数估计、统计量的渐进性质中会经常用到, 作业题 **Grimmett7.2.5(a)** 给出了

其简单形式, 我们来看几个相关的例子:

例题 1.49 设均值为 μ 的随机变量列 $\{X_n\}$ 独立同分布且二阶矩有限, 令 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$, 证明:

$$\sum_{k=1}^{n}(X_k-\mu)\Big/\sqrt{\sum_{k=1}^{n}(X_k-\overline{X})^2}\xrightarrow{\mathrm{D}}N(0,1).$$

证明 设其方差为 σ^2, 即证

$$\sum_{k=1}^{n}\frac{X_k-\mu}{\sqrt{n}\sigma}\Big/\sqrt{\sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{n\sigma^2}}\xrightarrow{\mathrm{D}}N(0,1).$$

注意到

$$\sum_{k=1}^{n}\frac{(X_k-\overline{X})^2}{n\sigma^2}=\frac{1}{n\sigma^2}\sum_{k=1}^{n}(X_k-\mu)^2-\frac{2}{n\sigma^2}(\overline{X}-\mu)\sum_{k=1}^{n}(X_k-\mu)+\frac{1}{\sigma^2}(\overline{X}-\mu)^2.$$

利用 CLT 知

$$\sum_{k=1}^{n}\frac{X_k-\mu}{\sqrt{n}\sigma}\xrightarrow{\mathrm{D}}N(0,1).$$

利用弱大数律知

$$\frac{1}{n\sigma^2}\sum_{k=1}^{n}(X_k-\mu)^2\xrightarrow{\mathrm{P}}1,\quad \frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)\xrightarrow{\mathrm{P}}0,\quad \overline{X}-\mu\xrightarrow{\mathrm{P}}0.$$

结合 **Slutsky** 引理知

$$\sum_{k=1}^{n}(X_k-\mu)\Big/\sqrt{\sum_{k=1}^{n}(X_k-\overline{X})^2}\xrightarrow{\mathrm{D}}N(0,1).$$

例题 1.50 随机变量列 $\{X_n\}$ 相互独立且满足

$$\mathbb{P}(X_n=\pm n)=\frac{1}{2n^3},\quad \mathbb{P}(X_n=\pm 1)=\frac{1}{2}-\frac{1}{2n^3},\quad n\in\mathbb{N}^*.$$

证明:

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k\xrightarrow{\mathrm{D}}N(0,1).$$

证明 引入截尾处理后的随机变量: $Y_i=\mathrm{sgn}X_i$, 注意到

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}Y_k+\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k-Y_k)$$

一方面, 利用 CLT 知,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}Y_k\xrightarrow{\mathrm{D}}N(0,1).$$

另一方面, 对 $\forall\varepsilon>0$, 有

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k-Y_k)\right|>\varepsilon\right)&\leqslant\frac{1}{\sqrt{n}\varepsilon}\mathbb{E}[|\sum_{k=1}^{n}(X_k-Y_k)|]\leqslant\frac{1}{\sqrt{n}\varepsilon}\sum_{k=1}^{n}\mathbb{E}[|X_k-Y_k|]\\&\leqslant\frac{1}{\sqrt{n}\varepsilon}\sum_{k=1}^{+\infty}\frac{1}{k^2}\xrightarrow{n\to+\infty}0.\end{aligned}$$

因此

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k-Y_k)\xrightarrow{\mathrm{P}}0.$$

结合 **Slutsky** 引理知,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k \xrightarrow{\mathrm{D}} N(0,1).$$

注 截尾法的妙用在这里又体现了出来. 这里也可以类似**例** 1.39一样转化.

例题 1.51 (Δ 方法) 随机变量列 $\{X_n\}$, 且存在常数 a 和 $\sigma^2>0$, 使得

$$\sqrt{n}(X_n-a)\xrightarrow{\mathrm{D}} N(0,\sigma^2).$$

若函数 g 在 a 处可导, 且 $g'(a)\neq 0$, 则

$$\sqrt{n}(g(X_n)-g(a))\xrightarrow{\mathrm{D}} N(0,\sigma^2(g'(a))^2).$$

证明 参考 苏淳, 冯群强: 概率论（第三版）, 例 **6.3.4**. 这也是 Slutsky 引理的一个重要应用.

例题 1.52 (Kolmogorov 不等式) 设 $\{X_k,1\leqslant k\leqslant n\}$ 独立并满足 $\mathbb{E}[X_k]=0,\mathbb{E}[X_k^2]<\infty$, 记 $S_n=\sum\limits_{k=1}^{n}X_k$. 对 $\forall\varepsilon>0$,

(1) 证明: $\mathbb{P}(\max\limits_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon)\leqslant\dfrac{1}{\varepsilon^2}\sum\limits_{k=1}^{n}\mathbb{E}[X_k^2]$.

(2) 若存在 $c>0$, 使得 $|X_k|\leqslant c,\forall 1\leqslant k\leqslant n$, 证明:

$$\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon)\geqslant 1-\frac{(\varepsilon+c)^2}{\sum\limits_{k=1}^{n}\mathbb{E}[X_k^2]}.$$

证明 (1) 记 $T=\inf\{k\in\mathbb{N}^*:|S_k|\geqslant\varepsilon\}$, 则

$$\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon)=\mathbb{P}(T\leqslant n)=\sum_{k=1}^{n}\mathbb{P}(T=k)=\sum_{k=1}^{n}\mathbb{E}[I_{\{T=k\}}]\leqslant\mathbb{E}\left[\left(\frac{|S_k|}{\varepsilon}\right)^2I_{\{T=k\}}\right]=\frac{1}{\varepsilon^2}\sum_{k=1}^{n}\mathbb{E}[S_k^2I_{\{T=k\}}]$$

而

$$\begin{aligned}\mathbb{E}[S_n^2I_{\{T=k\}}]&=\mathbb{E}[(S_k+S_n-S_k)^2I_{\{T=k\}}]=\mathbb{E}[S_k^2I_{\{T=k\}}]+\mathbb{E}[(S_n-S_k)^2I_{\{T=k\}}]+2\mathbb{E}[S_k(S_n-S_k)I_{\{T=k\}}]\\&\geqslant\mathbb{E}[S_k^2I_{\{T=k\}}].\end{aligned}$$

其中 $\mathbb{E}[S_k(S_n-S_k)I_{\{T=k\}}]=\mathbb{E}[S_kI_{\{T=k\}}]\mathbb{E}[S_n-S_k]=0$, 这是因为 $\{T=k\}$ 只决定于 $\{X_1,\cdots X_k\}$ ($\{T=k\}\in\sigma(X_1,\cdots X_k)$). 因此

$$\mathbb{P}(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon)=\frac{1}{\varepsilon^2}\sum_{k=1}^{n}\mathbb{E}[S_k^2I_{\{T=k\}}]\leqslant\frac{1}{\varepsilon^2}\sum_{k=1}^{n}\mathbb{E}[S_n^2I_{\{T=k\}}]=\frac{1}{\varepsilon^2}\mathbb{E}[S_n^2I_{\{T\leqslant n\}}]\leqslant\frac{\mathbb{E}[S_n^2]}{\varepsilon^2}=\frac{1}{\varepsilon^2}\sum_{k=1}^{n}\mathbb{E}[X_k^2].$$

(2) 我们有

$$\begin{aligned}\mathbb{E}[S_n^2I_{\{T=k\}}]&=\mathbb{E}[S_k^2I_{\{T=k\}}]+\mathbb{E}[(S_n-S_k)^2I_{\{T=k\}}]\leqslant(\varepsilon+c)^2\mathbb{P}(T=k)+\mathbb{E}[(S_n-S_k)^2]\mathbb{P}(T=k)\\&\leqslant\mathbb{P}(T=k)(\mathbb{E}[S_n^2]+(\varepsilon+c)^2),\end{aligned}$$

其中 $\mathbb{E}[S_k^2I_{\{T=k\}}]\leqslant(\varepsilon+c)^2\mathbb{P}(T=k)$ 是因为在 $\{T=k\}$ 下有 $|S_k|\leqslant|S_{k-1}+X_k|\leqslant\varepsilon+c$. 对 k 从 1 至 n 累加得

$$\mathbb{E}[S_n^2I_{\{T\leqslant n\}}]\leqslant\mathbb{P}(T\leqslant n)(\mathbb{E}[S_n^2]+(\varepsilon+c)^2).$$

另一方面,

$$\mathbb{E}[S_n^2I_{\{T\leqslant n\}}]=\mathbb{E}[S_n^2]-\mathbb{E}[S_n^2I_{\{T>n\}}]\geqslant\mathbb{E}[S_n^2]-\varepsilon^2\mathbb{P}(T>n)=\mathbb{E}[S_n^2]-\varepsilon^2+\varepsilon^2\mathbb{P}(T\leqslant n).$$

因此

$$\mathbb{P}(T \leqslant n) \geqslant \frac{\mathbb{E}[S_n^2] - \varepsilon^2}{\mathbb{E}[S_n^2] + (\varepsilon + c)^2 - \varepsilon^2} = 1 - \frac{(\varepsilon + c)^2}{\mathbb{E}[S_n^2] + (\varepsilon + c)^2 - \varepsilon^2} \geqslant 1 - \frac{(\varepsilon + c)^2}{\mathbb{E}[S_n^2]}.$$

注 这里引入了新的概念停时: $T = \inf\{1 \leqslant k \leqslant n : |S_k| \geqslant \varepsilon\}$, 这是一个概率尤其是随机过程中非常基本且重要的工具 (直观理解: 调节 “信息” 变化).

利用 Kolmogorov 不等式可以得到一级数定理:

定理 1.29 (一级数定理)

设 $\{X_n\}$ 独立, $\mathbb{E}[X_n] = 0$, $\sum_{n=1}^{\infty} \mathbb{E}X_n^2 < \infty$, 则 $\sum_{n=1}^{\infty} X_n$ a.s. 收敛.

证明 记 $S_n = \sum_{k=1}^{n} X_k$, 利用 Kolmogorov 不等式 (例 1.52), 有

$$\mathbb{P}(\max_{M \leqslant m \leqslant N} |S_m - S_M| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}[|S_N - S_M|^2]}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{m=M+1}^{N} \mathbb{E}[X_m^2].$$

令 $N \uparrow +\infty$, 则有

$$\mathbb{P}(\max_{m \geqslant M} |S_m - S_M| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} \sum_{m=M+1}^{+\infty} \mathbb{E}[X_m^2] \xrightarrow{M \to +\infty} 0.$$

结合命题 1.5, 得

$$\max_{m \geqslant M} |S_m - S_M| \xrightarrow{\mathrm{P}} 0 \implies S_m \text{ 为 a.s. Cauchy 列} \implies S_m \text{ a.s. 收敛} \implies \sum_{n=1}^{\infty} X_n \text{ a.s. 收敛}.$$

1.17.4 用截尾术证明弱大数律

课堂上我们利用特征函数及连续性定理方法证明了弱大数律 (先证明依分布收敛于 μ, 再依概率收敛等价转换). 这里我们再展示另一种证明方法——截尾术:

定理 1.30 (弱大数律)

设 $X_1, \cdots, X_n$ 独立同分布, $\mu = \mathbb{E}[X_1]$ 且 $\mathbb{E}[|X_1|] < +\infty$, 记 $S_n = \sum_{k=1}^{n} X_k$, 则

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

证明 引入经截断处理的随机变量:

$$X_n^{(1)} = X_n I_{\{|X_n| \leqslant M\}}, \quad X_n^{(2)} = X_n I_{\{|X_n| > M\}}.$$

其中 $X_n = X_n^{(1)} + X_n^{(2)}, M > 0$. 并设 $S_n^{(1)} = \sum_{i=1}^{n} X_i^{(1)}, S_n^{(2)} = \sum_{i=1}^{n} X_i^{(2)}$. 则对 $\forall \varepsilon > 0$, 有

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{(S_n^{(1)} - \mathbb{E}[S_n^{(1)}]) + (S_n^{(2)} - \mathbb{E}[S_n^{(2)}])}{n}\right| > \varepsilon\right) \\ &\leqslant \mathbb{P}\left(\left|\frac{S_n^{(1)} - \mathbb{E}[S_n^{(1)}]}{n}\right| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\left|\frac{S_n^{(2)} - \mathbb{E}[S_n^{(2)}]}{n}\right| > \frac{\varepsilon}{2}\right)\end{aligned}$$

现在我们分别处理这两部分:

$$\mathbb{P}\left(\left|\frac{S_n^{(1)}-\mathbb{E}[S_n^{(1)}]}{n}\right|>\frac{\varepsilon}{2}\right)\leqslant\frac{\operatorname{Var}(S_n^{(1)})}{(\frac{1}{2}\varepsilon n)^2}=\frac{4\operatorname{Var}(X_1^{(1)})}{\varepsilon^2 n}\leqslant\frac{4\mathbb{E}[(X_1^{(1)})^2]}{\varepsilon^2 n}\leqslant\frac{4M^2}{\varepsilon^2 n}\xrightarrow{n\to+\infty}0.$$

同时有

$$\mathbb{P}\left(\left|\frac{S_n^{(2)}-\mathbb{E}[S_n^{(2)}]}{n}\right|>\frac{\varepsilon}{2}\right)\leqslant\frac{\mathbb{E}|S_n^{(2)}-\mathbb{E}[S_n^{(2)}]|}{\frac{\varepsilon}{2}n}\leqslant\frac{4\mathbb{E}[|X_1|I_{\{|X_1|>M\}}]}{\varepsilon}\xrightarrow{M\to+\infty}0,$$

其中

$$\mathbb{E}|S_n^{(2)}-\mathbb{E}[S_n^{(2)}]|\leqslant\mathbb{E}[|S_n^{(2)}|]+|\mathbb{E}[S_n^{(2)}]|\leqslant 2\mathbb{E}[|S_n^{(2)}|]\leqslant 2\sum_{k=1}^n\mathbb{E}[|X_k^{(2)}|]=2n\mathbb{E}[|X_1^{(2)}|].$$

因此令 $n\to+\infty$, 再令 $M\to+\infty$, 有

$$\mathbb{P}\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)\to 0.$$

故 $\dfrac{S_n}{n}\xrightarrow{P}\mu$.

注

1. 分别处理并运用 Markov 不等式等矩不等式. 因为 $X_n^{(1)}$ 有界, 因此其常用于高阶矩方法估计, 而尾巴部分 $X_n^{(2)}$ 通过 $\mathbb{E}[|X|]<\infty$ 用一阶矩来控制.
2. 也可以这样截尾: $Y_n=X_nI_{\{|X_n|\leqslant n\}}$. 由**例** 1.39 知 $\dfrac{1}{n}\sum_{k=1}^n(X_k-Y_k)\xrightarrow{\text{a.s.}}0$. 因此

$$\frac{1}{n}\sum_{k=1}^n X_k\xrightarrow{\mathrm{P}}\mathbb{E}[X]\Longleftrightarrow\frac{1}{n}\sum_{k=1}^n Y_k\xrightarrow{\mathrm{P}}\mathbb{E}[X].$$

而

$$\frac{1}{n}\sum_{k=1}^n\mathbb{E}[Y_k]=\frac{1}{n}\sum_{k=1}^n\mathbb{E}[X_1I_{\{|X_1|\leqslant k\}}]\to\mathbb{E}[X]\ (a_n\to a\Longrightarrow\frac{1}{n}\sum_{k=1}^n a_k\to a)$$

只需证 $\dfrac{1}{n}\sum_{k=1}^n(Y_k-\mathbb{E}[Y_k])\xrightarrow{\mathrm{P}}0$, 由 CHebyshev 不等式, 只需证 $\dfrac{1}{n^2}\sum_{k=1}^n\mathbb{E}[Y_k^2]\xrightarrow{n\to+\infty}0$, 即证

$$\sum_{i=1}^\infty\frac{1}{i^2}\mathbb{E}[Y_i^2]<\infty.$$

记 $B_{ij}=\{j-1\le X_i<j\}$, 则

$$\begin{aligned}\sum_i\frac{1}{i^2}\mathbb{E}[Y_i^2]&=\sum_{i=1}^\infty\frac{1}{i^2}\sum_{j=1}^i\mathbb{E}[Y_i^2I_{B_{ij}}]\le\sum_{i=1}^\infty\frac{1}{i^2}\sum_{j=1}^i j^2\mathbb{P}(B_{ij})\\&=\sum_{i=1}^\infty\frac{1}{i^2}\sum_{j=1}^i j^2\mathbb{P}(B_{1j})\overset{\text{Fubini}}{=\!=\!=}\sum_{j=1}^\infty\left(\sum_{i=j}^\infty\frac{1}{i^2}\right)j^2\mathbb{P}(B_{1j})\\&\le\sum_{j=1}^\infty\frac{2}{j}\cdot j^2\mathbb{P}(B_{1j})=2\sum_{j=1}^\infty j\mathbb{P}(B_{1j})\\&\le 2(1+\mathbb{E}[X_1])<\infty\end{aligned}$$

利用**注 2** 的证明思路, 我们还可以得到弱大数律最精确的结果, 因此弱大数律并不是真正 “弱” 爆了.

定理 1.31

设 $X_1,\cdots,X_n$ 独立同分布, 且满足

$$x\mathbb{P}(|X_1|>x)\xrightarrow{x\to+\infty}0.$$

记 $S_n=\sum_{k=1}^{n}X_k,\ \mu_n=\mathbb{E}[X_1I_{\{|X_1|\leqslant n\}}]$ 则

$$\frac{S_n}{n}-\mu_n\xrightarrow{P}0.$$

证明 具体可以参考 **Durrett Theorem 2.2.12**.

1.17.5 再谈 a.s. 收敛和强大数律

这节我们着重讲一讲证明 a.s. 收敛的处理技巧. 首先 Borel-Cantelli 引理是一个最基本也是必要的工具, 关键是怎么去使用? 这里时常要辅助**子序列**, **截尾**, 独立复制与对称化等等方法来处理. 另外, 如何估计 $\mathbb{P}(|X_n-X|>\varepsilon)$ 也是一门学问, 一般来说要借助各种矩不等式或其他方法及放缩技巧解决. Markov 不等式 (包括其推论和 Chebyshev 不等式) 是最常见的一种, 对于随机变量来说, 矩的阶数越高带来的信息越多 (高阶矩存在直接推出低阶矩存在), 所以对矩选取合适的阶数也是至关重要的. 本节主要从选 "矩" 的角度运用 Markov 不等式和子序列方法这两个层面进行讲解, 最后再重新回到上课已讲过的强大数律中.

子序列方法

1. 先找子列 $\{n_k\}$, 结合 Borel-Cantelli 引理, 满足
$$\sum_k\mathbb{P}(|X_{n_k}-X|>\varepsilon)<\infty,\forall\varepsilon>0\implies X_{n_k}\xrightarrow{\text{a.s.}}X.$$
2. 对于一般项 n, 寻找子列中离其前后最近两的项消除其影响. 设 $n_k\leqslant n<n_{k+1}$, 下面两种是比较常见的处理方法:
   - $\sup\limits_{n_k\leqslant n<n_{k+1}}|X_n-X_{n_k}|\xrightarrow{\text{a.s.}}0.$
   - 如果这里 X_n 是 $\dfrac{S_n}{n}$ 型的, 则可以利用单调性控制 (前提 $S_{n_k}\leqslant S_n\leqslant S_{n_k+1}$):
   $$\frac{S_{n_k}}{n_{k+1}}\leqslant\frac{S_n}{n}\leqslant\frac{S_{n_k+1}}{n_k}.$$
   同时为了利用夹逼原理, 还要满足两侧均满足 a.s 收敛且收敛值相同.

   满足以上两条件之一, 再结合 $X_{n_k}\xrightarrow{\text{a.s.}}X$ 就能推出 $X_n\xrightarrow{\text{a.s.}}X$.

我们通过几个例子来展现子序列方法:

例题 1.53 随机变量 $X_1,\cdots,X_n$ 独立同分布, $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\dfrac{1}{2}$, 记 $S_n=\sum_{k=1}^{n}X_k$, 证明:

$$\frac{S_n}{n}\xrightarrow{\text{a.s.}}0.$$

证明 先证: $\dfrac{S_{n^2}}{n^2}\xrightarrow{\text{a.s.}}0$. 对 $\forall\varepsilon>0$, 利用 Markov 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^2n^4}\mathbb{E}[S_{n^2}^2]=\frac{1}{\varepsilon^2n^4}\sum_{k=1}^{n^2}\mathbb{E}[X_k^2]=\frac{1}{\varepsilon^2n^2}\implies\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right|>\varepsilon\right)<+\infty.$$

由 Borel-Cantelli 引理知, $\dfrac{S_{n^2}}{n^2} \xrightarrow{\text{a.s.}} 0$.

回到一般: 对 $\forall k \in \mathbb{N}^*$, $\exists n \in \mathbb{N}^*$, 使得 $n^2 \leqslant k < (n+1)^2$. 而

$$\sup_{n^2 \leqslant k<(n+1)^2} \frac{|S_k - S_{n^2}|}{n^2} = \sup_{n^2 \leqslant k<(n+1)^2} \frac{\left|\sum\limits_{i=n^2+1}^{k} X_i\right|}{n^2} \leqslant \frac{(n+1)^2-n^2}{n^2} = \frac{2n+1}{n^2} \xrightarrow{\text{a.s.}} 0(n \to +\infty)$$

因此

$$\sup_{n^2 \leqslant k<(n+1)^2} \left|\frac{S_k}{k}\right| \leqslant \sup_{n^2 \leqslant k<(n+1)^2} \frac{|S_k - S_{n^2}|}{n^2} + \frac{|S_{n^2}|}{n^2} \xrightarrow{\text{a.s.}} 0(n \to +\infty) \implies \frac{S_k}{k} \xrightarrow{\text{a.s.}} 0.$$

注 这里只展现子序列的方法, 也可以通过更高阶矩的估计来完成.

