



概率论 & 数理统计

Probability and Mathematical Statistics

关舒文
华南理工大学

Latest Update : 2020 年 6 月 7 日

符号说明

- $A \cap B, AB$ 事件 A 与事件 B 的积事件
- $A \cap B = \emptyset$ 事件 A 与 B 互不相容 (互斥)
- $A \cup B$ 事件 A 与事件 B 的和事件
- \overline{A} 事件 A 的对立事件 (逆事件)
- $A - B, A \setminus B$ 事件 A 与事件 B 的差事件
- $A \subset B$ 事件 B 包含事件 A

-  Color #0097e6
-  Color #7158e2
-  Color #007500
-  Color #ED4C67
-  Color #EA7500
-  Color #1289A7
-  Color #EA2027
-  Color #bf0040
-  Color #833471
-  Color #006266
-  Color #5758BB

目录

符号说明	i
1 概率论的基本概念	1
1.1 随机现象与随机试验	2
1.2 样本空间, 随机事件	2
1.2.1 样本空间	2
1.2.2 随机事件	2
1.2.3 事件间的关系与事件的运算	3
1.3 频率与概率	4
1.3.1 频率	4
1.3.2 概率	5
1.4 独立试验概型	7
附录	8
索引	8



第 1 章 概率论的基本概念

定义 1.0.1 概率论研究的对象

1. **确定性现象** 满足一定条件时, 该类现象的结果可以预见 (并不一定所有内容都可预测, 结果部分具有确定性, 部分具有偶然性);
2. **偶然性现象** 结果无法预见的现象 (部分现象可研究, 部分现象仍是临界问题, 目前无法研究, 但可以找到其特征);
3. **随机现象** 一部分偶然性现象可以在基本相同的条件下 **重复观测或重复试验**, 并且随着 **观测或试验次数的增多**, 不同结果的出现次数呈现出**统计规律**.

在大量的重复试验或观察中所呈现的固有规律性, 称为**统计规律性**. 概率论是研究随机现象的统计规律的科学.

推论 1.0.1 随机现象的判别标准

通过判断事件是否可以在基本相同的条件下重复观测或重复试验, 并且随着观测或试验次数的增多, 不同结果的是否会出现次数呈现出统计规律.

定义 1.0.2 概率论的特点

- 用确定的数学去研究随机现象;
- 研究非确定的现象: 用于数理金融, 控制论, 寿险精算, 统计物理学等, 由于科技手段的限制, 或者经济成本的考量, 很多事件适合从随机角度去研究;
- 以确定的数学为工具: 排列组合, 微积分, 线性代数等, 本质上是要利用 **数学工具** 来量化这种非确定性的程度.

例题 1.0.1

甲、乙两人赌技相同, 各出赌注 300 元. 约定: 谁先胜三局, 则拿走全部 600 元. 现已赌了三局, 甲二胜一负, 因故要中止赌博, 问这 600 元要如何分才算公平?

解法 1: 以已经获胜的局数作为分赌注的依据甲: $600 \times 2/3 = 400$, 乙: $600 \times 1/3 = 200$.
存在问题: 不具有可扩展性, 若只赌一盘, 则不公平.

解法 2: 设想继续赌下去, 甲、乙获胜的可能性分别是多大, 以此作为分赌注的依据. 最多还需要两局, 结果为四种情况: 甲甲, 甲乙, 乙甲, 乙乙. 由于甲乙赌技相同, 这四种情况出现的可能性相等. 前三种情况下甲最终获胜. 因此甲、乙最终获胜的可能性大小为 3 : 1, 即甲分得 450 元, 乙分得 150 元.

推广: 将这一分配方案进行扩展. 假设只赌了 1 局并且甲获胜, 该如何分配?

将 4 局全部比完, 可罗列出以下情形 *cf. Fig 1.1*:

