

概率论与数理统计

- | | | |
|--------|------------------|--------|
| 1. 第一讲 | 随机事件与概率 | P1-3 |
| 2. 第二讲 | 一维随机变量及其分布 | P4-10 |
| 3. 第三讲 | 多维随机变量及其分布 | P11-20 |
| 4. 第四讲 | 数学期望、方差、协方差与相关系数 | P21-34 |
| 5. 第五讲 | 大数定律与中心极限定理 | P35-41 |
| 6. 第六讲 | 数理统计的基本概念 | P42-50 |
| 7. 第七讲 | 参数估计 | P51-61 |
| 8. 第八讲 | 假设检验 | P62-70 |

复习

做题顺序: 书 + 笔记 + 辅导讲义 + 30讲 + 300题

笔记(原)

笔记(题)

第一讲	✓	✓	✓	✓	✓
第二讲	✓	✓	✓	✓	✓
第三讲	✓	✓	✓	✓	✓
第四讲	✓	✓	✓	✓	✓
第五讲	✓	✓	✓		
第六讲	✓	✓	✓		
第七讲	✓	✓	✓		
第八讲	✓	✓	✓		

 电气1804
 2018/10/8
 陈惠源

① 德摩根定律

$$\overline{A+B} = \bar{A}\bar{B}$$

或的非=非的与

$$\overline{A\bar{B}} = \bar{A}B$$

$$\overline{AB} = \bar{A} + \bar{B}$$

与的非=非的或

$$\overline{A\bar{B}} = \bar{A} + B$$

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

古典概型

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

几何概型

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

$$\text{条件概率 } P(A|B) = \frac{P(AB)}{P(B)}$$

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

$$\text{独立公式 } A+A_1=\Omega \quad A \cap A_1=\emptyset$$

$$P(AB) = P(A|B) \cdot P(B) = P(A) \cdot P(B|A)$$

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

$$P(ABC) = P(A) \cdot P(B|A) \cdot P(C|AB)$$

$$\text{独立性 } P(AB) = P(A) \cdot P(B)$$

②例: 独立性: ① 若两事件大于0的事件相独立, 则全木改。

$$P(AB) = P(A) \cdot P(B) = P(A) \cdot P(B|A)$$

$$\therefore \frac{P(AB)}{P(A)} = \frac{P(B)}{1} \quad \therefore P(AB) > 0$$

$$\text{若 } P(AB) = P(A)P(B)$$

$$P(B) = P(A) \cdot P(B|A)$$

$$P(B|A) = P(B) \cdot P(A)$$

则AB - (两两独立)

$$\text{若 } P(AB|C) = P(A|C)P(B|C)$$

则A, B, C 相互独立

② (A, B) (A, C) (A, D) - 组独立, 其他不独立

$$\text{若 } A, B, C \text{ 相互独立, 则 } P(ABC) = P(A)P(B)P(C) = P(AB)P(C)$$

(若两两独立, 则不一定相互独立)

若两两独立, 未必相互独立

$$\therefore (A, B, C) (AB, C) (\bar{A}B, C) (\bar{A}\bar{B}, C) \dots \text{也都相互独立.}$$

③例 对任意两事件A, B, 则若 $AB = \emptyset$, 则AB一定不相互独立

(X) 可能 $P(A)=0$ 或 $P(B)=0$

$$\text{④ 若 } A, A_1, \dots, A_n \text{ 两两独立 } P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

$$\text{⑤ 若 } A, B, C \text{ 相互独立, 则要求 } \begin{matrix} P(AB) = P(A) \cdot P(B) \\ P(AC) = P(A) \cdot P(C) \\ P(BC) = P(B) \cdot P(C) \\ P(ABC) = P(A) \cdot P(B) \cdot P(C) \end{matrix} \quad \text{4个等式成立.}$$

(n个事件相互独立, 需 $2^n - n - 1$ 个等式成立)

$$\text{③例 } P(AB) = P(\bar{A}\bar{B}) \text{ 且 } P(A) = P \quad P(B) = \underline{\hspace{2cm}}$$

$$P(\bar{A}\bar{B}) = P(\bar{A} + \bar{B}) = 1 - P(AB)$$

$$= 1 - P(A)P(B) + P(AB) = P(AB)$$

$$\therefore 1 - P$$

$$\text{若 } AB = \bar{A}\bar{B}, \text{ 则 } AB = \underline{\hspace{2cm}} \quad A+B = \underline{\hspace{2cm}}$$

$$A\bar{B} = \bar{A}B = \emptyset = AB$$

$$\therefore P(AB) = 0 \quad AB = \emptyset \quad A+B = \Omega$$

$$\text{若 } P(A) = P(B) = \frac{1}{2} \quad P(A+B) = 1 \quad \text{则 } \underline{\hspace{2cm}}$$

(C) 相互独立事件

$$A, A+B = \Omega \quad A, AB = \emptyset$$

$$(P(\bar{A} + \bar{B}) = 1)$$

(X) 不能判定事件
Y: $\{0, 1\}$ 的分布
 $P(X=0) = 0$ 但 X 与 Y 事件相互独立

例: 设工厂A和工厂B的产品次品率分别为1%和2%, 现从由A厂占60%, B厂占40%的一批产品中随机抽取一件, 发现是次品, 则该次品是A厂生产的概率

$C = \{\text{取的产品是A厂生产的}\}$

则 $\bar{C} = \{\text{取的产品是B厂生产的}\}$

$D = \{\text{取的是次品}\}$

$$P(D|C) = 0.01$$

$$P(D|\bar{C}) = 0.02$$

$$P(C) = \frac{3}{5} \quad P(\bar{C}) = \frac{2}{5}$$

$$P(C|D) = \frac{P(D)}{P(D)} = \frac{P(C) \cdot P(D|C)}{P(C) \cdot P(D|C) + P(\bar{C}) \cdot P(D|\bar{C})} = \frac{0.01 \times \frac{3}{5}}{0.01 \times \frac{3}{5} + 0.02 \times \frac{2}{5}} = \frac{3}{7}$$

④ 互不相容事件又相互独立事件 $EP AB = \phi$

(小概: 若A, B互不相容, 则 $AB = \phi$ 且 $AC\bar{B}$ $A\bar{B} = A$)

例: 设AB为随机事件, A, B互不相容, $P(AB) = \frac{1}{2}$ $P(A) = \frac{1}{3}$ 则 $P(AB|\bar{C}) =$

$$\frac{P(ABC)}{P(C)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

⑤ 大眼都能得出事件

若A, B满足 $P(B|A) = 1$, 则

(B)

$$\frac{P(AB)}{P(A)} = 1$$

A. ACB

B. $P(A-B|A) = 0$

A, B (为两两独立且两两不相容事件, 若 $P(A|BC) = \frac{1}{6}$ 且 $P(B|C) = 0$ 则 $P(A) =$

$$\begin{aligned} P(A+B+C) &= P(A) + P(B) + P(C) \\ &= P(A) + P(B) + P(C) \\ &= P(A) + P(B) + P(C) \end{aligned}$$

⑥ 独立事件 A, B相互独立 且 $P(A) = 0.5$, $P(AB) = 0.3$ 则 $P(B-A) =$

$$\text{解法: 设 } P(AB) = x \quad \therefore P(B-A) = 0.2$$



解法2: $P(AB) = 0.3 = P(A)P(B)$ 相互独立

$$\therefore P(A) = 0.6 \quad P(AB) = 0.3$$

$$P(A) = 0.3 + x$$

$$P(B) = 0.5 \quad x = 0.3$$

$$P(AB) = x = 0.5(0.3 + x)$$

⑦ 全概率统计 两个盒子, 第一个盒子2红1白, 第二个盒子1红2白, 现从两个盒子

中任取一球放盒一起再取一球, 则该球为红的概率? 若该球为白

从第一个盒子中取出的为红白球

① A_0 取0个红 A_1 取1个红 A_2 取2个红 为取红 $P(A_0) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$ $P(A_1) = \frac{2}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{2}$ $P(A_2) = \frac{1}{3}$

$$P(B|A_0) = 0$$

$$P(B|A_1) = \frac{1}{2}$$

$$\therefore P(B) = P(B|A_0)P(A_0) + P(B|A_1)P(A_1) + P(B|A_2)P(A_2) = 0 + \frac{1}{2} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{3} = \frac{7}{12}$$

* 若袋中有2个白球, B 个黑球, 分有放回与不放回两种情况, 自袋中随机抽取

(I) 从袋中取出的第 k 个球为白球 ($k \leq 2+B$)

(II) 从袋中取出 $a+b$ 个球中, 恰含 a 个白, b 个黑 ($a \leq 2, b \leq B$)

则有放回 $P_A = \frac{2}{2+B}$

不放回 $P(A) = \frac{C_{2+B-1}^{2-1}}{C_{2+B}^{2-1}} = \frac{2}{2+B}$

$$P(B) = \left(\frac{a}{a+b} \right)^a \left(\frac{b}{a+b} \right)^b$$

$$P(B) = \frac{C_2^a C_B^b}{C_{2+B}^{a+b}}$$

① 对样本空间中的每一个点都有实数 $X(\omega)$ 对应, 则称 $X(\omega)$ 为一个随机变量, 并简记为 X .

分布函数 $F(x) = P(X \leq x)$ 为随机变量 X 的分布函数

分布函数的性质

② 离散型随机变量

1) 分布列/分布律

写法1: $P(X=x_k) = p_k, k=1, 2, \dots$

常见类型 1) = 二项分布 binomial

写法2:	X	x_1	x_2	x_3
	P	p_1	p_2	p_3

随机变量

$X \sim B(n, p) P(X=k) = \binom{n}{k} p^k q^{n-k}$
 $k=0, 1, 2, \dots$

写法3: $X \sim \begin{bmatrix} x_1 & x_2 & x_3 \\ p_1 & p_2 & p_3 \end{bmatrix}$

2) 泊松分布 Poisson

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

(2) 分布函数

$F(x) = P(X \leq x) = \begin{cases} 0 & x < x_1 \\ p_1 & x_1 \leq x < x_2 \\ p_1 + p_2 & x_2 \leq x < x_3 \end{cases}$

3) 几何分布 Geometric

$X \sim G(p) P(X=k) = p q^{k-1}$
 $k=1, 2, \dots$

③ 连续型随机变量

1) 分布函数 $F(x) = \int_{-\infty}^x f(t) dt$

($\ast$ 分布函数单调, $F(x)$ 取值在 $[0, 1]$)

2) 概率密度函数 $f(x) = F'(x)$

性质 $\int_{-\infty}^{+\infty} f(x) dx = 1$
 $f(x) \geq 0$

$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(t) dt$
 $= F(x_2) - F(x_1)$

题型一: 由概率密度函数求分布函数

题型二: 由分布函数求概率密度函数

$f(x) = \begin{cases} \frac{x}{2} & 0 \leq x < 1 \\ 1-x & 1 \leq x < 2 \\ 0 & \text{其他} \end{cases}$

$F(x) = \begin{cases} A+B e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

(1) 求 X 的分布函数 $F(x)$

(1) 求常数 A, B 的值

(2) 求概率 $P(1 < X < 1.5)$

(2) 密度函数 $f(x)$ 的表达式

$F(x) = \begin{cases} 0 & x \leq 0 \\ \int_{-\infty}^x f(t) dt = \int_0^x t dt = \frac{1}{2} x^2 & 0 < x < 1 \\ \int_{-\infty}^x f(t) dt = \int_0^1 t dt + \int_1^x (1-t) dt \\ = \frac{1}{2} + 2x - \frac{1}{2} x^2 - \frac{1}{2} & 1 \leq x < 2 \\ = \frac{1}{2} x - \frac{1}{2} x^2 + 1 & 1 \leq x < 2 \\ \int_{-\infty}^x f(t) dt = \frac{1}{2} + \frac{1}{2} = 1 & x \geq 2 \end{cases}$

$\lim_{x \rightarrow 0} F(x) = 0 \therefore A = 1$
 $\lim_{x \rightarrow +\infty} F(x) = 1 \quad B = -1$
 $f(x) = F'(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

$P(1 < X < 1.5) = F(1.5) - F(1) = \frac{7}{8}$

3) 常见离散型随机变量分布

① 均匀分布 uniform

② 指数分布 Exponential

③ 正态分布 normal

$X \sim U(a, b) f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{其他} \end{cases}$

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

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

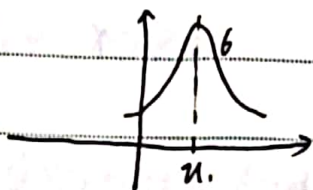
$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$

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

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

① 随机变量函数中心, 6 题小在战

如图像



若 $X \sim N(u, b^2)$

$$Y = ax + b$$

$$Y \sim N(au + b, (ab)^2)$$

标准正态分布的密度函数用 $\phi(x)$ 表示

分布函数用 $F(x)$ 表示

$$\phi(0) = \frac{1}{\sqrt{2\pi}}$$

$$\phi(x) = -\phi(-x)$$

$$P(a < X < b) = \phi(b) - \phi(a)$$

非标准正态分布 $X \sim N(u, b^2)$

$$P(X < y) = F(y) = \Phi\left(\frac{y-u}{b}\right)$$

$$X \sim N(1, 4) \quad P(X < 2.5) = F(2.5) = \Phi\left(\frac{2.5-1}{2}\right) = \Phi(0.75)$$

② 随机变量函数的分布

离散型

X	-1	0	1	2
$Y = 2X+1$	-1	1	3	5
$Y = X^2$	1	0	1	4
P	0.1	0.2	0.3	0.4

$$\therefore Y \sim \begin{cases} -1 & 0.1 \\ 1 & 0.2 \\ 3 & 0.3 \\ 5 & 0.4 \end{cases}$$

$$Y \sim \begin{cases} 0 & 0.2 \\ 1 & 0.4 \\ 4 & 0.4 \end{cases}$$

连续型

① 反函数法: 当已知 X 范围, X 的密度函数, 求 Y 的密度函数

反法: (1) 由 X 范围求 Y 的取值范围
(2) 在 D_Y 内求 Y 的分布函数 $F_Y(y)$
(3) 对 $F_Y(y)$ 求导, 得出 Y 的密度函数.

