

考研概率论

枫聆

2021 年 7 月 12 日

目录

1	随机事件和概率	2
1.1	基本定义	2
1.2	基本性质和运算法则	4
1.3	更深的思考和技巧	7
2	随机变量及其概率分布	8
2.1	随机变量及其分部函数	8
2.2	常用分布	10

随机事件和概率

基本定义

Definition 1.1. 对随机现象进行观察或者实验被称为**随机试验**当且仅当满足以下条件

1. 可以在相同的条件下**重复实验**;
2. 所得的可能结果不止一个, 且所有可能结果都能**事前已知**;
3. 每次具体实验之前**无法预知**出现的结果.

Definition 1.2. **随机试验**的每一可能的结果为被称为**样本点**, 所有**样本点**构成的集合被称为**样本空间**.

Definition 1.3. **样本空间**的任一子集被称为**随机事件**. 其中每个单点集被称为**基本事件**. 事件 Ω 被称为**必然事件**当且仅当每次试验必有 Ω 中某一样本点发生. 特别地, 把空集 \emptyset 称为**不可能事件**.

Definition 1.4. 若事件 A 的发生**必然导致**事件 B 发生, 则称事件 B 包含事件 A , 记为 $B \supset A$. 若 $A \supset B$ 和 $A \subset B$ 同时成立, 则称事件 A 和事件 B 相等, 记为 $A = B$.

Definition 1.5. 给定事件 A 和 B , 它们的交记为 $A \cap B$ 或者 AB , 表示其所有的公共样本点构成的事件. 这样事件的发生, 将导致**事件 A 和 B 同时发生**. 它们的并记为 $A \cup B$, 表示它们所有样本点放在一起构成的事件, 这样的事件发生将导致**至少事件 A 和 B 其中一个发生**.

Definition 1.6. 给定事件 A 和 B , 若它们的交 $AB = \emptyset$, 则称事件 A 和 B **互斥**或者**互不相容**. 若它们的并 $A \cup B = \Omega$, 且 $AB = \emptyset$, 则称事件 A 和 B 为**对立事件**或者**互逆事件**, 记为 $\bar{A} = B$ 或者 $\bar{B} = A$.

Definition 1.7. 给定事件 A 和 B , 它们的差记为 $A - B$, 表示事件 A 有而 B 没有的样本点, 通俗地来讲表示**事件 A 发生而事件 B 不发生**的样本点组成的新事件.

Definition 1.8. 设试验 E 的样本空间为 Ω , real-valued 函数 $P: \mathcal{A} \rightarrow \mathbb{R}$ 被为一个概率函数, 其中 \mathcal{A} 被称为**输入空间**或者**事件空间**, 即样本空间的**幂集**. 当其满足如下条件 (Kolmogorov axioms) 时

1. 对于任意的 $A \in \mathcal{A}$, 有 $P(A) \geq 0$;
2. $P(\Omega) = 1$;
3. 对于一个两两不相交的事件可数序列 A_1, A_2, \dots , 有 $P(\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} P(A_i)$ 成立.

这个 real-value P 其实看做一个代数形式, 其需要满足 3 个公理.

Definition 1.9. 给定事件 A 和 B , 且 $P(A) > 0$, 称

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

为在事件 A 发生的条件下事件 B 发生的条件概率.

Definition 1.10. 若事件 A, B 满足等式

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

则称 A 与 B 相互独立. 从条件概率看两个事件的独立性, 也就是其中一个发生的概率是不会影响另一个发生的概率. 推广至 n 个事件 A_1, \dots, A_n 相互独立, 需要 $\binom{n}{2} + \binom{n}{3} + \dots + \binom{n}{n} = 2^n - n - 1$ 等式成立, 即设任意的 $1 < k \leq n$, 对任意 $1 \leq i_1 \leq \dots \leq i_k \leq n$ 满足等式

$$P(A_{i_1} \dots A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

Definition 1.11. 当试验的样本空间由 n 个有限样本点构成, 且每个样本点的发生具有相同的可能性, 即若事件 A 由 n_A 个样本点组成, 则事件 A 对应的概率为

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

称这样的有限等可能试验中事件 A 的概率 $P(A)$ 为古典型概率.

