1.2 频率与概率公理化

随机事件在一次试验中可能发生、也可能不发生, 我们通常关心随机事件发生的可能性究竟有多大, 最好能用介于 0 和 1 之间的一个数来进行刻画. 为此, 我们首先引入频率, 用以描述随机事件发生的频繁程度, 然后介绍刻画随机事件发生可能性大小的数, 即事件的概率.

1.2.1 频率

定义 1.1 随机事件 A 在相同条件下重复进行的 n 次试验中出现了 n_A 次, 则称

$$f_n(A) = n_A/n$$

为事件 A 在 n 次试验中发生的 **频率**, 并称 n_A 为事件 A 发生的 **频数**.

直观而言, 事件的频率在一定程度上反应了事件发生的可能性, 若事件发生的频率越大, 则事件 A 发生越频繁, 因而事件在一次试验中发生的可能性越大, 反之亦然. 根据上述的定义可知频率具有如下性质:

1° 对任意事件 A 有 $f_n(A) \in [0,1]$;

2° 对必然事件 Ω 有 $f_n(\Omega) = 1$;

3° 对 k 个互不相容的事件 $A_1, A_2, \ldots, A_k$ 有

$$f_n(A_1 \cup A_2 \cup \cdots \cup A_k) = f_n(A_1) + f_n(A_2) + \cdots + f_n(A_k).$$

性质 1° 和 2° 根据定义显然成立. 对互不相容的事件 $A_1, A_2, \ldots, A_k$, 并事件 $A_1 \cup A_2 \cup \cdots \cup A_k$ 发生的频数等于每个事件 A_i $(i \in [k])$ 发生的频数之和, 由此可知性质 3° 成立.

频率在实际中往往表现出一定的随机性, 例如, 在相同条件下进行两轮 n 次试验, 每轮试验中事件 A 发生的频率往往不同; 其次, 随着试验次数 n 的增加, 事件 A 发生的频率 $f_n(A)$ 会发生一定的变化, 表现出一定的随机性.

尽管频率会随着试验次数 n 的变化表现出一定的随机性, 但在大量重复的试验中, 事件的频率通常在一个确定的常数 p 附近摆动, 而且随着试验次数的增大, 摆幅越来越小, 频率也越来越稳定于常数 p, 将这种规律称为 **频率的稳定性**. 例如, 历史上多人进行重复投掷硬币的试验, 下面给出了一些人的试验统计结果:

表 1.2 历史上多人重复投掷硬币的试验结果

实 验 者	投掷总数	正面朝上的频数	正面朝上的频率
德摩根	2048	1061	0.5181
蒲 丰	4040	2048	0.5069
K. 皮尔逊	12000	6019	0.5016
K. 皮尔逊	24000	12012	0.5005

图 1.2 任意投掷硬币, 正面朝上频率的趋势

也可以利用计算机产生随机数对投掷硬币的试验进行仿真, 图 1.2 给出了实验结果. 这些实验结果均表明, 尽管对不同的投掷总数, 正面朝上的频率不并相同, 但随着投掷次数的增加, 正面朝上的频率越来越接近一个常数 (0.5), 即频率逐渐稳定于 0.5. 这种频率的稳定性即通常所说的统计规律性, 是随机事件本身所固有的客观属性, 可用于度量事件发生的可能性大小.

定义 1.2 随机事件 A 在大量重复试验中发生的频率总是稳定地在一个常数 p 附近摆动, 且随着试验次数的增加而摆幅逐渐越小, 则称常数 p 为事件 A 发生的 **概率**, 记为 $P(A)=p$.

该定义又称 **概率的统计定义**, 其概率称为 **统计概率**, 提供了计算随机事件发生的概率的一种方法, 即当试验次数足够多时, 可用频率来给出事件概率的近似值. 然而概率的统计定义存在数学上的不严谨性, 而在实际中几乎不可能每一个事件做大量重复的试验来计算频率, 进而近似概率, 以此刻画事情发生的可能性. 受到频率的稳定性及其性质的启发, 下一节给出严谨的概率公理化定义.

1.2.2 概率公理化

20 世纪 30 年代, 前苏联数学家柯尔莫哥洛夫 (A. Kolmogorov) 提出了概率论的公理化体系, 通过基本的性质给出了概率的严格定义, 期望建立媲美于欧氏几何公理化的理论体系.

