第 1 章　随机事件与概率

对自然界和人类社会存在的各种现象进行观察, 会发现有一类现象在一定条件下是必然发生的, 常被称为 **必然现象**, 又称 **确定性现象**. 例如, 太阳从东边升起; 成熟的苹果会从树上掉下来; 在标准大气压下, 水在 0°C 以下会结冰, 加热到 100°C 以上会沸腾; 平面三角形两边之和大于第三边; 等等. 这些现象发生的条件与结果之间具有确定性关系, 可用确切的数学函数进行描述.

然而在自然界和人类社会也往往存在着另一类不确定的现象. 例如, 我们今晚能否观察到流星; 随意投掷一枚硬币, 可能正面朝上, 也可能反面朝上; 当你穿过马路时, 遇见的信号灯可能是绿色, 也可能是红色; 两位相恋的人最终能否走在一起; 等等. 这些现象在一定条件下可能出现这种结果, 也可能出现那种结果, 出现的结果并不唯一, 而事先不能确定哪种结果会出现, 常被称为 **随机现象**, 这类现象发生的条件与结果之间具有不确定性关系, 无法通过确切的数学函数进行刻画.

随机现象尽管在一次观察中无法确定哪种结果发生, 表现出不确定性或偶然性. 然而经过人们长期的研究后发现: 在大量重复的实验中, 随机现象的结果却表现出固有的规律性, 即 **统计规律性**. 例如多次重复投掷一枚硬币, 得到的正面朝上的次数和反面朝上的次数几乎总是差不多. 因此随机现象通常表现出二重属性:

- **偶然性**: 对随机现象进行一次观察, 其结果表现出不确定性;
- **必然性**: 对随机现象进行大量重复观察, 其结果呈现出固有的统计规律性.

概率论与数理统计是研究和揭示随机现象统计规律性的一门学科, 应用几乎遍及所有的科学技术领域、行业生产、国民经济生活等, 如法国著名数学家拉普拉斯 (P. S. Laplace, 1794-1827) 所言: “对生活的大部分, 最重要的问题实际上都是概率问题”. 图灵奖得主 Y. LeCun 在其自传中指出: “历史上大多数重要成果的出现都是偶然事件... 所有的努力都是为了提高概率”. 而对现实生活中的每个人而言: 所有的努力都是为了提高成功的概率.

1.1 随机事件及其运算

为研究和揭示随机现象的规律, 通常需要在相同的条件下重复进行一系列实验和观察, 常被称为 **随机试验**, 或简称为 **试验**. 一般用 E 或 E_1, E_2, $E_3, \ldots$ 表示, 本书所提及的试验均是随机试验. 下面给出一些随机试验的例子:

E_1: 随意抛一枚硬币, 观察正面朝上还是反面朝上.

E_2: 随意抛一枚骰子, 观察出现的点数.

E_3: 统计某地区一年内出生的婴儿数量.

E_4: 随机选取一盏电灯, 测试其寿命.

这些试验具有一些共有的特点: 试验在相同的条件下可重复进行, 具有多种结果, 我们已知每次试验所有可能的结果, 但在每次试验之前不确定出现哪种结果. 例如抛硬币有正面/反面朝上两种结果, 在相同的条件下可以重复进行, 且每次试验前不确定正面/反面朝上. 概括而言, 随机试验具有以下三个特点:

- **可重复**: 在相同的条件下试验可重复进行;
- **多结果**: 试验结果不唯一, 所有可能发生的结果事先明确已知;
- **不确定**: 试验前无法预测或确定哪一种结果会发生.

1.1.1 样本空间与随机事件

随机试验尽管在每次试验前不能确定发生的结果, 但其所有可能发生的结果却是事先已知的. 将随机试验 E 所有可能的结果构成的集合称为试验 E 的 **样本空间**, 记为 Ω. 样本空间 Ω 中的每个元素, 即试验 E 的每种结果, 称为 **样本点**, 记为 ω.

例如前一页所述的四种试验, 其样本空间分别为:

试验 E_1 的样本空间为 $\Omega_1 = \{$正面, 反面$\}$, 样本点分别为 ω_1=正面, ω_2=反面.

