

第 1 章 随机事件与概率

概率论与数理统计研究的对象是随机现象。概率论是研究随机现象的模型 (即概率分布)，数理统计是研究随机现象的数据收集与处理。

在一定的条件下，并不总是出现相同结果的现象称为**随机现象**(Random phenomenon)；只存在一个结果的现象称为**确定性现象**(Deterministic phenomenon)；对在相同的条件下，以重复的随机现象的观察、记录和实验的研究称为**随机试验**(Random trial)。概率论与数理统计主要研究能大量重复的随机现象和随机试验，但也十分注意研究不能重复的随机现象。

概率 (Probability) 是随机现象的**不确定性**的一种度量，概率模型就是对随机现象的数学描述。

1.1 随机事件与概率

内容提要

- 样本空间与随机事件
- 联合样本空间
- 事件间的关系与运算
- 概率公理
- 序贯模型
- 使用概率建模的方法步骤

1.1.1 样本空间与随机事件

样本点是抽样 (Sample) 的最基本单元，认识随机现象首先要列出它的样本空间。

定义 1.1 (样本空间与样本点)

随机现象的一切可能基本结果组成的集合称为样本空间 (Sample space)，记为 $\Omega = \{\omega\}$ ，其中 ω 表示基本结果，称为样本点 (Sample point)。

将样本点的个数为有限个或可列的情况归为一类，称为**离散样本空间**(Discrete sample space)；而将样本点的个数为不可列无限个的情况归为另一类，称为**连续样本空间**(Continuous sample space)。

例题 1.1 在研究英文字母的使用情况时，则可能出现的结果为 A, B, \dots, Z 和空格 “[]”，将样本空间选定为 $\Omega = \{\text{空格}, A, B, \dots, Z\}$ 是合适的，它是一个离散的样本空间；

例题 1.2 考察一晶体管的寿命，用变量 t 来表示晶体管开始工作到首次发生故障的时间，那么“晶体管寿命”的样本空间就是所有的非负实数 $\Omega = \{t \mid t \geq 0\}$ ，它是一个连续的样本空间。

在确定样本空间的时候，不同的试验结果必须是**相互排斥的**，即在试验过程中只能产生**唯一**的一个结果。需要注意的是随机现象的样本空间要包含**至少两个样本点**，若将确定性现象也一并考虑，那么仅含有一个样本点的样本空间对应的确定性现象。

定义 1.2 (随机事件)

样本空间的一个子集称作随机事件，简称事件 (Event)，通常用大写的英文字母 A, B, C, \dots 表示。

例如掷骰子所出现点数的样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$ ，可以定义集合 $\{1, 3, 5\}$ 为事件“掷到奇数点”，定义集合 $\{4, 5, 6\}$ 为事件“掷到的点数大于 3”……

当子集 A 中某个样本点出现了，就说事件 A 发生了，或者说：

命题 1.1

事件 A 发生当且仅当 A 中某个样本点出现了。

由样本空间 Ω 中的**单个元素**组成的子集称为**基本事件**(Elementary event)，而样本空间 Ω 的最大子集 (即 Ω 本身) 称为**必然事件**(Certain event)，样本空间 Ω 的最小子集 (即空集 \emptyset) 称为**不可能事件**(Impossible event)。

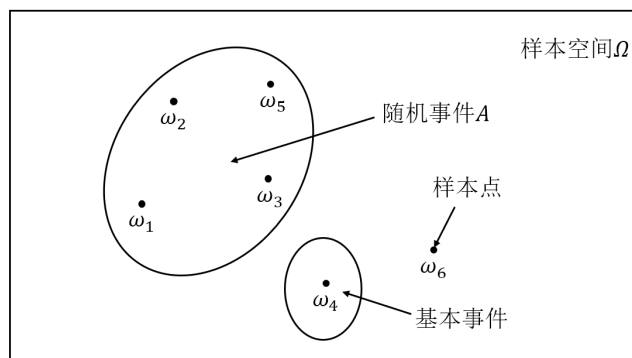


图 1.1: 样本空间、样本点与随机事件

1.1.2 联合样本空间

同一个随机现象，由于研究的“侧重点”不同，通常会定义不同的样本空间，这些“侧重点”就是我们研究随机现象所包含的**指标**(Indicator)、**维度**(Dimension) 或者**因子**(Factor)，**不同维度 (指标或者因子) 本质上描述还是同一个随机现象**。

比如要研究“某城市的交通事故”，如果研究的侧重点在于“交通事故的严重程度”，那么样本空间就为 $\Omega_1 = \{\text{轻微碰撞, 重度碰撞, 严重碰撞}\}$ ；如果研究的侧重点在于“一年中发生交通事故的次数”，那么样本空间就为 $\Omega_2 = \{0, 1, 2, \dots\}$ ；比如研究的侧重点在于“交通事故发生的时间段”，那么样本空间就为 $\Omega_3 = \{\text{凌晨, 早高峰, 中午, 晚高峰, 夜间}\} \dots$ 。但是一旦明确了研究的“侧重点”(维度)，则样本空间就被确定下来，此时样本空间是唯一且固定的。这意味着不能在同一个维度中使用多个样本空间，否则会导致分析上的混乱。

当然，也可以基于不同的研究维度所定义的样本空间，构建**联合样本空间**，或者称为**重构样本空间**，即将一个较大的、复杂的样本空间细化为更详细的子集来捕捉更具体的信息。比如要同时研究交通事故的严重程度和时间段，此时的样本空间则为：

$$\Omega = \Omega_1 \times \Omega_3 = \{(\text{轻微碰撞, 凌晨}), (\text{严重碰撞, 早高峰}), \dots\}$$

其中每个样本点是一个**二元组**，包含了交通事故的严重程度和时间段； $\Omega_1 \times \Omega_3$ 为两个集合的**Descartes 集**。

重构样本空间是对样本空间的不同层次的描述，但这并没有改变样本空间的唯一性，只是对它进行了更细致的表述，以适应不同的分析需要。选择样本空间时，重要的是要确保它**包含所有可能的结果**。同一个随机现象，不同样本空间的选择往往意味着对问题的不同维度的不同描述。一旦确定了样本空间，所有事件和概率计算必须基于该样本空间进行，不能在同一维度中混用多个样本空间来描述同一个随机现象。

1.1.3 事件间的关系与运算

下面的讨论总是假设事件 A 与 B 在**同一个样本空间 Ω** 中进行 (即同一个随机现象)，一般使用 Venn 图来表示事件间的关系与运算。事件间的关系与集合间的关系一样，主要有以下几种：

1. 包含关系：如果属于 A 的样本点必属于 B ，则称 A 被包含在 B 中，或称 B 包含 A ，记为 $A \subset B$ ，或 $B \supset A$ 。用概率论的语言说：事件 A 发生必然导致事件 B 发生。对于任一事件 A ，必有 $\emptyset \subset A \subset \Omega$ ；
2. 相等关系：如果事件 A 与事件 B 满足：属于 A 的样本点必属于 B ，而且属于 B 的样本点必属于 A ，即 $A \subset B$ 且 $B \supset A$ ，则称事件 A 与 B 相等，记为 $A = B$ 。从集合论观点看，两个事件相等就意味着这两事件是**同一个集合**；
3. 互不相容关系：如果 A 与 B **没有相同的样本点**，则称 A 与 B 互不相容。用概率论的语言说： A 与 B 互不相容就是**事件 A 与事件 B 不可能同时发生**。

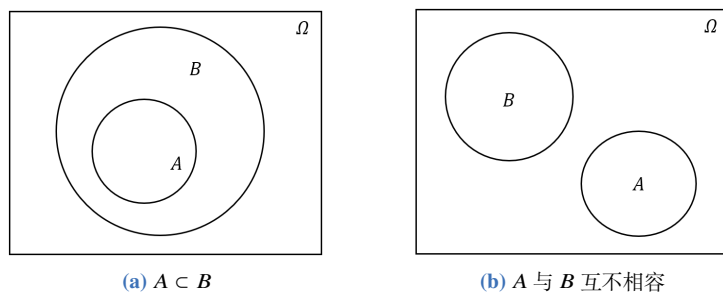


图 1.2: 事件间的关系

事件的运算与集合的运算相当，有并、交、差和余四种运算。

1. 事件 A 与 B 的差：记为 $A - B$ ，也可以记作 $A\bar{B}$ 。其含义为“由在事件 A 中而不在 B 中的样本点组成的新事件”，或用概率论的语言说“事件 A 发生而 B 不发生”；

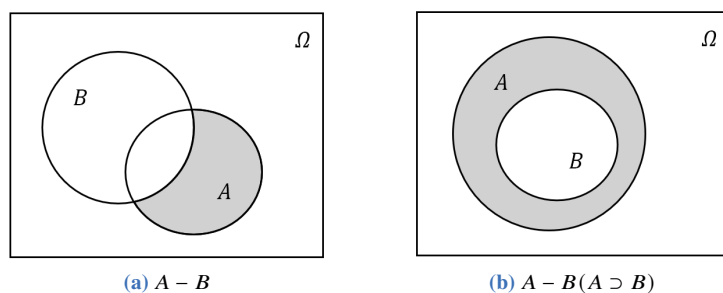


图 1.3: 事件的差集运算

2. 事件 A 与 B 的并：记为 $A \cup B$ ，也可以记作 $A + B$ 。其含义为“由事件 A 与 B 中所有的样本点 (相同的只计一次) 组成的新事件”，或用概率论的语言说“事件 A 与 B 中至少有一个发生”；
3. 事件 A 与 B 的交：记为 $A \cap B$ ，或简记为 AB 。其含义为“由事件 A 与 B 中公共的样本点组成的新事件”，或用概率论的语言说“事件 A 与 B 同时发生”；
4. 对立事件 (余集)：将事件 A 的对立事件记为 \bar{A} ，即“由在 Ω 中而不在 A 中的样本点组成的新事件”，或用概率论的语言说“ A 不发生”，即 $\bar{A} = \Omega - A$ 。

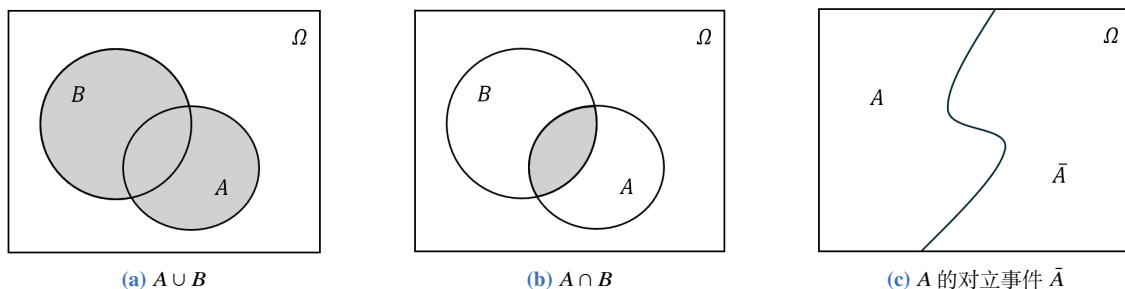


图 1.4: 事件的交、并与余集运算

注意，对立事件是相互的，即 A 的对立事件是 \bar{A} ，而 \bar{A} 的对立事件是 A ，即 $\bar{\bar{A}} = A$ 。必然事件 Ω 与不可能事件 \emptyset 互为对立事件，即 $\bar{\Omega} = \emptyset$ ， $\bar{\emptyset} = \Omega$ 。

