

概率论

第 1 章 事件的概率

组合公式

事件的运算

第 2 章 随机变量及概率分布

2.1 一维随机变量

2.1.1 离散型随机变量

- 1 二项分布
- 2 泊松分布
- 3 超几何分布
- 4 负二项分布
- 5 几何分布
- 5' 几何分布

2.1.2 连续型随机变量

- 1 正态分布
- 2 指数分布
- 3 威布尔分布
- 4 均匀分布
- 5 对数正态分布
- 6 柯西分布
- 7 拉普拉斯分布

2.2 多维随机变量

2.2.1 离散性随机向量

- 1 多项分布

2.2.2 连续型随机向量

- 1 矩形均匀分布
- 2 二维正态分布
- 3 多元正态分布

2.2.3 边缘分布

- 1 概念解释
- 2 多项分布
- 3 二维正态分布

2.3 条件概率分布与随机变量的独立性

2.3.1 条件概率分布的概念

2.3.2 离散性随机变量的条件概率分布

2.3.3 连续性随机变量的条件概率分布

2.3.4 随机变量的独立性

2.4 随机变量的函数的概率分布

2.4.1 离散性分布

2.4.2 连续型分布

1 单变量函数

1.1 严格单调

1.2 幂函数

2 多变量函数

2.4.3 随机变量和的密度函数

2.4.4 随机变量商的密度函数

2.4.5 多个随机变量的排序

1 独立同分布的最值

2 两项最值联合分布

3 次序统计量的分布

注意事项

第 3 章 随机变量的数字特征

3.1 数学期望与中位数

3.1.1 数学期望的定义

3.1.2 数学期望的性质

3.1.3 条件数学期望 (条件均值)

- 3.1.4 中位数
- 3.2 方差与矩
 - 3.2.1 方差和标准差
 - 3.2.2 原点矩与中心矩
 - 3.2.3 基于矩的系数
- 3.3 协方差与相关系数
 - 3.3.1 协方差
 - 3.3.2 相关系数
- 3.4 大数定理和中心极限定理
 - 3.4.1 大数定理
 - 3.4.2 中心极限定理
- 3.5 母函数
 - 3.5.1 母函数的定义
 - 3.5.2 常见分布的母函数
 - 3.5.3 母函数的性质
 - 3.5.4 母函数的定理
- 3.6 特征函数
 - 3.6.1 特征函数的定义
 - 3.6.2 常见分布的特征函数
 - 3.6.3 特征函数的性质
- 3.7 矩量母函数
 - 3.7.1 相关定义
 - 3.7.2 常见分布的矩量母函数
 - 3.7.3 矩量母函数的性质
- 例题

第 1 章 事件的概率

组合公式

由 $(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$ 知

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$$
$$\binom{n}{0} - \binom{n}{1} + \cdots + (-1)^{-1} \binom{n}{n} = 0$$

由 $(1+x)^{m+n} = (1+x)^m (1+x)^n$ 知

$$\binom{m+n}{k} = \sum_{i=0}^k \binom{m}{i} \binom{n}{k-i}$$

特别地, 当 $m = k = n$ 时,

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

第一式还可以写作下式, 并可以从直观上理解.

$$\sum_{k_1+k_2=k} \binom{n_1}{k_1} \binom{n_2}{k_2} = \binom{n_1+n_2}{k_1+k_2}$$

此外还有下式, 也可以从直观上理解.

$$\sum_{n_1+n_2=n} \binom{n_1}{k_1} \binom{n_2}{k_2} = \binom{n_1+n_2+1}{k_1+k_2+1}$$

多项式系数: $\frac{n!}{r_1! \cdots r_k!}$.

利用第一式 (杨辉恒等式) 数归得第二式 (或直观理解)

$$\binom{n+m}{m} + \binom{n+m}{m+1} = \binom{n+m+1}{m+1}$$
$$\sum_{r=0}^m \binom{n-1+r}{r} = \binom{n+m}{m}$$

由负指数二项展开式 $(1-x)^{-r} = \sum_{i=0}^{\infty} \binom{-r}{i} (-x)^i = \sum_{i=0}^{\infty} \binom{i+r-1}{r-1} x^i$ 知

$$\sum_{i=0}^{\infty} \binom{i+r-1}{r-1} = 0$$
$$\sum_{i=0}^{\infty} \binom{i+r-1}{r-1} (-1)^i = 2^{-r}$$
$$p^{-r} = \sum_{i=0}^{\infty} \binom{i+r-1}{r-1} (1-p)^i$$

原式两边求导, 并令 $x = 1-p$, 得

$$rp^{-r-1} = \sum_{i=0}^{\infty} i \binom{i+r-1}{r-1} (1-p)^{i-1}$$

[Stirling 数](#), [拆分数](#), [装箱问题](#), [Burnside 定理与 Polya 定理](#)

事件的运算

- 记号
 - $A + B \equiv A \cup B$.
 - $AB \equiv A \cap B$.
 - $A - B \equiv \overline{AB}$.
- 加法
 - 交换律: $A + B = B + A$.
 - 结合律: $(A + B) + C = A + (B + C)$.
于是可定义 $A + B + C = (A + B) + C$.
 - 自加: $A + A = A$.
- 乘法
 - 交换律: $AB = BA$.
 - 结合律: $(AB)C = A(BC)$.
于是可定义 $ABC := (AB)C$.
 - 自乘: $AA = A$.
- 分配律
 - 加法与乘法: $(A + B)C = AC + BC$.
 - 减法与乘法: $(A - B)C = AC - BC$.
本质为 $ABC = (AC)(BC)$, 这个式子在推导中是有用的.
 - 乘法与加法: $(AB) + C = (A + C)(B + C)$.
 - 乘法与减法: $(AB) - C = (A - C)(B - C)$.
- 减法
 - $A - B = \overline{AB} \neq A + (-B)$.
 - $A \subseteq B \Leftrightarrow A - B = \emptyset$.
 - $A = B \Leftrightarrow A - B = B - A = \emptyset$.
- 无消去律
 - $A + B = A + C \not\Rightarrow B = C$.
 - $A - B = A - C \not\Rightarrow B = C$.
 - 注: $A + AB = A$.
- 混合运算
 - $(A + B) - C \neq A + (B - C)$. (因为减法本质上是乘法)
因此 $A + B - C$ 没有意义, 除非定义运算顺序或优先级.
 - $A - (B + C) = A - B - C = (A - C) - (B - C)$.
 - $A - (B - C) = (A - B)C = AC - BC$.
- 负号 (补集)
 - $-(A + B) = (-A)(-B)$.
 - $-(A - B) = B - A = (-A)B$.

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

理论上可以这么写,实际上用 \overline{A} 的符号会更方便.

- 互斥

- A 与 B 互斥 $\Leftrightarrow AB = \emptyset \Leftrightarrow P(AB) = 0$.
- $AC = BC \Leftrightarrow A - B$ 与 $B - A$ 均与 C 互斥 $\Leftrightarrow P(\overline{A}BC) = P(\overline{A}\overline{B}C) = 0$.
当且仅当 $C = \Omega$ 时,可由此推出 $A = B$.
- A 与 B 互斥 $\Rightarrow AC$ 与 BC 互斥 $\Rightarrow P(C(A + B)) = P(AC) + P(BC)$.

- 对立

- A 与 B 对立 $\Leftrightarrow AB = \emptyset$ 且 $A + B = \Omega$.
- $\overline{A_1 A_2 \cdots A_n} = \overline{A_1} + \overline{A_2} + \cdots + \overline{A_n}$.
- $\overline{A_1 + A_2 + \cdots + A_n} = \overline{A_1} \overline{A_2} \cdots \overline{A_n}$.

