

第一章

概率空间和随机事件

自然界中的有些现象,在某些条件下必然发生,我们称之为确定性现象.例如,每天太阳会从东方升起,无支撑的重物会从高处落下,等等.还有很多现象,在条件给定的情况下是否发生并不确定.例如,掷一枚均匀硬币,观察落下后是正面朝上还是反面朝上,掷之前我们不能确定将会出现哪种情况.又如,某处某个时刻的气温,事先我们无法知道确切的结果,只有到了那个时刻测量后才能够知道.这种在给定条件下是否发生具有不确定性的现象称为随机现象.

在本章中,我们首先对随机试验、古典概型、几何概型等做简单的介绍.在此基础上,引入概率空间的概念,定义条件概率,证明全概率公式和贝叶斯公式,并给出事件和试验的独立性等性质的严格数学描述.作为对上述内容的扩展,我们还将介绍抽象测度的概念和基本性质.这些构成后面各章的理论基础.

1.1 随机试验和事件

1.1.1 随机试验

对于随机现象的观测和分析构成随机试验,其基本特征是在相同的条件下多次重复,且所有可能发生的结果构成的集合是确定不变的.随机试验在大量重复试验中会表现出规律性.

例 1.1.1 连续投掷一枚质地均匀的硬币,每次的结果是不确定的,出现正面和反面的可能性均为 $1/2$. 用 S_n 表示前 n 次投掷中出现正面的次数. 根据经验,当 n 很大时,比值 S_n/n 总是很接近于一次投掷中正面出现的可能性 $1/2$,从而显现出规律性. 这个规律可以利用本课程后面将要学习的大数定律给出严格的数学解释.

例 1.1.2 英国生物统计学家高尔顿 (Galton) 设计的钉板装置是用来研究随机现象的统计规律的模型. 钉板结构如图 1.1 所示,木板上钉有多层铁钉,每层相邻铁钉间的距离相同,且上层每颗钉的水平位置恰好位于下层两颗钉的正中间. 从入口处放进一个直径略小于两钉距离的玻璃圆球,圆球每次碰到铁钉后以 $1/2$ 的概率向左或向右滚下,直到滚到底部的某个格子内为止. 把大量的圆球不断从入口处放下. 当球的数目很大时,它们会在底部堆成中间高两边低的形状,从而显现出规律性. 这个规律可利用本课程后面将要证明的中心极限定理给出严格的数学解释.

为了准确地描述随机现象,我们需要使用集合论的语言和记号. 我们知道,集合是由一些元素构成的总体,其元素的个数可以是有限的,也可以是可数无限或不可数无限的. 有限集和可数无限集统称为可数集. 由一个集合 A 中的部分或全部元素构成的另一个集合 B 称为 A 的子集. 不包含任何元素的集合称为空集,记作 $\emptyset$. 我们规定空集是任何集合的子集.

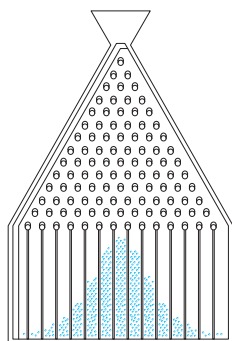


图 1.1 高尔顿钉板试验

本课程假定读者熟悉关于集合的基本概念, 运算和记号. 例如 $A = B$, $A \subseteq B$, $A \supseteq B$, $A \cap B$, $A \cup B$, $A \setminus B$ 等. 如果有一个明确的基本集合 Ω , 对于子集 $A \subseteq \Omega$, 我们称 $\Omega \setminus A$ 为 A 的补集或余集, 并将其记作 A^c .

某个随机试验所有可能的结果构成的 (非空) 集合称为它的样本空间, 通常用大写希腊字母 Ω 表示. 称 Ω 中的元素为样本点, 用小写希腊字母 ω 表示. 我们关心的随机事件可以用 Ω 的某些子集来表示. 某个随机事件 $A \subseteq \Omega$ 的概率是区间 $[0, 1]$ 中的一个数 $\mathbf{P}(A)$, 它给出 A 发生的可能性大小的描述. 我们称样本空间 Ω 为必然事件并规定 $\mathbf{P}(\Omega) = 1$, 而称空集 $\emptyset$ 为不可能事件并规定 $\mathbf{P}(\emptyset) = 0$. 事件 A 的余集 A^c 也称为它的对立事件, 显然 $\mathbf{P}(A) + \mathbf{P}(A^c) = 1$.

假设在一个随机试验中, 我们关心的仅仅是事件 $A \subseteq \Omega$ 是否发生, 并且已知 $\mathbf{P}(A) = p$ 和 $\mathbf{P}(A^c) = 1 - p$, 其中 $0 \leq p \leq 1$. 这样的随机试验称为伯努利 (Bernoulli) 试验. 在伯努利试验中, 事件 A 发生称为成功, 事件 A^c 发生称为失败. 对于伯努利试验, 为了简单我们可以将样本空间取为 $\Omega = \{1, 0\}$, 其中 1 表示 A 发生, 而 0 表示 A^c 发生.

例 1.1.3 掷一枚硬币, 用 1 表示结果正面朝上, 用 0 表示反面朝上. 则试验的样本空间为 $\Omega = \{1, 0\}$. 显然, 这是一种伯努利试验. 若硬币质地均匀, 则有 $\mathbf{P}(\{1\}) = \mathbf{P}(\{0\}) = 1/2$.

例 1.1.4 投掷骰子试验的样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$. 在这个随机试验中, 我们感兴趣的事件包括 $A = \{6\}$, $B = \{2, 4, 6\}$ 等. 若骰子质地均匀, 则每个点数出现的概率均为 $1/6$.

例 1.1.5 记录一台电话交换机在某个时段内收到的呼唤次数. 这个试验的样本空间可取为 $\Omega = \{0, 1, 2, \dots\}$.

例 1.1.6 射手连续向某个目标射击, 直至命中为止, 考虑射击的总次数. 这个试验的样本空间为 $\Omega = \{1, 01, 001, 0001, \dots\}$.

例 1.1.7 观察北京夏季某日的最低气温 x 和最高气温 y (单位: $^{\circ}\text{C}$). 这个试验的样本空间可取为 $\Omega = \{(x, y) : 0 < x \leq y < 50\}$.

考虑以 Ω 为样本空间的随机试验. 在这个随机试验中, 随机事件之间的基本关系

可以用集合的关系及运算描述. 具体地说, 若 $A \subseteq \Omega$ 和 $B \subseteq \Omega$ 是两个事件, 则我们有

$A \cap B$ 表示 A 和 B 同时发生;

$A \cup B$ 表示 A 和 B 中至少有一个发生;

$A \setminus B$ 表示 A 发生但 B 不发生;

$A \subseteq B$ 表示 A 发生则 B 发生.

为简单起见, 我们经常写 $A \cap B$ 为 AB . 显然 $AB = \emptyset$ 当且仅当 A 和 B 不同时发生, 此时称 A 和 B 互不相容或互斥. 我们称 A 和 $A^c := \Omega \setminus A$ 互为对立事件. 注意, 有些教科书将对立事件 A^c 写为 $\bar{A}$.

随机试验可以在相同的条件下重复多次, 多次重复试验构成联合随机试验. 假设我们进行了 n 次重复伯努利试验, 用 S_n 表示事件 A 在这 n 次试验中发生的次数. 对于 $0 \leq k \leq n$, 显然 “ $S_n = k$ ” 或者 “事件 A 发生 k 次” 是上述联合随机试验中的随机事件. 所以 S_n 的取值具有随机性. 我们称 $F_n(A) := S_n/n$ 为 A 发生的频率. 不难看出, 频率有如下性质:

(1) 非负性: $F_n(A) \geq 0$;

(2) 规范性: $F_n(\Omega) = 1$;

(3) 可加性: 若 $AB = \emptyset$, 则 $F_n(A \cup B) = F_n(A) + F_n(B)$.

根据实践经验, 当 n 很大时, $F_n(A)$ 会稳定在事件 A 发生的概率附近, 这个性质称为频率稳定性.

1.1.2 古典概率模型

当样本空间有限时, 我们可以自然地基于可能性假设建立一个概率模型. 这是最早研究的概率模型, 在概率论的发展中占有重要的地位. 由于这类模型比较简单, 它们非常有助于初学者理解概率论中的一些概念.

定义 1.1.1 若某个随机试验的样本空间 Ω 只包含有限个样本点, 且各样本点出现的概率相等, 则称这个随机试验为古典概率模型, 简称古典概型. 分别记 Ω 和其子集 A 中所包含的样本点的个数为 $n(A)$ 和 $n(\Omega)$. 则事件 A 发生的概率为

$$\mathbf{P}(A) = \frac{n(A)}{n(\Omega)}. \quad (1.1.1)$$

根据上述定义, 对于古典概型计算事件 A 的概率, 关键是求出 Ω 中样本点的总数和 A 中样本点的个数. 为此通常需要利用排列组合的工具, 而且往往有一定的技巧性. 对于整数 $n \geq m \geq 1$, 我们用 A_n^m 表示从 n 个取 m 个的所有排列数, 用 C_n^m 或 $\binom{n}{m}$ 表示从 n 个取 m 个的所有组合数, 即

$$A_n^m = n(n-1) \cdots (n-m+1), \quad C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

例 1.1.8 将一枚质量均匀的硬币重复投掷两次, 求出现正、反面各一次的概率.

解 用 1 表示结果正面朝上, 0 表示反面朝上. 则试验的样本空间为

$$\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\}.$$

它总共有 4 个样本点. 再记事件 { 正、反面各一次 } 为 B . 则 $B = \{(1, 0), (0, 1)\}$, 它有 2 个样本点. 按照古典概型的概率计算公式, 有 $P(B) = 2/4 = 1/2$. $\square$

例 1.1.9 从连续整数 0 至 9 中无放回地取 3 个数依次排列, 利用古典概型计算 3 个数构成 3 位偶数的概率.

解 样本空间 Ω 可取为 0 至 9 这 10 个数中任意 3 个的所有排列, 它包含 A_{10}^3 个样本点. 用 B 表示事件 {3 个数形成 3 位偶数}. 考虑到三位偶数的百位数不能够为 0, 事件 B 发生可分为如下两种情况: (i) 第 3 个数为 0 而第 1 和第 2 个数为 $\{1, 2, \dots, 9\}$ 这 9 个数中任意 2 个数的排列; (ii) 第 3 个数为 $\{2, 4, 6, 8\}$ 中的某个数 c , 第 1 个数为 $\{1, 2, \dots, 9\} \setminus \{c\}$ 这 8 个数中的某个数 a , 而第 2 个数为 $\{0, 1, 2, \dots, 9\} \setminus \{a, c\}$ 这 8 个数中的某个数 b . 根据这样的考虑, 我们有

$$P(B) = \frac{A_9^2 + 4 \times 8 \times 8}{A_{10}^3} = \frac{9 \times 8 + 4 \times 8 \times 8}{10 \times 9 \times 8} = \frac{9 + 4 \times 8}{10 \times 9} = \frac{41}{90}. \quad \square$$

例 1.1.10 (生日问题) 假定一年为 365 天. 利用古典概型计算 r ($1 \leq r \leq 365$) 个人生日互不相同的概率.

解 记 $B = \{r \text{ 个人生日互不相同}\}$. 则

$$P(B) = \frac{C_{365}^r \cdot r!}{365^r}.$$

特别地, 当 $r = 55$ 时, 可算得 $P(B) = 0.01$. 这个概率很小, 直观上有些意外! $\square$

例 1.1.11 袋子中有 r 个红球和 b 个黑球. 现将球依次取出, 求 $R_k = \{\text{第 } k \text{ 次取到红球}\}$ 的概率.

解 (1) 同色球可辨. 将球依次取出相当于把 $r+b$ 个不同的球进行排列. 这样 Ω 中样本点的个数为 $(r+b)!$. 事件 R_k 发生也就是第 k 位放红球, 有 r 种方式, 而其余 $r+b-1$ 个球在 $r+b-1$ 个位置随意排列, 共有 $(r+b-1)!$ 种排法. 故有

$$P(R_k) = \frac{r \cdot (r+b-1)!}{(r+b)!} = \frac{r}{r+b}.$$

(2) 同色球不可辨. 将球依次取出相当于在 $r+b$ 个位置中挑选 r 个位置放红球. 这样 Ω 中样本点个数为 C_{r+b}^r . 第 k 次取到红球相当于第 k 个位置必须是红球, 而在余下的 $r+b-1$ 个位置中挑 $r-1$ 个位置放其余的红球. 故有 C_{r+b-1}^{r-1} 种放置方法. 所以

$$P(R_k) = \frac{C_{r+b-1}^{r-1}}{C_{r+b}^r} = \frac{(r+b-1)! \cdot r! \cdot b!}{(r+b)! \cdot (r-1)! \cdot b!} = \frac{r}{r+b}. \quad \square$$

上例中用两种方法算出的概率 $\mathbf{P}(R_k)$ 相同, 且这个概率与 k 无关. 这说明我们平时采用的抽签方式是公平的.