试验 E_2 的样本空间为 $\Omega_2 = \{1,2,3,4,5,6\}$, 样本点分别为 $\omega_1 = 1$, $\omega_2 = 2$, $\ldots$, $\omega_6 = 6$.

试验 E_3 的样本空间为 $\Omega_3 = \{0,1,2,\ldots\}$, 样本点为任意非负整数.

试验 E_4 的样本空间为 $\Omega_4 = \{t\colon t \geqslant 0\}$, 样本点为任意非负数.

包含有限个样本点的样本空间称为 **有限样本空间**, 如样本空间 Ω_1 和 Ω_2. 包含无限但可列多个样本点的样本空间称为 **可列样本空间**, 如样本空间 Ω_3. 有限样本空间和无限可列样本空间统称为 **离散样本空间**. 包含无限不可列个样本点的样本空间称为 **不可列样本空间**, 如样本空间 Ω_4.

在随机试验中, 通常关心具有某些特性的样本点构成的集合, 称之为 **随机事件**, 简称为 **事件**, 一般用大写字母 $A, B, C, \ldots$ 表示. 随机事件的本质是集合, 由单个或某些样本点所构成的集合, 是样本空间 Ω 的子集. 如果随机试验的结果是事件 A 中包含的元素, 则称 **事件 A 发生**.

只包含一样本点的事件称为 **基本事件**, 包含两个或两个以上样本点的事件称为 **复合事件**. 样本空间 Ω 包含所有样本点, 是其自身的子集, 每次试验必然发生, 因而称 Ω 为 **必然事件**. 另一方面, 如果某事件在每次试验中都不发生, 则该事件不可能包含任何样本点, 我们用空集符号 $\emptyset$ 表示, 且称 $\emptyset$ 为 **不可能事件**.

例 1.1 随机试验 E: 抛一枚骰子观察其出现的点数, 其样本空间 $\Omega = \{1, 2, \ldots, 6\}$, 则:
事件 A 表示抛骰子的点数为 2, 则 $A = \{2\}$ 为基本事件;
事件 B 表示抛骰子的点数为偶数, 则 $B = \{2, 4, 6\}$;
事件 C 表示抛骰子的点数大于 7, 则 $C = \emptyset$ 为不可能事件;
事件 D 表示抛骰子的点数小于 7, 则 $D = \Omega$ 为必然事件.

1.1.2 随机事件的关系与运算

随机事件的本质是样本空间的子集, 因此随机事件的关系与运算可类比于集合论的关系与运算. 下面默认随机试验的样本空间为 Ω, 用 A, B, A_i $(i = 1, 2, \ldots)$ 表示样本空间 Ω 中的随机事件.

1) **包含事件** 若事件 A 发生必将导致事件 B 发生, 则称 **B 包含 A**, 记为 $A \subset B$ 或 $B \supset A$.
   若 $A \subset B$ 且 $B \subset A$, 则称事件 A 与 B **相等**, 记为 $A = B$.

2) **事件的并/和** 若事件 A 和 B 中至少有一个发生所构成的事件称为 **事件 A 与 B 的并** (或 **和**) **事件**, 记为 $A \cup B$, 即
$$A \cup B = \{\omega\colon \omega \in A \text{ 或 } \omega \in B\}.$$
   类似地, 事件 $A_1, A_2, \ldots, A_n$ 中至少有一个发生所构成的事件称为事件 $A_1, A_2, \ldots, A_n$ 的并事件, 记为
$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = \{\omega\colon \exists\, i \in [n] \text{ s.t. } \omega \in A_i\}\ ,$$
   称 $\bigcup_{i=1}^{\infty} A_i$ 为可列个事件 $A_1, A_2, \cdots$ 的并事件.

3) **事件的交/积** 若事件 A 和 B 同时发生所构成的事件称为 **事件 A 与 B 的交** (或 **积**) **事件**, 记为 $A \cap B$ 或 AB, 即
$$AB = A \cap B = \{\omega\colon \omega \in A \text{ 且 } \omega \in B\}\ .$$
   类似地, 事件 $A_1, A_2, \ldots, A_n$ 同时发生所构成的事件称为事件 $A_1, A_2, \ldots, A_n$ 的交事件, 记为
$$A_1 A_2 \cdots A_n = \bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n = \{\omega\colon \forall i \in [n] \text{ s.t. } \omega \in A_i\}\ ,$$
   称 $\bigcap_{i=1}^{\infty} A_i$ 为可列个事件 $A_1, A_2, \cdots$ 的交事件.