命题 1.2

A 与 B 互为对立事件的充分必要条件是： $A \cap B = \emptyset$ ，并且 $A \cup B = \Omega$ 。

命题(1.2)也可作为对立事件的另一种定义，即如果事件 A 与 B 满足： $A \cap B = \emptyset$ ，且 $A \cup B = \Omega$ ，则称 A 与

B 互为对立事件, 记为 $\bar{A} = B$, $\bar{B} = A$ 。

 **笔记** 对立事件一定是互不相容事件, 即 $A \cap \bar{A} = \emptyset$, 但是互不相容事件不一定是对立事件。

1.1.4 事件运算的性质

设事件 A, B, C 属于同一个样本空间 Ω , 那么事件 A, B, C 满足以下性质:

1. 交换律:

$$A \cup B = B \cup A, \quad AB = BA \quad (1.1)$$

2. 结合律:

$$A \cup (B \cap C) = (A \cup B) \cap C, \quad A(BC) = (AB)C \quad (1.2)$$

3. 分配律:

$$(A \cup B) \cap C = AC \cup BC, \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C) \quad (1.3)$$

4. De Morgan 公式:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B} \quad (1.4)$$

De Morgan 公式可推广至 n 个事件以及可列个事件的情况:

$$\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i, \quad \overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i \quad (1.5)$$

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \bar{A}_i, \quad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \bar{A}_i \quad (1.6)$$

1.1.5 概率的公理化

概率的**公理化定义**(Axiomatic definition of probability) 是概率论发展的一个重要里程碑, 它奠定了**现代概率论**的基础, 彻底改变了过去基于经验和直觉的概率观念。

18 世纪早期, Jacob Bernoulli 在古典概率论中, 将概率定义为

$$P(A) = \frac{\text{有利结果数}}{\text{总结果数}} \quad (1.7)$$

尽管这一定义在简单场景中有效, 但它依赖于**等可能性**的假设, 在更多复杂场景下无法推广。这一时期的概率论主要围绕着**有限样本空间的等可能事件**展开。

18 世纪晚期至 19 世纪, 频率学派的概率观开始取代古典定义。Richard von Mises 等人将概率定义为某一事件在大量重复试验中发生的**相对频率**, 即

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad (1.8)$$

其中, n_A 是事件 A 发生的次数, n 是试验的总次数。这一方法突破了古典定义在非等可能情境中的局限性, 但也面临一些哲学上的争议, 特别是在如何定义“**无限次重复**”时遇到了困难。

与频率学派同时期, Bayes 提出了基于先验知识更新概率的概念, 即贝叶斯定理。贝叶斯学派强调概率是对某一随机现象的**主观信念程度**, 主张通过**已知证据**(即已经发生的事件的信息) 来更新这一信念:

$$P(A | B) = \frac{P(A)P(B | A)}{P(B)} \quad (1.9)$$

Bayes 学派认为概率不一定源自重复试验, 而是基于观察归纳和对先验知识的修正。

1900 年 Hilbert 提出要建立概率的公理化定义以解决这个问题, 即以**最少的几条本质特性出发去刻画概率的概念**。1933 年 Kolmogorov 首次提出了**基于集合论和测度论**的公理化概率定义, 标志着概率论进入现代阶段。这个定义既概括了历史上几种概率定义中的共同特性, 又避免了各自的局限性和含混之处。**不管什么随机现象, 只有满足该定义中的三条公理, 才能说它是概率。**

Kolmogorov 的公理化体系使用了集合论的语言，将概率视为对一个集合（样本空间）的子集赋予的**测度**。他提出的三条公理已经形成了现代概率论的基础。

但在这之前，我们必须首先明确哪些集合可以被称为“事件”，并且构造一个包含在样本空间的“事件集合”。直觉上我们所构造的集合应该对事件的各种运算（即余集、交集和并集）封闭，并且样本空间 Ω 本身和它的余集 \emptyset 都应该在这个“事件集合”中。

定义 1.3 (事件域)

设 Ω 为一个样本空间， \mathcal{F} 是 Ω 的子集所组成的集合类。若 \mathcal{F} 满足：

1. $\Omega \in \mathcal{F}$;
2. 若 $A \in \mathcal{F}$ ，则对立事件 $\bar{A} \in \mathcal{F}$;
3. 若 $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$ ，则可列并 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

则称 \mathcal{F} 是一个事件域 (Event field)，或者 σ -域 或者 σ -代数。



在概率论中，称 (Ω, \mathcal{F}) 为**可测空间**(Measurable space)，在**可测空间上才能定义概率**，这时 \mathcal{F} 上都是有概率可言的事件。

σ -代数用于定义哪些集合可以称为“事件”，并确保可以在这些事件上合理地定义和计算概率。通过 σ -代数，我们能对各种复杂事件进行**可数次或者有限次运算**(余集、交集和并集)，并确保这些操作后的结果依然是事件，即仍然在 σ -代数中。

笔记

1. 在概率论中， σ -代数指的就是那些可以定义概率的事件的集合（并且满足余集和可列并封闭）。至于在其他学科中，有其他学科所赋予的**内涵**；
2. 由条件 (2) 可知， $\emptyset \in \mathcal{F}$ ，并且 $\bigcup_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \bar{A}_i \in \mathcal{F}$ 。

一维实数轴 \mathbb{R} 上的**Borel 集**是常见的 σ -代数，它是由一切有界开区间 (a, b) ，通过至多可数次的（有限或者可数）余集运算、交集运算和并集运算得到的集合。

在实际应用中， σ -代数提供了一个框架，它被用于定义一个复杂事件系统，使其可以进行概率分析，这允许我们根据不同的应用场景定义哪些事件是可以赋予概率的。通常一个较大的集合 Ω 可能包含无穷多个子集，而只有 σ -代数中的事件才是我们能够赋予概率的“可测事件”。不仅如此， σ -代数还为事件提供了一个良好的数学结构，还确保了我们对于事件操作的一致性。由于 σ -代数对余集和可数并封闭，所以在定义事件的概率时，无论我们是直接计算事件的概率，还是通过余集或并集的方式来间接计算概率，都能够得到一致的结果。

例题 1.3 假设每周末你和家人会一起看电影，选择电影时大家有不同的偏好。你们要从一个电影库中选择一部合适的电影，而每部电影可以按照几个维度进行分类，例如：

1. 电影类型：动作片、喜剧片、恐怖片、科幻片等。
2. 电影时长：短片（小于 1 小时）、中片（1-2 小时）、长片（大于 2 小时）。
3. 语言：中文、英语、西班牙语、其他语言。

这些条件构成了一个电影的“样本空间”，并且这些维度构成了更多的分类，从而形成更复杂的样本空间。这样具体的样本空间就为

$$\Omega = \{(\text{动作片}, \text{长片}, \text{英语}), (\text{喜剧片}, \text{中片}, \text{中文}), \dots\}$$

下面来定义具体的“事件”。“事件”在这个场景下指的是你和家人一起选择的某些符合条件的电影，例如：定义事件 A 为“选择动作片或者喜剧片”，即 $A = \{\text{动作电影}, \text{喜剧电影}\}$ ；定义事件 B 为“选择英语电影”，即 $B = \{\text{英语电影}\}$ 。

现在我们通过 σ -代数的封闭性操作，构造更多的“电影选择事件”。 σ -代数对余集、并集、交集封闭，这使得我们可以生成更复杂的选择组合。例如不想看动作片或者喜剧片，那么

$$\bar{A} = \Omega \setminus A = \{\text{爱情片}, \text{恐怖片}, \dots\}$$

如果想选择“动作片或英语电影”，那么

$$A \cup B = \{\text{动作电影, 英语电影}\}$$

如果只想选择“英语的动作片或英语的喜剧片”，那么

$$A \cap B = \{\text{英语动作电影, 英语喜剧电影}\}$$

所有这些可能的“电影选择”方式，经过 σ -代数的构造操作，通过余集、并集和交集操作，可以生成任意复杂的电影选择规则，就构成了一个满足 σ -代数性质的系统，这样的组合就是 σ -代数的封闭性体现。

在实际中，那些无法精确测量、描述过于复杂或超出我们观察能力的事件，通常不属于 σ -代数，因此我们无法为它们赋予概率。这些事件往往涉及无限小的区间、无穷复杂的现象或超精确的测量，在实际概率分析中无法处理，属于不可测事件。例如一个颜色的 RGB 值精确为某个特定数值、在无穷多次抛硬币结果中，硬币的正反面组成某个非常特定的无穷序列。

定义 1.4 (概率的公理化定义)


设 Ω 为一个样本空间， \mathcal{F} 为 Ω 上的事件域， $P(A)$ 为定义在 \mathcal{F} 上的实值函数。若 $P(A)$ 满足：

1. 非负性： $\forall A \in \mathcal{F}, P(A) \geq 0$;
2. 正则性： $P(\Omega) = 1$;
3. 可列可加性： $\forall A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$ ，且 $A_1, A_2, \dots, A_n, \dots$ 两两互不相容，则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (1.10)$$

则称 $P(A)$ 为事件 A 的概率 (Probability)。

称 (Ω, \mathcal{F}, P) 为**概率空间**(Probability space)，它为研究随机现象提供了一个严谨的数学框架，使我们能够合理地处理、计算和分析各种随机事件的概率。无论是处理简单的随机事件，还是处理复杂的随机变量，都可以在这个框架下进行。

 **笔记** 从定义可以看出，概率 P 是事件 A 的实函数，并且概率 $P(A)$ 是有界的，即 $0 \leq P(A) \leq 1$ 。

这一公理化体系与**Lebesgue 测度**相结合，使得概率论的基础变得更加坚实，能够处理复杂的事件和连续概率空间问题。Kolmogorov 的公理化定义克服了以往各个学派的局限性，使得概率论从经验和直觉的框架走向了严谨的数学体系。

命题 1.3 (样本空间的分割)

在样本空间 Ω 中，若有 n 个事件 D_1, D_2, \dots, D_n ，满足：

1. D_1, D_2, \dots, D_n 两两互不相容；
2. $\bigcup_{i=1}^n D_i = \Omega$ 。

则称 D_1, D_2, \dots, D_n 为样本空间 Ω 的一个分割 (Division)。

可列个互不相容的事件 $D_1, D_2, \dots, D_n, \dots$ 若满足 $\bigcup_{i=1}^{\infty} D_i = \Omega$ ，那么 $D_1, D_2, \dots, D_n, \dots$ 也可以成为样本空间 Ω 的一个分割。

1.1.6 序贯模型

序贯模型(Sequential Models) 通常指一类用于处理顺序数据或在数据采集过程中逐步更新模型的统计方法。其核心思想是基于当前数据点以及之前的观测数据，实时或分阶段地更新估计、推断或决策。

序贯模型的应用包括以下几个方面：

1. 对于涉及时间序列、文本序列等具有时间或顺序依赖的数据，序贯模型用来捕捉这些数据中前后顺序之间的依赖关系。例如时间序列分析、隐马尔可夫模型 (Hidden Markov Model) 和循环神经网络 (Recurrent neural network)；