例 1.1.12 袋中有 r 个红球和 b 个黑球, 从中随机取出 n 个. 求取出的球中恰有 k 个红球的概率 ($k = 0, 1, 2, \dots, n$).

解 令 $E_k = \{\text{摸出的球中恰有 } k \text{ 个红球}\}$. 我们需要对取球的方式进行讨论.

(1) 无放回抽取. 此时

$$\mathbf{P}(E_k) = \frac{C_r^k C_b^{n-k}}{C_{r+b}^n}.$$

(2) 有放回抽取. 此时

$$\mathbf{P}(E_k) = \frac{C_n^k r^k b^{n-k}}{(r+b)^n} = C_n^k \left(\frac{r}{r+b}\right)^k \left(\frac{b}{r+b}\right)^{n-k}.$$

(请读者说明上述两种方式对应的样本空间, 并对计算的合理性作出解释.)

在古典概型中, 用 $\mathcal{F}$ 表示样本空间 Ω 的所有子集构成的集类. 则 (1.1.1) 式在 $\mathcal{F}$ 上定义了一个取值于 $[0, 1]$ 的函数 $\mathbf{P}$, 它在 $A \in \mathcal{F}$ 的取值就是概率值 $\mathbf{P}(A)$. 不难看出, 该集函数具有如下基本性质:

- (1) 非负性: 对于任何 $A \in \mathcal{F}$ 有 $\mathbf{P}(A) \geq 0$;
- (2) 规范性: $\mathbf{P}(\Omega) = 1$;
- (3) 有限可加性: 对任何两两不交的 $A_1, A_2, \dots, A_n \in \mathcal{F}$ 有

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbf{P}(A_i).$$

1.1.3 几何概率模型

有时候, 我们需要考虑样本点的个数无限, 甚至不可数的随机试验, 其各个样本点的出现仍然具有某种“等可能”性. 此时我们无法像古典概型那样, 通过样本点个数的比值来定义随机事件的概率. 几何概型为我们提供了处理此类问题的常用的方法. 令 $\mathbb{R}^d$ 表示 d ($d \geq 1$) 维欧几里得 (Euclid) 空间, 简称欧氏空间.

定义 1.1.2 设随机试验的样本空间 Ω 是欧氏空间 $\mathbb{R}^d$ 的某个具有体积 $m(\Omega)$ 的子集, 且 $0 < m(\Omega) < \infty$. 若集合 $A \subseteq \Omega$ 具有体积 $m(A)$, 则规定 A 的概率为

$$\mathbf{P}(A) = \frac{m(A)}{m(\Omega)}. \quad (1.1.2)$$

称这样的随机试验为几何概率模型, 简称几何概型.

例 1.1.13 向一个半径为 0.5 m 的圆形靶子投飞镖. 根据几何概型计算飞镖落入以靶心为中心、以 0.25 m 为半径的圆内的概率.

解 根据圆面积的计算公式, 事件 $A = \{\text{飞镖落入以靶心为中心、以 0.25 m 为半径的圆内}\}$ 的概率为

$$\mathbf{P}(A) = \frac{\pi \cdot 0.25^2}{\pi \cdot 0.5^2} = \frac{1}{4}. \quad \square$$

在几何概型中, 用 $\mathcal{F}$ 表示样本空间 Ω 的所有可求体积的子集构成的集类. 则 (1.1.2) 式在集类 $\mathcal{F}$ 上定义了一个取值于 $[0, 1]$ 的函数 $\mathbf{P}$, 它在 $A \in \mathcal{F}$ 的取值就是概率值 $\mathbf{P}(A)$. 直观上, 几何概率应该具有如下基本性质:

- (1) 非负性: 对于任何 $A \in \mathcal{F}$ 有 $\mathbf{P}(A) \geq 0$;
- (2) 规范性: $\mathbf{P}(\Omega) = 1$;
- (3) 可数可加性: 对任何两两不交的 $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$ 有

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

根据实变函数论中的结论, 我们不能对欧氏空间 $\mathbb{R}^d$ 的所有子集合理地定义其体积; 参见 [22]. 所以 $\mathcal{F}$ 可能不包含 Ω 的所有子集.

例 1.1.14 设平面上布满间隔为 t 的平行线, 现向平面任意投放一根长为 l ($l < t$) 的针, 根据几何概型求事件 $E = \{\text{针与某直线相交}\}$ 的概率.

解 针的位置是由它的中点到最近直线的距离 ρ 以及针和直线的夹角 θ 决定. 考虑到对称性, 我们可取样本空间为 $\Omega = \left\{(\rho, \theta) : 0 \leq \rho \leq \frac{t}{2}, 0 \leq \theta \leq \frac{\pi}{2}\right\}$. 根据针与直线的几何位置, 有 $E = \left\{(\rho, \theta) : 0 \leq \rho \leq \frac{l}{2} \sin \theta\right\}$, 如图 1.2 所示. 由此可得

$$\mathbf{P}(E) = \frac{m(E)}{m(\Omega)} = \left(\frac{\pi}{2} \cdot \frac{t}{2}\right)^{-1} \int_0^{\frac{\pi}{2}} \frac{l}{2} \sin \theta d\theta = \frac{2l}{\pi t}. \quad \square$$

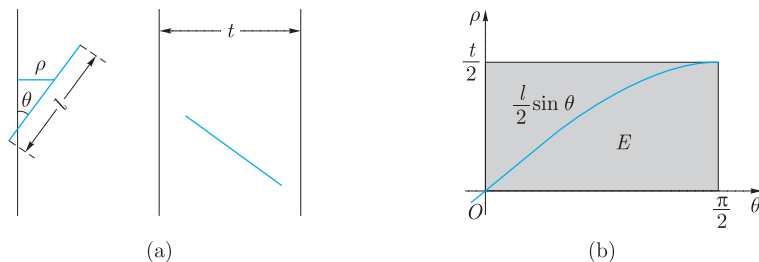


图 1.2 布丰投针试验

上例中的随机试验称为布丰 (Buffon) 投针试验. 重复该试验 n 次, 将事件 E 发生

的次数记为 S_n 次. 根据本课程后面证明的大数定律, 当 $n \rightarrow \infty$ 时, 在一定意义下有

$$\frac{S_n}{n} \rightarrow \mathbf{P}(E) = \frac{2l}{\pi t}.$$

因此, 当 n 很大时, 我们有近似关系

$$\frac{S_n}{n} \approx \frac{2l}{\pi t} \quad \text{或} \quad \pi \approx \frac{2ln}{tS_n}.$$

上式给出了一种利用概率模型进行近似计算的方法, 这类方法称为蒙特卡罗 (Monte Carlo) 方法.

例 1.1.15 在单位圆上任作一弦, 求 $E = \{\text{弦长大于 } \sqrt{3}\}$ 的概率.

解 本例题对于“任作一弦”的含义没有给出准确描述. 下面我们对其作出三种解读, 并分别给出问题的解答.

(1) 考虑由其中点确定的弦. 取单位圆的内接正三角形及该三角形的内切圆, 如图 1.3(a) 所示. 不难算出, 三角形的边长为 $2\sin(\pi/3) = \sqrt{3}$, 而小圆的半径为 $1/2$. 当且仅当弦的中点落在小圆内时, 弦的长度大于三角形的边长 $\sqrt{3}$. 单位圆盘的面积是 π , 而小圆盘的面积是 $\pi/2^2$. 基于这样的解读, 我们可以取 $\Omega = \{\text{单位圆盘}\}$ 和 $E = \{\text{半径为 } 1/2 \text{ 的同心圆盘}\}$. 这样得到

$$\mathbf{P}(E) = \frac{\pi/2^2}{\pi} = \frac{1}{4}.$$

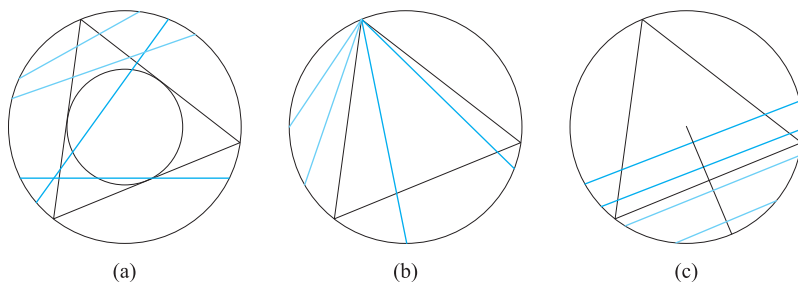


图 1.3 贝特朗奇论

(2) 考虑一个端点固定的弦. 以该端点为一个顶点作单位圆的内接正三角形, 如图 1.3(b) 所示. 注意, 三角形交于该顶点的两条边与圆在此处的切线分别成 60° 角和 120° 角. 由此可知, 当弦与切线成的角度介于 60° 和 120° 之间时其长度大于三角形的边长 $\sqrt{3}$. 这样, 我们可以取 $\Omega = [0, 180]$ 和 $E = [60, 120]$. 于是有

$$\mathbf{P}(E) = \frac{120 - 60}{180} = \frac{60}{180} = \frac{1}{3}.$$

(3) 考虑与某个半径垂直的弦. 取单位圆的内接正三角形及与该三角形的一条边垂直的半径. 则该半径被三角形的边平分, 如图 1.3(c) 所示. 当且仅当弦的中点落在半径

上三角形内时, 弦的长度大于三角形的边长 $\sqrt{3}$. 这样, 我们可以取 $\Omega = \{ \text{半径上的所有点} \}$ 和 $E = \{ \text{半径上到圆心距离小于 } 1/2 \text{ 的点} \}$. 于是有

$$P(E) = \frac{1}{2}.$$

在上例中, 基于“任作一弦”的三种解读方式, 对于同一个问题得出了三种完全不同的答案, 这是概率论发展史上一个著名的例子, 称为贝特朗 (Bertrand) 奇论. 那么, 上述哪一个答案是正确的呢? 这个例子对我们有哪些启发呢? 请读者查阅相关资料, 认真思考一下.

1.1.4 概率公理化体系

古典概型和几何概型对概率论早期的发展起了重要作用. 但是这两种模型都基于对样本点的“等可能”假定, 而且古典概型只适用于有限样本空间, 这使它们的应用有很大的局限性. 由于对于“等可能”假定的不同解读, 几何概型在应用中会产生像贝特朗奇论那样的问题. 事实上, 当随机试验有不可数无穷多个可能的结果时, 有时很难客观地规定“等可能”确切的含义, 这反映出几何概型的逻辑基础是不够严密的.

到 19 世纪下半叶, 概率论的发展及其在物理学等学科中的应用已经突破了概率的古典定义, 但关于概率的一般定义却始终未能明确化和严格化. 这种情况严重阻碍了概率论的进一步发展和应用. 在 1900 年的世界数学家大会上, 希尔伯特提出了建立物理学、概率论等公理化体系的问题.

进入 20 世纪后, 抽象测度和积分理论逐步发展完善. 随着研究的深入, 人们发现事件的运算与集合的运算完全类似, 而概率与测度有相同的性质. 在这种背景下, 柯尔莫哥洛夫 (Andrey N. Kolmogorov) 在 1933 年出版了《概率论基础》一书, 他在书中给出了概率的测度论式的定义和严密的公理化体系. 该公理化体系以概率空间概念为基础, 首先对事件与概率的最基本性质和关系作出规定, 在此基础上导出概率的其他性质. 这些性质从客观实际中抽象出来, 既概括了古典概率和几何概率的基本特性, 又避免了各自的局限性和含混之处. 柯尔莫哥洛夫概率公理化体系是概率论发展史上一个重要的里程碑, 为现代概率论的发展打下了坚实的基础.

本教材后面将采用柯尔莫哥洛夫的公理化体系, 在此基础上系统地介绍概率论的基本知识.

练习题

1. 设 A, B 和 C 是三个事件. 用它们给出下列事件的表示:

- (1) 至少有一个事件发生; (2) 至少有两个事件发生;

- (3) 所有事件都不发生; (4) 至多一个事件发生;
 (5) 至多两个事件发生; (6) 恰有两个事件发生.
2. 从 $0, 1, 2, 3, 4, 5$ 中任取 4 个数排成一行. 求恰好构成四位偶数的概率.
3. 在桥牌游戏中四人各从 52 张纸牌中分得 13 张. 求 4 张 A 集中在一个人手里的概率.
4. 信封里有 8 把钥匙, 其中只有一把能打开房门, 现取出钥匙试开. 分有放回和无放回两种情形, 求 3 次以内能打开房门的概率.
5. 袋子里有 m 个白球, n 个黑球 ($m, n \geq 1$).
- (1) 从袋中随机取两个球, 求两球同色的概率;
- (2) 从袋中随机取一个球, 记录颜色后放回, 从中再取一球, 求两次取的球同色的概率;
- (3) 证明 (2) 的概率始终大于 (1) 的概率.
6. 在区间 $(0, 1)$ 中随机地取出三个数, 以它们作为三条线段的长度. 求这些线段能连接成一个三角形的概率.
7. 在打好方格的白纸上投掷一枚直径为 1 的硬币. 问方格的边长至少应为多少才能使硬币与方格的边不交的概率小于 $1/100$.