4) **事件的差** 若事件 A 发生但事件 B 不发生所构成的事件称为 **事件 A 与 B 的差**, 记 $A - B$,
$$A - B = A - AB = A\bar{B} = \{\omega\colon \omega \in A \text{ 且 } \omega \notin B\}.$$

5) **对立/逆事件** 对事件 A 而言, 所有不属于事件 A 的基本事件所构成的事件称为 **事件 A 的对立事件** 或 **逆事件**, 记为 $\bar{A}$, 即 $\bar{A} = \Omega - A$.
   根据定义可知 $\bar{\bar{A}} = A$, $\bar{A} \cap A = \emptyset$ 和 $\Omega = A \cup \bar{A}$.

6) **互不相容/互斥事件** 若事件 A 和事件 B 不能同时发生, 即 $A \cap B = \emptyset$, 则称事件 A 和 B 是 **互不相容的** 或 **互斥的**.
   若事件 $A_1, A_2, \ldots, A_n$ 中任意两事件不可能同时发生, 即对任意 $i \neq j$ 有 $A_i \cap A_j = \emptyset$ 成立, 则称 n 个事件 $A_1, A_2, \ldots, A_n$ 是 **互不相容的** 或 **互斥的**, 类似地定义可列个互不相容的事件.
   对立的事件是互不相容的, 但互不相容的事件并不一定是对立事件.

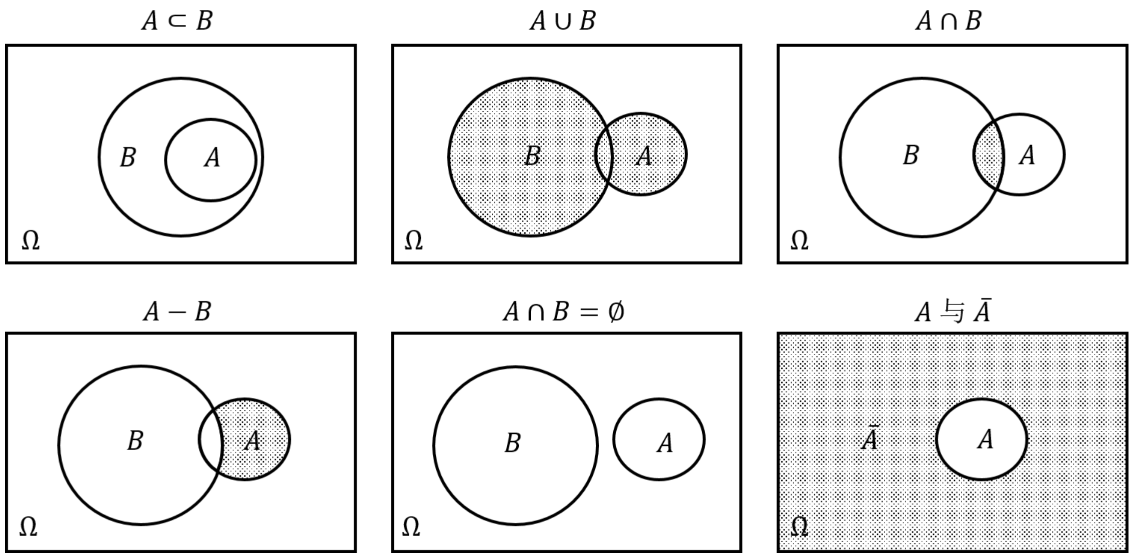


图 1.1 事件关系或运算通过韦恩图表示, $A \cup B$, $A \cap B$, $A - B$, $\bar{A}$ 分别为图中阴影部分所示

如图 1.1 所示, 借助集合论的韦恩图 (Venn Diagram) 可直观地表示事件之间的关系或运算. 例如, 在 $A \subset B$ 的图示中, 矩形表示样本空间 Ω, 椭圆 A 和 B 分别表示事件 A 和 B, 椭圆 B 包含椭圆 A 则表示事件 $A \subset B$; 在 $A \cup B$ 的图示中阴影部分表示并事件 $A \cup B$.