Definition 1.12. 当试验的样本空间是某区域 (该区域可以是一维, 二维或者三维等等), 以 $L(\Omega)$ 表示其几何度量, $L(\Omega)$ 有限, 且试验结果出现在 Ω 中任何区域的可能性只与该区域几何度量成正比, 事件 A 的样本点所表示的区域为 Ω_A , 则事件 A 的概率为

$$P(A) = \frac{L(\Omega_A)}{L(\Omega)}.$$

称这种样本点个数无限但是其几何度量上的等可能试验中事件 A 的概率 $P(A)$ 为几何型概率.

Definition 1.13. 把一随机试验独立重复做若干次, 即同一事件在各次试验中出现的概率相同. 这个过程称为独立重复试验.

Definition 1.14. 如果每次试验只有两个结果 A 和 \bar{A} , 则称这种试验为伯努利试验, 将伯努利试验独立重复进行 n 次, 称为 n 重伯努利试验. 设在每次试验中, 概率 $P(A) = p$ ($0 < p < 1$), 则在 n 重伯努利试验中事件 A 发生 k 次的概率为

$$C_n^k p^k (1-p)^{n-k}, \quad k = 1, 2, \dots, n,$$

其又称为二项概率公式.

基本性质和运算法则

Proposition 1.15. 事件相关的运算法则

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 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. 对偶律 $\overline{A \cup B} = \bar{A} \cap \bar{B}$; $\overline{A \cap B} = \bar{A} \cup \bar{B}$; $\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i$; $\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i$

Proposition 1.16. 概率分布相关性质

1. $P(\emptyset) = 0$;
2. $P(\bar{A}) = 1 - P(A)$;
3. $A \subset B$, 则 $P(A) \leq P(B)$;
4. $0 \leq P(A) \leq 1$;

证明. (2) 因为 $P(\Omega) = P(A \bar{A}) = P(A) + P(\bar{A})$.

(1) 可以马上通过 (2) 直接得到.

(3) P 满足单调性, 可以根据 Kolmogorov axioms(3) 很自然地可以得到.

(4) $P(\emptyset) \leq P(A) \leq P(\Omega)$.

□

Proposition 1.17. 五大概率公式

1. 加法公式

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

2. 减法公式

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

3. 乘法公式 当 $P(A) > 0$ 时,

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

当 $P(A_1 A_2 \cdots A_{n-1}) > 0$ 时,

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

4. **全概率公式** 设 B_1, B_2, \dots, B_n 满足 $\bigcup_{i=1}^n B_i = \Omega$, $B_i B_j = \emptyset$ 且 $P(B_k) > 0$, $k = 1, 2, \dots, n$, 则对任意事件有

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

其中称满足 $\bigcup_{i=1}^n B_i = \Omega$, $B_i B_j = \emptyset$ 的 B_1, B_2, \dots, B_n 为一个**完备事件组**. 通常把 $P(B_1), P(B_2), \dots, P(B_n)$ 叫做**先验概率**. 全概率公式的意义在于可以将复杂的事件 A 划分为简单互斥事件 AB_1, AB_2, \dots, AB_n , 再结乘法公式计算出 A 的概率.

5. **贝叶斯公式** 设 B_1, B_2, \dots, B_n 满足 $\bigcup_{i=1}^n B_i = \Omega$, $B_i B_j = \emptyset$ 且 $P(A) > 0, P(B_k) > 0$, $k = 1, 2, \dots, n$, 则

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

贝叶斯公式的意义在于在事件 A 已经发生的条件下, 贝叶斯公式可以用来寻找导致 A 发生各种“原因” B_i 的概率. 其中 $P(B_j|A)$ 被称为**后验概率**.

证明. (2) $P(A) = P(AB \cup A\bar{B}) = P(AB) + P(A - B)$.

(1) $P(A \cup B) = P((A - AB) \cup (B - AB) \cup AB)$, 再根据 (2) 有

$$\begin{aligned} P((A - AB) \cup (B - AB) \cup AB) &= P(A) - P(AB) + P(B) - P(AB) + P(AB) \\ &= P(A) + P(B) - P(AB). \end{aligned}$$

实际上这里还是要分类讨论一下当 $A = B$ 的时候.

(3) 条件概率的另一种写法.

(4) $P(A) = P(\bigcup_{i=1}^n AB_i) = \sum_{i=1}^n P(AB_i)$, 再用 (3) 替换一下即可.

