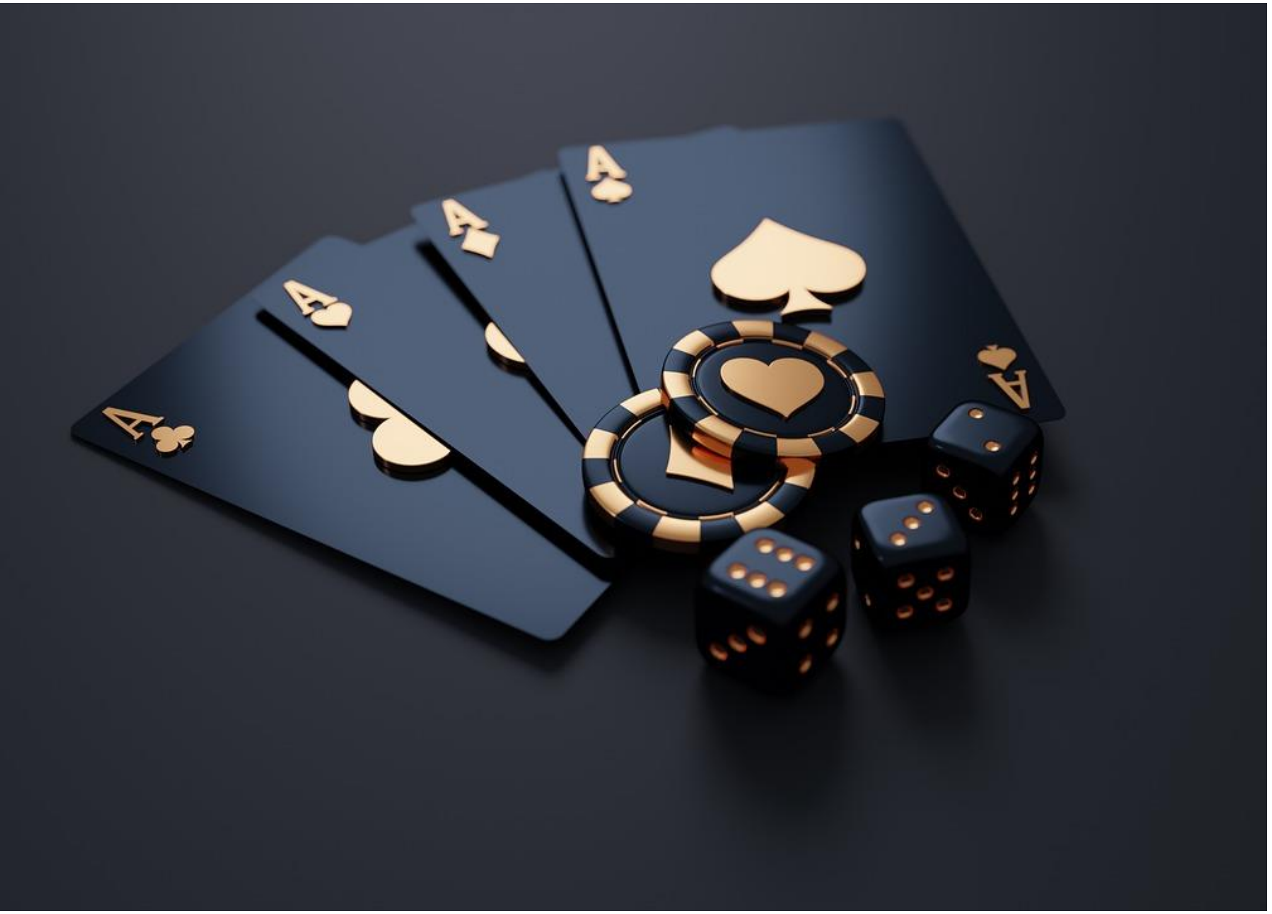
高中统计学 第壹课

概率的定义与计算

何濯羽 2024年2月7日

目录



Section 1

概率的定义

古典派 VS 频率派 VS 主观派

概 率 的 定 义

“**概率是什么？ What is *probability*?**”

这不仅是一个数学问题；这也是哲学问题。

根据针对“概率”的定义方式，我们可以将统计学家们分为三个主流学派：

- 古典派（古典概率）
- 频率派（试验概率）
- 主观派（主观概率）

古典学派

The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.

—— Pierre-Simon Laplace, *Théorie Analytique des Probabilités* (1812)

Jacob Bernoulli
(1655-1705)

Pierre-Simon Laplace
(1749-1827)

古典派认为：

未知的概率均为"等概率"。

例如，因为我们不知道掷出一枚普通硬币后哪一面更可能朝上，所以我们应该赋予正面和反面相同的概率，即

$$\Pr(\text{正面朝上}) = \Pr(\text{反面朝上}) = \frac{1}{2}$$

古典学派

The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.

—— Pierre-Simon Laplace, *Théorie Analytique des Probabilités* (1812)

Jacob Bernoulli
(1655-1705)

Pierre-Simon Laplace
(1749-1827)

古典概率（classical probability，或称“古典概型”）的**缺陷**：

- “未知的概率均为等概率”是循环定义，并没有回答“概率是什么？”
- 古典概型没有办法处理已知的非等概率的情况。例如，我们被告知这是一枚不均匀的硬币，那么正面朝上的概率是多少?
- 勃兰特悖论（Bertrand’s paradox）

频 率 学 派

The probable is that which for the most part happens.