2. 在贝叶斯框架中，序贯模型指在每个时间点或阶段逐步更新模型的参数或预测分布。即随着新数据的到来，不断使用新的证据修正先验信息形成后验，从而改进对未知参数的估计。这种更新是动态的，可以处理逐步到来的数据；
3. 在统计实验设计中，序贯模型也指根据实验过程中逐步获取的数据来动态调整实验的设计或数据采集过程。这种方法与传统的固定样本大小不同，实验者可以根据中间结果调整样本规模或停止实验。例如序贯假设检验在逐步收集样本的过程中，实时决定是否拒绝或接受假设。

许多试验本身具有序贯的特征。例如连续抛掷一枚硬币连续抛三次，或者连续观察一只股票 5 天，又或者在一个通信接收设备上接收 8 位数字。常用**序贯树形图**来刻画样本空间中的试验结果。

图(1.5)为连续两次抛掷有 4 个面的正四面体，其样本空间有两种等价的刻画方法。在这个试验中，可能的结果是全体有序对 (i, j) ，其中 i 表示第一次抛掷骰子得到的数字， j 表示第二次抛掷骰子得到的数字。试验结果可用图(1.5a)的二维格点表示，也可以用图(1.5b)的序贯树形图表示，后者的优点是可以表示试验的序贯特征 (Sequential feature)。在序贯树形图中，每个可能的试验结果可以用一个末端的树叶表示，或等价地用与树叶相联系的由根部到树叶的一个路径表示。左图中阴影部分代表事件 $\{(1, 4), (2, 4), (3, 4), (4, 4)\}$ ，它表示第 2 次抛掷得到 4。同一个事件可以在右图中用圆点标示的叶子集合表示，注意序贯树形图中的每一个结点可以代表一个事件，这个事件就是由这个结点出发的所有的叶子构成的事件。

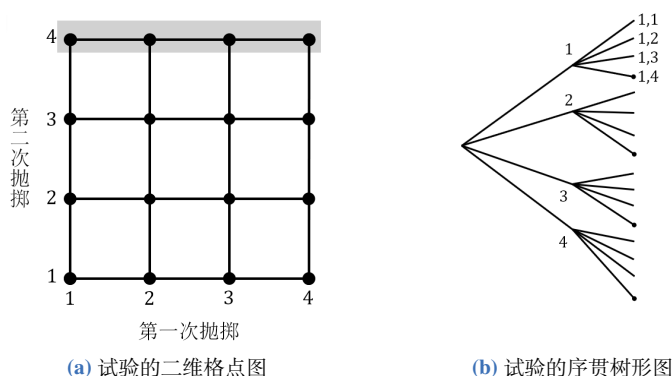


图 1.5: 试验的序贯表示

1.1.7 概率建模的步骤

概率论可以用来分析现实世界的许多不确定现象，这个过程通常分成两个阶段：

1. 从一个**适当的样本空间**中给出概率分布，从而建立概率模型。在这个阶段没有关于建立模型的一般规则，只要建立的概率分布符合概率公理化即可。可能有些人会怀疑所建立模型的真实性的，有时人们宁愿使用“错误”的模型，其理由是“错误”的模型比“正确”的模型简单且易于处理，这种处理问题的态度在科学和工程学中很普遍。因此在实际工作中，选择的模型往往既要**准确、简单**又要兼顾**易操作性**。此外统计学家还依据历史数据和过去相似试验的结果，利用统计方法确定模型，比如贝叶斯统计；
2. 我们将在完全严格的概率模型之下进行推导，计算某些事件的概率或推导出一些十分有趣的性质，(1) 的任务是**建立现实世界与数学的联系**，而 (2) 则是严格限制在**概率公理之下的逻辑推理**。如果涉及的计算很复杂或概率分布的陈述不简明，推理和理解就会遇到困难，但是所有的问题将会有有一个准确的答案不会产生歧义，只要有足够高的能力，所有的困难都将迎刃而解。

1.2 概率的性质

内容提要

- 概率的性质
- 事件的差的概率

- 加法公式

1.2.1 概率的性质

设事件 A, B 属于同一个样本空间 Ω , 由概率的公理化定义容易推出概率的以下性质:

1. 不可能事件 \emptyset 的概率为 0: 即 $P(\emptyset) = 0$;

证明 由于

$$\Omega = \Omega \cup \emptyset \cup \cdots \cup \emptyset \cup \cdots$$

由概率的可列可加性,

$$P(\Omega) = P(\Omega) + P(\emptyset) + \cdots + P(\emptyset) + \cdots$$

由于 $P(\Omega) = 1$, 因此

$$P(\emptyset) + \cdots + P(\emptyset) + \cdots = 0$$

又由于概率的非负性, 于是 $P(\emptyset) = 0$. □

2. 有限可加性: 若有 n 个两两互不相容的事件 A_1, A_2, \cdots, A_n , 则

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (1.11)$$

证明 由于

$$A_1 \cup A_2 \cup \cdots \cup A_n = A_1 \cup A_2 \cup \cdots \cup A_n \cup \emptyset \cup \cdots \cup \emptyset \cup \cdots$$

因此由可列可加性, 有

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1) + P(A_2) + \cdots + P(A_n) + P(\emptyset) + \cdots + P(\emptyset) + \cdots \\ &= P(A_1) + P(A_2) + \cdots + P(A_n) \end{aligned}$$

□

3. 对立事件的概率: 对任意事件 A , 其对立事件 \bar{A} 的概率为

$$P(\bar{A}) = 1 - P(A) \quad (1.12)$$

证明 根据对立事件的定义,

$$\Omega = A \cup \bar{A} \quad (\text{或者 } \bar{A} = \Omega - A)$$

因此 $P(\Omega) = P(A) + P(\bar{A}) = 1$, 即

$$P(\bar{A}) = 1 - P(A)$$

□

例题 1.4 抛一枚硬币 5 次, 求既出现正面又出现反面的概率。

解 设事件 A 为抛一枚硬币 5 次既出现正面又出现反面, 可以考虑事件 A 的对立事件: 抛一枚硬币 5 次都出现正面 (记为事件 B) 或者都出现反面 (记为事件 C), 这样考虑的问题就极大地被化简了。这时 $\bar{A} = B \cup C$, 并且 B 与 C 是互不相容事件, 容易计算出

$$P(B) = P(C) = \frac{1}{2^5}$$

因此

$$P(A) = 1 - P(\bar{A}) = 1 - P(B \cup C) = 1 - P(B) - P(C) = \frac{15}{16}$$

1.2.2 事件的差的概率

命题 1.4

对于任意两个事件 A 与 B ，其差 $A - B$ 的概率为

$$P(A - B) = P(A) - P(AB) \quad (1.13)$$

证明 由事件的结合律，有

$$A = A \cup (B \cap \bar{B}) = AB \cup A\bar{B} = AB \cup (A - B)$$

又由于 AB 与 $A - B$ 互不相容，那么

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

即 $P(A - B) = P(A) - P(AB)$ ，结论得证。□

笔记 事件 $A - B$ 也可以写作 $A\bar{B}$ ， $P(A - B)$ 实际上计算的是 $P(A\bar{B})$ 的值。

推论 1.1

对于任意两个事件 A 与 B ，若 $B \subset A$ ，则其差 $A - B$ 的概率为

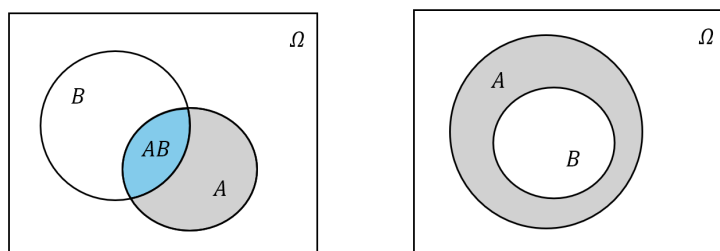
$$P(A - B) = P(A) - P(B) \quad (1.14)$$

证明 由于 $B \subset A$ ，因此 $AB = B$ ，于是

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

□

从几何直观上看，求事件 $A - B$ (图中灰色部分) 的概率只需用事件 A 的“面积”减去事件 AB 的“面积”：



(a) 任意的事件 A 与 B

(b) $B \subset A$

图 1.6: 事件 $A - B$ 的概率的 Venn 图示意

例题 1.5 某城有 n 辆卡车，车牌号从 1 到 n 。有一个外地人到该城去，把遇到的 m 辆车子的牌号抄下 (可能重复抄到某些车牌号)，求抄到的最大号码正好为 k 的概率 ($1 \leq k \leq n$)。

解 设事件 A_i 为遇到的 m 辆车子的最大牌号为 k 。若直接考虑计算 A_k 的概率会比较复杂，因此考虑“遇到的 m 辆车子的最大牌号小于等于 k ”，将其设为事件 B_k ，那么 A_k 则可以表示为 $A_k = B_k - B_{k-1}$ ，并且 $B_{k-1} \subset B_k$ 。事件 B_k 比起 A_k 的概率要好求得得多。根据古典概型可求得事件 B_k 的概率为

$$P(B_k) = \frac{k^m}{n^m}, \quad k = 1, 2, \dots, n$$

从而

$$P(A_k) = P(B_k - B_{k-1}) = P(B_k) - P(B_k B_{k-1}) = P(B_k) - P(B_{k-1}) = \frac{k^m - (k-1)^m}{n^m}$$

例题 1.6 在日常生活中, 我们经常会遇到诈骗电话的骚扰。设事件 A 为接到陌生电话, 并且 $P(A) = 0.9$; 再设事件 B 为接到诈骗电话, 那么事件 AB 就表示为接到陌生电话为诈骗电话, 根据一些专业机构的调查可估计出 $P(AB) = 0.7$, 于是接到陌生电话不是诈骗电话 (记为事件 $A - B$) 的概率则为

$$P(A - B) = P(A) - P(AB) = 0.2$$

即接到陌生电话不是诈骗电话的概率为 0.2, 也就是说接到的陌生电话中只有 20% 的比例不是诈骗电话。

推论 1.2 (概率的单调性)

对于任意两个事件 A 与 B , 若 $B \subset A$, 则 $P(B) \leq P(A)$ 。

笔记 推论(1.2)的逆命题不成立, 即若 $P(B) \leq P(A)$ 无法推出 $B \subset A$ 。

1.2.3 加法公式

定理 1.1

对于任意的两个事件 A 与 B , 其交事件 $A \cup B$ (即事件 A 或 B 至少发生一个) 的概率为

$$P(A \cup B) = P(A) + P(B) - P(AB) \quad (1.15)$$

证明 由于

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

又由于 $A \cap (B - AB) = \emptyset$, 即 A 与 $B - AB$ 互不相容, 则由有限可加性, 以及 $AB \subset B$, 有

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

□

从几何直观上看, 求事件 $A \cup B$ 的概率就是把事件 A 的“面积”加事件 B 的“面积, 但是由于多加了一次事件 AB (图中蓝色部分) 的面积, 因此必须再额外减去:

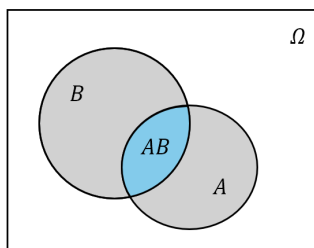


图 1.7: 事件 $A \cup B$ 的概率的 Venn 图示意

例题 1.7 一个人向 A 公司投递简历的概率为 0.6, 记作事件 A ; 向 B 公司投递简历的概率为 0.2, 记作事件 B 。那么事件 AB 则为既向 A 公司投递简历也向 B 公司投递简历, 其概率为 0.1, 那么这个人至少向其中一家投递简历的概率为

$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.7$$

推论 1.3

对于 n 个任意事件 A_1, A_2, \dots, A_n , 其交事件的概率为 $\bigcup_{i=1}^n A_i$ 的概率为

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + (-1)^{n-1} P(A_1 A_2 \cdots A_n) \quad (1.16)$$

其中一共包含 $\binom{n}{r}$ 个和项。

推论 1.4 (半可加性)

对于任意两个事件 A 与 B , 有

$$P(A \cup B) \leq P(A) + P(B) \quad (1.17)$$

对于任意 n 个事件 A_1, A_2, \dots, A_n , 有

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad (1.18)$$



1.3 确定概率的方法

内容提要

□ 频率的稳定性

□ 主观概率

□ 古典概型

□ 几何概型

□ Monte Carlo 方法

□ Bertrand 悖论

1.3.1 频率的稳定性

确定概率的频率方法是在**大量重复试验**中, 用频率的稳定值去获得概率的一种方法, 其基本思想是:

1. 与考察事件 A 有关的随机现象可大量重复进行。
2. 在 n 次重复试验中, 记 $n(A)$ 为事件 A 出现的次数, 又称 $n(A)$ 为事件 A 的**频数**, 称

$$f_n(A) = \frac{n(A)}{n} \quad (1.19)$$

为事件 A 出现的**频率**。

3. 人们的长期实践表明: 随着试验重复次数 n 的增加, 频率 $f_n(A)$ 会稳定在某一常数 a 附近, 我们称这个常数为频率的稳定值, 这个频率的稳定值就是我们所求的概率, 这个性质被称为**频率的稳定性**。因此就可以用事件 A 发生的频率来近似事件 A 发生的概率:

$$P(A) = \lim_{n \rightarrow \infty} f_n(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n} \quad (1.20)$$

容易验证: 由(1.20)式定义的“概率”满足非负性与正则性; 当事件 A 与 B 不相容时, $n(A \cup B) = n(A) + n(B)$, 因此

$$P(A \cup B) = \lim_{n \rightarrow \infty} \frac{n(A \cup B)}{n} = \lim_{n \rightarrow \infty} \frac{n(A)}{n} + \lim_{n \rightarrow \infty} \frac{n(B)}{n} = P(A) + P(B)$$

因此(1.20)式定义的“概率”的确为可测概率。

随机现象有其偶然性的一面, 也有其必然性的一面。这种必然性表现为大量试验中随机事件出现的频率的稳定性, 即一个随机事件出现的频率常在某个固定的常数附近摆动, 这种规律性我们称之为**统计规律性**(Statistical regularity)。频率的稳定性说明随机事件发生的可能性大小是随机事件本身固有的、不随人们意志而改变的一种客观属性, 因此可以对它进行度量。

例题 1.8 有人曾对 1930 年至 1988 年世界各地的 53274 场重大足球比赛作了统计: 在判罚的 15382 个点球中有 11172 个命中, 由此可得罚球点球命中率的估计值为

$$P(\text{点球命中}) = \frac{\text{点球命中的次数}}{\text{判罚点球的次数}} = \frac{11172}{15382} \approx 0.726$$

1.3.2 确定概率的主观方法

在现实世界里有一些随机现象是不能重复的或不能大量重复的, 统计界的贝叶斯学派认为: 一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信念, 这样给出的概率称为主观概率。**主观给定的概率**

要符合概率的公理化定义。

例如有一个学者以 90% 的把握断言《伊里亚特》和《奥德赛》是由同一作者创作的，所谓概率为 90% 的把握是学者的**主观信念**；再比如周五晚上的热门电影，你可能会觉得电影票已经卖完的概率是 60%，这是判断基于你对这个时间段和电影受欢迎程度的理解。

1.3.3 古典概型

古典概型是最早且最直观的一种概率模型，它适用于那些有限个且等可能事件的情境。所谓“古典”，意思是它起源于概率论的早期发展阶段，其基本思想如下：

1. 所涉及的随机现象只有**有限个样本点**；
2. 每个样本点发生的**可能性相等**(称为等可能性)；
3. 若事件 A 含有 k 个样本点，则事件 A 的概率为

$$P(A) = \frac{\text{事件}A\text{所含样本点的个数}}{\Omega\text{中所有样本点的个数}} = \frac{k}{n} \quad (1.21)$$

下面来解释一下基本概型的基本思想的内涵，设 $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ ，依据基本思想 (2)，有

$$P(\omega_1) = P(\omega_2) = \dots = P(\omega_n) = \frac{1}{n} \quad (1.22)$$

对于任意事件 A ，它总可以表示为样本点之交，即 $A = \omega_{i_1} \cup \omega_{i_2} \cup \dots \cup \omega_{i_m}$ ，其中 $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_m}$ 两两互不相容，于是

$$P(A) = P(\omega_{i_1}) + P(\omega_{i_2}) + \dots + P(\omega_{i_m}) = \frac{m}{n}$$

所以在古典概型中，事件 A 的概率的分母是样本点的总数 n ，而分子是事件 A 中所包含的样本点的个数。由于 $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_m}$ 的出现必导致 A 的出现，即它们的出现对 A 的出现“有利”，因此习惯上常称 $\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_m}$ 是 A 的“有利场合”。

容易验证：(1.21)式所定义的“概率”满足非负性与正则性，上面也证明了(1.21)式的可加性，因此(1.21)式所定义的“概率”的确为可测概率。

在古典概型中，求事件 A 的概率归结为计算 A 中含有的样本点的个数和中含有的样本点的总数，所以在计算中经常用到排列组合工具。

例题 1.9 考虑不放回抽样模型，一批产品共有 N 件，其中 M 件是不合格品， $N - M$ 件是合格品。从中随机取出 n 件 ($n \leq N$)，试求事件 A_m = “取出的 n 件产品中有 m 件不合格品” 的概率 ($m \leq M, n - m \leq N - M$)。

解 我们先计算样本空间 Ω 中样本点的个数，从 N 件中抽取 n 件产品，由于不限抽取次数，因此样本点的个数就为 $\binom{N}{n}$ 。又由于是随机抽取的，因此每一件产品抽取到的可能性都是相等的。

接下来先计算事件 A_0, A_1 的概率，然后再求出 A_m 的概率。事件 A_0 为取出的 n 件产品中没有任何不合格品，也就是 n 件产品全部都是从 $N - M$ 件产品中抽取的，于是

$$P(A_0) = \frac{\binom{N-M}{n}}{\binom{N}{n}} \quad (1.23)$$

事件 A_1 为取出的 n 件产品中仅有 1 件不合格品，也就是说要先从 M 件不合格品抽取 1 件不合格品，再从 $N - M$ 件合格品中抽取 $n - 1$ 件合格品，那么

1. 从 M 件不合格品抽取 1 件不合格品，一共有 $\binom{M}{1}$ 种取法；
2. 从 $N - M$ 合格品中抽取 $n - 1$ 件合格品，一共有 $\binom{N-M}{n-1}$ 种取法；

根据乘法原理，取出的 n 件产品中仅有 1 件不合格品的取法一共有 $\binom{M}{1}\binom{N-M}{n-1}$ 种，于是

$$P(A_1) = \frac{\binom{M}{1}\binom{N-M}{n-1}}{\binom{N}{n}} \quad (1.24)$$

于是事件 A_m 为取出的 n 件产品中有 m 件不合格品，要先从 M 件不合格品中抽取 m 件不合格品，一共有 $\binom{M}{m}$ 种取法，再从 $N - M$ 件合格品种抽取 $n - m$ 件合格品，一共用 $\binom{N-M}{n-m}$ 种取法，由乘法原理，事件 A_m 的取

法一共有 $\binom{M}{m}\binom{N-M}{n-m}$ 种, 于是

$$P(A_m) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\frac{N}{n}}, \quad m = 1, 2, \dots, r, \quad r = \min\{n, M\} \quad (1.25)$$

例题 1.10 设有 n 个球, 每个球都等可能地被放到 N 个不同盒子中的任一个, 每个盒子所放球数不限。试求

1. 在指定的 $n(n \leq N)$ 个盒子中各有一球的概率 p_1 ;
2. 恰好有 $n(n \leq N)$ 个盒子各有一球的概率 p_2 。

解 由于每一个球有 N 种放法, 因此 n 个球有 N^n 种放法。

1. 只需考虑 n 个球放在 n 个盒子并且每个盒子有且只有一个球的放法, 第 1 个球在 n 个盒子有 n 种放法, 第 2 个球在剩下的 $n-1$ 个盒子有 $n-1$ 种放法……到第 n 个球在剩下的第 n 个盒子只有 1 种放法, 由乘法原理, n 个球放在 n 个盒子并且每个盒子有且只有一个球的放法共有 $n!$ 种。于是

$$p_1 = \frac{n!}{N^n} \quad (1.26)$$

2. (2) 与 (1) 的区别在于 (1) 的 n 个盒子是提前指定好的, 而 (2) 的 n 个盒子是在试验时随机选取的, 并且一共由 $\binom{N}{n}$ 种选取方法。接着再把 n 个球放入这 n 个各自并且每个盒子有且只有一个球, 我们在 (1) 已经求出了一共有 $n!$ 放法, 于是根据乘法原理, 恰好有 n 个盒子各有一球的放法有 $\binom{N}{n} \cdot n! = A_N^n$, 从而

$$p_2 = \frac{A_N^n}{N^n} = \frac{N!}{N^n(N-n)!} \quad (1.27)$$

例题 1.11 下面我们用例(1.10)的 (2) 的结果考虑这个问题: n 个人的生日全不相同的概率 p_n 是多少。由 (2) 易知

$$p_n = \frac{365!}{365^n(365-n)!} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right)$$

1. 当 n 很小或者不是很大的时候, 右式的交叉项 $\frac{i}{365} \times \frac{j}{365} (i, j = 1, 2, \dots, n-1, i \neq j)$ 可以忽略不计, 于是

$$p_n \approx 1 - \frac{1+2+\cdots+(n-1)}{365} = 1 - \frac{n(n-1)}{730}$$

2. 当 n 足够大的时候 ($n > 27$), 由泰勒公式 $\ln(1-x) \approx -x$, 则

$$\ln p_n \approx -\frac{1+2+\cdots+(n-1)}{365} = -\frac{n(n-1)}{730}$$

例如一个班上 40 人, 那么这 40 人的生日都不相同的概率约为 $\exp(-\frac{40 \times 39}{730}) = 11.8\%$, 也就是说一个班上至少有 2 人生日相同的概率约为 88.2%! 进一步计算可知当 $n \geq 23$ 时, $1 - p_n > 0.5$, 即当人数大于 22 人时, 至少有 2 人生日相同的概率已经超过了 50%; 当 $n \geq 60$ 时, 几乎可以肯定至少有 2 人生日在同一天!