(5) $P(B_j|A) = \frac{P(B_j A)}{P(A)}$, 用 (3) 和 (4) 分别替换分子和分母即可. □

下面都是一些高中学过的排列组合的性质.

Definition 1.18. 加法原理 若完成一件事, 有 n 类方式, 第一类方式有 m_1 种解决方法, 第二类方式有 m_2 种解决方法, 如此定义下去即第 i 类方法有 m_i 种解决方法. 那么完成这件事就一共有 $m_1 + m_2 + \dots + m_n$ 种不同的方法.

Definition 1.19. 乘法原理 若完成一件事分成 n 个步骤, 其中第 i 步有 m_i 种不同的方法, 必须依次完成每一步之后才能进行下一步, 那么完成这件事就一共有 $m_1 m_2 \dots m_n$.

Definition 1.20. 排列数公式 从 n 个元素中取 m 个元素出来进行排列, 则不同排列的总数为

$$P_n^m = \frac{n!}{(n-m)!} = n \times (n-1) \times \cdots (n-k+1).$$

特别地, 若 $m = n$ 时, $P_n^n = n!$ 其被称为**全排列**; 若有放回的取, 则不同的排列总数为 n^n ;

Definition 1.21. 组合数公式 从 n 个元素中取 m 个元素组成一组, 即不管其顺序, 则不同的组合总数为

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{P_n^m}{m!}.$$

Proposition 1.22. 如果把 n 个不同的元素分成 k 组 ($1 \leq k \leq n$), 使得第 i 组有 n_i 个元素, 那么 $\sum_{i=1}^k n_i = n$, 组内不考虑元素的排列, 那么不同的分法总数有

$$\frac{n!}{n_1!n_2! \cdots n_k!}.$$

证明. 实际上就是

$$\begin{aligned} \binom{n}{n_1} \binom{n-n_1}{n_2} \cdots \binom{n-n_1-\cdots-n_{k-1}}{n_k} &= \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!((n-n_1)-n_2)!} \cdots \frac{(n-n_1-\cdots-n_{k-1})!}{n_k!((n-n_1-\cdots-n_{k-1})-n_k)!} \\ &= \frac{n!}{n_1!n_2! \cdots n_k!}. \end{aligned}$$

□

Proposition 1.23. 常用的组合数公式

$$C_n^k = C_n^{n-k}$$

$$C_{n+1}^k = C_n^k + C_n^{k-1}$$

$$\sum_{i=0}^n C_n^i = 2^n$$

$$C_{n+m}^k = \sum_{i=0}^k C_n^i C_m^{k-i}$$

证明. (1) 比较 trivial.

(2) 用自然语言来解释, 就是说我在 $n+1$ 个元素里面取 k 等价于我先考虑在 n 个元素里面取 k , 然后现在又来了一个新的元素 e , 考虑多出的取法显然要包括这个 e , 那么现在我只要再去原来 n 个元素里面取 $k-1$ 就够了. 这就是第二个等式的含义. 代数证明就略过了...

(3) 直接考虑 $(1+1)^n$ 的展开式就够了.

(4) 实际上也比较 trivial, 即考虑 k 分别在 m 个元素和 n 元素里面取.

□

更深的思考和技巧

Annotation 1.24. 零概率事件 $P(A) = 0$ 与不可能事件 $A = \emptyset, P(A) = 0$ 不是一个概念.

Annotation 1.25. 如何理解事件之间独立性 事件 A 和 B 独立, 则下面等式成立

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

你可能总想深究其更本质的含义, 可能你会联想到当事件 A 和 B 独立时, 那么 A 和 B 之间有什么关系呢? 实际上并不能确定 A 和 B 有什么特别地确切的关系, 所以你把上述等式理解为在描述一个代数结构就好了, 从这个代数结构可以引发一些有趣的性质 i.e. 例如化简某些运算条件概率, 所以就特别地把这个代数结构提出来了. 但是当 $P(A) > 0, P(B) > 0$ 时, 有这样一种关系: 互斥不独立, 独立不互斥, 证明也是很显然的.

Annotation 1.26. 当 $P(A) = 0$ 时, 如何计算 $P(AB)$? 条件概率的定义中特别指明了 $P(A) > 0$, 那么 $P(A)$ 等于 0 时候, 条件概率 $P(B|A)$ 是未定义的. 那么此时如果我们要计算 $P(AB)$ 应该怎么办呢? 肯定不能使用乘法公式了. 我们可以利用 P 的单调性, 因为 $AB \subseteq A$, 所以 $P(AB) \leq P(A)$, 所以 $P(AB) = 0$.

