

考研概率论

枫聆

2021 年 7 月 31 日

目录

| | | |
|----------|-------------------|-----------|
| 1 | 随机事件和概率 | 3 |
| 1.1 | 样本空间与事件 | 3 |
| 1.2 | 古典概型和几何概率 | 4 |
| 1.3 | 概率空间 | 7 |
| 1.4 | 条件概率和事件独立性 | 10 |
| 1.5 | 更深的思考和技巧 | 14 |
| 2 | 随机变量及其概率分布 | 15 |
| 2.1 | 随机变量及其分部函数 | 15 |
| 2.2 | 常用分布 | 18 |
| 2.3 | 随机变量的函数的分布 | 23 |
| 3 | 多维随机变量及其分布 | 25 |
| 3.1 | 多维分布 | 25 |
| 3.2 | 边缘分布 | 27 |
| 3.3 | 条件分布和随机变量的独立性 | 28 |
| 3.4 | 多维随机变量的函数分布 | 29 |
| 4 | 随机变量的数字特征 | 29 |
| 4.1 | 数学期望 | 29 |
| 4.2 | 方差、协方差和矩 | 32 |
| 4.3 | 常用分布的数字特征 | 37 |

| | | |
|----------|----------------------------|-----------|
| 5 | 大数定理和中心极限定理 | 38 |
| 6 | 数理统计的基本概念 | 40 |
| 6.1 | 总体、样本、统计量和样本数字特征 | 40 |
| 6.2 | 统计三大分布 | 41 |
| 7 | 参数估计 | 42 |
| 7.1 | 点估计 | 42 |
| 7.2 | 矩估计 | 42 |
| 7.3 | 极大似然估计 | 42 |

随机事件和概率

样本空间与事件

Definition 1.1. 对随机现象进行观察或者实验被成为**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. 特别地, 当 A_1, A_2, \dots, A_n 两两无不相容时, 并称为和, 记做

$$A_1 + A_2 + \dots + A_n,$$

或者 $\sum_{i=1}^n A_i$.

Definition 1.8. 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.

Proposition 1.9. **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 $\overline{A \cup B} = \bar{A} \cap \bar{B}$; $\overline{A \cap B} = \bar{A} \cup \bar{B}$; $\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i$; $\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i$

古典概型和几何概率

Definition 1.10. 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 1.11. 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 1.12. 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 1.13. 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.

Definition 1.14. 将每次伯努利试验结果推广至 A_1, A_2, \dots, A_r , 而 $P(A_i) = p_i, i = 1, 2, \dots, r$, 且

$$p_1 + p_2 + \dots + p_r = 1, p_i \geq 0.$$

当 $r = 2$ 时, 即为伯努利试验. 将这种推广的伯努利试验独立重复进行 n 次, 设在 n 次试验中 A_1 出现 k_1 次, A_2 出现 k_2 , \dots , A_r 出现 k_r 次的概率为

$$\frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r},$$

其中 $k_i \geq 0$, 且 $k_1 + k_2 + \dots + k_r = n$. 该式被称为多项分布, 即它是 $(p_1 + p_2 + \dots + p_r)^n$ 的展开式的一般项.

Definition 1.15. 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 1.16. 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 $m_1 m_2 \cdots m_n$.

Definition 1.17. 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

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

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 1.18. 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 1.19. 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 1.20. 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.

□

概率空间

Annotation 1.21. 对事件和概率的长期研究，发觉事件的运算与集合运算完全相似，概率与测度有相同的性质，这个事实随着当时在实变函数论中关于勒贝格测度和积分的研究以及一般抽象测度和积分理论的发展而日益明确起来。

另外，19 世纪末以来，数学的各个分支广泛流行着一股公理化潮流，这个流派主张把最基本的假定公理化，其他结论则由它们经过演绎导出。

在这种背景下，1933 年，前苏联数学家科尔莫戈罗夫提出了概率论公理化结构，这个结构综合了前人成果，明确定义了基本概率，使概率论成为严谨的数学分支。

Definition 1.22. 若把事件的全体记为 \mathcal{F} ，它是由 Ω 的一些子集构成的子集族。当 \mathcal{F} 满足以下条件时

1. $\Omega \in \mathcal{F}$;
2. 若 $A \in \mathcal{F}$ ，则 $\bar{A} \in \mathcal{F}$;
3. 若 $A_n \in \mathcal{F}$, $n = 1, 2, \dots$ ，则 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ 。

我们称 \mathcal{F} 是空间 Ω 上的一个 σ 代数。

Definition 1.23. 给定 Ω 上某个非空集合 A ，通过对 A 里面的元素进行可数并，可数交和取补操作构造的包含 A 本身最小 σ 代数称为 A 生成的 σ 代数。

Annotation 1.24. 包含 A 的 σ 代数是一定存在，i.e. Ω 。将包含 A 的所有 σ 代数取交得到的就是包含 A 的最小 σ 代数。

Definition 1.25. 由某个拓扑空间上的一些开集生成的 σ 代数称为 Borel 代数，构成 Borel 代数的集合称为 Borel set。特别地，由 \mathcal{R} 上全体的左开右闭（左闭右开）区间构成的集合族所产生的 σ 代数中的集合称为一维 Borel set，记为 \mathcal{F}_1 。

若 x, y 表示任意实数，由于

$$\begin{aligned}\{x\} &= \bigcap_{n=1}^{\infty} [x, x + \frac{1}{n}) \\ (x, y) &= [x, y) - \{x\} \\ [x, y] &= [x, y) + \{y\} \\ (x, y] &= [x, y) + \{y\} - \{x\}\end{aligned}$$

因此 \mathcal{F}_1 中包含一切开区间，闭区间，单个实数，可列个实数，以及由它们经过可列次并，交运算而得出的集合。

Definition 1.26. 定义在事件 \mathcal{F} 上的一个实值函数 P 称为概率函数或者简称为概率，如果它满足如下三个条件 (Kolmogorov axioms)

1. 对于任意的 $A \in \mathcal{F}$, 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.(可数可加性或者可列可加性)

Proposition 1.27. 概率相关性质

1. $P(\emptyset) = 0$;
2. 概率具有有限可加性;
3. $P(\bar{A}) = 1 - P(A)$;
4. 若 $A \subset B$, 则 $P(B - A) = P(B) - P(A)$;
5. 若 $A \subset B$, 则 $P(A) \leq P(B)$; (单调性)
6. $0 \leq P(A) \leq 1$;

证明. (1) 因为 $\Omega = \Omega + \emptyset + \dots$, 所以

$$P(\Omega) = P(\Omega) + P(\emptyset) + \dots.$$

因此 $P(\Omega) = 0$.

(2) 对有限互斥事件 $A_i, i = 1, 2, \dots, n$, 因为

$$A_1 + A_2 + \dots + A_n = A_1 + A_2 + \dots + A_n + \emptyset + \emptyset + \dots,$$

由可数可加性和性质 1 有

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

(3) 因 $A \cup \bar{A} = \Omega$, 且 $A\bar{A} = \emptyset$, 于是由性质 3

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

(4) 若 $A \subset B$, 则 $A \cup (B - A) = B$, 于是

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

故 $P(B - A) = P(B) - P(A)$.

(5) 性质 4 的推论, 由 $P(B - A) \geq 0$, 即得 $P(B) \geq P(A)$.

(6) 由性质 5 可知, 对任意的事件 A , 有 $A \subset \Omega$, 即 $P(A) \leq P(\Omega) = 1$. □

Proposition 1.28. 概率加法公式

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

证明. 因 $A \cup B = A \cup (B - AB)$, 且 $A \cap (B - AB) = \emptyset$, 于是由性质 2 有

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

因 $AB \subset B$, 那么用一下性质 4, 即可得到

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

□

Corollary 1.29. 布尔不等式

$$P(A \cup B) \leq P(A) + P(B).$$

Corollary 1.30. Bonferroni 不等式

$$P(AB) \geq P(A) + P(B) - 1.$$

Proposition 1.31. 一般加法公式

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{1, \dots, n} P(A_i) - \sum_{\substack{i < j \\ i, j = 1, \dots, n}} P(A_i A_j) + \sum_{\substack{i < j < k \\ i, j, k = 1, \dots, n}} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n).$$

证明. 可用数学归纳法证明.

□

Proposition 1.32. 概率减法公式

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

证明. 因 $A = A(B \cup \bar{B}) = AB \cup A\bar{B} = AB \cup (A - B)$, 且 $AB \cap (A - B) = \emptyset$, 于是

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

移项可得原式.

□

Definition 1.33. 在科尔莫戈洛夫的概率论公理化结构中, 称三元组总体 (Ω, \mathcal{F}, P) 为**概率空间**, 其中 Ω 是样本空间, \mathcal{F} 是事件域, P 是概率.

条件概率和事件独立性

Definition 1.34. 设 (Ω, \mathcal{F}, P) 是一个概率空间, 给定事件 $A \in \mathcal{F}$ 和 $B \in \mathcal{F}$, 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**(conditional probability).