例题 1.54 设零均值随机变量 $X_1, \cdots, X_n$ 两两不相关, $\sup\limits_n \mathbb{E}[X_n^2] \leqslant M < \infty$, 记 $S_n = \sum\limits_{k=1}^{n} X_k$, 定义 $D_n = \sup\limits_{n^2 \leqslant k<(n+1)^2} |S_k - S_{n^2}|$, 证明:

(1) $\dfrac{D_n}{n^2} \xrightarrow{\text{a.s.}} 0$;

(2) $\dfrac{S_n}{n} \xrightarrow{\text{a.s.}} 0$.

证明 (1) 注意到

$$\mathbb{E}[D_n^2] = \mathbb{E}(\sup_{n^2 \leqslant k<(n+1)^2} |S_k - S_{n^2}|)^2 \leqslant \sum_{k=n^2+1}^{(n+1)^2} \mathbb{E}[|S_k - S_{n^2}|^2] \leqslant (2n+1)\mathbb{E}[|S_{(n+1)^2-1} - S_{n^2}|^2].$$

故对 $\forall \varepsilon > 0$, 利用 Markov 不等式, 有

$$\mathbb{P}\left(\left|\frac{D_n}{n^2}\right| > \varepsilon\right) \leqslant \frac{\mathbb{E}[D_n^2]}{\varepsilon^2 n^4} \leqslant \frac{2n+1}{\varepsilon^2 n^4}\mathbb{E}[|S_{(n+1)^2-1} - S_{n^2}|^2] \leqslant \frac{(2n+1)^2 M}{n^4 \varepsilon^2} \leqslant \frac{C}{n^2} \xrightarrow{n\to+\infty} 0.$$

其中 C 为某个正常数. 因此 $\sum\limits_{n=1}^{+\infty} \mathbb{P}\left(\left|\dfrac{D_n}{n^2}\right| > \varepsilon\right) < +\infty$. 由 Borel-Cantelli 引理, $\dfrac{D_n}{n^2} \xrightarrow{\text{a.s.}} 0$.

(2) 先证: $\dfrac{S_{n^2}}{n^2} \xrightarrow{\text{a.s.}} 0$. 对 $\forall \varepsilon > 0$, 利用 Markov 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2 n^4}\mathbb{E}[S_{n^2}^2] = \frac{1}{\varepsilon^2 n^4}\sum_{k=1}^{n^2}\mathbb{E}[X_k^2] \leqslant \frac{M}{\varepsilon^2 n^2} \implies \sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| > \varepsilon\right) < +\infty.$$

由 Borel-Cantelli 引理, $\dfrac{S_n}{n^2} \xrightarrow{\text{a.s.}} 0$.

回到一般: 对 $\forall k \in \mathbb{N}^*$, $\exists n \in \mathbb{N}^*$, 使得 $n^2 \leqslant k < (n+1)^2$. 因此

$$\sup_{n^2 \leqslant k<(n+1)^2} \left|\frac{S_k}{k}\right| \leqslant \frac{D_n}{n^2} + \frac{|S_{n^2}|}{n^2} \xrightarrow{\text{a.s.}} 0(n \to +\infty) \implies \frac{S_k}{k} \xrightarrow{\text{a.s.}} 0.$$

注 **一个致命的细节错误**: $\mathbb{E}(\sup\limits_{n^2 \leqslant k<(n+1)^2} |S_k - S_{n^2}|)^2 \leqslant \sup\limits_{n^2 \leqslant k<(n+1)^2} \mathbb{E}[|S_k - S_{n^2}|^2]$. 反例大家自行思考.

例题 1.55 设事件 $A_1, A_2 \cdots$ 两两独立, 且 $\sum\limits_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. 证明:

$$\frac{\sum\limits_{k=1}^{n} I_{A_k}}{\sum\limits_{k=1}^{n} \mathbb{P}(A_k)} \xrightarrow{\text{a.s.}} 1.$$

证明 记 $S_n=\sum_{k=1}^{n} I_{A_k}, \mathbb{E}[S_n]=\sum_{k=1}^{n}\mathbb{P}(A_k), \mathbb{E}[S_n]\uparrow+\infty$. 则对 $\forall\varepsilon>0$, 利用 Chebyshev 不等式, 有

$$\mathbb{P}\left(\left|\frac{S_n}{\mathbb{E}[S_n]}-1\right|>\varepsilon\right)\leqslant\frac{\mathrm{Var}(S_n)}{\varepsilon^2(\mathbb{E}[S_n])^2}=\frac{\sum_{k=1}^{n}\mathrm{Var}(I_{A_k})}{\varepsilon^2(\mathbb{E}[S_n])^2}\leqslant\frac{\sum_{k=1}^{n}\mathbb{E}[I_{A_k}]}{\varepsilon^2(\mathbb{E}[S_n])^2}=\frac{1}{\varepsilon^2\mathbb{E}[S_n]}.$$

定义子列 $\{n_k\}$, 满足 $n_k=\inf\{n:\mathbb{E}[S_n]\geqslant k^2\}$, 利用 $k^2\leqslant\mathbb{E}[S_{n_k}]\leqslant\mathbb{E}[S_{n_k-1}+I_{A_{n_k}}]<k^2+1$, 则有

$$\sum_{k=1}^{+\infty}\mathbb{P}\left(\left|\frac{S_{n_k}}{\mathbb{E}[S_{n_k}]}-1\right|>\varepsilon\right)\leqslant\sum_{k=1}^{+\infty}\frac{1}{\varepsilon^2\mathbb{E}[S_{n_k}]}\leqslant\sum_{k=1}^{+\infty}\frac{1}{\varepsilon^2k^2}<+\infty.$$

由 Borel-Cantelli 引理知 $\frac{S_{n_k}}{\mathbb{E}[S_{n_k}]}\xrightarrow{\text{a.s.}}1$.

回到一般: 对 $\forall n\in\mathbb{N}^*$, 存在 $k\in\mathbb{N}^*$, 使得 $n_k\leqslant n<n_{k+1}$. 注意到

$$\frac{S_{n_k}}{\mathbb{E}[S_{n_{k+1}}]}\leqslant\frac{S_n}{\mathbb{E}[S_n]}\leqslant\frac{S_{n_{k+1}}}{\mathbb{E}[S_{n_k}]}.$$

利用 $\frac{\mathbb{E}[S_{n_{k+1}}]}{\mathbb{E}[S_{n_k}]}\in\left[1,\frac{(k+1)^2+1}{k^2}\right]$ 知

$$\frac{S_{n_{k+1}}}{\mathbb{E}[S_{n_k}]}=\frac{S_{n_{k+1}}}{\mathbb{E}[S_{n_{k+1}}]}\cdot\frac{\mathbb{E}[S_{n_{k+1}}]}{\mathbb{E}[S_{n_k}]}\xrightarrow{\text{a.s.}}1(k\to+\infty),\qquad\frac{S_{n_k}}{\mathbb{E}[S_{n_{k+1}}]}\xrightarrow{\text{a.s.}}1.$$

夹逼得 $\frac{S_n}{\mathbb{E}[S_n]}\xrightarrow{\text{a.s.}}1$.

选"矩": 没有固定的套路, 但要学会大胆尝试!

考虑随机变量 $X_1,\cdots X_n$, 记 $S_n=\sum_{k=1}^{n}X_k$, 一般可以先中心化处理故不妨设它们零均值. 对 $\delta>0$, 往往题目待证 $\frac{S_n}{n^\delta}\xrightarrow{\text{P}}0$ 或 $\frac{S_n}{n^\delta}\xrightarrow{\text{a.s.}}0$, 进而会涉及到: 对 $\forall\varepsilon>0$, $\mathbb{P}\left(\left|\frac{S_n}{n^\delta}\right|>\varepsilon\right)$ 的估计问题. 常见的一种处理方法就是利用题干条件已知的矩信息, 通过 Markov 不等式解决. 具体为:

$$\mathbb{P}\left(\left|\frac{S_n}{n^\delta}\right|>\varepsilon\right)\leqslant\frac{\mathbb{E}[|S_n|^r]}{\varepsilon^r n^{\delta r}},r>0\text{待定}$$

我们只需要找到一个满足要求的正数 r(一般选整数), 若证依概率收敛, 说明 $\frac{\mathbb{E}[|S_n|^r]}{\varepsilon^r n^{\delta r}}\xrightarrow{n\to+\infty}0$ 即可; 若证几乎处处收敛, 利用 Borel-Cantelli 引理知, 只需说明 $\sum_{n=1}^{+\infty}\frac{\mathbb{E}[|S_n|^r]}{\varepsilon^r n^{\delta r}}<+\infty$ 即可.

难点就在于如何利用题干已知的矩信息选取 r, 我们给一个例子让大家自行体会:

例题 1.56 随机变量 $X_1,\cdots,X_n$ 独立同分布, $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=-1)=\frac{1}{2}$, 证明:

(1) $\frac{1}{n^2}\sum_{k=1}^{n}kX_k\xrightarrow{\text{a.s.}}0$;

(2) 对 $\forall\delta>0$, 均有 $\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{k=1}^{n}X_k\xrightarrow{\text{a.s.}}0$.

思路: (1) 记 $T_n=\sum_{k=1}^{n}kX_k$, 待定 $r\in\mathbb{N}^*$, 考察: $\frac{\mathbb{E}[|T_n|^r]}{n^{2r}}$, 看 $\sum_{n=1}^{+\infty}\frac{\mathbb{E}[|T_n|^r]}{n^{2r}}$ 是否收敛.

- $r=1$, $\frac{\mathbb{E}[|T_n|]}{n^2}\leqslant\frac{1}{n^2}\sum_{k=1}^{n}k\mathbb{E}[|X_k|]=\frac{n+1}{2n}$, 排除;
- $r=2$, $\frac{\mathbb{E}[T_n^2]}{n^4}=\frac{1}{n^4}\left(\sum_{k=1}^{n}k^2\mathbb{E}[X_k^2]+\sum_{1\leqslant i\neq j\leqslant n}ij\mathbb{E}[X_iX_j]\right)=\frac{1}{n^4}\sum_{k=1}^{n}k^2=O(n^{-1})$, 排除;

- $r=3$, $\frac{\mathbb{E}[|T_n|^3]}{n^6}\leqslant\frac{1}{n^6}\left(\sum_{k=1}^n k^3\mathbb{E}[|X_k|^3]+C_1\sum_{1\leqslant i\neq j\leqslant n}i^2j\mathbb{E}[|X_i|^2|X_j|]+\sum_{1\leqslant i\neq j\neq k\leqslant n}ijk\mathbb{E}[|X_i||X_j||X_k|]\right)=$ $\frac{1}{n^6}\left(\sum_{k=1}^n k^3+O(n^5)+O(n^6)\right)=O(1)$, 排除;
- $r=4$, $\frac{\mathbb{E}[T_n^4]}{n^8}=\frac{1}{n^8}\left(\sum_{k=1}^n k^4\mathbb{E}[X_k^4]+C_1\sum_{1\leqslant i\neq j\leqslant n}i^3j\mathbb{E}[X_i^3X_j]+C_2\sum_{1\leqslant i\neq j\leqslant n}i^2j^2\mathbb{E}[X_i^2X_j^2]\right)$ $+\frac{1}{n^8}\left(C_3\sum_{1\leqslant i\neq j\neq k\leqslant n}i^2jk\mathbb{E}[X_i^2X_jX_k]+\sum_{1\leqslant i\neq j\neq k\neq l\leqslant n}ijkl\mathbb{E}[X_iX_jX_kX_l]\right)$ $=\frac{1}{n^8}\left(\sum_{k=1}^n k^4+C_2\sum_{1\leqslant i\neq j\leqslant n}i^2j^2\right)\leqslant\frac{1}{n^8}\left(\sum_{k=1}^n k^4+C_2\sum_{1\leqslant i\leqslant n}i^2\sum_{1\leqslant j\leqslant n}j^2\right)=O(n^{-2})$, 成立.

证明 (1) 记 $T_n=\sum_{k=1}^n kX_k$, 结合思路, 对 $\forall\varepsilon>0$, 利用 Markov 不等式, 我们有

$$\sum_{n=1}^{+\infty}\mathbb{P}\left(\left|\frac{T_n}{n^2}\right|>\varepsilon\right)\leqslant\sum_{n=1}^{+\infty}\frac{\mathbb{E}[T_n^4]}{\varepsilon^4n^8}\leqslant\sum_{n=1}^{+\infty}\frac{C}{\varepsilon^4n^2}<+\infty,$$

其中 C 为大于 0 的某个常数. 利用 Borel-Cantelli 引理即得 $\frac{T_n}{n^2}\xrightarrow{\text{a.s.}}0$.

(2) 记 $S_n=\sum_{k=1}^n X_k$, 利用矩方法 (定理 1.18), 有

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]\xrightarrow{n\to+\infty}\gamma_k,\qquad\gamma_{2m-1}=0,\quad\gamma_{2m}=(2m-1)!!.$$

取正偶数 k, 使得 $\delta k>1$, 对 $\forall\varepsilon>0$, 当 n 充分大时, 有

$$\mathbb{P}\left(\left|\frac{S_n}{n^{\frac12+\delta}}\right|>\varepsilon\right)\leqslant\frac{1}{\varepsilon^kn^{\delta k}}\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right]\leqslant\frac{2\gamma_k}{\varepsilon^kn^{\delta k}}\Longrightarrow\sum_{n=1}^{+\infty}\mathbb{P}\left(\left|\frac{S_n}{n^{\frac12+\delta}}\right|>\varepsilon\right)\leqslant\sum_{n=1}^{+\infty}\frac{2\gamma_k}{\varepsilon^kn^{\delta k}}<+\infty.$$

利用 Borel-Cantelli 引理即得 $\frac{S_n}{n^{\frac12+\delta}}\xrightarrow{\text{a.s.}}0$.

思考: (1) 中阶数 n^{-2} 能否放得再更紧一些 (找到更精确的收敛速度)?

注 高阶矩展开往往会带有 “组合计数” 的想法在里面, 另外注意偶阶矩和奇阶矩处理时微妙的区别.

最后回到强大数律:

定理 1.32 (强大数律)

设 $X_1,\cdots,X_n$ 独立同分布, $\mu=\mathbb{E}[X_1]$ 且 $\mathbb{E}[|X_1|]<+\infty$, 记 $S_n=\sum_{k=1}^n X_k$, 则

$$\frac{S_n}{n}\xrightarrow{\text{a.s.}}\mu.$$

回顾证明思路: 先设 $X_i\geqslant 0$:

1. **截尾术**: 引入 $Y_n=X_nI_{\{X_n<n\}}=\begin{cases}X_n & X_n<n\\ 0 & X_n\geq n\end{cases}$, 则

$$\frac1n\sum_{i=1}^n X_i\xrightarrow{\text{a.s.}}\mu\Longleftrightarrow\frac1n\sum_{i=1}^n Y_i\xrightarrow{\text{a.s.}}\mu.$$

2. **子序列方法: 找一个 a.s. 收敛的子列**: 对 $\alpha > 1$, 令 $\beta_k = [\alpha^k]$, 则

$$\alpha^k \leq \beta_k < \alpha^k + 1, \quad \frac{\beta_{k+1}}{\beta_k} \to \alpha, k \to \infty$$

且

$$\frac{1}{\beta_n}\sum_{k=1}^{\beta_n} Y_k \xrightarrow{\text{a.s.}} \mu.$$

3. **a.s. 收敛对一般 n 都成立**: 设 $S_n' = \sum_{k=1}^{n} Y_k$, 则 S_n' 关于 n 单调增, 取 $\beta_m \leq n < \beta_{m+1}$, 由

$$\frac{S'_{\beta_m}}{\beta_{m+1}} \leq \frac{S'_n}{n} \leq \frac{S'_{\beta_{m+1}}}{\beta_m}$$

从而

$$\frac{\beta_m}{\beta_{m+1}}\frac{S'_{\beta_m}}{\beta_m} \leq \frac{S'_n}{n} \leq \frac{S'_{\beta_{m+1}}}{\beta_{m+1}}\frac{\beta_{m+1}}{\beta_m}$$

所以

$$\frac{1}{\alpha}\mu \leq \liminf_{n\to\infty}\frac{S'_n}{n} \leq \limsup_{n\to\infty}\frac{S'_n}{n} \leq \alpha\mu, \text{a.s.}$$

令 $\alpha \downarrow 1$ 知,

$$\lim_{n\to\infty}\frac{S'_n}{n} = \mu, \text{a.s.}$$

最后, $X_i = X_i^+ - X_i^-$ 即可.

注 实际上, 强大数律的证明就是这节讲过的各种技巧的组合: 截尾、子序列方法等.

✍ **练习 1.29 Grimmett 7.3.8, 7.4.2, 7.11.18, 7.11.19, 7.11.22**

✍ **练习 1.30** 在函数空间 $L^1(\Omega, \mathcal{F}, \mathbb{P})$ 上定义

$$d_1(X,Y) = \mathbb{E}[|X-Y|], \quad d_2(X,Y) = \mathbb{E}[|X-Y| \wedge 1], \quad d_3(X,Y) = \mathbb{E}\left[\frac{|X-Y|}{1+|X-Y|}\right]$$

(1) d_1, d_2, d_3 是 Ω 上的度量.

(2) $X_n \xrightarrow{1} X \iff d_1(X_n, X) \to 0$.

(3) $X_n \xrightarrow{\text{P}} X \iff d_2(X_n, X) \to 0 \iff d_3(X_n, X) \to 0$.

✍ **练习 1.31** 设对 $\forall i = 1, 2, \cdots, m$, 有 $X_{n,i} \xrightarrow{\text{P}} X_i$, 函数 $g: \mathbb{R}^m \to \mathbb{R}$ 连续, 则

$$g(X_{n_1}, \cdots, X_{n,m}) \xrightarrow{\text{P}} g(X_1, \cdots, X_m).$$

✍ **练习 1.32** 设 $\{X_n\}$ 为非负随机变量列, $\mathbb{E}[X_1] = \mu$, $S_n = \sum_{k=1}^{n} X_k$, 令 $N(t) = \max\{n : S_n \leqslant t\}$, 证明:

(1) $\mathbb{P}(\lim_{n\to+\infty} N(t) = \infty) = 1$.

(2) $\dfrac{N(t)}{t} \xrightarrow{\text{a.s.}} \dfrac{1}{\mu}$.

✍ **练习 1.33** 设 $\{X_n\}$ 独立同服从参数为 p 的 Bernoulli 分布, 令 $Y_k = X_k X_{k+1}$, $S_n = \sum_{k=1}^{n} Y_k$. 证明:

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} p^2.$$

第 2 章 进阶内容

2.1 乘积空间浅引

问题: 同时抛掷 2 枚硬币与抛掷 1 枚硬币两次的概率空间的关系?

事实: $\forall i \in I, \mathcal{F}_i \subset 2^{\Omega}$ 为 σ 代数, 则 $\bigcap\limits_{i\in I} \mathcal{F}_i$ 亦为 σ 代数.

例题 2.1 $\Omega = \{1, 2, \cdots, n\}, A_i = \{i\}(n > 2), \mathcal{F}_i = \{\emptyset, \Omega, A_i, A_i^c\}$, 则 $\mathcal{F}_1 \cup \mathcal{F}_2 = \{\emptyset, \Omega, A_1, A_1^c, A_2, A_2^c\}$, 明显 $A_1 \cup A_2 = \{1, 2\} \notin \mathcal{F}_1 \cup \mathcal{F}_2$, 非 σ 代数.

由小 σ 代数到大 σ 代数, 方式之一: 对 $\mathcal{F}, \mathcal{G} \subset 2^{\Omega}$, 称包含 $\mathcal{F}, \mathcal{G}$ 的最小 σ 代数为 $\mathcal{F}$ 与 $\mathcal{G}$ 生成的 σ 代数, 记为 $\sigma(\mathcal{F} \cup \mathcal{G})$.

典型例子: 1 维 Borel 域: $\mathbb{R}$ 上形如 $(a, b]$ 区间生成的 σ 代数, 记为 $\mathcal{B}(\mathbb{R})$, 其中每个元素称为 Borel 集. $\{b\} = \bigcap\limits_{n} \left(b - \dfrac{1}{n}, b\right], (a, b) = (a, b]\backslash\{b\}, [a, b] = \{a\} \cup (a, b], [a, b) = \{a\} \cup (a, b)$. 即一切开区间, 闭区间, 半开半闭区间均为 Borel 集.

n 维 Borel 域: $\mathbb{R}^n$ 上形如 $(a_1, b_1] \times \cdots \times (a_n, b_n]$ 生成的 σ 域, 记为 $\mathcal{B}(\mathbb{R}^n)$

乘积空间: $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ 构造更大空间?

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

$$\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_i \in \mathcal{F}_i, i = 1, 2\}$$

未必为 σ 代数.

例题 2.2 接例子(2.1). $\mathcal{F}_1 \times \mathcal{F}_2 = \{\emptyset, \Omega \times \Omega, A_1 \times A_2^c, A_1^c \times A_2, A_1^c \times A_2^c, \cdots\}$, 显见 $A_1 \times A_2 \in \mathcal{F}_1 \times \mathcal{F}_2$, 但

$$\begin{aligned}
(A_1 \times A_2)^c &= \Omega \times \Omega \backslash \{1, 2\} \\
&= \{2, 3, \cdots, n\} \times \Omega \cup \Omega \times \{1, 3, 4, \cdots, n\} \\
&\notin \mathcal{F}_1 \times \mathcal{F}_2
\end{aligned}$$

记 $\mathcal{F} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ 看作 $\Omega = \Omega_1 \times \Omega_2$ 的 σ 代数, 概率测度? 引入 $\mathcal{F}_1 \times \mathcal{F}_2$ 上函数

$$\mathbb{P}_{12} : \mathcal{F}_1 \times \mathcal{F}_2 \to \mathbb{R}, \mathbb{P}_{12}(A_1 \times A_2) := \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$$

特别地, $\mathcal{F}_1 \times \mathcal{F}_2$ 中元素的不交并, 若 Ω_1, Ω_2 有限, $\mathcal{F}_i = 2^{\Omega_i}$, 这时 $\mathbb{P}_{12}$ 良定.