—— Aristotle, *Rhetoric* (4th century BCE)

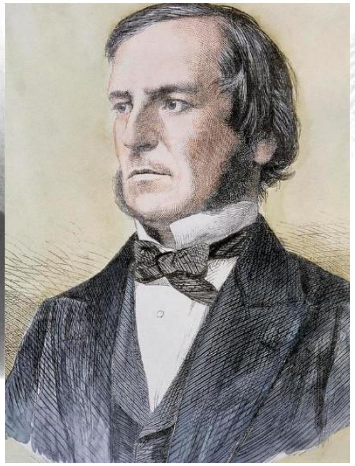
19世纪30年代，频率派崛起，开始对古典概率进行全面轰炸。

频率派认为：我们只能通过记录试验结果，然后计算各种结果的频率，去**估计**概率。当试验次数增大到无穷大时，各种结果的频率无限逼近它们各自的概率，即"**频率存在稳定性**"。

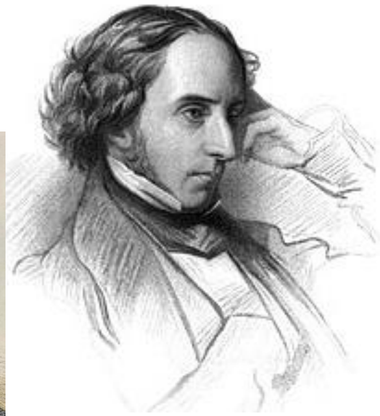
Jakob Friedrich Fries
(1773-1843)

Ronald Aylmer Fisher
(1890-1962)

Antoine Augustin Cournot
(1801-1877)

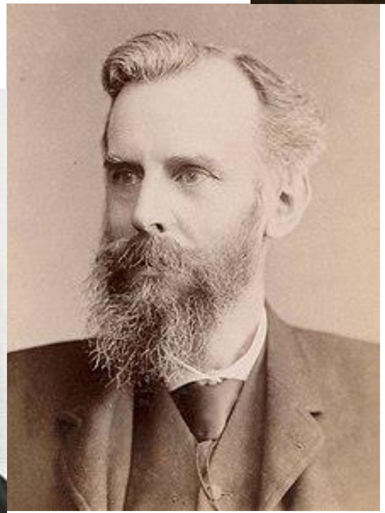


George Boole
(1815-1864)



Robert Leslie Ellis
(1817-1859)

Joseph Bertrand
(1822-1900)



John Venn
(1834-1923)

频率学派

The probable is that which for the most part happens.

—— Aristotle, *Rhetoric* (4th century BCE)

频率派的观点有三个基本特征：

- **试验性**：我们必须通过试验才能计算频率。
- **重复性**：每次重复试验的程序必须完全相同。
- **误差性**：频率只是概率的估计值，存在误差。

例如，一枚硬币落地时正面朝上的概率是

$$\Pr(\text{正面朝上}) = \boxed{\lim_{n\to\infty} \mathrm{Fr}_n(\text{正面朝上})}$$

投掷 ∞ 次相同硬币后计算得到的正面朝上的频率

频率学派

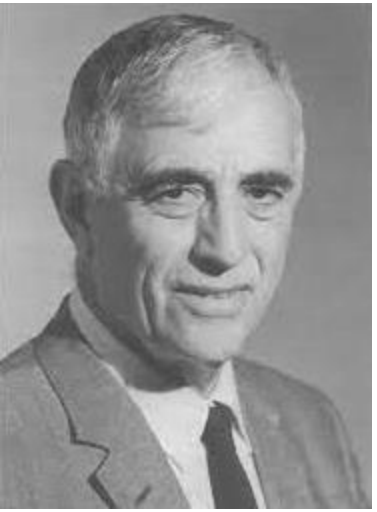
The probable is that which for the most part happens.

—— Aristotle, *Rhetoric* (4th century BCE)

试验概率（frequentist probability）的**缺陷**：

- 为了准确估计概率，我们需要重复试验 ∞ 次。现实中， ∞ 次如何实现?
- 每次试验的过程必须完全相同，这在现实中难以实现。例如，每次抛掷硬币时，硬币沾上的手汗都不相同。
- 一些事件无法进行重复试验。例如，如何估计富士火山喷发的概率?

主观学派



Bernard Osgood Koopman
(1900-1981)

Frank Plumpton Ramsey
(1903-1930)

主观派以个人信念强度（degree of belief）为基础，认为“概率”是人们根据有效的证据对某一事件发生的可能性的**主观评价**。简言之，事件 A 发生的概率等于个人对事件 A 发生的信念强度。