- 条件概率

- 定义: $P(A | B) = P(AB)/P(B)$.
- 全概率公式: 若两两互斥的 B_i 之交为必然事件,则 $P(A) = P(B_1)P(A | B_1) + P(B_2)P(A | B_2) + \cdots$.
- 贝叶斯公式:

$$\begin{aligned} \blacksquare P(B | A) &= \frac{P(B)P(A | B)}{P(A)}. \\ \blacksquare P(B_i | A) &= \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A | B_i)}{\sum P(B_j)P(A | B_j)}. \end{aligned}$$

- 几率

- 定义: $O(A) = \frac{P(A)}{1 - P(A)} = \frac{P(A)}{P(\overline{A})}$.
- 贝叶斯公式: $O(B | A) = \frac{P(B | A)}{P(\overline{B} | A)} = \frac{P(B)P(A | B)}{P(A)P(\overline{B} | A)} = \frac{P(B)P(A | B)}{P(\overline{B})P(A | \overline{B})} = O(B) \frac{P(A | B)}{P(A | \overline{B})}$.
- 贝叶斯因子: $BF = \frac{P(A | B)}{P(A | \overline{B})}$, 故 $O(B | A) = BF \cdot O(B)$.

- 促进作用的性质

- 具有对称性: A 促进 B , 则 B 促进 A , 即
 $P(A | B) > P(A) \Leftrightarrow P(B | A) > P(B)$.
- 不具有传递性: B 促进 A 且 C 促进 B 不能推出 C 促进 A , 即
 $P(A | B) > P(A), P(B | C) > P(B) \not\Rightarrow P(A | C) > P(A)$.
- 若 B 和 C 都促进 A , 则 $B + C$ 一定促进 A , 但 BC 和 $B - C$ 不一定促进 A , 即
 $P(A | B) > P(A), P(A | C) > P(A) \Leftrightarrow P(A | B + C) > P(A)$.
- 若 B 促进 A , 则 \overline{B} 抑制 A , B 抑制 \overline{A} , \overline{B} 促进 \overline{A} , 即
 $P(A | B) > P(A) \Leftrightarrow P(A | \overline{B}) < P(A) \Leftrightarrow P(\overline{A} | B) < P(\overline{A}) \Leftrightarrow P(\overline{A} | \overline{B}) > P(\overline{A})$.
 $P(A | B) = P(A) \Leftrightarrow P(A | \overline{B}) = P(A) \Leftrightarrow P(\overline{A} | B) = P(\overline{A}) \Leftrightarrow P(\overline{A} | \overline{B}) = P(\overline{A})$.

- 独立

- A 与 B 独立 $\Leftrightarrow P(AB) = P(A)P(B) \Leftrightarrow P(A | B) = P(A) \cdot (P(B) \neq 0)$
- 两两独立: $\forall i, j (1 \leq i, j \leq n, i \neq j) : P(A_i A_j) = P(A_i)P(A_j)$.
- 相互独立: $\forall 1 < k \leq n, 1 \leq i_1 < i_2 < \cdots < i_k \leq n : P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$.
- 相互独立 \Rightarrow 两两独立, 反之不一定成立.
- 独立事件的任一部分也独立.

- 若 A_1, A_2, \dots, A_n 独立, $B_i = A_i$ 或 \bar{A}_i , 则 B_1, B_2, \dots, B_n 也独立.
- 独立事件的概率
 - 乘法: $P(\prod E_i) = \prod P(E_i)$.
 - 加法: $P(\sum E_i) = 1 - P(\prod \bar{E}_i) = 1 - \prod P(\bar{E}_i)$.
 - 实例:

$$P(E_0 + E_1 E_2) = 1 - P(\bar{E}_0 \bar{E}_1 \bar{E}_2) = 1 - (1 - P(E_0))(1 - P(E_0 E_1))$$

$$= P(E_0) + P(E_1)P(E_2) - P(E_0)P(E_1)P(E_2).$$
- 运算定理
 - 加法定理 (并集): $P(A + B) = P(A) + P(B) \Leftrightarrow AB = \emptyset \Leftrightarrow P(AB) = 0$.
 - 减法定理 (差集): $P(A - B) = P(A) - P(B) \Leftrightarrow A \supseteq B \Leftrightarrow P(B - A) = 0$.
 - 乘法定理 (交集): $P(AB) = P(A)P(B) \Leftrightarrow A$ 与 B 独立.
 - 加法推论 (补集): $P(A^c) = P(\bar{A}) = 1 - P(A)$ 恒成立.
- 表为互斥事件
 - $$\sum A_i = A_1 + \bar{A}_1 A_2 + \dots + \bar{A}_1 \bar{A}_2 \dots \bar{A}_{n-1} A_n.$$
 - $A + B = A + (B - A).$
 - $A + B + C = A + (B - A) + (C - B - A).$
 - 设 $f(m) = \sum_{k,l} \left(\prod_{i=1}^m A_{k_i} \prod_{j=1}^{n-m} \bar{A}_{l_j} \right)$, 则 $\sum_{i=1}^n A_i = \sum_{m=1}^n f(m)$.
 - $A + B = (B - A) + (A - B) + AB.$
 - $$A + B + C = ABC + (BC - A) + (AC - B) + (AB - C) + (A - B - C) + (B - A - C) + (C - A - B)$$
 - 综合应用
 - $A + B = A + (B - A) = (A - B) + B = (B - A) + (A - B) + AB,$
 即 $A + B = A + \bar{A}B = \bar{A}B + A\bar{B} + AB.$
 - $P(A + B) = P(A) + P(\bar{A}B) = P(\bar{A}B) + P(A\bar{B}) + P(AB).$
- 容斥原理
 - $P(A + B) = P(A) + P(B) - P(AB)$, 或 $P(AB) = P(A) + P(B) - P(A + B).$
 - $P(\bar{A}\bar{B}) = P(\bar{A} + \bar{B}), P(\bar{A}\bar{B}) + P(\bar{A}B) = P(\bar{A}) + P(\bar{B}).$
 - $P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC).$
- 恒等式
 - 化简含括号的运算
 - $(A + B) + (A - B) = A + B.$
 - $(A + B) - (A - B) = B.$
 - $(A - B) + (B - A) = (A + B) - AB.$
 - $(A - B) - (B - A) = A - B.$
 - 有用的概率恒等式
 - $A - B = A - AB$, 或 $\bar{A}\bar{B} = \overline{AAB}.$
 - $P(A - B) = P(A) - P(AB)$. (利用减法定理)
 - $P(AB) = P(A) - P(\bar{A}B).$
 - $P(\bar{A}\bar{B}) = P(A) - P(AB) = P(A) - P(B) + P(\bar{A}B).$
- 例题

- 欧拉装错信封: $P_n = \sum_{k=0}^n \frac{(-1)^k}{k!}$.
- 先胜 n 局者为胜, 甲 a 胜 b 负, 则甲胜的概率为: $P_n(a, b) = \sum_{i=1}^{n-b} p^{n-a-1+i} (1-p)^{n-b-i} \binom{2n-a-b-1}{n-a-1+i}$.
- 设 n 个独立事件 A_1, A_2, \dots, A_n 的概率分别为 p_1, p_2, \dots, p_n , 记 $p = p_1 + p_2 + \dots + p_n$, 则
 - A_1, A_2, \dots, A_n 都不发生的概率小于 e^{-p} .
 - A_1, A_2, \dots, A_n 中至少发生 k 个的概率小于 $p^k/k!$.
- 蒲丰投针问题: $p = \frac{2l}{\pi a}$.

概率的公理化定义

1. 非负性.
2. 规范性.
3. 可列可加性.

定理 1 独立事件的交与并

若 A 和 B 均与 C 独立, 则 AB 与 C 独立 $\Leftrightarrow A+B$ 与 C 独立.