Annotation 1.27. 取部分对立事件不影响独立性 若 $P(AB) = P(A)P(B)$, 有下面等式成立

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

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

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

我们来证明第一个等式

$$\begin{aligned} P(AB) &= P(A)P(B) \\ &= (1 - P(\bar{A}))(1 - P(\bar{B})) \\ &= 1 - P(\bar{A}) - P(\bar{B}) + P(\bar{A})P(\bar{B}) \end{aligned}$$

$$\begin{aligned} P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cup \bar{B}) &= P(\bar{A})P(\bar{B}) \\ P(\bar{A}\bar{B}) &= P(\bar{A})P(\bar{B}) \end{aligned}$$

其余的等式应用类似的手法来证明. 由此在 A 和 B 相互独立的情况下, 延伸出来一些有用的等式

$$\begin{aligned} P(A \cup B) &= 1 - P(\bar{A}\bar{B}) = 1 - P(\bar{A})P(\bar{B}) \\ P(A - B) &= P(A\bar{B}) = P(A)P(\bar{B}) \end{aligned}$$

这些等式也可以推广为 n 个事件相互独立.

随机变量及其概率分布

随机变量及其分部函数

Definition 2.1. 在样本空间 Ω 上的实值函数 $X = X(\omega), \omega \in \Omega$, 称 $X(\omega)$ 为**随机变量**, 简记为 X .

Annotation 2.2. 随机变量的概念引入是为了通过函数的 image(实数) 来描述 preimage, 即某一样本空间中的样本点, i.e. $P(X < 1)$.

Definition 2.3. 如果一个随机变量的可能取值是有限多个或者可数无穷多个, 则称它为**离散型随机变量**.

Definition 2.4. 设离散型随机变量 X 的可能取值是 $x_k (k = 1, 2, \dots)$, X 取各可能值的概率为

$$P\{X = x_k\} = p_k, k = 1, 2, \dots,$$

称上式为离散型随机变量 X 的概率分部或者分部律, 其中 P 是一个概率函数.

Proposition 2.5. 分部律的性质如下

1. $p_k \geq 0, k = 1, 2, \dots;$

2. $\sum_{k=1}^{\infty} p_k = 1.$

证明. (2) $1 = P[\bigcup_{k=1}^{\infty} \{X = x_k\}] = \sum_{k=1}^{\infty} P\{X = x_k\}$, 这里说明了 $\{X = x_i\} \cap \{X = x_j\} = \emptyset, i \neq j$. □

Definition 2.6. 设 X 是一个随机变量, 对于任意实数 x , 函数

$$F(x) = P\{X \leq x\}, -\infty < x < +\infty,$$

称为随机变量 X 的分布函数 (累积分布函数或者 cumulative distribution function).

Proposition 2.7. 分布函数性质如下

1. $0 \leq F(x) \leq 1; \lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0; \lim_{x \rightarrow +\infty} F(x) = F(+\infty) = 1;$

2. $F(x)$ 是单调函数, 即当 $x_1 < x_2$ 时, $F(x_1) \leq F(x_2);$

3. $F(x)$ 是右连续的, 即 $F(x+0) = F(x);$

4. 对任意的 $x_1 < x_2$, 有 $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1);$

5. 对任意的 x , 有 $P\{X = x\} = F(x) - F(x-0);$

证明. (3) 从 (1)(2) 可知 $F(x)$ 是单调有界的, 若存在间断点, 那么只能是第一类间断点, 即 $F(x)$ 的任意一点 x_0 处的右极限 $F(x_0 + 0)$ 是存在的. 现在来证明 $F(x + 0) = F(x)$. 这里需要用一下函数极限用数列的表示方法”若 $f(x)$ 在 x_0 处有极限当且仅当任意极限 $\lim x_n = x_0$ 的数列其对应函数值极限 $\lim f(x_n)$ 存在且相等”. 这里取一个单调减的数列 $x_1 > x_2 > \cdots > x_n > \cdots > x_0$, 即证 $\lim_{n \rightarrow \infty} F(x_n) = F(x_0)$. 而

$$\begin{aligned} F(x_1) - F(x_0) &= P\{x_0 < X \leq x_1\} = P\left[\bigcup_{i=1}^{\infty} \{x_{i+1} < X \leq x_i\}\right] = \sum_{i=1}^{\infty} P\{x_{i+1} < X \leq x_i\} \\ &= \sum_{i=1}^{\infty} F(x_i) - F(x_{i+1}) = F(x_1) - \sum_{i=1}^{\infty} F(x_{i+1}), \end{aligned}$$