Proposition 1.35. 乘法公式 (今后出现条件概率 $P(B|A)$ 时, 都假定 $P(A) > 0$)

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

Corollary 1.36. 当 $P(A) > 0, P(B) > 0$ 时, 有

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

Proposition 1.37. 条件概率 $P(B|A)$ 具有概率的三个基本性质

1. $P(B|A) \geq 0$ (非负性);
2. $P(\Omega|A) = 1$ (规范性);
3. $P(\sum_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$ (可列可加性);

利用这三个基本性质也可以导出其他的一些性质

1. $P(\emptyset|B) = 0$;
2. $P(B|A) = 1 - P(\bar{B}|A)$;
3. $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) - P(B_1 B_2|A)$.

Proposition 1.38. 乘法公式可以推广到 n 个事件之交

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

这里要求 $P(A_1 A_2 \cdots A_{n-1}) > 0$.

Proposition 1.39. 全概率公式 设 B_1, B_2, \cdots, B_n 满足 $\bigcup_{i=1}^n B_i = \Omega$, $B_i B_j = \emptyset$ 且 $P(B_k) > 0$, $k = 1, 2, \cdots, n$, 则对任意事件有

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

称满足 $\bigcup_{i=1}^n B_i = \Omega$, $B_i B_j = \emptyset$ 的 B_1, B_2, \cdots, B_n 为一个**完备事件组**.

证明. 由概率的有限可列可加性有

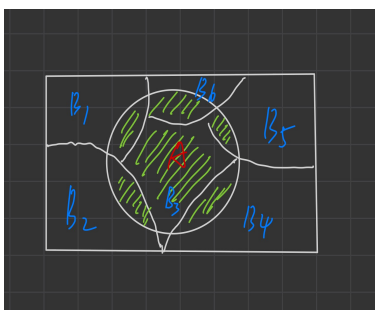
$$P(A) = \bigcup_{i=1}^n P(AB_i),$$

再利用乘法公式即得

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

□

Annotation 1.40. (全概率公式的 motivation) 概率论的重要研究课题之一是希望从已知的简单事件的概率推算出未知的复杂的事件的概率. 为了达到这个目的, 经常把一个复杂的事件分解成若干不相容的简单事件之和, 再用过分别计算这些简单事件的概率, 最后利用概率的可加性得到最终的结果.



Definition 1.41. 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

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

前提条件可以换成“若事件 A 能且只能与两两互不相容的事件 B_1, B_2, \dots, B_n 之一同时发生”, 这个条件实际上更有实际意义. 其中 $P(B_i)$ 称为先验概率, 条件概率 $P(B_i|A)$ 称为后验概率.

证明. 由于 $P(A) > 0, P(B_k) > 0$, 那么此时有

$$P(B_k A) = P(A)P(B_k|A) = P(B_k)P(A|B_k).$$

故

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

在利用全概率公式即有

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

□

Annotation 1.42. 贝叶斯公式中先验概率 $P(B_i)$ 反应了各种“原因”发生的可能的性大小，一般是以往的经验总结，在当前试验前就已经知道了. 若当前试验产生了事件 A ，这个信息将有助于探讨事件发生的原因. 后验概率 $P(B_i|A)$ 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.

Example 1.43. (贝叶斯决策) 为了判定一个字母是“C” 还是“O”，通常采取抽取它的一个特征 X . 然后再根据这个特征作出判决，这时贝叶斯决策常用的方法之一.

以 A_1, A_2 分别表示被检验的字母为 C 或者 O 这一事件，它们的先验概率 $P(A_1)$ 及 $P(A_2)$ 应预先给定，此外要通过试验确定 $P(X|A_1)$ 或者 $P(X|A_2)$ ，由贝叶斯公式得

$$P(A_i|X) = \frac{P(A_i)P(X|A_i)}{\sum_{i=1} P(A_i)P(X|A_i)}$$

其中 $i = 1, 2$. 若 $P(A_1|X) > P(A_2|X)$ ，则做出决策，具有特征 X 的字母是 C .

Definition 1.44. 若事件 A, B 满足等式

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

则称 A 与 B **相互独立**.

Annotation 1.45. 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

Corollary 1.46. 若事件 A 与 B 独立，则下列各对立事件也相互独立

$$\{\bar{A}, B\}, \{A, \bar{B}\}, \{\bar{A}, \bar{B}\}$$

证明. 由于

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

同理可得 $\{A, \bar{B}\}$. 特别因为 \bar{A} 和 B 独立，马上可以得到

$$P(\bar{A}) = P(\bar{A}B \cup \bar{A}\bar{B}) = P(\bar{A})P(B) + P(\bar{A}\bar{B}).$$

移项即可得到 $P(\bar{A}\bar{B}) = P(\bar{A})P(\bar{B})$. □

Definition 1.47. 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

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

Annotation 1.48. 任意事件两两独立并不能推出它们相互独立.

Proposition 1.49. 若 A, B 是两个相互独立的事件, 则

$$P(A - B) = P(A)P(\overline{B}).$$

证明.

$$P(A - B) = P(A) - P(AB) = P(A)(1 - P(B)) = P(A)P(\overline{B}).$$

□

Proposition 1.50. 若 A_1, A_2, \dots, A_n 是 n 个相互独立的事件, 则

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\overline{A_1})P(\overline{A_2}) \cdots P(\overline{A_n}).$$

证明. 由于 $\overline{A_1 \cup A_2 \cup \dots \cup A_n} = \overline{A_1} \overline{A_2} \cdots \overline{A_n}$, 因此

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= 1 - P(\overline{A_1} \overline{A_2} \cdots \overline{A_n}) \\ &= 1 - P(\overline{A_1})P(\overline{A_2}) \cdots P(\overline{A_n}). \end{aligned}$$

□

更深的思考和技巧

Annotation 1.51. 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 1.52. 如何理解事件之间独立性 事件 A 和 B 独立, 则有下面等式成立

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

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 1.53. 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 1.54. 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

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

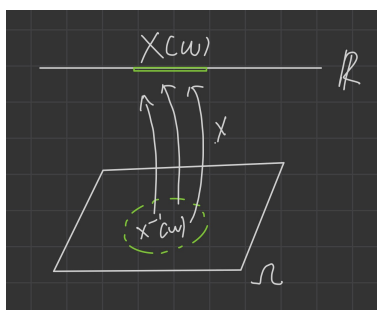
这些等式也可以推广为 n 个事件相互独立.

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 . X 表示样本点到 \mathbb{R} 上一一映射.

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. $P(X < 1)$. 随机变量是样本点的函数, 因此在试验前我们只能知道它可能取那些值, 而不确定它将取何值, 这就是随机的含义.



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**.

Annotation 2.4. 离散型随机变量可以理解为 \mathbb{R} 上的点.

Definition 2.5. 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 (质量) 分布**(probability mass function) 或者分部律, 其中 P 是一个概率函数.

Proposition 2.6. 分部律充要条件

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 2.7. 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 2.8. **分布函数的本质意义** 给定随机变量 X 的分布函数 $F(x)$, 则对任意的 $x_1 < x_2$, 有

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1).$$

证明.

$$P\{x_1 < X \leq x_2\} = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1).$$

□

Annotation 2.9. 从上式的证明中我们知道只要对一切实数 x 给出了概率 $P\{X \leq x\}$, 就能算出 X 落入某个区间 $(a, b]$ 的概率, 特别地再利用概率的性质还可以算出 X 属于 \mathbb{R} 上某些相当复杂的点集. 为了计算概率, 必要要求随机变量具有可测性, 而分布函数的引进则把对于随机变量的概率计算转换为了对分布函数的数值运算.

Proposition 2.10. 分布函数性质如下

1. $F(x)$ 是单调函数, 即当 $x_1 < x_2$ 时, $F(x_1) \leq F(x_2)$ (**分部函数充要条件之一**);
2. $0 \leq F(x) \leq 1$; $\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$; $\lim_{x \rightarrow +\infty} F(x) = F(+\infty) = 1$ (**分部函数充要条件之一**);
3. $F(x)$ 是右连续的, 即 $F(x+0) = F(x)$ (**分部函数充要条件之一**);
4. 对任意的 x , 有 $P\{X = x\} = F(x) - F(x-0)$;

证明. (1) 若 $x_2 > x_1$, 则

$$F(x_2) - F(x_1) = P\{x_1 < X \leq x_2\} \geq 0.$$

(2) 因

$$\begin{aligned} P\{-\infty < X < +\infty\} &= \sum_{n=-\infty}^{+\infty} P\{n \leq X \leq n+1\} \\ &= \sum_{n=-\infty}^{+\infty} [F(n+1) - F(n)] \\ &= \lim_{n \rightarrow +\infty} F(n) - \lim_{m \rightarrow -\infty} F(m) = 1 \end{aligned}$$

上面只是说明了有理数趋于无穷的极限, 需要推广任意实数上. 因为 $F(x)$ 的单调性 ($F(\lfloor x \rfloor) \leq F(x) \leq F(\lceil x \rceil)$), 所以 $\lim_{x \rightarrow -\infty} F(x) = \lim_{m \rightarrow -\infty} F(m)$, $\lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow +\infty} F(n)$ 存在. 因为 $0 \leq F(x) \leq 1$, 故

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

(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, 实际也可以从单调有界来看出从 x_0 右边趋于 x_0 时候 $F(x)$ 极限存在. 现在来证明 $F(x + 0) = F(x)$. 这里可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

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

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

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

(4) 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. □

Proposition 2.11. 特殊点和区间的概率分布

$$\begin{aligned} P\{X = x\} &= F(x) - F(x - 0) \\ P\{X < x\} &= F(x - 0) \\ P\{X > x\} &= 1 - F(x) \\ P\{X \geq x\} &= 1 - F(x - 0) \end{aligned}$$

Annotation 2.12. 上述更贴切的说明了分布函数是一种分析性质良好的函数, 给定了分布函数就能计算处各种事件的概率.

