随手译的 Grimmett

屑作

作者：Xuxuayame

组织：Xuxuayame Advanced Institute

时间：September 1, 2022

版本：0.1

Illustrator：ももこ

Muaaaaaaaaaaaaaa

目录

第 1 章 事件与其概率

摘要：任何包含随机性的实验可以作为概率空间来建模。这样一个空间包含了由实验的可能结果组成的集合 Ω，事件组成的集合 $\mathcal{F}$，以及概率测度 $\mathbb{P}$。本章探究了概率空间的定义与基本性质，并引入了条件概率与独立性的概念，并举了许多建模与计算的例子。

1.1 导论

我们的大部分生活都基于一种信念"未来是极其不可预测的"。举个例子，像投骰子或轮盘赌这样的概率游戏，如果它们的结果能被提前预知，那将不会出现什么赌徒。我们使用诸如"随机"或"概率"这样的词汇来表达我们对偶然行为的信念，而且我们试图通过游戏或者其他经历，来赋予这些词汇定量或定性的含义。我们对关于概率的命题的主要认识依赖于大量的概念，其中有些比其它更为合理。概率的数学理论将包含那些在普通理性理解中表达和隐含的对机会的概念。这样一种理论将把这些概念描述为公理的集合，并直接导出与实际试验相容的推论。本章包含这样构造的基本要素。

1.2 事件与集合

许多生活中的命题有着这样的形式"A 的可能性（或概率）为 p"，这里 A 是某些事件（比如"明天出太阳"，"剑桥赢了划船比赛"，⋯⋯）而 p 是一个数或描述数量的形容词（比如"八分之一"，"低"，⋯⋯）。A 的发生与否取决于牵涉到的一连串的情况。这些情况称为实验 *(Experiment)* 或试验 *(Trial)*，实验的结果称为结果 *(Outcome)*。总的来说，在实验结束之前我们无法提前准确预知实验结果，我们只能做到列举出可能的结果。

定义 1.1. 样本空间

实验所有可能的结果组成的集合称为样本空间 *(Sample space)*，记为 Ω。♣

例 1.1 轻抛一枚硬币。有两种可能的结果，正面（记为 H）与反面（记为 T），因此 $\Omega=\{H,T\}$。我们可能对以下事件的发生感兴趣：

1. 结果朝上；
2. 结果要么朝上，要么朝下；
3. 结果又朝上又朝下（这似乎极不可能发生）；
4. 结果不朝上。

例 1.2 扔一次骰子。有六种可能的结果，取决于朝上的数字是 $1,2,3,4,5$ 或 6。因此 $\Omega=\{1,2,3,4,5,6\}$。我们可能对以下事件的发生感兴趣：

1. 结果是 1；
2. 结果是偶数；
3. 结果是偶数，但不超过 3；

4. 结果不是偶数。

我们很容易看见这些例子中的每个事件都能描述为对应的样本空间 Ω 的子集 A。在第一个例子中，事件可以重写为：

1. $A = \{H\}$,
2. $A = \{H\} \cup \{T\}$,
3. $A = \{H\} \cap \{T\}$,
4. $A = \{H\}^c$。

以及第二个例子可以写为：

1. $A = \{1\}$,
2. $A = \{2, 4, 6\}$,
3. $A = \{2, 4, 6\} \cap \{1, 2, 3\}$,
4. $A = \{2, 4, 6\}^c$。

Ω 的子集 A 的补集 *(Complement)* 在这里和下文均记为 A^c。从现在开始，Ω 的单元素子集，比如 $\{H\}$，通常书写会去掉外面的括号。

今后我们将事件 *(Events)* 视作样本空间 Ω 的子集。每当我们对事件 A 和 B 感兴趣，我们理所当然也会关心 $A \cup B$，$A \cap B$ 和 A^c，分别代表 "A 或 B"，"A 且 B" 与 "否 A"。我们称事件 A 和 B 是不交的 *(Disjoint)* 若它们的交集为空集 $\varnothing$。$\varnothing$ 称为不可能事件 *(Impossible event)*，Ω 称为必然事件 *(Certain event)*，因为 Ω 的某些成分必然会发生。

因此事件是 Ω 的子集，但是否需要 Ω 所有的子集都是事件呢？答案是否定的，但某些原因在这里讨论太过困难。将事件的集族视作 $\mathcal{P}(\Omega)$ 的一个子集族 $\mathcal{F}$ 对我们而言已然足够。这个子集族应当具有一些确定的性质，以吻合我们前面的讨论：

1. 若 $A, B \in \mathcal{F}$，则 $A \cup B \in \mathcal{F}$ 且 $A \cap B \in \mathcal{F}$;
2. 若 $A \in \mathcal{F}$，则 $A^c \in \mathcal{F}$;
3. 空集 $\varnothing \in \mathcal{F}$。

Ω 的任何满足这三个条件的子集族 $\mathcal{F}$ 称为域 *(Field)*。由域 $\mathcal{F}$ 的上述性质可以得到：

$$\text{若 } A_1, A_2, \cdots, A_n \in \mathcal{F}, \text{ 那么 } \bigcup_{i=1}^{n} A_i \in \mathcal{F}$$

也就是说，$\mathcal{F}$ 在有限并从而在有限交下封闭（见习题 1.3）。如果 Ω 是有限集，这自然很好，但我们在 Ω 为无限集时需要多处理一点这样的一般情况，正如下面的例子所展示的：

例 1.3 重复抛出一枚硬币直到正面朝上。我们关心这一事件发生所抛掷的次数。试验所有可能结果组成的集合为 $\Omega = \{\omega_1, \omega_2, \omega_3, \cdots\}$，这里 ω_i 指的是前 $i-1$ 次均为朝下，第 i 次朝上。我们可能会试着给 $A = \{\omega_2, \omega_4, \omega_6, \cdots\}$，即偶数次首次正面朝上，赋予概率。而这是 Ω 的可数多元素的组合，且我们需要这样的集合属于 $\mathcal{F}$ 以便我们讨论其概率。

因此我们也需要事件的集族在可数并运算下是封闭的。任何拥有这些性质的 Ω 的子集称为 $\sigma-$ 域。

符号	集合术语	概率论术语
Ω	对象的集族	样本空间
ω	Ω 的元素	基本事件，结果
A	Ω 的子集	A 中结果发生
A^c	A 的补集	A 中结果未发生
$A \cap B$	交集	A 和 B 都发生
$A \cup B$	并集	A 或 B 发生，或皆之
$A \setminus B$	差集	A 发生，B 不发生
$A \triangle B$	对称差	A 或 B 发生，但不皆之
$A \subseteq B$	包含	A 发生则 B 发生
$\varnothing$	空集	不可能事件
Ω	全空间	必然事件

表 1.1: 集合论与概率论的术语

定义 1.2. sigma-域

我们称 Ω 的子集族 $\mathcal{F}$ 为$\sigma-$ 域 (σ–field)，若它满足以下条件：

1. $\varnothing \in \mathcal{F}$;
2. 若 $A_1, A_2, \cdots \in \mathcal{F}$，则 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;
3. 若 $A \in \mathcal{F}$，则 $A^c \in \mathcal{F}$。

♣

由习题 1.3，$\sigma-$ 域在可数交运算下也是封闭的。以下是一些 $\sigma-$ 域的例子。

例 1.4 和 Ω 有关的最小的 $\sigma-$ 域为 $\mathcal{F} = \{\varnothing, \Omega\}$。

例 1.5 若 A 为 Ω 的任何子集，那么 $\mathcal{F} = \{\varnothing, A, A^c, \Omega\}$ 为 $\sigma-$ 域。

例 1.6 Ω 的幂集，即 Ω 的所有子集组成的集合，记作 $\{0,1\}^{\Omega}$，显然是 $\sigma-$ 域。不过出于超出本书范围的原因，当 Ω 为无限集时，它的幂集对于概率而言太大而不能合理地为其元素赋予概率大小。

总而言之，对任何实验我们可以联系到一个二元对 $(\Omega, \mathcal{F})$，这里 Ω 为所有可能实验结果，或者说基础事件组成的集合，而 $\mathcal{F}$ 为囊括了所有我们感兴趣的事件的 Ω 的子集的 $\sigma-$ 域。今后，称集合 A 为事件与断言 A 属于 $\sigma-$ 域是等价的。我们经常将有关事件组合的命题翻译成集合论的术语，比如，A 和 B 均发生写作 $A \cap B$。表 1.1 是翻译对照表。

以下为 1.2 节练习。

1. 设 $\{A_i \mid i \in I\}$ 为集族。证明德摩根律 (*De Morgan's Laws*)[1]：
$$\left(\bigcup_i A_i\right)^c = \bigcap_i A_i^c, \quad \left(\bigcap_i A_i\right)^c = \bigcup_i A_i^c$$
2. 设 A 和 B 为某个 $\sigma-$ 域 $\mathcal{F}$ 中的元素。证明 $\mathcal{F}$ 含有集合 $A \cap B$, $A \setminus B$ 和 $A \triangle B$。
3. 传统的淘汰赛（如 Wimbledon[2]）以 2^n 名选手开局，有 n 个回合。假设没有附加赛，初始抽签表已经决定。试对所有可能结果的样本空间给予一个准确的描述。
4. 设 $\mathcal{F}$ 是 Ω 的子集的一个 $\sigma-$ 域，假设 $B \in \mathcal{F}$。证明 $\mathcal{G} = \{A \cap B \mid A \in \mathcal{F}\}$ 是 B 的子集的 $\sigma-$ 域。

[1]奥古斯塔斯 · 德摩根因给出第一个清晰的数学归纳原理的命题而广为人知。他赞扬概率论为：“我们研究的动机是用精神锻炼上的满足去替代有害地享受不道德的刺激。”

[2]温布尔登锦标赛，俗称温布尔登，是世界上最古老的网球锦标赛，被广泛认为是最负盛名的。(译者注)

5. 下面的式子哪些是始终成立的？对于那些不一直成立的，指出它们何时成立：
   (a). $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$；
   (b). $A \cap (B \cap C) = (A \cap B) \cap C$；
   (c). $(A \cup B) \cap C = A \cup (B \cap C)$；
   (d). $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$。

1.3 概率

我们希望能够讨论事件发生的可能性。假设我们大量重复某试验 N 次，尽可能保持初始条件一致，并假设 A 在每次重复试验中可能发生也可能不发生。对于绝大多数科学的实验而言，我们的经验是 A 发生次数的比例在 N 增大时稳定在某个值附近，也就是说，记 $N(A)$ 为 N 次试验中 A 发生的次数，比值 $N(A)/N$ 随着 N 增长表现为收敛到某个常数极限值。我们可以考虑将这个比值的最终值作为 A 在任意某次试验中发生的概率 $\mathbb{P}(A)$[3]，然而实际比值可能并不会连贯地变化，而且此时我们的直觉在这种层面上失效了，但我们这里不应讨论它。在实践中，N 可以取一个大但有限的数，而比值 $N(A)/A$ 可以作为对 $\mathbb{P}(A)$ 的估计。显然，这个比值介于零和一之间，若 $A = \varnothing$ 则 $N(\varnothing) = 0$，比值为 0，而 $A = \Omega$ 时，$N(\Omega) = N$，比值为 1。更进一步，设 A 和 B 是不交的事件，在每次试验中二者均可能发生也可能不发生。那么

$$N(A \cup B) = N(A) + N(B)$$

因而比值 $N(A \cup B)/N$ 是二者 $N(A)/N$, $N(B)/N$ 的和。现在我们认为这些比值代表了相应事件的概率，则上述关系变为

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \quad \mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\Omega) = 1$$

这里的讨论指出了概率函数 $\mathbb{P}$ 应当是*有限可加的 (Finitely additive)*，也就是说

$$\text{若 } A_1, A_2, \cdots, A_n \text{ 为不交事件，则 } \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

例 1.3 指出 $\mathbb{P}$ 应当更广泛地是*可数可加的 (Countably additive)*，因为那种情况下对应的性质应当对可数个不交事件 $A_1, A_2, \cdots$ 成立。

这些关系对于弄清楚概率函数 $\mathbb{P}$ 被应用到事件的集合上而想要具有的性质是充足的。任何这样对 $\mathcal{F}$ 的元素可能性的赋值称为概率测度。有些人不正式地将 $\mathbb{P}$ 指代“概率分布”，特别是当样本空间有限或者可数时。这一举动最好被避免，因为术语“概率分布”是第二章中有另外目的而保留的。

定义 1.3. 概率测度

$(\Omega, \mathcal{F})$ 上的概率测度 (Probability measure)$\mathbb{P}$ 是一个函数 $\mathbb{P} : \mathcal{F} \to [0, 1]$，满足

1. $\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\Omega) = 1$；

[3]这里粗浅的对概率的讨论在诸多地方是不足够的，质疑的读者可能会关心这门学科里面哲学与经验的方面（见附录三）。

2. 若 $A_1, A_2, \cdots$ 为 $\mathcal{F}$ 的不交元素，即 $A_i \cap A_j = \varnothing$ 对所有 $i \neq j$ 成立，则
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