定义 1.3 (概率公理化定义) 随机试验 E 所对应的样本空间 Ω 中每一个随机事件 A, 均赋予一实数 $P(A)$, 且满足以下条件:

1° **非负性**: 对任意事件 A 有 $P(A)\geqslant 0$;

2° **规范性**: 对样本空间 Ω 有 $P(\Omega)=1$;

3° **可列可加性**: 若 $A_1,A_2,\ldots,A_n,\ldots$ 是可列无穷个互不相容的事件, 即 $A_iA_j=\emptyset\ (i\neq j)$, 有

$$P(A_1\cup A_2\cup\cdots\cup A_n\cup\cdots)=P(A_1)+P(A_2)+\cdots+P(A_n)+\cdots;$$

则称 $P(A)$ 为随机事件 A 的 **概率**.

由定义可知概率 $P(A)$ 是在样本空间 Ω 中所有随机事件做构成的集合上定义的实值函数. 概率的三条公理 (非负性、规范性和可列可加性) 简明扼要地刻画了概率的定义, 为现代概率论奠定了基础, 公理化体系是概率论发展历史上的一个里程碑, 从此概率论被公认为数学的一个分支.

在后续章节中将证明当 $n \to \infty$ 时频率 $f_n(A)$ 在一定意义下接近概率 $P(A)$, 因此用概率 $P(A)$ 刻画事件 A 发生的可能性大小是合理的. 我们将根据概率的三条公理化定义, 推导概率的重要性质.

性质 1.1　对不可能事件 $\emptyset$ 有 $P(\emptyset) = 0$.

证明　令 $A_i = \emptyset\ (i = 1, 2, \ldots)$, 则有 $\emptyset = \cup_{i=1}^{\infty} A_i$ 且 $A_i \cap A_j = \emptyset\ (i \neq j)$. 根据公理 3° 有

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

再根据公理 1° 可知 $P(\emptyset) = 0$.

注: 不可能事件 $\emptyset$ 的概率为 0, 但概率为 0 的事件并不一定是不可能事件; 同理, 必然事件 Ω 的概率为 1, 但概率为 1 的事件并不一定是必然事件. 反例参考后面所学的连续函数或几何概型在一个样本点处的概率.

性质 1.2 (有限可加性)　若 $A_1, A_2, \cdots, A_n$ 是两两不相容事件, 则

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

证明　令 $A_i = \emptyset\ (i > n)$, 则有 $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i$, 且 $A_1, A_2, \cdots, A_n, A_{n+1}, \ldots$ 是两两互不相容事件. 根据公理 3° 可知

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) = \sum_{i=1}^{n} P(A_i)$$

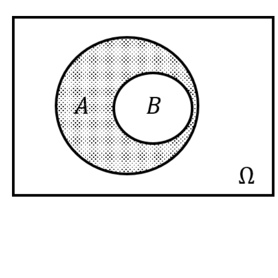
性质得证.

性质 1.3　对任意事件 A, 有 $P(\bar{A}) = 1 - P(A)$.

证明　由于 $\Omega = \bar{A} \cup A$, 以及事件 A 与 $\bar{A}$ 互不相容, 根据有限可加性有 $1 = P(\Omega) = P(A) + P(\bar{A})$.

性质 1.4　若事件 $B \subset A$, 则有 $P(B) \leqslant P(A)$ 和 $P(A - B) = P(A) - P(B)$.

证明　　若 $B \subset A$, 如右图所示有 $A = B \cup (A - B)$, 根据定义可知 B 与 $A - B$ 互不相容。由有限可加性有



$$P(A) = P(B) + P(A - B).$$

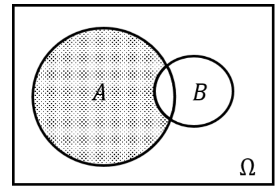
再根据公理 1° 有 $P(A - B) = P(A) - P(B) \geqslant 0$, 从而得到 $P(A) \geqslant P(B)$.

性质 1.5 对任意事件 A 和 B, 有

$$P(A-B)=P(A)-P(AB)=P(A\cup B)-P(B).$$

证明 根据 $A=(A-B)\cup(AB)$, 以及 $A-B$ 与 AB 互斥, 有

$$P(A)=P(A-B)+P(AB).$$



再根据 $A\cup B=(A-B)\cup B$, 以及 $A-B$ 与 B 互斥, 有

$$P(A-B)=P(A\cup B)-P(B).$$

性质 1.6 (容斥原理 Inclusion-Exclusion Principle) 对任意随机事件 A 和 B 有

$$P(A\cup B)=P(A)+P(B)-P(AB).$$

证明 因 $A\cup B=(A-B)\cup(AB)\cup(B-A)$, 以及 $A-B$, $B-A$, AB两两互不相容, 由有限可加性可知

$$P(A\cup B)=P(A-B)+P(B-A)+P(AB).$$

再将 $P(A-B)=P(A)-P(AB)$ 和 $P(B-A)=P(B)-P(AB)$ 代入上式即可完成证明.