对应: $A_i \in \mathcal{F}_i, A_1 \longleftrightarrow A_1 \times \Omega_2, A_2 \longleftrightarrow \Omega_1 \times A_2$, 在 $(\Omega_1 \times \Omega_2, \sigma(\mathcal{F}_1 \times \mathcal{F}_2), \mathbb{P}_{12})$ 看 $A_1 \times A_2$

$$\mathbb{P}_{12}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) = \mathbb{P}_{12}(A_1 \times \Omega_2) \cdot \mathbb{P}_{12}(\Omega_1 \times A_2)$$

表明 $A_1 \times \Omega_2$ 与 $\Omega_1 \times A_2$ 独立.

例题 2.3 掷硬币 2 次. $\Omega = \{HH, HT, TH, TT\} = \{H, T\} \times \{H, T\}$

$$\mathbb{P}_{12}(HH) = \mathbb{P}_1(H)\mathbb{P}_2(H) = \frac{1}{4}, \mathbb{P}_{12}(HT) = \frac{1}{4}.$$

2.2 互素概率问题浅谈

先看这样一个简单的例子:

例题 2.4 对 $N \geqslant 1$, 记 $\mathbb{P}_N$ 为 $\Omega_N = \{1, 2, \cdots, N\}$ 上均匀概率测度, 通过模 q 的余数可定义随机变量

$$\pi_q : \Omega_N \to \mathbb{Z}_q = \{0, 1, \cdots q-1\}.$$

对两个不同的素数 q_1 和 q_2, 证明 π_{q_1} 和 π_{q_2}渐近独立, 即 $\forall a_i \in \mathbb{Z}_{q_i}$, 有

$$\lim_{N\to+\infty} \mathbb{P}_N(\pi_{q_1} = a_1, \pi_{q_2} = a_2) = \lim_{N\to+\infty} \mathbb{P}_N(\pi_{q_1} = a_1) \times \lim_{N\to+\infty} \mathbb{P}_N(\pi_{q_2} = a_2).$$

证明 我们有

$$\mathbb{P}_N(\pi_{q_i} = a_i) = \begin{cases} \left[\dfrac{N}{q_i}\right] \cdot \dfrac{1}{N}, & a_i = 0, \\ \left(\left[\dfrac{N-a_i}{q_i}\right] + 1\right) \cdot \dfrac{1}{N}, & 1 \leqslant a_i \leqslant q-1, \\ 0, & \text{else.} \end{cases}$$

因为 q_1, q_2 均为素数, 利用中国剩余定理可知, $\exists 0 \leqslant a_3 \leqslant q_1 q_2 - 1$, 使得

$$\{\pi_{q_1} = a_1, \pi_{q_2} = a_2\} = \{\pi_{q_1 q_2} = a_3\}.$$

因此

$$\mathbb{P}_N(\pi_{q_1} = a_1, \pi_{q_2} = a_2) = \left[\frac{N-a_3}{q_1 q_2}\right] \cdot \frac{1}{N} \text{ or } \left(\left[\frac{N-a_3}{q_1 q_2}\right] + 1\right) \cdot \frac{1}{N}$$

若 $a_i \notin [0, q_i - 1]$, 则题中两式等于 0 显然成立; 若 $a_i \in [0, q_i - 1]$, 对 $N \to +\infty$, 则题中两式等于 $\dfrac{1}{q_1 q_2}$, 亦成立.

如果我们考虑整个正整数集 $\Omega = \{1, 2, \cdots\}$, 则概率测度 $\mathbb{P}$ 不能定义为 Ω 上的均匀测度, 因为如果成立, 我们在 $\mathbb{P}$ 下随机取一个正整数 k, 记 $\mathbb{P}(\{k\}) = a \geqslant 0$. 因为 $\mathbb{P}$ 满足可列可加性, 故有

$$\mathbb{P}(\Omega) = \sum_{k=1}^{\infty} \mathbb{P}(\{k\}) = 0 \text{ or } +\infty,$$

与概率的规范性 $\mathbb{P}(\Omega) = 1$ 矛盾.

因此我们要对测度 $\mathbb{P}$ 重新定义, 比如下面这个例子, 顺手还得到一些有意思的结果:

例题 2.5 对正整数集 $\Omega = \{1, 2, \cdots\}$, 引入概率测度

$$\mathbb{P}(\{n\}) = \frac{1}{\zeta(s) \cdot n^s}, \quad \zeta(s) = \sum_{n=1}^{\infty} n^{-s}, s > 1.$$

对正整数 q, $A_q = \{mq : m \in \Omega\}$.

(1) 对任意不同的素数 $p_1, p_2, \cdots, p_t$, 证明 $A_{p_1}, A_{p_2}, \cdots A_{p_t}$ 相互独立.

(2) 用概率方法证明欧拉公式:

$$\zeta(s) = \prod_{i=1}^{\infty} \left(1 - \frac{1}{p_i^s}\right)^{-1}.$$

其中所有素数 $p_1 < p_2 < \cdots$.

(3) 在概率测度 $\mathbb{P}$ 下, 从所有正整数中独立地随机选取两数 a, b, 证明 a, b 互素的概率 $\dfrac{1}{\zeta(2s)}$.(可推广到任意 n 个数互素)

注 $\sum_{n=1}^{+\infty} n^{-s}$ 收敛保证了概率测度 $\mathbb{P}$ 的良定.

证明 (1) 这里仅考虑 $\mathbb{P}(A_{p_1}\cdots A_{p_t})=\prod_{k=1}^{n}\mathbb{P}(A_{p_k})$. 注意到

$$\mathbb{P}(A_{p_i})=\sum_{k=1}^{\infty}\frac{1}{\zeta(s)\cdot(kp_i)^s}=\frac{1}{\zeta(s)p_i^s}\sum_{k=1}^{\infty}\frac{1}{k^s}=\frac{1}{p_i^s}$$

及

$$\mathbb{P}(A_{p_1}\cdots A_{p_t})=\sum_{k=1}^{\infty}\frac{1}{\zeta(s)\cdot(kp_1p_2\cdots p_t)^s}=\frac{1}{\zeta(s)(p_1p_2\cdots p_t)^s}\sum_{k=1}^{\infty}\frac{1}{k^s}=\frac{1}{(p_1p_2\cdots p_t)^s}$$

(2) 注意到

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{+\infty}A_{p_i}\right)&=1-\mathbb{P}\left(\bigcap_{i=1}^{\infty}A_{p_i}^c\right)=1-\lim_{n\to+\infty}\mathbb{P}\left(\bigcap_{i=1}^{n}A_{p_i}^c\right)\\&=1-\lim_{n\to+\infty}\prod_{i=1}^{n}\mathbb{P}(A_{p_i}^c)=1-\prod_{i=1}^{+\infty}(1-\mathbb{P}(A_{p_i}))\\&=1-\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^s}\right).\end{aligned}$$

而

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty}A_{p_i}\right)=\mathbb{P}(\{1\}^c)=1-\frac{1}{\zeta(s)},$$

两式联立后即得欧拉公式.

(3) 记事件 B_{p_j} 表示 b 能被 p_j 整除, 则有 $\mathbb{P}(A_{p_k}B_{p_l})=\mathbb{P}(A_{p_k})\mathbb{P}(B_{p_l})$. 记事件 C 表示 a,b 互素, 则有

$$C=\left(\bigcup_{i=1}^{\infty}(A_{p_i}\cap B_{p_i})\right)^c=\bigcap_{i=1}^{\infty}(A_{p_i}\cap B_{p_i})^c.$$

因此

$$\mathbb{P}(C)=\prod_{i=1}^{\infty}\mathbb{P}((A_{p_i}\cap B_{p_i})^c)=\prod_{i=1}^{\infty}(1-\mathbb{P}(A_iB_i))=\prod_{i=1}^{\infty}(1-\mathbb{P}(A_i)\mathbb{P}(B_i))=\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^{2s}}\right)=\frac{1}{\zeta(2s)}.$$

练习 2.1 对 $N\in\mathbb{N}^*$, 记 $\mathbb{P}_N$ 为 $\Omega_N=\{1,2,\cdots,N\}$ 上均匀概率测度, 在 $\mathbb{P}_N$ 下从 Ω_N 中独立地随机选取两数 a,b, 记 a,b 互素的概率为 P_N. 则当 $N\to+\infty$ 时, $P_N\to\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^2}\right)=\frac{6}{\pi^2}$. 其中 p_i 为素数且依序从小到大排列.

相关链接: https://en.wikipedia.org/wiki/Coprime_integers#Probabilities

2.3 概率方法举例

例题 2.6 (涂色问题) 平面上的 n 个点和连接各点之间的连线叫做一个完全图, 记作 G. 点称作图的顶点, 顶点之间的连线叫做边, 共有 $\binom{n}{2}$ 条. 给定一个整数 k, G 中任意 k 个顶点连同相应的边构成一个有 k 个顶点的完全子图, G 中共有 $\binom{n}{k}$ 个这样的子图, 记作 $G_i, i=1,\cdots,\binom{n}{k}$. 现将图 G 的每条边涂成红色或蓝色. 当 $\binom{n}{k}<2^{\frac{k(k-1)}{2}-1}$ 时, 问: 是否有一种涂色方法, 使得没有一个子图 G_i 的 $\binom{k}{2}$ 条边是同一颜色的.

解 我们对 G 的边进行随机涂色, 每条边为红色和蓝色的概率是 $\frac{1}{2}$. 记事件 A_i 表示子图 G_i 各边的颜色

相同, 则 $\bigcap_{i=1}^{\binom{n}{k}} A_i^c$ 表示没有一个子图 G_i 的 $\binom{k}{2}$ 条边是同一颜色的. 利用

$$\mathbb{P}(A_i)=\mathbb{P}(G_i\text{的各边均为红色})+\mathbb{P}(G_i\text{的各边均为蓝色})=\frac{2}{2^{\binom{k}{2}}}=2^{-\frac{k(k-1)}{2}+1}$$

及 $\binom{n}{k}<2^{\frac{k(k-1)}{2}-1}$, 有

$$\mathbb{P}\left(\bigcup_{i=1}^{\binom{n}{k}} A_i\right)\leqslant\sum_{i=1}^{\binom{n}{k}}\mathbb{P}(A_i)=\binom{n}{k}2^{-\frac{k(k-1)}{2}+1}<1$$

故 $\mathbb{P}\left(\bigcap_{i=1}^{\binom{n}{k}} A_i^c\right)>0$. 从而存在一种涂色方法, 使得没有一个子图 G_i 的 $\binom{k}{2}$ 条边是同一颜色的.

例题 2.7 (Balancing vectors) 设向量 $\boldsymbol{v_1},\boldsymbol{v_2},\cdots,\boldsymbol{v_n}\in\mathbb{R}^n$.

(1) 若对任意 $1\leqslant i\leqslant n$ 满足 $|\boldsymbol{v_i}|=1$. 证明: 存在 $\varepsilon_i\in\{-1,1\}$, 使得

$$\left|\sum_{i=1}^n\varepsilon_i\boldsymbol{v_i}\right|\leqslant\sqrt{n}\ (\text{这里换成}\geqslant\text{亦成立}).$$

(2) 若对任意 $1\leqslant i\leqslant n$ 满足 $|\boldsymbol{v_i}|\leqslant 1$. 对 $\boldsymbol{w}=\sum_{i=1}^n p_i\boldsymbol{v_i}$, 其中 $p_i\in[0,1]$. 证明: 存在 $\varepsilon_i\in\{0,1\}$, 使得

$$\left|\sum_{i=1}^n\varepsilon_i\boldsymbol{v_i}-\boldsymbol{w}\right|\leqslant\frac{\sqrt{n}}{2}.$$

证明 (1) 令 ε_i i.i.d 且满足 $\mathbb{P}(\varepsilon_i=1)=\mathbb{P}(\varepsilon_i=-1)=\frac{1}{2}$. 记 $X=\left|\sum_{i=1}^n\varepsilon_i\boldsymbol{v_i}\right|$. 则有

$$\begin{aligned}\mathbb{E}[X^2]&=\sum_{1\leqslant i,j\leqslant n}\boldsymbol{v}_i\cdot\boldsymbol{v}_j\mathbb{E}[\varepsilon_i\varepsilon_j]=\sum_{i=1}^n\boldsymbol{v}_i\cdot\boldsymbol{v}_i\mathbb{E}[\varepsilon_i^2]+\sum_{1\leqslant i\neq j\leqslant n}\boldsymbol{v}_i\cdot\boldsymbol{v}_j\mathbb{E}[\varepsilon_i]\mathbb{E}[\varepsilon_j]=\sum_{i=1}^n|\boldsymbol{v}_i|^2\\&=n.\end{aligned}$$

因此, 一定存在 $\varepsilon_i\in\{-1,1\}$, 使得

$$\left|\sum_{i=1}^n\varepsilon_i\boldsymbol{v_i}\right|\leqslant\sqrt{n}.$$

(2) 令 ε_i 相互独立且满足 $\mathbb{P}(\varepsilon_i=1)=p_i,\mathbb{P}(\varepsilon_i=0)=1-p_i$. 记 $X=\left|\sum_{i=1}^n\varepsilon_i\boldsymbol{v_i}-\boldsymbol{w}\right|$. 注意到

$$X^2=\left|\sum_{i=1}^n(\varepsilon_i-p_i)\boldsymbol{v_i}\right|^2=\sum_{1\leqslant i,j\leqslant n}\boldsymbol{v}_i\cdot\boldsymbol{v}_j(\varepsilon_i-p_i)(\varepsilon_j-p_j).$$

因此有

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{1\leqslant i,j\leqslant n} \boldsymbol{v}_i\cdot\boldsymbol{v}_j\mathbb{E}[(\varepsilon_i-p_i)(\varepsilon_j-p_j)]\\ &= \sum_{i=1}^n \boldsymbol{v}_i\cdot\boldsymbol{v}_i\mathbb{E}[(\varepsilon_i-p_i)^2]+\sum_{1\leqslant i\neq j\leqslant n}\boldsymbol{v}_i\cdot\boldsymbol{v}_j\mathbb{E}[(\varepsilon_i-p_i)]\mathbb{E}[(\varepsilon_j-p_j)]\\ &= \sum_{i=1}^n|\boldsymbol{v}_i|^2\mathrm{Var}(\varepsilon_i)\leqslant\sum_{i=1}^n\frac{|\boldsymbol{v}_i|^2}{4}\\ &\leqslant \frac{n}{4}.\end{aligned}$$

故一定存在 $\varepsilon_i\in\{0,1\}$, 使得

$$\left|\sum_{i=1}^n\varepsilon_i\boldsymbol{v_i}-\boldsymbol{w}\right|\leqslant\frac{\sqrt{n}}{2}.$$

例题 2.8 (Erdös, 1965. 组合数论) 证明: 对每个非零整数集合 $B=\{b_1,\cdots,b_n\}$, 都存在一个 sum-free 子集 A, 使得 $|A|>\frac{n}{3}$. 其中 **sum-free** 集 A 指的是: 不存在 $a_1,a_2,a_3\in A$, 使得 $a_1+a_2=a_3$.

证明 令 p 为 $3k+2$ 型素数且满足 $p>2\max\limits_{1\leqslant i\leqslant n}|b_i|$, 设 $C=\{k+1,k+2,\cdots,2k+1\}$, 可知 C 是 $\mathbb{Z}_p$ 的 sum-free 子集且满足

$$\frac{|C|}{p-1}=\frac{k+1}{3k+1}>\frac{1}{3}.$$

现在 $\{1,2,\cdots,p-1\}$ 中随机均匀选取整数 x, 并定义 $d_i\equiv xb_i(\mathrm{mod}\ p)$, $1\leqslant d_i\leqslant p-1$. 我们知道, 对固定的 b_i, 让 x 取遍 $1,2,\cdots,p-1$, 则 d_i 取遍 $\mathbb{Z}_p$ 中的非零元, 因此

$$\mathbb{P}(\{d_i\in C\})=\frac{|C|}{p-1}>\frac{1}{3}.$$

所以 $|B|$ 中满足 $\{d_i\in C\}$ 所对应的元素 b_i 的总数的期望值大于 $\frac{n}{3}$. 因此, 存在 $1\leqslant x\leqslant p-1$, 使得 B 中有一个子集 A 满足 $|A|>\frac{n}{3}$ 且

$$xa\equiv(\mathrm{mod}\ p)\in C,\ \forall a\in A.$$

最后证明 A 是 sum-free 的. 否则, 若存在 $a_1,a_2,a_3\in A$, 使得 $a_1+a_2=a_3$, 则有

$$xa_1+xa_2\equiv xa_3(\mathrm{mod}\ p),$$

这与 C 是 $\mathbb{Z}_p$ 的 sum-free 子集相矛盾. 这样我们就完成了证明.

2.4 一维简单随机游走的双吸收壁模型及其应用

在课堂上我们已解决如下问题:

问题: 粒子在 $t=0$ 时处于 $x=k$, 在 $x=a,b$ 处各有一个吸收壁 (移动到 a 或 b 处时停止). 现在粒子进行一维简单随机游走, 求粒子在 $x=a,b$ 处吸收概率. $(a<k<b)$

解 记 p_k为初始处于$x=k$且最后在$x=a$吸收概率, p_k^*为初始处于$x=k$且最后在$x=b$吸收概率, 第一步之后全概率公式

$$p_k=pp_{k+1}+qp_{k-1}, a<k<b$$

边界条件: $p_a=1, p_b=0$, 记 $r=\frac{q}{p}$, 则

$$p_{k+1}-p_k=r(p_k-p_{k-1})$$

解之

$$p_k = \begin{cases} \dfrac{r^{k-a} - r^{b-a}}{1 - r^{b-a}}, & r \neq 1, \\ 1 - \dfrac{k-a}{b-a}, & r = 1. \end{cases}$$

$$p_k^* = \begin{cases} \dfrac{1 - r^{k-a}}{1 - r^{b-a}}, & r \neq 1, \\ \dfrac{k-a}{b-a}, & r = 1. \end{cases}$$

这就是一维简单随机游走的双吸收壁模型, 下面我们继续探究其性质.

例题 2.9 双吸收壁模型下, 记 T 为粒子初始状态在 k 下被吸收时经过的总时间, 求 $\mathbb{E}_k[T]$.

解 取条件期望, 有

$$\mathbb{E}_k[T] = p(\mathbb{E}_{k+1}[T] + 1) + q(\mathbb{E}_{k-1}[T] + 1) = p\mathbb{E}_{k+1}[T] + q\mathbb{E}_{k-1}[T] + 1.$$

记 $r = \dfrac{q}{p}$, 化简得

$$\mathbb{E}_{k+1}[T] - \mathbb{E}_k[T] = r(\mathbb{E}_k[T] - \mathbb{E}_{k-1}[T]) - \frac{1}{p}$$

再利用 $0 = \mathbb{E}_b[T] - \mathbb{E}_a[T] = \sum\limits_{k=a}^{b-1}(\mathbb{E}_{k+1}[T] - \mathbb{E}_k[T])$, 可得

$$\mathbb{E}_{k+1}[T] - \mathbb{E}_k[T] = \begin{cases} \dfrac{1}{p}\left(\dfrac{r^{k-a}(b-a)}{1 - r^{b-a}} + \dfrac{1}{r-1}\right), & r \neq 1, \\ b + a - 2k - 1, & r = 1. \end{cases}$$

所以

$$\mathbb{E}_k[T] = \mathbb{E}_k[T] - \mathbb{E}_a[T] = \sum_{t=a}^{k-1}(\mathbb{E}_{t+1}[T] - \mathbb{E}_t[T]) = \begin{cases} \dfrac{1}{q-p}\left[(k-a) - (b-a)\left(\dfrac{1 - r^{k-a}}{1 - r^{b-a}}\right)\right], & r \neq 1, \\ (k-a)(b-a-k), & r = 1. \end{cases}$$

例题 2.10 一维简单随机游走 $S_n, S_0 = 0$. 记 $\tau_m = \inf\{n \geqslant 1 : S_n = m\}$. 证明:

$$\mathbb{P}(\tau_1 < +\infty) = \begin{cases} 1, & r \leqslant 1, \\ \dfrac{1}{r}, & r > 1. \end{cases} \qquad \mathbb{E}[\tau_1] = \begin{cases} +\infty, & r \geqslant 1, \\ \dfrac{1}{2p-1}, & r < 1. \end{cases}$$

这道题其实是单吸收壁, 但我们可以在 $x = -n, 1 (n \in \mathbb{N}^*)$ 两处均放置吸收壁, 然后让 $n \to +\infty$ 来进行转化. 双吸收壁模型下记 $p^{(n)}$ 为最后在 $x = 1$ 处吸收概率, $T^{(n)}$ 为粒子被吸收时经过的总时间. 注意到

$$\lim_{n \to +\infty} p^{(n)} = \mathbb{P}(\tau_1 < +\infty), \qquad \lim_{n \to +\infty} \mathbb{E}[T^{(n)}] = \mathbb{E}[\tau_1],$$

这里单调性 (单调收敛定理) 保证其成立. 再结合上述两问题即可.