例: $f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$ 求 $Y = X^2 + 1$ 的密度函数

例: X 服从 $E(1, 2)$ 的分布, 求 $Y = X^2$ 的密度函数

$$x \in (0, 1) \quad Y \in [1, 2]$$

$$x \in (1, 2) \quad Y \in [1, 4]$$

$$\text{当 } Y < 1 \text{ 时 } F_Y(y) = 0$$

$$f(x) = \begin{cases} \frac{1}{3} & 1 < x < 2 \\ 0 & \text{其他} \end{cases}$$

$$\text{当 } 1 \leq Y \leq 2 \text{ 时 } F_Y(y) = P(Y \leq y) = P(X^2 + 1 \leq y)$$

$$\text{当 } Y \leq 0 \quad F_Y(y) = 0$$

$$= P(-\sqrt{y-1} \leq X \leq \sqrt{y-1})$$

$$\text{当 } Y \in [0, 1] \quad F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$$

$$= \int_{-\sqrt{y-1}}^{\sqrt{y-1}} 2x dx = y-1$$

$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{3} dx = \frac{2}{3}\sqrt{y}$$

$$\text{当 } Y > 2 \text{ 时 } F_Y(y) = 1$$

$$\text{当 } Y \in [1, 4] \quad F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$$

$$\therefore f_Y(y) = \begin{cases} 1 & 1 < y < 2 \\ 0 & \text{其他} \end{cases}$$

$$= P(-\sqrt{y-1} \leq X \leq \sqrt{y-1}) = \int_{-\sqrt{y-1}}^{\sqrt{y-1}} \frac{1}{3} dx$$

② 公式法 (变量代换法) ~~交叉~~

$$f_Y(y) = f_X(h(y)) |h'(y)|$$

$$\therefore f_Y(y) = \begin{cases} \frac{1}{3\sqrt{y}} & 1 < y < 4 \\ 0 & \text{其他} \end{cases}$$

例: $f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$ 求 $Y = 2X - X^2 + 1$ 的密度函数

$$f_Y(y) = f_X\left(\frac{y}{2}\right) \cdot \left|\frac{1}{2}\right| = \frac{3}{8} y^2 \quad Y \in (0, 2)$$

$$f_Y(y) = f_X(y) \cdot \left|\frac{1}{2}\right| = \frac{3}{2} y^2 \quad Y \in (0, 1]$$

例 * 二项分布 $X \sim B(n, p)$ $P(X=k) = C_n^k p^k q^{n-k}$ $k=0, 1, 2, \dots, n$

泊松分布 $X \sim P(\lambda)$ $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $k=0, 1, 2, \dots, n$

① 设泊松变量 X 的分布律为 $P(X=k) = C \frac{\lambda^k}{k!}$ $k=0, 1, 2, \dots$ $\lambda > 0$ 则 $C =$ _____

$$k \neq 0 \therefore \text{不是泊松分布}$$

$$1 = \sum_{k=0}^{\infty} C \frac{\lambda^k}{k!} = C \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = C \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - \frac{\lambda^0}{0!} \right)$$

$$= C(e^{\lambda} - 1)$$

$$\therefore C = \frac{1}{e^{\lambda} - 1}$$

② 设随机变量 X 的分布函数为 $F(x)$, 引入函数 $F_1(x) = F(x)$, $F_2(x) = F'(x)$, $F_3(x) = 1 - F(x)$, $F_4(x) = F(F(x))$ 则其中分布函数的是 _____

A. $F_1(x)$ $F_3(x)$

B. $F_2(x)$ $F_4(x)$

(B)

判断 $F(x)$ 为分布函数

- 1) $F(0) = 0$ $F(+\infty) = 1$
- 2) 单调不减
- 3) 右连续

当 $q=1$ $F(x) = F(x)$ $x \rightarrow -\infty$ $F(x) = 1$ 不满足

$F_3(x) = 1 - F(x)$ 是右连续 $\therefore (X)$

③ 设随机变量 X 的概率密度为 $f(x)$, 则下列也是概率密度的 _____

A. $f(2x)$

B. $f'(x)$

C. $2xf(x')$

D. $3x^2 f(x')$

判断 $f(x)$ 为概率密度

$$\begin{cases} \int_{-\infty}^{+\infty} f(x) dx = 1 \\ f(x) \geq 0 \end{cases}$$

$$\therefore \int_{-\infty}^{+\infty} f(x) dx = 1$$

$$\int_{-\infty}^{+\infty} f(x) dx = \frac{1}{2} \int_{-\infty}^{+\infty} f(2x) d(2x) = \frac{1}{2} (\neq 1)$$

(D)

对 B. 假设 $X \sim U(0, 2)$ $f(x) = \begin{cases} \frac{1}{2} & 0 < x < 2 \\ 0 & \text{其他} \end{cases}$

$f(x) = \begin{cases} \frac{1}{4} & 0 < x < 2 \\ 0 & \text{其他} \end{cases}$

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \frac{1}{4} dx = \int_0^2 \frac{1}{4} dx = \frac{1}{2} (\neq 1) \therefore \text{不满足}$$

$2xf(x')$ 当 $x < 0$, $f(x) < 0$ $\therefore (C)$

$3x^2 f(x')$ ≥ 0

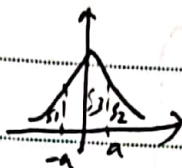
$$\int_{-\infty}^{+\infty} f(x) d(3x^2) = \int_{-\infty}^{+\infty} f(x) dx = 1$$

$\therefore \checkmark$

例: 设随机变量 X 的密度函数 $f(x)$ 满足 $f(x) = f(-x)$, $F(x)$ 是 X 的分布函数, 则对任意实数 a , 下列等式成立的 ()

A. $F(a) = 1 - \int_0^a f(x) dx$

B. $F(a) = \frac{1}{2} - \int_0^a f(x) dx$



$$F(-a) = \int_{-\infty}^{-a} f(x) dx$$

$$= \int_{-\infty}^0 f(x) dx + \int_0^{-a} f(x) dx$$

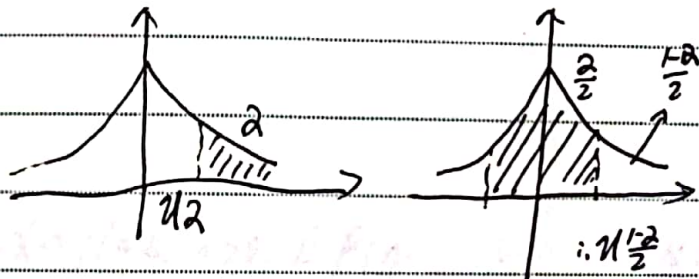
$$= \frac{1}{2} - \int_0^a f(x) dx = \frac{1}{2} - \int_0^a f(x) dx$$

或 $F(a) = S_1$, $F(a) = 1 - S_2 = 1 - F(-a)$

$$S_3 = \int_0^a f(x) dx = \frac{1}{2} - S_2 = \frac{1}{2} - F(-a)$$

例: 设随机变量 X 服从正态分布 $N(0,1)$ 对任意实数 $(0 < a < 1)$, 数 u_2 满足 $P(X > u_2) = a$, 若 $P(X < u_2) = a$, 则 u_2 等于

A. u_2 B. u_2^2



(B)

例: 设 $f(x)$ 为标准正态分布的概率密度, $f_1(x)$ 为 $[1,3]$ 上的均匀分布的概率密度, 若

$$f(x) = \begin{cases} a f_1(x) & x \leq 0 \\ b f_1(x) & x > 0 \end{cases} \quad (a > 0, b > 0) \quad \text{则}$$

(A)

A. $2a + 3b = 4$

B. $3a + 2b = 4$

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 a f_1(x) dx + \int_0^{\infty} b f_1(x) dx = \frac{1}{2} a + \frac{3}{2} b = 1 \Rightarrow 2a + 3b = 4$$

例: 随机变量 X 的概率密度 $f(x) = Ae^{-x^2/4}$ ($-\infty < x < \infty$), 则常数 $A =$

$$A \int_{-\infty}^{\infty} e^{-x^2/4} dx = Ae^{\frac{1}{4}} \int_{-\infty}^{\infty} e^{-t^2} dt = Ae^{\frac{1}{4}} \sqrt{\pi} = 1 \Rightarrow A = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4}}$$

* 例: 设随机变量 X 满足 $X^2 \sim N(0,1)$, 记标准正态分布函数为 $\Phi(x)$, 则 $P(1 < X < 2) =$

$$P(1 < X < 2) = P(1 < X^2 < 4) = P(0 < X^2 < 4) = \Phi(1) - \Phi(-1) = \Phi(1) - 0.5$$

例: 随机变量 X 的概率密度 $f(x) = \frac{1}{k!} e^{-x}$, $k=0,1,2,\dots$ 则常数 $k =$

$$f(x) = \frac{\lambda^k}{k!} e^{-\lambda}$$

例: 设随机变量 X 的概率密度为 $f(x) = \begin{cases} \lambda x^{k-1} & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$ ($k=1,2,\dots$) 则 $P(1 < X < 2) =$

$$\lambda = 1, 2, \dots$$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt = \int_0^x \lambda t^{k-1} dt = \frac{\lambda}{k} x^k$$

$$P(1 < X < 2) = P(1 < X^2 < 4) = \Phi(1) - \Phi(-1) = \Phi(1) - 0.5$$

例: 若 $f(x) = F(x)$ 且 $F(0) = 1$

则

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

$$F(x) = \begin{cases} e^x & x < 0 \\ 1 & x \geq 0 \end{cases}$$

例: 由根根接度分布函数

$$f(x) = \begin{cases} 2(1 - \frac{1}{x^2}) & 1 \leq x \leq 2 \\ 0 & \text{其他} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 1 \\ \int_1^x 2(1 - \frac{1}{t^2}) dt & 1 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

$$= 2x + \frac{2}{x} - 4$$

例: 设某项(印刷)错误的个数 X 服从泊松分布, 已知有 4 条只有两条(印刷)错误的页数相同, 则印刷机抽查页无错误的概率

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \lambda e^{-\lambda} = \frac{\lambda^2}{2} e^{-\lambda}$$

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \lambda = 2$$

$$P(X=0) = e^{-2} = e^{-8}$$

例: 已知 $X \sim U(a, b)$, $a > 0$, 且 $P(0 < X < 3) = \frac{1}{4}$ $P(X > 4) = \frac{1}{2}$, 求 X 的概率密度及 $P(1 < X < 5)$

$$\therefore a > 0$$

$$\therefore P(0 < X < 3) = P(a < X < 3) = \frac{1}{4}$$

$$P(X > 4) = P(4 < X < b) = \frac{1}{2}$$

$$\therefore \frac{3-a}{b-a} = \frac{1}{4} \quad \frac{b-4}{b-a} = \frac{1}{2}$$

$$\therefore \begin{cases} 2a+b=10 \\ \frac{a+b}{2}=4 \end{cases} \quad \therefore a=2 \quad b=6$$

$$\therefore f(x) = \begin{cases} \frac{1}{4} & 2 \leq x \leq 6 \\ 0 & \text{其他} \end{cases}$$

$$\therefore P(1 < X < 5) = P(2 < X < 5) = \frac{3}{4}$$

例: 设随机变量 X 的概率密度为 $f_X(x) = \begin{cases} 1+x & -1 \leq x < 0 \\ 1-x & 0 \leq x \leq 1 \\ 0 & \text{其他} \end{cases}$ 求 $Y = X^2 + 1$ 的分布函数

$$F_Y(y) = P(Y \leq y) = P(X^2 + 1 \leq y) = P(-\sqrt{y-1} \leq X \leq \sqrt{y-1}) = \int_{-\sqrt{y-1}}^{\sqrt{y-1}} f_X(x) dx$$

$$\text{当 } y < 1 \text{ 时 } F_Y(y) = 0$$

$$\text{当 } 1 \leq y < 2 \text{ 时 } F_Y(y) = \int_{-\sqrt{y-1}}^{\sqrt{y-1}} f_X(x) dx = \int_{-\sqrt{y-1}}^0 (1+x) dx + \int_0^{\sqrt{y-1}} (1-x) dx$$

$$= 2\sqrt{y-1} - y + 1$$

$$\frac{1}{2} \leq y < 2$$

$$F_Y(y) = 1$$

特殊构造函数

例: 随机变量 X 的概率密度函数 $f(x) = \begin{cases} \frac{1}{9}x^2 & 0 < x < 3 \\ 0 & \text{其它} \end{cases}$

已知随机变量 $Y = \begin{cases} 2 & X \leq 1 \\ X & 1 < X \leq 2 \\ 1 & X \geq 2 \end{cases}$ 则求 Y 的分布函数, 求联合分布 $P(X \leq Y)$

$$F_X(x) = \begin{cases} \frac{1}{27}x^3 & 0 < x < 3 \\ 0 & \text{其它} \end{cases}$$

$$F_Y(y) = P(Y \leq y)$$

$$y \in [0, 2]$$

$$\text{当 } y < 1 \quad F_Y(y) = 0$$

$$\text{当 } y \geq 2 \quad F_Y(y) = 1$$

$$\text{当 } 1 \leq y < 2 \quad F_Y(y) = P(X \geq 2) + P(1 < X \leq y)$$

$$= 1 - F_X(2) + F_X(y) - F_X(1)$$

$$= 1 - \frac{8}{27} + \frac{y^3}{27} - \frac{1}{27}$$

$$= \frac{y^3 + 18}{27}$$

$$P(X \leq Y) = P(X \leq 2) = \int_0^2 \frac{x^2}{9} dx = \frac{8}{27}$$

3.1(1) = 二维随机变量概念

(2) = 二维随机变量的联合分布函数 $F(x, y) = P(X \leq x, Y \leq y)$

(3) 联合分布函数的性质

$0 \leq F(x, y) \leq 1, F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0, F(+\infty, +\infty) = 1$ $F(x, y)$ 关于 x, y 单调不减

(4) 边缘分布函数 $F_X(x) = F(x, +\infty)$ $F_Y(y) = F(+\infty, y)$

(5) 联合分布函数与边缘分布函数关系

3.2(1) = 维离散型随机变量

(2) = 维离散型随机变量的联合分布列与联合分布律 $P_{ij} = P(X=x_i, Y=y_j)$

联合分布函数

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P_{ij}$$

$X \backslash Y$	y_1	y_2	...	y_j	$P_{i \cdot}$
x_1					
x_2					
x_i					
$P_{\cdot j}$					1

(3) 离散型随机变量的边缘分布

$$P(X=x_i) = P(X=x_i, Y<+\infty) = P_{i \cdot} = P(X=x_i, Y=y_1) + P(X=x_i, Y=y_2) + \dots + P(X=x_i, Y=y_j)$$

$$P(Y=y_j) = P(X<+\infty, Y=y_j) = P_{\cdot j}$$

(4) 求联合分布列, 边缘分布列方法

(X, Y) 可能的取值 $(0, 0), (0, 1), (1, 0), (1, 1)$ $\therefore P(X=0, Y=0) = P_{00}, P(X=0, Y=1) = P_{01}, \dots$

或者 X 可能的取值 $0, 1, 2, 3, Y$ 可能的取值 $0, 1, 2$ $\therefore$ 表格得 P_{ij}

(5) 维离散型随机变量独立性

$$P(X=x_i, Y=y_j) = P(X=x_i) P(Y=y_j)$$

(1) 维离散型随机变量的条件分布列

$$P_{ij} = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} = P(X=x_i | Y=y_j)$$

3.3(1) = 维连续型随机变量联合分布函数 $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dy dx$

(b) = 维连续型随机变量的独立性

(2) = 维连续型随机变量的联合密度函数 $f(x, y)$

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

(3) 小性质 $f(x, y) \geq 0$

或 $F(x, y) = F_X(x) F_Y(y)$, 则随机变量 X, Y 独立

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dy dx = 1$$

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

(1) 常值 = 维连续型随机变量的分布

(4) 边缘密度函数 $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$

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

(2) = 维分布 四面体体积为 $\frac{1}{6}$

边缘分布函数 $F_X(x) = \int_{-\infty}^x \int_{-\infty}^{+\infty} f(x, y) dy dx$

$$f(x, y) = \begin{cases} \frac{1}{6} & (x, y) \in D \\ 0 & (x, y) \notin D \end{cases}$$

(5) $P(a < X < b) = F(b) - F(a) = \int_a^b f(x) dx$

(3) = 维正态分布 $(X, Y) \sim (x_1, x_2, \sigma_1, \sigma_2, \rho)$

若 X, Y 独立, 则 $\rho = 0$

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) dy dx$$

离散型 $P_{ij} = P(X=x_i, Y=y_j)$

连续型 $F(x, y) = P(X \leq x, Y \leq y) = P(a_1 < X < a_2, b_1 < Y < b_2)$

例: 设二维连续变量 (X, Y) 的分布函数为 $F(x, y)$, 则 (X, Y) 的分布函数 $F_1(x, y) =$ $F(x, y)$

$$F(x, y) = P(X \leq x, Y \leq y)$$

$$F_1(x) = P(X \leq x, Y \leq y) = P(Y \leq y, X \leq x) = F(y, x)$$

例: 设随机变量 X, Y 独立同分布, 且 X 的分布函数为 $F(x)$, 则 $Z = \max(X, Y)$ 的分布函数为

- A. $F^2(x)$ B. $F(x)F(y)$

独立同分布: $X \sim F(x) \quad F(x, y) = F_1(x)F_2(y)$
 $Y \sim F(y) \quad = F(x)F(y)$

$$F_Z(x) = P(Z \leq x) = P(X \leq x, Y \leq x) = F^2(x)$$

二维连续变量: 分布函数也是一元函数, 故不满足

例: 设随机变量 $X_i (i=1, 2, 3, 4)$ 相互独立, 且均服从分布 $B(1, \frac{1}{2})$, 则行列式 $X = \begin{vmatrix} x_1 & x_2 \\ x_3 & x_4 \end{vmatrix}$ 的分布函数为