Proposition 2.13. 给定离散型随机变量的分布律 $P\{X = x_i\} = p_i, i = 1, 2, \cdots$, 它的分布函数 $F(x)$ 为

$$F(x) = P\{X < x\} = \sum_{x_k < x} p(x_k)$$

Definition 2.14. 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.

Annotation 2.15. 连续随机变量可以理解为 \mathbb{R} 上的区间. 自然地, 事件运算对应了区间集合运算.

Proposition 2.16. 连续型随机变量的分布函数 $F(x)$ 是在 $(-\infty, +\infty)$ 上连续的.

Proposition 2.17. 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

□

Annotation 2.18. 上述性质表明, 一个事件的概率为 0, 这件事并不一定是不可能事件; 同样地, 一个事件概率等于 1, 这件事也不一定是必然事件.

Proposition 2.19. 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 2.17 中 (2) 易得.

□

常用分布

Definition 2.20. 若随机变量 X 只取常数 c , 即 $P\{X = c\} = 1$, 那么这时的分布函数为

$$F(x) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}.$$

则称 X 服从退化分布, 又称单点分布.

Definition 2.21. 如果随机变量 X 的分布律为

| | | |
|-----|-------|-----|
| X | 0 | 1 |
| P | $1-p$ | p |

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 2.22. 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 2.23. 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.

Proposition 2.24. **二项分布分析性质** 若 $X \sim B(n, p)$, 当 n 固定时, $P\{X = k\}$ 先随 k 增加而增大, 达到某一极值后又逐渐下降, 其在 $k = \lfloor (n+1)p \rfloor$ 取得极值.

证明. 我们来考察

$$\frac{P\{X = k\}}{P\{X = k-1\}} = \frac{C_n^k p}{C_n^{k-1} q} = \frac{(n+1-k)p}{kq} = \frac{(n+1)p - (1-q)k}{kq} = 1 + \frac{(n+1)p - k}{kq}$$

因此

当 $k < (n+1)p$ 时, $P\{X = k\} > P\{X = k-1\}$;

当 $k = (n+1)p$ 时, $P\{X = k\} = P\{X = k-1\}$;

当 $k > (n+1)p$ 时, $P\{X = k\} < P\{X = k-1\}$;

因为 $(n+1)p$ 不一定是整数, 所以存在整数 m , 使得 $(n+1)p - 1 < m \leq (n+1)p$, 即当 $k = m$ 时 $P\{X = k\}$ 达到最大值. \square

Definition 2.25. 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 2.26. 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 $p(0 < p < 1)$, 则在第 k 次试验时才首次试验成功的概率服从几何分布.

Proposition 2.27. **几何分布具有无记忆性**, 即若假定前 m 次没有成功, 设随机变量 X' 为为了达到首次成功还需要做的试验数, 则 X' 依然服从几何分布.

证明.

$$P\{X' = k\} = P\{X = m + k | X > m\} = \frac{P\{X = m + k\}}{P\{X > m\}} = \frac{pq^{m+k-1}}{q^m} = pq^{k-1}.$$

\square

Definition 2.28. 若随机变量 X 的分布律为

$$P\{X = k\} = C_{k-1}^{r-1} p^r q^{k-r}, k = r, r+1, \dots,$$

其中 $r \geq 1$, 则称 X 服从**巴斯卡分布**. 特别地当 $r = 1$ 时, 就是几何分布.

Annotation 2.29. 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 $p(0 < p < 1)$ ，那么第 r 次成功时的试验次数的概率服从巴斯卡分布. 若考虑第 r 次成功时的失败试验次数的概率，即

$$P\{X = k\} = \binom{r+k-1}{r-1} p^r q^k.$$

上述概率分布被称为**负二项分布**，其中的组合数项可以写作

$$(-1)^k \frac{(-r)(-r-1)(-r-2)\cdots(-r-k+1)}{k!} = (-1)^k \binom{-r}{k},$$

那么原式就为

$$P\{X = k\} = \binom{-r}{k} p^r (-q)^k.$$

这就是”负”字的来源.

Definition 2.30. 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 2.31. 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 2.32. 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，则称随机变量 X 服从参数为 λ 的泊松分布，记为 $X \sim \mathbb{P}(\lambda)$.

Annotation 2.33.

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

其中 $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Theorem 2.34. 泊松定理 设 $\lambda > 0$ 是一个常数， n 是任意整数，设 $np_n = \lambda$ ，则对于任一固定的非负整数 k ，有

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Annotation 2.35. 应用泊松定理来做二项式近似计算要求 n 较大, p 较少, 并且 np 大小适中, 则

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!},$$

其中 $\lambda = np$.

Proposition 2.36. 若我们关注的随机变量 X 是表示某个事件 A 发生的次数 k 且 $X \sim P(\lambda)$, 假定它具有下面三个性质

1. **平稳性** 在 $[t_0, t_0 + t)$ 中 A 发生个数只与时间间隔长度有关而与时间起点 t_0 无关. 若以 $P_k(t)$ 表示在长度为 t 的时间区间中 A 发生的 k 次的概率, 那么对任意的 t 有

$$\sum_{k=0}^{\infty} P_k(t) = 1$$

成立. 过程的平稳性表示了它的概率规律不随时间的推移而改变.

2. **独立增量性 (无后效性)** 在 $[t_0, t_0 + t)$ 中 A 发生的 k 次这一事件与时刻 t_0 以前发生的事件独立. 独立增量性表明在互不相交的时间区间内过程进行的相互独立性.
3. **普通性** 在充分小的时间间隔内, A 最多发生一次. 即, 若记

$$\psi(t) = \sum_{k=2}^{\infty} P_k(t) = 1 - P_0(t) - P_1(t),$$

那么有

$$\lim_{t \rightarrow 0} \frac{\psi(t)}{t} = 0.$$

换句话说, 就是在同一时间瞬时 A 不可能发生 2 次以上. 由上述条件求出来的 $P_k(t)$ 固定 t 就是泊松分布.

Definition 2.37. 如果连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{other} \end{cases},$$

则称 X 在区间 $[a, b]$ 上服从**均匀分布**, 记做 $X \sim U[a, b]$. 其的分布函数为

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}.$$

Proposition 2.38. 设 $X \sim U[a, b]$, 则对 $a \leq c < d \leq b$, 有

$$P\{c \leq x \leq d\} = \frac{d-c}{b-a}.$$

Definition 2.39. 如果连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases},$$

其中 $\lambda > 0$ 是个常数, 则称 X 服从参数为 λ 的**指数分布**, 记做 $X \sim E(\lambda)$. 其分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}.$$

Proposition 2.40. 设 $X \sim E(\lambda)$, 则有

1. 当 $t > 0$ 时, $P\{X > t\} = \int_t^{+\infty} \lambda e^{-\lambda t} dt$;
2. 当 $t > 0, s > 0$ 时, $P\{X > t + s | X > s\} = \frac{P\{t+s\}}{P\{s\}} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda(s)}} = e^{-\lambda t} = P\{X > t\}$. 此性质称为指数分布具有“无记忆性”.

Definition 2.41. 如果连续性随机变量的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

其中 μ, σ 为常数且 $\sigma > 0$, 则称 X 服从参数为 μ, σ 的**正态分布**, 记做 $X \sim N(\mu, \sigma^2)$. 当 $\mu = 0, \sigma^2 = 1$ 时, 即 $X \sim N(0, 1)$, 称 X 服从**标准正态分布**, 此时用 $\varphi(x)$ 表示 X 的概率密度, 即

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

当 $X \sim N(\mu, \sigma^2)$, 其分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

当 $X \sim N(0, 1)$, 其分布函数为

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Annotation 2.42. **正态分布的图像性质** 正态分布概率密度图像是关于 $x = \mu$ 对称的. 当 σ 越小时, 分布越集中在 $x = \mu$ 附近; σ 越大时, 分布就越平坦. 若 X 服从 $N(\mu, \sigma^2)$, 在一次试验中 X 几乎都落在 $(\mu - 3\sigma, \mu + 3\sigma)$.