由集合 Ω，其子集的 $\sigma-$ 域 $\mathcal{F}$ 与 $(\Omega, \mathcal{F})$ 上的概率测度 $\mathbb{P}$ 构成的三元组 $(\Omega, \mathcal{F}, \mathbb{P})$，称为概率空间 *(Probability space)*。♣

概率测度是 $(\Omega, \mathcal{F})$ 上所谓测度 *(Measure)* 的一个特别例子。测度是一个函数 $\mu : \mathcal{F} : [0, +\infty)$，满足 $\mu(\varnothing) = 0$ 和上面的第二条。而测度 μ 为概率测度若 $\mu(\Omega) = 1$。

我们可以将概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 与任何实验联系起来，而与实验相关的所有问题都可以在这个空间中重新构建。对某些事件 A，去探求概率 $\mathbb{P}(A)$ 的数值大抵是自然的，而对这个问题的答案应当被包含在对实验的描述中。举个例子，断言公平地抛掷一枚硬币等价于说朝上和朝下具有相同的发生概率；事实上，这正是公平性的定义。

例 1.7 一枚硬币，可能并不均匀，抛出一次。我们可以取 $\Omega = \{H, T\}$ 和 $\mathcal{F} = \{\varnothing, H, T, \Omega\}$，以及一种可能的概率测度 $\mathbb{P} : \mathcal{F} \to [0, 1]$ 由以下给定：
$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(H) = p, \quad \mathbb{P}(T) = 1 - p, \quad \mathbb{P}(\Omega) = 1,$$
这里 p 为区间 $[0, 1]$ 中给定的实数。若 $p = \frac{1}{2}$，则我们说这枚硬币是公平的，或均匀的。

例 1.8 抛掷一次骰子。我们可以取 $\Omega = \{1, 2, 3, 4, 5, 6\}$，$\mathcal{F} = \{0, 1\}^{\Omega}$，而概率测度 $\mathbb{P}$ 为
$$\mathbb{P}(A) = \sum_{i \in A} p_i, \ \forall\ A \subset \Omega$$
这里 $p_1, p_2, \cdots, p_6$ 为区间 $[0, 1]$ 中给定的数，且和为 1。而 i 朝上的概率为 p_i。若 $p_i = \frac{1}{6}$ 对每个 i 都成立，则骰子是公平的，这种情况下
$$\mathbb{P}(A) = \frac{1}{6}|A|, \ \forall\ A \subset \Omega$$
这里 $|A|$ 表示 A 的势。

三元组 $(\Omega, \mathcal{F}, \mathbb{P})$ 表示了典型的概率空间。我们现在给出一些它的简单但是重要的性质。

引理 1.1

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$，
2. 若 $B \supset A$，那么 $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$，
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$，
4. 更一般地，若 $A_1, A_2, \cdots, A_n$ 为事件，则
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots$$
$$+ (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$
♡

证明

1. $A \cup A^c = \Omega$ 且 $A \cap A^c = \varnothing$，故 $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1$。

2. $B = A \cup (B \setminus A)$。而这是不交并，故

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$$

3. $A \cup B = A \cup (B \setminus A)$，为不交并。故由上一结论得

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B \setminus (A \cap B)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\end{aligned}$$

4. 证明由 n 归纳，留作习题（见习题 1.3.4）

在引理 1.1 的 2 中，$B \setminus A$ 表示在 B 中但不在 A 中的元素的集合。为了写出 $\mathbb{P}(B \setminus A)$ 的大小，我们需要 $B \setminus A \in \mathcal{F}$，即 $\mathbb{P}$ 的定义域；当 A 和 B 均在 $\mathcal{F}$ 中时这始终正确，而证明此事是习题 1.2.2 的一部分。注意，每个证明均是将事件表示为不交并，再应用 $\mathbb{P}$。有时计算事件的交比计算它们的并要简单许多；我们等会就会发现引理的 4 十分有用了。下一个关于 $\mathbb{P}$ 的性质则更有技术性，比如说 $\mathbb{P}$ 是连续的集合函数；这一性质本质上等价于 $\mathbb{P}$ 是可数可加的而不仅仅是有限可加的（见问题 1.8. 16）。

引理 1.2

设 $A_1, A_2, \cdots$ 为一事件递增列，使得 $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$，并记 A 为其极限：

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i$$

则 $\mathbb{P}(A) = \lim_{i \to \infty} \mathbb{P}(A_i)$。

类似地，若 $B_1, B_2, \cdots$ 为一事件递减列，使得 $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$，则

$$B = \bigcap_{i=1}^{\infty} B_i = \lim_{i \to \infty} B_i$$

满足 $\mathbb{P}(B) = \lim_{i \to \infty} \mathbb{P}(B_i)$。

♡

证明 $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$ 为一族不交事件的并。因此，由定义 1.3，

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A_1) + \sum_{i=1}^{\infty} \mathbb{P}(A_{i+1} \setminus A_i) \\ &= \mathbb{P}(A_1) + \lim_{n \to \infty} \sum_{i=1}^{n-1} [\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)] \\ &= \lim_{n \to \infty} \mathbb{P}(A_n)\end{aligned}$$

要说明一族递减的事件列的情况，取补集并运用刚才结论（留作习题）。

概括一下，与概率有关的命题隐含着与实验的关联，实验的结果并不完全可预测。对任意这样的实验我们可以关联到一个概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$，它所具有的性质与我们前面对于概率的概念的讨论是一致的。

现在最后提一些术语。一个事件 A 被称作是*空的 (Null)* 若 $\mathbb{P}(A) = 0$。若 $\mathbb{P}(A) = 1$，我们说 A *几乎必然 (Almost surely)* 发生。空事件不应与不可能事件混淆。我们身边无时无刻发生着空时间，即使它们的概率为零；毕竟，飞镖击中目标随便一个点的概率是多少？也就是说，不可能事件是空的，但空事件不必不可能。

以下为 1.3 节练习。

1. 设 A 和 B 为事件，且 $\mathbb{P}(A)=\frac{3}{4}$，$\mathbb{P}(B)=\frac{1}{3}$。说明 $\frac{1}{12}\le\mathbb{P}(A\cap B)\le\frac{1}{3}$，并举例说明两个极端均有可能。对 $\mathbb{P}(A\cup B)$，找到相应的界。
2. 重复抛掷一枚均匀硬币。说明迟早出现正面朝上的概率为一，说明出现任意给定的有限 H,T 列的概率为一。解释和墨菲定律（Murphy's Law）[4]的联系。
3. 有六对杯子和碟子：有两对红色，两对白色，两对带星星纹样。若杯子被随意放置在碟子上（每碟一杯），求没有杯子和碟子配对的概率。
4. 设 $A_1,A_2,\cdots,A_n$ 为事件且 $n\ge 2$，证明
$$\mathbb{P}\left(\bigcup_{i=1}^{n}A_i\right)=\sum_i\mathbb{P}(A_i)-\sum_{i<j}\mathbb{P}(A_i\cap A_j)+\sum_{i<j<k}\mathbb{P}(A_i\cap A_j\cap A_k)-\cdots$$
$$+(-1)^{n+1}\mathbb{P}(A_1\cap A_2\cap\cdots\cap A_n)$$
每袋玉米片中都可能找到剑桥大学的近五位副校长的塑料半身像，而任意给定的袋子中包含任意特定的副校长的概率为 $\frac{1}{5}$，与其它袋子无关。说明近三位副校长的每一员在一批六袋玉米片中被收集的概率是 $1-3(\frac{4}{5})^6+3(\frac{3}{5})^6-(\frac{2}{5})^6$。
5. 设 A_r，$r\ge 1$ 为事件。满足 $\mathbb{P}(A_r)=1$ 对所有 r 成立。说明 $\mathbb{P}(\bigcap_{r=1}^{\infty}A_r)=1$。
6. 假设事件 A_r，$1\le r\le n$ 中至少一个发生，但绝不超过两个发生。若 $\mathbb{P}(A_r)=p$ 且 $\mathbb{P}(A_r\cap A_s)=q$，$r\ne s$，说明 $p\ge\frac{1}{n}$ 而 $q\le\frac{2}{n}$。
7. 假设事件 A_r，$1\le r\le n$，$n\ge 3$ 中至少一个发生，但绝不超过三个发生。至少两个发生的概率为 $\frac{1}{2}$。若 $\mathbb{P}(A_r)=p$，$\mathbb{P}(A_r\cap A_s)=q$，$r\ne s$。且 $\mathbb{P}(A_r\cap A_s\cap A_t)=x$，$r<s<t$，说明 $p\ge\frac{3}{2n}$ 且 $q\le\frac{4}{n}$。

1.4 条件概率

诸多关于概率的命题具有如下形式“若 B 发生，则 A 的概率为 p”，这里 B 和 A 为事件（比如分别为“明天下雨”和“公交车准点到”）而 p 为概率，和之前一样。为了在我们的理论中包含这点，我们简要回顾在前文部分的开头中对比例的讨论。一个实验重复 N 次，并且在每次实验中我们观测 A 和 B 的发生与否，现在，假设我们仅关心 B 发生下的结果；其它所有实验均被忽略。在这更小的实验范围内，A 发生次数的比例为 $N(A\cap B)/N(B)$，因为每个实验中 B 都发生了。然而，
$$\frac{N(A\cap B)}{N(B)}=\frac{N(A\cap B)/N}{N(B)/N}$$
若我们现在认为这些比值为概率，我们看到 B 发生的前提下 A 发生的概率理应定义为 $\mathbb{P}(A\cap B)/\mathbb{P}(B)$。

概率的直觉引导出同样的结论。给定事件 B 发生，这种情况下 A 发生当且仅当 $A\cap B$ 发生。因此给定 B 下 A 的条件概率应当与 $\mathbb{P}(A\cap B)$ 成比例，也就是说它相等于 $\alpha\mathbb{P}(A\cap B)$ 对某些常数 $\alpha=\alpha(B)$ 成立。而给定 B 下 Ω 的条件概率必然是 1，于是因此 $\alpha\mathbb{P}(\Omega\cap B)=1$，导出 $\alpha=1/\mathbb{P}(B)$。

[4]Anything that can go wrong will go wrong.

我们将这些记号正式化如下。

定义 1.4. 条件概率

若 $\mathbb{P}(B) > 0$，那么给定 B 下 A 发生的条件概率 *(Conditional probability)* 定义为

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

我们记条件概率为 $\mathbb{P}(A \mid B)$，读作"给定 B 下 A 发生的概率"，有时也说"在 B 条件下 A 发生的概率"。

例 1.9 投掷两个骰子。假定第一个骰子结果为 3，那么总数超过 6 的概率是多少？答案显然是 $\frac{1}{2}$，因为第二个必须是 $4, 5, 6$。然而，让我们花点功夫分析一下。显然 $\Omega = \{1, 2, 3, 4, 5, 6\}^2$，即所有有序对 (i, j)，$i, j \in \{1, 2, 3, 4, 5, 6\}$ 组成的集合。而且我们可以取 $\mathcal{F}$ 为 Ω 的所有子集构成的集合，而对任意 $A \subseteq \Omega$ 有 $\mathbb{P}(A) = |A|/36$。设 B 为事件"第一个摇到 3,"，而 A 为事件"总数超过 6"。那么

$$B = \{(3, b) \mid 1 \le b \le 6\}, \quad A = \{(a, b) \mid a + b \ge 6\}, \quad A \cap B = \{(3, 4), (3, 5), (3, 6)\},$$

且

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{|A \cap B|}{|B|} = \frac{3}{6}$$

例 1.10 一个家庭有两个孩子。假定至少有一个为男孩，那么两个均为男孩的概率有多大？年长的和年少的均有可能是男性或女性，因此这里有四种可能的性别组合，我们假设它们是等可能的。因此我们可以很清楚地表示样本空间为

$$\Omega = \{GG, GB, BG, BB\}$$

这里 $\mathbb{P}(GG) = \mathbb{P}(BB) = \mathbb{P}(GB) = \mathbb{P}(BG) = \frac{1}{4}$。由条件概率的定义，

$$\begin{aligned}\mathbb{P}(BB \mid \text{至少一个男孩}) &= \mathbb{P}(BB \mid GB \cup BG \cup BB) \\ &= \frac{\mathbb{P}(BB \cap (GB \cup BG \cup BB))}{\mathbb{P}(GB \cup BG \cup BB)} \\ &= \frac{\mathbb{P}(BB)}{\mathbb{P}(GB \cup BG \cup BB)} = \frac{1}{3}\end{aligned}$$