例：我个人相信兰溪市第一中学2024届全部学生考上重点大学的概率为70%。

主观学派

主观概率（subjective probability）的**缺陷**：

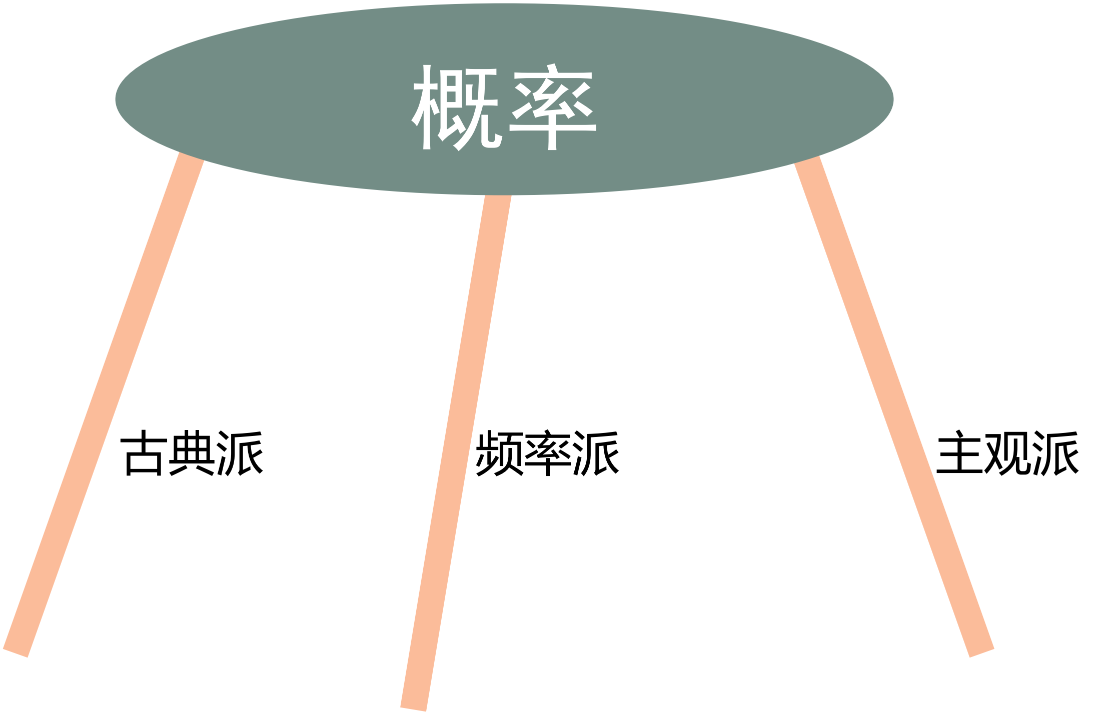
- X 主观概率充满了个人偏见，与科学的客观性相冲突。
- X 人们很难对某个主观概率达成共识。

概率的定义

三个主流概率学派的分歧是“解释上”的分歧，而非“数学上”的分歧。

实际上，概率论与统计学发展至今，三个学派已经不再泾渭分明。

例如，我们在日后将要学习的线性回归模型与假设检验中，既可以看见频率派的身影（使用实际数据），也可以看见主观派的思想（选择置信度）。



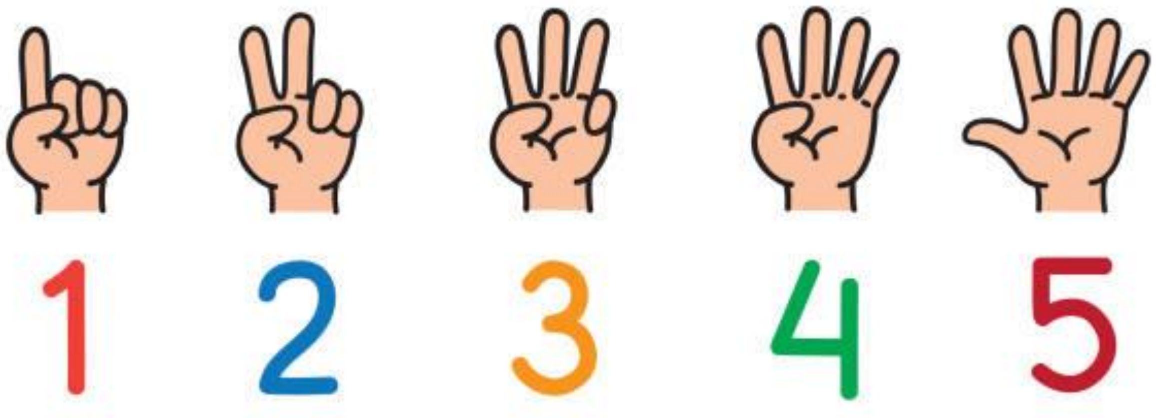
Section 2

计数基本法则

加法原理 VS 乘法原理

排列 VS 组合

计 数 的 目 的



古典派认为“未知的概率即为等概率”。为了计算出古典概型中的概率，我们必须知道某一事件可能出现的结果的个数。这就要求我们必须掌握**计数基本法则**（fundamental principle of counting）：

- 加法原理
- 乘法原理
- 排列数的计算
- 组合数的计算

计数（counting）是一个许多人在入学启蒙前就已经具备的能力。然而，计数对于人类而言，从不是一件简单的任务……

加法原理

假设我们可以选择两个模式：在模式 A 中，我们有 a 种不同的方法完成任务；在模式 B 中，我们有 b 种不同的方法完成任务。我们不可能同时选择两个模式，两个模式中的方法均不相同。请问：完成任务总共有几种不同的方法?

$$\boldsymbol{N = a + b}$$

假设我们有 m_k 种不同的方法在第 k 种模式下完成任务（其中 $k = 1, 2, \ldots, n$），我们只能选择一种模式，且不同模式下的方法彼此不同。请问：完成任务总共有几种方法?

$$\boldsymbol{N = m_1 + m_2 + \cdots + m_n = \sum_{i=1}^{n} m_i}$$

加法原理适用的情境有以下特点：

- 每种模式下的每个方法都能完成任务，不依赖于其它方法或模式。
- 同一模式下，两种方法不能同时使用。
- 不同模式下，没有完全相同的方法存在。

乘法原理 ✕

如果完成一个任务必须经过 n 个步骤，第 k 步可以通过 m_k 种不同方法来完成，那么完成这个任务总共有多少种不同的流程？

$$\boldsymbol{N = m_1 \cdot m_2 \cdots m_n = \prod_{i=1}^{n} m_i}$$

乘法原理适用的情境有以下特点：

- 完成任务的步骤有先后顺序，且每一步对完成任务都是必不可少的。
- 完成每个步骤的方法具有独立性，不受其它步骤的方法的影响。

例 题

1）假设我们要从甲市到乙市，只有陆路和水路两种方式。走陆路总共有 3 条线路，走水路总共有 2 条线路。请问：我们从甲市到乙市总共有多少种不同的线路?