根据定义可知道事件还满足下面的规律, 相关证明读者可参考集合的运算规律.

- 交换律: $A \cup B = B \cup A$, $A \cap B = B \cap A$;
- 结合律: $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$;
- 分配律: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$;
- 德摩根 (De Morgen) 律: $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$.

上面的四条规律对有限个或可列个事件均成立, 例如对德摩根律有

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}, \qquad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}, \qquad \overline{\bigcup_{i=1}^{+\infty} A_i} = \bigcap_{i=1}^{+\infty} \overline{A_i}, \qquad \overline{\bigcap_{i=1}^{+\infty} A_i} = \bigcup_{i=1}^{+\infty} \overline{A_i}\ .$$

此外, 若事件 $A \subset B$, 有 $AB = A$ 和 $A \cup B = B$ 成立.

例 1.2 设 A, B, C 为三个随机事件, 则有

- 事件 A 与 B 同时发生, 而事件 C 不发生的事件可表示为 $AB\bar{C}$ 或 $AB - C$;
- 这三个事件中至少有一个发生的事件可表示为 $A \cup B \cup C$;
- 这三个事件中恰好有一个发生的事件可表示为 $(A\bar{B}\bar{C}) \cup (\bar{A}B\bar{C}) \cup (\bar{A}\bar{B}C)$;

- 这三个事件中至多有一个发生的事件可表示为 $(A\bar{B}\bar{C}) \cup (\bar{A}B\bar{C}) \cup (\bar{A}\bar{B}C) \cup (\bar{A}\bar{B}\bar{C})$ 或 $\overline{AB \cup AC \cup BC}$;
- 这三个事件中至少有两个发生的事件可表示为 $AB \cup AC \cup BC$;
- 这三个事件中至多有两个发生的事件可表示为 $\bar{A} \cup \bar{B} \cup \bar{C}$;
- 这三个事件中恰好有两个发生的事件可表示为 $AB\bar{C} \cup AC\bar{B} \cup BC\bar{A}$.

例 1.3 设 A, B, C 为三个随机事件, 证明

$$(\bar{A} \cup B)(A \cup B)(\bar{A} \cup \bar{B})(A \cup \bar{B}) = \emptyset \qquad \text{和} \qquad (A - B) \cup (B - C) = (A \cup B) - BC\ .$$

证明 根据事件的分配律有 $(\bar{A} \cup B)(A \cup B) = (A \cap \bar{A}) \cup B = B$ 以及 $(\bar{A} \cup \bar{B})(A \cup \bar{B}) = \bar{B}$, 由此可得 $(\bar{A} \cup B)(A \cup B)(\bar{A} \cup \bar{B})(A \cup \bar{B}) = B \cap \bar{B} = \emptyset$.

根据事件的差 $A - B = A\bar{B}$ 可得 $(A - B) \cup (B - C) = (A\bar{B}) \cup (B\bar{C})$. 根据事件的分配律和德摩根律有

$$\begin{aligned}(A \cup B) - BC &= (A \cup B)\overline{BC} = (A \cup B) \cap (\bar{B} \cup \bar{C}) \\ &= (A\bar{B}) \cup (A\bar{C}) \cup (B\bar{B}) \cup (B\bar{C}) = (A\bar{B}) \cup (A\bar{C}) \cup (B\bar{C}).\end{aligned}$$

由此可知 $((A - B) \cup (B - C)) \subset ((A \cup B) - BC)$. 只需进一步证明 $A\bar{C} \subset (A\bar{B}) \cup (B\bar{C})$, 对任意 $x \in A\bar{C}$, 有 $x \in A$ 且 $x \in \bar{C}$, 再根据 $x \in B$ 或 $x \in \bar{B}$ 有 $x \in A\bar{B}$ 或 $x \in B\bar{C}$ 成立.

事件的关系与运算可类比于集合的关系与运算, 表 1.1 简要地给出了概率论和集合论之间对应关系, 即概率统计中事件的关系与运算可通过集合的方式进行描述.