一个很流行但不正确的答案是 $\frac{1}{2}$。这是另外一个问题的正确答案：对一个有着两个孩子的家庭，假定年少的是男孩，则两个均为男孩的概率是多少？在这种情况下，

$$\begin{aligned}\mathbb{P}(BB \mid \text{年少的是男孩}) &= \mathbb{P}(BB \mid GB \cup BB) \\ &= \frac{\mathbb{P}(BB \cap (GB \cup BB))}{\mathbb{P}(GB \cup BB)} = \frac{\mathbb{P}(BB)}{\mathbb{P}(GB \cup BB)} = \frac{1}{2}\end{aligned}$$

一个常见的很危险的论据包含了假设

$$\mathbb{P}(BB \mid \text{一个孩子为男生}) = \mathbb{P}(\text{其他孩子为男生})$$

为什么这是没道理的？[Hint: 考虑样本空间]

下面的引理在概率论中十分重要。一族事件 $B_1, B_2, \cdots, B_n$ 被称为集合 Ω 的分割 *(Partition)* 若

$$B_i \cap B_j = \varnothing,\ i \neq j, \quad \text{以及} \quad \bigcup_{i=1}^{n} B_i = \Omega$$

每个基本事件 $\omega \in \Omega$ 恰属于 Ω 的分割的一个集合。

引理 1.3

对任意事件 A 和 B 满足 $0 < \mathbb{P}(B) < 1$，

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

更一般地，设 $B_1, B_2, \cdots, B_n$ 为 Ω 的分割，满足 $\mathbb{P}(B_i) > 0$ 对所有 i 成立。那么

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

♡

证明 $A = (A \cap B) \cup (A \cap B^c)$。这是不交并，故

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \\ &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)\end{aligned}$$

第二部分类似（见问题 1.8.10）

例 1.11 假设我们有两个瓮，每个都装了一堆有色球。第一个瓮装了两个白球和三个蓝球，而第二个瓮装了三个白球和四个蓝球。随意从第一个瓮中摸取一个球并放入第二个瓮中，再随机从第二个瓮中摸取一个球并验视。它是蓝球的概率多大？除非另有规定，我们假设从任何瓮里摸球，各种结果出现是等可能的。读者会松一口气，因为我们不再需要详细地描述 $(\Omega, \mathcal{F}, \mathbb{P})$ 了；我们有自信在我们需要的时候能够详细描述之。显然，最后摸出的球的颜色取决于从第一个瓮中摸出的球的颜色。所以我们以此为条件。设 A 为事件“最后的球为蓝色”，B 为事件“第一个球为蓝色”。那么，由引理 1.3,

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

我们很容易找到这些概率的值：

$$\begin{gathered}\mathbb{P}(A \mid B) = \mathbb{P}(A \mid \text{第二个瓮装有三个白球和五个蓝球}) = \frac{5}{8}, \\ \mathbb{P}(A \mid B^c) = \mathbb{P}(A \mid \text{第二个瓮装有四个白球和四个蓝球}) = \frac{1}{2}, \\ \mathbb{P}(B) = \frac{3}{5}, \qquad \mathbb{P}(B^c) = \frac{2}{5}\end{gathered}$$

因此

$$\mathbb{P}(A) = \frac{5}{8} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{2}{5} = \frac{23}{40}$$

没准备好的读者可能会吃惊于书中突然出现的瓮。在十七、十八世纪，彩票通常要从瓮里抽取纸条，而投票经常和将纸条或小球放入瓮中有关。在今日的法国，*aller aux urnes* 指的就是投票。因此许多伯努利家族的人和其他人用装有不同颜色小球的瓮来模拟出生、婚姻、死亡、流体、气体等等并非是那么不自然的事。

例 1.12 只有两家工厂生产 zoggle[5]，而工厂 I 生产的 zoggle 有 20% 有缺陷，工厂 II 生产的 zoggle 有 5% 有缺陷。工厂 I 每周生产 zoggle 的量是工厂 II 的两倍。随机从两个工厂的产品中抽取一个 zoggle，它合格的概率有多大？显然合格与否取决于生产的工厂。令

[5]大概是一种玩具，但我查不到。

A 为事件 “选取的 zoggle 合格”，B 为事件 “它来自工厂 I”。如前面所讨论的，

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c) \\ &= \frac{4}{5} \cdot \frac{2}{3} + \frac{19}{20} \cdot \frac{1}{3} = \frac{51}{60}\end{aligned}$$

如果抽取的 zoggle 有缺陷，有多大的可能性它来自工厂 I？在我们的记号中它只是 $\mathbb{P}(B \mid A^c)$。不过

$$\mathbb{P}(B \mid A^c) = \frac{\mathbb{P}(B \cap A^c)}{\mathbb{P}(A^c)} = \frac{\mathbb{P}(A^c \mid B)\mathbb{P}(B)}{\mathbb{P}(A^c)} = \frac{\frac{1}{5} \cdot \frac{2}{3}}{1 - \frac{51}{60}} = \frac{8}{9}$$

本节最后讨论一个警示性的例子。在计算条件概率时，延续逻辑的错误并非不合传统。明确的定义与记号的缺少导致了诸多概率学家误入歧途，甚至包括 Boole，他被 Russell 归功于发现纯数学，被其他人归功于某些计算的逻辑基础。广为人知的" 囚徒悖论" 也阐述了某些危险。

例 1.13 囚徒悖论（Prisoner’s paradox）：在一个黑暗的国度，三名囚犯未经审判便承受牢狱之灾。他们的看守告诉他们，这个国家的独裁者决定随机释放他们中的一员，并处死另外两人，但他无法告诉任何一名囚犯他的命运如何。囚犯 A 于是知道他活下来的机会是 $\frac{1}{3}$。为了得到信息，他私底下请求看守告诉他将死之人的名字（除他自己外），而看守说出来囚犯 B 的名字。现在囚犯 A 对自己存活机会的估计是多少？它可以是 $\frac{1}{2}$：毕竟，他知道生还者要么是 A，要么是 C，除此外他一无所知。它也可以是 $\frac{1}{3}$：毕竟，根据规则，B 和 C 至少有一人会被处死，因此额外的信息没法对 A 早期的概率计算产生有效的影响。读者如何看待此事？悖论的解决方案在于两种情况 B 或 C 都可能发生。

这个悖论的一个替代构想是 Monty Hall 问题，是 1990 年 Marilyn vos Savant（和其他许多人）在 Parade 杂志上提出的问题；见习题 1.4.5。

以下为 1.4 节练习

1. 证明当 $\mathbb{P}(A)\mathbb{P}(B) \neq 0$ 时有 $\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A)\mathbb{P}(A)/\mathbb{P}(B)$。说明若 $\mathbb{P}(A \mid B) > \mathbb{P}(A)$，有 $\mathbb{P}(B \mid A) > \mathbb{P}(B)$。
2. 对事件 $A_1, A_2, \cdots, A_n$，若其满足 $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$，证明

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

3. 一个男人拥有五枚硬币，其中两枚两面都是正面图案，一枚两面都是背面图案，剩下两枚是正常硬币。他闭上眼睛，随机摸取一枚硬币，并抛掷它。硬币朝下的面为正面图案的概率是多少？
   他睁开眼睛并发现硬币正面朝上；朝下的面也是正面图案的概率有多大？
   他再次闭上眼睛，并再次抛掷硬币。朝下的面是正面图案的概率有多大？
   他睁开眼睛并发现硬币正面朝上；朝下的面也是正面图案的概率有多大？
   他扔掉硬币，又随意摸取一枚，并抛掷它。它正面朝上的概率有多大？
4. 你怎么看待下面 Lewis Carrol 对一瓮不容两同色球的 “证明” ？假设瓮含有两个球，每个球可能是黑色或者白色；因此，显然有，$\mathbb{P}(BB) = \mathbb{P}(BW) = \mathbb{P}(WB) = \mathbb{P}(WW) = \frac{1}{4}$。我们加入一个黑球，于是 $\mathbb{P}(BBB) = \mathbb{P}(BBW) = \mathbb{P}(BWB) = \mathbb{P}(BWW) = \frac{1}{4}$。接下来我们随意摸出一个球；球为黑色的概率是（利用条件概率）

$1 \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} = \frac{2}{3}$。然而，如果从三个球中随机摸取一个球为黑球的概率为 $\frac{2}{3}$，那么必然有 2 个黑球和 1 个白球，也就是说一开始瓮中有一个黑球和一个白球。

5. **The Monty Hall problem: goats and cars**
   (a) 残酷的命运推搡着你成为了某游戏秀中的一名参赛者；你必须从三扇门中选择一扇。一扇门后面是崭新的车，另外两扇后面是苍老的山羊。你作出了选择，但你选择的那扇门并未立刻开启。相反，主持人打开了另一扇门，里面是一只山羊，然后他给了你改选第三扇门（还未开启并未被选择）的机会。设 p 为第三扇门后为汽车的（条件）概率。p 的值取决于主持人的方案，设计方案使得 $p = \frac{1}{2}$，$p = \frac{2}{3}$。说明，对 $\alpha \in [\frac{1}{2}, \frac{2}{3}]$，存在一个方案使得 $p = \alpha$。改选第三扇门是明智之举吗？
   (b) 在问题中的变体中，主持人被批准打开被选择的第一扇门，并无论门后是什么，都将送给你作为奖励。如果他选择打开另一扇门，那么必定是因为这扇门后是山羊。设 p 是未开之门后面为车的概率，前提条件是主持人已经打开第二扇门。设计方案使得 $p = 0$ 和 $p = 1$，并推断对任意 $\alpha \in [0, 1]$，存在一个方案使得 $p = \alpha$。
6. **The prosecutor's fallacy**[6]：设 G 为事件"被指控者有罪"，T 为事件"某些证词属实"。一些律师假设 $\mathbb{P}(G \mid T) = \mathbb{P}(T \mid G)$ 并据此争论。说明这成立当且仅当 $\mathbb{P}(G) = \mathbb{P}(T)$。
7. **瓮**：有 n 个瓮，第 r 个瓮中含有 $r-1$ 个红球，$n-r$ 个品红球。你随便选取了一个瓮并不放回地随意拿走了两个球。试求以下事件的概率：
   (a) 第二个球是品红色的；
   (b) 假设第一个球是品红色的，第二个球也是品红色的。

1.5 独立性

一般而言，某些事件 B 的发生会影响其它事件 A 发生的概率，原来的概率 $\mathbb{P}(A)$ 修正为 $\mathbb{P}(A \mid B)$。若概率保持不变，也就是说 $\mathbb{P}(A \mid B) = \mathbb{P}(A)$，则我们称 A 和 B 是"独立的"。这只有在 $\mathbb{P}(B) > 0$ 时才是良好定义的。关于条件概率的定义 1.4 引导我们有如下定义：

定义 1.5. 独立性

我们称事件 A 和 B 是独立的 *(Independent)*，若

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

更一般地，一族 $\{A_i \mid i \in I\}$ 被称作独立的，若

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

[6]这次诉讼所犯之错发生在 1894 年著名的 Dreyfus 案。

对 I 的所有有限子集 J 均成立。♣

注 学生常见的错误是作出谬误命题：A 和 B 是独立的，如果 $A \cap B = \varnothing$。

若族 $\{A_i \mid i \in I\}$ 满足性质：

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j),\ \forall\, i \neq j$$

则称其为成对 *(Pairwise)* 独立的。成对独立的族不一定独立，正如下面例子所展示的。

例 1.14 设 $\Omega = \{abc, acb, cab, cba, bca, bac, aaa, bbb, ccc\}$，且 Ω 中的每个基本事件均具有相同的概率 $\frac{1}{9}$。设 A_k 为第 k 个字母为 a。以下事实留作习题：族 $\{A_1, A_2, A_3\}$ 成对独立但不独立。

例 1.15 回顾例 1.12，事件 A 和 B 显然不独立，因为 $\mathbb{P}(A \mid B) = \frac{4}{5}$ 但 $\mathbb{P}(A) = \frac{51}{60}$。

例 1.16 从一包 52 张卡牌中随机摸取一张，每一张均具有等概率 $\frac{1}{52}$。我们断言所选牌的花色与其点数是独立的。举个例子

$$\mathbb{P}(K) = \frac{4}{52}, \quad \mathbb{P}(K \mid \text{黑桃}) = \frac{1}{13}$$

或者说

$$\mathbb{P}(\text{黑桃}K) = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = \mathbb{P}(\text{黑桃})\mathbb{P}(K)$$

设事件 C 满足 $\mathbb{P}(C) > 0$。条件概率测度 $\mathbb{P}(\cdot \mid C)$ 对应着条件独立 *(Conditional independence)* 的想法。两个事件 A 和 B 被称作关于 C 条件独立 *(Conditionally independent given C)*，若

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C) \tag{1.1}$$

它可以自然扩展到一族事件的情况。[然而，见习题 1.5.5]

以下为 1.5 节练习

1. 设 A 和 B 为独立事件，说明 A^c 和 B 是独立的，并推断 A^c 和 B^c 独立。
2. 我们抛掷骰子 n 次。设 A_{ij} 为事件“第 i 次抛掷与第 j 次抛掷有相同点数”。说明事件 $\{A_{ij} \mid 1 \leq i < j \leq n\}$ 成对独立但不独立。
3. 重复抛掷一枚均匀硬币。说明以下两个命题等价：
   (a) 不同抛掷的结果是独立的。
   (b) 对任意给定的 H, T 列，序列在前 m 次投掷中出现的概率是 2^{-m}，这里 m 为序列的长度。
4. 设 $\Omega = \{1, 2, \cdots, p\}$，$p$ 为素数，$\mathcal{F}$ 为 Ω 所有子集组成的集合，$\mathbb{P}(A) = \frac{|A|}{p}$，$\forall\, A \in \mathcal{F}$。说明，若 A 和 B 为独立事件，则 A 和 B 至少有一者要么是 $\varnothing$，要么是 Ω。
5. 说明 A 和 B 关于 C 条件独立既不蕴含 A 和 B 独立，也不被其蕴含。对何种事件 C，满足对任意的 A 和 B，事件 A 和 B 独立当且仅当它们关于 C 条件独立。
6. **Safe or sorry?**：某种预防的形式在一年份的治疗中据说是 90% 有效的。如果不同年份中有效程度是独立的，说明治疗更有可能在 7 年内失效。
7. **Families.**：Jane 有三个孩子，每个孩子都等可能是男孩或女孩，且与其他孩子无关。

定义事件:

$$A = \{所有孩子均具有相同性别\},$$
$$B = \{至多有一个男孩\},$$
$$C = \{这个家庭包含一个男孩和一个女孩\},$$

(a) 说明 A 独立于 B，B 独立于 C。

(b) A 独立于 C 吗?