解： $N = 3 + 2 = 5$　　　　【加法原理】

2）假设我们要从甲市到乙市，必须在丙市进行中转，而从甲市到丙市共有 3 条线路，从丙市到乙市共有 2 条线路。请问：我们从甲市到乙市总共有多少种不同的线路?

解： $N = 3 \times 2 = 6$　　　　【乘法原理】

排列与组合

排列组合的核心是计算给定要求的排列或者组合可能出现的情况总数。

$$A_n^m$$

VS

$$C_n^m$$

排列（arrangement 或 permutation）是指从给定个数的元素中取出指定个数的元素进行排序。

从n个不同元素中，任取m个不同的元素按照一定顺序排成一列，这称为一个“排列”。我们把从n个不同元素中取出m个元素的所有可能排列的个数记作 A_n^m。

组合（combination）是指从给定个数的元素中取出指定个数的元素，但不考虑排序。

从n个不同元素中，任取m个元素合并为一组，这称为一个“组合”。我们把从n个不同元素中取出m个元素的所有可能组合的个数记作 C_n^m。

推导排列公式

例 某位老师购买了 3 个水果（苹果🍎、橙子🍊、西瓜🍉），而班级中有 8 名学生。老师决定将 3 个水果分别赠送给班级中任意 3 名学生。请问：总共有多少种不同的赠送方案?

我们可以将这一情景理解为需要分步完成的一个任务：

1) 这位老师先从 8 名学生中挑选 1 名学生，将🍎送给他。这一步中，老师总共有 8 种可能的选择。
2) 这位老师先从剩下的 7 名学生中挑选 1 名学生，将🍊送给他。这一步中，老师总共有 7 种可能的选择。
3) 这位老师先从剩下的 6 名学生中挑选 1 名学生，将🍉送给他。这一步中，老师总共有 6 种可能的选择。

根据乘法原理，这位老师的赠送方案总共有

$$8 \times 7 \times 6 = 336 \text{种}$$

我们也可以将这一情节理解为“排列”：这位老师从 8 名学生中任意挑选 3 名学生并进行排序，这些学生按顺序获得相应的水果。那么，赠送方案的种数就是排列数 A_8^3。

因此，我们得到

$$A_8^3 = 8 \times 7 \times 6 = 336$$

即 A_8^3 等于以 8 为始的 3 个递减的相邻整数的连乘。

推导排列公式

此时，假设该老师购买了 m 个不同的水果，而班级中有 n 名学生。老师决定将这些水果分别赠送给班级中任意 m 名学生。请问：总共有多少种不同的赠送方案?

$$\mathbf{A}_{n}^{m} = n \cdot (n-1) \cdot (n-2) \cdots (n-m+1), \qquad n, m \in \mathbb{N}, m \leq n$$

注意：连乘的最后一项是 $(n-m+1)$ 而非 $(n-m)$。

1808年，法国数学家 Christian Kramp 发明了**阶乘**（factorial）符号 !，并定义

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$$

上文展示的排列公式可以用阶乘表示为：

$$\mathbf{A}_{n}^{m} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-m+1) \cdot (n-m) \cdot (n-m-1) \cdots 1}{(n-m) \cdot (n-m-1) \cdots 1} = \frac{n!}{(n-m)!}$$

推导组合公式

例 某位老师购买了 3 个相同的苹果🍎🍎🍎，而班级中有 8 名学生。老师决定将 3 个🍎分别赠送给班级中任意 3 名学生。请问：总共有多少种不同的赠送方案?

假如老师已经选定 3 名同学（例如，甲、乙、丙）获得水果。在**排列**问题（3 个不同的水果）中，老师有多少种赠送方案?

$$A_3^3 = 3 \times 2 \times 1 = 6\text{ 种}$$

我们可以轻易地列出所有可能的排列：

甲乙丙、甲丙乙、乙甲丙、乙丙甲、丙甲乙、丙乙甲

然而，在**组合**问题（3 个相同的水果）中，以上 6 种排列算作同一种赠送方案，因为不论甲、乙、丙的顺序如何，他们得到的都是相同的🍎。顺序根本不重要!

事实上，排列是在组合的基础上进行排序：从 8 名学生任意抽选 3 人构成一个组合，共有 C_8^3 种组合；然后，对每一个组合内的学生进行排序，在一个组合内共有 A_3^3 种排序方案；总排列数记为 A_8^3。

因此，我们拥有以下关系式：

$$A_3^3 \cdot C_8^3 = A_8^3$$

推导组合公式

此时，假设该老师购买了 m 个相同的水果，而班级中有 n 名学生。老师决定将这些水果分别赠送给班级中任意 m 名学生。请问：总共有多少种不同的赠送方案?

$$\mathbf{C}_{\boldsymbol{n}}^{\boldsymbol{m}}=\binom{\boldsymbol{n}}{\boldsymbol{m}}=\frac{\mathbf{A}_{\boldsymbol{n}}^{\boldsymbol{m}}}{\mathbf{A}_{\boldsymbol{m}}^{\boldsymbol{m}}},\qquad \boldsymbol{n},\boldsymbol{m}\in\mathbb{N},\boldsymbol{m}\leq\boldsymbol{n}$$