1.2 可测空间和概率空间

柯尔莫哥洛夫的概率论公理化体系的基础是概率空间, 它给出了我们前面提到的样本空间、随机事件、概率等概念的严格数学定义.

1.2.1 可测空间

定义 1.2.1 设 Ω 为非空集, 而 $\mathcal{F}$ 为 Ω 的子集构成的集类. 如果

- (i) $\Omega \in \mathcal{F}$;
 (ii) 若 $A \in \mathcal{F}$, 则 $A^c \in \mathcal{F}$;
 (iii) 若 $A, B \in \mathcal{F}$, 则 $A \cup B \in \mathcal{F}$,

那么称 $\mathcal{F}$ 为 Ω 上的集代数.

定义 1.2.2 称非空集 Ω 的子集类 $\mathcal{F}$ 是 σ 代数, 是指定义 1.2.1 中的 (i), (ii) 和下面的性质成立: (iii)' 若 $A_1, A_2, \dots \in \mathcal{F}$, 则 $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$. 此时称 $(\Omega, \mathcal{F})$ 为可测空间, 称 $\mathcal{F}$ 中的集合为可测集.

命题 1.2.1 非空集 Ω 的子集类 $\mathcal{F}$ 是 σ 代数的充要条件是定义 1.2.1 中的 (i),

(ii) 和下面的性质成立: (iii)'' 若 $A_1, A_2, \dots \in \mathcal{F}$, 则 $\bigcap_{k=1}^{\infty} A_k \in \mathcal{F}$.

证明 只需证明, 在性质 (i) 和 (ii) 成立的情况下, 性质 (iii)' 和 (iii)'' 等价. 设性质 (i) 和 (ii) 成立. 由 $\bigcap_{k=1}^{\infty} A_k = \left(\bigcup_{k=1}^{\infty} A_k^c \right)^c$ 知 (iii)' $\Rightarrow$ (iii)''; 而由 $\bigcup_{k=1}^{\infty} A_k = \left(\bigcap_{k=1}^{\infty} A_k^c \right)^c$ 知 (iii)'' $\Rightarrow$ (iii)'. $\square$

例 1.2.1 令 Ω 为非空集. 则 Ω 的所有子集构成的集类 $\mathcal{F}_{\infty}$ 是 Ω 上的 σ 代数, 且它在 Ω 的所有 σ 代数中是最大的. 特别地, 若 Ω 是包含 n 个元素的有限集, 则 $\mathcal{F}_{\infty}$ 中共有 2^n 个子集.

例 1.2.2 令 Ω 为非空集. 易知 $\mathcal{F}_0 := \{\emptyset, \Omega\}$ 是 Ω 上的集代数和 σ 代数, 且它在 Ω 的所有集代数和 σ 代数中是最小的.

例 1.2.3 令 Ω 为非空集, 而 A 为 Ω 的非空真子集. 则 $\mathcal{F} := \{\emptyset, A, A^c, \Omega\}$ 是 Ω 的集代数和 σ 代数, 且它在 Ω 的所有包含 A 的集代数和 σ 代数中是最小的.

例 1.2.4 令 $\Omega = \{1, 2, 3\}$ 和 $A = \{1, 2\}$. 定义 $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$. 则单点集 $\{1\}$ 不属于 $\mathcal{F}$. 所以 Ω 上的集代数和 σ 代数未必包含其所有子集.

例 1.2.5 设 $(\Omega, \mathcal{F})$ 为可测空间. 给定非空集 $E \in \mathcal{F}$ 记 $\mathcal{E} = \{E \cap A : A \in \mathcal{F}\}$. 易证 $\mathcal{E}$ 是 E 上的 σ 代数, 故 $(E, \mathcal{E})$ 是可测空间. 注意 $\mathcal{E} \subseteq \mathcal{F}$.

命题 1.2.2 设 $\{\mathcal{F}_{\alpha} : \alpha \in I\}$ 是 Ω 上的一族 σ 代数, 其中 I 为指标集. 则 $\mathcal{F} := \bigcap_{\alpha \in I} \mathcal{F}_{\alpha}$ 也是 Ω 上的 σ 代数.

证明 因为每个集类 $\mathcal{F}_{\alpha}$ 都满足定义 1.2.1 中的性质 (i), (ii) 和定义 1.2.2 中的性质 (iii)', 易证 $\mathcal{F}$ 也满足这些性质. $\square$

给定非空集 Ω 的子集类 $\mathcal{G}$, 令

$$\sigma(\mathcal{G}) = \cap \{ \mathcal{F} : \mathcal{F} \text{ 为 } \Omega \text{ 上的 } \sigma \text{ 代数且 } \mathcal{F} \supseteq \mathcal{G} \}.$$

由命题 1.2.2 知 $\sigma(\mathcal{G})$ 是 σ 代数, 我们称之为由 $\mathcal{G}$ 生成的 σ 代数. 它是 Ω 上的包含 $\mathcal{G}$ 的最小的 σ 代数. 特别地, 若 $\mathcal{G}$ 是 σ 代数, 则 $\mathcal{G} = \sigma(\mathcal{G})$.

如果 Ω 的可数个子集 $B_1, B_2, \dots$ 两两不交且满足 $\bigcup_{n=1}^{\infty} B_n = \Omega$, 我们就称之为 Ω 的一个可数划分.

例 1.2.6 设 $\mathcal{G} := \{B_1, B_2, \dots\}$ 是 Ω 的可数划分. 则 $\sigma(\mathcal{G})$ 是由 $\mathcal{G}$ 中的集合的所有可能的并构成的集类 (规定零个集合的并为空集).

命题 1.2.3 设 $\mathcal{E}$ 和 $\mathcal{G}$ 是 Ω 上的集类且 $\mathcal{E} \subseteq \mathcal{G}$. 则 $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{G})$.

证明 因为 $\mathcal{E} \subseteq \mathcal{G}$, 所以

$\{\mathcal{F} : \mathcal{F} \text{ 为 } \Omega \text{ 上的 } \sigma \text{ 代数且 } \mathcal{F} \supseteq \mathcal{E}\} \supseteq \{\mathcal{F} : \mathcal{F} \text{ 为 } \Omega \text{ 上的 } \sigma \text{ 代数且 } \mathcal{F} \supseteq \mathcal{G}\},$

故知 $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{G})$. □

定义 1.2.3 令 $\mathcal{O}^d$ 为 d 维欧氏空间 $\mathbb{R}^d$ 的所有开集构成的集类. 我们称 $\mathcal{B}^d := \sigma(\mathcal{O}^d)$ 为 $\mathbb{R}^d$ 的博雷尔 (Borel) σ 代数, 称 $\mathcal{B}^d$ 中的集合为博雷尔可测集, 简称博雷尔集. 称 $(\mathbb{R}^d, \mathcal{B}^d)$ 为欧氏可测空间或博雷尔可测空间.

命题 1.2.4 令 $\mathcal{S}^d$ 是 $\mathbb{R}^d$ 上所有形如 $(a, b] = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_d, b_d]$ 的左开右闭区间构成的集类. 则 $\mathcal{B}^d = \sigma(\mathcal{S}^d)$.

证明 由于 $(a, b) = (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d) \in \mathcal{O}^d \subseteq \sigma(\mathcal{O}^d) = \mathcal{B}^d$, 而 $\mathcal{B}^d$ 对于可数交运算封闭, 故有

$$(a, b] = \bigcap_{k=1}^{\infty} (a_1, b_1 + 1/k) \times (a_2, b_2 + 1/k) \times \cdots \times (a_d, b_d + 1/k) \in \mathcal{B}^d,$$

因而 $\mathcal{S}^d \subseteq \mathcal{B}^d$. 由命题 1.2.3 有 $\sigma(\mathcal{S}^d) \subseteq \sigma(\mathcal{B}^d) = \mathcal{B}^d$. 另一方面, 因为 $\mathcal{O}^d$ 中的每个开集都可写为至多可数个形如 $(a, b]$ 的区间的并, 而 $\sigma(\mathcal{S}^d)$ 作为 σ 代数对于可数并运算封闭, 所以 $\sigma(\mathcal{S}^d) \supseteq \mathcal{O}^d$. 再由命题 1.2.3 有 $\sigma(\mathcal{S}^d) \supseteq \sigma(\mathcal{O}^d) = \mathcal{B}^d$. 综上得

$$\sigma(\mathcal{S}^d) = \mathcal{B}^d. \quad \square$$

事实上, 易证 $\mathcal{B}^d$ 也可以由 $\mathbb{R}^d$ 中的开区间, 或闭区间, 或左闭右开区间生成. 为简化记号, 我们后面将写写 $\mathcal{O} = \mathcal{O}^1$, $\mathcal{C} = \mathcal{C}^1$ 和 $\mathcal{B} = \mathcal{B}^1$.

1.2.2 测度及其基本性质

定义 1.2.4 设 $(\Omega, \mathcal{F})$ 是可测空间, 而 μ 为定义在 σ 代数 $\mathcal{F}$ 上取值于 $[0, \infty]$ 的函数. 我们称 μ 为一个测度, 是指 $\mu(\emptyset) = 0$ 且下面的可数可加性或 σ 可加性成立: 对任何两两不交的集合列 $\{A_1, A_2, \cdots\} \subseteq \mathcal{F}$ 有

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

此时, 若存在两两不交的集合列 $\{\Omega_1, \Omega_2, \cdots\} \subseteq \mathcal{F}$ 使得 $\bigcup_{k=1}^{\infty} \Omega_k = \Omega$ 且对每个 $k \geq 1$ 有 $\mu(\Omega_k) < \infty$, 则称 μ 为 σ 有限测度. 特别地, 若 $\mu(\Omega) < \infty$, 则称 μ 为有限测度; 而若 $\mu(\Omega) = 1$, 则称 μ 为概率测度或概率.

在可数可加性中取 $A_{n+1} = A_{n+2} = \cdots = \emptyset$, 我们得到有限可加性: 对任何两两不交的 $A_1, A_2, \cdots, A_n \in \mathcal{F}$ 有

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k). \quad (1.2.1)$$

进而容易证明测度 μ 的如下性质:

- 单调性: 若 $A, B \in \mathcal{F}$ 且有 $A \subseteq B$, 则 $\mu(A) \leq \mu(B)$;
- 可减性: 若 $A, B \in \mathcal{F}$ 且有 $A \subseteq B$ 和 $\mu(A) < \infty$, 则 $\mu(B \setminus A) = \mu(B) - \mu(A)$.

例 1.2.7 对于 $\omega \in \Omega$ 和 $A \in \mathcal{F}$, 当 $\omega \in A$ 时令 $\delta_\omega(A) = 1$, 当 $\omega \in A^c$ 时令 $\delta_\omega(A) = 0$. 显然 δ_ω 是 $(\Omega, \mathcal{F})$ 上的概率测度. 称 δ_ω 为集中于 $\omega \in \Omega$ 的单点概率测度或狄拉克 (Dirac) 测度.

例 1.2.8 设 μ 是可测空间 $(\Omega, \mathcal{F})$ 上的 σ 有限测度, 而 $E \in \mathcal{F}$ 为非空集. 对于 $A \in \mathcal{F}$ 令 $\mu_E(A) = \mu(E \cap A)$. 则 μ_E 也是 $(\Omega, \mathcal{F})$ 上的 σ 有限测度. 称 μ_E 为 μ 在 E 上的限制. 我们也经常将 μ_E 视为 $(E, \mathcal{E})$ 上的测度, 其中 $\mathcal{E}$ 如例 1.2.5 给出.

例 1.2.9 设 $(\Omega, \mathcal{F})$ 和 $(E, \mathcal{E})$ 如例 1.2.5 给出. 再设 ν 是 $(E, \mathcal{E})$ 上的 σ 有限测度. 对于 $A \in \mathcal{F}$ 令 $\mu(A) = \nu(E \cap A)$. 则 μ 是 $(\Omega, \mathcal{F})$ 上的 σ 有限测度.

为简单起见, 下面假定 μ 是可测空间 $(\Omega, \mathcal{F})$ 上的有限测度.