表 1.1 概率论与集合论之间相关概念的对应关系

符 号	概 率 论	集 合 论
Ω	必然事件, 样本空间	全集
$\emptyset$	不可能事件	空集
ω	基本事件	元素
A	随机事件	子集
$\bar{A}$	事件 A 的对立事件	集合 A 的补集
$\omega \in A$	事件 A 发生	元素 ω 属于集合 A
$A \subset B$	事件 A 发生导致 B 发生	集合 B 包含集合 A
$A = B$	事件 A 与 B 相等	集合 A 与 B 相等
$A \cup B$	事件 A 与 B 的并	集合 A 与 B 的并集
$A \cap B$	事件 A 与 B 的交	集合 A 与 B 的交集
$A - B$	事件 A 与 B 的差	集合 A 与 B 的差集
$AB = \emptyset$	事件 A 与 B 互不相容	集合 A 与 B 无相同元素

1.1.3 可测空间*

设 Ω 是一个样本空间, 用 2^Ω 表示样本空间 Ω 所有子集所构成的集合, 称为 Ω 的 **幂集**, 即样本空间 Ω 上所有事件所构成的集合. 对可列的样本空间, 将幂集 2^Ω 中的元素都看作事件没任何不妥; 但对无限不可列样本空间, 一般情形下不将样本空间 Ω 的一切子集都作为事件, 这将对概率的计算带来不可克服的困难. 例如, 在几何概型 (见 1.3.2节) 中将不可测集作为事件则难以计算概率. 为了更好地刻画随机事件, 本节引入可测空间.

定义 1.1 设 Ω 是一个样本空间且 $\Sigma \subseteq 2^\Omega$, 若 Σ 满足以下三个条件:

- 必然事件 $\Omega \in \Sigma$;
- 若任意 $A \in \Sigma$, 则有补集 $\bar{A} \in \Sigma$;
- 若任意 $A_i \in \Sigma$ $(i = 1, 2, \cdots)$, 则有 $\bigcup_{i=1}^{+\infty} A_i \in \Sigma$,

则称 Σ 是样本空间 Ω 的 σ **代数** (又称 σ **域**), 称 Σ 中的元素 (一个子集) 是 **可测集**, 以及 (Ω, Σ) 是一个 **可测空间**.

σ 代数 Σ 本质上是一个集合, 其每一个元素也是集合, 即 Σ 是 Ω 一些子集所构成的集合. 若 Σ 是一个 σ 代数, 则 Σ 中每个元素都是可测集. 根据可测空间定义可知

$$\emptyset \in \Sigma, \qquad \bigcup_{i=1}^{n} A_i \in \Sigma, \qquad \bigcap_{i=1}^{n} A_i \in \Sigma, \qquad \bigcap_{i=1}^{+\infty} A_i \in \Sigma.$$

给定样本空间 Ω, 最小的 σ 代数为 $\Sigma = \{\emptyset, \Omega\}$, 最大的 σ 代数为 $\Sigma = 2^\Omega$.

所关注的非空事件集合 $\mathcal{F} \subset 2^\Omega$ 有时不一定满足 σ 代数, 此时可以构造包含 $\mathcal{F}$ 的最小 σ 代数:

定义 1.2 给定样本空间 Ω 和非空事件集合 $\mathcal{F} \subset 2^\Omega$, 记 $\sigma(\mathcal{F})$ 为包含 $\mathcal{F}$ 的最小 σ 代数. 即若 Σ 是一个 σ 代数且 $\mathcal{F} \subset \Sigma$, 则有 $\sigma(\mathcal{F}) \subset \Sigma$.

例如当 $\mathcal{F} = \{A\}$ 时, 即集合 $\mathcal{F}$ 仅包含单一事件 A, 则最小 σ 代数为 $\sigma(\mathcal{F}) = \{\emptyset, A, \bar{A}, \Omega\}$. 对一般的事件集合 $\mathcal{F}$, 有最小 σ 代数 $\sigma(\mathcal{F}) = \bigcap_{\mathcal{F} \subset \Sigma} \Sigma$.