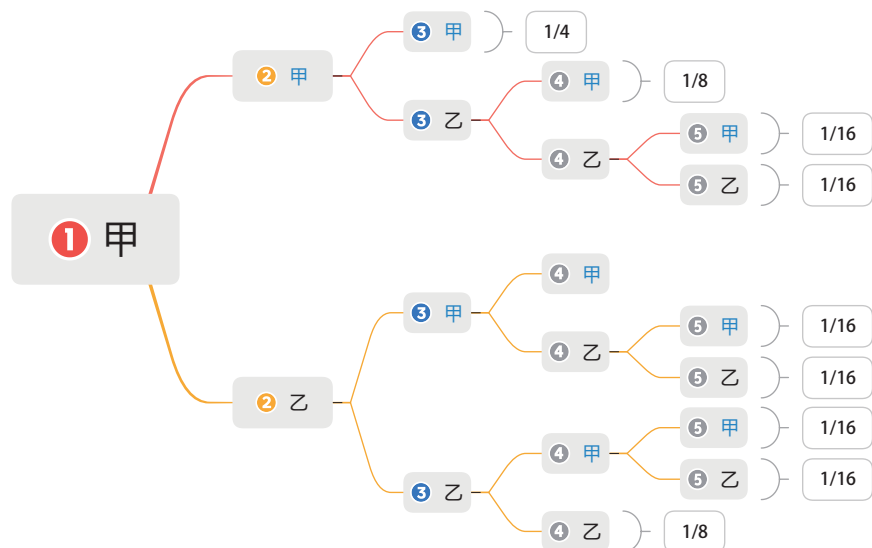


图 1.1: 可能发生的情形

定义 1.0.3

- 概率论 (实践): 对随机现象有基本认知的前提下, 进行演绎推理;
- 数理统计 (理论): 试图通过试验来认知随机现象, 处理问题的思路往往来自概率论的有关结果.

1.1 随机现象与随机试验**定义 1.1.1 随机试验**

对随机现象的一次观测或实际实验, 统称为**随机试验**, 其具有以下特点 (判断依据):

1. 可以在相同的条件下重复进行;
2. 每次实验的可能结果不止一个, 并且能够事先明确实验的所有可能结果;
3. 进行一次试验之前不能确定哪一个结果会出现.

1.2 样本空间, 随机事件**1.2.1 样本空间****定义 1.2.1 样本空间**

我们将随机试验 E 的**所有可能结果**组成的集合称为 E 的**样本空间**, 记为 S . 其中样本空间的元素, 即 E 的每个结果, 称为样本点.

1.2.2 随机事件**定义 1.2.2 随机事件**

试验 E 的样本空间 S 的子集为 E 的**随机事件**或称**事件**^①. 在每次实验中, 当且仅当这一中的样本点出现时, 称这一**事件发生**.

特别地, 有一个样本点组成的**单点集**, 称为**基本事件**.

定义 1.2.3

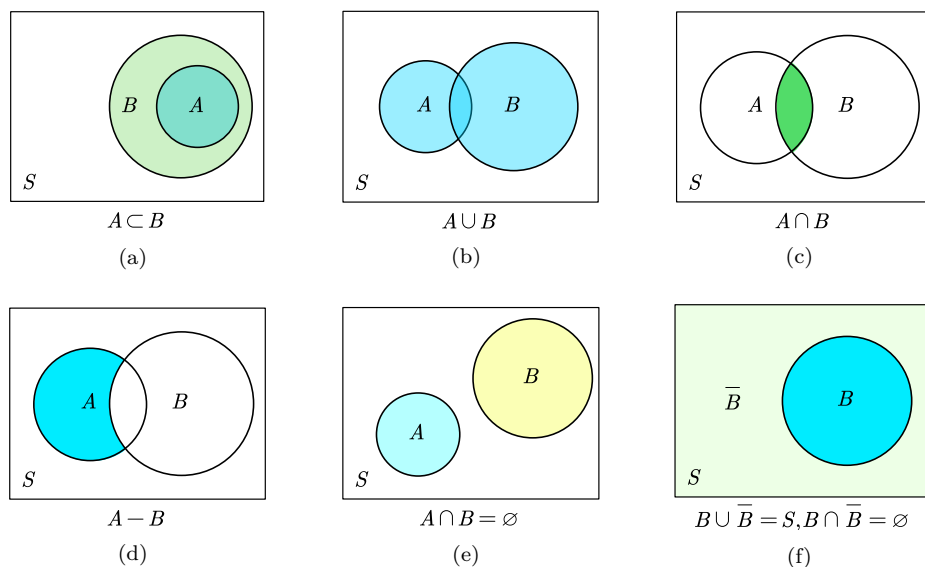
对于样本空间 S , 其包含的所有样本点即 S 是样本空间的全集, 也为其最大的子集. 空集 \emptyset 不包含任何样本点, 它也作为样本空间的子集. 根据事件的定义1.2.2:

- S 称为**必然事件**;
- \emptyset 称为**不可能事件**.

1.2.3 事件间的关系与事件的运算

事件是一个集合, 事件间的关系与运算自然需要并遵循集合论. 下面我们根据“事件发生的含义”以及集合的关系与运算, 给出他们在概率论中的定义:

定义 1.2.4



左图可以直观地表示上述各种关系. 长方形所围绕的空间为**样本空间 S** , 圆 A 与圆 B 分别表示事件 A 于事件 B , 阴影表示**事件与事件间**以及**事件与样本空间**间的关系.