定理 1.2.5 (次可加性) 对任何 $A_1, A_2, \dots \in \mathcal{F}$ 有

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

证明 定义 $B_1 = A_1$ 和 $B_n = A_1^c A_2^c \cdots A_{n-1}^c A_n$ ($n \geq 2$). 易证 $B_1, B_2, \dots \in \mathcal{F}$ 两两不交且 $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. 由可数可加性和单调性得

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n).$$

因 $\mu(B_n) \leq \mu(A_n)$, 故得欲证不等式. □

定理 1.2.6 (两集合的加法定理) 对于任何 $A, B \in \mathcal{F}$ 有

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B). \quad (1.2.2)$$

证明 注意 $A \cup B = A \cup (B \setminus A)$, 其中 $A, B \setminus A \in \mathcal{F}$ 不交. 由测度的有限可加性,

$$\mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A).$$

再由可减性得

$$\mu(B \setminus A) = \mu(B \setminus (A \cap B)) = \mu(B) - \mu(A \cap B),$$

从而 (1.2.2) 式成立. □

定理 1.2.7 (推广的加法定理) 对于任何 $A_1, A_2, \dots, A_n \in \mathcal{F}$ ($n \geq 2$), 有

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{1 \leq i \leq n} \mu(A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i A_j) +$$

$$\sum_{1 \leq i < j < k \leq n} \mu(A_i A_j A_k) + \cdots + (-1)^{n-1} \mu(A_1 A_2 \cdots A_n). \quad (1.2.3)$$

证明 利用数学归纳法证明. 由定理 1.2.6 知 (1.2.3) 式对 2 个集合的情况成立. 现设该等式对 n 个集合的情况成立. 由 (1.2.2) 式得

$$\begin{aligned} \mu\left(\bigcup_{k=1}^{n+1} A_k\right) &= \mu\left(\bigcup_{k=1}^n A_k\right) + \mu(A_{n+1}) - \mu\left(A_{n+1} \cap \bigcup_{k=1}^n A_k\right) \\ &= \mu\left(\bigcup_{k=1}^n A_k\right) + \mu(A_{n+1}) - \mu\left(\bigcup_{k=1}^n A_{n+1} A_k\right). \end{aligned} \quad (1.2.4)$$

利用归纳假设, 上式右端第 1 项可以表示为 (1.2.3) 式, 而第 3 项可以表示为

$$\begin{aligned} \mu\left(\bigcup_{k=1}^n A_{n+1} A_k\right) &= \sum_{1 \leq i \leq n} \mu(A_{n+1} A_i) - \sum_{1 \leq i < j \leq n} \mu(A_i A_j A_{n+1}) + \\ &\quad \sum_{1 \leq i < j < k \leq n} \mu(A_i A_j A_k A_{n+1}) + \cdots + (-1)^{n-1} \mu(A_1 A_2 \cdots A_n A_{n+1}). \end{aligned}$$

将上述表示代入 (1.2.4) 式知 (1.2.3) 式对 $n+1$ 个集合的情况成立. 由此即得欲证结果. $\square$

定理 1.2.8 (下连续性) 若 $\{A_n\} \subseteq \mathcal{F}$ 为不降集合序列, 则

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

证明 令 $A_0 = \emptyset$ 和 $B_n = A_n \setminus A_{n-1}$, $n \geq 1$. 则 $B_1, B_2, \cdots \in \mathcal{F}$ 两两不交且 $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. 利用可数可加性得

$$\begin{aligned} \mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \\ &= \sum_{n=1}^{\infty} [\mu(A_n) - \mu(A_{n-1})] \\ &= \lim_{k \rightarrow \infty} \sum_{n=1}^k [\mu(A_n) - \mu(A_{n-1})]. \end{aligned}$$

因 $\mu(A_0) = \mu(\emptyset) = 0$, 故得欲证等式. $\square$

定理 1.2.9 (上连续性) 若 $\{A_n\} \subseteq \mathcal{F}$ 为不增集合序列, 则

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

证明 注意 $A_1^c, A_2^c, \dots \in \mathcal{F}$ 且 $A_1^c \subseteq A_2^c \subseteq \dots$. 根据 μ 的下连续性及可减性有

$$\mu\left(\bigcup_{n=1}^{\infty} A_n^c\right) = \lim_{n \rightarrow \infty} \mu(A_n^c) = \mu(\Omega) - \lim_{n \rightarrow \infty} \mu(A_n).$$

再次利用可减性得

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu\left(\Omega \setminus \left(\bigcap_{n=1}^{\infty} A_n\right)^c\right) = \mu(\Omega) - \mu\left(\bigcup_{n=1}^{\infty} A_n^c\right).$$

综合以上二式即得欲证等式. $\square$

定义 1.2.5 称博雷尔可测空间 $(\mathbb{R}^d, \mathcal{B}^d)$ 上的测度 μ 为博雷尔测度. 如果对于任何有界区间 $[a, b] \subseteq \mathbb{R}^d$ 都有 $\mu[a, b] = \mu([a, b]) < \infty$, 那么称该测度是局部有限的.

显然, 局部有限测度是 σ 有限的.

例 1.2.10 根据实变函数中的结论, 存在博雷尔 σ 代数 $\mathcal{B}^d$ 上的唯一的测度 λ , 使得对任何区间 $(a, b] = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_d, b_d] \in \mathcal{S}^d$ 有

$$\lambda(a, b] = (b_1 - a_1)(b_2 - a_2) \cdots (b_d - a_d), \quad a_i \leq b_i \in \mathbb{R}, \quad i = 1, 2, \dots, d.$$

我们称 λ 为 $\mathbb{R}^d$ 上的勒贝格 (Lebesgue) 测度. 勒贝格测度是反映欧几里得空间中子集的长度、面积或者体积的严格数学概念. 显然, 勒贝格测度是局部有限的. 对于勒贝格测度存在性的证明感兴趣的读者, 可以参考实分析方面的著作, 例如 [22] 等.

1.2.3 概率空间的概念

定义 1.2.6 给定可测空间 $(\Omega, \mathcal{F})$ 和 $\mathcal{F}$ 上的概率测度 $\mathbf{P}$, 我们称 $(\Omega, \mathcal{F}, \mathbf{P})$ 为概率空间, 其中 Ω 称为样本空间, 而 $\mathcal{F}$ 称为事件域. 事件域 $\mathcal{F}$ 中的集合称为随机事件, 简称事件. 对于事件 $A \in \mathcal{F}$, 称 $\mathbf{P}(A)$ 为 A 的概率.

概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 是描述随机试验的数学模型. 这里样本空间 Ω 表示所有可能的结果的集合. 令 $A \in \mathcal{F}$ 为一个事件. 假设试验的结果是某个 $\omega \in \Omega$. 若 $\omega \in A$, 则表示 A 在试验中发生了, 否则表示 A 在试验中没有发生. 概率测度 $\mathbf{P}$ 给出了事件域 $\mathcal{F}$ 中所有事件的概率. 注意有 $\mathbf{P}(\Omega) = 1$ 和 $\mathbf{P}(\emptyset) = 0$. 我们分别称 Ω 和 $\emptyset$ 为必然事件和不可能事件.

给定概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$, 我们称 $A \subseteq \Omega$ 为 $\mathbf{P}$ 零集, 是指存在 $N \in \mathcal{F}$ 满足 $N \supseteq A$ 且 $\mathbf{P}(N) = 0$. 称 $(\Omega, \mathcal{F}, \mathbf{P})$ 是完备概率空间, 是指 $\mathcal{F}$ 包含一切 $\mathbf{P}$ 零集.

例 1.2.11 考虑关于事件 $A \subseteq \Omega$ 的伯努利试验. 令 $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$. 设成功概率为 $\mathbf{P}(A) = p$, 其中 $0 \leq p \leq 1$, 故有

$$\mathbf{P}(\emptyset) = 0, \quad \mathbf{P}(\Omega) = 1, \quad \mathbf{P}(A) = p, \quad \mathbf{P}(A^c) = 1 - p.$$

我们称 $(\Omega, \mathcal{F}, \mathbf{P})$ 为伯努利概率空间.

例 1.2.12 (古典概率空间) 设 Ω 为有限集. 令 $\mathcal{F}$ 为 Ω 的所有子集构成的 σ 代数. 用 $n(A)$ 表示集合 $A \in \mathcal{F}$ 中元素的个数, 并令

$$\mathbf{P}(A) = \frac{n(A)}{n(\Omega)}.$$

则 $(\Omega, \mathcal{F}, \mathbf{P})$ 构成一个概率空间, 称为古典概率空间.

例 1.2.13 (可数概率空间) 设 $\Omega = \{\omega_i : i \in I\}$ 为非空可数集, 其中 I 为其指标集. 令 $\mathcal{F}$ 为 Ω 的所有子集构成的 σ 代数. 取非负实数 $p_i, i \in I$ 满足 $\sum_{i \in I} p_i = 1$. 对于 $A = \{\omega_i : i \in J\} \in \mathcal{F}$, 其中 $J \subseteq I$, 令

$$\mathbf{P}(A) = \sum_{i \in J} p_i.$$

则 $(\Omega, \mathcal{F}, \mathbf{P})$ 为一个概率空间, 称为可数概率空间. 特别地, 若 Ω 为有限集, 则称 $(\Omega, \mathcal{F}, \mathbf{P})$ 为有限概率空间. 古典概率空间是典型的有限概率空间.

例 1.2.14 设 $\Omega \in \mathcal{B}^d$ 满足 $0 < \lambda(\Omega) < \infty$. 令 $\mathcal{F} = \Omega \cap \mathcal{B}^d$. 对 $A \in \mathcal{F}$ 定义

$$\mathbf{P}(A) = \frac{\lambda(A)}{\lambda(\Omega)}.$$

则 $(\Omega, \mathcal{F}, \mathbf{P})$ 构成一个概率空间, 称为几何概率空间.

将事件域定义为一个 σ 代数为事件和概率的运算提供了很大的便利. 给定一系列事件序列 $\{A_n\} \subseteq \mathcal{F}$, 分别称

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \quad \liminf_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \quad (1.2.5)$$

为 $\{A_n\}$ 的上极限和下极限. 容易证明如下性质:

- $\omega \in \limsup_{n \rightarrow \infty} A_n$ 当且仅当有无穷多个 $k \geq 1$ 使得 $\omega \in A_k$;
- $\omega \in \liminf_{n \rightarrow \infty} A_n$ 当且仅当有某个 $n \geq 1$ 使当 $k \geq n$ 时有 $\omega \in A_k$.

上述性质可以解释为: 事件 $\limsup_{n \rightarrow \infty} A_n$ 发生表示事件序列 $A_1, A_2, \dots$ 中有无穷多个事件发生, 而事件 $\liminf_{n \rightarrow \infty} A_n$ 发生表示有某个 $n \geq 1$ 使得事件 $A_n, A_{n+1}, A_{n+2}, \dots$ 都发生. 由此易知

$$\limsup_{n \rightarrow \infty} A_n \supseteq \liminf_{n \rightarrow \infty} A_n.$$

为简单起见, 若事件序列 $\{A_n\}$ 单调不降, 我们写 $A_n \uparrow \bigcup_n A_n$; 而若 $\{A_n\}$ 单调不增, 写

$$A_n \downarrow \bigcap_n A_n.$$

例 1.2.15 取 $\Omega = \{ \text{全体非负整数} \}$. 对 $n \geq 1$ 令 $A_{2n} = \{ \text{全体非负偶数} \}$ 和 $A_{2n-1} = \{ \text{全体非负奇数} \}$, 则有

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \Omega, \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n = \emptyset.$$

1.2.4 概率的计算

在概率空间的定义中, 我们将概率视为一种具有若干特殊性质的集函数. 这体现了数学学科的整体性, 使概率论的发展可以更有效地利用其他数学分支的方法和工具. 下面给出了几个利用概率的一般性质进行概率计算的例子.

例 1.2.16 假设 n 对夫妇任意在一排 $2n$ 个椅子上就座. 求有夫妇不相邻的概率.

解 记 $A = \{ \text{有夫妇不相邻} \}$. 显然 $A^c = \{ \text{所有夫妇都相邻} \}$ 的概率比较容易计算:

$$\mathbf{P}(A^c) = \frac{n! \cdot 2^n}{(2n)!} = \frac{1}{(2n-1)!!}.$$

从而 $\mathbf{P}(A) = 1 - \frac{1}{(2n-1)!!}$. □

例 1.2.17 (布丰投针试验的续) 向画满间隔为 t 的平行直线的桌面上任投一直径为 l ($l < t$) 的半圆形纸片 (图 1.4). 求纸片与某直线相交的概率.

解 (1) 记 $E = \{ \text{半圆形纸片与某直线相交} \}$. 设想把半圆形纸片拼成一个完整的圆形纸片. 再令 $F = \{ \text{新的半圆形纸片与某直线相交} \}$. 则有 $E \cup F = \{ \text{直径为 } l \text{ 的圆形与某直线相交} \}$. 因 $l < t$, 我们有

$$\begin{aligned} E \cap F &= \{ \text{两半圆分界线与某直线相交} \} \\ &= \{ \text{长为 } l \text{ 的线段与某直线相交} \}. \end{aligned}$$

由关于布丰投针问题的结果得 $\mathbf{P}(E \cap F) = 2l/\pi t$.