(c) 如果男女机会不均等，这些结果还成立吗?

(d) 如果 Jane 有四个孩子，这些结果还成立吗?

8. **Galton's paradox.**: 你翻动三枚均匀硬币。至少两个是一致的，第三枚朝上的正反都是等可能的。因此 $\mathbb{P}(全都一样) = \frac{1}{2}$。你同意吗?

9. 摇晃两个均匀骰子。说明事件 "它们和为 7" 独立于第一个骰子摇出来的点数。

1.6 完备性与乘积空间

这部分在第一遍阅读时应当被删去，但我们在后面会用到其内容。它仅包含一个关于完备概率空间与乘积空间的梗概；读者需要在其它地方来收获更详细的讲解（见 Billingsley 1995）。我们需要以下结果。

引理 1.4

若 $\mathcal{F}$ 与 $\mathcal{G}$ 为两个 Ω 上的 $\sigma-$ 域，则它们的交 $\mathcal{F} \cap \mathcal{G}$ 亦为 $\sigma-$ 域。更一般地，如果 $\{\mathcal{F}_i \mid i \in I\}$ 为一族 Ω 上的 $\sigma-$ 域，那么 $\mathcal{G} = \bigcap_{i\in I} \mathcal{F}_i$ 亦为 $\sigma-$ 域。♡

证明并不困难，并留作习题。注意它们的并 $\mathcal{F} \cup \mathcal{G}$ 不一定为 $\sigma-$ 域，虽然它可以被扩展为唯一的最小的 $\sigma-$ 域，写作 $\sigma(\mathcal{F} \cup \mathcal{G})$，如下。设 $\{\mathcal{G}_i \mid i \in I\}$ 为所有包含 $\mathcal{F}$ 和 $\mathcal{G}$ 的 $\sigma-$ 域构成的族；它是非空的，因为 Ω 的幂集在其中。那么 $\mathcal{G} = \bigcap_{i\in I} \mathcal{G}_i$ 为唯一的最小的包含 $\mathcal{F} \cup \mathcal{G}$ 的 $\sigma-$ 域。

1、完备性

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 为概率空间。任何概率为零的事件 A，即 $\mathbb{P}(A) = 0$，称为零事件 *(Null event)*。看起来让一个零事件 A 的任何子集 B 都为零的是合理的，但这可能毫无意义，因为 B 可能并不是事件，从而 $\mathbb{P}(B)$ 不一定被定义了。

定义 1.6. 完备概率空间

我们称一个概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 是完备的 *(Complete)*，若零事件的所有子集都是事件。♣

任何不完备的空间都可以完备化。设 $\mathcal{N}$ 为 $\mathcal{F}$ 中所有零事件的子集构成的族，并令 $\mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{N})$ 为最小的包含 $\mathcal{F}$ 和 $\mathcal{N}$ 的 $\sigma-$ 域。可以说明，$\mathbb{P}$ 的定义域可以以一种显然的方式从 $\mathcal{F}$ 拓展到 $\mathcal{G}$；$(\Omega, \mathcal{G}, \mathbb{P})$ 称为 $(\Omega, \mathcal{F}, \mathbb{P})$ 的完备化 *(Completion)*。

2、乘积空间

本章所讨论的概率空间经常是围绕单个实验的结果构造出来的，但当我们需要将好几个独立的实验的结果结合起来形成一个空间时（见例 1.3 与 1.9），新的情况便自然产生了。我们如何推广呢？

假设两个实验分别有相应的概率空间 $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$。则共同考虑下，这对实验的样本空间为 $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}$。而合适的 $\sigma-$ 域构造起来更为复杂。当然，它应当包含 $\Omega_1 \times \Omega_2$ 中所有具有形式 $A_1 \times A_2 = \{(a_1, a_2) \mid a_1 \in A_1, a_2 \in A_2\}$ 的子集，这里 A_1 和 A_2 分别是 $\mathcal{F}_1$ 与 $\mathcal{F}_2$ 的典型元素。然而，这样集合构成的族 $\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 \mid A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$ 一般而言并不是 $\sigma-$ 域。通过引理 1.4，存在唯一最小的 $\sigma-$ 域 $\mathcal{G} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2) \subset \Omega_1 \times \Omega_2$ 包含 $\mathcal{F}_1 \times \mathcal{F}_2$ 。而我们现在所需要的只是一个 $(\Omega_1 \times \Omega_2, \mathcal{G})$ 上的合适的概率函数。设 $\mathbb{P}_{12} : \mathcal{F}_1 \times \mathcal{F}_2 \to [0, 1]$ 由下面给出：

$$\mathbb{P}_{12}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2),\ \forall\, A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2 \tag{1.2}$$

可以证明 $\mathbb{P}_{12}$ 的定义域可以由 $\mathcal{F}_1 \times \mathcal{F}_2$ 拓展到整个 $\mathcal{G} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$。而随之而来的概率空间 $(\Omega_1 \times \Omega_2, \mathcal{G}, \mathbb{P}_{12})$ 称为 $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ 的*乘积空间 (Product space)*。更多概率空间的乘积空间可以类似构造。而测度 $\mathbb{P}_{12}$ 有时被称为 “乘积测度”，因为它的定义式 (1.2) 假设了两个实验是独立的。当然还有其它很多测度可以应用到 $(\Omega_1 \times \Omega_2, \mathcal{G})$ 上。

在诸多简单的情况中，这种技术活是不必要的。假设 Ω_1, Ω_2 有限，它们的 $\sigma-$ 域包含所有子集，即例 1.3 和 1.9 的情况。那么 $\mathcal{G}$ 也包含 $\Omega_1 \times \Omega_2$ 的所有子集。

1.7 应用实例

暂时略

1.8 问题

暂时略

第 2 章 随机变量及其分布

摘要：由随机性统治的数值对应于概率空间上的函数，称为随机变量。随机变量的取值可能会受概率的影响，而相关的概率由一个函数来描述，称为分布函数。我们会讨论两类重要的随机变量，即离散型随机变量与连续性随机变量。平均数定律，或者说大数定律，指出在对独立试验的长时运行中，事件发生的比例将收敛到该事件在任意单次试验中发生的概率。这一结果为基于重复实验的对概率的哲学观点提供了数学基础。最后列举了随机变量及其分布的应用实例，并以随机向量与 Monte Carlo 模拟结尾。

2.1 随机变量

我们不应该一直对单次实验本身感兴趣，更应该去考虑它的随机输出的一些结果。举个例子，许多赌徒比起赢而言更在乎输。这样的结果，如果被赋值，可以当做从 Ω 到实直线 $\mathbb{R}$ 的函数，而这些函数称为“随机变量”。

例 2.1 两次抛掷一枚均匀的硬币：$\Omega = \{HH, HT, TH, TT\}$。对 $\omega \in \Omega$，设 $X(\omega)$ 为正面朝上的次数，使得

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0$$

现在假设一名赌徒在实验结果上下注一英镑。他累计赌博，使得他的资产每次正面朝上时翻倍，而正面朝下时输光。他随后的财富 W 为一个随机变量，由以下给出：

$$W(HH) = 4, \quad W(HT) = W(TH) = W(TT) = 0$$

在实验结束，结果 $\omega \in \Omega$ 被获知后，随机变量 $X : \Omega \to \mathbb{R}$ 便取某些值。一般而言数值更有可能落在 $\mathbb{R}$ 的某些特定的子集，而不是其它子集，这取决于概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 与函数 X 它自己。我们希望能够描述 X 可能的取值的可能性的分布。例 2.1 暗示我们可以通过函数 $f : \mathbb{R} \to [0, 1]$ 去做这件事：

$$f(x) = X\text{等于}x\text{的概率}$$

但这样普遍而论是不合适的。我们更倾向于用*分布函数 (Distribution function)* $F : \mathbb{R} \to \mathbb{R}$：

$$F(x) = X\text{不超过}x\text{的概率}$$

更严谨地说，这是

$$F(x) = \mathbb{P}(A(x)) \tag{2.1}$$

这里 $A(x) \subset \Omega$ 由 $A(x) = \{\omega \in \Omega \mid X(\omega) \leq x\}$ 给出。然而，$\mathbb{P}$ 是事件族 $\mathcal{F}$ 上的函数，我们不能在 $A(x)$ 不属于 $\mathcal{F}$ 的时候讨论 $\mathbb{P}(A(x))$，于是我们有下面定义。

定义 2.1. 随机变量

随机变量 (Random variable) 为函数 $X : \Omega \to \mathbb{R}$，满足性质 $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$ 对任意 $x \in \mathbb{R}$ 成立。这样的函数称为$\mathcal{F}-$ 可测的 *($\mathcal{F}-$measurable)*。♣

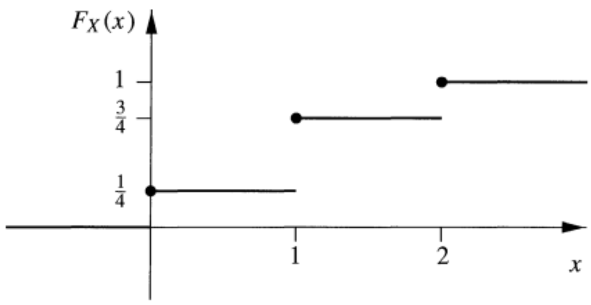

图 2.1: 例 2.1 与例 2.2 中随机变量 X 的分布函数 F_X

如果你愿意，你可以忽略这个定义里技术性的细节而简单地把随机变量当做 Ω 到 $\mathbb{R}$ 的函数。我们应当始终用大写字母，比如 X, Y, Z，来表示一般的随机变量，而用小写字母，如 x, y, z，来表示变量可能的取值。不要在你的手稿中弄混这些记号。

每个随机变量都有一个分布函数，由 (2.1) 给出；分布函数十分重要与实用。

定义 2.2. 分布函数

随机变量 X 的分布函数 (*Distribution function*) 为一个函数 $F : \mathbb{R} \to [0,1]$ 由 $F(x) = \mathbb{P}(X \le x)$ 给出。♣

这是 (2.1) 式明显的简写。事件 $\{\omega \in \Omega \mid X(\omega) \le x\}$ 通常简写为 $\{\omega \mid X(\omega) \le x\}$ 或 $\{X \le x\}$。在有必要强调 X 的角色时，我们写 F_X。

例 2.2 重新审视例 2.1，X 的分布函数 F_X 为

$$F_X(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{4} & 0 \le x < 1, \\ \frac{3}{4} & 1 \le x < 2, \\ 1 & x > 2, \end{cases}$$

并展示在图 2.1 中。而 W 的分布函数 F_W 由下面给出：

$$F_W(x) = \begin{cases} 0 & x < 0, \\ \frac{3}{4} & 0 \le x < 4, \\ 1 & x \ge 4, \end{cases}$$