证明

• 法一

1. 必要性

$$\begin{aligned}
 P((A+B)C) &= P(AC+BC) \\
 &= P(AC) + P(BC) - P(ABC) \\
 &= P(C)(P(A) + P(B) - P(AB)) \\
 &= P(C)P(A+B).
 \end{aligned}$$

2. 充分性

$$\begin{aligned}
 P(ABC) &= P((AC)(BC)) \\
 &= P(AC) + P(BC) - P((A+B)C) \\
 &= P(C)(P(A) + P(B) - P(A+B)) \\
 &= P(C)P(AB).
 \end{aligned}$$

• 法二

1. 必要性

$$\begin{aligned}
 P((A+B)C) &= P((A-B)C) + P((B-A)C) + P(ABC) \\
 &= P(AC-ABC) + P(BC-ABC) + P(ABC) \\
 &= P(A)P(C) - P(AB)P(C) + P(B)P(C) - P(AB)P(C) + P(AB)P(C) \\
 &= P(C)(P(A) + P(B) - P(AB)) \\
 &= P(C)P(A+B).
 \end{aligned}$$

2. 充分性

$$\begin{aligned}
 P(ABC) &= P((A+B)C) - P((A-B)C) - P((B-A)C) \\
 &= P(A+B)P(C) - P(AC-ABC) - P(BC-ABC) \\
 &= P(A+B)P(C) - P(A)P(C) + P(ABC) - P(B)P(C) + P(ABC) \\
 &= P(AC) + P(BC) - P((A+B)C) \\
 &= P(C)(P(A) + P(B) - P(A+B)) \\
 &= P(C)P(AB).
 \end{aligned}$$

推论 增加互斥条件的充分条件

若 A 和 B 均与 C 独立, 且 A 与 B 互斥, 则 AB 与 $A + B$ 均与 C 独立.

定理 2 相互独立的充要条件

设 $0 < P(A) < 1$, 则 $P(B | A) = P(B | \bar{A})$ 是事件 A, B 相互独立的充要条件.

证明

1. 必要性

$$\begin{aligned} P(B) &= P(AB) + P(\bar{A}B) \\ &= P(A)P(B | A) + P(\bar{A})P(B | \bar{A}) \\ &= P(B | A) = P(AB)/P(A). \end{aligned}$$

2. 充分性

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

第 2 章 随机变量及概率分布

2.1 一维随机变量

建议使用常见且最不容易混淆的词语 (已用黑体标出):

- **离散型随机变量**
 - **概率分布律**, 或概率分布, 或概率律, 或分布律.
Law of Probability Distribution.
 - **分布列**, 分布表.
Probability Distribution.
 - **类比数列的列表法**.
 - 适用于一维或二维.
 - **概率质量函数**, 或概率函数.
pmf, Probability Mass Function.
 - **类比数列的通项公式**.
 - 取值即概率.
 - **累积分布函数**, 或分布函数, 或累积函数.
CDF, Cumulative Distribution Function.
 - 单调, 有界, **右连续**.
- **连续型随机变量**
 - **概率密度函数**, 或概率函数, 或密度函数.
pdf, Probability Density Function.
 - 取值非负, 与 x 轴围成面积为 1.
 - 定积分后为概率.
 - **累积分布函数**, 或分布函数, 或累积函数.
CDF, Cumulative Density Function.
 - 单调, 有界, **处处连续**.
- **互补累积分布函数**, 或生存函数 (Survival Function),
或残存函数 (Survivor Function), 或可靠性函数 (Reliable Function)
CCDF, Complementary Cumulative Distribution Function.

此外, **各分布的定义在不同资料上可能不同, 请注意区分**.

常见分布更详细的信息请见笔记附录 ([源代码](#), [PDF](#), 或 [HTML](#))

2.1.1 离散型随机变量

1 二项分布

$$X \sim B(n, p).$$

理解: 事件发生的概率为 p , 则重复 n 次试验, 事件发生的次数为 x .

$$\text{概率分布: } P(X = i) = b(i; n, p) = \binom{n}{i} p^i (1 - p)^{n-i}.$$

$$\text{最可能数: } x = \lfloor (n + 1)p \rfloor.$$

2 泊松分布

$$X \sim P(\lambda).$$

理解: 单位时间内事件平均发生 λ 次, 则某一段单位时间内发生的次数为 x .

$$\text{概率分布: } P(X = i) = \lim_{n \rightarrow \infty} b(i; n, \frac{\lambda}{n}) = \frac{e^{-\lambda} \lambda^i}{i!}.$$

当二项分布满足 $n > 50, p < 0.1, np < 5$ 时, 用泊松分布近似效果较好.

3 超几何分布

$$X \sim H(N, n, M).$$

理解: N 件产品中有 M 件次品, 从总体中抽 n 件时次品的数量 m .

$$\text{概率分布: } P(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}.$$

4 负二项分布

$$X \sim NB(r, p).$$

理解: 合格率为 p , 抽取到 r 个合格产品时, 抽到的不合格产品的个数 x .

$$\text{概率分布: } P(X = i) = d(i; r, p) = \binom{i+r-1}{r-1} p^r (1-p)^i.$$

5 几何分布

$$X \sim GE(p).$$

理解: 合格率为 p , 抽取到第一个合格产品时, 抽到的不合格产品的个数 x .

$$\text{概率分布: } P(X = i) = p(1-p)^i.$$

几何分布具有无记忆性.

5' 几何分布

$$X \sim G(p).$$

理解: 合格率为 p , 抽取到第一个合格产品时, 抽到的总产品的个数 x .

$$\text{概率分布: } P(X = i) = p(1-p)^{i-1}.$$

几何分布具有无记忆性.

2.1.2 连续型随机变量

1 正态分布

$$X \sim N(\mu, \sigma^2).$$

$$\text{概率密度函数: } f(x) = (\sqrt{2\pi}\sigma)^{-1} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$$\text{标准正态分布: } Y = (X - \mu)/\sigma \sim N(0, 1).$$

3σ 原则: 0.6826, 0.9544, 0.9974.

$$\text{上 } \alpha \text{ 分位数: } \Phi(z_\alpha) = 1 - \alpha.$$

2 指数分布

$X \sim E(\lambda)$.

概率密度函数: $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

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

指数分布具有无记忆性, 即 $P(X > m + t | X > m) = P(X > t)$.

3 威布尔分布

概率密度函数: $f(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^\alpha}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

分布函数: $F(x) = \begin{cases} 1 - e^{-\lambda x^\alpha}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

4 均匀分布

$X \sim R(a, b)$.

概率密度函数: $f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b, \\ 0, & x < a \text{ 或 } x > b. \end{cases}$

分布函数: $F(x) = \begin{cases} 0, & x \leq a, \\ (x-a)/(b-a), & a < x < b, \\ 1, & x \geq b. \end{cases}$

5 对数正态分布

$\ln X \sim N(\mu, \sigma^2)$.

概率密度函数: $f(x, \mu, \sigma) = \begin{cases} (x\sqrt{2\pi}\sigma) \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], & x > 0, \\ 0, & x \leq 0. \end{cases}$

6 柯西分布

$X \sim C(\gamma, x_0)$.

概率密度函数: $f(x; x_0, \gamma) = \frac{1}{\pi} \cdot \frac{\gamma}{(x - x_0)^2 + \gamma^2} \quad (-\infty < x < +\infty)$.

7 拉普拉斯分布

$X \sim \text{La}(\mu, b)$.

概率密度函数: $f(x) = \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}}$.

2.2 多维随机变量

2.2.1 离散性随机向量

1 多项分布

$X = (X_1, \dots, X_n) \sim M(N; p_1, \dots, p_n)$.

$$P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \frac{N!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}.$$

多项分布的边缘分布是二项分布.

$$(X_1, X_2, \dots, X_n) \sim M(N; p_1, p_2, \dots, p_n) \Rightarrow X_1 + X_2 \sim B(N; p_1 + p_2).$$

2.2.2 连续型随机向量

1 矩形均匀分布

2 二维正态分布

$$X = (X_1, X_2) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho).$$

$$f(x_1, x_2) = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1} \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{(x_1-a)^2}{\sigma_1^2} - \frac{2\rho(x_1-a)(x_2-b)}{\sigma_1\sigma_2} + \frac{(x_2-b)^2}{\sigma_2^2} \right) \right].$$

当且仅当 $\rho = 0$ 时, X_1 和 X_2 独立.

其它性质

- 二维正态分布的**边缘分布**是正态分布.
- 二维正态分布的**条件分布**是正态分布.