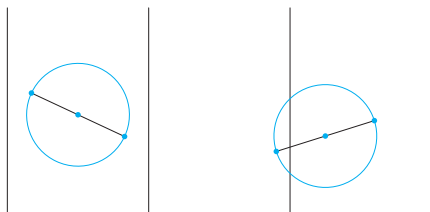


图 1.4 布丰投针试验

(2) 因 $E \cup F = \{ \text{直径为 } l \text{ 的圆形纸片与平行线相交} \}$, 只要考虑圆心与最近平行线的距离和半径的关系. 取样本空间 $\Omega = [0, t/2]$. 则 $E \cup F = \{ r \in \Omega : r \leq l/2 \}$, 故 $\mathbf{P}(E \cup F) = l/t$.

(3) 根据加法公式有

$$\mathbf{P}(E) + \mathbf{P}(F) = \mathbf{P}(E \cup F) + \mathbf{P}(E \cap F) = \frac{\pi l + 2l}{\pi t}.$$

因 $\mathbf{P}(E) = \mathbf{P}(F)$ 得

$$\mathbf{P}(E) = \frac{\pi l + 2l}{2\pi t}. \quad \square$$

例 1.2.18 参加集会的 n 个人将帽子放在一起, 会后每个人任取一顶帽子戴上. 求 $B_n = \{n \text{ 个人中至少一人戴对自己的帽子}\}$ 和 $B_n^c = \{n \text{ 个人中无人戴对自己的帽子}\}$ 的概率.

解 将参加集会的人和帽子对应地用数 $1, 2, \dots, n$ 编号. 样本空间 Ω 取为这 n 个数的排列 $(r_1, r_2, \dots, r_n)$ 所有可能的形式, 它共有 $n!$ 个样本点. 记 $A_i = \{\text{第 } i \text{ 个人戴对自己帽子}\} = \{(r_1, r_2, \dots, r_n) : r_i = i\}$. 则不难算出

$$\begin{aligned} \mathbf{P}(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n} \quad (1 \leq i \leq n), \\ \mathbf{P}(A_i A_j) &= \frac{1}{n(n-1)} \quad (1 \leq i < j \leq n), \quad \dots \end{aligned}$$

注意 $B_n = \bigcup_{i=1}^n A_i$. 根据加法公式,

$$\begin{aligned} \mathbf{P}(B_n) &= \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(A_i A_j) + \\ &\quad \sum_{1 \leq i < j < k \leq n} \mathbf{P}(A_i A_j A_k) + \dots + (-1)^{n-1} \mathbf{P}(A_1 A_2 \dots A_n) \\ &= \sum_{i=1}^n \frac{1}{n} - \sum_{1 \leq i < j \leq n} \frac{1}{n(n-1)} + \dots + (-1)^{n-1} \frac{1}{n!} \\ &= \frac{1}{1!} - \frac{1}{2!} + \dots + (-1)^{n-1} \frac{1}{n!}, \end{aligned}$$

因而

$$\mathbf{P}(B_n^c) = 1 - \mathbf{P}(B_n) = \sum_{i=0}^n (-1)^i \frac{1}{i!}. \quad \square$$

例 1.2.19 在例 1.2.18 的情况下, 求 $E_{n,k} = \{n \text{ 个人中恰有 } k \text{ 个戴对自己的帽子}\}$ 的概率.

解 注意 $E_{n,k} = \{(r_1, r_2, \dots, r_n) : \text{对 } k \text{ 个标号 } \alpha \text{ 有 } r_\alpha = \alpha \text{ 而对其余 } n-k \text{ 个标号 } \alpha \text{ 有 } r_\alpha \neq \alpha\}$. 令 $m(A)$ 为事件 $A \subseteq \Omega$ 所含的样本点的总数. 不难发现 $m(E_{n,k}) = C_n^k m(E_{n-k,0})$. 根据例 1.2.18 的结论有

$$m(E_{n-k,0}) = (n-k)! \mathbf{P}(E_{n-k,0}) = (n-k)! \sum_{i=0}^{n-k} (-1)^i \frac{1}{i!},$$

因而

$$m(E_{n,k}) = \frac{n!}{k!} \sum_{i=0}^{n-k} (-1)^i \frac{1}{i!}.$$

所以有

$$\mathbf{P}(E_{n,k}) = \frac{1}{k!} \sum_{i=0}^{n-k} (-1)^i \frac{1}{i!}.$$

练习题

1. 假设某个随机试验包含可数无穷多个不同的结果, 且每个结果出现的概率均严格大于零. 解释为何不会所有结果发生的概率都相同.

2. 设 $(\Omega, \mathcal{F})$ 为可测空间. 给定非空集 $E \subseteq \Omega$, 记 $\mathcal{E} = \{E \cap A : A \in \mathcal{F}\}$.

(1) 证明 $\mathcal{E}$ 是 E 上的 σ 代数;

(2) 是否有 $\mathcal{E} \subseteq \mathcal{F}$?

3. 设 $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2), \dots, (\Omega_n, \mathcal{F}_n, \mu_n)$ 是 n 个有限测度空间, 其中 $\Omega_1, \Omega_2, \dots, \Omega_n$ 两两不相交. 令 $\Omega = \bigcup_{k=1}^n \Omega_k$.

(1) 证明 $\mathcal{F} := \{A \subseteq \Omega : A \cap \Omega_k \in \mathcal{F}_k, 1 \leq k \leq n\}$ 是 Ω 上的 σ 代数;

(2) 对 $A \in \mathcal{F}$ 令 $\mu(A) = \sum_{k=1}^n \mu_k(A \cap \Omega_k)$. 证明 μ 是 $(\Omega, \mathcal{F})$ 上的有限测度.

4. 设 $(\Omega, \mathcal{F}, \mathbf{P})$ 为概率空间. 证明对 $A, B \in \mathcal{F}$ 有 $\mathbf{P}(A \cap B) \geq \mathbf{P}(A) + \mathbf{P}(B) - 1$.

5. 设事件 A, B 满足 $\mathbf{P}(A) = 3/4$ 和 $\mathbf{P}(B) = 1/3$. 证明 $1/12 \leq \mathbf{P}(A \cap B) \leq 1/3$.

6. 对于随机事件序列 $A_1, A_2, \dots \in \mathcal{F}$, 证明: 若对一切 $k \geq 1$ 有 $\mathbf{P}(A_k) = 1$, 则 $\mathbf{P}\left(\bigcap_{k=1}^{\infty} A_k\right) = 1$.

1.3 条件概率

1.3.1 条件概率与乘法公式

定义 1.3.1 假设 $(\Omega, \mathcal{F}, \mathbf{P})$ 为概率空间. 再设 $B \in \mathcal{F}$ 满足 $\mathbf{P}(B) > 0$. 对事件 $A \in \mathcal{F}$ 定义

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}. \quad (1.3.1)$$

称 $\mathbf{P}(A|B)$ 为已知 B 发生的情况下 A 的条件概率.

对 $A \in \mathcal{F}$ 记 $\mathbf{P}_B(A) = \mathbf{P}(A|B)$. 利用 (1.3.1) 式易证 $\mathbf{P}_B$ 也是可测空间 $(\Omega, \mathcal{F})$ 上的概率测度. 我们称 $\mathbf{P}_B$ 为给定事件 B 的条件概率测度, 而称 $(\Omega, \mathcal{F}, \mathbf{P}_B)$ 为给定事件 B 的条件概率空间.

另外, 由 (1.3.1) 式易得下面的乘法公式:

$$\mathbf{P}(AB) = \mathbf{P}(B)\mathbf{P}(A|B), \quad A, B \in \mathcal{F}. \quad (1.3.2)$$

为了运算方便, 我们规定当 $\mathbf{P}(B) = 0$ 时上式右端理解为零. 这样, 对于乘法公式 (1.3.2) 可以不须假定 $\mathbf{P}(B) > 0$.

利用数学归纳法可以证明:

定理 1.3.1 (乘法定理) 对任何 $A_1, A_2, \dots, A_n \in \mathcal{F}$ 有

$$\mathbf{P}(A_1 A_2 \cdots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1) \cdots \mathbf{P}(A_n|A_1 \cdots A_{n-1}).$$

基于乘法公式, 有时可以通过条件化和递推方法计算某些复杂的概率.

例 1.3.1 假设 n 对夫妇在 $2n$ 个椅子上随机地就座. 求 $A_n = \{ \text{丈夫全坐在妻子右方} \}$ 的概率.

解 将夫妇用数字 $1, 2, \dots, 2n$ 编号. 记 $B_1 = \{1 \text{ 号丈夫在妻子右方} \}$, 则

$$\mathbf{P}(B_1) = \frac{C_{2n}^2 (2n-2)!}{(2n)!} = \frac{1}{2}.$$

利用乘法公式,

$$\mathbf{P}(A_n) = \mathbf{P}(A_n|B_1)\mathbf{P}(B_1) = \frac{1}{2}\mathbf{P}(A_{n-1}).$$

利用上面递推关系立得 $\mathbf{P}(A_n) = 1/2^n$. □

1.3.2 全概率公式

根据之前引入的概念, 如果事件序列 $\{B_1, B_2, \dots\} \subseteq \mathcal{F}$ 两两不交且满足 $\bigcup_{n=1}^{\infty} B_n = \Omega$, 则 $\{B_1, B_2, \dots\}$ 构成 Ω 的一个可数划分.

定理 1.3.2 (全概率公式) 假设 $\{B_n\} \subseteq \mathcal{F}$ 为 Ω 的可数划分. 则对任何 $A \in \mathcal{F}$ 有

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(B_n)\mathbf{P}(A|B_n).$$

证明 由 $\{B_n\}$ 两两不交知 $\{B_n A\}$ 两两不交. 显然 $A = \bigcup_{n=1}^{\infty} B_n A$. 由可数可加性及乘法公式有

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(B_n A) = \sum_{n=1}^{\infty} \mathbf{P}(B_n) \mathbf{P}(A|B_n).$$

故得欲证等式. □

全概率公式是很常用的公式. 特别地, 当事件 A 的发生涉及某种因素时, 我们往往会根据该因素的不同情况, 分别计算 A 的条件概率, 然后利用全概率公式求得 $\mathbf{P}(A)$. 与乘法公式的情况类似, 当 $\mathbf{P}(B_n) = 0$ 时, 我们规定 $\mathbf{P}(B_n) \mathbf{P}(A|B_n) = 0$, 因此这里不须假定 $\mathbf{P}(B_n) > 0$.

例 1.3.2 设袋中有 r 个红球和 b 个黑球. 每次从袋中任取 1 球, 查看颜色后连同 s 个同色球一起放回. 以 R_n 表示第 n 次取出红球. 证明 $\mathbf{P}(R_n) = r/(r+b)$.

证明 首先, 容易看出 $\mathbf{P}(R_1) = r/(r+b)$, 即欲证结果对于 1 次摸球试验成立. 假定该结果对于 $n-1$ 次摸球试验成立. 则

$$\mathbf{P}(R_n|R_1) = \frac{r+s}{r+s+b}, \quad \mathbf{P}(R_n|R_1^c) = \frac{r}{r+s+b}.$$

利用全概率公式, 有

$$\begin{aligned} \mathbf{P}(R_n) &= \mathbf{P}(R_1) \mathbf{P}(R_n|R_1) + \mathbf{P}(R_1^c) \mathbf{P}(R_n|R_1^c) \\ &= \frac{r}{r+b} \cdot \frac{r+s}{r+s+b} + \frac{b}{r+b} \cdot \frac{r}{r+s+b} \\ &= \frac{r}{r+b}. \end{aligned}$$

根据数学归纳法即得结论. □

例 1.3.3 甲乙二人轮流掷骰子, 甲先掷, 以后谁掷出 1 点时则交给对方掷, 否则此人继续掷. 求 $A_n = \{\text{第 } n \text{ 次由甲掷}\}$ 的概率.

解 记 $p_n = \mathbf{P}(A_n)$. 由题意知 $p_1 = 1$. 利用全概率公式,

$$\begin{aligned} p_n &= \mathbf{P}(A_{n-1}) \mathbf{P}(A_n|A_{n-1}) + \mathbf{P}(A_{n-1}^c) \mathbf{P}(A_n|A_{n-1}^c) \\ &= \frac{5}{6} p_{n-1} + \frac{1}{6} (1 - p_{n-1}) = \frac{2}{3} p_{n-1} + \frac{1}{6}. \end{aligned}$$

根据上面递推关系,

$$p_n = \frac{2}{3} \left(\frac{2}{3} p_{n-2} + \frac{1}{6} \right) + \frac{1}{6} = \left(\frac{2}{3} \right)^2 p_{n-2} + \frac{1}{6} \times \frac{2}{3} + \frac{1}{6} = \cdots$$

$$\begin{aligned}
&= \left(\frac{2}{3}\right)^{n-1} p_1 + \frac{1}{6} \sum_{i=1}^{n-1} \left(\frac{2}{3}\right)^{i-1} = \left(\frac{2}{3}\right)^{n-1} + \frac{1}{6} \cdot \frac{1 - \left(\frac{2}{3}\right)^{n-1}}{1 - \frac{2}{3}} \\
&= \left(\frac{2}{3}\right)^{n-1} + \frac{1}{2} \left[1 - \left(\frac{2}{3}\right)^{n-1}\right] = \frac{1}{2} \left[1 + \left(\frac{2}{3}\right)^{n-1}\right]. \quad \square
\end{aligned}$$

1.3.3 贝叶斯公式

首先, 我们看两个例子.