1.3.4 几何概型

几何概型是古典概型的推广, 用于描述在连续空间中随机事件的发生概率。在几何概型中, 事件的概率是基于几何特征(如长度、面积或体积)来计算的。几何概型其基本思想是:

1. 一个随机现象的样本空间 Ω 充满某个区域, 其度量(长度、面积或体积等)大小可用 S_Ω 表示;
2. 任意一点落在度量相同的子区域内(可能位置不同)是等可能性的;
3. 若事件 A 为 Ω 中的某个子区域, 且其度量大小可用 S_A 表示, 则事件 A 的概率为

$$P(A) = \frac{S_A}{S_\Omega} \quad (1.28)$$

容易验证, (1.28)式所定义的“概率”满足非负性和正则性; 当事件 A 与 B 互不相容时, 显然有 $S_{A \cup B} = S_A + S_B$, 于是

$$P(A \cup B) = \frac{S_{A \cup B}}{S_\Omega} = \frac{S_A}{S_\Omega} + \frac{S_B}{S_\Omega} = P(A) + P(B)$$

因此(1.28)式所定义的“概率”的确为可测概率。

求几何概率的关键是对样本空间和所求事件 A 用图形描述清楚(一般用平面或空间图形), 然后计算出相关

图形的度量(一般为面积或体积)。

例题 1.12 甲乙两人约定在下午 6 时到 7 时之间在某处会面, 并约定先到者应等候另一个人 20 分钟, 过时即离去。求两人能会面的概率。

解 记事件 A 为甲乙两人能会面。分别用 x 和 y 分别表示甲、乙两人到达约会地点的时间(以 min 为单位), 在平面上建立 xOy 直角坐标系。 (x, y) 的所有可能取值在边长为 60 的正方形区域内, 其面积为 $S_{\Omega} = 60^2$, 而事件 $A = \text{“两人能够会面”}$ 相当于

$$|x - y| \leq 20$$

其面积为 $S_A = 60^2 - 40^2$ 。根据几何概型有

$$P(A) = \frac{S_A}{S_{\Omega}} = \frac{60^2 - 40^2}{60^2} = 0.5556$$

例题 1.13 平面上画有间隔为 $d(d > 0)$ 的等距平行线, 向平面任意投掷一枚长为 $l(l < d)$ 的针, 求针与任一平行线相交的概率。

解 记事件 A 为针与任意平行线相交, 设 x 为针的中点到最近平行线的垂直距离, φ 为针与最近平行线所形成的夹角, 于是样本空间 Ω 就为

$$\Omega = \left\{ (x, \varphi) \mid 0 \leq x \leq \frac{d}{2}, 0 \leq \varphi \leq \pi \right\}$$

易知针与平行线相交的充分必要条件是:

$$x \leq \frac{l}{2} \sin \varphi$$

于是由几何概型可知

$$P(A) = \frac{S_A}{S_{\Omega}} = \frac{\int_0^{\pi} \frac{l}{2} \sin \varphi \, d\varphi}{\frac{d}{2} \pi} = \frac{2l}{d\pi}$$

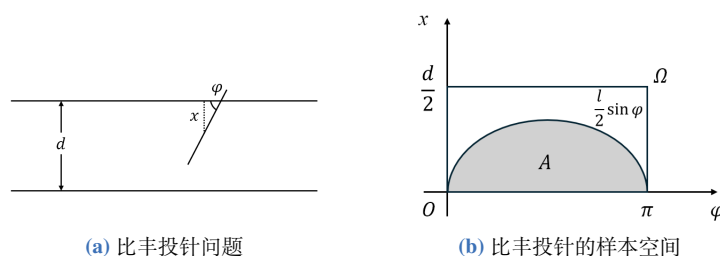


图 1.8: 比丰投针

其中 l, d 为已知量, 只需将 π 代入即可计算出 $P(A)$ 的值; 反过来若已经用事件 A 发生的频率 $\frac{n}{N}$ 估计出了 $P(A)$ 的值, 则可以反解出 π 来, 即

$$\pi \approx \frac{2lN}{dn} \quad (1.29)$$

这是一个颇为奇妙的方法: 只要设计一个随机试验, 使一个事件的概率与某个未知数有关, 然后通过重复试验, 以频率估计概率, 即可求得未知数的近似解。

一般来说试验次数越多, 则求得的近似解就越精确, 随着计算机的出现, 人们便可利用计算机来大量重复地模拟所设计的随机试验。这种方法得到了迅速的发展和广泛的应用, 人们称这种方法为**随机模拟法**, 也称为**Monte Carlo 法**。

例题 1.14 在圆内任取一条弦, 问其长度超过该圆内接等边三角形的边长的概率是多少?

解 该问题可以有三种不同的解法:

1. 假定弦的中点在直径上均匀分布, 直径上的点组成样本空间 Ω_1 。由于对称性可只考察某指定方向的弦, 作一条直径垂直于这个方向。显然只有交直径于 $\frac{1}{4}$ 与 $\frac{3}{4}$ 之间的弦才能超过正三角形的边长, 如此所求概率为 $\frac{1}{2}$ 。

- 假定弦的另一活动端点在圆周上均匀分布，圆周上的点组成样本空间 Ω_2 。由于对称性可让弦的一端点固定，让另一端点在圆周上作随机移动。若在固定端点作一切线，则与此切线夹角在 60° 与 120° 之间的弦才能超过正三角形的边长，如此所求概率为 $\frac{1}{3}$ 。
- 假定弦的中点在大圆内均匀分布，大圆内的点组成样本空间 Ω_3 。圆内弦的位置被其中点唯一确定。在圆内作一同心圆，其半径仅为大圆半径的一半，则大圆内弦的中点落在小圆内，此弦长才能超过正三角形的边长，如此所求概率为 $\frac{1}{4}$ 。

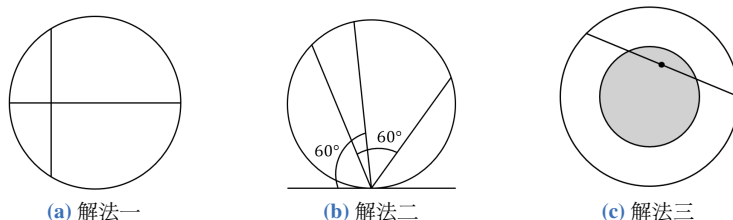


图 1.9: Bertrand 悖论的三种解法

笔记 可见上述三个答案是针对三个不同样本空间引起的，它们都是正确的，Bertrand 悖论引起人们注意，在定义概率时要事先明确指出样本空间是什么！

1.4 条件概率

内容提要

- ☐ 条件概率
- ☐ 优势比
- ☐ 乘法公式
- ☐ 全概率公式
- ☐ Bayes 公式

1.4.1 条件概率的定义

条件概率(Conditional probability) 是在给定**部分信息**的基础上对试验结果一种推断，当然也可以由试验结果反推试验原因的一些性质与可能性，这就是贝叶斯统计的研究内容了。

用更确切的话说，给定一个试验与这个试验相对应的样本空间和概率分布，假设我们已经知道给定的事件 B 发生了，希望知道另一个给定的事件 A 发生的可能性。因此我们要构造一个新的概率分布，它顾及了事件 B 已经发生的信息，求出任何事件 A 发生的概率。

定义 1.5 (条件概率)

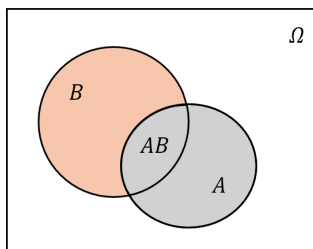
设 A 与 B 是样本空间 Ω 中的任意两个事件，若 $P(B) > 0$ ，则称

$$P(A | B) = \frac{P(AB)}{P(B)} \quad (1.30)$$

为“在 B 发生的条件下 A 发生”的概率，简称为条件概率 (Conditional probability)。

为了与条件概率相对应， $P(A)$ 可以被称作**无条件概率**(Unconditional probability)。条件概率 $P(A | B)$ 可理解为由于事件 B 的发生，**样本空间从 Ω 变为了 $\Omega_B = B$** 。从几何直观上看，无条件概率 $P(A)$ 就是事件 A (图中灰色部分) 占样本空间 Ω 的比例，条件概率 $P(A | B)$ 就是事件 AB 的“面积”占事件 B (图中橙色部分加上事件 AB 的灰色部分) 的“面积”的比例：

笔记 在通常意义下，若 $P(B) = 0$ ，那么条件概率 $P(A | B)$ 则无法定义或者无意义，也就是说若事件 B 根本不可能发生，那么以事件 B 作为前提来讨论事件 A 的发生情况就不再有意义。不过在测度论的框架下，对于**零概率事件**的条件概率有更为复杂的处理方法，如 Radon-Nikodym 导数，但这是在更抽象的数学背景下讨论的。

图 1.10: 条件概率 $P(A | B)$ 的 Venn 图示意

例题 1.15 在肝癌普查中发现, 某地区的自然人群中, 每十万人内平均有 40 人患原发性肝癌, 有 34 人甲胎球蛋白高含量, 32 人既患原发性肝癌又出现甲胎球蛋白高含量。设事件 A 为患肝癌, 事件 B 为甲胎球蛋白高含量, 那么 $P(A) = 0.0004$, $P(B) = 0.00034$, $P(AB) = 0.00032$, 从而

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{0.00032}{0.00034} = 0.9412, \quad P(B | A) = \frac{P(AB)}{P(A)} = \frac{0.00032}{0.0004} = 0.8$$

$P(B | A) = 0.8$ 这说明患肝癌的人群中有 80% 的人会有甲胎球蛋白高含量的症状, 因此检测甲胎球蛋白的含量能够很好地判断一个人是否患肝癌; 而 $P(A | B) = 0.9412$ 说明若出现了甲胎球蛋白高含量的症状, 那么对一个人患肝癌的概率判断的正确性高达 94.12%。

由于事件 B 的发生, 使事件 A 发生的概率由 0.0004 一下子上升到 0.9412。可见**事件发生的概率, 与条件有关, 也与已知信息有关**。

命题 1.5

设 $P(B) > 0$, 若

1. $\forall A \in \mathcal{F}$, 有 $P(A | B) \geq 0$;
2. $P(\Omega | B) = 1$;
3. 若 $\forall A_1, A_2, \dots, A_n \in \mathcal{F}$, 且 A_i 与 A_j 之间两两之间互不相容 ($i, j = 1, 2, \dots, n; i \neq j$), 则

$$P\left(\bigcup_{i=1}^n A_i | B\right) = \sum_{i=1}^n P(A_i | B) \quad (1.31)$$

则条件概率 $P(A | B)$ 是定义在概率空间 (Ω, \mathcal{F}, P) 上的概率。

因此类似于无条件概率, 条件概率也可由三个基本性质导出其他些性质, 例如:

1. $P(\emptyset | B) = 0$;
2. $P(A | B) = 1 - P(\bar{A} | B)$;
3. 对于任意两个事件 A_1 与 A_2 , 有 $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B) - P(A_1 A_2 | B)$ 。