若 $(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则给定 $X = x$ 时 Y 的条件分布为

$$N(b + \rho\sigma_2\sigma_1^{-1}(x-a), \sigma_2^2(1-\rho^2)).$$

- 二维正态分布的**边缘分布的和**仍为正态分布
若 $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则 $Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$.
- **独立**的正态分布的**联合分布**是正态分布.
正态分布的联合分布**不一定**是二维正态分布.
- 相互独立的正态分布的和仍为正态分布
若 $X_i \sim N(\mu_i, \sigma_i^2)$, 则 $X_1 + \dots + X_n \sim N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$.
- 若 $Y = X_1 + X_2$ 服从正态分布, X_1, X_2 独立, 则 X_1, X_2 也是正态分布. ★

3 多元正态分布

设 (X_1, X_2, \dots, X_n) 为 n 元随机变量, 令

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix},$$

其中 \mathbf{C} 为**协方差矩阵**. 如果 (X_1, X_2, \dots, X_n) 的概率密度函数为

$$f(x_1, x_2, \dots, x_n) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}}$$

则称 (X_1, X_2, \dots, X_n) 是参数为 $\boldsymbol{\mu}, \mathbf{C}$ 的 n 元正态变量.

其它性质

- n 维正态分布的**边缘分布**是正态分布.
- n 维正态分布的**条件分布**是正态分布.
- n 维正态分布的**边缘分布的和**是正态分布.
- n 维随机变量 (X_1, X_2, \dots, X_n) 服从 n 维正态分布的充要条件是:

$$\forall l_i \in \mathbb{R} (i = 1, 2, \dots, n) : l_1 X_1 + l_2 X_2 + \dots + l_n X_n \sim N(\mu, \sigma^2).$$

- 若 Y_1, Y_2, \dots, Y_m 都是 n 维正态分布分量 $X_i (i = 1, 2, \dots, n)$ 的**线性函数**, 则 (Y_1, Y_2, \dots, Y_m) 服从 m 维正态分布.

- n 维正态分布各分量相互对立充要条件是它们两两不相关.

2.2.3 边缘分布

1 概念解释

- 随机向量的分布可以决定其任一分量的边缘分布, 但反之不亦然.
- 随机向量也叫作其边缘分布的 **联合分布**.
- 类似的有二维的边缘分布.

2 多项分布

$(X_1, \dots, X_n) \sim M(N; p_1, \dots, p_n)$ 关于 X_1 的边缘分布为 $M(N, p_1)$.

3 二维正态分布

$(X_1, X_2) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$ 关于 X_1 和 X_2 的边缘分布分别是 $N(a, \sigma_1^2)$ 和 $N(b, \sigma_2^2)$.

2.3 条件概率分布与随机变量的独立性

2.3.1 条件概率分布的概念

2.3.2 离散性随机变量的条件概率分布

1 多项分布

在给定 $X_2 = k_2$ 的条件下, X_1 的条件分布为 $B(N - k_2, p_1/(1 - p_2))$.

2.3.3 连续性随机变量的条件概率分布

$$f_1(x_1 | a \leq X_2 \leq b) = \int_a^b f(x_1, t_2) dt_2 / \int_a^b f_2(t_2) dt_2$$

$$\begin{aligned} f(x_1, x_2) &= f_2(x_2) f_1(x_1 | x_2) \\ f(x_1, \dots, x_n) &= g(x_1, \dots, x_k) h(x_{k+1}, \dots, x_n | x_1, \dots, x_k) \end{aligned}$$

$$f_1(x_1) = \int_{-\infty}^{+\infty} f_2(x_2) f_1(x_1 | x_2) dx_2$$

正态变量的条件分布仍为正态. 正态分布条件分布的中心位置是

$$m(x_1) = b + \rho \sigma_2 \sigma_1^{-1} (x_1 - a).$$

2.3.4 随机变量的独立性

两个变量的独立 $\Leftrightarrow f_1(x_1) = f_1(x_1 | x_2)$.

定义 3.1 连续型随机变量的相互独立 (独立)

$$X_1, X_2, \dots, X_n \text{ 相互独立 (独立)} \Leftrightarrow f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

定理 3.1

连续变量独立 \Leftrightarrow 对应的事件独立.

定理 3.2

若连续型随机向量 (X_1, X_2, \dots, X_n) 的概率密度函数 $f(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \cdots g_n(x_n)$, 则 X_1, X_2, \dots, X_n 相互独立, 且 $f_i(x_i) = Cg_i(x_i)$.

定理 3.3

若 X_1, X_2, \dots, X_n 相互独立,

$$Y_1 = g_1(X_1, X_2, \dots, X_m), Y_2 = g_2(X_{m+1}, X_{m+2}, \dots, X_n),$$

则 Y_1 和 Y_2 独立.

定义 3.2 离散性随机变量的相互独立

X_1, X_2, \dots, X_n 相互独立 (独立) 等价于

$$\forall a_1, a_2, \dots, a_n: P(X_1 = a_1, \dots, X_n = a_n) = P(X_1 = a_1) \cdots P(X_n = a_n).$$

示性函数

$$X = \begin{cases} 1, & \text{当事件 } A \text{ 发生时,} \\ 0, & \text{当事件 } A \text{ 不发生时.} \end{cases}$$

2.4 随机变量的函数的概率分布

2.4.1 离散性分布

1. 多项分布 $(X_1, X_2, \dots, X_n) \sim M(N; p_1, p_2, \dots, p_n) \Rightarrow X_1 + X_2 \sim B(N; p_1 + p_2)$.

2. 二项分布 $X_1 \sim B(n_1, p), X_2 \sim B(n_2, p) \Rightarrow X_1 + X_2 \sim B(n_1 + n_2, p)$.

3. 泊松分布 $X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2) \Rightarrow X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$.

2.4.2 连续型分布

1 单变量函数

1.1 严格单调

若 X 有密度函数 $f(x)$, $Y = g(X)$ 且该函数**严格单调**, 令 $X = h(Y)$, 则 Y 的概率密度函数为

$$l(y) = f(h(y)) |h'(y)|.$$

• $X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$.

• 若 $X \sim f(x)$, 则 $aX + b \sim \frac{1}{a} f\left(\frac{x-b}{a}\right)$.

1.2 幂函数

若 X 有密度函数 $f(x)$, $Y = X^n$, 其中 n 为偶数, 则 Y 的概率密度函数为

$$l(y) = \left| \frac{y^{\frac{1}{n}-1}}{n} \right| \left[f(y^{\frac{1}{n}}) + f(-y^{\frac{1}{n}}) \right]. \quad (n \text{ 是偶数})$$

• 若 $X \sim N(0, 1)$, 则 $Y = X^2$ 的密度函数为 $l(y) = \begin{cases} \left(\sqrt{2\pi y} \right)^{-1} e^{-y/2}, & y > 0, \\ 0, & y \leq 0. \end{cases}$

2 多变量函数

以两个为例, 多变量是类似的.

$$\begin{cases} Y_1 = g_1(X_1, X_2) \\ Y_2 = g_2(X_1, X_2) \end{cases} \Rightarrow \begin{cases} X_1 = h_1(Y_1, Y_2) \\ X_2 = h_2(Y_1, Y_2) \end{cases}$$

则雅可比行列式为

$$J(y_1, y_2) = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{vmatrix},$$

概率密度函数

$$l(y_1, y_2) = f(h_1(y_1, y_2), h_2(y_1, y_2)) |J(y_1, y_2)|.$$

2.4.3 随机变量和的密度函数

设 (X_1, X_2) 的联合密度函数为 $f(x_1, x_2)$, 则 $Y = X_1 + X_2$ 的密度函数为

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

- 法一: 固定 y 后积分得分布函数, 再对 y 求导得上式.
- 法二: 补充 $Y_2 = X_1$, 利用 2.4.2.2

- 二维正态分布的边缘分布的和仍为正态分布

若 $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则 $Y = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$.

- 相互独立的正态分布的和仍为正态分布

若 $X_i \sim N(\mu_i, \sigma_i^2)$, 则 $X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$.

- 若 $Y = X_1 + X_2$ 服从正态分布, X_1, X_2 独立, 则 X_1, X_2 也是正态分布.