X_1	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

$X = x_1 x_4 - x_2 x_3$

$P(X=1, X_4=1) = P(X_1=1)P(X_2=1) = \frac{1}{4}$

$\therefore \begin{vmatrix} X_1 & X_4 \\ P & \end{vmatrix} \begin{vmatrix} 1 \\ \frac{1}{4} \end{vmatrix}$ 同理 X_2, X_3

X	-1	0	1
P	$\frac{3}{16}$	$\frac{5}{8}$	$\frac{3}{16}$

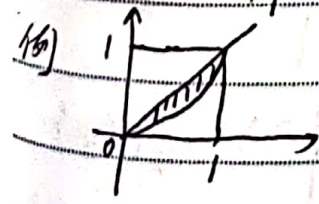
* 二维正态分布 $X \sim N(0, 1) \quad Y \sim N(1, 1) \quad \text{则 } X+Y \sim N(1, 2)$

$$aX+bY \sim (a\mu_1+b\mu_2, a^2\sigma_1^2+2ab\sigma_1\sigma_2\rho+b^2\sigma_2^2)$$

若独立则 $\rho=0$

max 函数 设二维连续变量 X, Y 相互独立, 且均服从相同 $[0, 1]$ 上均匀分布, 则 $P(\max(X, Y) \leq 1) =$ $\frac{1}{9}$

$$P(\max(X, Y) \leq 1) = P(X \leq 1, Y \leq 1) = P(X \leq 1)P(Y \leq 1) = \frac{1}{9}$$



面积 $Y=X$ 与 $Y=X^2$ 围成

$$\int_0^1 \int_{x^2}^x 1 dy dx = \int_0^1 (x-x^2) dx = \frac{1}{6}$$

若 (X, Y) 在该面积内均匀分布, 则 $f(x, y) = \begin{cases} 6 & (x, y) \in G \\ 0 & \text{其他} \end{cases}$

$$P(0 < X < \frac{1}{2}, 0 < Y < \frac{1}{2}) = \int_0^{\frac{1}{2}} \int_{x^2}^x 6 dy dx = \frac{1}{2}$$

* 二维正态分布 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ 若相互独立

则 $f(x, y) = f_X(x)f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 若相互独立, 则关于 X , 关于 Y 均

$P(X < Y) = P(Y < X) = \frac{1}{2} P(X < Y) \therefore P(X < Y) = \frac{1}{2}$

例: X 在 $(0,1)$ 内服从均匀分布, Y 在 $(0,1)$ 内服从均匀分布,

(1) 求 X, Y 的联合概率密度 (2) Y 的边缘概率密度 (3) 求 $P\{X+Y>1\}$

$$① f_{X,Y} = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{其它} \end{cases}$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \begin{cases} \frac{1}{x} & 0 < y < x \\ 0 & \text{其它} \end{cases}$$

$$\therefore f(x,y) = \begin{cases} \frac{1}{x} & 0 < y < x \\ 0 & \text{其它} \end{cases}$$

* 此时 $f(x,y)$ 的定义域为 $0 < y < x < 1$, $\therefore f(x,y) = \begin{cases} \frac{1}{x} & 0 < y < x < 1 \\ 0 & \text{其它} \end{cases}$

$$\therefore \int_0^1 \int_0^x \frac{1}{x} dy dx = 1 \quad \therefore \text{在 } 0 < x < 1 \text{ 以外 } f(x,y) = 0$$

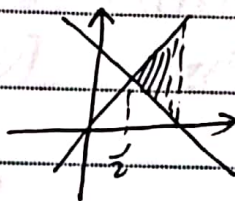
$$② f_Y(y) = \int_y^1 \frac{1}{x} dx = \begin{cases} -\ln y & 0 < y < 1 \\ 0 & \text{其它} \end{cases}$$

2. 求 $f(x,y) \xrightarrow{\text{边缘}} f_Y(y)$

$$③ P\{X+Y>1\} = \iint_{x+y>1} f(x,y) dy dx$$

$$= \int_{\frac{1}{2}}^1 \int_{1-x}^x \frac{1}{x} dy dx$$

$$= 1 - \ln 2$$



泊松分布的可加性 X, Y 相互独立, 且均服从 λ_1, λ_2 的泊松分布, 则 $Z = X+Y$ 服从 $\lambda = \lambda_1 + \lambda_2$ 的泊松分布

Γ 分布 X, Y 相互独立, 且服从参数为 p 的 Γ 分布, 则 $Z = X+Y$ 服从 $P(\theta = k) = (k-1)! p^{k-1} (1-p)^2$

例: 设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y) = \begin{cases} 1 & 0 < x < 1, 0 < y < x \\ 0 & \text{其它} \end{cases}$

求 (1) 求 (X,Y) 的边缘概率密度 $f_X(x), f_Y(y)$

(2) $Z = 2X - Y$ 的概率密度 $f_Z(z)$

(3) $P\{Y \leq \frac{1}{2} | X \leq \frac{1}{2}\}$

$$① f_X(x) = \int_0^x 1 dy = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{其它} \end{cases}$$

$$② F_Z(z) = P(Z \leq z) = P(2X - Y \leq z) = \iint_{2x-y \leq z} f(x,y) dx dy$$

$$= \iint_{2x-y \leq z} f(x,y) dx dy$$

$$f_Y(y) = \int_y^1 1 dx = \begin{cases} 1 - \frac{y}{2} & 0 < y < 2 \\ 0 & \text{其它} \end{cases}$$

$\therefore$ 讨论 z 的范围 $z < 0$ $F_Z(z) = 0$

$$0 < z \leq 2 \quad F_Z(z) = \iint_{2x-y \leq z} dx dy = 1 - (1 - \frac{z}{2})^2$$

$$= z - \frac{z^2}{4}$$

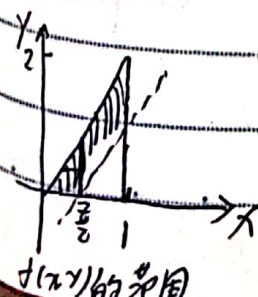
$$z > 2 \quad F_Z(z) = 1$$

$$③ P(Y \leq \frac{1}{2} | X \leq \frac{1}{2}) = \frac{P(Y \leq \frac{1}{2}, X \leq \frac{1}{2})}{P(X \leq \frac{1}{2})}$$

$$= \frac{\frac{3}{4} - \frac{\pi}{4}}{\frac{3}{4}} = \frac{3}{4}$$

$$\therefore F_Z(z) = \begin{cases} 0 & z < 0 \\ z - \frac{z^2}{4} & 0 < z \leq 2 \\ 1 & z > 2 \end{cases}$$

$$\therefore f_Z(z) = \begin{cases} 1 - \frac{z}{2} & 0 < z \leq 2 \\ 0 & \text{其它} \end{cases}$$



$f(x,y)$ 的范围

例 设随机变量 X, Y 相互独立, X 的根分布为 $P(X=i) = \frac{1}{3} (i=-1, 0, 1)$, Y 的根密度为 $f_Y(y) = \begin{cases} 1, & 0 \leq y < 1 \\ 0, & \text{其他} \end{cases}$

记 $Z = X + Y$

(1) 求 $P\{Z \leq \frac{1}{2} | X=0\}$

(2) 求 Z 的根密度 $f_Z(z)$.

$$\textcircled{1} \begin{array}{c|ccc} X & -1 & 0 & 1 \\ \hline P & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}$$

$$P\{Z \leq \frac{1}{2} | X=0\} = P\{X+Y \leq \frac{1}{2} | X=0\}$$

$$= P(Y \leq \frac{1}{2})$$

$$= F(\frac{1}{2})$$

$$= \frac{1}{2}$$

$$F_Y(y) = \int_0^y f_Y(y) dy = y$$

$$\textcircled{2} F_Z(z) = P(X+Y \leq z) = P(Y \leq z+1, X=-1) + P(Y \leq z, X=0) + P(Y \leq z-1, X=1)$$

$$= \frac{1}{3} [F_Y(z+1) + F_Y(z) + F_Y(z-1)] \quad z \in [-1, 2]$$

$$\therefore f_Z(z) = F'_Z(z) = \frac{1}{3} [f_Y(z+1) + f_Y(z) + f_Y(z-1)]$$

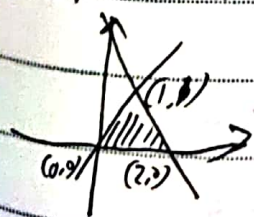
$$= \begin{cases} \frac{1}{3} & z \in (-1, 2) \\ 0 & \text{其他} \end{cases}$$

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi} \quad \int_{-\infty}^{+\infty} e^{-t^2} dy = \int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$$

例: 设二维随机变量 (X, Y) 服从区域 G 上的均匀分布, G 由 $x = X, x+y=2, y=0$ 围成的三角形

(1) 求 X 的根密度 $f_X(x)$

(2) 求条件根密度 $f_{Y|X}(y|x)$



$$f(x,y) = \begin{cases} 1 & 0 \leq y \leq 2-x \\ 0 & \text{其他} \end{cases}$$

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

$$= \int_y^{2-y} 1 dx = \begin{cases} 2-2y & 0 \leq y \leq 1 \\ 0 & \text{其他} \end{cases}$$

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y) dy$$

$$\text{当 } x < 0 \text{ 或 } x > 2 \text{ 时 } f_X(x) = 0$$

$$\text{当 } 0 \leq x < 1 \text{ 时 } f_X(x) = \int_0^x 1 dy = x$$

$$\text{当 } 1 \leq x < 2 \text{ 时 } f_X(x) = \int_{2-x}^0 1 dy = 2-x$$

$$\therefore f_{Y|X}(y|x) = \frac{1}{2-x} \quad 0 \leq y \leq 2-x$$

* min函数 设相互独立的两随机变量 X, Y 均服从 $E(1)$ 分布, 则 $P\{1 < \min(X, Y) < 2\}$ 的值为

$$F(x) = 1 - e^{-x}$$

$$P\{1 < \min(X, Y) < 2\} = P\{\min(X, Y) > 1\} - P\{\min(X, Y) \geq 2\}$$

$$= P\{X > 1, Y > 1\} - P\{X \geq 2, Y \geq 2\}$$

$$= P(X > 1)P(Y > 1) - P(X \geq 2)P(Y \geq 2)$$

$$= e^{-1} \cdot e^{-1} - e^{-2} \cdot e^{-2} = e^{-2} - e^{-4}$$

$$(X, Y) \sim F(x, y), P\{Z = \max(X, Y)\}$$

$$F_{\max}(z) = P\{\max(X, Y) \leq z\}$$

$$= P\{X \leq z, Y \leq z\}$$

$$\text{若相互独立} = F_X(z)F_Y(z)$$

$$F_{\min}(z) = P\{\min(X, Y) \leq z\}$$

$$= P(X \leq z) + P(Y \leq z) - P(X \leq z, Y \leq z)$$

$$= 1 - (1 - F_X(z))(1 - F_Y(z))$$

max函数 设两随机变量 X, Y 相互独立, 且均服从 $[0, 3]$ 上的均匀分布, $P\{\max(X, Y) \leq 1\} =$

$$P\{\max(X, Y) \leq 1\} = P\{X \leq 1, Y \leq 1\} = P(X \leq 1)P(Y \leq 1)$$

$$P\{\min(X, Y) < 2\} = P\{\min(X, Y) > 1\} - P\{\min(X, Y) \geq 2\}$$

$$= P\{X > 1, Y > 1\} - P\{X \geq 2, Y \geq 2\}$$

若相互独立

$$F_{\max}(z) = [F_X(z)]^2$$

$$F_{\min}(z) = 1 - [1 - F_X(z)]^2$$

* 若 $F_1(x), F_2(x)$ 为两个分布函数, 其相应的概率密度 $f_1(x), f_2(x)$ 为连续函数, 则下列函数是概率密度的()

A. $f_1(x)f_2(x)$ B. $f_1(x)F_2(x) + f_2(x)F_1(x)$

$F_1(x), F_2(x)$ 为分布函数, 则 $F_1(x) \cdot F_2(x)$ 为分布函数 则 $f_1(x)f_2(x) + f_2(x)f_1(x)$ 为概率密度函数

例: 设两随机变量 X, Y 相互独立, X 服从标准正态分布 $X \sim N(0, 1)$, Y 服从参数 $\lambda = \frac{1}{2}$ 的指数分布 $P(Y=0) = P(Y=1) = \frac{1}{2}$, 记 $F_Z(z)$ 为随机变量 $Z = XY$ 的分布函数, 则函数 $F_Z(z)$ 的间断点个数

$$Y \sim \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$X \sim N(0, 1)$$

$$F_Z(z) = P(Z \leq z) = P(XY \leq z) = P(Y=0)P(XY \leq z|Y=0) + P(Y=1)P(XY \leq z|Y=1)$$

$$= \frac{1}{2}P(0 \leq z|Y=0) + \frac{1}{2}P(X \leq z|Y=1)$$

$$= \frac{1}{2}P(0 \leq z) + \frac{1}{2}P(X \leq z)$$

$$= \begin{cases} \frac{1}{2}\phi(z) & z < 0 \\ \frac{1}{2} + \frac{1}{2}\phi(z) & z \geq 0 \end{cases}$$

$\therefore$ 1个间断点

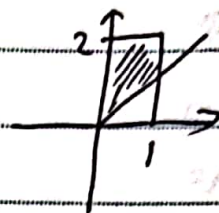
例: 设 X, Y 相互独立, $X \sim U(0,1), Y \sim U(0,1)$ 则 $P(X < Y) =$ _____

$$f(x) = 1$$

$$f(y) = \frac{1}{2}$$

$$f(x, y) = \begin{cases} \frac{1}{2} & 0 \leq x < 1, 0 \leq y < 1 \\ 0 & \text{其他} \end{cases}$$

$$\int_0^1 \int_0^1 \frac{1}{2} dy dx = \frac{1}{2} (2 - \frac{1}{2}) = \frac{3}{4}$$



例: 设随机变量 X 的概率密度函数为 $f(x) = \begin{cases} \frac{1}{4} & -2 \leq x \leq 0 \\ \frac{1}{2} & 0 \leq x \leq 1 \\ 0 & \text{其他} \end{cases}$, 记 $F(x, y)$ 为二维随机变量 (X, X^2) 的分布函数, 则 $F(-1, 4) =$ _____

$$F(-1, 4) = P(X \leq -1, X^2 \leq 4) = P(-2 \leq X \leq -1) = \int_{-2}^{-1} \frac{1}{4} dx = \frac{1}{4}$$

例: 设 X, Y 相互独立, $X \sim N(0,1), Y \sim U(-1,1)$, 求 $Z = X + Y$ 的概率密度函数 $f_z(z)$

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

$$f(y) = \frac{1}{2}$$

$$f(x, y) = f_X(x) f_Y(y) = \begin{cases} \frac{1}{2} \phi(x) & -\infty < x < +\infty, -1 \leq y \leq 1 \\ 0 & \text{其他} \end{cases}$$

$$f_z(z) = \int_{-\infty}^{+\infty} f(z-y, y) dy$$

卷积公式!!