类似地, 对三个随机事件 A,B,C 有

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC).$$

对 n 个随机事件 $A_1,A_2,\ldots,A_n$ 有

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1\cdots A_n).$$

对 n 个随机事件 $A_1,A_2,\ldots,A_n$ 的容斥原理可进一步简写为

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{r=1}^{n}(-1)^{r+1}\sum_{i_1<\cdots<i_r}P(A_{i_1}\cdots A_{i_r}).$$

性质 1.7 (Union Bound) 对事件 $A_1,A_2,\ldots A_n$ 有

$$P(A_1\cup A_2\cup\cdots\cup A_n)\leqslant P(A_1)+P(A_2)+\cdots+P(A_n).$$

证明 我们利用数学归纳法进行证明. 当 $n=2$ 时, 由容斥原理有

$$P(A\cup B)=P(A)+P(B)-P(AB)\leqslant P(A)+P(B). \tag{1.1}$$

假设当 $n=k$ 时性质成立, 对 $n=k+1$ 有

$$\begin{aligned}P(A_1\cup\cdots\cup A_{k+1})&=P((A_1\cup\cdots\cup A_k)\cup A_{k+1})\\&\leqslant P(A_1\cup\cdots\cup A_k)+P(A_{k+1})\leqslant P(A_1)+\cdots+P(A_k)+P(A_{k+1}),\end{aligned}$$

这里第一个不等式成立是根据式 (1.1), 而第二个不等式成立是根据归纳假设. 完成证明.

根据数学归纳法可类似推得到 Bonferroni 不等式: 对事件 $A_1,A_2,\ldots A_n$ 有

$$\begin{aligned}P\left(\bigcup_{i=1}^n A_i\right)&\leqslant\sum_{i=1}^n P(A_i);\\P\left(\bigcup_{i=1}^n A_i\right)&\geqslant\sum_{i=1}^n P(A_i)-\sum_{i<j}P(A_iA_j);\\P\left(\bigcup_{i=1}^n A_i\right)&\leqslant\sum_{i=1}^n P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k);\end{aligned}$$

可依次类推相关不等式.

例 1.4 设 $P(A)=p$, $P(B)=q$, $P(AB)=r$, 用 p,q,r 分别表示事件的概率: 1) $P(\bar A\cup\bar B)$, 2) $P(\bar AB)$; 3) $P(\bar A\cup B)$; 4) $P(\bar A\cap\bar B)$.

解 对问题 1), 根据事件的对偶律有

$$P(\bar A\cup\bar B)=P(\overline{AB})=1-r.$$

对问题 2), 根据差事件的定义

$$P(\bar AB)=P(B-A)=P(B)-P(AB)=q-r.$$

对问题 3), 根据容斥原理有

$$P(\bar A\cup B)=P(\bar A)+P(B)-P(\bar AB)=1-p+q-(q-r)=1-p+r.$$

对问题 4), 根据对偶律与容斥原理有

$$P(\bar A\cap\bar B)=P(\overline{A\cup B})=1-P(A\cup B)=1+r-p-q.$$

例 1.5 设三个随机事件 A, B, C 满足 $P(A) = P(B) = P(C) = 1/4$, $P(AB) = 0$, $P(AC) = P(BC) = 1/16$, 求事件 A, B, C 中至少有一个事件发生的概率.

解 首先根据三个事件的容斥原理有

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \\ &= 3/4 - 1/8 + P(ABC). \end{aligned}$$

根据 $P(AB) = 0$ 和 $ABC \subset AB$ 可知

$$0 \leqslant P(ABC) \leqslant P(AB) = 0$$

由此可知事件 A, B, C 中至少有一个事件发生的概率为 5/8.

1.3 古典概型与几何概型

本节介绍两种历史较为久远的经典概率模型: 古典概型与几何概型.

1.3.1 古典概型

首先研究一类简单的随机现象, 它是概率论早期最重要的研究对象, 其发展在概率论中具有重要的意义, 并在产品质量抽样检测等问题中具有广泛的应用.