Proposition 2.43. **正态分布到标准正态分布的转换** 设 $X \sim N(\mu, \sigma^2)$, 则 $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

证明. $Z = \frac{X-\mu}{\sigma}$ 的分布函数为

$$P\{Z \leq x\} = P\left\{\frac{X-\mu}{\sigma} \leq x\right\} = P\{X \leq \mu + \sigma x\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\mu+\sigma x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

令 $u = \frac{t-\mu}{\sigma}$, 那么 $t = \sigma u + \mu$, 于是

$$P\{Z \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du = \Phi(x).$$

由此 $Z \sim N(0, 1)$. □

Corollary 2.44. 若 $X \sim N(\mu, \sigma^2)$, 则其分布函数 $F(x)$ 可写成

$$F(x) = P\{X \leq x\} = P\left\{\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right\} = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Proposition 2.45. 设 $X \sim N(\mu, \sigma^2)$, 其分布函数为 $F(x)$, 则

1. $P\{a < x \leq b\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$;
2. 概率函数 $f(x)$ 关于 $x = \mu$ 对称, $\varphi(x)$ 是偶函数;
3. $\Phi(-x) = 1 - \Phi(x)$, $\Phi(0) = \frac{1}{2}$;
4. 当 $X \sim N(0, 1)$ 且 $a > 0$ 时, $P\{|X| \leq a\} = 1 - 2\Phi(-a) = 2\Phi(a) - 1$.

证明. (4)

$$I(a) = \int_0^\infty e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}}$$

□

随机变量的函数的分布

Annotation 2.46. ??? CDF 的导数是否等于 PDF?

$$F(x)' = f(x)$$

Definition 2.47. $y = f(x)$ 实际上是一个可测函数.

Theorem 2.48. 设随机变量 X 具有概率密度 $f_X(x)$ 对任意实数都有定义. 若给定 $g(x)$ 是严格单调的, 其反函数 $g^{-1}(x)$ 有连续导数(有可能怕破坏 $f_X(x)$ 的连续性?). 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] | [g^{-1}(y)]' | & \alpha < y < \beta \\ 0, & \text{otherwise} \end{cases},$$

其中 $\alpha = \min(g(-\infty), g(+\infty)), \beta = \max(g(-\infty), g(+\infty))$.

证明. 若 $g(x)$ 是严格单调增大, 且其反函数 $g^{-1}(x)$ 是有连续导数. 给定任意实数 $a \in (\alpha, \beta)$, 那么

$$F_Y(y) = P\{Y \leq a\} = P\{X \leq g^{-1}(a)\} = F_X(g^{-1}(a))$$

对 $F_Y(y)$ 关于 y 求导, 即有 $f_Y(y) = f_X[g^{-1}(y)][g^{-1}(y)]'$. 若给定实数 $a \leq \alpha$, 那么显然有 $F_Y(a) = 0$; 当 $a \geq \beta$, 有 $F_Y(a) = 1$.

同理若 $g(x)$ 是严格单调递减, 若给定任意实数 $a \in (\alpha, \beta)$, 那么

$$F_Y(y) = P\{Y \leq a\} = P\{X \geq g^{-1}(a)\} = 1 - F_X(g^{-1}(a)),$$

对 $F_Y(y)$ 关于 y 求导, 即有 $f_Y(y) = -f_X[g^{-1}(y)][g^{-1}(y)]'$. 综合两式便得命题中的概率密度函数. \square

Corollary 2.49. 设随机变量 X 具有概率密度 $f_X(x)$ 对任意实数都有定义. 若 $g(x)$ 在不相互重叠的区间 I_1, I_2, \dots 上逐段严格单调, 其反函数分别为 $h_1(y), h_2(y), \dots$, 而且 $h'_1(y), h'_2(y), \dots$ 均为连续函数. 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \sum_i f_X[h_i(y)] |h'_i(y)|.$$

多维随机变量及其分布

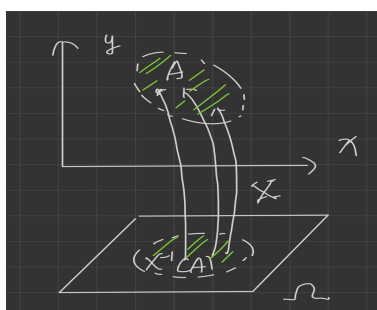
多维分布

Definition 3.1. 若随机变量 $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ 定义在同一概率空间 (Ω, \mathcal{F}, P) 上, 则称

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

构成一个 n 维随机向量, 亦称 n 维随机变量.

Annotation 3.2. 等待些许感受



Definition 3.3. 称 n 元函数

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

为 n 维随机变量 $X = (X_1, X_2, \dots, X_n)$ 的(联合) 分布函数.

Proposition 3.4. 多元分布函数的基本性质

1. 关于每个变元是单调不减;
2. $0 \leq F(x_1, x_2, \dots, x_n) \leq 1$;
3. $F(x_1, x_2, \dots, -\infty, \dots, x_n) = 0, F(+\infty, +\infty, \dots, +\infty) = 1$;
4. 关于每个变元右连续;
5. 对任意 $x_1 < x_2, y_1 < y_2$, 都有

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0.$$

上述实际上就是概率 $P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$.

Definition 3.5. 如果每一个 X_i 都是一个离散型随机变量, $i = 1, \dots, n$, 则称 $X = (X_1, \dots, X_n)$ 为一 n 维离散随机变量. 设 X_i 的所有可能取值 (有限或可数个) 为 $\{a_{i1}, a_{i2}, \dots\}, i = 1, \dots, n$, 则称

$$p(j_1, j_2, \dots, j_n) = P(X_1 = a_{1j_1}, X_2 = a_{2j_2}, \dots, X_n = a_{nj_n}), j_1, \dots, j_n = 1, 2, \dots$$

为 n 维离散型随机变量 X 的**概率函数**或者**分布律**.

Proposition 3.6. n 维离散型随机变量的概率函数具有如下性质

$$1. p(j_1, j_2, \dots, j_n) \geq 0, j_i = 1, 2, \dots, i = 1, 2, \dots, n;$$

$$2. \sum_{j_1, \dots, j_n} p(j_1, \dots, j_n) = 1.$$

Example 3.7. 多项分布

Definition 3.8. 若对 n 维随机变量 $X = (X_1, \dots, X_n)$ 的分布函数存在 \mathbb{R}^n 上的非负函数 $f(x_1, x_2, \dots, x_n)$, 对任意的实数 x_1, x_2, \dots, x_n , 都有

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(v, u, \dots, w) dv du \dots dw.$$

则称 $f(x)$ 为 X 的**概率密度函数**.

Definition 3.9. 如果二维连续随机变量 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} \frac{1}{S}, & (x, y) \in G \\ 0, & \text{other} \end{cases},$$

其中 S 是 \mathbb{R}^2 上有限区域 G 的面积, 则称 (X, Y) 服从区域 G 上的**均匀分布**.

Example 3.10. **以原点为圆心的单位圆上的均匀分布** 设二维随机变量 (X, Y) 在区域 D 上服从均匀分布, 其中

$$D = \{(x, y) \mid x^2 + y^2 \leq 1\}.$$

那么 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1 \\ 0, & \text{other} \end{cases}$$

X 的边缘密度为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \begin{cases} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{1}{\pi} dy, & -1 \leq x \leq 1 \\ 0, & \text{other} \end{cases} = \begin{cases} \frac{2\sqrt{1-x^2}}{\pi}, & -1 \leq x \leq 1 \\ 0, & \text{other} \end{cases}$$

条件概率密度 $f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, f_X(x) > 0$, 即 $-1 \leq x \leq 1$. 因此当 $-1 < x < 1$ 时有

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{2\sqrt{1-x^2}}, & -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2} \\ 0, & \text{other} \end{cases}$$

Definition 3.11. 如果二维连续随机变量 (X, Y) 的概率密度为

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_1\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right]\right),$$

其中 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 均为常数, $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$, 则称 (X, Y) 服从参数为 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的二项分布, 记为 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$.

Proposition 3.12. 若 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$, 则

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \\ f_Y(y) &= \frac{1}{2\pi\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \end{aligned}$$

即 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$

边缘分布

Definition 3.13. 多维随机变量中, 只包含其中部分变量的概率分布被称为**边缘分布**.

Definition 3.14. 二维随机变量 (X, Y) 的分布函数为 $F(x, y)$, 分别称 $F_X(x) = P\{X \leq x\}$ 和 $F_Y(y) = P\{Y \leq y\}$ 为 (X, Y) 的一维**边缘 (累积) 分布函数**. 即有下述等式成立

$$\begin{aligned} F_X(x) &= P\{X \leq x\} = P\{X \leq x, -\infty < y < +\infty\} = F(x, +\infty), \\ F_Y(y) &= P\{Y \leq y\} = P\{-\infty < x < +\infty, Y \leq y\} = F(+\infty, y) \end{aligned}$$

Proposition 3.15. 给定二维离散型随机变量 (X, Y) , (X, Y) 的可能取值为 $(x_i, y_j), i, j = 1, 2, \dots$, (X, Y) 的概率质量函数为 $P\{X = x_i, Y = y_j\} = p_{ij}$. 则关于 X 和 Y 的一维**边缘概率质量函数**或**边缘分布律**分别为

$$\begin{aligned} p_X(x_i) &= P(X = x_i) = \sum_{j=1}^{+\infty} p_{ij} \\ p_Y(y_j) &= P(Y = y_j) = \sum_{i=1}^{+\infty} p_{ij} \end{aligned}$$

Proposition 3.16. 给定二维连续型随机变量 (X, Y) 的概率密度函数为 $f(x, y)$. 则关于 X 和 Y 的一维**边缘概率密度函数**分别为

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx \end{aligned}$$

Proposition 3.17. 边缘分布律不能决定联合分布律; 边缘概率密度不能决定联合概率密度.