对有限或可列的样本空间 Ω, 一般考虑 σ 代数 $\Sigma = 2^\Omega$; 而当样本空间 Ω 时实数集 $\mathbb{R}$ 时, 一般考虑博雷尔 σ 代数, 即由有限或可列个开区间 (或闭区间) 构成的 σ 代数, 记为 $\mathfrak{R}_1$, 即

$$\begin{aligned}
\mathfrak{R}_1 &= \sigma\left(\{(a,b)\colon a < b \in \mathbb{R}\}\right) = \sigma\left(\{[a,b]\colon a < b \in \mathbb{R}\}\right) = \sigma\left(\{(-\infty,b), b \in \mathbb{R}\}\right) \\
&= \sigma\left(\{(-\infty,b], b \in \mathbb{R}\}\right) = \sigma\left(\{(a,+\infty), a \in \mathbb{R}\}\right) = \sigma\left(\{[a,+\infty), a \in \mathbb{R}\}\right).
\end{aligned}$$

上式成立的原因有 $[a,b] = \bigcap_{n=1}^{\infty}(a + 1/n, b - 1/n)$, $(a,b) = [a,b] \setminus a \setminus b$, 以及可列次的逆、并、交等运算. 类似可定义 n 维博雷尔 σ 代数 $\mathfrak{R}_n$.

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. 为此首先引入频率, 用以描述随机事件发生的频繁程度, 在此基础上引入事件的概率.

1.2.1 频率

定义 1.3 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**.

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. 根据上面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 发生的频数之和, 由此可知性质 3° 成立.

频率在实际中通常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.

尽管频率表现出一定的随机性, 但经过大量重复的试验, 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**. 例如历史上有多人做过重复投掷硬币的试验, 下表列出了其中一些试验统计结果:

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

我们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 给出了相应的试验结果. 这些研究结果均表明: 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常数 1/2, 即频率逐渐稳定于 1/2. 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 1.4 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 $P(A) = p$.

图 1.2 任意投掷硬币, 正面朝上频率的趋势

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.

另一方面, 概率的统计定义存在着数学上的不严谨性, 在实际中也不太可能每一个事件做大量重复试验来计算频率, 以此近似概率. 受到频率的稳定性及其性质的启发, 下面我们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, 建立可媲美于欧氏几何公理化的理论体系.

定义 1.5 (概率公理化) 在可测空间 (Ω, Σ) 上, 若函数 $P:\Sigma\to R$ 满足以下条件:

1° **非负性**: 对任意 $A\in\Sigma$ 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 $A_1, A_2, \ldots, A_n, \ldots\in\Sigma$ 是可列个互不相容的事件, 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**, 称 (Ω, Σ, P) 为 **概率测度空间** 或 **概率空间** (probability space).

概率 $P(A)$ 是定义在可测空间 (Ω, Σ) 上的实值函数, 满足非负性、规范性和可列可加性三条公理, 该定义简明扼要地刻画了概率的本质, 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.

根据概率公理化的定义, 可以推导出很多概率的性质.

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 $P(\emptyset)=0$.

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几何概型或连续随机变量.

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

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

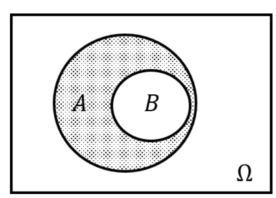
性质得证.

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



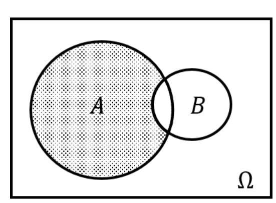
$$P(A)=P(B)+P(A-B).$$

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) 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

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

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得到下列不等式:

推论 1.1 (Bonferroni 不等式) 对事件 $A_1, A_2, \ldots A_n$ 有

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

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) $P(\bar{A}B)$; 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

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

对问题 3), 根据容斥原理有

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

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

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

$$= \quad 3/4 - 1/8 + P(ABC).$$

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 1.6 (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 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, 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 $\{\omega_{i_1}, \omega_{i_1}, \ldots, \omega_{i_k}\}$, 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.

计算古典概率的本质是计数 (Counting), 计数是组合学研究的重要内容, 我们将在下一节详细的介绍各种计数方法, 这里仅仅介绍一些基本原理和排列组合:

- **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.

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.