图 1.2: 事件间的关系韦恩图

设试验 E 的样本空间为 S , 而 $A, B, A_k (k = 1, 2, \dots)$ 是 S 的子集, 则有:

1. 若 $A \subset B$, 则称**事件 B 包含事件 A** , **事件 A 发生必导致事件 B 发生**, cf. Fig 1.2a. 特别地, 若 $A \subset B$ 且 $B \subset A$, 即 $A = B$, 则称事件 A 与事件 B 相等.
2. 事件 $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$ 称为事件 A 与事件 B 的**和事件**, cf. Fig 1.2b. **当且仅当 A, B 中至少有一件发生时**, 事件 $A \cup B$ 发生.

类似地, 我们称 $\bigcup_{k=1}^n A_k$ 为 n 个事件 A_1, A_2, \dots, A_n 的**和事件**, 称 $\bigcup_{k=1}^{\infty} A_k$ 为可列个事件 A_1, A_2, \dots 的**和事件**.

3. 事件 $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$ 称为事件 A 与事件 B 的**积事件**, cf. Fig 1.2c. **当且仅当 A, B 同时发生时**, 事件 $A \cap B$ 发生.

$A \cap B$ 又记作 AB

^①严格地说, 事件是指 S 中满足某些条件的子集当且仅当 S 是由有限个元素或由可列无限个元素组成的, 每个子集才可以看作一个事件.

类似地, 我们称 $\bigcap_{k=1}^n A_k$ 为 n 个事件 A_1, A_2, \dots, A_n 的 **积事件**; 称 $\bigcap_{k=1}^{\infty} A_k$ 为可列个事件 A_1, A_2, \dots 的 **积事件**.

$A - B$ 又记作 $A \setminus B$

4. 事件 $A - B = \{x \mid x \in A \text{ and } x \notin B\}$ 称为事件 A 与事件 B 的**差事件**, cf. Fig 1.2d, **当且仅当 A 发生 B 不发生时**, 事件 $A - B$ 发生.
5. 若 $A \cap B = \emptyset$, 则称事件 A 与 B **互不相容的或互斥的**, cf. Fig 1.2e. **事件 A 与事件 B 不能同时发生**. 基本事件是两两互不相容的.
6. 若 $A \cup B = S$ 且 $A \cap B = \emptyset$, 则称事件 A 与事件 B 互为**逆事件**或称**对立事件**, cf. Fig 1.2f. 这指的是对每次试验而言, **事件 A, B 中必有一个发生, 且仅有一个发生**. A 的对立事件可以记为 \bar{A} 且 $\bar{A} = S - A$.

定理 1.2.1

进行事件运算时我们常用到以下定律. 设 A, B, C 为事件, 则有

1. 交换律:

$$A \cup B = B \cup A;$$

$$A \cap B = B \cap A$$

2. 结合律:

$$A \cup (B \cup C) = (A \cup B) \cup C;$$

$$A \cap (B \cap C) = (A \cap B) \cap C.$$

3. **分配律**: 若 $(B \cup C) \cap A$ 发生, 则 A 发生且 B 和 C 至少一个发生, 则 $A \cap B$ 或 $A \cap C$ 发生. 于是有 $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. 反过来, 若 $(A \cap B) \cup (A \cap C)$ 发生, 则有 $A \cap B$ 或 $A \cap C$ 中至少有一个发生, 从而 $A \cap (B \cup C)$ 一定发生, 由此可得 $A \cap (B \cup C) \supset (A \cap B) \cup (A \cap C)$, 可得:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

因为 $\overline{A \cup B}$ 表示“ A 和 B 中至少一个发生的否命题”, 即 A 和 B 均不发生, 由此可理解德摩根律. 同时可以通过 Fig 1.2b 进一步理解与记忆 **德摩根律**.

4. 德摩根律 (对偶律):

$$\overline{A \cup B} = \bar{A} \cap \bar{B};$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}.$$

1.3 频率与概率

为了找到一个合适的数来表征事件在一次实验中发生的可能性大小, 首先我们引入 **频率**, 它描述了事件发生的 **频繁程度**, 进而可以引出表征事件在一次试验中发生的可能性大小的数——**概率**.