自由度为 n 的皮尔逊卡方密度与 **卡方分布** $X \sim \chi_n^2$

$$k_n(x) = \begin{cases} \frac{e^{-x/2} x^{(n-2)/2}}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

- 若 X_1, X_2, \dots, X_n 相互独立, 且有公共分布 $N(0, 1)$ (**独立同分布 iid**), 则 $Y = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi_n^2$.
- 若 $X_1 \sim \chi_m^2$ 与 $X_2 \sim \chi_n^2$ 独立, 则 $X_1 + X_2 \sim \chi_{m+n}^2$.
- 若 X_1, X_2, \dots, X_n 相互独立, 且都服从指数分布 $E(\lambda)$, 则 $X = 2\lambda(X_1 + X_2 + \cdots + X_n) \sim \chi_{2n}^2$.
- $E(\chi_n^2) = n$.
- $E(\chi_n^2)^{-1} = \frac{1}{n-2}$.
- $E(\chi_n^2)^k = \frac{2^k \Gamma\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right)} (k \in \mathbb{Z})$.
- $\text{Var}(\chi_n^2) = 2n$.

注意到方差是均值的两倍, 可以以此检验是否为卡方分布. 🍌

2.4.4 随机变量商的密度函数

设 (X_1, X_2) 的联合密度函数为 $f(x_1, x_2)$, 则 $Y = X_1/X_2$ 的密度函数为

$$l(y) = \int_0^{+\infty} x f(xy, x) dx.$$

- 法一: 固定 y 后积分得分布函数, 再对 y 求导得上式.
- 法二: 补充 $Y_2 = X_2$, 利用 2.4.2.2

设 X_1, X_2 独立, $X_1 \sim \chi_n^2$, $X_2 \sim N(0, 1)$, $Y = X_2/\sqrt{X_1/n}$, 则 Y 的概率函数为

$$t_n(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}.$$

称为自由度为 n 的 t 分布.

- $E(t_n) = 0$ ($n > 1$).
- $\text{Var}(t_n) = \frac{n}{n-2}$ ($n > 2$).

设 X_1, X_2 独立, $X_1 \sim \chi_n^2$, $X_2 \sim \chi_m^2$, $Y = \frac{X_2}{m} / \frac{X_1}{n}$, 则 Y 的概率密度函数为

$$f_{m,n}(y) = m^{m/2} n^{n/2} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} y^{m/2-1} (my+n)^{-(m+n)/2} \quad (y > 0)$$

称为自由度为 (m, n) 的 F 分布.

- $E(f_{m,n}) = \frac{n}{n-2}$ ($n > 2$).
- $\text{Var}(f_{m,n}) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$.

2.4.5 多个随机变量的排序

1 独立同分布的最值

若 X_1, X_2, \dots, X_n 独立同分布, 且有分布函数 $F(x)$ 和密度函数 $f(x)$, 则

$$\begin{aligned} Y = \max(X_1, X_2, \dots, X_n) &\sim nF^{n-1}(x)f(x), \\ Z = \min(X_1, X_2, \dots, X_n) &\sim n[1 - F(x)]^{n-1}f(x). \end{aligned}$$

2 两项最值联合分布

设独立随机变量 X 和 Y 的分布函数分别为 $F_X(x)$ 和 $F_Y(y)$, 则最小值与最大值的联合分布函数为

$$F(u, v) = \begin{cases} F_X(u)F_Y(v) + F_X(v)F_Y(u) - F_X(u)F_Y(u), & u \leq v, \\ F_X(v)F_Y(v), & u > v. \end{cases}$$

证明

当 $u > v$ 时, $F(u, v) = F(v, v)$. 以下均假设 $u \leq v$.

法一: 分类讨论

$$\begin{aligned} F(u, v) &= P\{\min(X, Y) \leq u, \max(X, Y) \leq v\} \\ &= P\{X \leq Y \leq u\} + P\{X \leq u \leq Y \leq v\} + \\ &\quad P\{Y \leq X \leq u\} + P\{Y \leq u \leq X \leq v\} \\ &= P\{X \leq u, Y \leq u\} + P\{X \leq u \leq Y \leq v\} + P\{Y \leq u \leq X \leq v\} \\ &= F_X(u)F_Y(u) + F_X(u)(F_Y(v) - F_Y(u)) + F_Y(u)(F_X(v) - F_X(u)) \\ &= F_X(u)F_Y(v) + F_X(v)F_Y(u) - F_X(u)F_Y(u). \end{aligned}$$

法二: 条件概率

$$\begin{aligned}
F(u, v) &= P\{\min(X, Y) \leq u, \max(X, Y) \leq v\} \\
&= P\{\min(X, Y) \leq u \mid \max(X, Y) \leq v\}P\{\max(X, Y) \leq v\} \\
&= (1 - P\{X, Y \in [u, v] \mid X \leq v, Y \leq v\})P\{X \leq v, Y \leq v\} \\
&= \left(1 - \frac{(F_Y(v) - F_Y(u))(F_X(v) - F_X(u))}{F_X(v)F_Y(v)}\right)F_X(v)F_Y(v) \\
&= F_X(u)F_Y(v) + F_X(v)F_Y(u) - F_X(u)F_Y(u).
\end{aligned}$$

法三：密度函数

$$\begin{aligned}
f(u, v) &= f_X(u)f_Y(v) + f_X(v)f_Y(u), \quad u \leq v. \\
F(u, v) &= \int_{-\infty}^u \int_{-\infty}^v f(u, v)I_{\{u \leq v\}} dv du \\
&= \int_{-\infty}^u \int_u^v f(u, v) dv du \\
&= \int_{-\infty}^u [f_X(u)F_Y(v) - f_X(u)F_Y(u) + \\
&\quad F_X(v)f_Y(u) - F_X(u)f_Y(u)] du \\
&= F_X(u)F_Y(v) + F_X(v)F_Y(u) - F_X(u)F_Y(u).
\end{aligned}$$

备注

本题十分经典，证明过程比结果更加重要；建议这些方法自己都手算一遍，确保真正理解。

第一种方法最为直观，但是难以处理更加复杂的问题；第二种方法技巧性比较强，但利用条件概率或者全概率的思路最为通用；第三种方法比较典型，但是很多时候我们是不知道密度函数的。

其中第三种方法可以作为已知联合密度、求联合分布的例题。联合分布在计算时，利用图像要比纯粹的代数更容易求解；而且利用示性函数简化表达式，可以让思路更加清晰。

3 次序统计量的分布

将独立同分布 $F(x)$ 的随机变量 X_1, X_2, \dots, X_n 重排为

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

则 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 称为顺序统计量。

性质 1 次序统计量是充分统计量，即在次序统计量下的条件分布与总体分布无关。

性质 2 $X_{(k)}$ 的密度函数为

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x)(1-F(x))^{n-k} f(x).$$

性质 3 $(X_{(i)}, X_{(j)}), (i < j)$ 的联合分布密度函数为

$$f_{ij}(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(x)(F(y)-F(x))^{j-i-1}(1-F(y))^{n-j} f(x)f(y).$$

注意事项

- 概率密度函数在某点的取值必为 0，如果非零，则不存在这样的概率密度函数。即混合型随机变量没有概率密度函数。
- 计算随机变量的函数的概率分布时，注意[单调性](#)、[值域](#)和[值域是否重叠](#)。

第 3 章 随机变量的数字特征

3.1 数学期望与中位数

3.1.1 数学期望的定义

3.1.2 数学期望的性质

基本性质

- 随机变量之和的期望

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

- 独立的随机变量之积的期望

$$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n).$$

- 随机变量函数的期望

$$E(g(X)) = \sum_i g(a_i)p_i \text{ 或 } \int_{-\infty}^{+\infty} g(x)f(x) dx \quad (\text{若求和或极限存在.})$$

- 期望的导数 (以下期望是对于 X 而言的).

$$\frac{d}{dt} E(g(X, t)) = E\left(\frac{d}{dt} g(X, t)\right).$$

用分布函数计算期望

🍌 设随机变量 X 只取非负值, 其分布函数为 $F(x)$, 则在以下两种情况下都有

$$E(X) = \int_0^{+\infty} [1 - F(x)] dx.$$

1. X 为连续型, 有概率密度函数 $f(x)$.
2. X 为离散型, 有分布 $P(X = k) = p_k (k = 0, 1, 2, \cdots)$.

证明

$$\begin{aligned} \int_0^{+\infty} [1 - F(x)] dx &= \int_0^{+\infty} \left[\int_0^{+\infty} f(y) dy - \int_0^x f(y) dy \right] dx \\ &= \int_0^{+\infty} \int_x^{+\infty} f(y) dy dx = \int_0^{+\infty} y f(y) dy = E(X) \\ \int_0^{+\infty} [1 - F(x)] dx &= \sum_{i=0}^{\infty} \int_i^{i+1} [1 - F(x)] dx = \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} p_j \\ &= p_1 + 2p_2 + 3p_3 + \cdots = E(X) \end{aligned}$$

注 上式对任何非负随机变量都成立, 但证明超出初等方法.