条件分布和随机变量的独立性

Definition 3.18. 设 (X, Y) 为二维离散型随机变量, 其全部的可能取值为 $(x_i, y_j), i, j = 1, 2, \dots$. (X, Y) 的联合分布律为 $P\{X = x_i, Y = y_j\} = p_{ij}$. 若给定事件 $\{Y = y_j\}$, 其概率 $P\{Y = y_j\} > 0$, 则称

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{\sum_{i=1}^{+\infty} p_{ij}}, i = 1, 2, \dots$$

为在给定 $Y = y_j$ 的条件下 X 的**条件分布律**(概率函数). 类似的, 若 $P\{X_i\} > 0$, 则称

$$P(Y = y_j | X = x_i) = \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} = \frac{p_{ij}}{\sum_{j=1}^{+\infty} p_{ij}}, j = 1, 2, \dots$$

为在给定条件 $X = x_i$ 下 Y 的**条件分布律**.

Definition 3.19. 设 (X, Y) 有概率密度 $f(x, y)$ 和分布函数 $F(x)$, 则

$$\begin{aligned} P\{X \leq x | Y = y\} &= \lim_{\Delta y \rightarrow 0} P\{X \leq x | y \leq Y \leq y + \Delta y\} \\ &= \frac{P\{X \leq x, y \leq Y \leq y + \Delta y\}}{P\{y \leq Y \leq y + \Delta y\}} \\ &= \frac{F(x, y + \Delta y) - F(x, y)}{F(+\infty, y + \Delta y) - F(+\infty, y)} \\ &= \frac{\int_{-\infty}^x \left[\int_y^{y+\Delta y} f(v, u) du \right] dv}{\int_{-\infty}^{+\infty} \left[\int_y^{y+\Delta y} f(v, u) du \right] dv} \quad \text{分子分母除以 } \Delta y \\ &= \frac{\int_{-\infty}^x f(v, y) dv}{\int_{-\infty}^{+\infty} f(v, y) dv} \end{aligned}$$

于是 $P\{X \leq x | Y = y\} = \frac{\int_{-\infty}^x f(v, y) dv}{f_Y(y)} = \int_{-\infty}^x \frac{f(v, y)}{f_Y(y)} dv$. 因此在给定 $Y = y$ 的条件下, X 的**条件概率密**为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, f_Y(y) > 0.$$

类似地, 在给定 $X = x$ 的条件下, Y 的**条件概率密度**为

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, f_X(x) > 0.$$

Definition 3.20. 设 X_1, X_2, \dots, X_n 为 n 个随机变量, 若对于任意的实数 x_1, x_2, \dots, x_n 都有

$$P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\} = P\{X_1 \leq x_1\} \cdots P\{X_n \leq x_n\},$$

则称 X_1, X_2, \dots, X_n 是相互独立的, 即

$$F(x_1, x_2, \dots, x_n) = F(x_1)F(x_2) \cdots F(x_n).$$

Proposition 3.21. 设 X_1, X_2, \dots, X_n 为 n 个离散型随机变量, 若它们是相互独立的, 则它们的联合分布律等于各种的边缘分布律的乘积, 即

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = P\{X_1 = x_1\}P\{X_2 = x_2\} \cdots P\{X_n = x_n\},$$

其中 (x_1, \dots, x_n) 为 (X_1, X_2, \dots, X_n) 的值域中的任意一个点.

Proposition 3.22. 设 X_1, X_2, \dots, X_n 为 n 个连续型随机变量, 若它们是相互独立的, 则它们的联合密度等于各自的边缘密度的乘积, 即

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \cdots f(x_n), \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

多维随机变量的函数分布

Proposition 3.23. 当 ξ, η 是相互独立的两个非负整数离散型随机变量, 各有分布律 $\{a_k\}$ 与 $\{b_k\}$, 那么 $\xi + \eta$ 有分布律

$$P\{\xi + \eta = n\} = \sum_{k=0}^n a_k b_{n-k}.$$

称此公式为离散卷积公式.

Proposition 3.24. 设 X_1, X_2, \dots, X_n 为 n 个连续型随机变量, (X_1, X_2, \dots, X_n) 的概率密度函数为 $f(x_1, x_2, \dots, x_n)$. 若随机变量 $Y = g(X_1, X_2, \dots, X_n)$, 那么对任意的 y , 有

$$P\{Y < y\} = \int \int \cdots \int_{g(x_1, x_2, \dots, x_n) < y} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n.$$

随机变量的数字特征

数学期望

Definition 4.1. 设 X 为一离散型随机变量, 其分布律为

$$P\{X = x_i\} = p_i, i = 1, 2, \dots.$$

如果级数 $\sum_{i=1}^{+\infty} x_i p_i$ 绝对收敛, 则称此级数为随机变量 X 的数学期望, 记做 $E(X)$, 即 $E(x) = \sum_{i=1}^{+\infty} x_i p_i$. 若 $\sum_{i=1}^{+\infty} x_i p_i$ 发散, 则称 X 的数学期望不存在.

Annotation 4.2. (为什么需要绝对收敛) 如果某个级数 $\sum_{i=1}^{+\infty} x_i p_i$ 只是收敛 (条件收敛), 而其绝对值构成的级数 $\sum_{i=1}^{+\infty} |x_i| p_i$ 并不收敛, 那么这个级数各项次序改排之后, 可以使得 $\sum_{i=1}^{+\infty} x_i p_i$ 并不收敛, 或者使它收敛于事先确定的任意值. 若 $\sum_{i=1}^{+\infty} x_i p_i$ 表示某个数学期望 $E(X)$, 这就意味着 $E(X)$ 取值与随机变量所取值的排列次序有关, 而 $E(X)$ 作为刻画 X 的某种特性的数值, 有其客观的意义, 不应与其人为排列次序有关.

Definition 4.3. 如果连续型随机变量 X 具有概率密度函数为 $f(x)$, 若积分

$$\int_{-\infty}^{+\infty} x f(x) dx$$

绝对收敛, 则称此积分为随机变量 X 的数学期望, 记为 $E(x)$, 即

$$E(x) = \int_{-\infty}^{+\infty} x f(x) dx.$$

Definition 4.4. 设 X 为一随机变量, 其分布函数为 $F(x)$, 如果

$$\int_{-\infty}^{+\infty} |x| dF(x)$$

收敛, 则称 $\int_{-\infty}^{+\infty} x dF(x)$ 为随机变量 X 的数学期望, 记为 $E(X)$.

Proposition 4.5. 给定随机变量 X 与 Y , 它们数学期望都存在分布为 $E(X)$ 与 $E(Y)$ 存在则有性质如下

1. 若 X 是一常数 c , 则 $E(c) = c$;
2. 任意给定常数 c , 则 $E(cX) = cE(X)$;
3. $E(X \pm Y) = E(X) \pm E(Y)$;
4. 若 $X \geq 0$, 则 $E(X) \geq 0$;

证明. (3) 若二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, 其边缘的概率密度分布为 $f_X(x), f_Y(y)$, 则

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dy dx + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy \\ &= E(X) + E(Y). \end{aligned}$$

□

Proposition 4.6. 若随机变量 X 和 Y 相互独立, 且它们的数学期望均存在. 则

$$E(XY) = E(X)E(Y).$$

证明. 设二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, 其边缘的概率分布为 $f_X(x), f_Y(y)$, 则

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dxdy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{+\infty} yf_Y(y) \left[\int_{-\infty}^{+\infty} xf_X(x)dx \right] dy \\ &= E(X)E(Y). \end{aligned}$$