并展示在图 2.2 中。这阐释了重点，即随机变量 X 的分布函数告诉了我们 X 的取值以及相应的可能性，而不是样本空间和事件族。

引理 2.1

分布函数 F 具有以下性质：

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

(b) 若 $x < y$ 则 $F(x) \le F(y)$，

(c) F 是右连续的，也就是说，$F(x+h) \to F(x)$，$h \to 0^+$。♡

证明

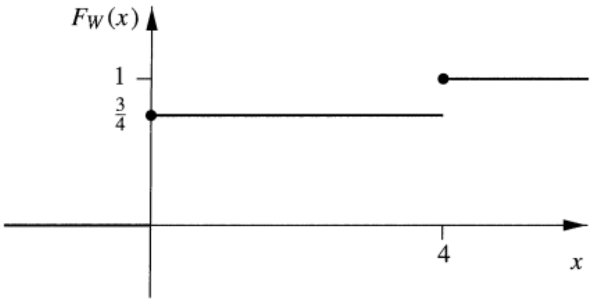

图 2.2: 例 2.1 与例 2.2 中随机变量 W 的分布函数 F_W

(a) 设 $B_n = \{\omega \in \Omega \mid X(\omega) \le -n\} = \{X \le -n\}$。序列 $B_1, B_2, \cdots$ 是递减的，且以空集为极限。因此，由引理 1.2，$\mathbb{P}(B_n) \to \mathbb{P}(\varnothing) = 0$。另一部分类似。

(b) 设 $A(x) = \{X \le x\}$，$A(x,y) = \{x < X \le y\}$。则 $A(y) = A(x) \cup A(x,y)$ 为不交并，于是由定义 1.3，

$$\mathbb{P}(A(y)) = \mathbb{P}(A(x)) + \mathbb{P}(A(x,y))$$

进而

$$F(y) = F(x) + \mathbb{P}(x < X \le y) \ge F(x)$$

(c) 留作练习。使用引理 1.2。

事实上，这个引理恰好刻画了分布函数。换句话说，F 为某个随机变量的分布函数当且仅当它满足上面引理的三个条件。

暂且我们可以忘记概率空间，专注于随机变量和它们的分布函数。X 的分布函数 F 包含了大量关于 X 的信息。

例 2.3 常值变量: 最简单的随机变量在全空间 Ω 上取常值。设 $c \in \mathbb{R}$ 并定义 $X : \Omega \to \mathbb{R}$ 为

$$X(\omega) = c, \ \forall\, \omega \in \Omega$$

分布函数 $F(x) = \mathbb{P}(X \le x)$ 为阶梯函数

$$F(x) = \begin{cases} 0 & x < c, \\ 1 & x \ge c. \end{cases}$$

稍微更广泛点，我们称 X 为常值的 *(几乎必然)(Constant(almost surely))* 若存在 $c \in \mathbb{R}$ 满足 $\mathbb{P}(X = c) = 1$。

例 2.4 伯努利变量: 考虑例 1.7。设 $X : \Omega \to \mathbb{R}$ 由以下给出:

$$X(H) = 1, \quad X(T) = 0.$$

那么 X 为最简单的非平凡随机变量，有两种可能的取值，0 和 1。它的分布函数 $F(x) =$

$\mathbb{P}(X \le x)$ 为

$$F(x) = \begin{cases} 0 & x < 0, \\ 1-p & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$

X 被称为有伯努利分布 *(Bernoulli distribution)*，有时记为 $\mathrm{Bern}(p)$。

例 2.5 示性函数：一类特别的 Bernoulli 变量在概率论中十分实用。设 A 为一个事件，并令 $I_A : \Omega \to \mathbb{R}$ 为 A 的示性函数 *(Indicator function)*，即

$$I_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \in A^c. \end{cases}$$

那么 I_A 为 Bernoulli 随机变量，取值为 1 与 0，概率分别是 $\mathbb{P}(A)$ 与 $\mathbb{P}(A^c)$。假设 $\{B_i \mid i \in I\}$ 为一组不交的事件，满足 $A \subset \bigcup_{i\in I} B_i$，那么

$$I_A = \sum_i I_{A\cap B_i} \tag{2.2}$$

为一个经常有用的恒等式。

引理 2.2

设 F 为 X 的分布函数。那么

(a) $\mathbb{P}(X > x) = 1 - F(x)$,

(b) $\mathbb{P}(x < X \le y) = F(y) - F(x)$,

(c) $\mathbb{P}(X = x) = F(x) - \lim_{y\to x^-} F(y)$。

证明 (a) 和 (b) 留作练习。

(c) 设 $B_n = \{x - \frac{1}{n} < X \le x\}$ 并使用引理 2.1 的证明方式。

注意将来会用的最后一个术语。我们说随机变量 X 与其分布函数 F 有两个“尾巴”，由以下给定：

$$T_1(x) = \mathbb{P}(X > x) = 1 - F(x), \quad T_2(x) = \mathbb{P}(X \le -x) = F(-x)$$

这里 x 为正充分大的数。我们将在后面看见 T_i 在 $x \to \infty$ 时衰减到零的速率对被称之为分布的“矩”的确切相关的数值的存在与否有重大的影响。

以下为 2.1 节练习：

1. 设 X 为给定的概率空间上的随机变量，并设 $a \in \mathbb{R}$。证明
   (i) aX 为随机变量。
   (ii) $X - X = 0$，即这个随机变量始终取值为 0，且 $X + X = 2X$。
2. 随机变量 X 有分布函数 F。那么 $Y = aX + b$ 的分布函数是什么？这里 a, b 为实常数。
3. 抛掷一枚均匀的硬币 n 次。证明，在合理的假设下，恰好 k 次正面朝上的概率为 $\binom{n}{k}(\frac{1}{2})^n$。当每次抛掷正面朝上的概率为 p 时，概率又是多少？
4. 证明若 F 和 G 为分布函数，$0 \le \lambda \le 1$，那么 $\lambda F + (1-\lambda)G$ 为分布函数。其积 FG

也是分布函数吗?

5. 设 F 为分布函数而 r 为正整数。证明以下均为分布函数:
   (a) $F^r(x)$,
   (b) $1-(1-F(x))^r$,
   (c) $F(x)+(1-F(x))\log(1-F(x))$,
   (d) $(F(x)-1)e+\exp(1-F(x))$

2.2 平均律

我们回顾 1.3 节对重复实验的讨论。在每 N 次重复实验中,我们观察给定的事件 A 是否发生,并且记 $N(A)$ 为 A 发生的总次数。概率论的一种可能的哲学上的支撑要求比值 $N(A)/N$ 当 $N\to\infty$ 时稳定在某个极限附近,可以理解为 "A 的概率"。我们的理论迄今为止与这样的要求一致吗?

怀着这样的问题,我们假设 $A_1,A_2,\cdots$ 为一列独立事件且具有等概率 $\mathbb{P}(A_i)=p$,这里 $0<p<1$。这样的假设当然需要相应的概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 的存在性,但我们不打算在这里深挖这个问题。我们考虑 A_i 为事件 "第 i 次实验中 A 发生"。记 $S_n=\sum_{i=1}^n I_{A_i}$,即 $A_1,A_2,\cdots,A_n$ 的示性函数的和。S_n 为一个随机变量,计算了 A_i, $1\le i\le n$ 发生的个数(当然 S_n 是 Ω 的一个函数,因为它是这样函数的和,而 S_n 是 $\mathcal{F}-$ 可测的,留作习题)。下面关于比例 $n^{-1}S_n$ 的结果在 1692 年前被 James Bernoulli 证明。

定理 2.1

当 $n\to\infty$ 时,$n^{-1}S_n$ 收敛到 p 的情况是,对任意 $\varepsilon>0$,

$$\mathbb{P}(p-\varepsilon\le n^{-1}S_n\le p+\varepsilon)\to 1,\quad n\to\infty$$

♡

随机变量敛散性的研究确实需要技术性(见第七章),而这也是这个定理谨慎的表述的原因。我们暂且鼓励读者把这个定理理解为简单地断言事件 $A_1,A_2,\cdots,A_n$ 发生次数的比例 $n^{-1}S_n$ 在 $n\to\infty$ 时收敛到它们的共同概率 p。我们之后会看见作出这样的陈述时小心谨慎是多么重要。

翻译成抛均匀骰子的情况,这个定理蕴含了正面朝上的比例(大概率)接近 $\frac{1}{2}$。作为考虑到学习随机变量敛散性时固有的难处的警告,我们指出,在抛均匀硬币的一个 "典型的" 序列中,大约一半的情况正面朝上的次数超过反面朝上的说法并不正确。

证明 假设我们重复抛一枚硬币,每次抛掷正面朝上的概率为 p。随机变量 S_n 和前 n 次抛掷正面朝上的次数 H_n 有相同的概率分布,也就是说 $\mathbb{P}(S_n=k)=\mathbb{P}(H_n=k)$, $\forall\, k$。于是我们有,对微小的正数 ε,

$$\mathbb{P}\left(\frac{1}{n}S_n\ge p+\varepsilon\right)=\sum_{k\ge n(p+\varepsilon)}\mathbb{P}(H_n=k)$$

由习题 2.1.3 我们有

$$\mathbb{P}(H_n=k)=\binom{n}{k}p^k(1-p)^{n-k},\ \forall\, 0\le k\le n,$$

因此

$$\mathbb{P}\left(\frac{1}{n}S_n \geq p+\varepsilon\right) = \sum_{k=m}^{n}\binom{n}{k}p^k(1-p)^{n-k} \tag{2.3}$$

这里 $m = [n(p+\varepsilon)]$，不小于 $n(p+\varepsilon)$ 的最小整数。下面的论述在概率论中是标准的。设 $\lambda > 0$，注意如果 $k \geq m$，那么 $e^{\lambda k} \geq e^{\lambda n(p+\varepsilon)}$。记 $q = 1-p$，我们有

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}S_n \geq p+\varepsilon\right) &\leq \sum_{k=m}^{n} e^{\lambda[k-n(p+\varepsilon)]}\binom{n}{k}p^k q^{n-k} \\ &\leq e^{-\lambda n\varepsilon}\sum_{k=0}^{n}\binom{n}{k}(pe^{\lambda q})^k(qe^{-\lambda p})^{n-k} \\ &= e^{-\lambda n\varepsilon}(pe^{\lambda q}+qe^{-\lambda p})^n\end{aligned}$$

由二项式定理。容易证明 $e^x \leq x + e^{x^2}$ 对 $x \in \mathbb{R}$ 成立。由此不等式，我们得到

$$\mathbb{P}\left(\frac{1}{n}S_n \geq p+\varepsilon\right) \leq e^{-\lambda n\varepsilon}[pe^{\lambda^2q^2}+qe^{\lambda^2p^2}]^n \leq e^{\lambda^2 n-\lambda n\varepsilon} \tag{2.4}$$

我们可以选取 λ 以最小化右边，即 $\lambda = \frac{1}{2}\varepsilon$，给出

$$\mathbb{P}\left(\frac{1}{n}S_n \geq p+\varepsilon\right) \leq e^{-\frac{1}{4}n\varepsilon^2},\ \varepsilon > 0 \tag{2.5}$$

这个不等式被称为 “Bernstein 不等式”。立刻有 $\mathbb{P}(n^{-1}S_n \geq p+\varepsilon) \to 0,\ n \to \infty$。精确类似的论证可以说明 $\mathbb{P}(n^{-1}S_n \leq p-\varepsilon) \to 0,\ n \to \infty$，因此这个定理被证明了。

Bernstein 不等式 (2.5) 是十分有力的，断言 S_n 超出其平均值一个 n 阶小量的概率随着 $n \to \infty$ 以指数级速度趋于零；这样一个不等式称作 “大偏差估计”。我们可以用这个不等式去证明比这个定理的推论多得多的东西。比起去估计 S_n 对于某些特定值 n 落在 $n(p-\varepsilon)$ 和 $n(p+\varepsilon)$ 之间的概率，让我们估计这件事对所有 n 足够大发生的概率。记 $A_n = \{p-\varepsilon \leq n^{-1}S_n \leq p+\varepsilon\}$，我们希望估计 $\mathbb{P}(\bigcap_{n=m}^{\infty} A_n)$。现在这个交的补事件为事件 $\bigcup_{n=m}^{\infty} A_n^c$，而这个并的概率，由 Boole 和 Bernstein 不等式，满足，

$$\mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n^c\right) \leq \sum_{n=m}^{\infty}\mathbb{P}(A_n^c) \leq \sum_{n=m}^{\infty} 2e^{-\frac{1}{4}n\varepsilon^2} \to 0,\ m \to \infty \tag{2.6}$$

于是得到

$$\mathbb{P}\left(p-\varepsilon \leq \frac{1}{n}S_n \leq p+\varepsilon, \forall\, n \geq m\right) \to 1,\ m \to \infty \tag{2.7}$$

以下为 2.2 节练习：

1. 你希望去询问一大群人中的每一个人一个回答 “是” 会令人尴尬的问题。下面的步骤被提出来以查明人群中尴尬的比例。每问一次问题，在提问者的视线外就抛一次硬币。如果回答本应该是 “否” 而硬币正面朝上，那么便回答 “是”。否则人们如实回答。你如何看待这个步骤?
2. 重复抛掷一枚硬币，每次抛掷正面朝上的概率为 p。设 H_n 和 T_n 分别为 n 次抛掷中正面朝上和反面朝上的次数。证明：对 $\varepsilon > 0$，
$$\mathbb{P}\left(2p-1-\varepsilon \leq \frac{1}{n}(H_n - T_n) \leq 2p-1+\varepsilon\right)$$
3. 设 $\{X_r \mid r \geq 1\}$ 为彼此独立且一致分布有分布函数 F 的观测值。描述并证明一个