例 1.3.4 某工厂的第 1, 2, 3 车间生产同一产品, 产量比分别为 0.5, 0.25, 0.25, 而次品率分别为 0.01, 0.01, 0.02. 现从这个厂的产品中任取一件, 求取到次品的概率.

解 令 $A = \{\text{取到一件次品}\}$ 和 $B_i = \{\text{取到的产品来自第 } i \text{ 车间}\}$. 由题意有

$$\mathbf{P}(B_1) = 0.5, \quad \mathbf{P}(B_2) = 0.25, \quad \mathbf{P}(B_3) = 0.25,$$

$$\mathbf{P}(A|B_1) = 0.01, \quad \mathbf{P}(A|B_2) = 0.01, \quad \mathbf{P}(A|B_3) = 0.02.$$

再根据全概率公式,

$$\begin{aligned}
\mathbf{P}(A) &= \mathbf{P}(B_1)\mathbf{P}(A|B_1) + \mathbf{P}(B_2)\mathbf{P}(A|B_2) + \mathbf{P}(B_3)\mathbf{P}(A|B_3) \\
&= 0.5 \times 0.01 + 0.25 \times 0.01 + 0.25 \times 0.02 = 0.0125. \quad \square
\end{aligned}$$

例 1.3.5 在例 1.3.4 的解答中, 假定 $A = \{\text{取到一件次品}\}$, 则在全概率公式的基础上利用条件概率定义得到

$$\mathbf{P}(B_1|A) = \frac{\mathbf{P}(AB_1)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B_1)\mathbf{P}(A|B_1)}{\mathbf{P}(A)} = \frac{0.5 \times 0.01}{0.0125} = 0.4.$$

由前面例子中的计算过程可以总结出一个重要的概率计算公式. 这就是下面定理给出的贝叶斯 (Bayes) 公式.

定理 1.3.3 (贝叶斯公式) 设 $\{B_1, B_2, \dots\} \subseteq \mathcal{F}$ 为 Ω 的一个划分, 而 $A \in \mathcal{F}$ 满足 $\mathbf{P}(A) > 0$. 则对任何 $n \geq 1$ 有

$$\mathbf{P}(B_n|A) = \frac{\mathbf{P}(B_n)\mathbf{P}(A|B_n)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B_n)\mathbf{P}(A|B_n)}{\sum_{k=1}^{\infty} \mathbf{P}(B_k)\mathbf{P}(A|B_k)}.$$

例 1.3.6 医学中的诊断错误包括完全漏诊和完全误诊: 把有病诊断为无病称为完全漏诊; 把无病诊断为有病称为完全误诊. 假设通过验血诊断某种疾病的完全误诊率和

完全漏诊率分别仅为 1% 和 3%. 也就是说, 若记

$$A = \{\text{验血结果为阳性}\}, \quad B = \{\text{受检者患此病}\},$$

则有 $P(A|B^c) = 0.01$ 和 $P(A^c|B) = 0.03$. 再设 $P(B) = 0.005$. 分别求 $P(B|A)$ 和 $P(B|A^c)$.

解 首先, 由全概率公式,

$$\begin{aligned} P(A) &= P(B)P(A|B) + P(B^c)P(A|B^c) \\ &= 0.005 \times (1 - 0.03) + (1 - 0.005) \times 0.01 \\ &= 0.004\,85 + 0.009\,95 = 0.014\,8, \end{aligned}$$

故有 $P(A^c) = 0.985\,2$. 再由贝叶斯公式有

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{0.005 \times (1 - 0.03)}{0.0148} = \frac{0.004\,85}{0.014\,8} \approx 0.328$$

和

$$P(B|A^c) = \frac{P(B)P(A^c|B)}{P(A^c)} = \frac{0.005 \times 0.03}{0.985\,2} = \frac{0.000\,15}{0.985\,2} \approx 0.000\,152. \quad \square$$

在上例中, 由于总患病率 $P(B) = 0.005$ 很低, 化验结果为阳性时, 受检者实际患病的条件概率 $P(B|A) = 0.328$ 并不很大. 同样的原因导致, 化验结果为阴性时, 受检者实际患病的条件概率 $P(B|A^c) = 0.000\,152$ 非常小.

*1.3.4 蒙提霍尔问题

有个请嘉宾选择奖品的游戏, 主办方在编号为 1, 2, 3 的三个门中等可能地随机选择一个, 放入一辆跑车, 另两个门里各放一只山羊. 节目的嘉宾不知道哪个门里是跑车. 主持人请嘉宾选择一个门, 门后的奖品由他获得. 嘉宾当然希望选中跑车!

问题 假如你是嘉宾, 选择了 1 号门, 在此门打开之前, 主持人首先打开了余下的门中的一个有山羊的门, 并给你重新选择的机会. 不妨假设他打开的是 3 号门, 你是坚持选 1 号门? 还是改选 2 号门?

这个问题来源于一个著名的电视节目, 曾经引起广泛关注和讨论, 被称为蒙提霍尔 (Monty Hall) 问题. 下面用两种方法给出问题的解答.

解 解法 1 因为放跑车的门是等可能随机抽取的, 所以 1 号门里有跑车的概率为 $1/3$, 有山羊的概率为 $2/3$. 如果决策是不改选, 那么你在 1 号门里有跑车的情况下获得跑车, 你成功的概率为 $1/3$; 如果决策是换 2 号门, 那么你在 1 号门里有山羊的情况下获得跑车, 你成功的概率为 $2/3$. 所以, 你应该改变选择.

解法 2 按照假设你已经选择了 1 号门. 用 A_1, A_2, A_3 分别表示 1, 2, 3 号门里有跑车, 用 B_1, B_2, B_3 分别表示主持人打开 1, 2, 3 号门. 因为你在选择 1 号门时不知道车在哪个门里, 你的选择不影响跑车在三个门里的概率, 所以

$$\mathbf{P}(A_1) = \mathbf{P}(A_2) = \mathbf{P}(A_3) = \frac{1}{3}. \quad (1.3.3)$$

主持人打开 1 号门之外一个放山羊的门, 有以下几种可能的情况:

(1) 跑车在 1 号门里, 主持人有 2, 3 号门可以打开, 故 $\mathbf{P}(B_3|A_1) = 1/2$;

(2) 跑车在 2 号门里, 主持人只有 3 号门可以打开, 故 $\mathbf{P}(B_3|A_2) = 1$;

(3) 跑车在 3 号门里, 主持人只有 2 号门可以打开, 故 $\mathbf{P}(B_3|A_3) = 0$.

利用全概率公式, 主持人打开 3 号门的概率为

$$\mathbf{P}(B_3) = \sum_{i=1}^3 \mathbf{P}(A_i) \mathbf{P}(B_3|A_i) = \frac{1}{3} \times \left(\frac{1}{2} + 1 + 0 \right) = \frac{1}{2}.$$

再根据贝叶斯公式, 在 3 号门打开的情况下 1 号和 2 号门里有跑车的条件概率分别为

$$\mathbf{P}(A_1|B_3) = \frac{\mathbf{P}(A_1) \mathbf{P}(B_3|A_1)}{\mathbf{P}(B_3)} = \frac{1}{3} \quad (1.3.4)$$

和

$$\mathbf{P}(A_2|B_3) = \frac{\mathbf{P}(A_2) \mathbf{P}(B_3|A_2)}{\mathbf{P}(B_3)} = \frac{2}{3}. \quad (1.3.5)$$

通过比较发现, 改选 2 号门是正确的决策. □

上面利用贝叶斯公式计算条件概率的解答方法包含了重要的思想. 在统计推断中, 称 (1.3.3) 式中的概率为先验概率, 而称 (1.3.4) 式和 (1.3.5) 式中的条件概率为后验概率. 后验概率是根据主持人打开有山羊的 3 号门这个信息, 对于前面的先验概率的修正.

这种通过获得信息利用贝叶斯公式得到后验概率, 并根据后验概率改进决策的过程非常近似于人类的学习模式, 也是贝叶斯方法许多应用的关键. 正是由于这个特点, 贝叶斯方法在机器学习领域发挥了重要作用, 已经成为学习型人工智能的理论基础.

练习题

1. 对于事件 A 和 B , 证明: 若 $\mathbf{P}(A|B) > \mathbf{P}(A)$, 则 $\mathbf{P}(B|A) > \mathbf{P}(B)$.

2. 假设有三个事件 A, B 和 C .

(1) 已知 $\mathbf{P}(A|C) > \mathbf{P}(B|C)$ 和 $\mathbf{P}(A|C^c) > \mathbf{P}(B|C^c)$. 问是否一定有 $\mathbf{P}(A) > \mathbf{P}(B)$? 若是, 加以证明; 否则举出反例.

(2) 已知 $\mathbf{P}(A|C) > \mathbf{P}(A|C^c)$ 和 $\mathbf{P}(B|C) > \mathbf{P}(B|C^c)$. 问是否一定有 $\mathbf{P}(A) > \mathbf{P}(B)$? 若是, 加以证明; 否则举出反例.

3. 考试中的一道多选题共有 5 个选项, 某同学知道正确答案的概率为 p , 他在不知道答案时猜中答案的概率为 $1/5$.

(1) 求该同学做对一道多选题的概率;

(2) 设对某道题该同学选对了答案. 求他确实知道正确答案的概率.

4. 盒子里有 15 个乒乓球, 其中 9 个是新球. 第一场比赛时从中取出了 3 个用完后放回盒中. 第二场比赛时再从中取出 3 个, 求这 3 个都是新球的概率.

5. 在信号传递中发送信号“0”或“1”占比分别为 $1/3$ 和 $2/3$. 由于外界干扰等原因, 接收到的信号可能出现错误. 已知出错的概率为 0.05. 设接收到的信号为“1”, 求发送的信号确实为“1”的概率.

6. 盒子里有标号为 $0, 1, \dots, k$ 的 $k+1$ 枚不均匀的硬币. 已知投掷第 i 号硬币正面朝上的概率为 i/k . 从盒中随机取一枚硬币重复投掷. 设前 n 次投掷的结果都是正面朝上. 求第 $n+1$ 次也是正面朝上的概率.

1.4 事件的独立性

随机事件和随机试验的独立性是非常重要的概念. 首先考虑两个随机事件的独立性.

1.4.1 两个事件的独立性

定义 1.4.1 给定概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$, 称事件 $A, B \in \mathcal{F}$ 是独立的, 是指它们满足

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B). \quad (1.4.1)$$

若 $\mathbf{P}(B) > 0$, 则事件 A, B 独立当且仅当 $\mathbf{P}(A|B) = \mathbf{P}(A)$. 类似地, 若 $\mathbf{P}(A) > 0$, 则事件 A, B 独立当且仅当 $\mathbf{P}(B|A) = \mathbf{P}(B)$. 直观上, 事件 A, B 相互独立意味着事件 B 的发生不会对 A 发生的概率造成影响, 而事件 A 的发生不会对 B 发生的概率造成影响.

独立事件与互斥事件是完全不同的概念. 事实上, 若事件 A, B 独立且 $\mathbf{P}(A) > 0$, $\mathbf{P}(B) > 0$, 则有 $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B) > 0$, 故它们不可能互斥. 反之, 若事件 A, B 互斥且 $\mathbf{P}(A) > 0$, $\mathbf{P}(B) > 0$, 则有 $0 = \mathbf{P}(AB) \neq \mathbf{P}(A)\mathbf{P}(B) > 0$, 故它们不可能独立.

例 1.4.1 若 $A \in \mathcal{F}$ 与自己独立, 则 $\mathbf{P}(A) = \mathbf{P}(AA) = \mathbf{P}(A)^2$, 故有 $\mathbf{P}(A) = 0$ 或 1.

例 1.4.2 袋中有 r 个红球和 b 个黑球. 现连续取球 2 次, 令 $R_i = \{\text{第 } i \text{ 次取到红球}\}$, $i = 1, 2$. 问 R_1 与 R_2 是否独立?

解 (1) 有放回抽取. 此时显然有 $\mathbf{P}(R_i) = r/(r+b)$ 和

$$\mathbf{P}(R_1 R_2) = \frac{r^2}{(r+b)^2} = \mathbf{P}(R_1)\mathbf{P}(R_2),$$

故 R_1 与 R_2 独立.

(2) 无放回抽取. 根据例 1.1.11 有 $\mathbf{P}(R_i) = r/(r+b)$. 所以

$$\mathbf{P}(R_1 R_2) = \frac{r(r-1)}{(r+b)(r+b-1)} \neq \frac{r^2}{(r+b)^2} = \mathbf{P}(R_1)\mathbf{P}(R_2),$$

故 R_1 与 R_2 不独立. □

容易证明下面的结论:

命题 1.4.1 下面的性质等价: 事件 A 与 B 独立; A 与 B^c 独立; A^c 与 B 独立; A^c 与 B^c 独立.