因此有 $F(x_0) = \sum_{i=1}^{\infty} F(x_{i+1}) = F(x + 0)$. 注意第二个等号使用了一个重要的极限

$$\{x_0 < X \leq x_1\} = \lim_{n \rightarrow \infty} \bigcup_{i=1}^n \{x_{i+1} < X \leq x_i\}$$

(5) 这里因为无法保证是右连续的, 所以减去 x 除的右极限的跃度就是这 x 这一点的概率. □

Definition 2.8. 如果对随机变量 X 的分布函数 $F(x)$, 存在一个非负可积函数 $f(x)$, 使得对任意的实数 x , 都有

$$F(x) = \int_{-\infty}^x f(t)dt, -\infty < x < +\infty,$$

那么称 X 为连续型随机变量, 函数 $f(x)$ 称为 X 的概率密度.

Proposition 2.9. 连续型随机变量的分布函数 $F(x)$ 是在 $(-\infty, +\infty)$ 上连续的.

Proposition 2.10. 概率密度函数 $f(x)$ 的性质如下

1. $f(x) \geq 0$ ($f(x)$ 是概率密度函数的充要条件之一);
2. 对于任意实数 x , 有 $P\{X = x\} = F(x) - F(x - 0) = 0$.
3. $F(+\infty) = \int_{-\infty}^{+\infty} f(t)dt = 1$ ($f(x)$ 是概率密度函数的充要条件之一);
4. 对任意实数 $x_1 < x_2$, 有 $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)dt$;
5. 在 $f(x)$ 的连续点处有 $F'(x) = f(x)$. (证明见高数积分上限函数一节)

证明. 证明见高数积分上限函数一节 □

Proposition 2.11. 若 X 是连续型随机变量, 则

$$P\{x_1 < X \leq x_2\} = P\{x_1 \leq < x_2\} = P\{x_1 < X < x_2\} = P\{x_1 \leq X \leq x_2\}.$$

证明. 由 proposition 2.10 中 (2) 易得. □

常用分布

Definition 2.12. 如果随机变量 X 的分布律为

$$\begin{array}{c|cc} X & 0 & 1 \\ \hline P & 1-p & p \end{array}$$

其中 $0 < p < 1$, 则称 X 服从参数为 p 的 $0-1$ 分布或者两点分布.

Definition 2.13. 如果随机变量 X 的分布律为

$$P\{X = k\} = C_n^k p^k q^{n-k}, k = 0, 1, 2, \dots, n,$$

其中 $0 < p < 1, q = 1 - p$, 则称 X 服从参数为 n, p 的二项分布, 记做 $X \sim B(n, p)$.

Annotation 2.14. 在 n 重伯努利试验中, 若每次试验成功率为 $p(0 < p < 1)$, 则在 n 次独立重复试验中成功的总次数 X 服从二项分布.

Definition 2.15. 如果随机变量 X 的分布律为

$$P\{X = k\} = pq^{k-1}, k = 1, 2, \dots,$$

其中 $0 < p < 1, q = 1 - p$, 则称 X 服从参数为 p 的几何分布, 或称 X 具有几何分布.

Annotation 2.16. 在独立地重复做一系列伯努利试验中, 若每次试验成功率为 $p(0 < p < 1)$, 则第 k 次试验时才首次试验成功的概率服从几何分布.

Definition 2.17. 如果随机变量 X 的分布律为

$$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = l_1, \dots, l_2,$$

其中 $l_1 = \max(0, n - N + M), l_2 = \min(M, n)$, 则称随机变量 X 服从参数 n, N, M 的超几何分布.

Annotation 2.18. 如果 N 件产品中含有 M 件次品, 从中任意一次取出 n 件, 令 $X =$ 抽取的 n 件产品中的次品件数, 则 X 服从参数 n, N, M 的超几何分布.

Definition 2.19. 如果随机变量 X 的分布律为

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots,$$

其中 $\lambda > 0$ 为常数.