1.4.2 乘法公式

乘法公式是计算**两个事件同时发生的概率**的工具, 它将两个事件的联合概率分解为两个步骤: 首先计算事件 B 发生的概率, 然后计算在 B 发生的前提下事件 A 发生的概率。这种分解能**从局部的角度理解复杂的联合概率**, 其意义在于它可以**分步骤计算复杂的联合概率**, 揭示了事件之间的相互依赖关系。

定理 1.2 (乘法公式)

设 A 与 B 是两个任意事件, 若 $P(B) > 0$, 那么

$$P(AB) = P(B)P(A | B) \quad (1.32)$$

由条件概率的定义式可知,

$$P(AB) = P(B)P(A | B) = P(A)P(B | A) \quad (1.33)$$

那么则有

$$P(A | B) = \frac{P(A)}{P(B)} \cdot P(B | A) \quad (1.34)$$

也就是说 $P(A | B)$ 与 $P(B | A)$ 的关系是依靠调整因子 $\frac{P(A)}{P(B)}$ 联系在一起的, 并且 $P(A | B)$ 与 $P(B | A)$ 的比例就是 $\frac{P(A)}{P(B)}$ 的值, 其意义就在于让我们能够清楚地了解是事件 A 影响事件 B 的因素大, 还是事件 B 影响事件 A 的因素大。

若事件 A 与事件 B 之间不是独立发生的两个事件, 乘法公式也体现了一个事件的发生如何影响另一个事件的发生概率。通过计算条件概率 $P(A | B)$ 可以准确估计联合概率。若事件 A 与事件 B 是独立的, 即 $P(A | B) = P(A)$, 则有 $P(AB) = P(A) \cdot P(B)$, 这正是独立事件的定义。

例题 1.16 今早天气预报播报的下雨概率为 70%, 记为事件 A ; 而根据以往的生活经验, 下雨时堵车的概率为 40%, 记作“堵车”事件 B 。那么今早既下雨又堵车的概率则为

$$P(AB) = P(A)P(B | A) = 0.28$$

推论 1.5

若 $P(A_1 A_2 \cdots A_{n-1}) > 0$, 那么

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \cdots P(A_n | A_1 A_2 \cdots A_{n-1}) \quad (1.35)$$

证明 由于

$$\begin{aligned} P(A_1 A_2 \cdots A_n) &= P(A_1 A_2 \cdots A_{n-1})P(A_n | A_1 A_2 \cdots A_{n-1}) \\ &= P(A_1 A_2 \cdots A_{n-2})P(A_{n-1} | A_1 A_2 \cdots A_{n-2})P(A_n | A_1 A_2 \cdots A_{n-1}) \\ &= \cdots \\ &= P(A_1) \cdots P(A_{n-1} | A_1 A_2 \cdots A_{n-2})P(A_n | A_1 A_2 \cdots A_{n-1}) \end{aligned}$$

又由于

$$P(A_1) \geq P(A_1 A_2) \geq \cdots \geq P(A_1 A_2 \cdots A_{n-1}) > 0$$

因此 $P(A_2 | A_1), P(A_3 | A_1 A_2), \cdots, P(A_n | A_1 A_2 \cdots A_{n-1})$ 都是有意义的。□

例题 1.17 你站在三个封闭的门前, 其中一个门后有奖品, 当然奖品在哪一个门后是完全随机的。当你选定一个门以后, 主持人打开其余两扇门中的一扇空门, 显示门后没有奖品。此时你可以有两种选择, 保持原来的选择或者改选另一扇没有被打开的门。

当你作出最后选择以后, 如果打开的门后有奖品, 这个奖品就归你。现在有三种策略:

1. 坚持原来的选择;
2. 改选另一扇没有被打开的门;
3. 你首先选择 1 号门, 若主持人打开的是 2 号门, 你则坚持原来的选择, 若主持人打开的是 3 号门, 你则选择 2 号门。

问最好的策略是什么? 现在计算在各种策略之下赢得奖品的概率。

解 设事件 A_i 为“第 i 次选择的门后有奖品”, 则

1. 此策略赢得奖品的概率完全取决于你初始选择的门, 因此赢得奖品的概率为 $P(A_1) = \frac{1}{3}$;
2. 第一步必须选择一扇门后没有奖品的门, 其概率为 $P(\bar{A}_1) = \frac{2}{3}$, 第二步你改选另一扇没有没打开的门, 其门后有奖品的概率为 $P(A_2 | \bar{A}_1) = 1$, 于是由乘法公式, 赢得奖品的概率为 $P(\bar{A}_1 A_2) = \frac{2}{3}$;
3. 在此策略中由于提供的信息不够充分(赢得奖品的概率依赖于主持人打开门的方式), 因此再分为以下两种情况讨论:
 - (a). 假设主持人始终会打开 2 号门(当奖品在 2 号门时, 主持人则会打开其他号门)。假定奖品在 1 号门后(其概率为 $\frac{1}{3}$), 主持人则会打开 2 号门, 这时赢得奖品的概率为 $P(A_1) = \frac{1}{3}$; 假定奖品在 2 号门后, 主持人打开 3 号门, 这时赢得奖品的概率为 $P(\bar{A}_1 A_2) = \frac{1}{3}$; 假定奖品在 3 号门后, 主持人打开 2 号门, 这时赢得奖品的概率为 0。于是由加法原理, 赢得奖品的概率为 $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$ 。

- (b). 假设主持人会随机地打开 2 号门或者 3 号门。假定奖品在 1 号门后 (其概率为 $\frac{1}{3}$), 主持人随机打开 2 号门或者 3 号门 (其概率为 $\frac{1}{2}$), 这时赢得奖品的概率为 $P(A_1) = \frac{1}{6}$; 假定奖品在 2 号门后, 主持人随机打开 2 号门或者 3 号门, 这时赢得奖品的概率为 $P(\bar{A}_1 A_2) = \frac{1}{3}$; 假定奖品在 3 号门后, 主持人随机打开 2 号门或者 3 号门, 这时赢得奖品的概率为 0。于是由加法原理, 赢得奖品的概率为 $\frac{1}{6} + \frac{1}{3} = \frac{1}{2}$ 。

1.4.3 全概率公式

全概率公式将事件的概率分解为多个条件下的概率, 从而提供了一种结构化的计算方式, 可以更有效地计算复杂情境下的概率。它特别适合处理多个互不相容的子事件, 在这些不同的子事件中包含另一个事件总的发生概率。

定理 1.3

设 A 是一个任意事件, 若 B_1, B_2, \dots, B_n 是样本空间 Ω 的一个分割, 即 B_1, B_2, \dots, B_n 两两之间互不相容, 且 $\bigcup_{i=1}^n B_i = \Omega$ 。若 $P(B_i) > 0 (i = 1, 2, \dots, n)$, 则 A 发生的概率为

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i) \quad (1.36)$$

证明 由于 B_1, B_2, \dots, B_n 构成样本空间 Ω 的一个分割, 于是

$$A = A\Omega = A \left(\bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (AB_i)$$

又由于 B_1, B_2, \dots, B_n 两两之间互不相容, 因此 $(AB_1) \cap (AB_2) \cap \dots \cap (AB_n) = \emptyset$, 从而 AB_1, AB_2, \dots, AB_n 两两之间也互不相容, 于是

$$P(A) = \sum_{i=1}^n P(AB_i)$$

由乘法公式

$$P(AB_i) = P(B_i)P(A | B_i) \quad i = 1, 2, \dots, n$$

因此

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

□

可以把全概率公式理解为“加权平均”的过程: 事件 A 的总概率是基于不同情况 B_1, B_2, \dots, B_n 下发生的可能性, 以及在这些情景下事件 A 发生的条件概率, 通过每种情况 $B_i (i = 1, 2, \dots, n)$ 下发生的概率 $P(B_i)$ 来对相对应的条件概率 $P(A | B)$ “加权”, 最终得到事件 A 的总概率。

笔记

1. B_1, B_2, \dots, B_n 不必为 Ω 的一个分割, 只需 B_1, B_2, \dots, B_n 是两两互不相容的事件, 且 $A \subset \bigcup_{i=1}^n B_i$ 即可;
2. 可列个 $B_1, B_2, \dots, B_n, \dots$ 是两两互不相容的事件, 且 $A \subset \bigcup_{i=1}^{\infty} B_i$, 则全概率公式依旧成立;
3. 要求 $\sum_{i=1}^n P(B_i) = 1$, 解释为对应的条件概率的权重, 体现各条件概率的重要程度或者贡献度。

例题 1.18 在社会调查中, 若调查内容包含有隐私性问题, 被调查者通常不愿意直接回答。例如在校园中调查学生接触过黄色书刊或者影像是否会影响其身心健康发展的课题, 从调查数据中估计出学生中解出过黄色书刊或者影像的比例 p , 为此需要制作两份不同的调查问卷:

1. 你的生日是否在 7 月 1 日之前?
2. 你是否接触过黄色书刊或者影像?

调查员只需向被调查者随机地出示这两份问卷的其中之一 (比如将问卷放入小球中, 让被调查者随机地抽取小

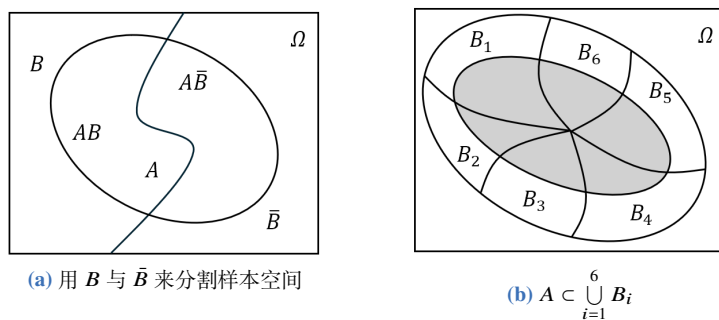


图 1.11: 不同形式的全概率公式

球); 被调查者只需回答“是”或者“否”, 然后让其在答卷上勾选自己所认同的答案即可。这样旁人就无法得知被调查者回答的是问题 (1) 还是问题 (2), 从而极大地打消了被调查者的心理顾虑, 使得调查结果更为准确。

下面假设一共收到了 n 张答卷, 其中有 k 张回答“是”, 但是我们无法得知 n 张答卷中有多少回答了问题 (2), 也不知道 k 中回答“是”的答卷中有多少回答了问题 (2)。但是我们可以提前设定的是:

1. 在 n 充分大时, 任意一人的生日在 7 月 1 日之前的概率应该为 0.5;
2. 控制问题 (2) 占总体的比率 π ;

于是我们可以利用全概率公式解出 p :

$$\begin{aligned} P(\text{是}) &= P(\text{问题 (1)})P(\text{是} | \text{问题 (1)}) + P(\text{问题 (2)})P(\text{是} | \text{问题 (2)}) \\ &= 0.5(1 - \pi) + \pi p \end{aligned}$$

其中 $P(\text{是})$ 可以用频率 $\frac{k}{n}$ 代替 (当 n 充分大时), 因此

$$\frac{k}{n} = 0.5(1 - \pi) + \pi p$$

从而

$$p = \frac{k/n - 0.5(1 - \pi)}{\pi}$$

例如在一次调查中, 设置问题 (2) 的比率为 $\pi = 0.6$ 。1583, 其中有 389 份答卷回答了“是”, 那么就可以估计出 p 的值:

$$p = \frac{389/1583 - 0.2}{0.6} = 0.0762$$

即大约有 7.62% 的学生接触过黄色书刊或者影像。

例题 1.19 双胞胎可能是同卵双生或者异卵双生, 同卵双生也叫单卵双生, 是由一个受精卵分裂成为两个完全一样的部分发育而来的, 因此同卵双胞胎含有相同的基因。异卵双生又叫二卵双生, 是由两个受精卵植入子宫发育而来的。异卵双胞胎像不同时间出生的兄弟姐妹一样, 基因多少是有些一样的, 为了知道双胞胎中同卵双生所占的比例, 统计学家首先要求市里每一家医院对双胞胎做记录, 同时对是否是同卵双生做标记, 然而医院告诉他判断一个新生儿是否是同卵双生并不是一件简单的事, 这关系到父母是否愿意自费给孩子做这项复杂而又昂贵的 DNA 检验, 经过一番思考之后, 统计学家只让医院提供标记着双胞胎是否是相同性别的所有双胞胎数据列表。当数据表明约有 64% 的双胞胎是性别相同时, 统计学家就宣称约有 28% 的双胞胎是同卵双生, 这是为什么呢?