$$= \int_{-1}^1 \frac{1}{2} \phi(z-y) dy$$

$$= \int_{z-1}^z \frac{1}{2} \phi(t) d(-t)$$

$$= \frac{1}{2} \left[\int_{-\infty}^{z+1} \phi(t) dt - \int_{-\infty}^{z-1} \phi(t) dt \right]$$

$$= \frac{1}{2} [\Phi(z+1) - \Phi(z-1)]$$

例: $f(x, y) = (1 + \sin x \sin y) p(x) p(y)$ $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 求

1) $f_X(x), f_Y(y)$

$$f_X(x) = \int_{-\infty}^{+\infty} (1 + \sin x \sin y) p(x) p(y) dy$$

$$= \int_{-\infty}^{+\infty} p(x) p(y) dy + \int_{-\infty}^{+\infty} \sin x \sin y p(x) p(y) dy$$

$$= p(x) \int_{-\infty}^{+\infty} p(y) dy + \sin x p(x) \int_{-\infty}^{+\infty} \sin y p(y) dy$$

$$= p(x)$$

奇函数

$$f_Y(y) = p(y)$$

	Y	0	1
X			
0	0.4	a	
1	b	0.1	

$\{X=0\}$ 与 $\{X+Y=1\}$ 不相容, 则 (A)

A $a=0.4$ $b=0.1$

B $a=0.2$ $b=0.3$

① $0.4+a+b=1$

$a+b=0.5$

② $P\{X=0, X+Y=1\}$

$=P\{X=0\}P\{X+Y=1\}$

$=0.4a$

$=P\{X=0\}P\{Y=1\}$

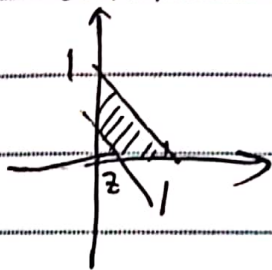
$=0.4b$

$P_i \cdot P_j = P_{ij}$

$\therefore 0.4a = 0.4b \therefore a=b$

例) $f(x,y) = \begin{cases} bx & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$

求 $Z=X+Y$ 的概率密度



$F_Z(z) = P\{Z \leq z\}$

$= P\{X+Y \leq z\}$

$= \iint_{x+y \leq z} bx \, dx \, dy$

$x+y \leq z$

$z < 0 \quad F_Z(z) = 0$

$0 < z < 1 \quad F_Z(z) = \int_0^z \int_0^{z-x} bx \, dy \, dx$

$= \frac{z^3}{6}$

$z > 1 \quad F_Z(z) = 1$

① 数学期望: 1) 离散型数学期望 $E(X) = \sum_{k=1}^{\infty} x_k p_k$ 判断 $E(X)$ 不存在 $\sum_{k=1}^{\infty} |x_k| p_k = +\infty$
 连续型数学期望 $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ $\int_{-\infty}^{+\infty} |x| f(x) dx = +\infty$

2) 常见随机变量的分布

= 二项分布 $X \sim B(n, p)$ $P(X=k) = \binom{n}{k} p^k q^{n-k}$ $E(X) = np$ $D(X) = npq$

泊松分布 $X \sim P(\lambda)$ $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $E(X) = \lambda$ $D(X) = \lambda$

几何分布 $X \sim G(p)$ $P(X=k) = p q^{k-1}$ $E(X) = \frac{1}{p}$ $D(X) = \frac{q}{p^2}$

均匀分布 $X \sim U(a, b)$ $f(x) = \frac{1}{b-a}$ $E(X) = \frac{a+b}{2}$ $D(X) = \frac{(b-a)^2}{12}$

指数分布 $X \sim E(\lambda)$ $f(x) = \lambda e^{-\lambda x}$ $E(X) = \frac{1}{\lambda}$ $D(X) = \frac{1}{\lambda^2}$

正态分布 $X \sim N(\mu, \sigma^2)$ $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $E(X) = \mu$ $D(X) = \sigma^2$

3) 随机变量的函数的数学期望 $Y = g(X)$ 类似 $X^2, |X|, e^X$ 的数学期望

离散型 $Y = g(X)$ $E(Y) = \sum_{k=1}^{\infty} g(x_k) p_k$

连续型 $Y = g(X)$ $E(Y) = \int_{-\infty}^{+\infty} g(x) f(x) dx$

4) 二维随机变量函数的数学期望 $Z = g(X, Y)$

离散型 $E(Z) = E[g(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) p_{ij}$

$E(X) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i p_{ij}$

$E(Y) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} y_j p_{ij}$

连续型 $E(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dy dx$

$E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dy dx$

5) 数学期望的性质

$E(C) = C$

$E(2X-3) = 2E(X)-3$

$E(CX) = C E(X)$

$E(X+3) = E(X)+3$

$E(X_1+X_2) = E(X_1)+E(X_2)$

若 X, Y 相互独立 $E(XY) = E(X)E(Y)$

$\sum_{k=1}^{\infty} \frac{x^k}{k!} = e^x - 1$

$E(g(X)h(Y)) = E(g(X)) E(h(Y))$

例: 判断数学期望是否存在 ① $P(X=k) = \frac{6}{\pi^2 k^2}$ $k=1, 2, \dots, n$ 的数学期望 ② 证 $f(x) = \frac{1}{x(1+x^2)}$ $x \in R$ 不存在数学期望

1) $E(X) = \sum_{k=1}^{\infty} x_k p_k = \sum_{k=1}^{\infty} \frac{6}{\pi^2 k^2} \cdot k = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = +\infty \therefore E(X)$ 不存在

2) $E(X) = \int_{-\infty}^{+\infty} |x| \cdot \frac{1}{(1+x^2)\pi} dx = \int_{-\infty}^{+\infty} \frac{|x|}{(1+x^2)\pi} dx = \frac{1}{\pi} \ln(1+x^2) \Big|_{-\infty}^{+\infty} = +\infty \therefore E(X)$ 不存在

例: 1) $X \sim P(5)$ 求 $E(X)$ 及 $E(\frac{1}{X+1})$

$$X \sim P(5) \quad P(X=k) = \frac{5^k}{k!} e^{-5}$$

$$E(X) = \sum_{k=0}^{\infty} k \frac{5^k}{k!} e^{-5} = e^{-5} \sum_{k=1}^{\infty} k \frac{5^k}{(k-1)!} = e^{-5} \cdot 5 \sum_{k=1}^{\infty} \frac{5^{k-1}}{(k-1)!} = 5e^{-5} \cdot e^5 = 5$$

$$E(\frac{1}{X+1}) = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{5^k}{k!} e^{-5} = \frac{e^{-5}}{5} \sum_{k=0}^{\infty} \frac{5^{k+1}}{(k+1)!} = \frac{e^{-5}}{5} \sum_{k=1}^{\infty} \frac{5^k}{k!} = \frac{e^{-5}}{5} (e^5 - 1) = \frac{1}{5}(1 - e^{-5})$$

2) $X \sim U(a, b)$ $Y = \frac{1}{b-a} \ln X$ 求 $E(Y)$

$$f(x) = \frac{1}{b-a}$$

$$E(Y) = \int_a^b \frac{1}{b-a} \ln x \cdot \frac{1}{b-a} dx = \frac{\pi}{(b-a)^2} \int_a^b x^2 dx = \frac{\pi}{24} (a+b)(a^2+b^2)$$

例: 离散型

$$E(Z) = \sum_{i=1}^n \sum_{j=1}^m g(x_i, y_j) p_{ij}$$

$$E(X) = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij}$$

$$E(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij}$$

连续型

$$E(Z) = E[g(x, y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy$$

$$E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dy dx$$

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

X \ Y	0	1	2	P _{i.}
-1	$\frac{1}{9}$	$\frac{2}{9}$	0	$\frac{1}{3}$
2	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$
P _{.j}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

例: $f(x, y) = \begin{cases} e^{-x} & 0 < x < +\infty \\ 0 & \text{其他} \end{cases}$ 求 $E(X), E(X^2), E(Y), E(Y^2)$

例: $E(X), E(X^2), E(XY), E(Y), E(Y^2), E(XY^2)$

$$E(X) = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij} = -1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3} = 1$$

$$E(X^2) = \sum_{i=1}^n \sum_{j=1}^m x_i^2 p_{ij} = (-1)^2 \cdot \frac{1}{9} + (-1)^2 \cdot \frac{2}{9} + (-1)^2 \cdot 0 + 2^2 \cdot \frac{2}{9} + 2^2 \cdot \frac{1}{9} + 2^2 \cdot \frac{2}{9} = 3$$

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{ij} = (-1) \cdot 0 \cdot \frac{1}{9} + (-1) \cdot 1 \cdot \frac{2}{9} + 2 \cdot 1 \cdot \frac{2}{9} + 2 \cdot 1 \cdot \frac{2}{9} = \frac{8}{9}$$

$$E(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij} = 0 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 2 \cdot \frac{2}{9} = \frac{7}{9}$$

$$E(Y^2) = \sum_{i=1}^n \sum_{j=1}^m y_j^2 p_{ij} = 0 \cdot \frac{1}{9} + 1 \cdot \frac{2}{9} + 4 \cdot \frac{2}{9} = \frac{11}{9}$$

$$E(XY^2) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j^2 p_{ij} = (-1) \cdot 0^2 \cdot \frac{1}{9} + (-1) \cdot 1^2 \cdot \frac{2}{9} + 2 \cdot 1^2 \cdot \frac{2}{9} + 2 \cdot 4 \cdot \frac{2}{9} = 7$$

$$E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} x e^{-x} dy dx = \int_0^{+\infty} x e^{-x} dx = 1$$

$$E(X^2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} x^2 e^{-x} dy dx = \int_0^{+\infty} x^2 e^{-x} dx = 2$$

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} xy e^{-x} dy dx = \int_0^{+\infty} x e^{-x} dx = 1$$

$$E(Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} y e^{-x} dx dy = \int_0^{+\infty} y e^{-y} dy = 1$$

$$E(Y^2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y^2 f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} y^2 e^{-x} dx dy = \int_0^{+\infty} y^2 e^{-y} dy = 2$$

$$E(XY^2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy^2 f(x, y) dx dy = \int_0^{+\infty} \int_0^{+\infty} xy^2 e^{-x} dy dx = \int_0^{+\infty} x y^2 e^{-x} dx = 7$$

①方差 $D(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$ 标准差 $\sigma(X) = \sqrt{D(X)}$

$$D(aX+b) = a^2 D(X)$$

$$D(C) = 0$$

$$D(X \pm Y) = D(X) \pm 2\text{COV}(X, Y) \pm D(Y) \quad \text{若相互独立} \quad D(X \pm Y) = D(X) + D(Y)$$

②协方差 $E[(X - E(X))(Y - E(Y))]$ 为随机变量 X, Y 的协方差, 记为 $\text{COV}(X, Y)$

若 X, Y 相互独立 $\text{COV}(X, Y) = 0 = E(XY) - E(X)E(Y)$

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

$$= \sqrt{D(X)D(Y)} \rho_{XY}$$

性质: $\text{COV}(X, Y) = \text{COV}(Y, X) \quad \text{COV}(X, C) = 0$ 若 X, Y 独立 则 $\text{COV}(X, Y) = 0$

$$\text{COV}(X, X) = D(X) \quad \text{COV}(aX+b, Y) = a\text{COV}(X, Y)$$

$$\text{COV}(aX, bY) = ab\text{COV}(X, Y)$$

$$\text{COV}(X_1 + X_2, Y) = \text{COV}(X_1, Y) + \text{COV}(X_2, Y)$$

$$D(aX+bY) = a^2 D(X) + b^2 D(Y) + 2ab\text{COV}(X, Y)$$

③相关系数 $\rho_{XY} = \frac{\text{COV}(X, Y)}{\sqrt{D(X)D(Y)}}$

$$|\rho_{XY}| \leq 1$$

当 $\rho_{XY} = 1$ 时 X 与 Y 有线性关系 $Y = aX + b, \quad \rho(Y = aX + b) = 1$

若 X, Y 相互独立, 则 $\rho_{XY} = 0$

若 $\rho_{XY} = 0$, 但 X, Y 未必相互独立 ($\rho_{XY} = 0 \Rightarrow$ 不线性相关, 但可能非线性相关: 不相互独立)

④二维正态分布 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

则 $\text{COV}(X, Y) = \sqrt{\sigma_1 \sigma_2} \rho = \sigma_1 \sigma_2 \rho$

(若 = 二维正态分布 则相互独立 $\Leftrightarrow X, Y$ 不相关)

⑤随机变量标准化 $X^* = \frac{X - E(X)}{\sqrt{D(X)}}$

$E(X^*) = 0 \quad D(X^*) = 1 \quad \rho_{XY} = \rho_{X^*Y^*} = E(X^*Y^*)$

⑥矩 $E(X^k)$ 为 X 的 k 阶原点矩

$E[(X - E(X))^k]$ 为 X 的 k 阶中心矩

$$\begin{aligned} \text{例 } \text{COV}(X, Y) &= \frac{1}{2}(\text{COV}(X, X) + \text{COV}(Y, Y)) \\ &= \frac{1}{2}(D(X) + D(Y)) \\ &= \frac{1}{2}(1 + 1) = 1 \end{aligned}$$

$$\text{例 } (X, Y) \sim N(1, 0, 9, 16, -\frac{1}{2})$$

设 $Z = \frac{X}{3} + \frac{Y}{4}$

(1) 求 Z 的概率密度函数

(2) X, Y 是否独立? 是否相关?