□

$$\begin{aligned} & \int_{-\infty}^{+\infty} y \left[\int_0^{+\infty} \frac{1}{x_1} f(x_1) g\left(\frac{y}{x_1}\right) dx_1 - \int_{-\infty}^0 \frac{1}{x_1} f(x_1) g\left(\frac{y}{x_1}\right) dx_1 \right] dy \\ & \int_{-\infty}^{+\infty} y \left[\int_0^{+\infty} \frac{1}{x_1} f(x_1) g\left(\frac{y}{x_1}\right) dx_1 \right] dy - \int_{-\infty}^{+\infty} y \left[\int_{-\infty}^0 \frac{1}{x_1} f(x_1) g\left(\frac{y}{x_1}\right) dx_1 \right] dy \\ & \int_0^{+\infty} \int_{-\infty}^{+\infty} y \frac{1}{x_1} f(x_1) g\left(\frac{y}{x_1}\right) dy dx_1 - \int_{-\infty}^0 \int_{-\infty}^{+\infty} y \frac{1}{x_1} f(x_1) g\left(\frac{y}{x_1}\right) dy dx_1 \\ & \int_0^{+\infty} x_1 f(x_1) \int_{-\infty}^{+\infty} \frac{y}{x_1} g\left(\frac{y}{x_1}\right) d\frac{y}{x_1} dx_1 + \int_{-\infty}^0 x_1 f(x_1) \int_{-\infty}^{+\infty} \frac{y}{x_1} g\left(\frac{y}{x_1}\right) d\frac{y}{x_1} dx_1 \quad x_1 < 0, t = \frac{y}{x_1} \text{ 积分上下限交换} \end{aligned}$$

$$E(X_2) \left(\int_0^{+\infty} x_1 f(x_1) dx_1 + \int_{-\infty}^0 x_1 f(x_1) dx_1 \right) = E(X_2)E(X_1).$$

Theorem 4.7. (随机变量函数期望) 设随机变量 X 为离散型, 有分布律 $P\{X = x_i\} = p_i, i = 1, 2, \dots$, 或者为连续型, 有概率密度函数 $f(x)$. 设 Y 随机变量 X 的函数 $Y = g(X)$ (g 是连续函数), 则

$$E(g(X)) = \sum_{i=1}^{+\infty} g(x_i)p_i, \left(\sum_{i=1}^{+\infty} g(x_i)p_i < \infty \right).$$

或

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx, \left(\int_{-\infty}^{+\infty} g(x)f(x)dx < \infty \right).$$

方差、协方差和矩

Definition 4.8. 设 X 是一个随机变量, 若 $E\{[X - E(X)]^2\}$ 存在, 则称 $E\{[X - E(X)]^2\}$ 为 X 的方差, 记为 $D(X)$ 或者 $\text{Var}(D)$, 即

$$D(X) = \text{Var}(D) = E\{[X - E(X)]^2\}.$$

称 $\sqrt{D(X)}$ 为标准差或者均方差, 记为 $\sigma(X)$.

Definition 4.9. 对离散型随机变量 X 有

$$D(X) = \sum_{i=0}^{+\infty} [x_i - E(X)]^2 p_i.$$

对连续型随机变量 X 有

$$D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 dx.$$

Proposition 4.10.

$$D(X) = E(X^2) - E^2(X).$$

因此有 $D(X) \leq E(X^2)$ 和 $E(X^2) \geq E^2(X)$.

Proposition 4.11. 给定随机变量 X 与 Y , 它们的方差均存在分别为 $D(X)$ 与 $D(Y)$. 给定常数 c , 则有如下性质

1. $D(c) = 0$;
2. $D(X + c) = D(X)$;
3. $D(cX) = c^2 D(X)$;
4. $D(X + Y) = D(X) + D(Y) + 2E[(X - E(X))(Y - E(Y))]$;
5. $D(X) = 0$ 当且仅当 $P(X = c) = 1$, 其中 $c = E(x)$.
6. 对任意常数 c 有, $D(X) \leq E^2(X - c)$, 其中等号成立当且仅当 $c = E(X)$.
7. 如果 X 和 Y 相互独立, a, b 为常数. 则 $D(aX + bY) = a^2 D(X) + b^2 D(Y)$.

证明. □

Definition 4.12. 设 X 为随机变量, c 为常数, k 为正整数. 若

$$E\{X^k\}$$

存在, 则称为 X 的 k 阶原点矩. 若

$$E\{[X - E(X)]^k\}$$

存在, 则称为 X 的 k 阶中心矩. 当 $k = 2$ 的 X 的中心矩就是 X 的方差.

Proposition 4.13. 设随机变量 X 的概率密度 $f(x)$, 若 f 关于某点 a 对称, 即

$$f(x + a) = f(x - a),$$

则 $E(X) = a$, 且 $E\{[X - E(X)]^3\} = 0$.

证明. 根据 $E(X)$ 的定义有

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^a (x + 2a - x)f(x)dx = 2aF(a) = a.$$

设 $Y = [X - E(X)]^3$, 则

$$E(Y) = \int_{-\infty}^{+\infty} (x - a)^3 f(x)dx = \int_{-\infty}^a (x - a)^3 f(x)dx + \int_a^{+\infty} (x - a)^3 f(2a - x)dx$$

令 $t = 2a - x$, 则

$$\int_{-\infty}^a (x - a)^3 f(x)dx + \int_{-\infty}^a (t - a)^3 f(t)dt = 0.$$

□

Annotation 4.14. 设 $\mu = E\{[X - E(X)]^3\}$. 若 $\mu > 0$ 则称 X 的分布右偏; 若 $\mu < 0$ 则称 X 的分布左偏. 对于正态分布而言 $\mu = 0$.

Definition 4.15. 称 $E\{[X - E(X)][Y - E(Y)]\}$ 为随机变量 X, Y 的协方差(Covariance), 记为 $\text{Cov}(X, Y)$.

Proposition 4.16.

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

证明.

$$E\{[X - E(X)][Y - E(Y)]\} = E(XY) - 2E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y).$$

□

Proposition 4.17. 若 c_1, c_2, c_3, c_4 都是常数, 则

$$\text{Cov}(c_1X + c_2, c_3Y + c_4) = c_1c_3\text{Cov}(X, Y).$$

证明.

$$\begin{aligned}
\text{Cov}(c_1X + c_2, c_3Y + c_4) &= E[(c_1X + c_2)(c_3Y + c_4)] - E(c_1X + c_2)E(c_3Y + c_4) \\
&= E(c_1c_3XY + c_1c_4X + c_2c_3Y + c_2c_4) - [c_1c_3E(X)E(Y) + c_1c_4E(X) + c_2c_3E(Y) + c_2c_4] \\
&= c_1c_3[E(XY) - E(X)E(Y)] = c_1c_3\text{Cov}(X, Y)
\end{aligned}$$

□

Proposition 4.18. 若随机变量 X, Y 独立, 则 $\text{Cov}(X, Y) = 0$.

证明. 由 4.16 和 4.6 可证.

□

Proposition 4.19. 给定随机变量 X, Y 的方差分别为 $D(X) = \sigma_1^2$ 与 $D(Y) = \sigma_2^2$, 则

$$[\text{Cov}(X, Y)]^2 \leq \sigma_1^2 \sigma_2^2.$$

等号成立当且仅当 X, Y 之间有严格线性关系, 即存在常数 a, b 使得 $Y = aX + b$.

证明. 首先考虑等式

$$E[t(X - E(X)) + (Y - E(Y))]^2 = \sigma_1^2 t^2 + 2\text{Cov}(X, Y)t + \sigma_2^2. \quad (1)$$

(1) 式左边是一个非负随机变量的期望, 所以它对任意的 t 都是非负的. 那么左边是个关于 t 的二次多项式, 非负就意味至多有且只有一个根, 因此

$$\sigma_1^2 \sigma_2^2 \leq [\text{Cov}(X, Y)]^2. \quad (2)$$

必要性 若 (2) 式等号成立, 那么 (1) 式 RHS 就可以写成 $(\sigma_1 t \pm \sigma_2)^2$. 其中 \pm 取决于 $\text{Cov}(X, Y)$ 的符号, 方便证明我们这里假设 $\text{Cov}(X, Y) > 0$, 则 (1) 式 RHS 为 $(\sigma_1 t + \sigma_2)^2$. 于是 $t = t_0 = -\frac{\sigma_2}{\sigma_1}$ 时, (1) 式就等于 0, 即

$$E[t_0(X - E(X)) + (Y - E(Y))]^2 = 0.$$

这里需要使用一个事实, 当随机变量 Z 只能取非负值时, 若 $E(Z) = 0$, 则 $Z = 0$. 证明它可以假设 $Z \neq 0$, 则必有 Z 必有取大于 0 值时的概率大于 0, 这将导致 $E(Z) > 0$. 于是这里有

$$t_0(X - E(X)) + (Y - E(Y)) = 0,$$

因此 X, Y 之间有严格线性关系.

充分性 若 $Y = aX + b$, 则 $\sigma_2^2 = D(aX + b) = a^2 \sigma_1^2$; 而 $E(Y) = E(aX + b) = aE(X) + b$. 于是

$$\text{Cov}(X, Y) = E\{(X - E(X))(aX + b - aE(X) - b)\} = a\sigma_1^2.$$

因此

$$[\text{Cov}(X, Y)]^2 = a^2 \sigma_1^4 = \sigma_1^2 (a^2 \sigma_1^2) = \sigma_1^2 \sigma_2^2.$$

□

Definition 4.20. 给定随机变量 X, Y , 则

$$\frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为 X, Y 的**相关系数**, 记为 ρ_{XY} .