证明 假设事件 A 与 B 独立, 则 $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$. 因此有

$$\begin{aligned}\mathbf{P}(AB^c) &= \mathbf{P}(A \setminus (AB)) = \mathbf{P}(A) - \mathbf{P}(AB) \\ &= \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A)[1 - \mathbf{P}(B)] \\ &= \mathbf{P}(A)\mathbf{P}(B^c),\end{aligned}$$

故 A 与 B^c 独立. 其他蕴涵关系的证明类似. □

例 1.4.3 概率为 0 或 1 的事件与任何事件独立. 事实上, 若 $\mathbf{P}(A) = 0$, 由概率的单调性知 $\mathbf{P}(AB) = 0$, 故 $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$, 即 A 与 B 独立. 若 $\mathbf{P}(A) = 1$, 则 $\mathbf{P}(A^c) = 0$, 由已证结果知 A^c 与 B 独立, 再由命题 1.4.1 知 A 与 B 也独立.

例 1.4.4 向区间 $[0, 1]$ 内随机地任投一粒子.

(i) 用 A_1 表示粒子落在 $[0, 1/2]$ 内, B_1 表示粒子落在子区间 $[1/4, 3/4]$ 内. 则

$$\begin{aligned}\mathbf{P}(A_1) &= 1 - \frac{1}{2} = \frac{1}{2}, \quad \mathbf{P}(B_1) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}, \\ \mathbf{P}(A_1 B_1) &= \frac{1}{2} - \frac{1}{4} = \frac{1}{4} = \mathbf{P}(A_1)\mathbf{P}(B_1).\end{aligned}$$

所以 A, B 不互斥且独立.

(ii) 用 A_2 表示粒子落在 $[1/5, 1/3]$ 内, B_2 表示粒子落在子区间 $[1/4, 1]$ 内. 则

$$\begin{aligned}\mathbf{P}(A_2) &= \frac{1}{3} - \frac{1}{5} = \frac{2}{15}, \quad \mathbf{P}(B_2) = 1 - \frac{1}{4} = \frac{3}{4}, \\ \mathbf{P}(A_2 B_2) &= \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \neq \frac{1}{10} = \mathbf{P}(A_2)\mathbf{P}(B_2).\end{aligned}$$

所以 A_2, B_2 不互斥且不独立.

定义 1.4.2 称概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 上的两个事件类 $\mathcal{A}$ 和 $\mathcal{C}$ 是独立的, 是指对于任何 $A \in \mathcal{A}$ 和 $C \in \mathcal{C}$, 事件 A 和 C 独立.

下面命题的证明用到了本书附录中介绍的邓肯类和邓肯类定理.

命题 1.4.2 设 $\mathcal{F}$ 的子类 $\mathcal{A}$ 和 $\mathcal{C}$ 对交运算封闭且相互独立. 则它们生成的 σ 代数 $\sigma(\mathcal{A})$ 和 $\sigma(\mathcal{C})$ 也相互独立.

证明 对 $A \in \mathcal{F}$ 记 $\mathcal{D}_A = \{B \in \sigma(\mathcal{F}) : \mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)\}$. 易证 $\mathcal{D}_A$ 是邓肯类. 根据假设, 当 $A \in \mathcal{A}$ 有 $\mathcal{D}_A \supseteq \mathcal{C}$. 再由邓肯类定理知 $\mathcal{D}_A \supseteq d(\mathcal{C}) = \sigma(\mathcal{C})$, 即 $A \in \mathcal{A}$ 与任意的 $B \in \sigma(\mathcal{C})$ 独立. 以此为基础, 重复上面推导可知当 $C \in \sigma(\mathcal{C})$ 有 $\mathcal{D}_C \supseteq d(\mathcal{A}) = \sigma(\mathcal{A})$. 这说明任意的 $A \in \sigma(\mathcal{A})$ 和 $C \in \sigma(\mathcal{C})$ 独立. $\square$

1.4.2 多个事件的独立性

定义 1.4.3 给定概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$, 称事件 $A_1, A_2, \dots, A_n \in \mathcal{F}$ ($n \geq 2$) 独立, 是指对任何 $2 \leq k \leq n$ 及 $1 \leq i_1 < i_2 < \dots < i_k \leq n$ 都有

$$\mathbf{P}(A_{i_1}A_{i_2}\cdots A_{i_k}) = \mathbf{P}(A_{i_1})\mathbf{P}(A_{i_2})\cdots\mathbf{P}(A_{i_k}). \quad (1.4.2)$$

显然, 当 $A_1, A_2, \dots, A_n$ 独立时, 其中的任意 m ($2 \leq m \leq n$) 个事件也独立. 注意 (1.4.2) 式中包括了 $2^n - n - 1$ 个式子.

例 1.4.5 考虑样本空间 $\Omega = \{1, 2, 3, 4\}$ 的古典概型. 记 $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{1, 4\}$. 显然

$$\mathbf{P}(A_1A_2) = \mathbf{P}(\{1\}) = \frac{1}{4} = \mathbf{P}(A_1)\mathbf{P}(A_2),$$

$$\mathbf{P}(A_2A_3) = \mathbf{P}(\{1\}) = \frac{1}{4} = \mathbf{P}(A_2)\mathbf{P}(A_3),$$

$$\mathbf{P}(A_3A_1) = \mathbf{P}(\{1\}) = \frac{1}{4} = \mathbf{P}(A_3)\mathbf{P}(A_1).$$

所以 A_1, A_2, A_3 两两独立. 但

$$\mathbf{P}(A_1A_2A_3) = \mathbf{P}(\{1\}) = \frac{1}{4} \neq \frac{1}{8} = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3),$$

故 A_1, A_2, A_3 不是相互独立的. 这说明两两独立不能推出相互独立.

例 1.4.6 考虑样本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 的古典概型. 记 $A_1 = \{1, 2, 3, 4\}$ 和 $A_2 = A_3 = \{4, 5, 6\}$. 则 $\mathbf{P}(A_1) = 2/3$, $\mathbf{P}(A_2) = \mathbf{P}(A_3) = 1/2$, 从而

$$\mathbf{P}(A_1A_2A_3) = \mathbf{P}(\{4\}) = \frac{1}{6} = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3).$$

但是

$$\mathbf{P}(A_2A_3) = \mathbf{P}(A_2) = \frac{1}{2} \neq \mathbf{P}(A_2)\mathbf{P}(A_3).$$

所以 A_1, A_2, A_3 不相互独立.

定理 1.4.3 事件 $A_1, A_2, \dots, A_n \in \mathcal{F}$ 独立当且仅当事件 $\hat{A}_1, \hat{A}_2, \dots, \hat{A}_n \in \mathcal{F}$ 独立, 其中 $\hat{A}_i = A_i$ 或 A_i^c .

证明 与命题 1.4.1 的证明类似, 从略. $\square$

下面的定理给出了两个常用的与独立性有关的概率计算公式.

定理 1.4.4 若 $A_1, A_2, \dots, A_n \in \mathcal{F}$ 独立, 则有

$$\mathbf{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbf{P}(A_i), \quad \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n [1 - \mathbf{P}(A_i)].$$

证明 显然, 第一个等式已经包括在定义 (1.4.2) 式中. 由定理 1.4.3 知 $A_1^c, A_2^c, \dots, A_n^c \in \mathcal{F}$ 独立, 故

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \mathbf{P}\left(\bigcap_{i=1}^n A_i^c\right) = 1 - \prod_{i=1}^n \mathbf{P}(A_i^c),$$

其中 $\mathbf{P}(A_i^c) = 1 - \mathbf{P}(A_i)$, 故得第二个等式. $\square$

在理论和应用中, 也经常需要考虑任意多的事件或事件类的独立性问题. 为此我们给出下面的

定义 1.4.4 设 $A_i, i \in I$ 是某概率空间上的一族随机事件, 其中 I 是任意的非空集. 如果对任何有限子集 $\{i_1, i_2, \dots, i_n\} \subseteq I$ 事件 $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ 都独立, 就称 $A_i, i \in I$ 是独立的.

定义 1.4.5 设 $\mathcal{A}_i, i \in I$ 是某概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 上的一族事件类, 其中 I 是任意的非空集. 如果对于任取的 $A_i \in \mathcal{A}_i, i \in I$, 事件 $A_i, i \in I$ 都独立, 就称 $\mathcal{A}_i, i \in I$ 是独立的.

1.4.3 二维乘积概率空间

定义 1.4.6 假设 $(\Omega_1, \mathcal{F}_1)$ 和 $(\Omega_2, \mathcal{F}_2)$ 是两个可测空间. 我们称乘积集

$$\Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

为 Ω_1 和 Ω_2 的乘积样本空间. 记

$$\mathcal{C} = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}. \quad (1.4.3)$$

称 $\mathcal{F}_1 \times \mathcal{F}_2 := \sigma(\mathcal{C})$ 为 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 的乘积 σ 代数, 称 $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ 为 $(\Omega_1, \mathcal{F}_1)$ 和 $(\Omega_2, \mathcal{F}_2)$ 的乘积可测空间.

例 1.4.7 设 $\Omega_1 = \Omega_2 = \{1, 2\}$, 并令 $\mathcal{F}_1 = \mathcal{F}_2 = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ 为 $\{1, 2\}$ 的所有子集构成的 σ 代数. 再令 $\mathcal{C}$ 由 (1.4.3) 式定义. 不难发现

$$\begin{aligned}\mathcal{C} = \{ & \emptyset, \{1\} \times \{1\}, \{1\} \times \{2\}, \{1\} \times \Omega_2, \\ & \{2\} \times \{1\}, \{2\} \times \{2\}, \{2\} \times \Omega_2, \\ & \Omega_1 \times \{1\}, \Omega_1 \times \{2\}, \Omega_1 \times \Omega_2 \}.\end{aligned}$$

该集类不是 σ 代数. 事实上,

$$(\{1\} \times \{1\}) \cup (\{2\} \times \{2\}) = \{(1, 1)\} \cup \{(2, 2)\} = \{(1, 1), (2, 2)\} \notin \mathcal{C},$$

故 $\mathcal{C}$ 对于并运算不封闭.

定理 1.4.5 对于概率空间 $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$, 存在乘积 σ 代数 $\mathcal{F}_1 \times \mathcal{F}_2$ 上的唯一的概率测度 $\mathbf{Q}$ 满足性质:

$$\mathbf{Q}(A_1 \times A_2) = \mathbf{P}_1(A_1)\mathbf{P}_2(A_2), \quad A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2. \quad (1.4.4)$$

证明 可数样本空间情况. 设 Ω_1 和 Ω_2 都是可数集. 不妨设 $\mathcal{F}_1$ 和 $\mathcal{F}_2$ 分别为 Ω_1 和 Ω_2 的所有子集构成的 σ 代数. 这样 $\mathcal{F}_1 \times \mathcal{F}_2$ 是 $\Omega_1 \times \Omega_2$ 的所有子集构成的 σ 代数. 根据 (1.4.4) 式可以唯一确定 $\mathbf{Q}$ 在 $\mathcal{F}_1 \times \mathcal{F}_2$ 中的单点集上的取值, 即

$$\mathbf{Q}(\{(\omega_1, \omega_2)\}) = \mathbf{P}_1(\{\omega_1\})\mathbf{P}_2(\{\omega_2\}), \quad \omega_1 \in \Omega_1, \omega_2 \in \Omega_2. \quad (1.4.5)$$

由于任何 $A \in \mathcal{F}_1 \times \mathcal{F}_2$ 至多包含可数个点, 利用可数可加性可以唯一确定 $\mathbf{Q}(A)$ 的值. 所以具有性质 (1.4.4) 的概率测度 $\mathbf{Q}$ 存在且唯一. $\square$

定理 1.4.5 的一般情况的证明超出了本课程的范围, 可参考测度论或高等概率论方面的著作, 如见 [19, 20].

定义 1.4.7 对于给定的概率空间 $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$, 我们称乘积 σ 代数 $\mathcal{F}_1 \times \mathcal{F}_2$ 上满足性质 (1.4.4) 的概率测度 $\mathbf{Q}$ 为 $\mathbf{P}_1$ 和 $\mathbf{P}_2$ 的乘积概率测度, 并记作 $\mathbf{P}_1 \times \mathbf{P}_2$. 称 $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbf{P}_1 \times \mathbf{P}_2)$ 为 $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ 的乘积概率空间.