定义 1.4 (古典概型) 如果试验 E 满足:

- 试验的结果只有有限种可能, 即样本空间 $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, 其中 w_i 为基本事件,
- 每种结果发生的可能性相同, 即 $P(\{\omega_i\}) = P(\{\omega_j\})$ $(i \neq j)$,

则称该类试验称为 **古典概型**, 又称 **等可能概型**.

根据上述定义以及 $P(\Omega) = 1$ 可知: 每个基本事件发生的概率为 $P(\{\omega_i\}) = 1/n$, 若事件 A 包含 k 个基本事件 $\{\omega_{i_1}, \omega_{i_1}, \dots, \omega_{i_k}\}$, 则事件 A 发生的概率为

$$P(A) = k/n = |A|/|\Omega|,$$

这里 $|A|$ 表示事件 A 包含的事件的个数. 很显然古典概型的概率满足概率公理化体系的三条公理.

在使用古典概型计算概率时需注意每个基本事件发生的可行性大小是否相等, 例如,

例 1.6 在相同条件下连续两次抛一枚均匀硬币, 此试验观察的结果: A) 两正面, B) 两反面, C)一正一反. 根据古典概型可知

$$P(A) = P(B) = P(C) = 1/3.$$

然而这种结论不正确, 因为这三个事件发生的可能性不同. 正确的理解是事件 $C=\{C_1,C_2\}$, 其中 C_1 表示先正后反的事件, C_2 表示先反后正的事件, 从而有

$$P(A)=P(B)=P(C_1)=P(C_2)=1/4.$$

古典概率计算的本质是计数 (Counting), 下面介绍一些基本的计数原理, 更为详细的计数方法将在下一节介绍. 首先介绍计数的两条基本原理:

- **加法原理**: 若一项工作可以用两种不同的过程 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 完成, 且过程 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 分别有 n_1 和 n_2 种方法, 则完成该工作有 n_1+n_2 种方法.
- **乘法原理**: 若一项工作需要依次通过 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 两过程, 且过程 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 分别有 n_1 和 n_2 种方法, 则完成该工作有 $n_1\times n_2$ 种方法.

上述两条原理可进一步推广到多个过程的情况.

下面介绍无放回的排列组合, 更为复杂的排列组合将在下一节介绍:

排列: 从 n 个不同的元素中无放回地取出 r 个元素进行排列, 此时既要考虑取出的元素, 也要顾及其排列顺序, 则有 $(n)_r=n(n-1)\cdots(n-r+1)$ 种不同的排列. 若 $r=n$ 时称全排列, 有 $n!$ 种.

组合: 从 n 个不同的元素中无放回地取出 r 个元素, 取出的元素之间无顺序关系, 共有 $\binom{n}{r}$ 种不同的取法, 其中

$$\binom{n}{r}=\frac{n!}{r!(n-r)!}=\frac{(n)_r}{r!},\quad \text{且记}\quad \binom{n}{0}=1.$$

这里 $\binom{n}{r}$ 称为 **组合数** 或 **二项系数**, 它是二项展开式 $(a+b)^n=\sum_{r=0}^{n}\binom{n}{r}a^rb^{n-r}$ 中项 a^rb^{n-r} 的系数.

例 1.7 将 n 个不同的球随机放入 N $(N\geqslant n)$ 个不同的盒子中, 事件 A 表示恰有 n 个盒子且每盒一球; 事件 B 表示指定的 n 个盒子中各有一球; 事件 C 表示指定一盒子恰有 m 个球. 求事件 A,B,C 发生的概率. (盒子的容量不限, 放入同一个盒子内的球无顺序排列区别)

解 将 n 只不同的球随机放入 N 个不同的盒子中, 共有 N^n 种不同的放法. 而对事件 A, 有 $(N)_n=N!/n!$ 种不同的放法, 因此

$$P(A)=\frac{(N)_n}{N^n}=\frac{N!}{N^n n!}.$$

对事件 B, 有 $n!$ 种不同的放法, 因此

$$P(B)=\frac{n!}{N^n}.$$

对事件 C, 可分为两步: 第一步在指定的盒子内放入 m 个球, 有 $\binom{n}{m}$ 种不同的放法; 第二步将剩下的 $n-m$ 个球放入 $N-1$ 个盒子, 有 $(N-1)^{n-m}$ 种不同的放法. 因此

$$P(C)=\frac{\binom{n}{m}(N-1)^{n-m}}{N^n}.$$