估计 $F(x)$ 的方法。

2.3 离散型与连续型变量

对随机变量的诸多研究都致力于分布函数，以引理 2.1 为特征。关于分布函数及其应用的一般理论是十分困难而抽象的，最好在这个阶段删去。它依赖于对 Lebesgue-Stieltjes 积分的构造的严格处理；这会在 5.6 节概括一下。然而，如果我们已经准备好将注意力限制到随机变量的确切的子类上，且具有一些令其可驾驭的性质，事情将变得简单许多。我们应当深入了解“离散型”随机变量和“连续型”随机变量。

定义 2.3. 离散型随机变量

随机变量 X 被称为离散的 (Discrete) 若其仅在 $\mathbb{R}$ 的某些可数子集 $\{x_1, x_2, \cdots\}$ 上取值。离散型随机变量 X 具有 (概率) 质量函数 ((Probability) mass function) $f : \mathbb{R} \to [0, 1]$，$f(x) = \mathbb{P}(X = x)$。 ♣

我们应当看见，离散型随机变量的分布函数在 $x_1, x_2, \cdots$ 处是跳跃间断的，而在它们之间则为常值；这样一个分布称为*元的 (Atomic)*。这与我们下面考虑的另一类重要的分布函数差异明显。

定义 2.4. 连续型随机变量

随机变量 X 被称为连续的 (Continuous) 若其分布函数可以表示为

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u, \quad x \in \mathbb{R}$$

这里 $f : \mathbb{R} \to [0, \infty)$ 为可积函数，称为 X 的 (概率) 密度函数 ((Probability) density function)。 ♣

连续型随机变量的分布函数当然是连续的 (事实上它是“绝对连续的”)。我们暂时只考虑离散型随机变量和连续型随机变量。还有另外一种随机变量，称为“*奇异的 (Singular)*”，关于它的讨论读者应当去其它地方查看。这种现象的一个常见例子是基于 Cantor 三分集的（见 Grimmett and Welsh 1986，或 Billingsley 1995）。其它的随机变量为离散型，连续型和奇异型随机变量的“混合”。注意“连续型”在这里属于是用词不当：描述 X 是连续型时，我们指的是它的分布函数的性质而不是随机变量 X 它自己。

例 2.6 离散型随机变量：例 2.1 中的随机变量 X 和 W 分别在 $\{0, 1, 2\}$ 和 $\{0, 4\}$ 上取值；它们都是离散的。

例 2.7 连续型随机变量：一根直竿被随意抛到水平面上，测得竿和正北方的夹角为 ω。结果为 $\Omega = [0, 2\pi)$ 中的数。暂时不要在意 $\mathcal{F}$；我们可以假设 $\mathcal{F}$ 包含了 Ω 的所有好的子集，包括那些开子区间如 (a, b)，这里 $0 \le a < b < 2\pi$。隐含的对称性暗示了概率测度 $\mathbb{P}$ 满足 $\mathbb{P}((a, b)) = (b - a)/(2\pi)$；也就是说，角度落到某些区间的概率是直接与区间长度成比例的。这是两个随机变量 X 和 Y：

$$X(\omega) = \omega, \qquad Y(\omega) = \omega^2.$$

注意 Y 是 X 的函数，因为 $Y = X^2$。X 和 Y 的分布函数为

$$F_X(x) = \begin{cases} 0 & x \le 0, \\ x/(2\pi) & 0 \le x < 2\pi, \\ 1 & x \ge 2\pi, \end{cases} \quad F_Y(y) = \begin{cases} 0 & y \le 0, \\ \sqrt{y}/(2\pi) & 0 \le y < 4\pi^2, \\ 1 & y \ge 4\pi^2 \end{cases}$$

为了说明这件事，设 $0 \le x > 2\pi$ 且 $0 \le y < 4\pi^2$。那么

$$\begin{aligned} F_X(x) &= \mathbb{P}(\{\omega \in \Omega \mid 0 \le X(\omega) \le x\}) \\ &= \mathbb{P}(\{\omega \in \Omega \mid 0 \le \omega \le x\}) = x/(2\pi), \\ F_Y(y) &= \mathbb{P}(\{\omega \mid Y(\omega) \le y\}) \\ &= \mathbb{P}(\{\omega \mid \omega^2 \le y\}) = \mathbb{P}(\{\omega \mid 0 \le \omega \le \sqrt{y}\}) = \mathbb{P}(X \le \sqrt{y}) \\ &= \sqrt{y}/(2\pi) \end{aligned}$$

随机变量 X 和 Y 是连续的，因为

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}\,u, \qquad F_Y(y) = \int_{-\infty}^{y} f_Y(u)\,\mathrm{d}\,u,$$

这里

$$f_X(u) = \begin{cases} 1/(2\pi) & 0 \le u \le 2\pi, \\ 0 & \text{其它情况}, \end{cases}$$

$$f_Y(u) = \begin{cases} u^{-\frac{1}{2}}/(4\pi) & 0 \le u \le 4\pi^2, \\ 0 & \text{其它情况}, \end{cases}$$

例 2.8 一个既不连续也不离散的随机变量：抛掷一枚硬币，正面朝上的概率为 $p(= 1 - q)$。若正面朝上则朝地面丢下一根杆，角度测量如例 2.7 所述。那么 $\Omega = \{T\} \cup \{(H, x) \mid 0 \le x < 2\pi\}$，记号显然。设 $X : \Omega \to \mathbb{R}$ 由下面给出

$$X(T) = -1, \qquad X((H, x)) = x.$$

随机变量 X 取值在 $\{-1\} \cup [0, 2\pi)$ 上（图 2.3 是其分布函数的草图）。我们称 X 是连续的除了“在 -1 处的点质量（或元）”。

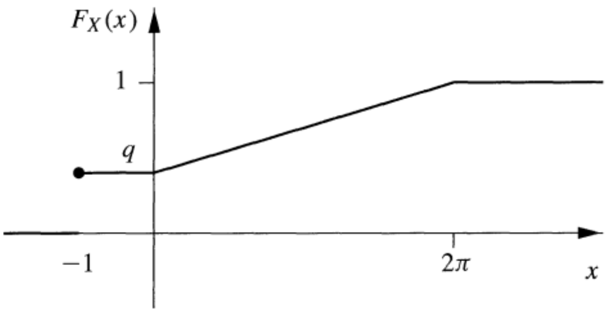

图 2.3: 例 2.8 中随机变量 X 的分布函数 F_X

以下为 2.3 节练习：

1. 设 X 为随机变量，分布函数为 F，并设 $a = (a_m \mid -\infty < a_m < \infty)$ 为严格单调

增的实数列，满足 $a_{-m} \to -\infty$ 和 $a_m \to \infty$，只要 $m \to \infty$。定义 $G(x) = \mathbb{P}(X \leq a_m)$，$a_{m-1} \leq x < a_m$，使得 G 为离散型随机变量的分布函数。选取序列 a 使得 $\sup_m |a_m - a_{m-1}|$ 越来越小，此时 G 如何变化？

2. 设 X 为随机变量，$g : \mathbb{R} \to \mathbb{R}$ 为连续且严格单调增的函数。证明 $Y = g(X)$ 为随机变量。
3. 设 X 为随机变量，拥有分布函数：
$$\mathbb{P}(X \leq x) = \begin{cases} 0 & x \leq 0, \\ x & 0 < x \leq 1, \\ 1 & x > 1. \end{cases}$$
设 F 为连续且严格单调增的分布函数。证明 $Y = F^{-1}(X)$ 为随机变量，且 F 为其分布函数。F 连续且/或严格单调增是必要的吗？
4. 证明，若 f 和 g 为密度函数，且 $0 \leq \lambda \leq 1$，那么 $\lambda f + (1-\lambda)g$ 为密度函数。积 fg 是密度函数吗？
5. 下面哪个是密度函数？对于那些是密度函数的函数，找到 c 与相应的分布函数。
   (a) $f(x) = \begin{cases} cx^{-d} & x > 1, \\ 0 & \text{其它情况}. \end{cases}$
   (b) $f(x) = ce^x(1+e^x)^{-2}$，$x \in \mathbb{R}$。

2.4 应用实例

例 2.9 向一个半径 3 的圆靶投掷飞镖。我们可以把受击点视为随机实验的结果；简单起见，我们应当假设玩家保证击中了靶子。将靶子中心设为 $\mathbb{R}^2$ 的原点，我们可以看见该实验的样本空间为
$$\Omega = \{(x,y) \mid x^2 + y^2 < 9\}.$$
不要在意事件族 $\mathcal{F}$。大致来说，我们假设飞镖落在某个区域 A 的概率是正比于其面积 $|A|$ 的。因此
$$\mathbb{P}(A) = |A|/(9\pi). \tag{2.8}$$
得分系统如下。目标靶被三个同心圆 C_1, C_2, C_3 分割，圆心在原点，半径依次为 $1,2,3$。这些圆将目标靶划分为三个环 A_1, A_2 与 A_3，这里
$$A_k = \{(x,y) \mid k-1 \leq \sqrt{x^2+y^2} < k\}.$$
我们假设玩家得到 k 分，当且仅当飞镖击中 A_k。则结果得分 X 为随机变量：
$$X(\omega) = k,\ \omega \in A_k$$
它的分布函数是什么？

解答： 显然
$$\mathbb{P}(X = k) = \mathbb{P}(A_k) = |A_k|/(9\pi) = \frac{1}{9}(2k-1),\ =1,2,3,$$

故 X 的分布函数为

$$F_X(r) = \mathbb{P}(X \le r) = \begin{cases} 0 & r < 1, \\ \frac{1}{9}\lfloor r \rfloor^2 & 1 \le r < 3, \\ 1 & r \ge 3, \end{cases}$$

这里 $\lfloor r \rfloor$ 表示不超过 r 的最大整数（见图 2.4）。

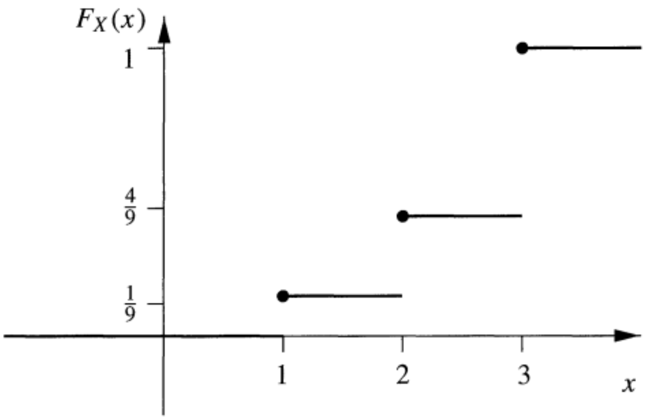

图 2.4: 例 2.9 中 X 的分布函数 F_X

例 2.10 继续例 2.9：让我们考虑另一种计分方式，玩家得分恰为受击点 ω 和靶心的距离。这一次得分 Y 为随机变量：

$$Y(\omega) = \sqrt{x^2 + y^2}, \quad \omega = (x, y).$$

Y 的分布函数是什么？

解答：对任意实数 r，令 C_r 为圆心 $(0,0)$，半径为 r 的圆盘，即

$$C_r = \{(x, y) \mid x^2 + y^2 \le r\}.$$

那么

$$F_Y(r) = \mathbb{P}(Y \le r) = \mathbb{P}(C_r) = \frac{1}{9}r^2, \quad 0 \le r \le 3.$$

如图 2.5 所示。

例 2.11 继续例 2.9：现假设玩家未能击中靶子的概率为 p；若他成功击中，则我们假设受击点的分布如方程 (2.8) 所描述。他的分数按如下方式决定。若他击中靶子，则他得分与受击点到中心的距离一致；若他未能击中靶子，则他得 4 分。他的得分 Z 的分布函数是什么？

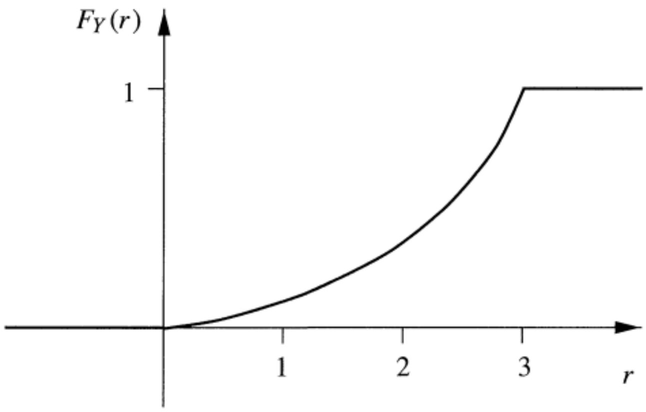

图 2.5: 例 2.10 中 Y 的分布函数 F_Y

解答：显然 Z 在区间 $[0,4]$ 中取值。利用引理 1.3 我们有

$$\begin{aligned}F_Z(r) &= \mathbb{P}(Z \le r)\\ &= \mathbb{P}(Z \le r \mid \text{击中靶子})\mathbb{P}(\text{击中靶子}) + \mathbb{P}(Z \le r \mid \text{未击中靶子})\mathbb{P}(\text{未击中靶子})\\ &= \begin{cases} 0 & r < 0,\\ (1-p)F_Y(r) & 0 \le r < 4,\\ 1 & r \ge 4,\end{cases}\end{aligned}$$