期望的不等式

若 X, Y 独立同分布且只取正值, 则

1. $E(X)E(X^{-1}) \geq 1$.
2. $E(X/Y) \geq 1, E(Y/X) \geq 1$.

证明 (1) 式由施瓦茨不等式 $E(X^2)E(Y^2) \geq [E(XY)]^2$ 即得, (2) 式由 (1) 式即得.

期望的等式

若 X_1, X_2, \dots, X_n 独立同分布且只取正值, 则

$$E\left(\frac{X_1}{X_1 + X_2 + \dots + X_n}\right) = \frac{1}{n}.$$

3.1.3 条件数学期望 (条件均值)

条件期望 $E(Y | x)$ 称为 Y 对 X 的回归函数.

$$E(Y | x) = \int_{-\infty}^{+\infty} y f(y | x) dy.$$

期望等于条件期望的期望

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} E(Y | x) f_1(x) dx \\ &= E[E(Y | X)] \end{aligned}$$

3.1.4 中位数

- 中位数总是存在, 均值则不然.
- 中位数可以不唯一 (有无穷多个).
- $E|X - a| \geq E|X - m|$, 其中 m 为中位数.

3.2 方差与矩

3.2.1 方差和标准差

符号说明 $\text{Var}(X) \equiv D(X) \equiv E[(X - \bar{X})^2]$.

基本性质

1. $\text{Var}(X) = E(X - EX)^2 = E(X^2) - (EX)^2$.
2. $\text{Var}(aX + b) = a^2 \text{Var}(X)$.
3. 独立随机变量: $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$.

注意 比较 $\text{Var}(nX_1) = n^2 \text{Var}(X_1)$, 而独立同分布时 $\text{Var}(X_1 + \dots + X_n) = n \text{Var}(X_1)$.

其它性质

1. $E[(X - c)^2] = \text{Var}(X) + (EX - c)^2$.
 - 由此可知 $\text{Var}(X) \leq E[(X - c)^2]$, 当且仅当 $c = EX$ 时取等.
 - 注: 令 $c = 0$ 即得基本性质 (1) 式.
2. 若 $X \in [a, b]$, 则 $\text{Var}(X) \leq \frac{(b - a)^2}{4}$.

证明: 令 $Y = \frac{X - a}{b - a} \in [0, 1]$, 则 $\text{Var}(Y) = E(Y^2) - E(Y)^2 \leq E(Y) - E(Y)^2 \leq \frac{1}{4}$.

3.2.2 原点矩与中心矩

矩又被称为动差 (Moment).

定义

1. X 关于 c 点的 k 阶矩: $E[(X - c)^k]$.
2. k 阶原点矩: $\alpha_k = E(X^k)$.

3. k 阶中心距: $\mu_k = E[(X - EX)^k]$.

特例

1. $E(X) = \mu = \alpha_1$.

2. $\text{Var}(X) = \sigma^2 = \mu_2$.

关系

$$\begin{aligned}\mu_0 &= 1 = \alpha_0 \\ \mu_1 &= 0 \neq \alpha_1 = \mu, \\ \mu_2 &= \sigma^2 = \alpha_2 - \mu^2, \\ \mu_3 &= \alpha_3 - 3\mu\alpha_2 + 2\mu^3, \\ \mu_4 &= \alpha_4 - 4\mu\alpha_3 + 6\mu^2\alpha_2 - 3\mu^4, \\ \mu_k &= \sum_{i=0}^k (-1)^i \binom{k}{i} \alpha_1^i \alpha_{k-i}.\end{aligned}$$

性质

1. $\mu_k(aX + b) = a^k \mu_k(X)$.

2. 独立随机变量: $\mu_k(X_1 + \cdots + X_n) = \mu_k(X_1) + \cdots + \mu_k(X_n)$.

3. $\alpha_k(aX + b) = \sum_{i=0}^k \binom{k}{i} a^i b^{k-i} \alpha_i(X)$.

3.2.3 基于矩的系数

偏度系数 $\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$.

峰度系数 $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

- 正态分布 $N(\mu, \sigma^2)$ 的峰度系数为 3, 故有时定义峰度系数为 $\mu_4 / \mu_2^2 - 3$.

变异系数 $c_v \equiv V_\sigma \equiv \frac{\sigma}{\mu}$.

- 备注
 - 也称为**标准差系数**, **标准离差率** 或 **单位风险**.
 - 有时使用记号 $C.V = \frac{SD}{MN} \times 100\%$.
- 优缺点
 - 无量纲参数, 用于比较两组数据的离散程度.
 - 当期望接近 0 时, 变异系数的精确度会下降.
 - 编译系数无法发展处类似均值的置信区间的工具.

3.3 协方差与相关系数

3.3.1 协方差

基本性质

1. $\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$.

2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

3. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

4. $\text{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2 \text{Cov}(X, Y)$.

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

6. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$. ★

7. $\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)$.

8. 若 X, Y 独立, 则 $\text{Cov}(X, Y) = 0$.

9. $[\text{Cov}(X, Y)]^2 \leq \sigma_1^2 \sigma_2^2$, 且当且仅当 $Y = a + bX$ 时取等.

施瓦茨不等式 $E(X^2)E(Y^2) \geq [E(XY)]^2$, 当且仅当具有线性关系即 $aX + bY = 0$ 时取等.

• 由 $E[(Y + tX)^2] \geq 0$ 的判别式小于零即得.

协方差矩阵 设 n 维随机变量 (X_1, X_2, \dots, X_n) 中 X_i 与 X_j 的协方差都存在, 且记作

$c_{ij} = \text{Cov}(X_i, X_j)$, $i, j = 1, 2, \dots, n$, 则协方差矩阵为

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}.$$

其中 $c_{ii} = \text{Var}(X_i)$. [多元正态分布中的应用](#)

3.3.2 相关系数

符号说明 $\text{Corr}(X, Y) = R(X, Y) = \rho_{XY}$.

基本性质

1. $\text{Corr}(X, Y) \equiv \text{Cov}(X, Y) / (\sigma_1 \sigma_2)$.

◦ 相关系数也可以记作

$$\rho(X, Y) = E \left[\left(\frac{X - E(X)}{\sqrt{\text{Var}(X)}} \right) \left(\frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}} \right) \right].$$

2. $\text{Corr}(X, Y) = \text{Corr}(Y, X)$.

◦ 相关系数矩阵是对称阵.

3. $\text{Corr}(a_1 X + b_1, a_2 Y + b_2) = \text{Corr}(X, Y)$.

◦ 相关系数不受单位影响.

4. 若 X, Y 独立, 则 $\text{Corr}(X, Y) = 0$.

◦ 若 $\text{Corr}(X, Y) = 0$ 则称 X 与 Y **不相关**.

5. $|\text{Corr}(X, Y)| \leq 1$, 且当且仅当 X 和 Y 有严格线性关系时取等.

◦ 相关系数又称为线性相关系数.

6. $\text{Corr}(X, Y) = \pm 1 \Leftrightarrow P \left(\frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}} = \pm \frac{X - E(X)}{\sqrt{\text{Var}(X)}} \right) = 1$.