组合不考虑排序，因此，我们将排列数除以同一组合中不同排序的总数即可得到组合数。

因为

$$\mathbf{A}_{\boldsymbol{n}}^{\boldsymbol{m}}=\frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{m})!},\qquad \mathbf{A}_{\boldsymbol{m}}^{\boldsymbol{m}}=\boldsymbol{m}!$$

所以

$$\mathbf{C}_{\boldsymbol{n}}^{\boldsymbol{m}}=\frac{\mathbf{A}_{\boldsymbol{n}}^{\boldsymbol{m}}}{\mathbf{A}_{\boldsymbol{m}}^{\boldsymbol{m}}}=\frac{\mathbf{A}_{\boldsymbol{n}}^{\boldsymbol{m}}}{\boldsymbol{m}!}=\frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{m})!\cdot\boldsymbol{m}!}$$

组合数的性质

（1）等一性：　$\mathrm{C}_n^0=\mathrm{C}_n^n=1$

（2）对称性：　$\mathrm{C}_n^m=\mathrm{C}_n^{n-m}$

（3）递推性：　$\mathrm{C}_n^m=\mathrm{C}_{n-1}^{m-1}+\mathrm{C}_{n-1}^m$

练习：证明组合数的这三个性质成立。

练习答案

对称性的证明：$C_n^{n-m}=\frac{A_n^{n-m}}{A_{n-m}^{n-m}}=\frac{n!}{[n-(n-m)]!\cdot(n-m)!}=\frac{n!}{m!\cdot(n-m)!}=C_n^m$

递推性的证明：$C_{n-1}^{m-1}+C_{n-1}^{m}=\frac{(n-1)!}{[(n-1)-(m-1)]!\cdot(m-1)!}+\frac{(n-1)!}{[(n-1)-m]!\cdot m!}$

$$=\frac{(n-1)!}{(n-m)!\cdot(m-1)!}+\frac{(n-1)!}{(n-m-1)!\cdot m!}$$

$$=\frac{n!}{(n-m)!\cdot m!}\cdot\frac{m}{n}+\frac{n!}{(n-m)!\cdot m!}\cdot\frac{n-m}{n}$$

$$=\frac{n!}{(n-m)!\cdot m!}\cdot\left(\frac{m}{n}+\frac{n-m}{n}\right)$$

$$=\frac{n!}{(n-m)!\cdot m!}=C_n^m$$

组合数求和公式

严格的数学证明需要使用数学归纳法（mathematical induction），在此不做展示。

$$\mathbf{C}_n^0 + \mathbf{C}_n^1 + \mathbf{C}_n^2 + \cdots + \mathbf{C}_n^n = 2^n$$

C_n^0 是从 n 个球中抽取 0 个球的组合数，C_n^1 是从 n 个球中抽取 1 个球的组合数，……，C_n^n 是从 n 个球中抽取 n 个球的组合数。这意味着，公式左侧等于从 n 个球中抽取任意个数的球时的组合数。

- 对于第 1 个球，我们可以抽取，也可以不抽取，所以有 2 种情况；
- 对于第 2 个球，我们可以抽取，也可以不抽取，所以有 2 种情况；
- ……
- 对于第 n 个球，我们可以抽取，也可以不抽取，所以有 2 种情况。

因此，我们一共有

$$\underbrace{2 \times 2 \times 2 \times \cdots \times 2}_{n\text{ 个 }2\text{ 相乘}} = 2^n \text{ 种组合}$$

软件知识

在 Microsoft Excel 中，我们可以

- 使用 FACT( ) 函数计算阶乘；
- 使用 PERMUT( ) 函数计算排列数；
- 使用 COMBIN( ) 函数计算组合数。

在 R 中，我们可以

- 使用软件自带的 factorial( ) 函数计算阶乘；
- 使用 combinat 包中的 combn( ) 函数计算排列数；
- 使用 combinat 包中的 permn( ) 函数计算组合数。

在 Python 中，我们可以

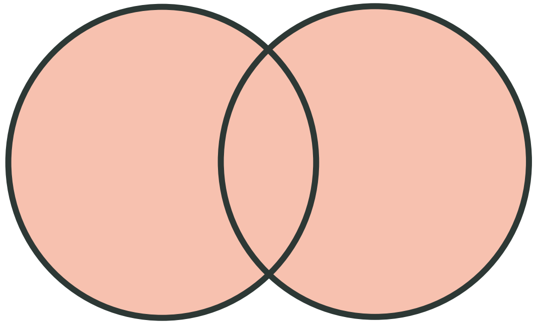
- 使用 math 包中的 factorial( ) 函数计算阶乘；
- 使用 itertools 包中的 permutations 程序列举出所有排列；
- 使用 itertools 包中的 combinations 程序列举出所有组合。

Section 3

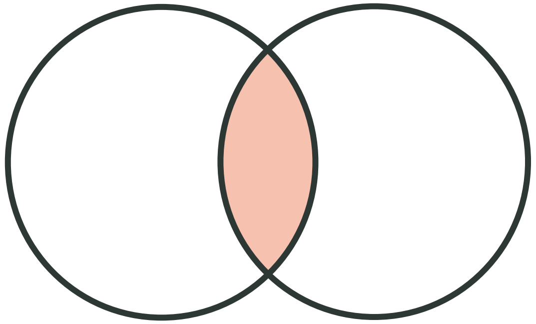
概率的三大公理

非负性 & 归一性 & 可加性

集合运算律



$A \cup B$



$A \cap B$

交换律： $A \cup B = B \cup A$，$A \cap B = B \cap A$

结合律： $A \cup (B \cup C) = (A \cup B) \cup C$，$A \cap (B \cap C) = (A \cap B) \cap C$