这里 F_Y 由例 2.10 给出，F_Z 图像如图 2.6 所示。

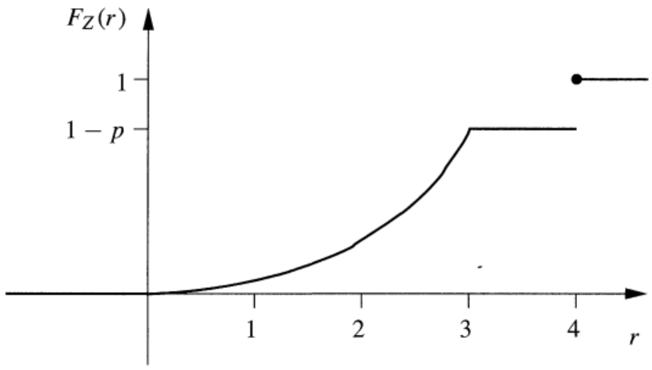

图 2.6: 例 2.11 中 Z 的分布函数 F_Z

以下为 2.4 节练习：

1. 设 X 为随机变量，有连续的分布函数 F。求下列随机变量的分布函数的表达式：
   (a) X^2,
   (b) $\sqrt{X}$,
   (c) $\sin X$,
   (d) $G^{-1}(X)$,
   (e) $F(X)$,
   (f) $G^{-1}(F(X))$,

   这里 G 为连续且严格单调增的函数。

2. **截断**：设 X 为随机变量，分布函数为 F，并设 $a<b$。作出“截断的”随机变量 Y 和 Z 的分布函数的草图：

$$Y=\begin{cases}a & X<a,\\ X & a\leq X\leq b,\\ b & X>b,\end{cases}\qquad Z=\begin{cases}X & |X|\leq b,\\ 0 & |X|>b,\end{cases}$$

指出这些分布函数在 $a\to-\infty,b\to\infty$ 时的行为如何。

2.5 随机向量

假设 X 和 Y 为概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量。它们的分布函数，F_X 和 F_Y，包含了它们相关概率的信息。但我们如何去封装关于它们互相关联的性质的信息呢？关键在于将 X 和 Y 视作一个取值在 $\mathbb{R}^2$ 上的“随机向量”(X,Y) 的分量，而不是作为不相关的随机变量，各自取值在 $\mathbb{R}$ 中。

例 2.12 Tontine[1]是一种方案，其中每一名订阅了相同基金的订阅者都会在他或她的有生之年从该基金中收到一笔年金，且年金会随着其他订阅者的逝去而增加。当所有订阅者都故去，基金将传递到法国政府手中（这是第一种这样的方案，由 Lorenzo Tonti 在 1653 年附近设计）。基金的表现取决于订阅者的寿命 $L_1,L_2,\cdots,L_n$（也取决于他们的财富），而我们可以将这些记为一个随机变量的向量 $(L_1,L_2,\cdots,L_n)$。

例 2.13 飞镖：一枚飞镖被投向传统的镖靶。击打的位置决定了到中心的距离 R，与竖直向上间的夹角 Θ（比如顺时针测量），以及分数 S。关于这个实验我们可以关联到随机向量 (R,Θ,S)，而且我们指出 S 是 (R,Θ) 的函数。

例 2.14 抛硬币：假设我们抛掷一枚硬币 n 次，并设 X_i 等于 0 或 1，这取决于第 i 次抛掷的结果是反面还是正面。我们考虑向量 $\mathbf{X}=(X_1,X_2,\cdots,X_n)$ 以描述这个复合实验的结果。正面的总数为 $\mathbf{X}$ 的分量之和。

一个单独的随机变量 X 有分布函数 F_X，定义为 $F_X(x)=\mathbb{P}(X\leq x)$, $x\in\mathbb{R}$。随机向量 $(X_1,X_2,\cdots,X_n)$ 的对应的“联合”分布函数为 $\mathbb{P}(X_1\leq x_1,X_2\leq x_2,\cdots,X_n\leq x_n)$，为 n 个实变量 $x_1,x_2,\cdots,x_n$ 的函数。为了辅助这个记号，我们介绍实数向量的一种序关系：对向量 $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ 和 $\mathbf{y}=(y_1,y_2,\cdots,y_n)$，我们记 $\mathbf{x}\leq\mathbf{y}$，若 $x_i\leq y_i$ 对每个 $i=1,2,\cdots,n$ 成立。

定义 2.5. 联合分布函数

概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 中的随机向量 $\mathbf{X}=(X_1,X_2,\cdots,X_n)$ 的联合分布函数 *(Joint distribution function)* 为函数 $F_{\mathbf{X}}:\mathbb{R}^n\to[0,1]$，定义为 $F_{\mathbf{X}}(\mathbf{x})=\mathbb{P}(\mathbf{X}\leq\mathbf{x})$, $\mathbf{x}\in\mathbb{R}^n$。♣

正如以往，表达式 $\{\mathbf{X}\leq\mathbf{x}\}$ 是事件 $\{\omega\in\Omega\mid\mathbf{X}(\omega)\leq\mathbf{x}\}$ 的缩写。联合分布函数与那些常规的分布函数有着相似的性质。举个例子，引理 2.1 变成如下形式：

[1]唐提养老金制，又称联合养老保险制

引理 2.3

随机向量 (X,Y) 的联合分布函数 $F_{X,Y}$ 有以下性质：

(a) $\lim_{x,y\to-\infty} F_{X,Y}(x,y)=0$, $\lim_{x,y\to\infty} F_{X,Y}(x,y)=1$,

(b) 若 $(x_1,y_1)\le(x_2,y_2)$，那么 $F_{X,Y}(x_1,y_1)\le F_{X,Y}(x_2,y_2)$,

(c) $F_{X,Y}$ 是上连续的，因为

$$F_{X,Y}(x+u,y+v)\to F_{X,Y}(x,y),\quad u,v\to 0^+$$

我们表述了引理在随机向量仅有两个分量 X 和 Y 的情况，但相应的结果在 n 个分量的情况也是成立的。引理的证明留作习题。更多的结果也是成立的。不需要很难就能看见

$$\lim_{y\to\infty} F_{X,Y}(x,y)=F_X(x)(=\mathbb{P}(X\le x)) \tag{2.9}$$

以及类似地

$$\lim_{x\to\infty} F_{X,Y}(x,y)=F_Y(y)(=\mathbb{P}(Y\le y)). \tag{2.10}$$

这个引理的 (a) 部分的加强版本告诉我们，我们可以从已知的联合分布函数中重新捕捉到 X 和 Y 的单独的分布函数。但反过来则不对：并不总是可以从 F_X 和 F_Y 单独的信息中计算 $F_{X,Y}$。函数 F_X 和 F_Y 称为 $F_{X,Y}$ 的“边缘”分布函数。

例 2.15 一名学校教师要求他或她班上的每位学生去抛掷两次一枚均匀的硬币并记录结果。勤奋的小学生 D 如此做了并记录了一对结果 (X_D,Y_D)。懒惰的小学生 L 只抛了一次硬币并将结果写了两遍，因此记录了一对 (X_L,Y_L)，这里 $X_L=Y_L$。显然 X_D,Y_D,X_L，和 Y_L 为具有相同分布函数的随机变量。然而，(X_D,Y_D) 与 (X_L,Y_L) 有着不同的联合分布函数。特别地，$\mathbb{P}(X_D=Y_D=\text{正面})=\frac{1}{4}$ 因为四种可能的结果中仅有一种只包含正面，然而 $\mathbb{P}(X_L=Y_L=\text{正面})=\frac{1}{2}$。

再一次，这里有两类随机向量特别有趣：“离散型”与“连续型”。

定义 2.6. 联合离散

概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量 X 和 Y 称为 (联合) 离散的 ((Jointly)discrete)，如果向量 (X,Y) 仅在 $\mathbb{R}^2$ 的某个可数子集中取值。联合离散型随机变量 X,Y 有 联合 (概率) 质量函数 (Joint(probability)mass function)$f:\mathbb{R}^2\to[0,1]$，定义为 $f(x,y)=\mathbb{P}(X=x,Y=y)$。

定义 2.7. 联合连续

概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量 X 和 Y 称为 (联合) 连续的 ((Jointly)continuous)，如果它们的联合分布函数可以表示为

$$F_{X,Y}(x,y)=\int_{u=-\infty}^{x}\int_{v=-\infty}^{y} f(u,v)\,\mathrm{d}v\,\mathrm{d}u \qquad x,y\in\mathbb{R},$$

对某些可积函数 $f:\mathbb{R}^2\to[0,\infty)$ 成立。f 称为 (X,Y) 的联合 (概率) 密度函数 (Jointly(Probability)density function)。

我们应当在后续章节中回到这些问题。同时这里有两个具体的例子。

例 2.16 三边硬币：我们被提供了一枚特别的三边硬币，每一次抛掷可能会出现 H(正面)，T(反面)，E(侧面)，每种都有 $\frac{1}{3}$ 的概率。设 H_n, T_n，和 E_n 为抛掷 n 次该硬币相应结果的数目。向量 (H_n, T_n, E_n) 为随机变量的向量，满足 $H_n + T_n + E_n = n$。若不同抛掷的结果互不干扰，那么不难看出

$$\mathbb{P}((H_n, T_n, E_n) = (h, t, e)) = \frac{n!}{h!t!e!}\left(\frac{1}{3}\right)^n$$

对任意和为 n 的非负整数三元组 (h, t, e) 成立。随机变量 H_n, T_n, E_n 为（联合）离散的，并称（联合）拥有三项分布。

例 2.17 飞镖：回到例 2.13 的丢飞镖，让我们假设镖靶上没有区域比起任何其它等面积的区域是被过分偏袒的。那么可以证明（见例 2.10）

$$\mathbb{P}(R \le r) = \frac{r^2}{\rho^2}, \quad \mathbb{P}(\Theta \le \theta) = \frac{\theta}{2\pi}, \qquad 0 \le r \le \rho,\ 0 \le \theta \le 2\pi,$$

这里 ρ 为镖靶的半径，进一步

$$\mathbb{P}(R \le r, \Theta \le \theta) = \mathbb{P}(R \le r)\mathbb{P}(\Theta \le \theta).$$

于是有

$$F_{R,\Theta}(r, \theta) = \int_{u=0}^{r}\int_{v=0}^{\theta} f(u, v)\,\mathrm{d}\,v\,\mathrm{d}\,v$$

这里

$$f(u, v) = \frac{u}{\pi\rho^2}, \qquad 0 \le u \le \rho,\ 0 \le v \le 2\pi.$$

(R, Θ) 是（联合）连续的。

以下为 2.5 节练习：

1. 抛掷两次一枚均匀的硬币。设 X 为正面朝上的次数，W 为事件 $\{X = 2\}$ 的示性函数。试求 $\mathbb{P}(X = x, W = w)$ 对所有合适的 x 和 w 的值。
2. 设 X 为 Bernoulli 随机变量，使得 $\mathbb{P}(X = 0) = 1 - p$，$\mathbb{P}(X = 1) = p$。令 $Y = 1 - X$ 而 $Z = XY$。求 $\mathbb{P}(X = x, Y = y)$ 与 $\mathbb{P}(X = x, Z = z)$ 对 $x, y, z \in \{0, 1\}$。
3. 随机变量 X 和 Y 有联合分布函数
$$F_{X,Y}(x, y) = \begin{cases} 0 & x < 0, \\ (1 - e^{-x})\left(\frac{1}{2} + \frac{1}{\pi}\tan^{-1} y\right) & x \ge 0. \end{cases}$$
证明 X 和 Y 是（联合）连续分布的。
4. 设 X 和 Y 有联合分布函数 F。证明
$$\mathbb{P}(a < X \le b, c < Y \le d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$
当 $a < b$ 且 $c < d$。
5. 设 X, Y 为离散型随机变量，取值为整数，且有联合质量函数 f。证明，对整数 x, y，
$$\begin{aligned} f(x, y) =& \mathbb{P}(X \ge x, Y \le y) - \mathbb{P}(X \ge x + 1, Y \le y) \\ & - \mathbb{P}(X \ge x, Y \le y - 1) + \mathbb{P}(X \ge x + 1, Y \le y - 1). \end{aligned}$$
从而求抛掷一枚均匀的骰子 r 次中出现的最小数与最大数的联合质量函数。
6. 函数 $F(x, y) = 1 - e^{-xy}$，$0 \le x, y < \infty$ 是某对随机变量的联合分布函数吗？