注 由此可计算 $\mathbb{P}(\tau_0), \mathbb{E}[\tau_0]$, 另外请读者自行验证: $\tau_2 - \tau_1$ 与 τ_1 独立同分布 (在 $\tau_1 < +\infty$ 意义下). 可类推得 $\tau_{n+1} - \tau_n, \tau_n - \tau_{n-1}, \cdots, \tau_2 - \tau_1, \tau_1$ 独立同分布 (在 $\tau_n < +\infty$ 意义下). 即为强马氏性. 由此可得

$$\mathbb{P}(\tau_n < +\infty) = \begin{cases} 1, & r \leqslant 1, \\ \dfrac{1}{r^n}, & r > 1. \end{cases} \qquad \mathbb{E}[\tau_n] = \begin{cases} +\infty, & r \leqslant 1, \\ \dfrac{n}{2p-1}, & r > 1. \end{cases}$$

因为

$$\begin{aligned}\mathbb{P}(\tau_n<\infty)&=\mathbb{P}(\tau_n<\infty,\tau_{n-1}<\infty)=\mathbb{P}(\tau_n-\tau_{n-1}<\infty,\tau_{n-1}<\infty)=\mathbb{P}(\tau_n-\tau_{n-1}<\infty)\mathbb{P}(\tau_{n-1}<\infty)\\&=\mathbb{P}(\tau_1<\infty)\mathbb{P}(\tau_{n-1}<\infty)\\&\ \vdots\\&=\mathbb{P}(\tau_1<\infty)^k.\end{aligned}$$

及

$$\mathbb{E}[\tau_n]=\mathbb{E}\left[\sum_{i=2}^{n}(\tau_i-\tau_{i-1})+\tau_1\right]=\sum_{i=2}^{n}\mathbb{E}\left[[\tau_i-\tau_{i-1}]+\mathbb{E}[\tau_1]\right]=n\mathbb{E}[\tau_1].$$

例题 2.11 考虑一质点, 它沿着按一个圆周排列的标以 $0,1,\cdots,m$ 的 $m+1$ 个节点移动. 在每一步质点等概率按顺时针或逆时针方向移动至下一个位置. 现在质点从 0 出发按上述规则移动, 直到节点 $1,2,\cdots,m$ 均被访问过位置. 求最后一个被访问的节点是 $i(1\leqslant i\leqslant m)$ 的概率.

解 考虑首次访问 i 的相邻两节点 $i-1,i+1$ 中之一 (该事件为必然事件).

若节点 $i-1$ 先于 $i+1$ 被访问, 因为 $i,i+1$ 均没有被访问, 所以 i 为最后一个访问点等价于 $i+1$ 在 i 之前被访问, 而其等价于一维简单对称随机游走中质点从 $x=1$ 出发并在 $x=0,m$ 两处均放置吸收壁的双吸收壁模型上求质点被 m 吸收的概率 (可以将节点 i 为头, 节点 $i+1$ 为尾的 $m+1$ 个点圆弧拉成一条线段), 而问题中已求得其概率为 $\dfrac{1}{m}$.

若节点 $i+1$ 先于 $i-1$ 被访问, 由对称性同理可知其概率为 $\dfrac{1}{m}$. 利用全概率公式即得最后一个被访问的节点是 $i(1\leqslant i\leqslant m)$ 的概率是 $\dfrac{1}{m}$.

2.5 $\mathbb{Z}^d$ 上简单对称随机游走的常返性

一个醉汉从某点出发, 他总有一个时刻能回到该点. 但是给鸟打一针醉药后, 它做不到. 当然我们需要假设他们醉到天花板, 一直游走且不会 "寿终". 该问题即为 $\mathbb{Z}^d$ 上简单对称随机游走的常返性问题.

例题 2.12 一质点从原点出发, 在 $\mathbb{Z}^d$ 上做简单对称随机游走, 若事件 "质点从原点出发后存在某个时刻回到原点" 发生的概率为 1, 则称该随机游走是常返的. 试探究 $\mathbb{Z}^d$ 上简单对称随机游走的常返性与维数 d 的关系.

解 我们先证明如下引理:

引理 2.1

条件同上. 记 $p_{00}^{(n)}$ 表示从原点出发的质点, 经过 n 步后又回到原点的概率, 则该随机游走是常返的当且仅当

$$\sum_{n=1}^{\infty}p_{00}^{(n)}=+\infty.$$

证明 定义 $N_0=\sum\limits_{n=1}^{\infty}I_{\{S_n=0\}}$ 表示质点经过点 0 的次数, $\tau_m^k=\inf\left\{n\geqslant\tau_m^{k-1}:S_n=m\right\}$, $\tau_m^0=0$, ρ_0 表示事件 "之后存在某个时刻回到原点" 发生的概率, 显然 $0\leqslant\rho_0\leqslant 1$. 由定义知, $\rho_0=1$ 等价于该随机游走常返. 由定义知, $\mathbb{P}(\tau_0^1-\tau_0^0<\infty)=\rho_0$, 同时我们有 $\tau_m^{k+1}-\tau_m^k$ 与 $\tau_m^k-\tau_m^{k-1}$ 独立同分布 (即为强马氏性, 感兴趣

者自行验证), 故 $\tau_m^{k+1}-\tau_m^k \stackrel{d}{=\!=} \tau_m^1-\tau_m^0$. 由此可推得

$$\begin{aligned}\mathbb{P}(N_0\geqslant k)&=\mathbb{P}(\tau_0^k<\infty)=\mathbb{P}(\tau_0^k<\infty,\tau_0^{k-1}<\infty)=\mathbb{P}(\tau_0^k-\tau_0^{k-1}<\infty,\tau_0^{k-1}<\infty)\\&=\mathbb{P}(\tau_0^k-\tau_0^{k-1}<\infty)\mathbb{P}(\tau_0^{k-1}<\infty)=\mathbb{P}(\tau_0^1-\tau_0^0<\infty)\mathbb{P}(\tau_0^{k-1}<\infty)\\&=\rho_0\mathbb{P}(\tau_0^{k-1}<\infty)\\&\ \vdots\\&=\rho_0^k.\end{aligned}$$

利用 Fubini 定理, 我们有

$$\sum_{n=1}^{\infty}p_{00}^{(n)}=\sum_{n=1}^{\infty}\mathbb{E}_0[I_{\{S_n=0\}}]=\mathbb{E}_0[N_0]=\sum_{k=1}^{\infty}\mathbb{P}(N_0\geqslant k)=\sum_{k=1}^{\infty}\rho_0^k.$$

而

$$\sum_{k=1}^{\infty}\rho_0^k=\begin{cases}+\infty, & \rho_0=1,\\ \dfrac{1}{1-\rho_0}<\infty, & 0\leqslant\rho_0<1.\end{cases}$$

故引理得证.

回到原题. 利用 Stirling 公式 $n!\sim\sqrt{2n\pi}\left(\frac{n}{\mathrm{e}}\right)^n$ $(n\to\infty)$, 当 $d=1$ 时, 有 $p_{00}^{(2n)}=\frac{1}{4^n}\binom{2n}{n}=O(n^{-\frac12})$, $d=2$ 时, 若质点在平面沿上下左右方向游走后回到原点, 则其向上与向下的步数相等, 向左与向右的步数相等, 且向上和向右的步数和为 n. 故有

$$p_{00}^{(2n)}=\frac{1}{4^{2n}}\sum_{k=0}^{n}\frac{((2n)!)}{k!k!(n-k)!(n-k)!}=\frac{1}{4^{2n}}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}^2=\frac{1}{16^n}\binom{2n}{n}^2=O(n^{-1}).$$

这里 $\sum_{k=0}^{n}\binom{n}{k}^2=\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k}=\binom{2n}{n}$. $d\geqslant 3$ 时, 我们有

$$\begin{aligned}p_{00}^{(2n+1)}&=0,\\ p_{00}^{(2n)}&=\sum_{\substack{i_1,\cdots,i_{d-1}\geqslant 0\\ i_1+\cdots+i_{d-1}\leqslant n}}\frac{(2n)!}{(i_1!)^2(i_2!)^2\cdots(i_{d-1}!)^2((n-i_1-\cdots-i_{d-1})!)^2}(2d)^{-2n}\\&=2^{-2n}\binom{2n}{n}\sum_{\substack{i_1,\cdots,i_{d-1}\geqslant 0\\ i_1+\cdots+i_{d-1}\leqslant n}}\left(d^{-n}\frac{n!}{i_1!i_2!\cdots i_{d-1}!(n-i_1-\cdots-i_{d-1})!}\right)^2\\&\leqslant 2^{-2n}\binom{2n}{n}\max_{\substack{i_1,\cdots,i_{d-1}\geqslant 0\\ i_1+\cdots+i_{d-1}\leqslant n}}\left(d^{-n}\frac{n!}{i_1!i_2!\cdots i_{d-1}!(n-i_1-\cdots-i_{d-1})!}\right)\\&\times\sum_{\substack{i_1,\cdots,i_{d-1}\geqslant 0\\ i_1+\cdots+i_{d-1}\leqslant n}}d^{-n}\frac{n!}{i_1!i_2!\cdots i_{d-1}!(n-i_1-\cdots-i_{d-1})!}\end{aligned}$$

一方面,

$$\sum_{\substack{i_1,\cdots,i_{d-1}\geqslant 0\\ i_1+\cdots+i_{d-1}\leqslant n}}\frac{n!}{i_1!i_2!\cdots i_{d-1}!(n-i_1-\cdots-i_{d-1})!}=(\underbrace{1+1+\cdots+1}_{d个})^n=d^n.$$

另一方面, 存在 $C_1>0$, 使得

$$\max_{\substack{i_1,\cdots,i_{d-1}\geqslant 0\\ i_1+\cdots+i_{d-1}\leqslant n}}\left(d^{-n}\frac{n!}{i_1!i_2!\cdots i_{d-1}!(n-i_1-\cdots-i_{d-1})!}\right)\leqslant C_1d^{-n}\frac{n!}{(\frac{n}{d}!)^d}$$

利用 Stirling 公式, 我们有

$$C_1 d^{-n}\frac{n!}{(\frac{n}{d}!)^d}\xrightarrow{n\to\infty} d^{-n}\frac{n^n}{((\frac{n}{d})^{n/d})^d}\sqrt{\frac{n}{(\frac{n}{d})^d}}\frac{1}{(2\pi)^{\frac{d-1}{2}}}\leqslant C_2 n^{-\frac{d-1}{2}},$$

其中 C_2 为大于 0 的常数. 再由 Stirling 公式知,

$$p_{00}^{(2n)}=C_2 2^{-2n}\binom{2n}{n}n^{-\frac{d-1}{2}}\leqslant C_3 n^{-\frac{d}{2}},$$

其中 C_3 为大于 0 的常数. 所以

$$p_{00}^{(n)}=O(n^{-\frac{d}{2}})\quad(n\to\infty).$$

因此

$$\sum_{n=1}^{\infty}p_{00}^{(n)}=\begin{cases}+\infty, & d=1,2,\\ <\infty, & d\geqslant 3.\end{cases}$$

由引理 2.1 知, $d=1,2$ 时该随机游走常返, $d\geqslant 3$ 时该随机游走不常返.

2.6 特征函数与矩

X 的特征函数: $\phi(t)=\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}]$. 一般地, $(\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}])'\neq\mathbb{E}[(\mathrm{e}^{\mathrm{i}tX})']=\mathbb{E}[\mathrm{i}X\mathrm{e}^{\mathrm{i}tX}]$.

例题 2.13 特征函数的导数不一定存在. 例如随机变量 X 满足 $\mathbb{P}(X=5^k)=2^{-k}, k=1,2,\cdots$, 则有

$$\phi(t)=\sum_{k=1}^{+\infty}\frac{1}{2^k}\mathrm{e}^{\mathrm{i}t5^k}.$$

而 $\phi(t)$ 处处不可导.

但这也不是无可救药, 如果随机变量的 k 阶矩存在, 仍然能得到特征函数前 k 阶导的信息, 进而能得到矩的信息.

定理 2.1

若 $\mathbb{E}[|X|^k]<\infty$, 则 $\forall j\leq k$ 有

$$\phi^{(j)}(t)=\int_{-\infty}^{+\infty}(\mathrm{i}x)^j\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x),\quad \phi^{(j)}(0)=i^j\mathbb{E}[X^j]$$

进而 $\phi(t)=\sum_{j=0}^{k}\frac{(it)^j}{j!}\mathbb{E}[X^j]+o(t^k), t\to 0$

♡

注 在一定条件下, 可以利用特征函数来计算矩.

对于一个随机变量, 其 k 阶矩存在可以推出其特征函数 k 阶可导. 而反之是否成立?

定理 2.2

设 k 为偶数, 若 $\phi(t)$ 在 0 处附近有 k 阶导数, 则 $\mathbb{E}[|X|^k]<\infty$.

♡

证明 先看 $k=2$. 假设 $\phi''(t)$ 在 0 附近存在, 则有

$$\phi''(0)=\lim_{h\to 0}\frac{\phi(h)+\phi(-h)-2\phi(0)}{h^2}=\lim_{h\to 0}\int_{-\infty}^{+\infty}\frac{\mathrm{e}^{\mathrm{i}xh}+\mathrm{e}^{-\mathrm{i}xh}-2}{h^2}\mathrm{d}F(x)=-2\lim_{h\to 0}\int_{-\infty}^{+\infty}\frac{1-\cos hx}{h^2}\mathrm{d}F(x).$$

利用 Fatou 引理, 我们有

$$\mathbb{E}[X^2]=\int_{-\infty}^{+\infty}x^2\mathrm{d}F(x)=2\int_{-\infty}^{+\infty}\lim_{h\to 0}\frac{1-\cos hx}{h^2}\mathrm{d}F(x)\leqslant\lim_{h\to 0}2\int_{-\infty}^{+\infty}\frac{1-\cos hx}{h^2}\mathrm{d}F(x)=-\phi''(0)<\infty.$$

所以 $k=2$ 时成立. 现在我们开始归纳: 假设当 $n=2k$ 时命题成立, 即 $\mathbb{E}[X^{2k}]<\infty$, 则有

$$\phi^{(2k)}(t)=\int_{-\infty}^{+\infty}(\mathrm{i}x)^{2k}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x).$$

设 $G(x):=\int_{-\infty}^{x}y^{2k}\mathrm{d}F(y)$, 则 $\dfrac{G(x)}{G(+\infty)}$ 为某个随机变量 Y 的分布函数, 其特征函数为

$$\phi_Y(t)=\frac{1}{G(+\infty)}\int_{-\infty}^{+\infty}x^{2k}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x)=\frac{(-1)^k\phi^{(2k)}(t)}{G(+\infty)}.$$

因此 $\phi_Y''(t)$ 在 0 附近存在, 故有

$$\int_{-\infty}^{+\infty}x^2\frac{\mathrm{d}G(x)}{G(+\infty)}<\infty\implies\frac{1}{G(+\infty)}\int_{-\infty}^{+\infty}x^{2k+2}\mathrm{d}F(x)<\infty\implies\mathbb{E}[X^{2k+2}]<\infty.$$

注 k 为奇数时上述结论不成立. 例如 X 满足 $\mathbb{P}(X=\pm j)=\dfrac{1}{2Cj^2\log j}$, $j=2,3,\cdots$, 则

$$\phi(t)=\frac{1}{C}\sum_{j=2}^{+\infty}\frac{\cos jt}{j^2\log j},$$

其中 $C=\sum_{j=2}^{+\infty}\dfrac{1}{j^2\log j}$ 为正则化常数. $\phi(t)$ 在 0 附近可导且连续, 但 $\mathbb{E}[|X|]=\sum_{j=2}^{+\infty}\dfrac{1}{j\log j}=+\infty$.

利用特征函数, 我们还可以通过矩来得到分布:

定理 2.3

若 $\limsup\limits_{n\to+\infty}\dfrac{(\mu_{2n})^{\frac{1}{2n}}}{2n}=r<\infty$, 则至多存在一个分布 F, 使得

$$\int_{-\infty}^{+\infty}x^n\mathrm{d}F(x)=\mu_n.$$

证明 见本文附录, 可以暂时忽略.

注 **Riesz** 条件 $\limsup\limits_{n\to+\infty}\dfrac{(\mu_{2n})^{\frac{1}{2n}}}{2n}=r<\infty$ 可以换成 **Carleman** 条件 $\sum_{k=1}^{+\infty}\dfrac{1}{\mu_{2k}^{1/2k}}=+\infty$. 上述结论依然成立.

据此显然有如下推论:

推论 2.1

对随机变量 X,Y, 若 $\limsup\limits_{n\to+\infty}\dfrac{(\mathbb{E}[X^{2n}])^{\frac{1}{2n}}}{2n}=r<\infty$, 且

$$\mathbb{E}[X^n]=\mathbb{E}[Y^n],\quad\forall n\in\mathbb{N}^*.$$

则 X 与 Y 同分布.

同时也能推出矩收敛定理 (**定理** 1.15), 这里不再赘述.

下面这个例子体现出随机变量在矩的信息转化成分布的过程中, 特征函数起到中转站的作用:

例题 2.14 随机变量 X 满足 $\mathbb{E}[|X|]<+\infty$. 若对任意有界光滑函数 f 均有

$$\lambda\mathbb{E}[f(X+1)]=\mathbb{E}[Xf(X)],$$

其中常数 $\lambda>0$. 证明: X 服从参数为 λ 的泊松分布.

证明 因为 $\mathrm{e}^{\mathrm{i}tx}=\cos tx+\mathrm{i}\sin tx$ 有界光滑, 故

$$\lambda\mathbb{E}[\mathrm{e}^{\mathrm{i}t(X+1)}]=\mathbb{E}[X\mathrm{e}^{\mathrm{i}tX}].$$

记 $\phi(t)=\mathbb{E}[\mathrm{e}^{\mathrm{i}tX}]$, 因为 $\mathbb{E}[|X|]<+\infty$, 故有

$$\phi'(t)=\mathbb{E}[\mathrm{i}X\mathrm{e}^{\mathrm{i}tX}].$$

从而

$$\mathrm{i}\lambda\mathrm{e}^{\mathrm{i}t}\phi(t)=\phi'(t).$$

结合初值条件 $\phi(0)=1$, 由常微分方程的理论可知,

$$\phi(t)=\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)},$$

由唯一性定理知, X 服从参数为 λ 的泊松分布.

2.7 连续性定理与弱收敛

课堂上给出了连续性定理的内容, 我们这节的目标是证明连续性定理. 在此之前先引入一些与之有关的定理, 引理和概念.

上节课我们讲了特征函数里的 Parseval 等式, 在这里就要派上其用场.

定理 2.4 (Parseval 等式)

对随机变量 X,Y, 有

$$\int_{-\infty}^{+\infty}\phi_X(t)\mathrm{d}F_Y(t)=\int_{-\infty}^{+\infty}\phi_Y(t)\mathrm{d}F_X(t).$$

我们取 Y 为 $[-u,u]$ 上的均匀分布, 则 $\phi_Y(t)=\dfrac{\sin ut}{ut}$, 利用 Parseval 等式, 则有

$$\begin{aligned}\frac{1}{2u}\int_{-u}^{u}\phi_X(t)\mathrm{d}t&=\int_{-\infty}^{+\infty}\frac{\sin ut}{ut}\mathrm{d}F_X(t)=\int_{|ut|<2}\frac{\sin ut}{ut}\mathrm{d}F_X(t)+\int_{|ut|\geqslant 2}\frac{\sin ut}{ut}\mathrm{d}F_X(t)\\&\leqslant\int_{|ut|<2}\mathrm{d}F_X(t)+\frac{1}{2}\int_{|ut|\geqslant 2}\mathrm{d}F_X(t)=1-\frac{1}{2}\mathbb{P}(|Xu|\geqslant 2)\\&=1-\frac{1}{2}\mathbb{P}\left(|X|\geqslant\frac{2}{u}\right).\end{aligned}$$

由此得到

命题 2.1

设 X 是随机变量, 则对 $\forall u>0$, 均有

$$\mathbb{P}(|X|\geqslant\frac{2}{u})\leqslant\frac{1}{u}\int_{-u}^{u}(1-\phi_X(t))\mathrm{d}t.$$

下面这个引理给出证明弱收敛的一个等价转化形式:

引理 2.2

$X_n\xrightarrow{D}X\Longleftrightarrow$ 对任意子列 $\{n_k\}$, 存在子子列 $\left\{n_k'\right\}\subset\{n_k\}$, 使得

$$X_{n_k'}\xrightarrow{D}X.$$

提示 左推右显然, 右推左反证即可.

一个重要的准备工作是要引入**胎紧**这一概念. 之前我们在数学分析中学过 “一致连续”, “一致收敛” 等概念, 其中的 “一致” 都是强调函数列 “整体” 的行为. 对于随机变量 X 的分布函数 F 而言, 由其定义

知

$$\mathbb{P}(|X| > M) = (F(-M) + 1 - F(M))) \xrightarrow{M\to+\infty} 0.$$

但当研究对象为分布函数列时，又会涉及到**一致性**的问题，这里我们通过胎紧性来描述:

定义 2.1

F_n 是一列分布函数. 称 F_n 是胎紧的, 若

$$\sup_n(F_n(-M) + 1 - F_n(M))) \xrightarrow{M\to+\infty} 0.$$

注 记 X_n 的分布函数为 F_n, 则

$$\{F_n\} \text{ 胎紧} \iff \sup_n \mathbb{P}(|X_n| > M) \xrightarrow{M\to+\infty} 0 \iff \limsup_{n\to+\infty} \mathbb{P}(|X_n| > M) \xrightarrow{M\to+\infty} 0$$

证明连续性定理的关键是下面的 Helly 定理.

定理 2.5 (Helly 选择定理)

设 F_n 为分布函数, 则存在子序列 $\{n_k\}$, 使得 $F_{n_k} \xrightarrow{v} F$.