解 首先统计学家假设同卵双生双胞胎性别总是相同的 (事实上同卵双生双胞胎几乎总是相同性别的)。又因为异卵双生双胞胎就相当于普通的兄弟姐妹, 所以性别相同的概率也有 $\frac{1}{2}$ 。令事件 I 表示“同卵双生双胞胎”, 令事件 SS 表示“双胞胎性别相同”, 他在双胞胎是否为同卵双生的条件下计算概率 $P(SS)$, 得到

$$\begin{aligned} P(SS) &= P(I)P(SS | I) + P(\bar{I})P(SS | \bar{I}) \\ &= 1 \times P(I) + \frac{1}{2} \times [1 - P(I)] \\ &= \frac{1}{2} + \frac{1}{2}P(I) \end{aligned}$$

于是当 $P(SS) \approx 0.64$ 时, $P(I) \approx 0.28$, 即当约有 64% 的双胞胎是性别相同时, 则约有 28% 的双胞胎是同卵双生。

1.4.4 Bayes 公式

首先我们先来探究 $P(A)$ 与 $P(\bar{A})$ 、 $P(A|B)$ 与 $P(\bar{A}|B)$ 的关系:

定义 1.6 (优势比)

设事件 A 发生的概率为 $P(A)$, 那么事件 A 的优势比 (Odds ratio) 则为

$$\frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1 - P(A)} \quad (1.37)$$

事件 A 的优势比告诉我们事件 A 发生的可能性是不发生时的可能性的倍数, 如果 $P(A) = \frac{2}{3}$, 那么 $P(A) = 2P(\bar{A})$ 。如果某事件的优势比等于 α , 那么通常称支持假设 (事件) 成立的优势比为 “ $\alpha:1$ ”。

考虑假设 (事件) H 以概率 $P(H)$ 成立, 如果我们发现了新的证据 E (假设 H 发生所依赖的辅助事件), 那么在 E 成立的条件下, H 成立和 H 不成立的条件概率分别为

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}, \quad P(\bar{H}|E) = \frac{P(\bar{H})P(E|\bar{H})}{P(E)}$$

因此, 引进证据 E 后, 假设 H 的新的优势比为

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(H)}{P(\bar{H})} \cdot \frac{P(E|H)}{P(E|\bar{H})} \quad (1.38)$$

即 H 的新的优势比值等于它原来的优势比的值乘以新的证据在 H 和 \bar{H} 之下的条件概率比值。如果

$$P(E|H) > P(E|\bar{H})$$

则 H 的优势比值是递增的。反之 H 的优势比是递减的。

例题 1.20 一个坛子里装了两枚 A 型硬币和一枚 B 型硬币, 当抛 A 型币时, 正面向上的概率是 $\frac{1}{4}$ 。当抛 B 型硬币时, 正面向上的概率是 $\frac{3}{4}$ 。随机从坛子里取一枚硬币掷, 假定掷出的结果是正面向上, 则所掷的是 A 型硬币的概率是多少?

解 令事件 A 为 “投掷的硬币为 A 型硬币”, 事件 B 为 “投掷的硬币为 B 型硬币”, 并且易知 $\bar{A} = B$, 用 head 表示 “投掷结果为正面向上”, 那么事件 A 的优势比为

$$\frac{P(A|\text{head})}{P(\bar{A}|\text{head})} = \frac{P(A)}{P(\bar{A})} \cdot \frac{P(\text{head}|A)}{P(\text{head}|\bar{A})} = \frac{2}{3}$$

即在硬币正面朝上的情况下事件 A 的优势比为 $\frac{2}{3}:1$, 或者说在硬币正面朝上的情况下事件 A 发生的概率为 $\frac{2}{5}$ 。

在没有任何信息的情况下, 我们判断事件 A 发生的概率为 $\frac{2}{3}$; 而在已经知道硬币正面朝上的信息下, 我们能 “硬币正面朝上” 这个事件的信息量更新判断事件 A 发生的概率为 $\frac{2}{5} < \frac{2}{3}$, 这说明 “硬币正面朝上” 这一个信息对我们判断硬币种类没有任何帮助, 反而还会被 “硬币正面朝上” 这个冗余的信息扰乱我们的视线。

下面考虑新的证据如何影响某个特定假设成立的概率的问题, 令事件 H 表示 “某个特定假设”, 而事件 E 表示 “新的证据”, 那么

$$P(H|E) = \frac{P(E|H)P(H)}{P(H)P(E|H) + P(E|\bar{H})[1 - P(H)]} \quad (1.39)$$

其中 $P(H)$ 为在新证据展示之前我们对假设成立的可能性的估值。新证据支持假设成立, 如果它使得假设成立的可能性增大, 即 $P(H|E) \geq P(H)$ 。于是

$$P(E|H) \geq P(H)P(E|H) + P(E|\bar{H})[1 - P(H)] \quad (1.40)$$

进一步

$$P(E|H) \geq P(E|\bar{H}) \quad (1.41)$$

即当假设成立时, 新证据发生的概率大于假设不成立时发生的概率, 就认为新证据支持假设。事实上, 已知新

的证据发生的条件下，假设成立的新概率和初始概率的关系可从下式看出：

$$P(H | E) = \frac{P(H)}{P(H) + [1 - P(H)] \frac{P(E|\bar{H})}{P(E|H)}} \quad (1.42)$$

(1.39)式的分母实际上为 $P(E) = P(H)P(E | H) + P(\bar{H})P(E | \bar{H})$ ，只包含了事件 H 与其对立事件 \bar{H} ，分子实际上为 $P(HE) = P(H)P(E | H)$ 。可以将其推而广之，于是就有下面的 Bayes 公式 (定理)：

定理 1.4 (Bayes 公式)

设 B_1, B_2, \dots, B_n 是样本空间 Ω 的一个分割，即 B_1, B_2, \dots, B_n 是两两之间互不相容的事件，且 $\bigcup_{i=1}^n B_i = \Omega$ 。若 $P(A) > 0$ ，且 $P(B_i) > 0 (i = 1, 2, \dots, n)$ ，那么

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)} \quad (1.43)$$

证明 由条件概率的定义

$$P(B_i | A) = \frac{P(AB_i)}{P(A)}, \quad i = 1, 2, \dots, n$$

分子与分母分别用乘法公式和全概率公式代入，即得

$$P(B_i | A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)}, \quad i = 1, 2, \dots, n$$

□

假定 B_1, B_2, \dots, B_n 是导致试验结果的各种“原因”， $P(B_i)$ 称之为**先验概率**(Prior)，是事件 B_i 没有任何经验或者知识的情况下的概率，它反映了各种“原因”发生的可能性，一般是对以往经验的总结。

现在在试验中发生了事件 A ，这个信息将有助于探讨事件发生的“原因”， $P(B_i | A)$ 称之为 B_i 的**后验概率**(Posterior)，是通过事件 A 发生的条件下，用事件 A 的经验或者知识来修正先验概率 $P(B_i)$ ，它反映了试验之后对各种“原因”发生的可能性大小的新知识。

$P(A | B_i)$ 称之为事件 A 发生的**似然概率**¹(Likelihood)，似然概率是对因果关系的一种量化。

例题 1.21 某地区居民的肝癌发病率为 0.0004，现用甲胎蛋白法进行普查。医学研究表明，化验结果是可能存有错误的。已知患有肝癌的人其化验结果 99% 呈阳性，而没患肝癌的人其化验结果 99.9% 呈阴性。现某人的检查结果呈阳性，问他真的患肝癌的概率是多少？

解 设事件 A 为患肝癌，事件 B 为化验结果为阳性。则

$$P(A) = 0.0004, \quad P(B | A) = 0.99, \quad P(\bar{B} | \bar{A}) = 0.999$$

于是 $P(B | \bar{A}) = 0.001$ ，因此

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B | A)}{P(A)P(B | A) + P(\bar{A})P(B | \bar{A})} = 0.2837$$

这说明在 10000 个人中真正患肝癌的人有 4 个，但是在剩下的 9996 人中会有 $9996 \times 0.001 = 9.996$ 人会在化验单上查出自己阳性，并且在真正患肝癌的 4 人中仅有 $4 \times 0.99 = 3.96$ 人能够在化验单上查出自己阳性。如果仅从这 $9.996 + 3.96 = 13.956$ 个查出自己阳性的人中看，真正患肝癌的比例约为 28.37%。

由此可见如果按照题设的方法检测是否患肝癌，其准确率并不高，因此进一步降低错检的概率是提高检验精度的关键。若给疑似患肝癌的人群再次进行复查，此时 $P(A) = 0.2837$ ，这时再用 Bayes 公式计算得

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B | A)}{P(A)P(B | A) + P(\bar{A})P(B | \bar{A})} = 0.9975$$

这就大大提高了甲胎蛋白法的准确率了。

¹ 概率 (Probability) 与似然概率 (Likelihood) 的区别在于：Probability 是给定参数的情况下预测某个事件发生的可能性；而 Likelihood 是给定数据推测参数的可能性。

例题 1.22 保险公司认为人可以分为两类，一类易出事故，另一类则不易出事故。统计表明一个易出事故者在一年内发生事故的的概率为 0.4，而对不易出事故者来说这个概率则减少到 0.2，若假定第一类人占人口的比例为 30%，现有一个新人来投保，那么该人在购买保单后一年内将出事故的的概率有多大？再假设一个新投保人在购买保单后一年内出了事故，那么他是易出事故者的概率是多大？