1.3.1 频率

观察右式可知事件 A 发生的频率是它 **发生的次数与试验次数之比**

定义 1.3.1 频率的定义

在相同的条件下, 进行了 n 次试验, 在这 n 次试验中, 事件 A 发生的次数 n_A 称为事件 A 发生的**频数**. 比值 n_A/n 称为事件 A 发生的**频率**, 记作 $f_n(A)$, 即有, 其大小

表示事件发生的频繁程度, 频率越大, 事件 A 发生越频繁, 在一次实验中发生的可能性越大. 因而我们发现频率可以反应事件发生的可能性大小.

$$f_n(A) = \frac{n_A}{n}.$$

性质 1.3.1 频率的性质

根据上述定义, 不难发现以下性质:

1. $0 \leq f_n(A) \leq 1$;
2. $f_n(S) = 1$;
3. 若 A_1, A_2, \dots, A_k 是两两互不相容的事件, 则

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

1.3.2 概率

大量试验证实, 当重复试验的次数 n 逐渐增大时, 频率 $f_n(A)$ 呈现出稳定性并逐渐稳定于某个常数. 这种频率稳定性即所说的统计规律性. 由于大量的试验成本巨大, 为了理论研究的需要, 我们从频率的稳定性与频率的性质得到启发, 给出如下能够表征事件发生可能性大小的概率的定义.

定义 1.3.2 概率的统计定义

在对某一随机现象进行大量的重复观测时, 随着观测次数 n 的增加, 某一事件 A 的频率 (cf. Def 1.3.1):

$$f_n(A) = \frac{n_A}{n}$$

会在某一个数 p (其中 $p \in [0, 1]$) 附近摆动, 我们称 p 为事件 A 发生频率的稳定值. 该值能够反应事件 A 发生的可能性大小, 我们称之为 A 发生的概率.

等可能概型 (古典概型)

定义 1.3.3 古典概型

我们将具有以下特点:

1. 试验的样本元素只包含有限个元素;
2. 试验中每个基本事件发生的可能性相同,

的实验称为等可能概型 (古典概型)

定义 1.3.4 概率的古典定义

设随机试验只有 $\omega_1, \omega_2, \dots, \omega_n$ 等 n 个结果, 每次实验有且仅有其中的一个发生, 并且每个结果发生的可能性大小相同, 我们可以定义事件 A 发生的概率为

$$P(A) = \frac{n_A}{n},$$

其中 n_A 称为事件 A 包含于 $\{\omega_1, \omega_2, \dots, \omega_n\}$ 中元素的个数.

定理 1.3.1 古典概型中事件 A 的概率的计算方法

概率的古典定义是一种计算概率 (等可能性) 的数学模型, 常称为古典概型.

若事件 A 包含 k 个基本事件, 即 $A = \{e_{i_1}\} \cup \{e_{i_2}\} \cup \cdots \cup \{e_{i_k}\}$, 这里 i_1, i_2, \cdots, i_k 是 $1, 2, \cdots, n$ 中某 k 个不同的数, 则有

$$P(A) = \sum_{j=1}^k P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A \text{ 包含的基本事件数}}{S \text{ 中基本事件的总数}}. \quad (1.3.1)$$

(1.3.1) 式称为等可能概型中事件 A 的概率计算公式

例题 1.3.1 放回抽样与不放回抽样

已知袋中有 a 只白球, b 只红球, k 个人依次在袋中取一只球分别作放回抽样和不放回抽样, 各求第 i ($i = 1, 2, \cdots, k$) 人取到白球 (事件 B) 的概率. ($k \leq a + b$)

解答:

(1) 放回抽样: 每个人可能性显然相等, 有

$$P(B) = \frac{a}{a+b}.$$

(2) 不放回抽样: 每个人各取一个球, 每种取法是一个基本事件, 共有

$$\underbrace{(a+b)}_{\text{第一个人}} \underbrace{(a+b-1)}_{\text{第二个人}} \cdots \underbrace{(a+b-k+1)}_{\text{第 } k \text{ 个人}} = A_{a+b}^k$$

个基本事件. 由对称性可知每个基本事件发生的可能性相同. 当事件 B 发生时, 第 i 个人取的应是白球, 它可能是 a 中的某一只, 所以有 a 种取法. 其余被取的 $k-1$ 只球可以是其余的 $a+b-1$ 只球中的任意 $k-1$ 只, 共有

$$(a+b-1)(a+b-2) \cdots [a+b-1-(k-1)+1] = A_{a+b-1}^{k-1}$$

种取法, 于是事件 B 中包含 $a \cdot A_{a+b-1}^{k-1}$ 个基本事件, 故由 (1.3.1) 式, 可得

$$P(B) = a \cdot A_{a+b-1}^{k-1} / A_{a+b}^k = \frac{a}{a+b}.$$

值得注意的是 $P(B)$ 与 i 无关, 即 k 个人取球, 尽管取球的先后顺序不一样, 但计算发现, 每个人取到白球的概率一致, 大家机会相同. 此处可以提出一个有意思的理解方法, 在后者未知前者取球结果的情况下, 每人各自取球等同于随机分发, 概率显然相等. 若后者已知前者的取球结果, 并且事件 B 已发生, 那么后者取到白球的可能性下降, 反之上升, 由于前者的取球结果必然为白球或红球, 即不能肯定前者取得的是白球 (前者取到白球并非必然事件), 那么对于后者来说这种可能上升可能下降的情况实际的总的取到白球概率仍然未变. 这样亦可以解释到为何放回抽样与不放回抽样的概率结果相同.