7. 最小二乘及其均方误差.

$$\begin{aligned} E[(Y - a - bX)^2] &\equiv E[(Y - m_2) - b(X - m_1) - c]^2 \\ &= \sigma_2^2 + b^2 \sigma_1^2 - 2b \text{Cov}(X, Y) + c^2 \\ &\geq \sigma_2^2 + b^2 \sigma_1^2 - 2b \text{Cov}(X, Y) \end{aligned}$$

$$b = \text{Cov}(X, Y) / \sigma_1^2 = \sigma_1^{-1} \sigma_2 \text{Corr}(X, Y) \equiv \sigma_1^{-1} \sigma_2 \rho$$

$$L(X) = m_2 - \sigma_1^{-1} \sigma_2 \rho m_1 + \sigma_1^{-1} \sigma_2 \rho X$$

$$\begin{aligned} E[(Y - L(X))^2] &= \sigma_2^2 + b^2 \sigma_1^2 - 2b \text{Cov}(X, Y) \quad (\text{由最上式}) \\ &= \sigma_2^2 (1 - \rho^2) \end{aligned}$$

二维正态分布

若 $(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$, 则

1. 即使允许用任何函数 $M(X)$ 逼近 Y , 则所得到的最佳逼近仍是 $L(X)$, 故只需考虑线性逼近已足够.
2. 对于二维正态分布, $\text{Corr}(X, Y) = \rho$, 即 $\text{Corr}(X, Y) = 0$ 可推出二者独立.

3.4 大数定理和中心极限定理

3.4.1 大数定理

依概率收敛 $Y_n \xrightarrow{P} a \Leftrightarrow \lim_{n \rightarrow \infty} P\{|Y_n - a| < \varepsilon\} = 1$.

- $X_n \xrightarrow{P} a \Rightarrow g(X_n) \xrightarrow{P} g(a)$. (其中 X_n 可为向量)

马尔科夫不等式 若 Y 为只取非负值的随机变量, 则对 $\forall \varepsilon > 0$, 有

$$P(Y \geq \varepsilon) \leq E(Y)/\varepsilon.$$

切比雪夫不等式 若 $\text{Var}(Y)$ 存在, 则

$$P(|Y - EY| \geq \varepsilon) \leq \text{Var}(Y)/\varepsilon^2.$$

- 切比雪夫不等式在参数估计的相合性判断中的[应用](#).

弱大数定理 设 X_1, X_2, \dots, X_n 是独立同分布的随机变量, 记它们的公共均值为 μ , 方差存在并记为 σ^2 , 则对 $\forall \varepsilon > 0$, 有

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

强大数定理 同上条件,

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1.$$

伯努利大数定理 即大数定理的特例 (频率收敛于概率)

$$\lim_{n \rightarrow \infty} P(|p_n - p| \geq \varepsilon) = 0.$$

- 大数定理中无需假定 X_i 的方差存在也可以证明 (即 **辛钦大数定理**), 不必同分布, 甚至可以不独立.

3.4.2 中心极限定理

应用

林德伯格-莱维定理 设 X_1, X_2, \dots, X_n 为独立同分布的随机变量, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$ ($0 < \sigma^2 < \infty$), 则对任何实数 x , 有

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{\sqrt{n}\sigma}(X_1 + X_2 + \dots + X_n - n\mu) \leq x\right) = \Phi(x) = \int_{-\infty}^x \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt.$$

- 因此, 任何独立同分布的大量随机变量之和近似服从正态分布.

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

- 误区: 变量本身的分布并未改变. 可以将大量随机变量分为若干组大量随机变量, 分别计算其和, 并验证其符合正态分布. 这个性质可用于参数估计等, 但不可用于检验随机变量的值是否被篡改.

棣莫弗-拉普拉斯定理 上式的特例, 当 $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ ($0 < p < 1$) 时, 对任何实数 x , 有

$$\lim_{n \rightarrow \infty} P \left(\frac{1}{\sqrt{np(1-p)}} (X_1 + X_2 + \cdots + X_n - np) \leq x \right) = \Phi(x).$$

- 或者说, 若随机变量 $\eta_n \sim B(n, p)$ ($n = 1, 2, \dots$), 则对任何实数 x , 有

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

估值公式

$$P(t_1 \leq X_1 + X_2 + \cdots + X_n \leq t_2) \approx \Phi(y_2) - \Phi(y_1).$$

其中 $y_i = (t_i - np) / \sqrt{np(1-p)}$, 或修正为

$$\begin{cases} y_1 = \left(t_1 - \frac{1}{2} - np \right) / \sqrt{np(1-p)}, \\ y_2 = \left(t_2 + \frac{1}{2} - np \right) / \sqrt{np(1-p)}. \end{cases}$$

3.5 母函数

3.5.1 母函数的定义

整值随机变量 即只取非负整值的随机变量.

若整值随机变量的概率分布为 $P\{X = k\} = p_k$, $k = 0, 1, 2, \dots$, 则其 **母函数** 为

$$G(s) := E(s^X) = \sum_{k=0}^{+\infty} p_k s^k.$$

- $G(1)$ 收敛且为 1 (而不是书中说的因为某种方式的计算结果为 1 而收敛), 且 $G(-1)$ 绝对收敛, 故 $G(s)$ 至少在 $[-1, 1]$ 上绝对收敛.

3.5.2 常见分布的母函数

- 对于 $X \sim B(n, p)$, 有

$$G(s) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} s^k = (ps + q)^n, \quad s \in (-\infty, +\infty).$$

- 对于 $X \sim P(\lambda)$, 有

$$G(s) = \sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} \cdot s^k = e^{\lambda(s-1)}, \quad s \in (-\infty, +\infty).$$

- 对于 $X \sim G(p)$, 有

$$G(s) = \sum_{k=1}^{+\infty} p q^{k-1} s^k = \frac{ps}{1 - qs}, \quad s \in \left(-\frac{1}{q}, \frac{1}{q} \right).$$

3.5.3 母函数的性质

- $p_k = \frac{G^{(k)}(0)}{k!}$, $k = 0, 1, 2, \dots$.
- $E(X) = \sum_{k=1}^{+\infty} k p_k = G'(1)$.
- $E(X_{(n)}) = G^{(n)}(1)$. (前者下标表示下降阶乘幂, 后者上标表示高阶导数)
- $\text{Var}(X) = G''(1) + G'(1) - [G'(1)]^2$.

若 X 的 k 阶矩存在, 则有

- $G^{(0)}(1) = 1$.
- $G^{(1)}(1) = E(X)$.
- $G^{(2)}(1) = E(X^2) - E(X)$.
- $G^{(3)}(1) = E(X^3) - 3E(X^2) + 2E(X)$.
- $G^{(4)}(1) = E(X^4 - 6X^3 + 11X^2 - 6X)$.

3.5.4 母函数的定理

对于可数集合 \mathcal{A} 及其大小函数 $f: \mathcal{A} \rightarrow \mathbb{N}$, 简记 $f(p) = |p|_{\mathcal{A}} = |p|$, 定义

$$a_n := |\{p \mid p \in \mathcal{A}, |p| = n\}|$$

$$|(a_1, a_2, \dots, a_n)| := \sum_{i=1}^n |a_i|$$

加法定理 如果 \mathcal{A} 与 \mathcal{B} 互斥, 则其无交并的母函数为 $C(x) = A(x) + B(x)$.

或者说, 若满足 (或定义) 了

1. $\mathcal{A} \cap \mathcal{B} = \emptyset$.
2. $\mathcal{A} \cup \mathcal{B} = \mathcal{C}$.
3. $|p|_{\mathcal{C}} = \begin{cases} |p|_{\mathcal{A}}, & p \in \mathcal{A}, \\ |p|_{\mathcal{B}}, & p \in \mathcal{B}, \end{cases}$

则有 $C(x) = A(x) + B(x)$.

乘法定理 如果有 $\mathcal{A} \times \mathcal{B} = \mathcal{C}$, 则有 $C(x) = A(x)B(x)$.

序列构造 若 $\mathcal{B} = \text{SEQ}(\mathcal{A}) := \bigcup_{n \in \mathbb{N}} \mathcal{A}^n$, 则有形式幂级数 $B(x) = \frac{1}{1 - A(x)}$.