Annotation 4.21. 形式上可以把相关系数视为“标准尺度下的协方差”. 协方差作为 $[X - E(X)][Y - E(Y)]$ 的均值, 依赖于 X, Y 的度量单位, 选择适当单位使得 X, Y 的方差都为 1, 则协方差就是相关系数. 这样就能更好的反应 X, Y 之间的关系, 不受所用单位的影响. 设 $D(X) = \sigma_1^2$ 与 $D(Y) = \sigma_2^2$.

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_1 \sigma_2} = \text{Cov}\left(\frac{X}{\sigma_1}, \frac{Y}{\sigma_2}\right).$$

此时 $D\left(\frac{X}{\sigma_1}\right) = D\left(\frac{Y}{\sigma_2}\right) = 1$.

Theorem 4.22. 给定随机变量 X, Y , 它们的协方差为 ρ_{XY} . 则

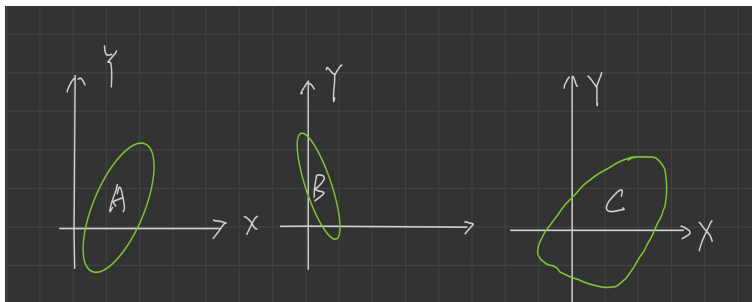
1. $-1 \leq \rho_{XY} \leq 1$;
2. $|\rho_{XY}| = 1$ 当且仅当 X 和 Y 有严格线性关系.

证明. 直接由4.19可得. □

Definition 4.23. 当 $\rho_{XY} = 0$ (或 $\text{Cov}(X, Y) = 0$) 时, 称 X, Y **不相关**.

Proposition 4.24. 若随机变量 X, Y 独立, 则 X, Y 不相关. 反过来并不一定成立.

Annotation 4.25. **如何理解随机变量的之间线性关系的程度?** 如果 $0 < |\rho_{X,Y}| < 1$, 则解释为 X, Y 之前有“一定程度的”线性关系而非严格的线性关系. 何谓“一定程度”的线性关系, 用下图来说明. 在下面的三个图中, 我们都假定 (X, Y) 都服从区域 A, B, C 内的均匀分布. 在这三个图中, X, Y 都无严格的线性关系, 因为由 X 的值并不能确定 Y 的值. 可是由这几个图我们都能感觉出, X, Y 之间存在一种线性趋势. 这种趋势, 在第一幅图中 X 增加时, Y 也倾向于增加 (实际这里 $\text{Cov}(X) > 0$, 关于这一点从4.19的 t_0 的取值也可以看的出来). 第二幅图这种线性趋势更加明显 (这里 $\text{Cov}(X) < 0$). 第三幅图相比于前两幅图只有一点线性倾向, 但是已经比较微弱 (这里 $\text{Cov}(X) > 0$ 但是接近于 0).



Proposition 4.26. 最小二乘下的线性逼近 设有两个随机变量 X, Y . 给定两个实数 a, b , 若要使得 $E\{[Y - (aX + b)]^2\}$ 达到最小, 则得

$$a = \frac{\text{Cov}(X, Y)}{D(X)}$$

和

$$b = E(Y) - E(X) \frac{\text{Cov}(X, Y)}{D(X)}.$$

其对应的最小值为

$$E\{[Y - (aX + b)]^2\} = (1 - \rho_{XY}^2)D(Y).$$

证明. 我们设 $E(X) = m_1, E(Y) = m_2, D(X) = \sigma_1^2, D(Y) = \sigma_2^2$. 于是

$$E\{[Y - (aX + b)]^2\} = E\{[(Y - m_2) - a(X - m_1) + m_2 - am_1 - b]^2\}.$$

记 $c = m_2 - am_1 - b$, 则

$$E\{[(Y - m_2) - a(X - m_1) + c]^2\} = \sigma_2^2 + a^2\sigma_1^2 - 2a\text{Cov}(X, Y) + c^2.$$

对上式分布求偏导并令它们等于 0, 可得 $c = 0$ 和 $a = \frac{\text{Cov}(X, Y)}{D(X)}$, 由此马上可到

$$b = m_2 - am_1 = E(Y) - E(X) \frac{\text{Cov}(X, Y)}{D(X)}.$$

带入 b, c 即得

$$E\{[Y - (aX + b)]^2\} = \sigma_2^2 + \frac{[\text{Cov}(X, Y)]^2}{\sigma_1^2} - 2 \frac{[\text{Cov}(X, Y)]^2}{\sigma_1^2} = \sigma_2^2(1 - \rho_{XY}^2).$$

□

常用分布的数字特征

表 3.1: 常见分布表

| 分布名称 | 参数 | 概率密度 | 期望 | 方差 | 特征函数 | 备注 |
|--------------------------|---|---|---------------------|--|-------------------------------------|------------------------|
| 退化分布 | c | $\binom{c}{1}$ | c | 0 | e^{ict} | 关于参数 c 具有再生性 |
| 二点分布 | p $(0 < p < 1)$ | $\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$ | p | pq | $q + pe^{it}$ | 为 $n = 1$ 的二项分布 |
| 二项分布 $B(n, p)$ | $n \geq 1$ $0 < p < 1$ | $\binom{n}{k} p^k q^{n-k}$ $k = 0, \dots, n$ | np | npq | $(q + pe^{it})^n$ | 关于 n 具有再生性 |
| 几何分布 | p $(0 < p < 1)$ | $q^{k-1} p, k = 1, 2, \dots$ | $\frac{1}{p}$ | $\frac{q}{p^2}$ | $\frac{pe^{it}}{1 - qe^{it}}$ | 无记忆 |
| 巴斯卡分布 | r, p $r \in \mathbb{N}$ $0 < p < 1$ | $\binom{k-1}{r-1} p^r q^{k-r},$ $k = r, r + 1, \dots$ | $\frac{r}{p}$ | $\frac{rq}{p^2}$ | $(\frac{pe^{it}}{1 - qe^{it}})^r$ | 关于 r 具有再生性 |
| 波松分布 $P(\lambda)$ | $\lambda (\lambda > 0)$ | $\frac{\lambda^k}{k!} e^{-\lambda},$ $k = 0, 1, \dots$ | λ | λ | $e^{\lambda(e^{it} - 1)}$ | 关于 λ 具有再生性 |
| 超几何分布 | $M, N, n \in \mathbb{N}$ | $\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ | $\frac{nM}{N}$ | $\frac{nM}{N} \frac{(N-M)}{N} \frac{N-n}{N-1}$ | | |
| 均匀分布 $U(a, b)$ | $a, b (a < b)$ | $\frac{1}{b-a} I_{a < x < b}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | $\frac{e^{itb} - e^{ita}}{it(b-a)}$ | |
| 正态分布 $N(a, \sigma^2)$ | a, σ^2 | $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ | a | σ^2 | $e^{iat - \frac{1}{2}\sigma^2 t^2}$ | 关于 a, σ^2 具有再生性 |
| 指数分布 | $\lambda (\lambda > 0)$ | $\lambda e^{-\lambda x} I_{x > 0}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | $(1 - \frac{it}{\lambda})^{-1}$ | |
| χ^2 分布 | $n (n \geq 1)$ | $\frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}$ $x > 0$ | n | $2n$ | $(1 - 2it)^{-n/2}$ | 关于 n 具有再生性 |

13

大数定理和中心极限定理

Theorem 5.1. (切比雪夫不等式) 设随机变量 X 具有数学期望 $E(X) = \mu$, 方差为 $D(X) = \sigma^2$, 则有

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

证明. 若 X 是连续性随机变量, 其概率密度为 $f(x)$, 则

$$P\{|X - \mu| \geq \varepsilon\} = \int_{|x-\mu| \geq \varepsilon} f(x)dx \leq \int_{|x-\mu| \geq \varepsilon} \frac{(x-\mu)^2}{\varepsilon^2} f(x)dx \leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = \frac{\sigma^2}{\varepsilon^2}$$

□

Annotation 5.2. 切比雪夫不等式用于在已知 $E(X)$ 和 $D(X)$ 的情况下估计概率 $P\{|X - \mu| \geq \varepsilon\}$.

Theorem 5.3. (马尔科夫不等式) 设 X 为非负随机变量, 则对任意数 $\varepsilon > 0$ 有

$$P\{X \geq \varepsilon\} \leq \frac{E(X)}{\varepsilon}.$$

证明. 若 X 是连续性随机变量, 其概率密度为 $f(x)$. 因 X 非负, 所以有 $y < 0, f(y) = 0$, 故

$$E(X) = \int_0^{+\infty} xf(x)dx \geq \int_{\varepsilon}^{+\infty} xf(x)dy.$$

当 $x \in [\varepsilon, +\infty]$, 有 $x \geq \varepsilon$. 于是有

$$E(X) \geq \int_{\varepsilon}^{+\infty} xf(x)dx \geq \varepsilon \int_{\varepsilon}^{+\infty} f(x)dx = \varepsilon P\{X \geq \varepsilon\}.$$

□

Theorem 5.4. (大数定理) 设 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的随机变量, 即 $E(X_i) = \mu, i = 1, 2, \dots$. 则对任意数 $\varepsilon > 0$ 有

$$\lim_{n \rightarrow \infty} P\{|\bar{X}_n - \mu| \geq \varepsilon\} = 0,$$

其中 $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.