记 $F_n \xrightarrow{v} F$ 为 F_n 淡收敛至 F, 指对 $\forall x \in C_F, F_n(x) \to F$, 但 F 不一定有分布函数 (F 右连续且单调递增但 $F(-\infty) = 0, F(+\infty) = 1$ 不一定成立).

证明 参考 **Durrett Theorem 3.2.12**.

命题 2.2

接上, 对任意收敛子序列 F_{n_k} 都收敛到分布函数的充要条件是 $\{F_n\}$ 胎紧.

证明 充分性: 设 $F_{n_k} \xrightarrow{v} F$, $\{F_n\}$ 胎紧, 即对 $\forall\varepsilon > 0$, 均有 $M_\varepsilon > 0$, 使得

$$\limsup_{n\to+\infty}(1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)) < \varepsilon.$$

记 $r < -M_\varepsilon, s > M_\varepsilon$ 为 F 的连续点. 因为 $F_{n_k}(r) \to F(r)$,$F_{n_k}(s) \to F(s)$, 故有

$$1 - F(s) + F(r) = \lim_{k\to+\infty}(1 - F_{n_k}(s) + F_{n_k}(r)) \leqslant \limsup_{n\to+\infty}(1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)) < \varepsilon.$$

再利用 F 的单调递增性知,

$$\limsup_{x\to+\infty}(1 - F(x) + F(-x)) < \varepsilon.$$

由 $\varepsilon > 0$ 的任意性知, F 是分布函数.

必要性: 反证: 假设 $\{F_n\}$ 不是胎紧的, 即 $\exists\varepsilon > 0$ 及子列 $\{n_k\}$ 使得

$$1 - F_{n_k}(k) + F_{n_k}(-k) \geqslant \varepsilon$$

对所有 $k \in \mathbb{N}^*$ 成立. 取子子列 $\{n_{k_j}\} \subset \{n_k\}$ 使得 $F_{n_{k_j}} \xrightarrow{v} F$. 记 $r < 0 < s$ 是 F 的连续点, 则有

$$1 - F(s) + F(r) = \lim_{j\to+\infty}(1 - F_{n_{k_j}}(s) + F_{n_{k_j}}(r)) \geqslant \liminf_{j\to+\infty}(1 - F_{n_{k_j}}(k_j) + F_{n_{k_j}}(-k_j)) \geqslant \varepsilon.$$

令 $s \to +\infty, r \to -\infty$, 则 F 不是分布函数.

准备工作已就绪, 现在我们来证明连续性定理:

定理 2.6 (Lévy-Cramér 连续性定理)

F_n 为分布函数, $\phi_n(t)=\int e^{itx}\mathrm{d}F_n$

(1) 若 F 为分布函数, 且 $F_n \xrightarrow{w} F$, 则 $\phi_n(t)\to\phi(t), \forall t\in\mathbb{R}$(且内闭一致收敛, 这里暂且不证), 这里 $\phi(t)=\int e^{itx}\mathrm{d}F$

(2) 若 $\phi(t):=\lim\limits_{n\to\infty}\phi_n(t)$ 存在, 且 $\phi(t)$ 在 $t=0$ 处连续, 则 ϕ 为某分布函数 F 的特征函数, 且 $F_n\xrightarrow{w}F$.

证明 对 (1) 设 X_n, X 的分布函数分别是 F_n, F, 则有 $X_n\xrightarrow{D}X$. 注意到 e^{itx} 有界连续, 因此 $\mathbb{E}[\mathrm{e}^{itX_n}]\to\mathbb{E}[\mathrm{e}^{itX}]$.

(2) 先利用引理 2.2, 即证: 对任意子列 $\{n_k\}$, 存在子子列 $\{n_k'\}\subset\{n_k\}$, 使得

$$F_{n_k'}\xrightarrow{w}F.$$

对任意子列 $\{n_k\}$, 我们分以下四个步骤完成证明:

1. 利用 **Helly** 选择定理 **(**定理 **2.5)** 得到淡收敛子子列 $\{n_k'\}$,
2. $\phi(t)$ 在 $t=0$ 处连续保证了 $\{F_n\}$ 的胎紧, 因为
$$\limsup_{n\to+\infty}\mathbb{P}(||X_n|>r)\leqslant\frac{r}{2}\limsup_{n\to+\infty}\int_{-\frac{2}{r}}^{\frac{2}{r}}(1-\phi_n(t))\mathrm{d}t=\frac{r}{2}\int_{-\frac{2}{r}}^{\frac{2}{r}}(1-\phi(t))\mathrm{d}t\xrightarrow{r\to+\infty}2(1-\phi(0))=0,$$
其中 F_n 是随机变量 X_n 的分布函数, 第一个不等式用到了命题 2.1, 后面的等号用到了特征函数的一致连续性.
3. $\phi(t)=\lim\limits_{n\to\infty}\phi_n(t)$ 保证了子子列收敛的唯一性.
4. 利用等价条件命题 2.2 即得证.

2.8 用 Linderberg 替换术证明 Linderberg-Feller 中心极限定理

[1] $X_1,X_2,\cdots,X_n$ 相互独立, $\mathbb{E}[X_k]=0, b_k^2=\mathrm{Var}(X_k), B_n=\sqrt{b_1^2+\cdots+b_n^2}$.

Linderberg 条件 (1922):

$$\forall\epsilon>0,\ \lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x|>\epsilon B_n}x^2\mathrm{d}F_k=0 \tag{L}$$

Linderberg-Feller CLT: 设 $\{X_i\}$ 相互独立, 满足(L), 则

$$\frac{1}{B_n^2}\max_{1\le k\le n}b_k^2\to 0 \tag{Feller}$$

且

$$\frac{1}{B_n}\sum_{k=1}^{n}X_k\xrightarrow{D}N(0,1)$$

1922 年 Linderberg 观察到

[1] 可参考 Tao 的讲义:
https://terrytao.wordpress.com/2010/01/05/254a-notes-2-the-central-limit-theorem/
https://terrytao.wordpress.com/2015/11/02/275a-notes-4-the-central-limit-theorem/
这两篇 notes 包含了 Lindeberg Replacement Tricks 和 Moments Method in CLT

(1) 当 $\{Y_k\}$ 为独立正态随机变量, $\mathbb{E}[Y_k]=0, \mathrm{Var}(Y_k)=b_k^2$ 时,

$$\mathbb{E}\left[g\left(\frac{1}{B_n}\sum_{k=1}^n Y_k\right)\right]=\mathbb{E}[g(G)]+o(1), G\sim N(0,1)$$

其中 g 是一个好的测试函数, 如有界连续函数.

(2) 对 $\{X_k\}$ 三阶矩存在时,

$$\mathbb{E}\left[g\left(\frac{1}{B_n}\sum_{k=1}^n X_k\right)\right]=\mathbb{E}\left[g\left(\frac{1}{B_n}\sum_{k=1}^n Y_k\right)\right]+o(1)$$

(3) 使用截断术去掉三阶矩假设

回顾: $X_n\xrightarrow{D}X\Leftrightarrow\forall g\in C_{\mathrm{b}}(\mathbb{R})$(有界连续函数全体), $\mathbb{E}[g(X_n)]\to\mathbb{E}[g(X)]\Leftrightarrow\forall g,g',g'',g'''\in C_{\mathrm{b}}(\mathbb{R}),\mathbb{E}[g(X_n)]\to\mathbb{E}[g(X)]$

引入 $\{Y_k\}$ 独立正态随机变量, 且与 $\{X_k\}$ 独立, $\mathbb{E}[Y_k]=0, b_k^2=\mathrm{Var}(Y_k)$ (思考: 两个分布函数 F,G, 试 CHECK 存在独立随机变量 X 与 Y, 使得对应分布函数为 F,G)

令 $\zeta_k=X_1+\cdots+X_{k-1}+Y_{k+1}+\cdots+Y_n$, 则

$$\zeta_n+X_n=\sum_{k=1}^n X_k:=S_n$$

$$\zeta_1+Y_1=\sum_{k=1}^n Y_k\overset{D}{=\!=}B_nY, Y\sim N(0,1)$$

更一般 $1\le k<n$ 有

$$\zeta_k+X_k=\zeta_{k+1}+Y_{k+1}$$

$$\mathbb{E}\left[g\left(\frac{S_n}{B_n}\right)\right]-\mathbb{E}[g(Y)]=\sum_{k=1}^n\left(\mathbb{E}\left[g\left(\frac{\zeta_k+X_k}{B_n}\right)\right]-\mathbb{E}\left[g\left(\frac{\zeta_k+Y_k}{B_n}\right)\right]\right)$$

相当于每一步做替换 $X_n\to Y_n, X_{n-1}\to Y_{n-1},\cdots,X_1\to Y_1$, 只要使总的误差足够小即可.

观察: 因为 ζ_k,X_k,Y_k 独立, 有

$$\mathbb{E}\left[g'\left(\frac{\zeta_k}{B_n}\right)(X_k-Y_k)\right]=0$$

$$\mathbb{E}\left[g''\left(\frac{\zeta_k}{B_n}\right)(X_k^2-Y_k^2)\right]=0$$

引入 $h(t)=\sup_x\{g(x+t)-g(x)-g'(x)t-\frac{1}{2}g''(x)t^2\}$ 则易知 $\exists K>0$, 使得 $h(t)\le Kt^2\wedge|t|^3$ 因此, 对 $g\left(\frac{\zeta_k+X_k}{B_n}\right), g\left(\frac{\zeta_k+Y_k}{B_n}\right)$ 在 $\frac{\zeta_k}{B_n}$ 点处做 Taylor 展开, 常数项、一次项、二次项均消去, 得

$$\left|\mathbb{E}\left[g\left(\frac{\zeta_k+X_k}{B_n}\right)-g\left(\frac{\zeta_k+Y_k}{B_n}\right)\right]\right|\le\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\right]+\mathbb{E}\left[h\left(\frac{Y_k}{B_n}\right)\right]$$

只须证明:

$$\sum_{k=1}^n\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\right]\to 0 \tag{2.1}$$

$$\sum_{k=1}^n\mathbb{E}\left[h\left(\frac{Y_k}{B_n}\right)\right]\to 0 \tag{2.2}$$

先证明 (2.2): 由 $h(t) \le K|t|^3$

$$\begin{aligned}
\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{Y_k}{B_n}\right)\right] &\le K\sum_{k=1}^{n}\mathbb{E}\left[\left|\frac{Y_k}{B_n}\right|^3\right] \\
&= K\sum_{k=1}^{n}\frac{b_k^3}{B_n^3}\mathbb{E}[|Y|^3] \\
&\le K\max_{1\le k\le n} b_k\sum_{k=1}^{n}\frac{b_k^2}{B_n^3}\mathbb{E}[|Y|^3] \\
&= K\max_{1\le k\le n} b_k\frac{1}{B_n}\mathbb{E}[|Y|^3] \to 0 \text{ (Feller)}
\end{aligned}$$

再来证明(2.1): 为了使用 Linderberg 条件, 做划分 $\{|X_k| \le \epsilon B_n\}, \{|X_k| > \epsilon B_n\}$, 对前一部分用三次项控制, 后一部分用二次项控制

$$\begin{aligned}
\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\right] &= \sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\left(I_{\{|X_k|\le\epsilon B_n\}}+I_{\{|X_k|>\epsilon B_n\}}\right)\right] \\
&\le K\sum_{k=1}^{n}\frac{1}{B_n^3}\mathbb{E}\left[|X_k|^3 I_{\{|X_k|\le\epsilon B_n\}}\right] \\
&\quad + K\sum_{k=1}^{n}\frac{1}{B_n^2}\mathbb{E}\left[|X_k|^2 I_{\{|X_k|>\epsilon B_n\}}\right] \\
&\le K\sum_{k=1}^{n}\frac{\epsilon}{B_n^2}\mathbb{E}\left[|X_k|^2 I_{\{|X_k|\le\epsilon B_n\}}\right] \\
&\quad K\sum_{k=1}^{n}\frac{1}{B_n^2}\mathbb{E}\left[|X_k|^2 I_{\{|X_k|>\epsilon B_n\}}\right] \\
&\le \epsilon K + K\sum_{k=1}^{n}\frac{1}{B_n^2}\int_{|x|>\epsilon B_n} x^2 \mathrm{d}F_k \to 0
\end{aligned}$$

最后一步使用了 Linderberg 条件, 证毕.

思考题: 设 $\{X_k\}$ 是独立同随机变量, 且 $\mathbb{E}[X_k] = 0, \mathrm{Var}(X_k) = 1, \mathbb{E}[X_k^3] < \infty, g \in C_b^3(\mathbb{R})$. 证明:

(1) $\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k\right)\right] - \mathbb{E}[g(Y)]\right| \leqslant \frac{C_1}{\sqrt{n}}$.

(2) $\left|\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}X_k \leqslant t\right) - \mathbb{P}(Y \leqslant t)\right| \leqslant C_2 n^{-\frac{1}{8}}$.

其中 $Y \sim N(0,1), C_1, C_2$ 为常数.

2.9 利用 Kolmogorov 三级数定理证明强大数律

命题 2.3

设 $\{X_n\}$ 独立, $\mathbb{E}[X_n] = 0$, 且存在 $C > 0$, 使得 $|X_n| \leqslant C$, 则

$$\sum_{n=1}^{\infty}\mathrm{Var}(X_n) < +\infty \Longleftrightarrow \sum_{n=1}^{\infty}X_n \text{ a.s. 收敛.}$$

♠

证明 $\Longrightarrow$: 见一级数定理 (定理 **1.29**);

$\Longleftarrow$: 反证: 设 $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) = +\infty$, 利用例 1.52有

$$\mathbb{P}(\max_{1\leqslant k\leqslant m} |X_{n+1} + \cdots + X_{n+k}| \geqslant \varepsilon) \geqslant 1 - \frac{(\varepsilon + c)^2}{\sum_{k=n+1}^{n+m} \mathbb{E}[X_k^2]} \xrightarrow{m\to+\infty} 1.$$

因此

$$\mathbb{P}(\sup_{k\geqslant 1} |X_{n+1} + \cdots + X_{n+k}| \geqslant \varepsilon) = 1 \implies \sum_{n=1}^{\infty} X_n \text{ a.s. 发散.}$$

矛盾.

更一般地, 有

命题 2.4

设 $\{X_n\}$ 独立, 且存在 $C > 0$, 使得 $|X_n| \leqslant C$, 则

$$\begin{cases} \sum_{n=1}^{\infty} \mathrm{Var}(X_n) < +\infty \\ \sum_{n=1}^{\infty} \mathbb{E}[X_n] < \infty. \end{cases} \iff \sum_{n=1}^{\infty} X_n \text{ a.s. 收敛.}$$

证明 $\Longrightarrow$: 利用一级数定理 (定理 **1.29**) 即可;

$\Longleftarrow$: 设 $\{X_n'\}$ 是 $\{X_n\}$ 的独立复制, 则 $\widetilde{X_n} := X_n - X_n'$ 是对称的随机变量, 且

$$\sum_{n=1}^{\infty} X_n \text{ a.s. 收敛} \iff \sum_{n=1}^{\infty} X_n' \text{ a.s. 收敛.}$$

因此 $\sum_{n=1}^{\infty} \widetilde{X_n}$ a.s. 收敛. 利用

$$\mathbb{E}[|X_n|] < \infty \implies \mathbb{E}[\widetilde{X_n}] = 0,$$

由命题 2.3得

$$\sum_{n=1}^{\infty} \mathrm{Var}(\widetilde{X_n}) < +\infty.$$

同时,

$$\mathbb{E}[X_n^2] < \infty \implies \mathbb{E}[\widetilde{X_n}^2] = \mathrm{Var}(\widetilde{X_n}) = 2\mathrm{Var}(X_n) \implies \sum_{n=1}^{\infty} \mathrm{Var}(X_n) < +\infty.$$

利用一级数定理 (定理 **1.29**), 结合 $\sum_{n=1}^{\infty} X_n$ a.s. 收敛, 得

$$\sum_{n=1}^{\infty} (X_n - \mathbb{E}[X_n]) \text{ a.s. 收敛} \implies \sum_{n=1}^{\infty} \mathbb{E}[X_n] < \infty.$$

由此我们得到三级数定理:

定理 2.7 (Kolmogorov 三级数定理)

设 $\{X_n\}$ 为相互独立的随机变量, 则 $\sum\limits_{n=1}^{\infty} X_n$ a.s. 收敛的充分条件是: 存在常数 $c>0$, 使得

$$\begin{cases} 1^{\circ}. \sum\limits_{n=1}^{\infty} \mathbb{P}(|X_n|>c)<\infty, \\ 2^{\circ}. \sum\limits_{n=1}^{\infty} \mathbb{E}[X_n I_{\{|X_n|\leqslant c\}}]<\infty, \\ 3^{\circ}. \sum\limits_{n=1}^{\infty} \mathrm{Var}(X_n I_{\{|X_n|\leqslant c\}})<\infty. \end{cases}$$

而必要条件是: 对 $\forall c>0$, 上述级数收敛.

证明 充分性: 令 $Y_n = X_n I_{\{|X_n|\leqslant c\}}$, 则由 1° 知

$$\sum_{n=1}^{+\infty} \mathbb{P}(X_n \neq Y_n) = \sum_{n=1}^{+\infty} \mathbb{P}(|X_n|>c)<\infty.$$

因此 $\sum\limits_{n=1}^{\infty} X_n$ 和 $\sum\limits_{n=1}^{\infty} Y_n$ 有相同的收敛性. 结合命题 2.4, 由 2° 和 3° 知 $\sum\limits_{n=1}^{\infty} Y_n$ a.s. 收敛. 所以 $\sum\limits_{n=1}^{\infty} X_n$ a.s. 收敛.

必要性: $\sum\limits_{n=1}^{\infty} X_n$ a.s. 收敛 $\implies X_n \xrightarrow{\text{a.s.}} 0 \iff \sum\limits_{n=1}^{+\infty} \mathbb{P}(|X_n|>c)<\infty$, 1° 成立. 因此 $\sum\limits_{n=1}^{\infty} X_n$ 和 $\sum\limits_{n=1}^{\infty} Y_n$ 有相同的收敛性, 故 $\sum\limits_{n=1}^{\infty} Y_n$ a.s. 收敛. 结合命题 2.4知 2° 和 3° 成立.

三级数定理本身非常有用, 在此不作列举. 现在我们要通过三级数定理来证明强大数律. 在此之前先引入非常重要的一个引理:

引理 2.3 (Kronecker 引理)

设 $\{a_n\}$ 为正实数序列且 $a_n \uparrow \infty$, $\{x_n\}$ 为实数列. 若 $\sum\limits_{n=1}^{\infty} \frac{x_n}{a_n}$ 收敛, 则

$$\lim_{n\to+\infty} \frac{1}{a_n}\sum_{k=1}^{n} x_n = 0.$$

证明 令 $b_n = \sum\limits_{k=1}^{n} \frac{x_k}{a_k}$, 则 $b_n \xrightarrow{n\to+\infty} \sum\limits_{k=1}^{\infty} \frac{x_k}{a_k} := b$. 故

$$\frac{1}{a_n}\sum_{k=1}^{n} x_k = \frac{1}{a_n}\sum_{k=1}^{n} a_k(b_k-b_{k-1}) = \frac{1}{a_n}\left(a_n b_n - \sum_{k=0}^{n-1} b_k(a_k-a_{k-1})\right) \xrightarrow{n\to+\infty} b-b=0.$$

我们来看强大数律的证明:

定理 2.8 (强大数律)

设 $X_1,\cdots,X_n,X$ 独立同分布, 且 $\mathbb{E}[|X|]<+\infty$, 则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} \mathbb{E}[X].$$

证明 不妨设 $\mathbb{E}[X]=0$, 要证 $\dfrac{\sum_{k=1}^{n} X_k}{n} \xrightarrow{\text{a.s.}} 0$. 由 **Kronecker** 引理 (引理 2.3) 知只需证: $\sum_{n=1}^{+\infty} \dfrac{X_n}{n}$ a.s. 收敛, 再由 **Kolmogorov** 三级数定理 (定理 2.7) 可知等价于证明

$$\begin{cases} 1^{\circ}. \displaystyle\sum_{n=1}^{\infty} \mathbb{P}(|X_n|>n)<\infty, \\ 2^{\circ}. \displaystyle\sum_{n=1}^{\infty} \mathbb{E}[\frac{X_n}{n} I_{\{|X_n|\leqslant n\}}]<\infty, \\ 3^{\circ}. \displaystyle\sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{X_n}{n} I_{\{|X_n|\leqslant n\}}\right)<\infty. \end{cases}$$

$1^{\circ}: \sum_{n=1}^{\infty} \mathbb{E}[I_{\{|X|>n\}}]=\mathbb{E}[\sum_{n=1}^{\infty} I_{\{|X|>n\}}]=\mathbb{E}[\sum_{n<|x|} 1] \leqslant \mathbb{E}[|X|]<\infty.$

$3^{\circ}: \sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{X_n}{n} I_{\{|X_n|\leqslant n\}}\right) \leqslant \sum_{n=1}^{\infty} \mathbb{E}\left(\frac{X_n}{n} I_{\{|X_n|\leqslant n\}}\right)^2=\sum_{n=1}^{\infty} \mathbb{E}[\frac{X^2}{n^2} I_{\{|X|\leqslant n\}}]=\mathbb{E}[\sum_{n=1}^{\infty} \frac{X^2}{n^2} I_{\{|X|\leqslant n\}}]=\mathbb{E}[X^2 \sum_{n\geqslant|X|} \frac{1}{n^2}] \leqslant \mathbb{E}[X^2 \cdot \frac{C}{|X|}] \leqslant C\mathbb{E}[|X|]<\infty.$

2°: 令 $Y_n=X_n I_{\{|X_n|\leqslant n\}}$, 则

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(\frac{Y_n}{n}\right)<\infty \Longrightarrow \sum_{n=1}^{\infty} \frac{Y_n-\mathbb{E}[Y_n]}{n} \text{ a.s. 收敛.}$$

再由 **Kronecker** 引理 (引理 2.3) 知

$$\frac{\sum_{k=1}^{n}(Y_k-\mathbb{E}[Y_k])}{n} \xrightarrow{\text{a.s.}} 0$$

结合

$$\frac{\sum_{k=1}^{n} \mathbb{E}[Y_k]}{n}=\frac{\sum_{k=1}^{n} \mathbb{E}[X I_{\{|X|\leqslant k\}}]}{n} \xrightarrow{n\to+\infty} \mathbb{E}[X].$$

可知

$$\frac{1}{n} \sum_{k=1}^{n} Y_k \xrightarrow{\text{a.s.}} \mathbb{E}[X].$$

再由 1° 知

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} \mathbb{E}[X].$$

2.10 信息熵 (entropy)

什么是信息? 对事件发生惊奇程度. 例如: 3 个骰子点数之和为偶数的概率是 $\frac{1}{2}$, 3 个骰子点数之和为 18 的概率是 $\frac{1}{6^3}$. 概率为 p 事件发生后, 用 $S(p)$ 表示惊奇程度, $S(p), 0<p\leq 1$ 有何要求?