设整值随机变量 $X \sim P\{X = k\} = a_k$ 和 $Y \sim P\{Y = k\} = b_k$ 相互独立, 且母函数分别为 $A(s)$, $B(s)$, 则 $Z = X + Y$ 的母函数为

$$G(s) = A(s)B(s) = \sum_{n=0}^{\infty} \sum_{k=0}^n a_k b_{n-k} s^n.$$

若 n 个整值随机变量独立同分布, 则 $G(s) = [G_1(s)]^n$.

由上, 若非负整值随机变量 X_i 独立同分布 $P\{X = k\} = a_k$, 则有

$$P\{X_1 + X_2 + \dots + X_n = N\} = \frac{\partial^N (\sum_{k=0}^{\infty} a_k x^k)^n}{n! \partial x^N} \Big|_{x=0}.$$

3.6 特征函数

3.6.1 特征函数的定义

设 X, Y 为实随机变量, 则称 $Z = X + iY$ 为 **复随机变量**.

设 X 是实随机变量, 则 X 的 **(一维) 特征函数** 为

$$\begin{aligned}
g(t) &= E(e^{itX}) \quad (-\infty < t < +\infty) \\
&= \begin{cases} \int_{-\infty}^{+\infty} e^{itx} f(x) dx, & \text{连续型随机变量,} \\ \sum p_k e^{itx_k} = G(e^{it}), & \text{离散性随机变量.} \end{cases} \\
&= \begin{cases} \int_{-\infty}^{+\infty} f(x) \cos tx dx + i \int_{-\infty}^{+\infty} f(x) \sin tx dx, & \text{连续型随机变量,} \\ \sum p_k \cos tx_k + i \sum p_k \sin tx_k, & \text{离散性随机变量.} \end{cases}
\end{aligned}$$

- 上述级数与广义积分绝对收敛.

3.6.2 常见分布的特征函数

常见分布	特征函数
$X \sim B(n, p)$	$g(t) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{itk} = (pe^{it} + q)^n$
$X \sim P(\lambda)$	$g(t) = \sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} e^{itk} = e^{\lambda(e^{it}-1)}$
$X \sim G(p)$	$g(t) = \sum_{k=1}^{+\infty} p q^{k-1} e^{itk} = \frac{pe^{it}}{1 - qe^{it}}$
$X \sim U(a, b)$	$g(t) = \int_a^b \frac{e^{itx}}{b-a} dx = \begin{cases} \frac{e^{ibt} - e^{iat}}{it(b-a)}, & t \neq 0, \\ 1, & t = 0. \end{cases}$
$X \sim E(\lambda)$	$g(t) = \int_0^{+\infty} e^{itx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - it}$
$X \sim N(\mu, \sigma^2)$	$g(t) = \int_{-\infty}^{+\infty} \frac{e^{itx}}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = e^{i\mu t} = e^{i\mu t - \frac{\sigma^2}{2} t^2}$

3.6.3 特征函数的性质

1. $g(0) = 1$.
2. $|g(t)| \leq 1$.
3. $g(-t) = \overline{g(t)}$.
4. $g(t)$ 在 $(-\infty, +\infty)$ 上一致连续.
5. $\forall n \in \mathbb{N}^+, \forall t_i \in \mathbb{R}, \forall z_i \in \mathbb{C} :$

$$\sum_{j=1}^n \sum_{k=1}^n g(t_j - t_k) z_j \bar{z}_k \geq 0.$$

6. 如果 X 的 n 阶原点矩存在, 则它的特征函数的 n 阶导数存在, 且

$$g^{(k)}(0) = i^k E(X^k), \quad k = 1, 2, \dots, n.$$

7. 若 X 的特征函数为 $g_X(t)$, 且 $Y = aX + b$ ($a, b \in \mathbb{R}$), 则

$$g_Y(t) = e^{ibt} g_X(at).$$

8. 如果 X_1 和 X_2 相互独立, 且特征函数分别为 $g_1(t), g_2(t)$, 则 $Y = X_1 + X_2$ 的特征函数为

$$g_Y(t) = g_1(t) g_2(t).$$

9. 如果两个随机变量有相同的特征函数, 那么它们具有相同的概率分布, 反之亦然.

类似的, 由性质 6 可得到 k 阶原点矩.

3.7 矩量母函数

矩量母函数又称**动差生成函数** (MGF, Moment, Generating Function).

3.7.1 相关定义

对于任意随机变量 ξ , 若下述求数学期望存在, 即求和或积分存在, 则矩量母函数 (mgf) 为

$$m_{\xi}(t) := E(e^{t\xi}) = \begin{cases} \sum e^{tx} P(X = x), & \text{离散性随机变量,} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) dx, & \text{连续性随机变量.} \end{cases}$$

其对数称为 **累积量生成函数**

$$R_{\xi}(t) := \ln m_{\xi}(t).$$

3.7.2 常见分布的矩量母函数

常见分布	矩量母函数
$\xi \sim \Gamma(\alpha, \beta)$	$m_{\xi}(t) = \frac{1}{(1 - \beta t)^{\alpha}}, \quad t < \frac{1}{\beta}.$
$X \sim \text{Ga}(\alpha, \beta)$	$m_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \quad t < \beta.$
$\xi \sim \chi_n^2$	$m_{\xi}(t) = \frac{1}{(1 - 2t)^{\frac{n}{2}}}, \quad t < \frac{1}{2}.$
$\xi \sim E(\lambda)$	$m_{\xi}(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$

3.7.3 矩量母函数的性质

1. 矩量性质

$$\begin{aligned} m_{\xi}(t) &= \int_{-\infty}^{+\infty} \left[1 + tx + \frac{(tx)^2}{2!} + \cdots + \frac{(tx)^n}{n!} + \cdots \right] f(x) dx \\ &= 1 + \mu_1 t + \frac{\mu_2}{2!} t^2 + \cdots + \frac{\mu_n}{n!} t^n + \cdots \end{aligned}$$

2. 求随机变量的原点矩

$$E(X^k) = \mu_k = \frac{d^k}{dt^k} m_X(t) \Big|_{t=0} = m_X^{(k)}(0).$$

$$1. E(X) = m'_X(0).$$

$$2. \text{Var}(X) = m''_X(0) - m'_X(0)^2.$$

$$3. \text{双边 Laplace 变换 } m_{\xi}(-t) = \mathcal{B}[f(x)](t).$$

4. 无论概率密度函数是否连续, 都有

$$m_{\xi}(t) = \int_{-\infty}^{+\infty} e^{tx} dF(x).$$

5. 如果两个随机变量具有相同的矩量母函数, 则它们具有相同的概率分布; 反之亦然.

6. 如果 X_1 和 X_2 相互独立, 且矩量母函数分别为 $m_{X_1}(t)$ 和 $m_{X_2}(t)$, 则 $Y = X_1 + X_2$ 的矩量母函数为

$$m_Y(t) = m_{X_1}(t) \cdot m_{X_2}(t).$$

例题

1. 分鞋问题

将 X 分为若干个同分布但不相互独立 X_i 相加, 则 $E(X) = \sum_{i=1}^n E(X_i)$.

由于不独立, 不可单独计算方差后相加, 但可以利用 $E(X^2) = \sum_{i,j=1}^n E(X_i X_j)$.

2. 装盒问题

将 r 个球随机地放入 n 个盒子中, 以 X 记空盒个数,

设恰有 k 个空盒的概率为 $p_k(r, n)$, 均值记为 $m(r, n)$, 则

$$\begin{aligned} p_k(r+1, n) &= p_k(r, n) \frac{n-k}{n} + p_{k+1}(r, n) \frac{k+1}{n} \\ m(r+1, n) &= \sum_{k=0}^n k p_k(r, n) + \frac{1}{n} \sum_{k=0}^n [(k+1)^2 p_{k+1} - k^2 p_k] - \sum_{k=0}^n \frac{k+1}{n} p_{k+1} \\ &= \left(1 - \frac{1}{n}\right) m(r, n) = n \left(1 - \frac{1}{n}\right)^r. \end{aligned}$$