证明. 设 $D(X_i) = \sigma^2, i = 1, 2, \dots$. 由切比雪夫不等式有

$$P\{|\bar{X}_n - \mu| \geq \varepsilon\} \leq \frac{D(\bar{X})}{\varepsilon^2} = \frac{n \frac{\sigma^2}{n^2}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

因此

$$\lim_{n \rightarrow \infty} P\{|\bar{X}_n - \mu| \geq \varepsilon\} = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0.$$

□

Definition 5.5. 大数定理的特例 若随机变量 $X_n \sim B(n, p) (n = 1, 2, \dots)$, 则对于任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right\} = 0.$$

频率趋于概率.

Theorem 5.6. 中心极限定理 设 $X_1, X_2, \dots, X_n, \dots$ 为独立同分布的随机变量, $E(X_i) = \mu, D(X_i) = \sigma^2, i = 1, 2, \dots$. 则对任意实数 x , 有

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{\sqrt{n}\sigma} \left(\sum_{i=1}^n X_i - n\mu\right) \leq x\right\} = \Phi(x)$$

这里 $\Phi(x)$ 是标准正态分布 $N(0, 1)$ 的分布函数.

Annotation 5.7. 注意 $\frac{1}{\sqrt{n}\sigma}(\sum_{i=1}^n X_i - n\mu)$ 的期望为 0 方差为 1, 这个操作称为 $X_1 + X_2 + \dots + x_n$ 的标准化, 即与 $N(0, 1)$ 的期望和方差符号.

Theorem 5.8. 中心极限定理的一个特例 设随机变量 $X_n \sim B(n, p) (n = 1, 2, \dots)$, 则对于任意的实数 x , 有

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{\sqrt{np(1-p)}}(X_n - np) \leq x\right\} = \Phi(x)$$

数理统计的基本概念

总体、样本、统计量和样本数字特征

Definition 6.1. 对推断统计而言, 主要有如下两大类:

1. **参数估计** 在概率模型 (分布) 假定下, 根据数据用一些方法对分布的位置参数进行估计.
2. **假设检验** 对某种假设做出推断, 即根据数据, 用一些方法对分布或者分布中的未知数参数进行检验.

Definition 6.2. 数理统计中所研究对象的某项数量指标的全体称为**总体**, 而把总体中的每个成员称为**个体**.

Definition 6.3. 若用 X 表示所研究对象的某一指标, 则总体即为 X (的分布). 从总体中抽取的 n 个随机变量 X_1, \dots, X_n 称为**样本**, n 称为**样本容量**, 而样本 X_1, \dots, X_n 的值 x_1, \dots, x_n 称为样本的**观察值**. 当 X_1, \dots, X_n 独立同分布时, 称为 X_1, \dots, X_n 为**简单样本** (后面若无特殊说明, 所说的样本都是指简单样本).

Proposition 6.4. 若总体 X 的分布为 $F(x)$, 则样本 X_1, \dots, X_n 的联合分布为

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F(x_i).$$

若 X 是离散型随机变量, 那么其概率分布为

$$P\{X_1 = x_1, \dots, X_n = x_n\} = \prod_{i=1}^n P\{X = x_i\}.$$

若 X 是连续随机变量, 那么其概率密度为

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i).$$

Definition 6.5. 关于样本 X_1, \dots, X_n 的不含未知参数的函数 $T = T(X_1, \dots, X_n)$ 称为**统计量**. 也就是只依赖样本的量称为统计量.

Definition 6.6. 常见的统计量

1. 样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, 反映了总体总结的信息.
2. 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, 反映了总体方差的信息.
3. 样本 k 阶原点矩 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$, 反映了总体 k 阶原点矩的信息.

4. 样本 k 阶中心矩 $B_k = \frac{1}{n} \sum_{i=1}^n X_i - \bar{X}^k$, 反映了总体 k 阶中心矩的信息. 特别地 $B_2 = \frac{n-1}{n} S^2$.

Definition 6.7. 将样本 X_1, \dots, X_n 的观测值 x_1, \dots, x_n 按大小排序, 得到 $x_{(1)} \leq \dots \leq x_{(n)}$, 则称

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{|x|}{n}, & x_{(1)} \leq x \leq x_{(n)} \\ 1, & x > x_{(n)} \end{cases}$$

为经验分布函数.

统计三大分布

Definition 6.8. 设 X_1, \dots, X_n 为相互独立且具有共同分布 $N(0, 1)$ 的随机变量, 则称 $X = \sum_{i=1}^n X_i^2$ 的分布服从自由度为 n 的 χ^2 分布, 记为 $X \sim \chi_n^2$.

Proposition 6.9. χ_n^2 分布的概率密度为

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{other} \end{cases}$$

证明.

□

参数估计

点估计

Definition 7.1. 设有一个总体 X , 以 $f(x, \theta_1, \dots, \theta_k)$ 记其概率密度函数 (若总体分布为连续型) 或者概率函数 (若总体分为离散型), 其中 $\theta_1, \dots, \theta_k$ 表示总体分布包含的 k 个未知参数, 用样本 X_1, \dots, X_n 构造的统计量 $\hat{\theta}_i(X_1, \dots, X_n)$ 来估计未知参数 θ_i 称为点估计, 统计量 $\hat{\theta}_i(X_1, \dots, X_n)$ 来估计未知参数称为估计量.

矩估计

Definition 7.2. 设总体 X 的分布有未知参数 $\theta_1, \dots, \theta_k$, 设其有 l 阶原点矩 a_l 存在, 设样本的 l 阶原点矩为 $A_l = \frac{1}{n} \sum_{i=1}^n X_i^l$. 于是

$$a_l = a_l(\theta_1, \dots, \theta_k) = A_l,$$

取 $l = 1, 2, \dots, k$, 就得到一个方程组, 解此方程组, 得其根为

$$\hat{\theta}_i = \hat{\theta}_i(X_1, \dots, X_n), i = 1, 2, \dots, k.$$

以 $\hat{\theta}_i$ 作为 θ_i 的一个估计, 这样的估计量叫做矩估计.

极大似然估计

Definition 7.3. 设总体 X 有分布 $f(X; \theta_1, \dots, \theta_k)$ (即概率密度或者概率分布), 给定样本 X_1, \dots, X_n , 则样本 (X_1, \dots, X_n) 的分布为

$$f(X_1; \theta_1, \dots, \theta_k) f(X_2; \theta_1, \dots, \theta_k) \cdots f(X_n; \theta_1, \dots, \theta_k),$$

记为 $L(X_1, \dots, X_n; \theta_1, \dots, \theta_k)$, 称其为似然函数. 这函数对不同的 $(\theta_1, \dots, \theta_k)$ 的取值, 反映了在观察结果 (X_1, \dots, X_n) 已知的条件下, $(\theta_1, \dots, \theta_k)$ 的各种值的似然程度. 应用似然程度最大的那个 $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ 作为 $(\theta_1, \dots, \theta_k)$ 的估计值. 这个估计 $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ 就叫做 $(\theta_1, \dots, \theta_k)$ 的极大似然估计.

Annotation 7.4. 多元函数求极值的场景.

Example 7.5. 设 X_1, \dots, X_n 是从正态总体 $N(\mu, \sigma^2)$ 中抽出的样本, 则似然函数为

$$L = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2}.$$

等式两边取对数有

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2.$$

于是

$$\begin{cases} \frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) = 0 \\ \frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (X_i - \mu)^2 = 0 \end{cases}$$

显然有 $\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. 我们还需要判定其是不是最大值, 其二阶导分别为

$$\begin{cases} \frac{\partial^2 L}{\partial \mu \partial \mu} = -\frac{n}{\sigma^2} \\ \frac{\partial^2 L}{\partial \mu \partial \sigma^2} = -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n (X_i - \mu) \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \mu} = -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n (X_i - \mu) \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n (X_i - \mu)^2 \end{cases}$$

故 L 有二阶连续偏导, 根据极值判定法则

$$L''_{\mu\mu}(\hat{\mu}, \hat{\sigma}^2) \cdot L''_{\sigma^2\sigma^2}(\hat{\mu}, \hat{\sigma}^2) - \left[L''_{\mu\sigma^2}(\hat{\mu}, \hat{\sigma}^2) \right]^2 = \frac{-n}{\hat{\sigma}^2} \cdot \frac{-n}{2(\hat{\sigma}^2)^2} > 0.$$

确实在 $(\hat{\mu}, \hat{\sigma}^2)$ 取得最大值.