解 令事件 A 为“投保人一年内出事故”，事件 B 为“投保人为易出事故者”，那么

$$P(A) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B}) = 0.26$$

另外，由乘法公式与条件概率公式，可得

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)P(A|B)}{P(A)} = 0.4615$$

例题 1.23 在回答一道多项选择题时，学生或者知道正确答案（记作事件 K ），或者就猜一个。令 p 表示他知道正确答案的概率，则 $1-p$ 表示猜的概率。假定学生猜中正确答案的概率为 $\frac{1}{m}$ ，其中 m 就是多项选择题的可选择答案数。那么该学生回答正确（记作事件 C ）的情况下，他知道正确答案的概率则为

$$\begin{aligned} P(K|C) &= \frac{P(KC)}{P(C)} = \frac{P(K)P(C|K)}{P(K)P(C|K) + P(\bar{K})P(C|\bar{K})} \\ &= \frac{p}{p + (1-p)\frac{1}{m}} = \frac{mp}{1 + (m-1)p} \end{aligned}$$

贝叶斯公式还可以用来进行因果推理，有许多“原因”可以造成某一“结果”。现在设我们观察到某一结果，希望推断造成这个结果出现的“原因”。当观察到结果 A 的时候，我们希望反推结果 A 是由原因 B_i 造成的概率 $P(B_i|A)$ 。 $P(B_i|A)$ 为由于代表新近得到的信息 A 之后 B_i 出现的概率。

用一句话总结，贝叶斯公式告诉我们，在试验之前对这些假设条件所作的判断（即 $P(B_i)$ ），可以如何根据试验的结果来进行修正。

例题 1.24 早上你突然发现自己感冒了（记作事件 A ），其原因可能是：前一天晚上着凉了（记作事件 B_1 ）、其他人感染给你（记作事件 B_2 ）和免疫力下降（记作事件 B_3 ）。已知前一天晚上着凉的的概率为 40%，其他人感染给你的概率为 30%，免疫力下降的概率为 30%。并且因为着凉而感冒的概率为 80%，被传染而感冒的概率为 90%，免疫力下降而感冒的概率为 70%。

于是由全概率公式，由于不同原因导致感冒的概率就为

$$P(A) = \sum_{i=1}^3 P(B_i)P(A|B_i) = 0.8$$

而感冒是由于着凉所引起的概率为

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = 0.4$$

同理可分别计算出感冒是由于被传染所引起的概率和感冒是由于免疫力下降所引起的概率为

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(A)} = 0.3375, \quad P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(A)} = 0.2625$$

由此可见，早上引起感冒的原因很有可能是因为前一天晚上着凉而导致的。

出现这种情况是由于 $P(B_i|A) \propto P(B_i)P(A|B_i)$ ，而先验概率 $P(B_1)$ 与似然概率 $P(A|B_1)$ 在三个原因中都较高，因此后验概率 $P(B_1|A)$ 自然也比较高。

例题 1.25 一架飞机失踪了，推测它等可能地坠落在 3 个区域。令 $1-\beta_i (i=1,2,3)$ 表示飞机坠落在第 i 个区域的情况下，飞机被发现的概率（ β_i 称为忽略概率，表示忽略飞机的概率，通常由该区域的地理和环境条件决定）。求在第 1 个区域没有发现飞机的条件下，飞机坠落在第 i 个区域的条件概率（ $i=1,2,3$ ）。

解 令事件 R_i 表示“飞机坠落在第 i 个区域”（ $i=1,2,3$ ），事件 E 表示“第 1 个区域没有发现飞机”。我们先来求 $R_1|E$ 的概率，此时

$$P(E) = P(R_1)P(E|R_1) + P(R_2)P(E|R_2) + P(R_3)P(E|R_3) = \frac{2}{3} + \frac{1}{3}\beta_1$$

于是

$$P(R_1|E) = \frac{P(R_1)P(E|R_1)}{P(E)} = \frac{\beta_1}{\beta_1 + 2}$$

接着求 $R_2 | E$ 与 $R_3 | E$ 的概率:

$$P(R_2 | E) = \frac{P(R_2)P(E | R_1)}{P(E)} = \frac{1}{\beta_1 + 2}$$

同理可得 $P(R_3 | E) = P(R_2 | E) = \frac{1}{\beta_1 + 2}$ 。

注意, 当搜索了第 1 个区域没有发现飞机时, 飞机坠落在第 j 个区域的条件概率会增大 ($j \neq 1$), 而落在第 1 个区域的概率会减小, 这是一个常识问题: 因为既然在第 1 个区域没有发现飞机, 当然飞机坠落在该区域的概率会减少, 而落在其他区域的概率会增大, 而且飞机坠落在第 1 个区域的条件概率是 β_i 的递增函数, 当增加时增大了飞机坠落在第 1 个区域的条件概率。类似地, $P(R_j | E)$ 是 β_i 的递减函数 ($j \neq 1$)。

例题 1.26 储物箱里有 3 种不同的一次性手电。第一种手电使用超过 100 小时的概率为 0.7, 而第二种和第三种手电相应的概率分别只有 0.4 和 0.3。假设箱子里的手电, 20% 为第一种, 30% 为第二种, 50% 第三种。那么

1. 随机挑一个手电, 能使用 100 小时以上的概率是多少?
2. 已知手电使用超过 100 小时, 那么它是第 i 种 ($i = 1, 2, 3$) 手电的条件概率是多少?

解 (1) 设事件 B 表示“手电能使用 100 小时以上”, 事件 A_i 表示“取到第 i 种手电” ($i = 1, 2, 3$)。于是

$$P(B) = \sum_{i=1}^3 P(A_i)P(B | A_i) = 0.41$$

(2) 已知手电使用超过 100 小时, 于是 A_i ($i = 1, 2, 3$) 的后验概率则为

$$P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{P(B)} = 0.3415$$

$$P(A_2 | B) = \frac{P(A_2)P(B | A_2)}{P(B)} = 0.2927$$

$$P(A_3 | B) = \frac{P(A_3)P(B | A_3)}{P(B)} = 0.3659$$

1.5 独立性

内容提要

□ 独立性的定义

□ 随机试验的独立性

1.5.1 独立性的定义

两个事件之间的独立性是指: 一个事件的发生不影响另一个事件的发生, 即事件 B 的发生并没有给事件 A 带来新的信息, 它没有改变事件 A 发生的概率。

另外从概率的角度看, 事件 A 的条件概率 $P(A | B)$ 与无条件概率的差别在于: 事件 B 的发生改变了事件 A 发生的概率, 也即事件 B 对事件 A 有某种“影响”。如果事件 A 与 B 的发生是相互不影响的, 则有 $P(A | B) = P(A)$ 和 $P(B | A) = P(B)$, 它们都等价于

$$P(AB) = P(A)P(B)$$

另外对 $P(B) = 0$ 或者 $P(A) = 0$, 上式仍然成立。为此, 我们用该式作为两个事件相互独立的定义。

定义 1.7 (两个事件的独立性)

设事件 A 与 B 是样本空间 Ω 中的任意两个事件, 如果

$$P(AB) = P(A)P(B) \quad (1.44)$$

成立, 则称事件 A 与 B 相互独立 (Independent), 简称 A 与 B 独立。否则称 A 与 B 不独立或相依 (Dependent)。



笔记 至于用 $P(AB) = P(A)P(B)$ 作为独立性的定义, 而不用 $P(A | B) = P(A)$ 的理由如下:

1. 独立性是事件之间的一种对称关系，也就是说若事件 A 与 B 独立，那么事件 B 与 A 也必须独立，因此我们需要一个对称性的定义。然而条件概率 $P(A|B)$ 是非对称的，它只考虑了已知事件 B 发生下 A 十分受影响，并且 $P(A|B)$ 并不总是等于 $P(B|A)$ ；
2. 条件概率 $P(A|B)$ 要求 $P(B) > 0$ ，在 $P(B) > 0$ 的情况下 $P(A|B)$ 才有意义。若 $P(B) = 0$ ，则会遇到无法定义事件 A 与 B 的独立性；
3. $P(AB) = P(A)P(B)$ 表明，两个事件的共同发生的概率仅仅是它们各自发生的概率的乘积，表明它们彼此之间没有任何影响。相比之下，条件概率 $P(A|B) = P(A)$ 只描述了在事件 B 已经发生的情况下 A 发生的概率与 B 无关，而无法直观地表达 A 和 B 同时发生的独立性；
4. $P(A|B) = P(A)$ 可以作为 $P(AB) = P(A)P(B)$ 的推论，即

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

例题 1.27 从一副 52 张的扑克牌中任取 1 张，设事件 A 为取到黑桃，事件 B 为取到 J ，则因为

$$P(A) = \frac{1}{4}, \quad P(B) = \frac{4}{52} = \frac{1}{13}$$

而事件 AB 表示“取到黑桃 J ”，故

$$P(AB) = \frac{1}{52} = P(A)P(B)$$

所以 A 与 B 相互独立。

命题 1.6

若事件 A 与 B 独立，则 A 与 \bar{B} 独立， \bar{A} 与 B 独立， \bar{A} 与 \bar{B} 独立。

证明 由于事件 A 与 B 独立，那么

$$P(AB) = P(A)P(B)$$

于是

$$P(A\bar{B}) = P(A - B) = P(A) - P(AB) = P(A)[1 - P(B)] = P(A)P(\bar{B})$$

即事件 A 与 \bar{B} 独立，同理事件 \bar{A} 与 B 也独立。

又由于

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

即事件 \bar{A} 与 \bar{B} 独立，结论得证。 \square

在实践中，通常 $P(AB)$ 与 $P(A)P(B)$ 的绝对误差不超过 10^{-3} 或者 10^{-4} 、相对误差不超过 1% 或者 5%，即

$$|P(AB) - P(A)P(B)| \leq 10^{-4} \quad \text{或者} \quad |P(AB) - P(A)P(B)| \leq 10^{-3}$$

以及

$$\frac{|P(AB) - P(A)P(B)|}{P(AB)} \leq 1\% \quad \text{或者} \quad \frac{|P(AB) - P(A)P(B)|}{P(AB)} \leq 5\%$$

即认为事件 A 与 B 相互独立，也可以根据实际需要选择合适的显著性水平。

定义 1.8 (多个事件的独立性)

设 A, B, C 是三个任意的事件，若有


$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{cases}$$

则称 A, B, C 两两独立；若还有

$$P(ABC) = P(A)P(B)P(C)$$

则称 A, B, C 相互独立



 **笔记** 事件 A, B, C 两两独立不能推出事件 A, B, C 相互独立；反过来事件 A, B, C 相互独立也不能推出事件 A, B, C 两两独立！

1.5.2 试验的独立性

定义 1.9

设有两个试验 E_1 和 E_2 ，假如试验 E_1 的任一结果(事件)与试验 E_2 的任一结果(事件)都是相互独立的事件，则称这两个试验相互独立。



例如掷一枚硬币(试验 E_1)与掷一颗骰子(试验 E_2)是相互独立的试验。

类似地可以定义 n 个试验 E_1, E_2, \dots, E_n 的相互独立性：如果 E_1 的任一结果、 E_2 的任一结果、 \dots 、 E_n 的任一结果都是相互独立的事件，则称试验 E_1, E_2, \dots, E_n 相互独立。

如果这几个独立试验还是相同的，则称其为 n 重独立重复试验。如果在 n 重独立重复试验中，每次试验的可能结果为两个： A 或者 \bar{A} ，则称这种试验为 n 重 Bernoulli 试验。

附录 A 计算机模拟

A.1 Buffon 投针的计算机模拟

A.2 三门问题的计算机模拟