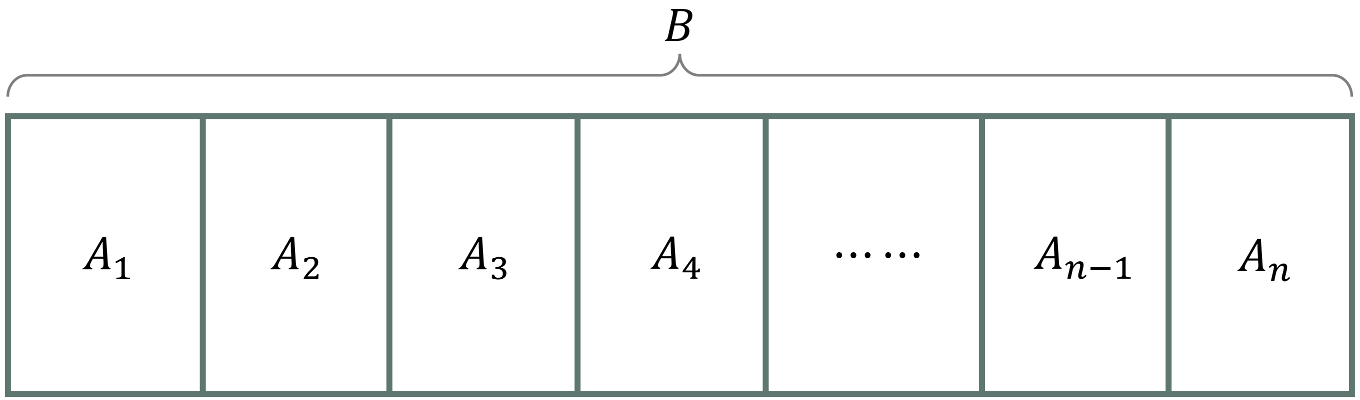
分配律： $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$，$A \cup (B \cap C) = (A \cap B) \cap (A \cup C)$

集合关系的定义

如果集合 A 与集合 B 满足 $A \cap B = \emptyset$，我们称集合 A 与集合 B 为**不相交集合**（disjoint sets，也称“并查集”）。如果集合 $A_1, A_2, \ldots, A_n$ 满足 $A_i \cap A_j = \emptyset\ \forall i \neq j$，我们称集合 $A_1, A_2, \ldots, A_n$ 为**不相交集合**。

如果集合 $A_1, A_2, \ldots, A_n$ 满足 $\bigcup_{i=1}^{n} A_i = B$，我们称集合$A_1, A_2, \ldots, A_n$为集合 B 的**完备集**（exhaustive collection）。

如果集合 $A_1, A_2, \ldots, A_n$ 既是不相交集合，也是集合 B 的完备集，那么我们称集合 $A_1, A_2, \ldots, A_n$ 为集合 B 的**划分**（partition）。



概率论的基石

试验（experiment）：可以在相同条件下重复进行且能够产生随机现象的行为。

结果（outcome）：一次试验产生的特定现象。

样本空间（sample space）：一个以某个试验所有的可能结果为元素的集合，一般用拉丁字母 S 或希腊字母 Ω 表示。

事件（event）：样本空间的任意子集，一般用大写拉丁字母（如 A、B、C 等）表示。

概率函数（probability function）：自变量为事件，因变量为实数，并且满足**概率三大公理**的函数，一般用 $\mathrm{P}(\ )$ 或 $\mathrm{Pr}(\ )$ 表示。

$$\mathrm{Pr}: \Omega \rightarrow \mathbb{R}$$

例 将投掷一颗均匀的骰子当作“试验”，落地后朝上的点数为“结果”。那么，样本空间为

$$S=\{1,2,3,4,5,6\}$$

针对该样本空间，我们有多少事件?

$$N=C_6^0+C_6^1+C_6^2+C_6^3+C_6^4+C_6^5+C_6^6=2^6=64$$

具体如下:

$$\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},$$

$$\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\},\{2,3\},\ldots$$

$$\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,2,6\},\{1,3,4\},\ldots$$

$$\{1,2,3,4\},\{1,2,3,5\},\{1,2,3,6\},\{1,2,4,5\},\ldots$$

$$\{1,2,3,4,5\},\{1,2,3,4,6\},\{1,2,3,5,6\},\{1,2,4,5,6\},\ldots$$

$$\{1,2,3,4,5,6\},\emptyset$$

注意：空集是所有集合的子集，空集代表着“**不可能事件**”。

事件的关系

事件（event）：样本空间的子集，一般用大写拉丁字母（如 A、B、C 等）表示。

互斥事件（disjoint events）：当事件 A 和事件 B 不可能同时发生时，我们称 A 和 B 为互斥事件。互斥事件的数学定义为

$$A \cap B = \emptyset$$

因此，互斥事件有如下特点：

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

互斥事件的相反情况是“相容事件”——即事件 A 和事件 B 有可能同时发生。

独立事件（independent events）：当事件 A 的发生与否和事件 B 的发生与否不会相互影响时，我们称 A 和 B 为独立事件。独立事件的数学定义为