解: $X \sim N(1, 9), Y \sim N(0, 16)$

$E(Z) = \frac{1}{3}E(X) + \frac{1}{4}E(Y) = \frac{1}{3} \times 1 + \frac{1}{4} \times 0 = \frac{1}{3}$

$D(Z) = \frac{1}{9}D(X) + \frac{1}{16}D(Y) + 2 \times \frac{1}{3} \times \frac{1}{4} \times \text{COV}(X, Y)$

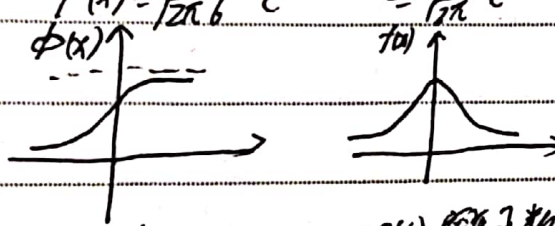
$= \frac{1}{9} \times 9 + \frac{1}{16} \times 16 + 2 \times \frac{1}{3} \times \frac{1}{4} \times (-\frac{1}{2} \times 3 \times 4) = 1 - 1 = 0$

例: X 服从 $\lambda=1$ 的泊松分布, $P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$

$$E(X) = \lambda = 1 \quad P(X) = \lambda = 1$$

$$E(X^2) = P(X) + [E(X)]^2 = 2$$

$$P(X=2) = \frac{1}{2!} e^{-1} = \frac{1}{2e}$$

例 标准正态分布函数 $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$
 $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

 $\phi(0) = \frac{1}{2}$
 $p(x)$ 密度函数关于 $x=0$ 对称

$X \sim N(\mu, \sigma^2)$ 的概率函数
 $F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
 $\therefore$ 对 $E(X) = \mu$
 $\text{若 } F(x) = (1) \Phi\left(\frac{x-\mu_1}{\sigma_1}\right) + (2) \Phi\left(\frac{x-\mu_2}{\sigma_2}\right)$
 $\text{则 } E(X) = (1)\mu_1 + (2)\mu_2$

$$\int_{-\infty}^{+\infty} p(x) dx = 1$$

$$\int_{-\infty}^{+\infty} x p(x) dx = E(X) = 0$$

例: 设随机变量 X 的分布函数 $F(x) = 0.5 \phi(x) + 0.5 \phi\left(\frac{x-4}{2}\right)$ 其中 $\phi(x)$ 为标准正态分布函数, 则 $E(X) =$

$$f(x) = 0.5 p(x) + \frac{1}{4} p\left(\frac{x-4}{2}\right)$$

$$\text{或 } E(X_1 + X_2) = E(X_1) + E(X_2)$$

$$0.5 \phi(x) \rightarrow E(X) = 0$$

$$E(X) = \int_{-\infty}^{+\infty} 0.5 x p(x) + \frac{1}{4} x p\left(\frac{x-4}{2}\right) dx$$

$$= \frac{1}{2} \int_{-\infty}^{+\infty} x p(x) dx + \int_{-\infty}^{+\infty} \frac{x-4}{2} p\left(\frac{x-4}{2}\right) d\left(\frac{x-4}{2}\right) + \int_{-\infty}^{+\infty} 2 p\left(\frac{x-4}{2}\right) d\left(\frac{x-4}{2}\right) \cdot 0.5 \phi\left(\frac{x-4}{2}\right) \rightarrow E(X) = 0.5 \cdot 4 = 2$$

$$= 0 + 0 + 2 = 2$$

$$\therefore E(X) = 2$$

例: 商店经营某商品, 每周进销量 X , 与需求量 Y 是相互独立的随机变量均服从 $(0, 20)$ 上的均匀分布, 商店每售出一单位商品获利 1000 元, 若 $Y > X$, 则从其它商店调货, 此时每件商品获利 500 元, 例: 利润期望值

令 Z 为利润

$$Z = g(X, Y) = \begin{cases} 1000Y & 10 \leq Y \leq X \leq 20 \\ 500X + 500Y & 10 \leq X < Y \leq 20 \\ 0 & \text{其他} \end{cases}$$

$$f(x, y) = \begin{cases} f(x) \cdot f(y) = \frac{1}{100} & 10 \leq x, y \leq 20 \\ 0 & \text{其他} \end{cases}$$

$$\begin{aligned} E(Z) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dy dx \\ &= \int_{10}^{20} \int_{10}^x 1000y \cdot \frac{1}{100} dy dx + \int_{10}^{20} \int_x^{20} (500x + 500y) \cdot \frac{1}{100} dy dx \\ &= 1416.7 \text{ 元} \end{aligned}$$

例: 设连续型随机变量 X 与 Y 相互独立且均服从 $N(0,1)$, X_1, X_2 的根号长度为 $\sqrt{2}$, Y 的根号长度为 $\sqrt{2}$, $Y_1 = \frac{1}{\sqrt{2}}(X_1 + X_2)$, $Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2)$ 则

A. $EY_1 = EY_2, DY_1 > DY_2$ B. $EY_1 > EY_2, DY_1 > DY_2$

(A)

$$DY_1 - DY_2 = EY_1^2 - (EY_1)^2 - EY_2^2 + (EY_2)^2$$

$$= EY_1^2 - EY_2^2$$

$$= \int_{-\infty}^{+\infty} y^2 \cdot \frac{1}{\sqrt{2}} [f_1(y) + f_2(y)] dy - \frac{1}{4} E(X_1 + X_2)^2$$

$$= \frac{1}{2} EY_1^2 + \frac{1}{2} EY_2^2 - \frac{1}{4} [EX_1^2 + 2EX_1EX_2 + EX_2^2]$$

$$= \frac{1}{4} EX_1^2 - \frac{1}{2} EX_1EX_2 + \frac{1}{4} EX_2^2$$

$$= \frac{1}{4} E(X_1 - X_2)^2 > 0$$

若 $\frac{1}{4} E(X_1 - X_2)^2 = 0$ 则 $D(X_1 - X_2) + [E(X_1 - X_2)]^2 = 0$

$\therefore D(X_1 - X_2) = 0, E(X_1 - X_2) = 0, \therefore X_1 = X_2$ 与 X_1, X_2 相互独立矛盾 $\therefore > 0$

$$E(Y_1) = \int_{-\infty}^{+\infty} y f_1(y) dy$$

$$= \int_{-\infty}^{+\infty} y \cdot \frac{1}{\sqrt{2}} [f_1(y) + f_2(y)] dy$$

$$= \frac{1}{2} \left[\int_{-\infty}^{+\infty} y f_1(y) dy + \int_{-\infty}^{+\infty} y f_2(y) dy \right]$$

$$= \frac{1}{2} (EX_1 + EX_2)$$

$$E(Y_2) = \frac{1}{2} (EX_1 - EX_2) \therefore EY_1 = EY_2$$

例: 设随机变量 X 服从标准正态分布 $N(0,1)$, $E(Xe^{2X}) = \underline{\quad\quad\quad} = 2e^2$

$$\begin{aligned} E(Xe^{2X}) &= \int_{-\infty}^{+\infty} xe^{2x} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 - 4x + 2}{2}} dx \cdot e^2 \\ &= e^2 \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}} dx \\ &= e^2 \cdot 2 = 2e^2 \end{aligned}$$

极大极小分布: 例: 设 X, Y 相互独立, 且 $EX \leq EY$ 有限, 记 $U = \max\{X, Y\}, V = \min\{X, Y\}$, 则 $E(UV) = \underline{\quad\quad\quad}$

$$U = \frac{X+Y+|X-Y|}{2}$$

$$V = \frac{X+Y-|X-Y|}{2}$$

$$\therefore E(UV) = E\left[\frac{(X+Y)^2 - (X-Y)^2}{4}\right] = E(XY) = E(X)E(Y)$$

例: X, Y 相互独立且均服从指数分布, 记 $U = \max\{X, Y\}, V = \min\{X, Y\}$

(1) 求 V 的分布函数 $F_V(v)$

$$F_V(v) = P(V \leq v) = P(X \leq v \cup Y \leq v) = P(X \leq v) + P(Y \leq v) - P(X \leq v, Y \leq v)$$

$$= 2F(v) - F(v)^2 \quad f(v) = 2e^{-2v} F(v)$$

$$F(v) = 1 - e^{-v}$$

例: 设随机变量 X, Y 的联合分布都在点 $(0,1), (1,0), (1,1)$ 为顶点的三角形区域服从均匀分布, 试求随机变量 $U = X+Y$ 的方差。



$$f(x,y) = \begin{cases} 2 & (x,y) \in G \\ 0 & (x,y) \notin G \end{cases}$$

$$\begin{aligned} E(U) &= \iint_G (x+y) \cdot 2 \, dy \, dx \\ &= \int_0^1 \int_{1-x}^1 (2x+2y) \, dy \, dx \\ &= \frac{4}{3} \end{aligned}$$

$$\begin{aligned} E(U^2) &= \iint_G 2(x+y)^2 \, dy \, dx \\ &= \frac{11}{6} \end{aligned}$$

$$D(U) = E(U^2) - (E(U))^2 = \frac{11}{6} - \frac{16}{9} = \frac{1}{18}$$

例: X 的概率密度 $f(x) = \frac{1}{2}e^{-|x|}, -\infty < x < +\infty$

(1) 求 X 数学期望与方差。

(2) X 与 $|X|$ 的协方差与相关性

(3) X 与 $|X|$ 是否相互独立。

$$\textcircled{1} E(X) = \int_{-\infty}^{+\infty} x f(x) \, dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{2} e^{-|x|} \, dx.$$

$$\textcircled{2} \text{Cov}(X, |X|) = E(X|X|) - E(X)E(|X|)$$

$\because x \cdot \frac{1}{2} e^{-|x|}$ 在 $(-\infty, +\infty)$ 上为奇函数

$$= \int_{-\infty}^{+\infty} x|x| f(x) \, dx = 0. \quad \text{奇函数}$$

$$\therefore E(X) = 0.$$

$\therefore X$ 与 $|X|$ 不相关

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x) \, dx = \int_0^{+\infty} x^2 e^{-x} \, dx = 2$$

$\textcircled{3}$ 不独立

$$\text{对于 } a \in (0, +\infty) \quad P(|X| \leq a) > 0 \quad P(X \leq a) < 1$$

$$P(X \leq a) P(|X| \leq a) < P(X \leq a)$$

$$P(X \leq a, |X| \leq a) = P(X \leq a) \neq P(X \leq a) P(|X| \leq a)$$

$\therefore$ 不独立

例: 求相关系数 ρ_{XY}

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

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

例: X, Y 均服从 $x^2 + y^2 \leq 1$ 内的均匀分布, 求 X, Y 相关系数

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

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

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

$$E(X) = \int_{-1}^{+1} x f_X(x) \, dx = \int_{-1}^{+1} x \cdot \frac{2}{\pi} \sqrt{1-x^2} \, dx = 0 \quad E(Y) = 0$$

$$E(XY) = \int_{-1}^{+1} \int_{-1}^{+1} xy f(x,y) \, dx \, dy = \iint_{x^2+y^2 \leq 1} xy \cdot \frac{1}{\pi} \, dx \, dy = 0 \quad \therefore \rho_{XY} = 0.$$

例 $f_X(x) = \begin{cases} \frac{1}{2} & -1 < x < 0 \\ \frac{1}{4} & 0 < x < 2 \\ 0 & \text{其他} \end{cases}$ 令 $Y = X^2$, $F(x, y)$ 为二维随机变量 (X, Y) 的分布函数

求 1) Y 的分布密度 $f_Y(y)$

2) $\text{cov}(X, Y)$

3) $F(x, y)$ 为 (X, Y) 分布函数求 $F(-\frac{1}{2}, y)$

① $Y = X^2$ $Y \leq 0$ 时 $F_Y(y) = 0$ $Y > 0$ 时 $F_Y(y) = F_Y(y) > 0$

$0 < Y < 1$ 时 $F_Y(y) = P(Y \leq X^2) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

$$= P(-\sqrt{y} \leq X < 0) + P(0 \leq X \leq \sqrt{y})$$

$$= \int_{-\sqrt{y}}^0 \frac{1}{2} dx + \int_0^{\sqrt{y}} \frac{1}{4} dx$$

$$= \frac{1}{2}\sqrt{y} + \frac{1}{4}\sqrt{y} = \frac{3}{4}\sqrt{y}$$

$$f_Y(y) = F'_Y(y) = \frac{3}{8\sqrt{y}}$$

$1 \leq Y < 4$ 时 $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

$$= P(-1 \leq X \leq 0) + P(0 \leq X \leq \sqrt{y})$$

$$= \int_{-1}^0 \frac{1}{2} dx + \int_0^{\sqrt{y}} \frac{1}{4} dx$$

$$= \frac{\sqrt{y}}{4} + \frac{1}{2}$$

$$f_Y(y) = F'_Y(y) = \frac{1}{8\sqrt{y}}$$

$$\therefore f_Y(y) = \begin{cases} \frac{3}{8\sqrt{y}} & 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & 1 \leq y < 4 \\ 0 & \text{其他} \end{cases}$$

② $\text{cov}(X, Y) = \text{cov}(X, X^2) = E(X^3) - E(X)E(X^2)$

$$E(X) = \int_{-\infty}^{+\infty} xf(x) dx = \int_{-1}^0 x \frac{1}{2} dx + \int_0^2 x \frac{1}{4} dx = \frac{1}{4}$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_{-1}^0 \frac{1}{2} x^2 dx + \int_0^2 x^2 \frac{1}{4} dx = \frac{5}{8}$$

$$E(X^3) = \frac{7}{8} \quad \therefore \text{cov}(X, Y) = \frac{7}{8} - \frac{1}{4} \times \frac{5}{8} = \frac{3}{8}$$

③ $F(-\frac{1}{2}, y) = P(X \leq -\frac{1}{2}, Y \leq y) = P(X \leq -\frac{1}{2}, X^2 \leq y) = P(-2 \leq X \leq -\frac{1}{2}) = P(-2 \leq X < -1) + P(-1 \leq X < -\frac{1}{2})$

$$= \frac{1}{4}$$

例 $X \sim P(1)$ $P\{E(\frac{X}{1+X})\} = \underline{\hspace{2cm}}$

$$E(\frac{X}{1+X}) = E(1 - \frac{1}{1+X}) = 1 - E(\frac{1}{1+X})$$

$$P(X=k) = \frac{1}{k!} e^{-1}$$

$$E(\frac{1}{1+X}) = \sum_{k=0}^{\infty} \frac{1}{1+k} \frac{1}{k!} e^{-1} = \sum_{k=0}^{\infty} \frac{1}{(k+1)!} e^{-1} = \sum_{k=1}^{\infty} \frac{1}{k!} e^{-1}$$

$$\therefore E(\frac{X}{1+X}) = e^{-1}$$

$$= 1 - \frac{1}{0!} e^{-1} = 1 - e^{-1}$$

例: 设 DX, DY 均且满 $\overline{D(X+Y)} = \overline{DX} + \overline{DY}$ 则 X, Y 相关系数 $\rho =$ _____

$$\begin{aligned}\overline{D(X+Y)} &= \overline{DX} + \overline{DY} \quad \therefore D(X+Y) = DX + DY + 2\overline{DXDY} \\ &= DX + DY + 2\text{cov}(X, Y) \\ &= DX + DY + 2\rho\overline{DXDY} \quad \therefore \rho = 1\end{aligned}$$

例: 随机变量 X 的分布函数 $F(x) = \begin{cases} 0 & x < -1 \\ 0.2 & -1 \leq x < 0 \\ 0.8 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$ 则 $E(X), E(2X+5), E(X^2), D(X^2)$.

附: 离散型分布

x	-1	0	1
p	0.2	0.6	0.2

x^2	0	1
p	0.6	0.4

$$E(X) = 0 \quad E(2X+5) = 2E(X)+5 = 5 \quad E(X^2) = 0.4 \quad D(X^2) = 0.24$$

例: X 服从参数为 λ 的泊松分布, 且知 $E[(X-1)(X-2)] = 1$, 则 $\lambda =$ _____

$$E[(X-1)(X-2)] = E(X^2 - 3X + 2) = E(X^2) - 3E(X) + 2 = 1 \quad \therefore \lambda = 1$$

例: X 的概率密度 $f(x) = \frac{1}{\pi(1+x^2)}$ ($-\infty < x < +\infty$) 求 $E(\min\{X, 1\})$

$$\begin{aligned}\text{关于对称} \quad \therefore E(\min\{X, 1\}) &= 2 \int_0^{+\infty} \frac{x}{\pi(1+x^2)} dx \\ &= 2 \left[\int_0^1 \frac{x}{\pi(1+x^2)} dx + \int_1^{+\infty} \frac{1}{\pi(1+x^2)} dx \right] \\ &= \frac{1}{\pi} \ln 2 + 2 \int_1^{+\infty} \frac{1}{\pi(1+x^2)} dx \\ &= \frac{1}{\pi} \ln 2 + \frac{2}{\pi} \arctan x \Big|_1^{+\infty} \\ &= \frac{1}{\pi} \ln 2 + \frac{1}{2}\end{aligned}$$

$X \backslash Y$	-1	0	1	$P_{i \cdot}$
0	0	$\frac{1}{3}$	0	$\frac{1}{3}$
1	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
$P_{\cdot j}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

1) $Z = XY$ 求 Z 的概率分布

2) 求 ρ_{XY}

$$P(Z = -1) = P(X = 1, Y = -1) = \frac{1}{3}$$

$$P(Z = 0) = P(X = 0, Y = -1) \cup P(X = 0, Y = 0) \cup P(X = 1, Y = 0) \cup P(X = 1, Y = 1) = \frac{1}{3} = \frac{0 - \frac{1}{3} \times 0}{\sqrt{\frac{1}{3} \times \frac{1}{3}}} = 0$$

$$\textcircled{2} \rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{DX} \sqrt{DY}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{DX} \sqrt{DY}}$$