例题 1.3.2 超几何分布

设有 N 件产品, 其中有 D 件是次品, 今从中任取 n 件, 问其中恰有 k ($k \leq D$) 件次品的概率是多少?

解答: 在 N 件产品中抽取 n 件 (不放回抽样), 所有可能的取法共有 $\binom{N}{n}$ ①, 每一种取法为一基本事件, 且由对称性知每个基本事件发生的可能性相同. 又因在 D 件

① 对于任意实数 a 以及非负整数 r , 定义

$$\binom{a}{r} = \frac{a(a-1) \cdots (a-r+1)}{r!},$$

特殊的, $\binom{a}{0} = 1$. 当 a 为正整数, 且 $r \leq a$ 时, $\binom{a}{r}$ 即为组合数, 即

$$\binom{a}{r} = C_a^r \quad (r \leq a).$$

此处只需要保证第 i 个人取到白球 (特定一个人取到白球), 其他人任意取球即可.

次品中选取 n 件, 所有可能的取法有 $\binom{D}{k}$ 种. 在 $N - D$ 件正品中取 $n - k$ 件所有可能的取法有 $\binom{N - D}{n - k}$ 种. 有乘法原理知在 N 件产品中取 n 件, 其中恰有 k 件次品的取法共有 $\binom{D}{k} \binom{N - D}{n - k}$ 种. 于是所求概率为

$$p = \frac{\binom{D}{k} \binom{N - D}{n - k}}{\binom{N}{n}}. \quad (1.3.2)$$

(1.3.2) 式即超几何分布的概率公式.

例题 1.3.3 实际推断原理

某接待站在某一周曾接待过 12 次来访, 已知所有这 12 次接待都是周二和周四进行的, 问是否可以推断接待时间是有规定的.

解答: 假设接待站的接待时间没有规定, 而各来访者在一周的任一天中去接待站是等可能的, 那么发生如题所述事件的概率为:

$$p = \frac{2^{12}}{7^{12}} = 0.000\ 000\ 3.$$

可能性非常小, 故接待时间应该是有规定的.

人们在长期的实践中总结出: 概率很小的事件在一次实验中实际上几乎不发生, 我们称之为实际推断原理.

根据实际推断原理, 现在概率很小的事件在一次试验中居然发生了, 因此我们就有理由怀疑假设的正确性, 从而推断出答案.

几何概型

定义 1.3.5 概率的几何定义

设随机试验是往区域 S 里投点, 点落到 S 中某子区域 G 的可能性大小只与 G 的度量大小有关, 而与 G 的形状和位置无关, 则定义

$$P(\text{点落到子区域} G) = \frac{|G|}{|S|},$$

其中 $|\cdot|$ 表示几何度量, 它可能是长度, 面积等度量.

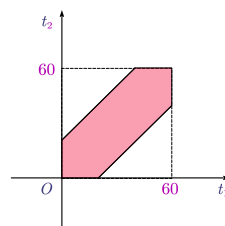
我们称这种计算概率的数学模型为几何概型.

例题 1.3.4 约会问题

两人相约在 7 点到 8 点间在某地会面, 先到者可以等候 20 分钟, 过时即离去. 失球这两人会面的概率.

解答: 两人可能在区间 $[0, 60]$ 的任一点到达, 设 x 和 y 分别是两人到达的时间, 则两人到达的时刻所对应的所有基本事件为二维区域 $D: [0, 60] \rightarrow [0, 60]$ 内的所有点, 如图. 而两人会面当且仅当

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



1.4 独立试验概型

区别于事件独立性, 我们可以提出试验独立的概念.

定义 1.4.1 试验的独立性

设有一组随机试验 E_1, E_2, \dots, E_n . 如果对 E_i 的任意结果 (事件) A_i , 都有

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n),$$

则称这组随机试验相互独立.

事件独立和试验独立的区别, 事件独立指的是同一试验中的事件, 试验独立指的是不同试验的事件.

索引

B

不可能事件, 3

必然事件, 3

J

基本事件, 3

S

事件发生, 3

随机事件, 3

T

统计概率性, 1

Y

样本点, 2

样本空间, 2