2.6 Monte Carlo 模拟

暂时略

2.7 问题

第 3 章 离散型随机变量

摘要：离散型随机变量的分布可以通过它的概率质量函数来确定。本章介绍了离散型随机变量关于独立性的关键概念。对离散型变量，也定义了期望，或者说平均值的概念，而引出了离散型随机变量的方差和矩的定义。介绍了联合分布，条件分布，和条件期望，伴随协方差和相关性的想法。展示了 Cauchy-Schwarz 不等式。对随机变量的和的分析引出了质量函数的卷积公式。在一定深度上学习了随机游走，包括反射原理，投票定理，停时定理，反正弦律

3.1 概率质量函数

回忆一下，称随机变量 X 是离散的，若它仅在某些可数集 $\{x_1, x_2, \cdots\}$ 上取值。它的分布函数 $F(x) = \mathbb{P}(X \leq x)$ 为跳跃函数；而它的质量函数同它的分布函数一样重要。

定义 3.1. （概率）质量函数

离散型随机变量 X 的 (概率) 质量函数 *((Probability)mass function)*[a] 为函数 $f : \mathbb{R} \to [0, 1]$，定义为 $f(x) = \mathbb{P}(X = x)$。

[a] 某些地方宽松地用 X 的质量函数指代其分布。。

♣

分布与质量函数有如下关系

$$F(x) = \sum_{i:x_i \leq x} f(x_i), \qquad f(x) = F(x) - \lim_{y \to x^-} F(y).$$

引理 3.1

概率质量函数 $f : \mathbb{R} \to [0, 1]$ 满足：

(a) 满足 $f(x) \neq 0$ 的 x 的集合是可数的，

(b) $\sum_i f(x_i) = 1$，这里 $x_1, x_2, \cdots$ 为 x 满足 $f(x) \neq 0$ 的值。

♡

证明 证明是显然的。

这个引理表征了概率质量函数。

例 3.1 二项分布：抛掷一枚硬币 n 次，每次正面朝上的概率为 $p(= 1 - q)$。那么 $\Omega = \{H, T\}^n$。正面的总数 X 在集合 $\{0, 1, 2, \cdots, n\}$ 中取值且为离散型随机变量。它的概率质量函数 $f(x) = \mathbb{P}(X = x)$ 满足

$$f(x) = 0, \quad x \notin \{0, 1, 2, \cdots, n\}.$$

设 $0 \leq k \leq n$，考虑 $f(k)$。Ω 中恰好 $\binom{n}{k}$ 个点给出了总数为 k 的正面；每个点的发生概率为 $p^k q^{n-k}$，故

$$f(k) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leq k \leq n.$$

称随机变量 X 拥有参数为 n 和 p 的二项分布 *(Binomial distribution)*，写作 $\text{bin}(n, p)$。它是 n 个 Bernoulli 变量的和 $X = Y_1 + Y_2 + \cdots + Y_n$ （见例 2.4）。

例 3.2 泊松分布：若取值在 $\{0,1,2,\cdots\}$ 上的随机变量 X 具有质量函数

$$f(k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots$$

这里 $\lambda>0$，则称 X 具有参数为 λ 的泊松分布 (*Poisson distribution*)。

以下为 3.1 节练习：

1. 常数 C 取何值时，下列定义了正整数 $1,2,\cdots$ 上的质量函数？
   (a) 几何的：$f(x)=C2^{-x}$。
   (b) 对数的：$f(x)=C2^{-x}/x$。
   (c) 平方倒数的：$f(x)=Cx^{-2}$。
   (d) “修正” Poisson 的：$f(x)=C2^x/x!$。
2. 对随机变量 X（依次）拥有上题中的质量函数，求：
   (i) $\mathbb{P}(X>1)$，
   (ii) X 的最可能的值，
   (iii) X 为偶数的概率。
3. 我们抛掷 n 枚硬币，每一枚正面朝上的概率均为 p，且相互独立。正面朝上的硬币会被重抛。第二轮抛掷后正面朝上的数目的质量函数是什么？
4. 设 S_k 为所有满足十进制展开恰好为 k 位数的正整数构成的集合（使得，比如说，$1024\in S_4$）。抛掷一枚硬币，直到正面首次朝上，并记 T 为所需抛掷数。从 S_T 中随意等概率抽取一个元素，记为 N。N 的质量函数是什么？
5. **对数凸性**：
   (a) 证明，若 X 为二项或 Poisson 随机变量，则质量函数 $f(k)=\mathbb{P}(X=k)$ 拥有性质 $f(k-1)f(k+1)\le f^2(k)$。
   (b) 证明，若 $f(k)=90/(\pi k)^4$，$k\ge 1$，则 $f(k-1)f(k+1)\ge f^2(k)$。
   (c) 求质量函数 f 满足 $f^2(k)=f(k-1)f(k+1)$，$k\ge 1$。

3.2 独立性

请记住，称事件 A 和 B “独立”，若 A 的发生不改变 B 随后发生的概率。更严谨地说，A 和 B 独立当且仅当 $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$。类似地，我们称离散型变量 X 和 Y 是“独立的”，如果 X 的数值不影响 Y 的分布。考虑到这一点我们有如下定义。

定义 3.2. 独立

离散型随机变量 X 和 Y 是独立的，如果事件 $\{X=x\}$ 与 $\{Y=y\}$ 对所有 x 和 y 都独立。♣

假设 X 在 $\{x_1,x_2,\cdots\}$ 中取值，Y 在 $\{y_1,y_2,\cdots\}$ 中取值。令

$$A_i=\{X=x_i\},\quad B_j=\{Y=y_j\}.$$

注意到 (见问题 2.7.2)X 和 Y 为示性变量 I_{A_i}, I_{B_j} 的线性组合，因为

$$X = \sum_i x_i I_{A_i}, \quad Y = \sum_j y_j I_{B_j}$$

随机变量 X 和 Y 独立当且仅当 A_i 和 B_j 对所有 i, j 都独立。类似的定义对离散型变量族 $\{X_1, X_2, \cdots, X_n\}$ 都成立。

例 3.3 Poisson 翻动：抛掷一次硬币，正面朝上概率为 $p = 1 - q$。记 X 和 Y 分别为正面与反面的次数。不出意料 X 和 Y 并不独立。毕竟，

$$\mathbb{P}(X = Y = 1) = 0, \quad \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = p(1 - p).$$

假设现在硬币被抛出 N 次，N 是个随机数字，满足参数为 λ 的 Poisson 分布。这是个卓越非凡的事实，结果中正面与反面的次数 X 和 Y 是独立的，因为

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

而由引理 1.3，

$$\begin{aligned}\mathbb{P}(X = x) &= \sum_{n \ge x} \mathbb{P}(X = x \mid N = n)\mathbb{P}(N = n) \\ &= \sum_{n \ge x} \binom{n}{x} p^x q^{n-x} \frac{\lambda^n}{n!} e^{-\lambda} = \frac{(\lambda p)^x}{x!} e^{-\lambda p};\end{aligned}$$

对 Y 有类似的结果成立，于是

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

若 X 为随机变量，$g : \mathbb{R} \to \mathbb{R}$，那么 $Z = g(X)$，由 $Z(\omega) = g(X(\omega))$ 定义，也是随机变量。我们将需要以下内容。

定理 3.1

若 X 和 Y 独立，$g, h : \mathbb{R} \to \mathbb{R}$，则 $g(X)$ 和 $h(Y)$ 也独立。♡

证明 留作习题。见问题 (3.11.1)

更一般地，我们称（离散型）随机变量的族 $\{X_i \mid i \in I\}$ 是独立的，如果事件族 $\{X_i = x_i\}$, $i \in I$，对所有可能的在 X_i 上取值的集合 $\{x_i \mid i \in I\}$ 的选择都独立。也就是说，$\{X_i \mid i \in I\}$ 为独立的族当且仅当

$$\mathbb{P}(X_i = x_i, \forall\, i \in J) = \prod_{i \in J} \mathbb{P}(X_i = x_i)$$

对所有集合 $\{x_i \mid i \in I\}$ 与 I 的所有有限子集 J 成立。给定事件 C，一族随机变量的条件独立的定义是类似于一族事件的条件独立的；见方程 (1.1)。

随机变量的独立族研究起来比不独立的情况容易得多，我们将会很快看到这点。值得一提的是成对独立的族不一定独立。

以下为 3.2 节练习：

1. 设 X 和 Y 为独立的随机变量，每个取值 -1 或 1 的概率均为 $\frac{1}{2}$，并设 $Z = XY$。证明 X, Y 和 Z 是成对独立的。它们独立吗？

2. 设 X 和 Y 为取值在正整数中的独立的随机变量，且具有相同的质量函数 $f(x) = 2^{-x}$，$x = 1, 2, \cdots$。求：
   (a) $\mathbb{P}(\min\{X, Y\} \leq x)$，
   (b) $\mathbb{P}(Y > X)$，
   (c) $\mathbb{P}(X = Y)$，
   (d) $\mathbb{P}(X \geq kY)$，k 为给定的正整数，
   (e) $\mathbb{P}(X \mid Y)$[1]，
   (f) $\mathbb{P}(X = rY)$，r 为给定的正有理数。
3. 设 X_1, X_2, X_3 为取值在正整数中的独立的随机变量，且具有相同的质量函数 $\mathbb{P}(X_i = x) = (1 - p_i)p_i^{x-1}$，对 $x = 1, 2, \cdots$ 与 $i = 1, 2, 3$。
   (a) 证明
   $$\mathbb{P}(X_1 < X_2 < X_3) = \frac{(1 - p_1)(1 - p_2)p_2 p_3^2}{(1 - p_2 p_3)(1 - p_1 p_2 p_3)}.$$
   (b) 求 $\mathbb{P}(X_1 \leq X_2 \leq X_3)$。
4. 三名玩家，A, B 和 C，轮流丢骰子；他们以 $ABCABCA \cdots$ 的顺序进行。
   (a) 证明，三名玩家中 A 首先投出 6，B 其次，C 第三的概率为 $216/1001$。
   (b) 证明，第一个 6 被 A 投出，第二个 6 被 B 投出，第三个 6 被 C 投出的概率为 $46656/753571$。
5. 设 X_r，$1 \leq r \leq n$，为关于 0 对称的独立的随机变量；也就是说，X_r 和 $-X_r$ 具有相同的分布。证明，对所有 x，$\mathbb{P}(S_n \geq x) = \mathbb{P}(S_n \leq -x)$，这里 $S_n = \sum_{r=1}^{n} X_r$。没有独立性的假设，结论还一定正确吗？

3.3 期望

设 $x_1, x_2, \cdots, x_N$ 为重复某些实验 N 次得到的数值结果。这些结果的平均值为
$$m = \frac{1}{N} \sum_i x_i$$
在进行这些实验之前我们可以将它们的结果用一列随机变量 $X_1, X_2, \cdots, X_N$ 来表示，而且我们应当假设这些变量是离散的，拥有共同的质量函数 f。那么，粗略地讲（见 1.3 节开头），对每个可能的取值 x，X_i 中大约有 $Nf(x)$ 恰好取之。所以平均数 m 大约为
$$m \simeq \frac{1}{N} \sum_x xNf(x) = \sum_x xf(x)$$
这里求和遍历了所有可能的 X_i 的取值。这一平均称为在质量函数 f 分布下的“期望”或“均值”。

[1]整除，不是条件概率

3.4 示性函数与配对

这一节较为轻松愉悦，打着一些对示性函数的说明的幌子。例 2.5 已给出了它的定义，从那以后它便偶尔出现。回忆一下，

$$I_A(\omega) = \begin{cases} 1 & \text{若}\omega \in A, \\ 0 & \text{若}\omega \in A^c, \end{cases}$$

而且 $\mathbb{E}I_A = \mathbb{P}(A)$。

例 3.4 对引理 1.1(3)(4) 的证明：注意到

$$I_A + I_{A^c} = I_{A\cup A^c} = I_\Omega = 1$$

而且 $I_{A\cap B} = I_A I_B$。因此

$$\begin{aligned} I_{A\cup B} &= 1 - I_{(A\cup B)^c} = 1 - I_{A^c\cap B^c} \\ &= 1 - I_{A^c}I_{B^c} = 1 - (1 - I_A)(1 - I_B) \\ &= I_A + I_B - I_A I_B \end{aligned}$$

取期望即得

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

更一般地，若 $B = \bigcup_{i=1}^n A_i$ 那么

$$I_B = 1 - \prod_{i=1}^n (1 - I_{A_i});$$

将右边乘开并取期望得

$$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap \cdots \cap A_n). \qquad (3.1)$$

这就是所谓的容斥原理 *(Inclusion-exclusion formula)*。

例 3.5 配对问题：