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

公理（axiom）是推导出一套特定演绎知识的基本假设。公理**不证自明**。其它所有的命题（例如定理、推论、法则）都必须借助这些基本假设才能被证明。

概率三大公理

① 非负性

$\Pr(A) \geq 0 \; \forall \; A \subseteq S$

② 归一性

$\Pr(S) = 1$

③ 可加性

如果 $\cap_{i=1}^{\infty} A_i = \emptyset$，那么

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

概率三大公理

① **非负性**：$\Pr(A) \geq 0 \ \forall \ A \subseteq S$

② **归一性**：$\Pr(S) = 1$

③ **可加性**：如果 $\cup_{i=1}^{\infty} A_i = \emptyset$，那么 $\Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{n} \Pr(A_i)$

根据概率三大公理，我们可以推导出：

- $\Pr(A) \leq 1 \ \forall A \subseteq S$；
- $\Pr(\emptyset) = 0$；
- $\Pr(A) = 1 - \Pr(\complement_S A) \ \forall A \subseteq S$；
- 如果 $A \subseteq B$，那么 $\Pr(A) \leq \Pr(B)$；
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$；
- 布尔不等式（Boole’s Inequality）：$\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$；
- 邦佛洛尼不等式（Bonferroni’s Inequality）：$\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1$

该等式可以通过韦恩图（Venn diagram）理解，而严格的数学证明需要借助“划分定理”（Partition Theorem），在此不做展示。

Section 4

条件概率的定义

语 言 定 义

假设 A 和 B 是我们正在研究的两个事件。

$\Pr(A \cap B)$

联合概率（joint probability）

两个事件 A 和 B 同时发生的概率

$\Pr(A), \Pr(B)$

边际概率（marginal probability）

单个事件（A 或 B ）发生的概率

$\Pr(A|B)$

条件概率（conditional probability）

事件 A 在事件 B 已发生的条件下发生的概率

数 学 定 义

假设 A 和 B 是我们正在研究的两个事件，A 在 B 发生的条件下发生的概率为

$$\boxed{\mathbf{Pr}(\boldsymbol{A}|\boldsymbol{B})} = \frac{\boxed{\mathbf{Pr}(\boldsymbol{A} \cap \boldsymbol{B})}}{\boxed{\mathbf{Pr}(\boldsymbol{B})}}$$

联合概率

边际概率

条件概率

注意：这个定义要求 $\Pr(B) > 0$，即事件 B 发生的概率必须大于 0。

例 题 (Sandholm & Saranti, 2018)

一家咖啡连锁店最近开发了一种新的主食用以搭配它的咖啡饮品。下表记录了最近 200 名顾客的选择。

	点了主食	没点主食	总和
点了饮料	50	140	190
没点饮料	10	0	10
总和	60	140	200

问题：

1) 从这 200 名顾客中随机选取一名，这名顾客同时购买了主食和饮料的概率是多少？
2) 如果我们从中随机选择一名购买了饮料的顾客，这名顾客恰好也购买了主食的概率是多少？
3) 如果我们从中随机选择一名购买了主食的顾客，这名顾客没有购买饮料的概率是多少？

解：

我们可以用 F 表示某位顾客购买了主食，用 D 表示某位顾客购买了饮料。

1）$\Pr(F \cap D) = \frac{50}{200} = \frac{1}{4}$

2）$\Pr(F|D) = \frac{\Pr(F \cap D)}{\Pr(D)} = \frac{\frac{50}{200}}{\frac{190}{200}} = \frac{5}{19}$

3）$\Pr(D^c|F) = \frac{\Pr(D^c \cap F)}{\Pr(F)} = \frac{\frac{10}{200}}{\frac{60}{200}} = \frac{1}{6}$

Section 5

贝叶斯法则

贝叶斯法则

条件概率的数学定义式为：

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad 和 \quad \Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

根据以上式子，我们可以推出

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B) = \Pr(B|A) \cdot \Pr(A)$$

$$联合概率 = 条件概率 \times 边际概率$$

我们可以以一种新的形式书写条件概率的定义式：

$$\Pr(A|B) = \frac{\mathbf{Pr(\boldsymbol{B}|\boldsymbol{A}) \cdot Pr(\boldsymbol{A})}}{\Pr(B)} \quad 和 \quad \Pr(B|A) = \frac{\mathbf{Pr(\boldsymbol{A}|\boldsymbol{B}) \cdot Pr(\boldsymbol{B})}}{\Pr(A)}$$