- 公理 1: $S(1)=0$. 必然事件不带来惊奇
- 公理 2: $S(p_1)>S(p_2), p_1<p_2$. 严格减, 越不可能发生, 越表达更多信息.

- 公理 3: $S(p)$ 为 p 的连续函数. p 的微小变化引起 $S(p)$ 微小变化.
- 公理 4: $S(pq) = S(p) + S(q), 0 < p, q \le 1$. E, F 独立, $\mathbb{P}(E) = p, \mathbb{P}(F) = q, E \cap F$ 惊奇度 $S(pq)$, 则 $S(pq) - S(p)$ 表示听到 E 发生后增加惊奇度, 但 E 与 F 独立, 故应该

$$S(pq) - S(p) = S(q).$$

定理 2.9

若 $S(p)$ 满足公理 $1-4$, 则 $\exists c > 0$, 使得 $S(p) = -c \log p$.

证明 $S(p^2) = 2S(p), \cdots, S(p^m) = mS(p), S(p^{\frac{1}{n}}) = \frac{1}{n}S(p), S(p^{\frac{m}{n}}) = \frac{m}{n}S(p)$, 公理 3 表明 $S(p^x) = xS(p)$, 取 $x = -\log p$, 则

$$S(p) = S\left(\frac{1}{e^x}\right) = xS\left(\frac{1}{e}\right) = -c\log p$$

这里 $c = S\left(\frac{1}{e}\right) > S(1) = 0$ (这里利用了公理 1,2)

离散型

X	x_1	x_2	$\cdots$	x_n
$\mathbb{P}$	p_1	p_2	$\cdots$	p_n

定义 2.2

Shannon 信息熵: $H(X) = -\sum_{i=1}^{n} p_i \log p_i$

联合熵:

$$H(X,Y) = -\sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j)$$

相对熵:

$$H_Y(X) = \sum_j H_{Y=y_j}(X) p_Y(y_j)$$

这里 $H_{Y=y_j}(x) = -\sum_i p(x_i \mid y_j) \log p(x_i \mid y_j)$

引理 2.4

$H(X,Y) = H(Y) + H_Y(X)$

证明

$$\begin{aligned}
H_Y(X) &= \sum_j \left(-\sum_i p(x_i \mid y_j) \log p(x_i \mid y_j)\right) p_Y(y_j) \\
&= -\sum_{i,j} p(x_i, y_j) \log\left(\frac{p(x_i, y_j)}{p(y_j)}\right) \\
&= H(X,Y) - H(Y)
\end{aligned}$$

注 注此定义可扩展到一般离散型

定理 2.10

$H_Y(X) \leq H(X)$, 取等号等价于 X 与 Y 独立

证明

$$\begin{aligned}
H_Y(X) - H(X) &= -\sum_i \sum_j \left(p(x_i \mid y_j) \log p(x_i \mid y_j)\right) p_Y(y_j) \\
&\quad + \sum_i \sum_j p(x_i, y_j) \log p_X(x_i) \\
&= \sum_{i,j} p(x_i, y_j) \log \left(\frac{p_X(x_i)}{p(x_i \mid y_j)}\right) \\
&\leq \sum_{i,j} p(x_i, y_j) \left(\frac{p_X(x_i)}{p(x_i \mid y_j)} - 1\right) \\
&= \sum_{i,j} p(x_i, y_j) p_X(x_i) p_Y(y_j) - \sum_{i,j} p(x_i, y_j) \\
&= 0
\end{aligned}$$

易知取等号条件.

注 注 $\mathbb{P}(X = x_i) = p_i, i = 1, \cdots, n$

$$H(X) = -\sum_{i=1}^n p_i \log p_i \qquad \left(\sum_i p_i = 1\right)$$

$$\max\{H(X)\} = H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) = \log n$$

还有

$$H\left(\frac{1}{n-1}, \cdots, \frac{1}{n-1}\right) < H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$$

熵表达一个系统混乱程度的一种度量.

连续性

X 有密度 f.

定义 2.3

$$H(X) = -\int_{-\infty}^{+\infty} f(x) \log f(x) \mathrm{d}x$$

$$H(X, Y) = -\int_{\mathbb{R}^2} f(x, y) \log f(x, y) \mathrm{d}x\mathrm{d}y$$

引入 $\eta(u) = -u \log u, u \geq 0$, 则

$$\begin{aligned}
\eta(u) - \eta(v) &= \eta'(v)(u - v) + \frac{1}{2}\eta''(\xi)(u - v)^2 \\
&\leq \eta'(v)(u - v) \leq -(\log v + 1)(u - v)
\end{aligned}$$

Gibbs 不等式: $u - u \log u \leq v - u \log v$

立即有: $\int (f - f\log f)\mathrm{d}x \le \int (g - f\log g)\mathrm{d}x$, 这里 f, g 均为密度, 进而

$$-\int f\log f\mathrm{d}x \le -\int f\log g\mathrm{d}x$$

定理 2.11

记 $D = \{x \in \mathbb{R} : f(x) > 0\}$

1. $D = (-\infty, +\infty), \mathbb{E}[X] = 0, \mathrm{Var}(X) = 1$, 正态分布时熵最大, 最大熵为 $\log\sqrt{2\pi e}$
2. $D = (0, +\infty), \mathbb{E}[X] = \frac{1}{\lambda}, \lambda > 0$, 指数分布 $\mathrm{Exp}(\lambda)$ 熵最大 $\log\frac{e}{\lambda}$
3. $D = [0, a]$, 均匀分布熵最大 $\log a$

证明 这是指在给定密度函数 (广义，可以是离散的) 的支撑集上，给定一些条件时，存在某种可以使熵极大化的概率分布, 这里只证第一条, 其他类似. 取 $g(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$, 则

$$\begin{aligned}
H(X) &\le -\int f\log\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right)\mathrm{d}x \\
&= -\int\left(\log\frac{1}{\sqrt{2\pi}} - \frac{1}{2}x^2\right)f\mathrm{d}x \\
&= \log\sqrt{2\pi} + \frac{1}{2}
\end{aligned}$$

为标准正态分布熵.

Boltzmann 熵: $S = k_B\log W$, 其中 S 为宏观系统熵, W 为微观状态数, k_B 为 Boltzmann 常数. 每一个发生可能性为 $\frac{1}{W}$, 共有 W 种可能结果, 熵为 $-\sum_{i=1}^{W}\frac{1}{W}\log\frac{1}{W} = \log W$

能级:$E_1, E_2, \cdots, E_n$

概率: $p_1, p_2, \cdots, p_n$

平均能量: $\sum_{i=1}^{n} p_i E_i = U$

最大熵时 $p_i = \frac{1}{Z}e^{-\beta E_i}$,Gibbs 分布. 配分函数 $Z = \sum_i e^{-\beta E_i}$, β 由 U 决定, 由 Lagrange 乘子法:

$$L = -\sum_i p_i\log p_i - \beta\left(\sum_i p_iE_i - U\right) - \lambda\left(\sum_i p_i - 1\right)$$

$$\frac{\partial L}{\partial p_i} = -(1 + \log p_i) - \beta E_i - \lambda = 0$$

$$\frac{\partial L}{\partial \beta} = -\sum_i p_iE_i + U = 0$$

$$\frac{\partial L}{\partial \lambda} = -\sum_i p_i + 1 = 0$$

即 $p_i = e^{-1-\lambda-\beta E_i}, \sum_i p_i = 1, \sum_i p_iE_i = U$

f 为密度, 约束: $\int_{\mathbb{R}} h(x)f(x)\mathrm{d}x = \bar{h}$ 能量何时 $H(f)$ 最大? (Gibbs 分布)

猜:

$$f_0(x) = \frac{e^{-ch(x)}}{\int e^{-ch(x)}\mathrm{d}x}, c \in \mathbb{R}$$

类似正态分布情形, 最大熵

$$H(f_0) = \ln Z + c\bar{h}.$$

2.11 随机矩阵初步

[2] PW Anderson 1972: More is different!

FJ Dyson 1962: Random Matrix Theory is a new kind of statistical machanics in which we renounce exact knowledge. Not of the state of the system,But of the system itself.

强烈推荐 Dyson 的科普文章 Birds and Frogs.

When randomness meets matrix.

$(\Omega, \mathcal{F}, \mathbb{P})$

随机变量 $X : \Omega \to \mathbb{R}$, 可测

随机向量 $\vec{X} : \Omega \to \mathbb{R}^n$, 可测映射

随机矩阵: 矩阵值随机变量或矩阵元为随机变量

随机矩阵 = 矩阵论 + 概率论, 无处不在

2.11.1 起源

- 统计 1928 Wishart 样本协方差阵

$$X_k = (X_{1k}, \cdots, X_{pk})^T$$

$$W = \frac{1}{n}\sum_{k=1}^{n} X_k X_k^T = \frac{1}{n} X X^T$$

这里 $X = (X_1, \cdots, X_n)$, $p \times n$ 矩阵, $\{X_{ij}\}$ 为独立同 $N(0,1)$, 矩阵 X 的矩阵元联合密度为

$$f(X) = \frac{1}{(\sqrt{2\pi})^{pn}} e^{-\frac{1}{2}\mathrm{tr}XX^T}$$

- 物理 1950s Wigner(1902-1995) Schrödinger 方程:

$$i\hbar\frac{\partial}{\partial t}\psi(t,\vec{x}) = H\psi(t,\vec{x})$$

$H = -\dfrac{\hbar}{2m}\Delta + V(\vec{x})$ 为哈密顿算子, $\psi(t,\vec{x})$ 为波函数, $\displaystyle\int |\psi|^2 \,\mathrm{d}\vec{x} = 1$.

1950s Wigner 发现粒子多, 强相互作用 $\to$ 复杂度增加 $\to$ 随机性引入. 利用 H 截断 $\to$ $n \times n$ 矩阵 H_n, $H_n^* = H_n$, 矩阵元为随机变量.

Wigner 观点: 用随机矩阵模拟复杂体系哈密尔顿算子

最初模型: $H_{2n+1} = (v_{ij})_{i,j=-n}^n$, $v_{ii} = 0, \forall i$, $\{v_{ij}\}_{i<j}$ 独立同分布于 Z,

$$\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$$

必须研究 $n \to \infty$ 时, H_n 谱的渐近性, 进而逼近 H 谱! 大维现象.

[2]可参考https://www.math.harvard.edu/media/feier.pdf
包含了老师上课讲的矩方法证明 Wigner 半圆律（以及推广的 Wishart 矩阵的渐进谱分布)

Connection with

- MIMO 无线通信
- 高维数据
- 统计与量子物理
- 自由概率论
- 表示论 群 S_n
- 数论 Riemann ζ 函数

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_{p}\left(1-\frac{1}{p^s}\right)^{-1}, \mathrm{Re}(s)>1$$

其中 p 是素数.

对称性:

$$\Gamma\left(\frac{s}{2}\right)\pi^{-\frac{s}{2}}\zeta(s)=\Gamma\left(\frac{1-s}{2}\right)\pi^{-\frac{1-s}{2}}\zeta(1-s)$$

Riemann Hypothesis:Non-trivial zeroes of $\zeta(s)$ lie on $\mathrm{Re}(s)=\frac{1}{2}$(critical line)

传说: Hilbert-Polya Conjecture:Non-trivial zeroes of $\zeta(s)$ are the spectrum of a self-adjoint Hamiltionian operator.

2.11.2 高斯正交系综 (Gaussian Orthogonal Ensemble,GOE)

$$X_n:\Omega\longrightarrow M_{n\times n}(\mathbb{R})$$

$$X_n(\omega)\longrightarrow(x_{ij}(\omega))_{i,j=1}^{n}$$

假设 $\{X_{ij}\}$ 为 iid $N(0,\sigma^2)$, 构造对称矩阵

$$A_n=\frac{1}{2}(X_n+X_n)^T$$

明显: $a_{ii}\sim N(0,\sigma^2), a_{ij}=a_{ji}\sim N\left(0,\frac{\sigma^2}{2}\right)$ $(i\neq j)$ 且 $\{a_{ij}\}_{i\leq j}$ 相互独立. A_n 矩阵元联合密度

$$\begin{aligned}f(A_n)&=\frac{1}{(\sqrt{2\pi\sigma^2})^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}a_{ii}^2}\cdot\frac{1}{(\sqrt{\pi\sigma^2})^{\frac{1}{2}n(n-1)}}e^{-\frac{1}{\sigma^2}\sum_{i<j}a_{ij}^2}\\&=\frac{1}{(\sqrt{2})^n}\cdot\frac{1}{(\sqrt{\pi\sigma^2})^{\frac{1}{2}n(n+1)}}e^{-\frac{1}{2\sigma^2}\mathrm{Tr}A_n^2}\end{aligned}$$

此时称 A_n 为一个 GOE 矩阵, 记 $A_n\sim\mathrm{GOE}_n(\sigma)$.GOE 具有在正交群作用下不变性, 即:

$$\forall Q\in O(n):=\{X\in M_{n\times n}:XX^T=I\}$$

令 $B_n=QA_nQ^{-1}$, 则 B_n 亦为对称阵.

Claim: $B_n\sim GOE_n(\sigma)$

因为 $B_n=\frac{1}{2}\left(QX_nQ^T+(QX_nQ^T)^T\right)$, 只须说明: QX_n 矩阵元为 iid $N(0,\sigma^2)$. 另一方面, 记 $X_n=(\vec{X_{n1}},\cdots,\vec{X_{nn}})$, 而对 $\forall k,\vec{X_{nk}}$ 为 $N(0,\sigma^2I)$, 可知 $Q\vec{X_{nk}}\sim N(0,\sigma^2I)$

定理 2.12

设 $A_n \sim \mathrm{GOE}_n(\sigma)$, 记 $X_1, \cdots, X_n$ 为 A_n 特征值, 则特征值联合密度为

$$\rho_n(x_1, \cdots, x_n) = \frac{1}{c_n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2} \cdot \prod_{1 \le j < k \le n} |x_k - x_j|$$

这里 $c_n = (2\pi)^{\frac{1}{2}n} \sigma^{\frac{1}{2}n(n+1)} \prod_{k=1}^{n} \frac{\Gamma(1+\frac{k}{2})}{\Gamma(1+\frac{1}{2})}$ 为归一化常数, $\prod_{1 \le j < k \le n} |x_k - x_j|$ 来自 Jacobian 行列式.

观察到, 两个特征值靠的很近时, 密度为 0, 说明特征值具有排斥效应, 非独立, 具有强相互作用. 虽然矩阵元独立, 但特征值不独立.

More is different！ 多则异.

当 $n = 2$ 时, 定理 (2.12)容易验证, 大致想法:

$$|\lambda I - A_n| = 0 \Rightarrow \lambda_{\pm} = \frac{1}{2}(a_{11} + a_{22} \pm \sqrt{\Delta})$$
$$\Delta = (a_{11} - a_{22})^2 + 4a_{12}$$

明显 $\Delta = (\lambda_+ - \lambda_-)^2$

$$|a_{11} - a_{22}| = \sqrt{(\lambda_+ - \lambda_-)^2 - 4a_{12}^2}$$

取 $\lambda_0 = a_{12}$, 算 $(a_{11}, a_{12}, a_{22}) \mapsto (\lambda_+, \lambda_-, \lambda_0)$ Jacobian $= \dfrac{\sqrt{(\lambda_+ - \lambda_-)^2 - 4\lambda_0^2}}{|\lambda_+ - \lambda_-|}$, 对 λ_0 积分后保留 λ_+, λ_- 即可.

2.11.3 半圆律

实 Wigner 矩阵: $A_n = (a_{ij}), A_n = A_n^T$, 满足:

- $\{a_{ii}\}$ 同分布与 Y
- $\{a_{ij}\}_{i<j}$ 同分布与 Z
- $\{a_{ij}\}_{i \le j}$ 相互独立
- $\mathbb{E}[Y] = \mathbb{E}[Z] = 0, \mathrm{Var}(Z) = 1, \mathrm{Var}(Y) < \infty$
- $\forall k \ge 3$ 时, $\mathbb{E}[|Y|^k], \mathbb{E}[|Z|^k] < \infty$

记

$$\gamma_k = \begin{cases} 0 & k = 2m+1 \\ \dfrac{1}{m+1}\dbinom{2m}{m} & k = 2m \end{cases}$$

对象: $\dfrac{A_n}{\sqrt{n}}$

半圆律: $\omega(x) = \dfrac{1}{2\pi}\sqrt{4 - x^2}, x \in [-2, 2]$, γ_k 即为半圆律的 k 阶矩.

定理 2.13

$\forall k = 0, 1, 2, \cdots$ 有

$$\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right] \to \gamma_k, n \to \infty$$

先算

$$\int x^k\omega(x)\mathrm{d}x = \int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}}\frac{x^k}{4\pi}I_{\{x^2+y^2\le 4\}}\mathrm{d}y\mathrm{d}x$$
$$= \frac{1}{4\pi}\iint_{x^2+y^2\le 4}x^k\mathrm{d}y\mathrm{d}x$$

引入 $x = r\cos\theta, y = r\sin\theta$

$$\int x^k\omega(x)\mathrm{d}x = \frac{1}{4\pi}\int_0^2 r^{k+1}\mathrm{d}r\int_0^{2\pi}\left(\frac{e^{i\theta}+e^{-i\theta}}{2}\right)^k\mathrm{d}\theta$$
$$= \frac{2\pi}{4\pi}\cdot\frac{2^{k+2}}{k+2}\cdot\frac{1}{2^k}\binom{k}{\frac{1}{2}k} = \gamma_k$$

直观上, x_j 为 A_n 特征值

$$\sum_{j=1}^n x_j^2 = \mathrm{Tr}A_n^2, \sum_{j=1}^n\mathbb{E}[x_j^2] = \sum_{i,j=1}^n\mathbb{E}[a_{ij}^2]$$

$$n\mathbb{E}[\overline{x}^2] = n(n-1)+n\mathrm{Var}(Y)$$

可知 $\mathbb{E}[\overline{x}^2] = n-1+\mathrm{Var}(Y), |\overline{x}|\sim\sqrt{n}$

证明

$$\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right] = \frac{1}{n^{1+\frac{1}{2}k}}\sum_{i_1,\cdots,i_k}^n\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]$$

设 $\mathrm{RHS} = \sum_{i_1,\cdots,i_k}^n\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]$

1. $k=1, \mathrm{RHS} = \sum_{i_1}\mathbb{E}[a_{i_1}a_{i_1}] = 0$ 因此 $k=1$ 时

$$\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right] = \frac{1}{n^{\frac{3}{2}}}\mathrm{RHS}\to 0 = \gamma_1$$

2. $k=2$,

$$\mathrm{RHS} = \sum_{i_1,i_2}\mathbb{E}[a_{i_1i_2}a_{i_2i_1}] = n(n-1)+n\mathrm{Var}(Y)$$

因此 $k=2$ 时

$$\frac{1}{n}\mathbb{E}\left[\mathrm{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right] = \frac{1}{n^2}\mathrm{RHS}\to 1 = \gamma_2$$

3. $k=3$,

$$\begin{aligned}\text{RHS} &= \sum_{i_1,i_2,i_3}\mathbb{E}[a_{i_1i_2}a_{i_2i_3}a_{i_3i_1}]\\ &= \sum_{i_1,i_3}\mathbb{E}[a_{i_1i_1}a_{i_1i_3}^2] : O(n) \quad i_1=i_2\\ &+\sum_{i_1,i_2}\mathbb{E}[a_{i_1i_2}^2a_{i_1i_1}] : 0 \quad i_1\neq i_2, i_1=i_3\\ &+\sum_{i_1,i_2}\mathbb{E}[a_{i_1i_2}^2a_{i_2i_2}] : 0 \quad i_1\neq i_2, i_2=i_3\\ &+\sum_{i_1,i_2,i_3}\mathbb{E}[a_{i_1i_2}a_{i_2i_3}a_{i_3i_1}] : 0 \quad i_1,i_2,i_3\text{全不同}\end{aligned}$$

因此 $k=3$ 时,

$$\frac{1}{n}\mathbb{E}\left[\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right]=\frac{1}{n^{\frac{5}{2}}}\text{RHS}\to 0=\gamma_3$$

4. $k=4$, $\text{RHS}=\sum_{i_1,\cdots,i_k}^{n}\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]$,

四条边 $\{i_1,i_2\},\{i_2,i_3\},\{i_3,i_4\},\{i_4,i_1\}$ 中若有一条边与其他不同, 则因期望为 0, 此项必为 0. 每条边若出现, 则非零贡献时, 其出现的次数至少为 2. 分析非消失项具体形式:

若 $i_3\neq i_1$, 对 $a_{i_1i_2}a_{i_2i_3}a_{i_3i_4}a_{i_4i_1}$ 有 $a_{i_1i_2}$ 与 $a_{i_2i_3}$ 不同, 此时 $a_{i_3i_4}$ 必为两者之一, 可知 $i_4=i_2$(否则 $i_1=i_4\neq i_3=i_2$, 此时 $a_{i_1i_2}a_{i_2i_2}a_{i_2i_1}a_{i_1i_1}$ 期望为 0).