例: 设 (X, Y) 服从二维正态分布, 则 $Z = X+Y, Y = X-Y$ 不相关的充要条件

$$A. E(X^2) - (EX)^2 = E(Y^2) - (EY)^2$$

Z, Y 不相关 $\Leftrightarrow \text{cov}(Z, Y) = 0$

$$\text{cov}(Z, Y) = E(ZY) - E(Z)E(Y)$$

$$= E(X^2) - E(Y^2) - (EX)^2 + (EY)^2$$

例: $(X, Y) \sim N(0, 0, 6, 6, 0)$, $Z_1 = 2X+Y, Z_2 = 2X-Y$ 则 Z_1, Z_2 相关性

$$DZ_1 = D(2X+Y) = 4DX + DY + 2 \cdot 2C = 40 + 10 = 50$$

$$\text{cov}(Z_1, Z_2) = E(Z_1 Z_2) - E(Z_1)E(Z_2)$$

$$DZ_1 = 50, DZ_2 = 50$$

$$= E(4X^2 - Y^2) - E(2X+Y)E(2X-Y)$$

$$= \frac{36^2}{50 \cdot 50} = \frac{3}{5}$$

$$= 4E(X^2) - E(Y^2) - 4(EX)^2 + (EY)^2$$

$$= 4D(X) - D(Y) = 36^2$$

① 切比雪夫不等式: 若随机变量 X 的数学期望 $E(X)=\mu$, $D(X)=\sigma^2$, 则对任意正数 ε 有

$$P\{|X-\mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2} \quad \text{或} \quad P\{|X-\mu| > \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

这个不等式说明 X 的方差 σ^2 越小, 则事件 $|X-\mu| < \varepsilon$ 发生的概率越大, X 取值越接近期望

② 大数定律: 频率收敛于概率, 通俗的讲, 100 名学生的平均身高比 5 名学生的平均身高更接近全校平均身高,

切比雪夫大数定律: $X_1, X_2, \dots, X_n$ 为独立同分布的随机数

$$\text{则 } \lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right\} = 1$$

(通俗讲, n 很大时, $X_1, X_2, X_3, \dots, X_n$ 的算术平均值无限趋近我们的数学期望平均值)

伯努利大数定律: $X_1, X_2, X_3, \dots, X_n$ 均服从二项分布 $B(n, p)$

$$\text{则 } \lim_{n \rightarrow \infty} P\left\{\left|\frac{nA}{n} - p\right| < \varepsilon\right\} = 1$$

辛钦大数定律: $X_1, X_2, \dots, X_n$ 独立同分布, 且 $E(X)=\mu$,

$$\text{则 } \lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right\} = 1$$

③ 中心极限定理: 个别随机变量数无限增加时, 它们的和趋于正态分布

① $X_1, X_2, \dots, X_n$ 独立同分布的随机变量 $E(X_i)=\mu$ $D(X_i)=\sigma^2$

$$\text{则 } \lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma^2} \leq x\right\} = \Phi(x) \quad \lim_{n \rightarrow \infty} P\left\{a < \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma^2} \leq b\right\} = \Phi(b) - \Phi(a)$$

② $X_1, X_2, \dots, X_n$ 独立同分布的随机变量, 且 $X_i \sim B(n, p)$

(用 n 个独立同分布的 $B(1, p)$)

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

例: 10000 盏灯, 夜晚每盏灯开灯概率 0.7, 则夜晚开灯在 6800-7200 盏的概率

设 X 为开灯盏数 $X \sim B(10000, 0.7)$ 为二项分布

$$E(X) = 10000 \times 0.7 = 7000 \quad D(X) = npq = 2100$$

$$P\{|X-7000| < 200\} \geq 1 - \frac{\sigma^2}{\varepsilon^2} = 1 - \frac{2100}{200^2} = 0.95$$

例: $X \sim P(2), Y \sim E(\frac{1}{2}), \text{且 } P_{XY} = \frac{1}{3}$ 试估计概率 $P\{|X-Y| > 5\}$

$$X \sim P(2) \quad E(X)=2 \quad D(X)=2$$

$$\text{令 } Z = X - Y$$

$$E(Z) = E(X-Y) = E(X) - E(Y) = 0$$

$$Y \sim E(\frac{1}{2}) \quad E(Y)=2 \quad D(Y)=4$$

$$D(Z) = D(X) + D(Y) - 2\text{Cov}(X, Y)$$

$$\text{Cov}(X, Y) = P_{XY} \sqrt{D_X D_Y} = \frac{1}{3}$$

$$= 2 + 4 - \frac{4}{3} = 6 - \frac{4}{3} = \frac{14}{3}$$

$$P\{|X-Y| > 5\} \leq \frac{\sigma^2}{\varepsilon^2} = \frac{14}{25} = \frac{14}{25}$$

例: 某餐厅时接待 400 人, 每人消费服从 $(20, 100)$ 上的分布, 且顾客消费额相互独立

1) 求该餐厅日均营业额

2) 求该餐厅日营业额在 760 元内概率

① 令 X_i 为第 i 位顾客的消费额, $X_i \sim U(20, 100)$

$$E(X_i) = \frac{20+100}{2} = 60$$

$$D(X_i) = \frac{(100-20)^2}{12} = \frac{1600}{3}$$

$$\sum_{i=1}^{400} X_i = \sum_{i=1}^{400} E(X_i) = 400 \times 60 = 24000 \text{ 元}$$

$$\begin{aligned} \text{② } P(-760 \leq \sum_{i=1}^{400} E(X_i) - 24000 \leq 760) &= \Phi\left(\frac{760}{\sqrt{400 \cdot \frac{1600}{3}}}\right) - \Phi\left(\frac{-760}{\sqrt{400 \cdot \frac{1600}{3}}}\right) \\ &= 2\Phi(0.645) - 1 = 0.90 \end{aligned}$$

例: 2500 人买保险, 1 人死之概率 0.002, 每人交 120 元, 一年内死亡, 则赔付 20000 元

① 亏本概率 ② 获利 100000 元概率

① 令 X 为死亡人数 $X \sim B(2500, 0.002)$ $E(X) = 5$ $D(X) = 4.99$

$2500 \times 120 \div 20000 = 15$ (人) 当死亡 15 人及以上时亏本

$$P(X \geq 15) = P\left(\frac{X - E(X)}{\sqrt{D(X)}} \geq \frac{15 - E(X)}{\sqrt{D(X)}}\right) = P\left(\frac{X - 5}{\sqrt{4.99}} \geq \frac{10}{\sqrt{4.99}}\right) = 1 - \Phi\left(\frac{10}{\sqrt{4.99}}\right) = 0.000004$$

② $(2500 \times 120 - 100000) = 10$ (人)

$$P(X \leq 10) = P\left(\frac{X - E(X)}{\sqrt{D(X)}} \leq \frac{10 - E(X)}{\sqrt{D(X)}}\right) = P\left(\frac{X - 5}{\sqrt{4.99}} \leq \frac{5}{\sqrt{4.99}}\right) = \Phi\left(\frac{5}{\sqrt{4.99}}\right) = 0.9871$$

例: 200 名工参加考试, 考试通过率 0.8, 试计算 200 人中有 150 人通过的的概率

设 $X_i = \begin{cases} 1 & \text{第 } i \text{ 人通过} \\ 0 & \text{第 } i \text{ 人未通过} \end{cases}$ $E(X_i) = np = 160$ $D(X_i) = npq = 32$

$$P\left(\sum_{i=1}^{200} X_i \geq 150\right) = 1 - \Phi\left(\frac{150 - np}{\sqrt{npq}}\right) = 1 - \Phi\left(\frac{150 - 160}{\sqrt{32}}\right) = \Phi(1.77) = 0.96$$

例: X, Y 的数学期望为 2, 2, 方差为 1, 4, 相关系数为 -0.5, 则 $P(|X+Y| \geq 6) \leq \frac{1}{12}$

$$E(X) = -2 \quad E(Y) = 2$$

$$E(X+Y) = E(X) + E(Y) = 0$$

$$P(|Z - u| \geq 6) \leq \frac{\sigma^2}{\varepsilon^2}$$

$$D(X) = 1 \quad D(Y) = 4$$

$$\begin{aligned} D(X+Y) &= D(X) + D(Y) + 2\rho\sqrt{D(X)D(Y)} \\ &= 1 + 4 + 2(-0.5) = 3 \end{aligned}$$

$$P(|Z| \geq 6) \leq \frac{3}{36} = \frac{1}{12}$$

例：一堆糖，每箱重量随机，平均每箱50kg，标准差5kg。若用最大载重5t的汽车运，试利用中心极限定理，说明每车两车最多装多少箱可以保证不超载概率大于0.977。（ $\Phi(2)=0.977$ ）

设 X_i 为第 i 箱重量

$$E(X_i) = 50$$

$$\sqrt{D(X_i)} = 5$$

$$P\left(\sum_{i=1}^n X_i \leq 5000\right) = P\left\{\frac{\sum_{i=1}^n X_i - nE(X)}{\sqrt{n} \cdot 5} \leq \frac{5000 - 50n}{\sqrt{n} \cdot 25}\right\} \approx \Phi\left(\frac{5000 - 50n}{5\sqrt{n}}\right) > 0.977 = \Phi(2)$$

$$\therefore \frac{5000 - 50n}{5\sqrt{n}} > 2$$

$$\therefore n < 98.0199 \quad \therefore \text{最多 } 98 \text{ 箱}$$

* 中心极限定理使用条件：独立同分布 且 $E(X)$, $D(X)$ 存在

大数定律的使用条件：独立同分布，且 $E(X)$, $D(X)$ 存在

独立大数定律，仅存在 $E(X)$ 即可

例：设随机变量 $X_1, X_2, \dots, X_n$ 相互独立，均服从参数为2的指数分布，当 $n \rightarrow \infty$ 时 $Y_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ 依概率收敛于

$$X_1^2, X_2^2, \dots, X_n^2 \text{ 也相互独立} \quad E(X) = \frac{1}{2} \quad D(X) = \frac{1}{4}$$

$$E(X^2) = D(X) + [E(X)]^2 = \frac{1}{2} \quad \therefore \frac{1}{2}$$

例：100个零件，损坏概率0.1，85个以上完好的概率。

$$\begin{pmatrix} \Phi(1.667) = 0.9522 \\ \Phi(1.647) = 0.95 \end{pmatrix}$$

设 X_i 为第 i 个零件

$$\begin{cases} X_i = 1 & \text{第 } i \text{ 个完好} \\ X_i = 0 & \text{第 } i \text{ 个损坏} \end{cases} \quad \begin{matrix} E(X_i) = 0.9 \\ D(X_i) = 0.09 \end{matrix}$$

$$P\left(\sum_{i=1}^{100} X_i > 85\right) = P\left(\frac{\sum_{i=1}^{100} X_i - 90}{\sqrt{9}} > \frac{85 - 90}{3}\right) = 1 - \Phi\left(-\frac{5}{3}\right) = \Phi\left(\frac{5}{3}\right) = 0.9522$$

例：若该系统有 n 个元件，要求80%以上完好，则 n 至少多大，80%以上完好的概率不小于95%，损坏概率0.1。

设 X_i 为第 i 个零件

$$\begin{cases} X_i = 1 & \text{第 } i \text{ 个完好} \\ X_i = 0 & \text{第 } i \text{ 个损坏} \end{cases} \quad \begin{matrix} E\left(\sum_{i=1}^n X_i\right) = 0.9n \\ D\left(\sum_{i=1}^n X_i\right) = 0.09n \end{matrix}$$

$$P\left(\sum_{i=1}^n X_i > 0.8n\right) = P\left(\frac{\sum_{i=1}^n X_i - 0.9n}{\sqrt{0.09n}} > \frac{0.8n - 0.9n}{\sqrt{0.09n}}\right) = 1 - \Phi\left(\frac{-0.1n}{\sqrt{0.09n}}\right) = \Phi\left(\frac{\sqrt{n}}{3}\right) \geq 0.95 \geq \Phi(1.645)$$

$$\therefore \frac{\sqrt{n}}{3} > 1.645 \quad n \geq 24.35 \quad \therefore n = 25 \text{ 个}$$

第六讲 数理统计的基本概念

- ① 数理统计:
- ① 抽样分布
 - ② 参数估计
 - ③ 假设检验

② 数理统计的三个基本概念: 总体, 样本, 统计量。

1) 在统计研究中, 往往研究个体的数量指标的分布情况, 故 所有数量指标的全体是总体, 总体往往用随机变量 X 或随机变量 X 的分布表示。

2) 使用 $X_1, X_2, X_3, \dots, X_n$ 表示样本, 样本要求: 代表性 $\rightarrow$ 每个个体有同等机会抽入样本, 即每样本与总体 X 有相同分布
独立性 $\rightarrow$ 每样本取值相互独立, 互不影响。

3) 使用 $x_1, x_2, \dots, x_n$ 表示样本值。

4) 统计量: 设 $X_1, X_2, \dots, X_n$ 是总体 X 的一样本, 若 $T(X_1, X_2, X_3, \dots, X_n)$ 是样本的函数, 且不含未知数, 则称为统计量。

③ 常见的统计量

样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

样本标准差 $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$

样本 k 阶矩 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$

样本中心矩 $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$

$E(X) = E(\bar{X}) = \mu$ $D(\bar{X}) = \frac{\sigma^2}{n}$ ① $X_1, X_2, \dots, X_n \sim N(0, 1)$ 则 $D(S^2) =$

$D(X) = \sigma^2$ $D(S^2) = \frac{2}{n-1}$ $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$ $D[\chi^2(n-1)] = 2(n-2)$ $\sigma=1$

$D(\bar{X}) = \frac{\sigma^2}{n}$ $\therefore D(S^2) = \frac{1}{(n-1)^2} D(n-1) S^2 = \frac{2}{n-1}$

$E(S^2) = D(X) = \sigma^2$

② $\bar{X} \sim N(0, \frac{1}{n})$ $\frac{\bar{X}}{\frac{1}{\sqrt{n}}} = \sqrt{n} \bar{X} \sim N(0, 1)$ $D(\sqrt{n} \bar{X}) = D(\chi^2(1)) = 2$

$D(\bar{X}) = \frac{1}{n} D(n \bar{X}) = \frac{2}{n}$

④ 样本的联合分布函数

$F_n(x_1, x_2, x_3, \dots, x_n) = F(x_1) \cdot F(x_2) \cdot \dots \cdot F(x_n) = \prod_{i=1}^n F(x_i)$

样本的联合密度函数

$f_n(x_1, x_2, x_3, \dots, x_n) = f(x_1) f(x_2) \cdot \dots \cdot f(x_n) = \prod_{i=1}^n f(x_i)$

⑤ 常用的统计量的分布 (正态分布, χ^2 分布(卡方分布), t 分布, F 分布)

1) 标准正态分布 $X \sim N(0, 1)$

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

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

2) χ^2 分布 (卡方分布)

定义: 设 $X_1, X_2, \dots, X_n$ 相互独立, 且均服从标准正态分布 $N(0, 1)$, 则称随机变量 $\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$ 服从自由度为 n 的 χ^2 分布

: 记作 $\chi^2 \sim \chi^2(n)$

性质: ① 可加性 $\chi_1^2 \sim \chi^2(n)$ $\chi_2^2 \sim \chi^2(m)$ 则 $\chi_1^2 + \chi_2^2 \sim \chi^2(n+m)$

(常写作 $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$ 则 $X+Y \sim \chi^2(n+m)$)

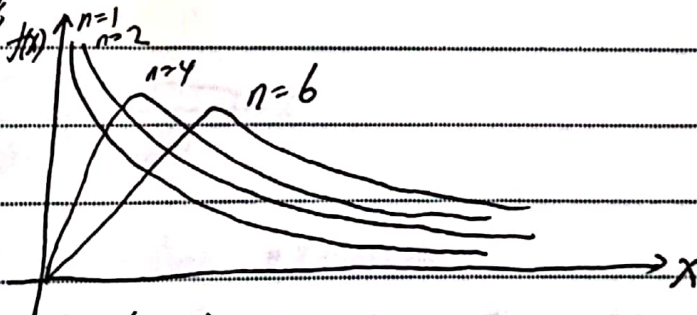
② $E(\chi^2)$ 若 $\chi^2 \sim \chi^2(n)$ 则 $E(\chi^2) = n$ $D(\chi^2) = 2n$

证明: $\because X_i \sim N(0, 1) \therefore E(X_i) = 0 \quad D(X_i) = 1$

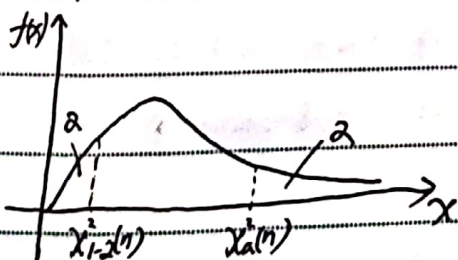
$$\therefore D(X) = E(X^2) - [E(X)]^2 \therefore E(X_i^2) = 1 \therefore E(\chi^2) = n$$

$$\therefore D(\chi^2) = E(\chi^4) - [E(\chi^2)]^2 \therefore E(\chi^4) = 3 \therefore D(\chi^2) = 2n$$

③ χ^2 分布的密度函数图形



④ χ^2 分布的上 α 分位点 (一般研究 $n \gg 2$)



设随机变量 $\chi^2 \sim \chi^2(n)$, 若对 $\alpha \in (0, 1)$, 存在实数 $\chi_{\alpha}^2(n)$

满足 $P\{\chi^2 > \chi_{\alpha}^2(n)\} = \alpha$, 则称 $\chi_{\alpha}^2(n)$ 为 χ^2 的上 α 分位点.