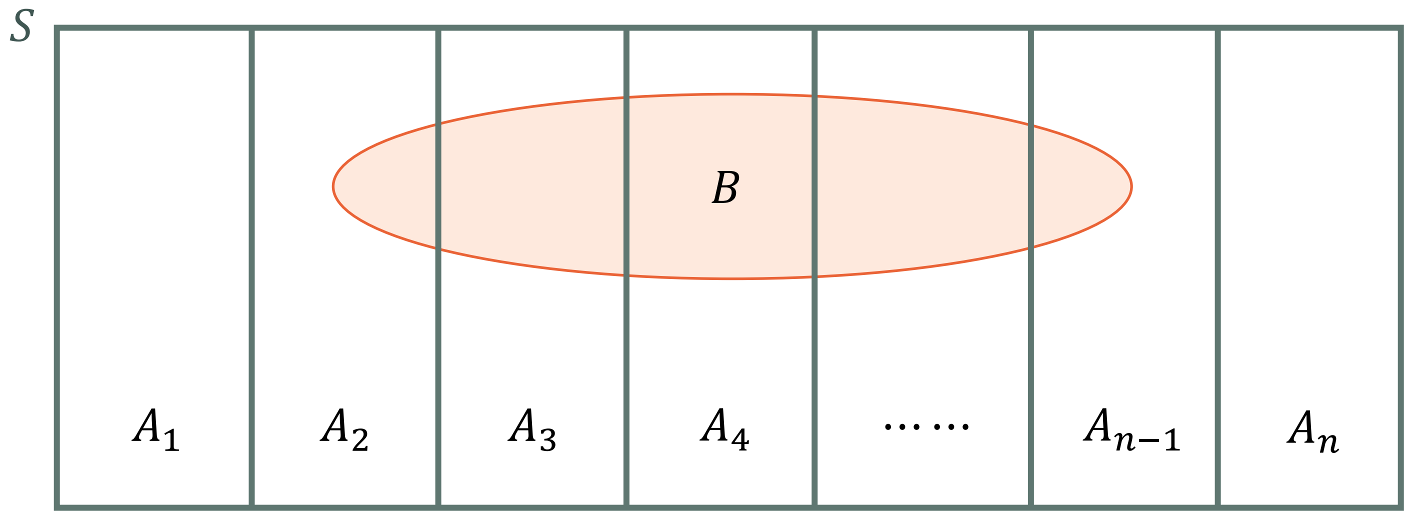
这被称为“贝叶斯法则”（Bayes’ Rule）或“贝叶斯定理”（Bayes’ Theorem）。

全概率定理

假设 A_i（$i = 1, 2, \ldots, n$）和 B 是同一样本空间 S 中的事件。如果 $A_1, A_2, \ldots, A_n$ 是样本空间 S 的**划分**，那么我们有

$$\mathbf{Pr}(\boldsymbol{B}) = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \mathbf{Pr}(\boldsymbol{B} \cap \boldsymbol{A}_{\boldsymbol{i}})$$

这个公式被称为“**全概率定理**”（Law of Total Probability，也称作“**全概率公式**”），它告诉我们任一事件 B 发生的概率等于 B 与其所处样本空间的每个划分事件同时发生的概率的总和。



全概率定理与贝叶斯法则

$$\mathbf{Pr}(\boldsymbol{B}) = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \mathbf{Pr}(\boldsymbol{B} \cap \boldsymbol{A_i})$$

用 A^c 表示 A 在样本空间 S 内的补集，我们可以写出一个“简化版”全概率定理：

$$\mathbf{Pr}(\boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{B} \cap \boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B} \cap \boldsymbol{A^c})$$

根据条件概率的定义式，我们有 $\Pr(B \cap A_i) = \Pr(B|A_i) \cdot \Pr(A_i)$。因此，全概率定理也可以写作

$$\mathbf{Pr}(\boldsymbol{B}) = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A_i}) \cdot \mathbf{Pr}(\boldsymbol{A_i}) \quad \text{或} \quad \mathbf{Pr}(\boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A}) \cdot \mathbf{Pr}(\boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A^c}) \cdot \mathbf{Pr}(\boldsymbol{A^c})$$

这样，贝叶斯法则可以被表示为：

$$\mathbf{Pr}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A}) \cdot \mathbf{Pr}(\boldsymbol{A})}{\mathbf{Pr}(\boldsymbol{B})} = \frac{\mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A}) \cdot \mathbf{Pr}(\boldsymbol{A})}{\mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A}) \cdot \mathbf{Pr}(\boldsymbol{A}) + \mathbf{Pr}(\boldsymbol{B}|\boldsymbol{A^c}) \cdot \mathbf{Pr}(\boldsymbol{A^c})}$$

例 题 (Zhang & Lim, 2021)

当机器在传送加密信息时，难免会发生一些错误。例如，摩斯代码使用“点”（dot）和“线”（dash）来传送信息，传送过程中一些“点”可能被误传为“线”，或者反之。假设一串摩斯代码中，点和线出现次数的比值为 $3:4$，而一个点被误传为一条线的概率为 $\frac{1}{8}$，反之亦然。请问，当我们接收到一个点时，这个点不是被误传的概率为多少?

解:

令 $SDot$ 表示“一个点被发送”，$RDot$ 表示“一个点被接收”，$SDash$ 表示“一条线被发送”，$RDash$ 表示“一条线被接收”。

根据贝叶斯法则和全概率定理，我们有

$$\Pr(SDot|RDot) = \frac{\Pr(RDot|SDot)\cdot\Pr(SDot)}{\Pr(RDot)} = \frac{\Pr(RDot|SDot)\cdot\Pr(SDot)}{\Pr(RDot|SDot)\cdot\Pr(SDot)+\Pr(RDot|SDash)\cdot\Pr(SDash)}$$

根据题意，可得

$$\Pr(SDot) = \frac{3}{7} \qquad \Pr(SDash) = \frac{4}{7} \qquad \Pr(RDot|SDot) = 1-\frac{1}{8} = \frac{7}{8} \qquad \Pr(RDot|SDash) = \frac{1}{8}$$

因此，

$$\Pr(SDot|RDot) = \frac{\frac{7}{8}\times\frac{3}{7}}{\frac{7}{8}\times\frac{3}{7}+\frac{1}{8}\times\frac{4}{7}} = \frac{21}{25}$$

What did we learn?

概率的定义

古典派 VS 频率派 VS 主观派

计数基本法则

排列 A_n^m VS 组合 C_n^m

概率三大公理

非负性 & 归一性 & 可加性

条件概率

定义 & 贝叶斯法则 & 全概率定理

感谢聆听

THANK YOU!