下面设 $i_3=i_1$, 非消失项典型形式

- Type1: $i_3=i_1, i_4\neq i_2\neq i_1$

$$\sum_{i_1\neq i_2\neq i_4}\mathbb{E}[a_{i_1i_2}^2a_{i_1i_4}^2]=n(n-1)(n-2)\ \text{贡献}1$$

- Type2: $i_3=i_1, i_4=i_2\neq i_1$

$$\sum_{i_1,i_2}\mathbb{E}[a_{i_1i_2}^4]=O(n^2)$$

- Type3: $i_3=i_1=i_2, i_4\neq i_2$

$$\sum_{i_1\neq i_4}\mathbb{E}[a_{i_1i_1}^2a_{i_1i_4}^2]=O(n^2)$$

- Type4: $i_1=i_2=i_3=i_4$

$$\sum_{i_1}\mathbb{E}[a_{i_1i_1}^4]=O(n)$$

当 Type1, $i_4=i_2$ 但 $i_1\neq i_3\neq i_2$ 时,

$$\sum_{i_1\neq i_2\neq i_3}\mathbb{E}[a_{i_1i_2}^2a_{i_2i_3}^2]=n(n-1)(n-2)\ \text{贡献}1$$

因此 $k=4$ 时

$$\frac{1}{n}\mathbb{E}\left[\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^k\right]=\frac{1}{n^3}\text{RHS}\to 2=\gamma_4$$

5. 一般 k:

$$\text{RHS} = \sum_{i_1,\cdots,i_k}^{n} \mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]$$

下标循环, 期望为 0, 独立性, 高阶矩有限, 非消失项至多 $\left[\frac{k}{2}\right]$ 个不同边, 因此 $i_1,\cdots,i_k$ 中至多有 $1+\left[\frac{k}{2}\right]$ 个自由顶点 (不相同的顶点数). 反证: 否则, 若至少有 $2+\left[\frac{k}{2}\right]$ 个不同顶点, 除第一个外每一个自由顶点带来一条新边, 即至少有 $1+\left[\frac{k}{2}\right]$ 条不同边, 矛盾!

立即知 $\text{RHS} = O\left(n^{1+\left[\frac{k}{2}\right]}\right)$. 进而 $k = 2m+1$ 时结论成立.

当 $k = 2m$ 时, RHS 求和式中主项由下面项构成: 每条边恰好出现 2 次, 且有 $(m+1)$ 个自由顶点. 这时称 $(i_1,i_2,\cdots,i_{2m})$ 为一个不相交的路径.

例. $k = 6$, 两个典型不相交路径

- $(i_1,i_2,i_3,i_4,i_3,i_2)$, i_1,i_2,i_3,i_4 互不相同,

$$\mathbb{E}[a_{i_1i_2}^2a_{i_2i_3}^2a_{i_3i_4}^2] = 1$$

- $(i_1,i_2,i_1,i_3,i_1,i_4)$, i_1,i_2,i_3,i_4 互不相同,

$$\mathbb{E}[a_{i_1i_2}^2a_{i_1i_3}^2a_{i_1i_4}^2] = 1$$

但 $(i_1,i_2,i_3,i_1,i_2,i_3)$,$i_1,i_2,i_3$ 互不相同, 虽然

$$\mathbb{E}[a_{i_1i_2}^2a_{i_2i_3}^2a_{i_3i_1}^2] = 1$$

但是自由顶点数少 1, 不是不相交路径. 计数问题: 同构意义下, $i_1,i_2,\cdots,i_{2m}$ 从 $1,2,\cdots,m+1$ 中取. 对不相交路径 $(i_1,i_2,\cdots,i_{2m})$, 可构造一一对应:$(a_1,a_2,\cdots,a_{2m})$, 定义:

$$a_j = \begin{cases} 1 & i_ji_{j+1}\text{在旅行}i_1,i_2,\cdots,i_{2m},i_1\text{中首次出现} \\ -1 & \text{否则} \end{cases}$$

这里先固定 $i_1 = 1$.

$k = 6$, $i_1,i_2,i_3,i_4,i_3,i_2,i_1$ 对应 $1,1,1,-1,-1,-1$

$i_1,i_2,i_1,i_3,i_1,i_4,i_1$ 对应 $1,-1,1,-1,1,-1$

CHECK: 1-1 对应? 练习

构造一个对称随机游走: 令 $S_i = S_{i-1}+a_i, i = 1,2,\cdots,2m, S_0 = 0$, 则 $S_{2m} = 0, S_i \geq 0, S_0 = 0, \forall i$. 满足条件的轨道数即为 Catalan 数, $\mathrm{C}_m = \frac{1}{m+1}\mathrm{C}_{2m}^m$. 参见之前讲过的母函数或反射原理.

总之, $\text{RHS} = n(n-1)\cdots(n-m)\mathrm{C}_m + O(n^m)$

Wigner 半圆律偶阶矩 C_m 组合诠释, $1,2,\cdots,2m-1,2m$ 两两配对且不相交的个数即为 C_m. $1,2,3,4$ 两两配对的个数为 3, 分别是 1-2,3-4 (non-crossing) 与 1-3,2-4(相交) 与 1-4,2-3(不相交).

标准正态分布 $2m$ 阶矩为 $1,2,\cdots,2m-1,2m$ 两两配对的个数, 故 $N(0,1)$ 四阶矩为 3, 半圆律四阶矩为 2.

练习: $1,2,\cdots,2m-1,2m$ 不相交两两配对数为 C_m.

2.11.4 Wishart 矩阵模型

$X = (x_{ij})_{i=1,\cdots,p;j=1,\cdots,n}$, $\{x_{ij}\}$iid $N(0,1)$, 假设 $n-p=\alpha \geq 0$ 固定, 试证

$$\frac{1}{p}\mathbb{E}\left[\mathrm{Tr}\left(\frac{1}{p}XX^T\right)^m\right] \to \mathrm{C}_m, p \to \infty$$

$m=2$ 时作业.

第 3 章 部分练习题提示与解答

练习 1.1

提示 注意到

$$\begin{aligned}\mathbb{P}\left(\bigcup_{r=1}^{n+1}A_r\right) &= \mathbb{P}\left(\bigcup_{r=1}^{n}A_r\right)+\mathbb{P}(A_{n+1})-\mathbb{P}\left(\bigcup_{r=1}^{n}(A_r\cap A_{n+1})\right)\\ &\geqslant \mathbb{P}\left(\bigcup_{r=1}^{n}A_r\right)+\mathbb{P}(A_{n+1})-\sum_{r=1}^{n}\mathbb{P}(A_r\cap A_{n+1}),\end{aligned}$$

再利用归纳法即可.

练习 1.2

解 记事件 A_n 表示第 n 次抛掷时由甲掷, 记 $p_n=\mathbb{P}(A_n)$. 利用全概率公式, 有

$$p_n=\mathbb{P}(A_n)=\mathbb{P}(A_{n-1})\mathbb{P}(A_n|A_{n-1})+\mathbb{P}(A_{n-1}^c)\mathbb{P}(A_n|A_{n-1}^c)=\frac{5}{6}p_{n-1}+\frac{1}{6}(1-p_{n-1})=\frac{2}{3}p_{n-1}+\frac{1}{6}.$$

利用 $p_1=1$, 整理并求解:

$$p_n-\frac{1}{2}=\frac{2}{3}(p_{n-1}-\frac{1}{2}) \implies p_n=\frac{1}{2}\left(\frac{2}{3}\right)^{n-1}+\frac{1}{2}.$$

练习 1.3

提示 不妨对任意 i, 第 i 名乘客对应第 i 个座位. 把上述 100 一般化为 $n(n\geqslant 2)$, 记最后一名乘客坐自己的座位的概率是 p_n. 用归纳法证明 $p_n=\frac{1}{2}$. 注意到: 若第 1 位乘客选择第 1 个座位, 则最后一名乘客一定坐到自己的座位; 若第 1 位乘客选择第 i 个座位 $(2\leqslant i\leqslant n-1)$, 则第 2 位至第 $i-1$ 位乘客都坐到自己的座位, 之后的情形等同于原题中乘客座位数量为 $n-i+1$ 时的情形; 若第 1 位乘客选择第 n 个座位, 则最后一名乘客一定坐不到自己的座位. 因此, 通过对第一位乘客选择的座位号取条件概率可以得到

$$p_{n+1}=\frac{1}{n+1}+\frac{1}{n+1}\sum_{k=2}^{n}p_k.$$

利用归纳假设即得概率是 $\frac{1}{2}$.

练习 1.4

解 (1) $U[0,1]$ 上的中位数是 $\frac{1}{2}$, $B\left(n,\frac{1}{2}\right)$ 的中位数是 $\begin{cases}[\frac{n-1}{2},\frac{n+1}{2}], n\text{为奇数}\\ \frac{n}{2}, n\text{为偶数}\end{cases}$. (自行验证)

(2) 定义 $m=\sup\left\{x:F(x)<\frac{1}{2}\right\}$, 由 $\lim\limits_{x\to-\infty}F(x)=0$, $\lim\limits_{x\to+\infty}F(x)=1$ 可知 m 存在. 由 m 的上确界定义知, 对 $\forall x<m$, 均有 $F(x)<\frac{1}{2}$. 若 $F(m)<\frac{1}{2}$, 由 F 的右连续性知, $\exists m_0>m$, 使得 $F(m_0)<\frac{1}{2}$, 这与 m 的上确界性矛盾. 因此 $F(m)\geqslant\frac{1}{2}$, 故 m 为 F 的中位数.

再定义 $M=\sup\left\{x:F(x)\leqslant\frac{1}{2}\right\}$. 对 $\forall x>M$, 均有 $F(x)>\frac{1}{2}$; 对 $\forall x<M$, 均有 $F(x)\leqslant\frac{1}{2}$. 由定义知 $M\geqslant m$, 所以 $F(M)\geqslant F(m)\geqslant\frac{1}{2}$. 所以 M 为中位数, 且中位数集合为闭区间 $[m,M]$.

练习 1.5

解 注意到

$$\mathbb{P}(S_n=j)=p\cdot\binom{j-1}{n-1}\cdot p^{n-1}(1-p)^{j-n}=\binom{j-1}{n-1}p^n(1-p)^{j-n}, \cdot j\geqslant n.$$

因而

$$\mathbb{P}(S_{n+1}=k|S_n=j)=\frac{\mathbb{P}(S_{n+1}=k,S_n=j)}{\mathbb{P}(S_n=j)}=\frac{\binom{j-1}{n-1}p^n(1-p)^{j-n}\cdot(1-p)^{k-j-1}p}{\binom{j-1}{n-1}p^n(1-p)^{j-n}}=(1-p)^{k-j-1}p$$

及因而

$$\mathbb{P}(S_n=k|S_{n+1}=j)=\frac{\mathbb{P}(S_n=k,S_{n+1}=j)}{\mathbb{P}(S_{n+1}=j)}=\frac{\binom{k-1}{n-1}p^n(1-p)^{k-n}\cdot(1-p)^{j-k-1}p}{\binom{j-1}{n}p^{n+1}(1-p)^{j-n-1}}=\frac{\binom{k-1}{n-1}}{\binom{j-1}{n}}.$$

练习 1.6

解 记收齐第 $i-1$ 张卡片之后到收齐第 i 张卡片间花费的券数为 η_i, 则有

$$X_n=\eta_1+\eta_2+\cdots+\eta_n.$$

同时, η_i 服从几何分布

$$\mathbb{P}(\eta_i=k)=\left(\frac{i-1}{n}\right)^{k-1}\frac{n-i+1}{n},\quad k=1,2,\cdots$$

因此

$$\mathbb{E}[\eta_i]=\frac{n}{n-i+1}.$$

故有

$$\mathbb{E}[X_n]=\mathbb{E}[\eta_1+\eta_2+\cdots+\eta_n]=\sum_{k=1}^n\mathbb{E}[\eta_k]=n\sum_{k=1}^n\frac{1}{n-k+1}=n\sum_{k=1}^n\frac{1}{k}.$$

利用

$$\ln n<\sum_{k=1}^n\frac{1}{k}<1+\ln(n+1)$$

夹逼即得

$$\lim_{n\to+\infty}\frac{\mathbb{E}[X_n]}{n\ln n}=1.$$

练习 1.8

解 (1) 利用方差的“投影”性质及 $\left|X-\dfrac{a+b}{2}\right|\leqslant\dfrac{b-a}{2}$, 我们有

$$\mathrm{Var}(X)=\min_{t\in\mathbb{R}}\mathbb{E}[(X-t)^2]\leqslant\mathbb{E}[(X-\frac{a+b}{2})^2]\leqslant\frac{(b-a)^2}{4}.$$

或者考虑中心化: $Y=X-\dfrac{b+a}{2}$, 则 $-\dfrac{b-a}{2}\leqslant Y\leqslant\dfrac{b-a}{2}$. 因此

$$\mathrm{Var}(X)=\mathrm{Var}(Y)\leqslant\mathbb{E}[Y^2]\leqslant\frac{(b-a)^2}{4}.$$

(2) 一方面, 由 Cauchy-Schwarz 不等式, 我们有

$$\mathbb{E}[X]\mathbb{E}[\frac{1}{X}]\geqslant(\mathbb{E}[\sqrt{X}\cdot\frac{1}{\sqrt{X}}])=1,$$

当 X 为常值时等号成立. 另一方面, 注意到 $\mathrm{Cov}\left(X,\dfrac{1}{X}\right)=1-\mathbb{E}[X]\mathbb{E}[\dfrac{1}{X}]$, 而

$$\left|\mathrm{Cov}\left(X,\frac{1}{X}\right)\right|\leqslant\sqrt{\mathrm{Var}(X)\mathrm{Var}\left(\frac{1}{X}\right)}\leqslant\sqrt{\frac{(b-a)^2}{4}\cdot\frac{(\frac{1}{a}-\frac{1}{b})^2}{4}}=\frac{(b-a)^2}{4ab},$$

因此

$$1-\mathbb{E}[X]\mathbb{E}[\frac{1}{X}]\geqslant-\frac{(b-a)^2}{4ab}\implies\mathbb{E}[X]\mathbb{E}[\frac{1}{X}]\leqslant\frac{(a+b)^2}{4ab},$$

当 $\mathbb{P}(X=a)=\mathbb{P}(X=b)=\dfrac{1}{2}$ 时等号成立.

练习 1.11

解 我们有

$$\mathbb{E}[K]=\mathbb{E}[\mathbb{E}[K|N]]=\mathbb{E}[\frac{1}{2}N]=\frac{1}{2}\mathbb{E}[N]=\frac{\lambda}{2}.$$

以上也可利用泊松分布在随机选择下的不变性得到 $K\sim P\left(\dfrac{\lambda}{2}\right)$. 而

$$\mathbb{P}(N=n)=\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda},\quad \mathbb{P}(K=k|N=n)=\frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^n.$$

所以

$$\mathbb{P}(N=n|K=k)=\frac{\mathbb{P}(K=k|N=n)\mathbb{P}(N=n)}{\mathbb{P}(K=k)}=\frac{\frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^n\cdot\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}}{\frac{(\frac{\lambda}{2})^k}{k!}\mathrm{e}^{-\lambda/2}}=\frac{(\frac{\lambda}{2})^{n-k}}{(n-k)!}\mathrm{e}^{-\lambda/2}.$$

因此

$$\mathbb{E}[N|K=k]=\sum_{n=k}^{\infty}n\mathbb{P}(N=n|K=k)=\sum_{n=k}^{\infty}(n-k+k)\frac{(\frac{\lambda}{2})^{n-k}}{(n-k)!}\mathrm{e}^{-\lambda/2}=k+\frac{\lambda}{2}.$$

故 $\mathbb{E}[N|K]=K+\dfrac{\lambda}{2}$.

练习 1.12

解 (1) 记 X 表示取出的总球数, 在装有 b 个蓝球和 r 个红球的瓮下记 $M_{b,r}=\mathbb{E}[X]$, 并令 $Y=I_{\{\text{取出的第一个球是蓝球}\}}$, 则有

$$\begin{aligned}M_{b,r}&=\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X|Y]]=\mathbb{E}[X|Y=0]\mathbb{P}(Y=0)+\mathbb{E}[X|Y=1]\mathbb{P}(Y=1)=\frac{r}{b+r}(1+M_{b,r-1})+\frac{b}{b+r}\\&=1+\frac{r}{b+r}M_{b,r-1},\end{aligned}$$

利用 $M_{b,0}=1$, 可归纳得

$$M_{b,r}=\frac{b+r+1}{b+1}\text{ (细节请自行补充)}.$$

(2) 令 M 为过程停止时瓮中取出的总球数, B_r 为移走所有红球时剩下的蓝球数量, B_b 为移走所有蓝球时剩下的红球数量. 则

$$B_r+B_b+M=b+r.$$

假设我们把瓮中所有的球取出并按序排好, 记到第一个蓝球被取出后取出的红球数为 T_r, 由 (1) 知 $\mathbb{E}[T_r]=\dfrac{b+r+1}{b+1}-1=\dfrac{r}{b+1}$, 利用头尾对称性可知

$$\mathbb{E}[B_b]=\mathbb{E}[T_r]=\frac{r}{b+1}.$$

同理, $\mathbb{E}[B_r]=\dfrac{b}{r+1}$. 所以

$$\mathbb{E}[M]=b+r-\mathbb{E}[B_r]-\mathbb{E}[B_b]=\frac{br}{b+1}+\frac{br}{r+1}.$$

注 (1) 也可以用 $\mathbb{E}[X]=\sum_{n=0}^{\infty}\mathbb{P}(X>n)$ 来求解, 或者同 (2) 把所有的球取出并按序排好然后考虑插空.

练习 1.13

解 一方面, 有

$$\mathbb{E}[\mathrm{Var}(Y|X)]=\mathbb{E}[\mathbb{E}[Y^2|X]]-\mathbb{E}[(\mathbb{E}[Y|X])^2]=\mathbb{E}[Y^2]-\mathbb{E}[(\mathbb{E}[Y|X])^2].$$

另一方面, 有

$$\mathrm{Var}(\mathbb{E}[Y|X]) = \mathbb{E}[(\mathbb{E}[Y|X])^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 = \mathbb{E}[(\mathbb{E}[Y|X])^2] - (\mathbb{E}[Y])^2.$$

上述两式相加即得证.

练习 1.16

解 (1) 旋转 45 度后即为在 1 维简单对称随机游走模型中求 $\mathbb{P}(S_0 = S_{2n} = 0, S_1 \geqslant 0, \cdots, S_{2n-1} \geqslant 0)$. 该问已在课堂中用数列母函数方法求解过. 也可利用作业题 **3.10.1**, 结合

$$\mathbb{P}(S_0 = S_{2n} = 0, S_1 \geqslant 0, \cdots, S_{2n-1} \geqslant 0) = 2\mathbb{P}(S_0 = 0, S_{2n+1} = -1, S_1 \geqslant 0, \cdots, S_{2n} \geqslant 0)$$

即可. 答案为 $\dfrac{\binom{2n}{n}}{4^n(n+1)}$.

(2) $C_{3,2} = \dfrac{12}{2^9} = \dfrac{3}{128}$. 记 $a_k = 8^k C_{k,2}$ 为符合题意的路径个数. 我们注意到

- 若首次与直线 $y = \dfrac{1}{2}x$ 相交于 (a_1, b_1), 则必经过 $(a_1 - 1, b_1)$;
- 若首次与直线 $y = \dfrac{1}{2}x + \dfrac{1}{2}$ 相交于 (a_2, b_2), 则必经过 $(a_2 - 1, b_2)$;
- 第一步必定向上走至 $(0, 1)$.

这里 a_i, b_i 为正整数. 对于一条从 $(0, 0)$ 到 $(2n, n)$ 且始终在直线 $y = \dfrac{x}{2}$ 及其上方运动的路径, 必有

$$\{(0,1) \to (a_2 - 1, b_2), \quad (a_2, b_2) \to (a_1 - 1, b_1), \quad (a_1, b_1) \to (2n, n)\}$$

一一对应, 其中 $b_1 = \dfrac{a_1}{2}, b_2 = \dfrac{a_2 + 1}{2}$ 且均为正整数. 故对于 $G(s) = \sum_{k=0}^{\infty} \dfrac{a_k}{8^k} s^k$, 有

$$a_n = \sum_{b_1, b_2} a_{b_1 - 1} a_{b_2 - b_1} a_{n - b_2}.$$

所以

$$\frac{a_n}{8^n} = \frac{1}{8} \sum_{b_1, b_2} \frac{a_{b_1 - 1}}{8^{b_1 - 1}} \frac{a_{b_2 - b_1}}{8^{b_2 - b_1}} \frac{a_{n - b_2}}{8^{n - b_2}}.$$

因此

$$G(s) = 1 + \frac{s}{8} G^3(s).$$

练习 1.18

提示 可以计算 $J^{-1} = -2(v^2 + 1)$. 求出联合密度后可直接得到独立性, 答案:

$$f_U(u) = \frac{1}{2} \mathrm{e}^{-\frac{u}{2}}, \ u > 0, \quad f_V(v) = \frac{1}{\pi(1 + v^2)}, v \in \mathbb{R}.$$

练习 1.19

提示 利用球坐标换元计算即可.