例: 若给 $\alpha = 0.05$, 求 $\chi_{0.05}^2(10)$ 与 $\chi_{1-0.05}^2(10)$

$$\chi_{0.05}^2(10) = 18.307 \quad \chi_{0.95}^2(10) = 3.940$$

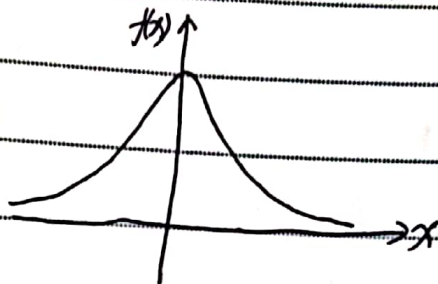
例: $X \sim N(0, 1)$, $X_1, X_2, \dots, X_n$ 是简单随机样本, 则 $X_1^2 + X_2^2$ 服从 $\chi^2(2)$ 分布

3) t分布 定义: $X \sim N(0,1)$, $Y \sim \chi^2(n)$ 且 X 与 Y 相互独立

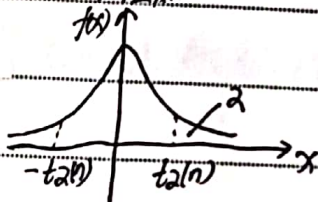
则称随机变量 $T = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 记作 $T \sim t(n)$

性质: ① t 分布为偶函数, $f(x) = f(-x)$, 当 $n \rightarrow \infty$ 时, $t(n)$ 分布近似于 $N(0,1)$ 分布。

② t 分布的图形



③ t 分布的上 α 分位点



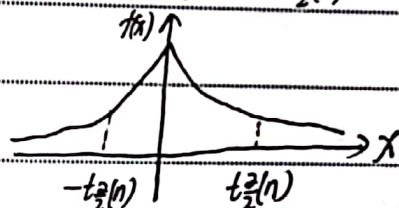
设 $T \sim t(n)$, 对给定的 $\alpha, \alpha \in (0,1)$, 有实数 $t_\alpha(n)$

满足 $P\{T > t_\alpha(n)\} = \alpha$, 则称 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位点。

$\because$ 偶函数 $\therefore t_{1-\alpha}(n) = -t_\alpha(n)$

即双侧面积为 α 的双侧分位点

t 分布的双侧分位点 $t_{\frac{\alpha}{2}}(n)$

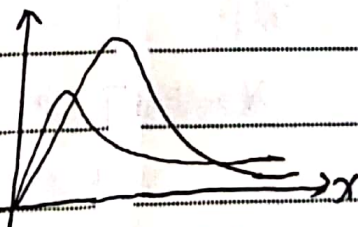


4) F 分布 定义: $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$ 且 X 与 Y 相互独立

则称随机变量 $F = \frac{X/n}{Y/m}$ 服从自由度为 n, m 的 F 分布, 记作 $F \sim F(n, m)$

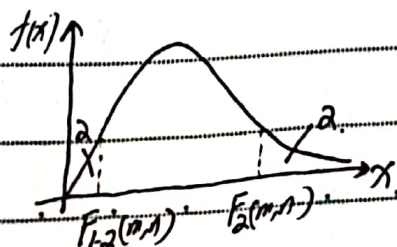
性质: ① $F \sim F(n, m)$, 则 $\frac{1}{F}$ 也服从 F 分布 $\frac{1}{F} \sim F(m, n)$

② F 分布的图形



③ 若 $t \sim t(n)$, 则 $t^2 \sim F(1, n)$ 分布

④ F 分布的上 α 分位点



设 $F \sim F(m, n)$, 对给定的 $\alpha, \alpha \in (0,1)$, 有实数 $F_\alpha(m, n)$

满足 $P\{F > F_\alpha(m, n)\} = \alpha$, 则称 $F_\alpha(m, n)$ 为 $F(m, n)$ 分布的上 α 分位点。

$$F_{1-\alpha}(m, n) = \frac{1}{F_\alpha(n, m)}$$

① 正态总体 $X \sim N(0, 1)$ 的抽样分布

当 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, X_3, \dots, X_n$ 为来自总体的样本, 样本均值为 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 则有

$$1) \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\text{即 } E(\bar{X}) = E(X) = \mu$$

$$D(\bar{X}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{n}$$

$$2) \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$3) \bar{X} \text{ 与 } S^2 \text{ 相互独立 } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$4) \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

$$5) \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$$

$$X \sim N(\mu, \sigma^2) \quad P(X \leq x) = F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

例: 若 $X \sim N(0, 1)$, $X_1, X_2, \dots, X_n$ 是样本, 则下列统计量服从什么分布

$$① X_1^2 + X_2^2$$

$$① X_1^2 + X_2^2 \sim \chi^2(2) \text{ 分布}$$

$$② \frac{X_1 - X_2}{\sqrt{X_3^2 + X_4^2}}$$

$$② X_1 \sim N(0, 1) \quad X_1 - X_2 \sim N(0, 2)$$

$$E(X_1) = E(X_2) = 0$$

$$X_3^2 + X_4^2 \sim \chi^2(2) \quad \frac{X_1 - X_2}{\sqrt{2}} \sim N(0, 1)$$

$$D(X_1 - X_2) = D(X_1) + D(X_2) = 2$$

$X_1, X_2, \dots, X_n$ 相互独立.

$$\therefore \frac{(X_1 - X_2)/\sqrt{2}}{\sqrt{(X_3^2 + X_4^2)/2}} \sim t(2)$$

$$③ \frac{X_2}{\sqrt{X_3^2 + X_4^2}}$$

$$③ X_2 \sim N(0, 1)$$

$$|X_2| = \sqrt{X_2^2/1} \quad \therefore \frac{X_2}{\sqrt{X_3^2 + X_4^2}} \sim t(1)$$

$$④ \frac{(\frac{n-1}{3}) \sum_{i=1}^3 X_i^2}{\sum_{i=4}^n X_i^2}$$

$$④ \sum_{i=1}^3 X_i^2 \sim \chi^2(3)$$

$$\sum_{i=4}^n X_i^2 \sim \chi^2(n-3)$$

$$\frac{(\frac{n-1}{3}) \sum_{i=1}^3 X_i^2}{\sum_{i=4}^n X_i^2} = \frac{\sum_{i=1}^3 X_i^2/3}{\sum_{i=4}^n X_i^2/(n-3)} \sim F(3, n-3)$$

例: $X \sim N(52, 6.3^2)$, $X_1, X_2, X_3, \dots, X_n$ 为样本, $\bar{X}$ 为样本均值.

(1) 求 $\bar{X}$ 的数学期望与方差

$$(2) P(50.8 < \bar{X} < 53.8)$$

$$① E(\bar{X}) = E(X) = 52$$

$$② P(50.8 < \bar{X} < 53.8) = P\left(\frac{50.8-52}{6.3/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{53.8-52}{6.3/\sqrt{n}}\right)$$

$$= \Phi(1.714) - \Phi(-1.714)$$

$$= 0.9564 - (1 - 0.9564)$$

$$= 0.8293$$

$$D(\bar{X}) = \frac{D(X)}{n} = \frac{6.3^2}{36} = 1.1025$$

例: 设总体 $X \sim E(\lambda)$, 则来自总体 X 的样本 $X_1, X_2, \dots, X_n$ 的联合概率密度 $f_n(x_1, x_2, \dots, x_n) =$ _____

$$X \sim E(\lambda) \quad f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad f_n(x_1, x_2, \dots, x_n) = \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^n x_i} & x_1, x_2, \dots, x_n > 0 \\ 0 & \text{其它} \end{cases}$$

$$= \prod_{i=1}^n f(x_i)$$

例: 总体 X 服从参数为 p 的 0-1 分布, 则来自总体 X 的简单随机样本 $X_1, X_2, \dots, X_n$ 的联合分布为

X	0	1
P	$1-p$	p

$$P(x) = \begin{cases} (1-p)^{1-x} p^x & \text{当 } x=0 \text{ 或 } x=1 \text{ 时} \\ 0 & \text{其它} \end{cases}$$

$$\text{联合分布 } P_n(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i) = \begin{cases} p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} & x_1, x_2, \dots, x_n > 0 \\ 0 & \text{其它} \end{cases}$$

例 $\chi^2 \sim \chi^2(n)$, 则 $E(\chi^2) =$ _____ $D(\chi^2) =$ _____

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

$$X_i \sim N(0, 1) \quad E(X_i) = 0 \quad D(X_i) = 1 \quad E(X_i^2) = 1$$

$$E(\chi^2) = E(X_1^2 + X_2^2 + \dots + X_n^2) = n$$

$$D(\chi^2) = 2n$$

例 X_1, X_2, X_3, X_4 是来自总体 $N(0, 2^2)$ 的简单随机样本, 已知 $\chi^2 = \frac{(X_1 - 2X_2)^2}{a} + \frac{(3X_3 - 4X_4)^2}{b}$ 服从 $\chi^2(n)$ 的分布,

则 $a =$ _____ $b =$ _____ $n =$ _____

$$\chi^2(n) = X_1^2 + X_2^2 \quad \text{其中 } X_1, X_2 \sim N(0, 1)$$

$$\therefore \frac{X_1 - 2X_2}{\sqrt{a}} \sim N(0, 1)$$

$$E(X_1 - 2X_2) = 0 \quad D(X_1 - 2X_2) = D(X_1) + 4D(X_2) = 20$$

$$\therefore \frac{X_1 - 2X_2}{\sqrt{20}} \sim N(0, 1) \quad \therefore a = 20$$

$$\text{同理 } E(3X_3 - 4X_4) = 0 \quad D(3X_3 - 4X_4) = 9D(X_3) + 16D(X_4) = 100$$

$$\therefore b = 100$$

$$\therefore n = 2$$

例: $X_1, X_2, \dots, X_n$ 为来自总体 $X \sim N(0, 1)$ 的样本, 则 $\frac{(n-1) \sum_{i=1}^n X_i^2}{2}$ 服从 _____ 分布 $\frac{\sum_{i=1}^n X_i^2}{2} \sim F(1, n-1)$

例: $X_1, X_2, \dots, X_n$ 为来自 $X \sim N(0, \sigma^2)$ 的样本, 则 $\chi^2 = \left(\frac{\sum_{i=1}^n X_i}{\sigma} \right)^2$ 服从 _____ 分布 $\frac{(\sum_{i=1}^n X_i)^2 / n}{(\sum_{i=1}^n X_i^2) / (n-1)} \sim F(1, n-1)$

例 X_1, X_2, X_3 相互独立, 且服从 $N(0, \sigma^2)$, 则 $\sqrt{\frac{2}{3}} \frac{X_1 + X_2 + X_3}{|X_2 - X_3|}$ 服从 _____ 分布

$$X_1 + X_2 + X_3 \sim N(0, 3\sigma^2)$$

$$(X_1 + X_2 + X_3) / \sqrt{3} \sigma \sim N(0, 1)$$

$$|X_2 - X_3| = \sqrt{(X_2 - X_3)^2}$$

$$\left(\frac{X_2 - X_3}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$$

$$\therefore |X_2 - X_3| / \sqrt{2}\sigma = \sqrt{\left(\frac{X_2 - X_3}{\sqrt{2}\sigma}\right)^2} \sim \sqrt{\chi^2(1)}$$

$$\therefore \sqrt{\frac{2}{3}} \frac{X_1 + X_2 + X_3}{|X_2 - X_3|} \sim \frac{X}{\sqrt{Y/1}} \sim t(1)$$

例: $X_1, X_2, \dots, X_n$ 来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 样本方差 S^2 , 则 $D(S^2) =$ _____

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1)$$

$$D[\chi^2(n-1)] = 2(n-1) = 2n-2$$

$$\therefore \left(\frac{n-1}{\sigma^2}\right)^2 D(S^2) = 2n-2$$

$$D(S^2) = \frac{2\sigma^4}{n-1}$$

例: 总体 X 服从分布 $N(\mu, \sigma^2)$, 抽取简单样本 $X_1, X_2, \dots, X_n$, 其样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, 求统计量

$Y = \sum_{i=1}^n (X_i + X_{n+1} - 2\bar{X})$ 的数学期望 EY .