乘积概率空间 $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbf{P}_1 \times \mathbf{P}_2)$ 描述了由 $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ 和 $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ 给出的两次独立试验构成的联合随机试验. 令 $A_1 \in \mathcal{F}_1$ 和 $A_2 \in \mathcal{F}_2$. 则 $A_1 \times \Omega_2 \in \mathcal{F}_1 \times \mathcal{F}_2$ 表示 { 第 1 次试验中事件 A_1 发生 }, 而 $\Omega_1 \times A_2 \in \mathcal{F}_1 \times \mathcal{F}_2$ 表示 { 第 2 次试验中事件 A_2 发生 }. 显然有

$$A_1 \times A_2 = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) \in \mathcal{F}_1 \times \mathcal{F}_2,$$

它表示 { 第 1 次和第 2 次试验中分别发生了事件 A_1 和 A_2 }. 由 (1.4.4) 式有

$$\mathbf{Q}(A_1 \times A_2) = \mathbf{Q}(A_1 \times \Omega_2)\mathbf{Q}(\Omega_1 \times A_2).$$

所以乘积概率空间 $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbf{P}_1 \times \mathbf{P}_2)$ 中的事件 $A_1 \times \Omega_2$ 和 $\Omega_1 \times A_2$ 独立.

例 1.4.8 设袋中有 r 个红球和 b 个黑球. 将 r 个红球用数字 $1, 2, \dots, r$ 编号, b 个黑球用数字 $r+1, r+2, \dots, r+b$ 编号. 定义样本空间

$$\Omega = \{1, \dots, r, r+1, \dots, r+b\},$$

并令 $(\Omega, \mathcal{F}, \mathbf{P})$ 为相应的古典概型概率空间, 即

$$\mathbf{P}(\{k\}) = \frac{1}{r+b}, \quad k \in \Omega. \quad (1.4.6)$$

用此概率空间就可描述一次取球的随机试验. 注意, 事件 $R = \{1, 2, \dots, r\}$ 表示取到红球, 而事件 $B = \{r+1, r+2, \dots, r+b\}$ 表示取到黑球. 根据古典概型的概率计算公式有

$$\mathbf{P}(R) = \frac{r}{r+b}, \quad \mathbf{P}(B) = \frac{b}{r+b}.$$

例 1.4.9 设袋中有 r 个红球和 b 个黑球. 像上例中那样, 将 r 个红球用数字 $1, 2, \dots, r$ 编号, b 个黑球用数字 $r+1, r+2, \dots, r+b$ 编号. 则有放回地两次取球的随机试验可用古典概型概率空间 $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ 描述, 其中

$$\tilde{\Omega} = \{(i, j) : 1 \leq i, j \leq r+b\}.$$

注意对每个样本点 $(i, j) \in \tilde{\Omega}$ 有

$$\tilde{\mathbf{P}}(\{(i, j)\}) = \frac{1}{(r+b)^2}. \quad (1.4.7)$$

注意, 事件 { 第 1 次取到红球 } 可表示为

$$\tilde{R}_1 = \{(i, j) : 1 \leq i \leq r, 1 \leq j \leq r+b\},$$

而事件 { 第 2 次取到黑球 } 可表示为

$$\tilde{B}_2 = \{(i, j) : 1 \leq i \leq r+b, r+1 \leq j \leq r+b\}.$$

根据古典概型的概率计算公式有

$$\tilde{\mathbf{P}}(\tilde{R}_1) = \frac{r(r+b)}{(r+b)^2} = \frac{r}{r+b}, \quad \tilde{\mathbf{P}}(\tilde{B}_2) = \frac{(r+b)b}{(r+b)^2} = \frac{b}{r+b}.$$

事件 { 第 1 次取到红球且第 2 次取到黑球 } 可表示为

$$\tilde{R}_1 \cap \tilde{B}_2 = \{(i, j) : 1 \leq i \leq r, r+1 \leq j \leq r+b\}.$$

不难看出,

$$\tilde{\mathbf{P}}(\tilde{R}_1 \cap \tilde{B}_2) = \frac{rb}{(r+b)^2}.$$

这说明事件 $\tilde{R}_1, \tilde{B}_2$ 独立. 若令 $(\Omega, \mathcal{F}, \mathbf{P})$ 为上例中定义的古典概型概率空间, 由 (1.4.6) 式和 (1.4.7) 式有

$$\tilde{\mathbf{P}}(\{(i, j)\}) = \frac{1}{(r+b)^2} = \mathbf{P}(\{i\})\mathbf{P}(\{j\}).$$

对比上式和 (1.4.5) 式知 $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}})$ 实际上是 $(\Omega, \mathcal{F}, \mathbf{P})$ 与自己的乘积概率空间.

*1.4.4 多维乘积概率空间

与二维乘积概率空间类似, 我们可以定义有限或可数维乘积概率空间, 它给出了由有限次或可数次独立试验构成的联合随机试验的数学描述.

定义 1.4.8 设 $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i), i = 1, 2, \dots, n$ 是 n 个概率空间. 我们称乘积集

$$\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n := \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \Omega_i; i = 1, 2, \dots, n\}$$

为 $\Omega_i, i = 1, 2, \dots, n$ 的乘积样本空间. 记

$$\mathcal{C} = \{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{F}_i; i = 1, 2, \dots, n\}. \quad (1.4.8)$$

称 $\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n := \sigma(\mathcal{C})$ 为 $\mathcal{F}_i, i = 1, 2, \dots, n$ 的乘积 σ 代数. 此时存在 $\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n$ 上的唯一的概率测度 $\mathbf{Q} := \mathbf{P}_1 \times \mathbf{P}_2 \times \cdots \times \mathbf{P}_n$ 满足性质:

$$\mathbf{Q}(A_1 \times A_2 \times \cdots \times A_n) = \mathbf{P}_1(A_1)\mathbf{P}_2(A_2) \cdots \mathbf{P}_n(A_n), \quad A_i \in \mathcal{F}_i; i = 1, 2, \dots, n. \quad (1.4.9)$$

称 $\mathbf{P}_1 \times \mathbf{P}_2 \times \cdots \times \mathbf{P}_n$ 为 $\mathbf{P}_i, i = 1, 2, \dots, n$ 的乘积概率测度, 并称

$$(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n, \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n, \mathbf{P}_1 \times \mathbf{P}_2 \times \cdots \times \mathbf{P}_n)$$

为 $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i), i = 1, 2, \dots, n$ 的乘积概率空间.

定义 1.4.9 设 $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i), i = 1, 2, \dots$ 是可数个概率空间. 我们称乘积集

$$\Omega_1 \times \Omega_2 \times \cdots := \{(\omega_1, \omega_2, \dots) : \omega_i \in \Omega_i; i = 1, 2, \dots\}$$

为 $\Omega_i, i = 1, 2, \dots$ 的乘积样本空间. 记

$$\begin{aligned} \mathcal{C} = \{A_1 \times A_2 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots : \\ A_i \in \mathcal{F}_i; i = 1, 2, \dots, n; n = 1, 2, \dots\}. \end{aligned} \quad (1.4.10)$$

称 $\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots := \sigma(\mathcal{C})$ 为 $\mathcal{F}_i, i = 1, 2, \dots$ 的乘积 σ 代数. 此时存在 $\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots$ 上的唯一的概率测度 $\mathbf{Q} := \mathbf{P}_1 \times \mathbf{P}_2 \times \cdots$ 满足性质: 对任意的 $A_i \in \mathcal{F}_i; i = 1, 2, \dots, n; n = 1, 2, \dots$,

$$\mathbf{Q}(A_1 \times A_2 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots) = \mathbf{P}_1(A_1)\mathbf{P}_2(A_2) \cdots \mathbf{P}_n(A_n). \quad (1.4.11)$$

称 $\mathbf{P}_1 \times \mathbf{P}_2 \times \cdots$ 为 $\mathbf{P}_i, i = 1, 2, \cdots$ 的乘积概率测度, 并称

$$(\Omega_1 \times \Omega_2 \times \cdots, \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots, \mathbf{P}_1 \times \mathbf{P}_2 \times \cdots)$$

为 $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i), i = 1, 2, \cdots$ 的乘积概率空间.

关于乘积概率测度 $\mathbf{P}_1 \times \mathbf{P}_2 \times \cdots \times \mathbf{P}_n$ 和 $\mathbf{P}_1 \times \mathbf{P}_2 \times \cdots$ 的存在唯一性, 可参考测度论或高等概率论的书籍. 特别地, 当 $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i) = (\Omega, \mathcal{F}, \mathbf{P})$ 与标号 i 无关时, 可将上面定义的乘积概率空间分别记作 $(\Omega^n, \mathcal{F}^n, \mathbf{P}^n)$ 和 $(\Omega^\infty, \mathcal{F}^\infty, \mathbf{P}^\infty)$, 并分别称之为 $(\Omega, \mathcal{F}, \mathbf{P})$ 的 n 维乘积概率空间和可数无穷维乘积概率空间.

例 1.4.10 令 $(\Omega_1, \mathcal{F}_1, \mathbf{P}_1)$ 为关于成功事件 $A \subseteq \Omega_1$ 的一次伯努利试验对应的概率空间, 其中 $\mathcal{F}_1 = \{\emptyset, A, A^c, \Omega\}$. 设成功概率为 $\mathbf{P}_1(A) = p$ ($0 < p < 1$). 考虑其可数无穷维乘积概率空间 $(\Omega, \mathcal{F}, \mathbf{P}) := (\Omega_1^\infty, \mathcal{F}_1^\infty, \mathbf{P}_1^\infty)$. 此概率空间所描述的随机试验为可数无穷次伯努利试验. 对 $k \geq 1$ 记

$$B_k = \Omega_1^{k-1} \times A \times \Omega_1^\infty,$$

则有

$$B_k^c = \Omega_1^{k-1} \times A^c \times \Omega_1^\infty.$$

根据 (1.4.10) 式和可数无穷维乘积 σ 代数的定义, 显然有 $B_k \in \mathcal{F}_1^\infty$ 和 $B_k^c \in \mathcal{F}_1^\infty$. 这两个事件分别表示第 k 次试验的结果为 A 和 A^c . 利用 (1.4.11) 式不难算出 $\mathbf{P}(B_k) = \mathbf{P}_1(A) = p$ 和 $\mathbf{P}(B_k^c) = \mathbf{P}_1(A^c) = 1 - p$. 另外, 易证事件 $B_1, B_2, \cdots$ 相互独立.

例 1.4.11 求在关于成功事件 A 的可数无穷次伯努利试验中事件 $E = \{A \text{ 终将发生}\}$ 的概率.

解 对于 $n \geq 1$ 令 $F_n = \{\text{前 } n \text{ 次都失败}\}$. 采用上例的记号, 我们有 $F_n = B_1^c \cap B_2^c \cap \cdots \cap B_n^c$ 和 $\mathbf{P}(F_n) = (1-p)^n$. 注意 $\{F_n\}$ 是单调下降的集序列, 且 $E^c = \bigcap_{n=1}^{\infty} F_n$. 由概率的上连续性得

$$\mathbf{P}(E) = 1 - \mathbf{P}(E^c) = 1 - \lim_{n \rightarrow \infty} \mathbf{P}(F_n) = 1 - \lim_{n \rightarrow \infty} (1-p)^n = 1.$$

这说明持续不断地将试验进行下去, 最终一定能够成功. □

练习题

1. 证明: (1) 若事件 A 与自身独立, 则 $\mathbf{P}(A) = 0$ 或 1 ;

(2) 若 $\mathbf{P}(A) = 0$ 或 1 , 则 A 与任意事件独立.

2. 对于某个家庭考察两个事件 $A = \{\text{至多有一个女孩}\}$ 和 $B = \{\text{男女孩都有}\}$. 设男女孩的出生率都是 $1/2$. 证明: 对 3 孩之家, 事件 A 和 B 独立; 而对 4 孩之家, 事件 A 和 B 不独立.

3. 证明: 事件 $A_1, A_2, \dots, A_n$ 独立的充要条件是下列 2^n 个等式成立:

$$\mathbf{P}(\hat{A}_1 \hat{A}_2 \cdots \hat{A}_n) = \mathbf{P}(\hat{A}_1) \mathbf{P}(\hat{A}_2) \cdots \mathbf{P}(\hat{A}_n),$$

其中 $\hat{A}_i$ 取 A_i 或 A_i^c .

4. 独立地重复投掷两枚骰子, 并计算点数之和. 求“和数为 5”出现在“和数为 7”之前的概率.

5. 设 E 和 F 是某随机试验中的互不相容事件. 证明: 若独立重复地进行这样的试验, 则 E 发生在 F 之前的概率为 $\mathbf{P}(E)/[\mathbf{P}(E) + \mathbf{P}(F)]$.

6. 重复投掷硬币, 设第一次出现正面的概率为 c ($0 \leq c \leq 1$), 之后每次与上次出现相同面的概率为 p ($0 < p < 1$). 求第 n 次出现正面的概率 p_n , 并讨论当 $n \rightarrow \infty$ 时 p_n 的极限.

提示: 根据全概率公式有 $p_n = (1 - p) + (2p - 1)p_{n-1}$, 于是有

$$\begin{aligned} p_n &= (1 - p)[1 + (2p - 1) + \cdots + (2p - 1)^{n-2}] + c(2p - 1)^{n-1} \\ &= 1/2 - (2p - 1)^{n-1}/2 + c(2p - 1)^{n-1}. \end{aligned}$$