练习 1.20

解 (2) 提示: 用 Fubini 定理证明: 对 $\forall a, b \in \mathbb{R}$, 有

$$\mathbb{E}[|X - b|] - \mathbb{E}[|X - a|] = \int_a^b (\mathbb{P}(X \leqslant x) - \mathbb{P}(X \geqslant x)) \mathrm{d}x$$

(3) 利用 Cauchy-Schwarz 不等式, 结合 (2) 可得

$$|\mathbb{E}[X] - m| = |\mathbb{E}[X - m]| \leqslant \mathbb{E}[|X - m|] \leqslant \mathbb{E}[|X - \mathbb{E}[X]|] \leqslant \sqrt{\mathrm{Var}(X)}.$$

练习 1.21 一方面,

$$\mathbb{P}(A_n > t) \le \sum_{k=1}^{n} \mathbb{P}(|Z_k| > t) = n\mathbb{P}(|Z_1| > t) \le Ant^{-1}\mathrm{e}^{-\frac{t^2}{2}},$$

其中 $A > 0$ 为常数. 因此

$$\begin{aligned}\mathbb{E}[A_n] &= \int_0^{\sqrt{2\log n}} \mathbb{P}(A_n > t)\mathrm{d}t + \int_{\sqrt{2\log n}}^{+\infty} \mathbb{P}(A_n > t)\mathrm{d}t \\ &\le \sqrt{2\log n} + An\int_{\sqrt{2\log n}}^{+\infty} t^{-1}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t \\ &\le \sqrt{2\log n} + A\int_{\sqrt{2\log n}}^{+\infty} t^{-1}\mathrm{d}t \\ &= \sqrt{2\log n} + o(1) \quad (n \to +\infty).\end{aligned}$$

另一方面, 对 $\forall \varepsilon > 0$, 有

$$\mathbb{E}[B_n] \ge (\sqrt{2\log n} - \varepsilon)\mathbb{P}(B_n > \sqrt{2\log n} - \varepsilon).$$

为证原命题, 只需证明

$$\sqrt{\log n}\mathbb{P}(B_n \le \sqrt{2\log n} - \varepsilon) \xrightarrow{n\to+\infty} 0,$$

因为

$$\mathbb{E}[B_n] \ge (\sqrt{2\log n} - \varepsilon)\left(1 - o\left(\frac{1}{\sqrt{\log n}}\right)\right) = \sqrt{2\log n} - \varepsilon + o(1),$$

再令 $\varepsilon \to 0$ 即可. 而

$$\begin{aligned}\mathbb{P}(B_n \le \sqrt{2\log n} - \varepsilon) &= \left(1 - \mathbb{P}(Z_1 > \sqrt{2\log n} - \varepsilon)\right)^n \\ &\le \exp\left(-n\mathbb{P}(Z_1 > \sqrt{2\log n} - \varepsilon)\right) \\ &\le \mathrm{e}^{-(\log n)^2} = o\left(\frac{1}{\sqrt{\log n}}\right),\end{aligned}$$

其中最后一个不等式是因为

$$\mathbb{P}(Z_1 > \sqrt{2\log n} - \varepsilon) = \frac{\mathrm{e}^{-\varepsilon^2/2} + o(1)}{2n\sqrt{\pi\log n}}\exp\left(\varepsilon\sqrt{2\log n}\right) \quad (n \to +\infty).$$

因此

$$\sqrt{\log n}\mathbb{P}(B_n \le \sqrt{2\log n} - \varepsilon) \xrightarrow{n\to+\infty} 0.$$

注意到 $B_n \le A_n$, 综上, 当 $n \to +\infty$ 时, 有

$$\mathbb{E}[A_n] = \sqrt{2\log n} + o(1), \quad \mathbb{E}[B_n] = \sqrt{2\log n} + o(1).$$

练习 1.24

提示: $\Longleftarrow$: 特征函数直接验证; $\Longrightarrow$: 设 X 的特征函数是 $\phi(t)$, 则有

$$\phi(at)\phi(bt) = \phi(\sqrt{a^2+b^2}t).$$

因为 X 二阶矩存在, 故 $\phi(t) = 1 + \mathrm{i}t\mathbb{E}[X] - \dfrac{t^2}{2} + o(t^2)$, 联立可推得 $\mathbb{E}[X] = 0$, 且三阶矩为零, 四阶矩存在. 进一步有

$$\phi(t) = 1 - \frac{t^2}{2} + \frac{t^4}{4!}\mathbb{E}[X^4] + o(t^4).$$

联立再一直重复上述操作即可.

练习 1.26

提示 注意到 $-\ln X_k$ 是参数为 1 的指数分布, 由中心极限定理得

$$\sqrt{n}(\ln Y_n - 1) \xrightarrow{D} N(0,1).$$

同时,

$$\lim_{n\to+\infty}\mathbb{P}(\sqrt{n}(\ln Y_n - 1) \le x) = \lim_{n\to+\infty}\mathbb{P}\left(Y_n \le \mathrm{e}^{\frac{x}{\sqrt{n}}+1}\right) = \lim_{n\to+\infty}\left(\frac{\sqrt{n}(Y_n-\mathrm{e})}{\mathrm{e}} \le \sqrt{n}(\mathrm{e}^{\frac{x}{\sqrt{n}}}-1)\right).$$

利用 $\lim\limits_{n\to+\infty}\sqrt{n}(\mathrm{e}^{\frac{x}{\sqrt{n}}}-1) = x$ 即可.

练习 1.27

提示 注意到 $X_n \sim P(n)$ 可分解为 $X_n = \sum\limits_{k=1}^{n} Y_k$, 其中 Y_k 独立同分布于 $P(1)$.

练习 1.28

提示 验证 Linderberg 条件或者 Lyapunov 条件. 以验证 Linderberg 条件为例, $B_n = \sqrt{\sum\limits_{k=1}^{n}\mu_k^2}$. 对 $\forall\varepsilon > 0$, 一方面,

$$\int_{\mu_k+\varepsilon B_n}^{+\infty}(x_k-\mu_k)^2\frac{1}{\mu_k}\mathrm{e}^{-\frac{x_k}{\mu_k}}\mathrm{d}x_k \xlongequal{y_k=\frac{x_k-\mu_k}{\mu_k}} \frac{\mu_k^2}{\mathrm{e}}\int_{\frac{\varepsilon B_n}{\mu_k}}^{+\infty} y_k^2\mathrm{e}^{-y_k}\mathrm{d}y_k \leqslant \frac{\mu_k^2}{\mathrm{e}}\int_{\frac{\varepsilon B_n}{\max\mu_k}}^{+\infty} y_k^2\mathrm{e}^{-y_k}\mathrm{d}y_k$$

所以

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{\mu_k+\varepsilon B_n}^{+\infty}(x_k-\mu_k)^2\frac{1}{\mu_k}\mathrm{e}^{-\frac{x_k}{\mu_k}}\mathrm{d}x_k = \frac{1}{\mathrm{e}}\int_{\frac{\varepsilon B_n}{\max\mu_k}}^{+\infty} y_k^2\mathrm{e}^{-y_k}\mathrm{d}y_k \xrightarrow{n\to+\infty} 0.$$

另一方面, 存在 $N\in\mathbb{N}^*$, 当 $n > N$ 时, 有 $\dfrac{\max\mu_k}{B_n} < \varepsilon$, 即 $\mu_k - \varepsilon B_n < 0$. 因此 Lindeberg 条件成立.

练习 1.30

证明 (3) 第一个 $\Longleftrightarrow$ 引入示性函数拆分成两部分, 分别估计即可. 第二个 $\Longleftrightarrow$ 见 **Grimmett 7.11.10**.

练习 1.31

提示 利用命题 1.6.

练习 1.32

证明 见 **Grimmett 7.5.2** .

练习 1.33

提示 先中心化:$Z_n = Y_n - p^2$, 然后算到四阶矩后用 Markov 不等式再利用 Borel-Cantelli 引理即可.

练习 2.1 记事件 $C^{(N)}$ 表示 a,b 互素, 事件 $A_i^{(N)}$ 表示素数 $p_i|\gcd(a,b)$, 并记 l 为使得 p_l 是不超过 N 的最大素数. 则对 $\forall 1\leqslant i_1 < i_2 < \cdots < i_k \leqslant l$, 有

$$\mathbb{P}\left(\bigcap_{n=i_1}^{i_k} A_n\right) = \frac{\left\lfloor\dfrac{N}{\prod_{n=i_1}^{i_k} p_n}\right\rfloor^2}{N^2}.$$

利用 Jordan 公式, 有

$$\mathbb{P}(C^{(N)})=\mathbb{P}\left(\left(\bigcup_{n=1}^{l}A_n^{(N)}\right)^c\right)=1-\mathbb{P}\left(\bigcup_{n=1}^{l}A_n^{(N)}\right)=1-\sum_{n=1}^{l}(-1)^n\sum_{1\leqslant i_1<\cdots<i_n\leqslant l}\mathbb{P}(A_{i_1}^{(N)}\cdots A_{i_n}^{(N)})$$

$$=1-\sum_{n=1}^{l}(-1)^n\sum_{1\leqslant i_1<\cdots<i_n\leqslant l}\frac{\left\lfloor\dfrac{N}{\prod_{k=i_1}^{i_n}p_k}\right\rfloor^2}{N^2}.$$

故

$$\left|\mathbb{P}(C^{(N)})-\prod_{i=1}^{l}\left(1-\frac{1}{p_i^2}\right)\right|=\left|\mathbb{P}(C^{(N)})-1+\sum_{n=1}^{l}(-1)^n\sum_{1\leqslant i_1<\cdots<i_n\leqslant l}\frac{\left(\dfrac{N}{\prod_{k=i_1}^{i_n}p_k}\right)^2}{N^2}\right|$$

$$=\frac{1}{N^2}\left|\sum_{n=1}^{l}(-1)^n\sum_{1\leqslant i_1<\cdots<i_n\leqslant l}\left(\left(\frac{N}{\prod_{k=i_1}^{i_n}p_k}\right)^2-\left\lfloor\frac{N}{\prod_{k=i_1}^{i_n}p_k}\right\rfloor^2\right)\right|$$

$$\leqslant\frac{1}{N^2}\sum_{n=1}^{l}\sum_{1\leqslant i_1<\cdots<i_n\leqslant l}\frac{2N}{\prod_{k=i_1}^{i_n}p_k}\leqslant\frac{2}{N}\prod_{i=1}^{l}\left(1+\frac{1}{p_i}\right)$$

$$\leqslant\frac{3}{N}\left(1+\frac{1}{3}\right)\cdot\left(1+\frac{1}{5}\right)\cdots\left(1+\frac{1}{p_l}\right)$$

$$\leqslant\frac{3}{N}\sqrt{\prod_{i=2}^{N+1}\left(1+\frac{1}{i}\right)}$$

$$\leqslant\frac{3\sqrt{N+1}}{N}\xrightarrow{N\to+\infty}0.$$

而由 $\sum_{i=1}^{\infty}\frac{1}{p_i^2}$ 绝对收敛知, $\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^2}\right)$ 收敛, 故有 $\lim_{N\to+\infty}\mathbb{P}(C^{(N)})=\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^2}\right)$.

最后我们证明:

$$\prod_{i=1}^{\infty}\frac{1}{1-\frac{1}{p_i^2}}=\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}.$$

更一般地, 欧拉乘积公式:

$$\zeta(s)=\sum_{n=1}^{\infty}\frac{1}{n^s}=\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^s}\right)^{-1}.$$

利用

$$\frac{1}{1-\frac{1}{p_i^2}}=1+\sum_{n=1}^{\infty}\frac{1}{(p_i^2)^n},$$

有

$$\prod_{p_i\leqslant N}\frac{1}{1-\frac{1}{p_i^2}}=\sum_{n=1}^{N}\frac{1}{n^2}+\sum_{n=N+1}^{\infty}{}'\frac{1}{n^2}$$

其中

$$\sum_{n=N+1}^{\infty}{}'\frac{1}{n^2}<\sum_{n=N+1}^{\infty}\frac{1}{n^2},$$

$\sum\limits_{n=N+1}^{\infty}{}' \frac{1}{n^2}$ 表示: 从 $n \geqslant N+1$ 起, 对每个 n 至多仅有一个 $\frac{1}{n^2}$ 在该级数内. 故有

$$0 < \prod_{p_i \leqslant N} \frac{1}{1-\frac{1}{p_i^2}} - \sum_{n=1}^{N} \frac{1}{n^2} < \sum_{n=N+1}^{\infty} \frac{1}{n^2}.$$

因为 $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$ 收敛, 由夹逼原理知,

$$\prod_{i=1}^{\infty} \frac{1}{1-\frac{1}{p_i^2}} = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

附录 A 一些可以忽略的定理证明

定理 A.1 (矩收敛定理)

设 $\limsup\limits_{k\to\infty}\dfrac{1}{2k}(\gamma_{2k})^{\frac{1}{2k}}=r<\infty$, 则至多存在一个分布函数 F, 使得

$$\gamma_k=\int x^k\mathrm{d}F, k=0,1,\cdots$$

证明 令 $\mu_k=\displaystyle\int|x|^k\,\mathrm{d}F(x)$, 则由 Cauchy-Schwarz 不等式:

$$\mu_{2k+1}\le\sqrt{\mu_{2k}\mu_{2k+2}}$$

因此

$$\limsup_{k\to\infty}\frac{1}{k}(\mu_k)^{\frac{1}{k}}=r<\infty$$

有如下事实:

$$e^{it}-\sum_{m=0}^{n-1}\frac{(it)^m}{m!}=\frac{i^n}{(n-1)!}\int_0^t(t-s)^{n-1}e^{is}\mathrm{d}s$$

$$\left|e^{it}-\sum_{m=0}^{n-1}\frac{(it)^m}{m!}\right|\le\frac{|t|^n}{n!},\forall t\in\mathbb{R}$$

于是:

$$\left|e^{i\theta x}(e^{itx}-\sum_{m=0}^{n-1}\frac{(itx)^m}{m!})\right|\le\frac{|tx|^n}{n!}$$

取期望:

$$\left|\phi(\theta+t)-\sum_{m=0}^{n-1}\frac{t^m}{m!}\phi^{(m)}(\theta)\right|\le\frac{|t|^n}{n!}\mu_n$$

又当 n 充分大时, 对 $\epsilon>0$ 有

$$\mu_n\le(r+\epsilon)^n n^n$$

利用 $\dfrac{n^n}{n!}\le e^n$ 知

$$\phi(\theta+t)=\phi(\theta)+\sum_{m=1}^{\infty}\frac{t^m}{m!}\phi^{(m)}(\theta)\tag{A.1}$$

当 $|t|<\dfrac{1}{er}$ 时, $\dfrac{|t|^n}{n!}\mu_n\le e^n(r+\epsilon)^n|t|^n\le\left(\dfrac{r+\epsilon}{r}\right)^n$, 故上述级数可以展开.

对 $\theta=0,\phi(0)=1,\phi^{(m)}(0)$ 由矩 γ_m 给出, 故决定 $\phi(t),|t|<\dfrac{1}{er}$. 由 (A.1), 再次取 $\theta=\pm\dfrac{1}{er}$ 时, 决定 $\phi(t),|t|<\dfrac{2}{er}$. 如此下去, 矩决定 $\phi(t)$, 进而确定 F.

定理 A.2 (Skorokhod 表示定理)

设 $X_n \xrightarrow{D} X$, 则存在 $(\Omega, \mathcal{F}, \mathbb{P})$, 其上 Y_n, Y 满足:

(1) Y_n 与 Y 同分布, Y 与 X 同分布

(2) $Y_n \xrightarrow{\text{a.s.}} Y$

证明 Fact: F 为分布函数, 定义 "反函数"

$$F^{-1}(y) := \sup\{x : F(x) < y\}, y \in (0,1)$$

则 $F^{-1}(y) \le x \Leftrightarrow y \le F(x)$, 且当 U 为 $(0,1)$ 上均匀分布时, $F^{-1}(U)$ 的分布函数为 F.

下面 CHECK $F^{-1}(y) > x \Leftrightarrow y > F(x)$

$\Leftarrow$, 由 $y > F(x)$ 及右连续性可知, $\exists \delta > 0$,s.t. $y > F(x+\delta)$, 再由 $F^{-1}(y)$ 定义知: $x + \delta \le F^{-1}(y)$, 从而 $x < F^{-1}(y)$.

$\Rightarrow$, 由定义知, $\exists x^* \in \{x : F(x) < y\}$, s.t. $x < x^*$, 从而 $F(x) \le F(x^*) < y$.

令 $Y_n(u) = F_n^{-1}(u), Y(u) = F^{-1}(u), u \in (0,1)$, 则

$$u \le F_n(x) \Leftrightarrow Y_n(u) \le x \tag{A.2}$$

$$u \le F(x) \Leftrightarrow Y(u) \le x \tag{A.3}$$

进而 X_n 与 Y_n, X 与 Y 同分布.

Step1. $\forall \epsilon > 0, u \in (0,1), x \in C_F$,s.t.

$$Y(u) - \epsilon < x < Y(u)$$

(A.3)表明 $F(x) < u$, 结合 $F_n(x) \to F(x)$, 可知对充分大 n, $F_n(x) < u$, 再由 (A.2)知, $Y_n(u) > x$, 从而

$$Y(u) - \epsilon < x < Y_n(u)$$

于是 $\liminf\limits_{n\to\infty} Y_n(u) \ge Y(u)$

Step2. $u < u' < 1, x \in C_F$,s.t. Y 在 u 连续

$$Y(u') < x < Y(u') + \epsilon$$

由(A.3)知 $u < u' \le F(x)$, 结合 $F_n(x) \to F(x)$, 可知对充分大 n,$F_n(x) > u$, 再由 (A.2)知, $Y_n(u) \le x$, 进而

$$Y_n(u) \le x < Y(u') + \epsilon$$

取上极限 $\limsup\limits_{n\to\infty} Y_n(u) \le Y(u')$, 因 Y 单调增, 可取 $u' \in C_F$ 且 $u' \downarrow u$, 于是 $\limsup\limits_{n\to\infty} Y_n(u) \le Y(u)$, 从而 $\lim\limits_{n\to\infty} Y_n(u) = Y(u)$ 在所有的连续点处成立, 因为 Y 是单调增, 不连续点可数, 所以不连续点的概率测度为 0, 证毕.

定理 A.3 (Slutsky 引理)

$X_n \xrightarrow{\text{D}} X, Y_n \xrightarrow{\text{P}} b, Z_n \xrightarrow{\text{P}} c$. 则

$$X_n Y_n + Z_n \xrightarrow{\text{D}} bX + c.$$

证明 $b=0$ 留给读者思考, 以下设 $b>0, Y_n \geqslant 0$. 对 $\forall 0<\varepsilon<b$, 均有

$$
\begin{aligned}
\mathbb{P}(X_nY_n+Z_n \leqslant x) &\leqslant \mathbb{P}(|Y_n-b|>\varepsilon)+\mathbb{P}(|Z_n-c|>\varepsilon)+\mathbb{P}(X_nY_n+Z_n \leqslant x, |Y_n-b| \leqslant \varepsilon, |Z_n-c| \leqslant \varepsilon) \\
&\leqslant \mathbb{P}(|Y_n-b|>\varepsilon)+\mathbb{P}(|Z_n-c|>\varepsilon)+\mathbb{P}\left(X_n \leqslant \frac{x-c+\varepsilon}{b-\varepsilon}\right) \\
&\xrightarrow{n\to+\infty} \mathbb{P}\left(X \leqslant \frac{x-c+\varepsilon}{b-\varepsilon}\right).
\end{aligned}
$$

取 $0<\varepsilon<b$ 使得 $\dfrac{x-c+\varepsilon}{b-\varepsilon} \in C_{F_X}$, 并令 $\varepsilon \downarrow 0$, 有

$$
\limsup_{n\to\infty} \mathbb{P}(X_nY_n+Z_n \leqslant x) \leqslant \mathbb{P}\left(X \leqslant \frac{x-c}{b}\right) = \mathbb{P}(bX+c \leqslant x)
$$

另一方面,

$$
\begin{aligned}
\mathbb{P}(X_nY_n+Z_n \leqslant x) &\geqslant \mathbb{P}(X_nY_n+Z_n \leqslant x, |Y_n-b| \leqslant \varepsilon, |Z_n-c| \leqslant \varepsilon) \\
&\geqslant \mathbb{P}\left(X_n \leqslant \frac{x-c-\varepsilon}{b+\varepsilon}, |Y_n-b| \leqslant \varepsilon, |Z_n-c| \leqslant \varepsilon\right) \\
&= \mathbb{P}\left(X_n \leqslant \frac{x-c-\varepsilon}{b+\varepsilon}\right) - \mathbb{P}\left(|Y_n-b|>\varepsilon\right) - \mathbb{P}\left(|Z_n-c|>\varepsilon\right) \\
&\xrightarrow{n\to+\infty} \mathbb{P}\left(X \leqslant \frac{x-c-\varepsilon}{b+\varepsilon}\right).
\end{aligned}
$$

取 $0<\varepsilon<b$ 使得 $\dfrac{x-c-\varepsilon}{b+\varepsilon} \in C_{F_X}$, 并令 $\varepsilon \downarrow 0$, 有

$$
\liminf_{n\to\infty} \mathbb{P}(X_nY_n+Z_n \leqslant x) \geqslant \mathbb{P}\left(X < \frac{x-c}{b}\right) = \mathbb{P}(bX+c<x) \xlongequal{x\in C_F} \mathbb{P}(bX+c \leqslant x).
$$