构造新的简单随机样本 $Z = (X_1 + X_{n+1}), (X_2 + X_{n+2}), \dots$

$$Z \sim N(2\mu, 2\sigma^2)$$

$$\text{新样本均值 } E\left(\frac{1}{n} \sum_{i=1}^n (X_i + X_{n+1})\right) = \frac{1}{n} \sum_{i=1}^n X_i = 2\bar{X}$$

$$\text{新样本方差 } D\left(\frac{1}{n-1} \sum_{i=1}^n (X_i + X_{n+1} - 2\bar{X})\right) = 2\sigma^2 = \frac{1}{n-1} Y$$

$$E\left(\frac{1}{n-1} Y\right) = 2\sigma^2 \therefore E(Y) = 2(n-1)\sigma^2$$

例: 设总体 X 的概率密度 $f(x) = \begin{cases} \frac{2x}{3\theta^2} & 0 < x < \theta \\ 0 & \text{其他} \end{cases}$, $X_1, X_2, \dots, X_n$ 为样本, 则 $E\left(\sum_{i=1}^n X_i^2\right) =$ _____

$$E\left(\sum_{i=1}^n X_i^2\right) = E(X_1^2 + X_2^2 + \dots + X_n^2) = n E(X^2) = n \int_0^\theta \frac{2x}{3\theta^2} x^2 dx = \frac{2n}{3\theta^2} \int_0^\theta x^3 dx = \frac{2}{3} n \theta^2$$

例: 设总体 X 的概率密度 $f(x) = \frac{1}{2} e^{-|x|}$ ($-\infty < x < +\infty$), $X_1, X_2, \dots, X_n$ 为样本, 样本方差为 S^2 , 则 $E(S^2) =$ _____

$$E(S^2) = D(X) = E(X^2) - [E(X)]^2$$

$$E(X) = \frac{1}{2} \int_{-\infty}^{+\infty} x e^{-|x|} dx = 0$$

$$E(X^2) = \frac{1}{2} \int_{-\infty}^{+\infty} x^2 e^{-|x|} dx = \frac{1}{2} \int_0^{+\infty} x^2 e^{-x} dx = 2$$

$$\therefore E(S^2) = 2$$

第七讲

参数估计

① 上讲的数理统计：已知样本信息推出总体信息，即已知样本分布推出总体分布。

这讲的参数估计：已知样本信息，已知总体分布类型，推出总体分布的参数。

② 本章内容

- 点估计 $\rightarrow$
 - 定义
 - 矩估计法 $\rightarrow$
 - 一阶矩 $\rightarrow$ 期望
 - 二阶矩 $\rightarrow$ 方差
 - 最大似然估计法
- $\rightarrow$ 对点估计优良性的评价 $\rightarrow$
 - 无偏性
 - 有效性
 - 一致性
- 区间估计 $\rightarrow$ 置信区间

③ 点估计：1) 矩法：设总体 X 的分布函数形式已知 $F(x, \theta)$ ，其中 θ 为未知参数，我们需要估计参数 θ 。

方法是：从总体中抽取样本 $X_1, X_2, \dots, X_n$ ，构造合适的统计量记为 θ 的估计量，记为 $\hat{\theta} = g(X_1, X_2, \dots, X_n)$ 。

代入样本值，得到 θ 的估计值 $\hat{\theta} = g(x_1, x_2, \dots, x_n)$ 。

2) 矩估计法：使用样本矩代替总体矩。

3) 一阶矩：即一个未知参数，设总体 X 分布为 $F(x, \theta)$ ， θ 为待估参数， $X_1, X_2, \dots, X_n$ 为总体的样本。

若总体 X 的数学期望 $E(X)$ 存在，则 $E(X)$ 为 θ 的函数 $h(\theta)$ ，由大数定律 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E(X) = h(\theta)$ 。

则得到 θ 的矩估计值。

离散型：例：设总体 X 的分布律

X	1	2	3
P_k	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中 θ 为未知参数， $\theta \in (0, 1)$ ，若已取样本值 $x_1=1, x_2=2, x_3=3$ ，求 θ 的矩估计值。

$$E(X) = \theta^2 + 4\theta(1-\theta) + 9(1-\theta)^2 = 3 - 2\theta = \bar{X} \quad \therefore \theta \text{ 的矩估计量为 } \hat{\theta} = \frac{3-\bar{X}}{2}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = 2 \quad \therefore \theta \text{ 的矩估计值为 } \hat{\theta} = \frac{1}{2}$$

连续型：例：设总体 X 的概率密度 $f(x, \theta) = \begin{cases} (\theta+1)x^\theta & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$ ，其中 $\theta > -1$ 为未知参数， $X_1, X_2, \dots, X_n$ 为来自总体 X 的样本，求 θ 的矩估计量。

$$E(X) = \int_{-\infty}^{+\infty} xf(x, \theta) dx = \int_0^1 x(\theta+1)x^\theta dx = \int_0^1 (\theta+1)x^{\theta+1} dx = \frac{\theta+1}{\theta+2} = \bar{X}$$

$$\therefore \theta \text{ 的矩估计量 } \hat{\theta} = \frac{1-2\bar{X}}{\bar{X}-1}$$

4) 二阶矩：当总体 X 的分布含两个未知参数时，则令 $E(X) = \bar{X}$ ， $E(X^2) = A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ 。

一般情况考虑 $X \sim U(a, b)$ 或 $X \sim N(\mu, \sigma^2)$ ，见例。

例: 设总体 $X \sim U(a, b)$, 其中 a, b 为未知参数, $X_1, X_2, \dots, X_n$ 为来自总体的一样本, 求 a, b 的矩估计量.

$$X \sim U(a, b)$$

$$E(X) = \frac{a+b}{2}$$

$$D(X) = \frac{(b-a)^2}{12}$$

$$E(X^2) = D(X) + [E(X)]^2 = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2$$

$$\text{令 } E(X) = \bar{X}$$

$$E(X^2) = A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

$$\therefore \hat{a} = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{b} = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

例: $X_1, X_2, \dots, X_n \sim U(a, a)$ 上的分布

的材料, 则 a 的矩估计量,

$$E(X) = \frac{a+a}{2} = a \quad \text{不含 } a: \text{ 用 } E(X^2)$$

$$E(X^2) = D(X) + [E(X)]^2 = \frac{a^2}{12} + a^2 = \frac{13a^2}{12}$$

$$E(X^2) = A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{a^2}{3}$$

$$\therefore \hat{a} = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}$$

例: 设总体 $X \sim N(\mu, \sigma^2)$, $X_1, X_2, \dots, X_n$ 为来自总体的一样本, 求 μ, σ^2 的矩估计量

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

$$E(X) = \mu$$

$$D(X) = \sigma^2$$

$$E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2$$

$$\text{令 } E(X) = \bar{X}$$

$$E(X^2) = A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

$$\Rightarrow \begin{cases} \mu = \bar{X} \\ \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

$$\therefore \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}$$

5) 最大似然估计法 (又称最大可能性估计)

① 原理: 最大似然原理, 即样本出现的结果最可能是概率最大的结果

例: $X \sim B(4, p)$, p 未知, X_1, X_2, X_3 为来自总体 X 的样本, 若样本观测值为 1, 2, 1, 则 p 取何值更为合理?

$$P(A) = P(X_1=1, X_2=2, X_3=1) = P(X_1=1) \cdot P(X_2=2) \cdot P(X_3=1)$$

当 $P(A)$ 值最大时最合理

$$= \binom{4}{1} p^1 (1-p)^3 \cdot \binom{4}{2} p^2 (1-p)^2 \cdot \binom{4}{1} p^1 (1-p)^3$$

$$f(x) = 96 p^4 (1-p)^8 \quad f'(x) = 96 p^4 (1-p)^7 (-1)$$

$$= 96 p^4 (1-p)^8$$

$\therefore$ 当 $p = \frac{1}{3}$ 时 $f(x)$ max, 即 $p = \frac{1}{3}$ 时最合理.

② 似然函数 $L(\theta)$: 设 $X_1, X_2, \dots, X_n$ 为来自总体 X 的简单随机样本, $X_1, X_2, \dots, X_n$ 为样本观测值, θ 为待估参数;

则 $L(\theta) = \prod_{i=1}^n p(X_i; \theta)$ 为参数 θ 的似然函数 (即样本 X_i 对应的概率 p 的乘积)

对于离散型: $L(\theta) = P(X_1=X_1, X_2=X_2, \dots, X_n=X_n) = P(X_1=X_1) \cdot P(X_2=X_2) \cdot \dots \cdot P(X_n=X_n) = \prod_{i=1}^n p(X_i; \theta)$

对于连续型: $L(\theta) = \prod_{i=1}^n f(X_i; \theta)$

③ 最大似然估计

对于似然函数 $L(\theta)$ 存在一个实数 $\hat{\theta}$, 使得 $L(\hat{\theta}) = \max L(\theta)$, 则 $\hat{\theta}$ 为参数 θ

的最大似然估计值, $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为 θ 的最大似然估计量。

(即通过求 $L(\theta)$ 的极大值, 求参数 θ 的最大似然估计值 $\hat{\theta}$)

例: 总体 X 的分布律

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

取样本 X_1, X_2, X_3 , 又恰的样本观测值为 1, 2, 3

则 θ 的最大似然估计值?

$$L(\theta) = \prod_{i=1}^n P(X_i; \theta) = \theta^2 \cdot 2\theta(1-\theta) \cdot (1-\theta)^2 = 2\theta^3(1-\theta)^3$$

$$\frac{d}{d\theta} L(\theta) = 6\theta^2(1-\theta)^2(1-2\theta) = 0$$

$$\therefore \theta \text{ 可能取值 } 0, 1, \frac{1}{2} \quad \because \theta \in (0, 1)$$

$$\therefore \hat{\theta} = \frac{1}{2} \text{ 为 } \theta \text{ 的最大似然估计值.}$$

X	X	0	1	2	3	6个样本观测值
P	θ^2	$2\theta(1-\theta)$	$\theta^2(1-\theta)^2$			3, 1, 3, 0, 3, 1, 2, 3

则 θ 的极大似然估计与最大似然估计值

$$1) E(X) = 2\theta(1-\theta) + 2\theta^2 + 3(1-\theta)^2 = 3 - \theta$$

$$\bar{x} = 3 + 1 + 3 + 0 + 3 + 1 + 2 + 3 = 2 \quad \text{令 } E(X) = \bar{x} \quad \therefore \hat{\theta} = \frac{1}{4}$$

$$2) L(\theta) = \prod_{i=1}^n P(X_i; \theta) = \theta^2 [2\theta(1-\theta)]^2 \cdot \theta^2 \cdot (1-\theta)^4 = \theta^6(1-\theta)^6$$

$$\ln L(\theta) = \ln \theta^6 + 6 \ln(1-\theta) \quad \frac{d}{d\theta} \ln L(\theta) = \frac{6}{\theta} - \frac{6}{1-\theta} = 0 \quad \hat{\theta} = \frac{1}{2}$$

注: 当计算似然函数 $L(\theta)$ 的极大值时, 往往比较复杂

例 $X \sim P(\lambda)$, $\lambda > 0$ 的未知参数, $X_1, X_2, \dots, X_n$ 为总体的样本, 求 λ 的最大似然估计量。

可以考察对数似然函数 $\ln L(\theta)$, $L(\theta)$ 与 $\ln L(\theta)$ 有相同极值, 即相同的 $\hat{\theta}$

$$X \sim P(\lambda) \quad P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$L(\lambda) = \prod_{i=1}^n P(X_i; \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\lambda}$$

$$\text{取对数似然函数 } \ln L(\lambda) = \sum_{i=1}^n x_i \ln \lambda - n\lambda - \sum_{i=1}^n \ln x_i!$$

$$\text{令 } \frac{d \ln L(\lambda)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0$$

$$\text{则 } \lambda = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\therefore \lambda \text{ 的最大似然估计量 } \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}$$

例: 总体 X 的概率密度 $f(x; \theta) = \begin{cases} (\theta+1)x^\theta & 0 < x < 1 \\ 0 & \text{其它} \end{cases}$, $X_1, X_2, \dots, X_n$ 为样本求 θ 的最大似然估计量。

$$f(x; \theta) = \begin{cases} (\theta+1)x^\theta & 0 < x < 1 \\ 0 & \text{其它} \end{cases}$$

$$L(\theta) = \prod_{i=1}^n f(X_i; \theta) = \prod_{i=1}^n (\theta+1)X_i^\theta = (\theta+1)^n \left(\prod_{i=1}^n X_i \right)^\theta$$

$$\text{取对数似然函数 } \ln L(\theta) = n \ln(\theta+1) + \theta \sum_{i=1}^n \ln X_i$$

$$\frac{d}{d\theta} \ln L(\theta) = \frac{n}{\theta+1} + \sum_{i=1}^n \ln X_i = 0$$

$$\therefore \theta \text{ 的最大似然估计量 } \hat{\theta} = \frac{n}{\sum_{i=1}^n \ln X_i} - 1$$

④ 点估计的无偏性评价标准

1) 无偏性: 若参数 θ 的估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 满足 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计量。

例: 设 $X_1, X_2, \dots, X_n$ 为总体 X 的一个样本, 已知 $E(X) = \mu$, $D(X) = \sigma^2$

1) 证样本均值 $\bar{X}$ 与样本方差 S^2 为 μ, σ^2 的无偏估计量

2) 判定 $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 是否是 σ^2 的无偏估计量。

$$\textcircled{1} E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n\mu = \mu$$

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \frac{1}{n-1} E\left[\sum_{i=1}^n X_i^2 - n\bar{X}^2\right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n (\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right] \\ &= \frac{1}{n-1} (n\mu^2 + n\sigma^2 - \mu^2 n - \sigma^2) \\ &= \sigma^2 \end{aligned}$$

$$\begin{aligned} \textcircled{2} E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] &= \frac{1}{n} \sum_{i=1}^n [E(X_i^2) - E(\bar{X}^2)] \\ &= \frac{1}{n} \left[\sum_{i=1}^n (\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right] \\ &= \frac{1}{n} (n\mu^2 + n\sigma^2 - \mu^2 n - \sigma^2) \\ &= \frac{n-1}{n} \sigma^2 \neq \sigma^2 \end{aligned}$$

$\therefore$ 不是无偏估计量

$\therefore \bar{X}$ 与 S^2 是 μ, σ^2 的无偏估计量。

例: 设 X_1, X_2, X_3 为来自总体 X 的样本, 则下列哪些估计量为总体 X 均值 μ 的无偏估计量。

$$\hat{\mu}_1 = \frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3$$

$$\hat{\mu}_2 = \frac{2}{5}X_1 + \frac{2}{5}X_2 + \frac{1}{5}X_3$$

$$\hat{\mu}_3 = \frac{1}{3}X_1 + \frac{2}{9}X_2 + \frac{1}{7}X_3$$

$$E(\hat{\mu}_1) = E\left(\frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{2}X_3\right) = \frac{1}{6}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{2}E(X_3) = \mu$$

$$E(\hat{\mu}_2) = E\left(\frac{2}{5}X_1 + \frac{2}{5}X_2 + \frac{1}{5}X_3\right) = \mu$$

$$E(\hat{\mu}_3) = \frac{44}{63}\mu \neq \mu$$

$\therefore \hat{\mu}_1, \hat{\mu}_2$ 为无偏估计量

2) 有效性. 当 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 均为 θ 的无偏估计时, 如果 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效.

例: 设 X_1, X_2, X_3 为总体 X 的一个样本, 则 $\hat{\mu}_1, \hat{\mu}_2$ 哪个更有效?

$$\hat{\mu}_1 = \frac{1}{6}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$$

$$\hat{\mu}_2 = \frac{2}{5}X_1 + \frac{2}{5}X_2 + \frac{1}{5}X_3$$

$$E(\hat{\mu}_1) = \frac{1}{6}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{6}E(X_3) = \mu$$

$$D(\hat{\mu}_1) = \frac{1}{36}D(X_1) + \frac{1}{9}D(X_2) + \frac{1}{36}D(X_3) = \frac{7}{18}\sigma^2$$

$$E(\hat{\mu}_2) = \frac{2}{5}E(X_1) + \frac{2}{5}E(X_2) + \frac{1}{5}E(X_3) = \mu$$

$$D(\hat{\mu}_2) = \frac{4}{25}D(X_1) + \frac{4}{25}D(X_2) + \frac{1}{25}D(X_3) = \frac{9}{25}\sigma^2$$

$\therefore \hat{\mu}_1$ 与 $\hat{\mu}_2$ 均为无偏估计量

$D(\hat{\mu}_1) < D(\hat{\mu}_2) \therefore \hat{\mu}_1$ 更有效.

3) 一致性 用于证明当样本量足够大时, $\hat{\theta}$ 趋近于 θ , 即 $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$

(一般不出题)

例: 设 $X_1, X_2, \dots, X_n$ 为来自总体 $X \sim B(1, p)$ 的样本, 记统计量 $T = \bar{X} - S^2$, 则 $E(T) = \underline{\hspace{2cm}}$. 若 $\bar{X} + KS^2$ 为 p 的无偏估计量, 则 $K = \underline{\hspace{2cm}}$.

$$E(T) = E(\bar{X}) - E(S^2) = E(X) - D(X) = np - npq = np^2$$

$$E(\bar{X} + KS^2) = E(\bar{X}) + KE(S^2) = np + Knpq = np(HK - qp)$$

∵ 无偏 ∴ $HK - qp = p \quad \therefore K = -1$

例: $X_1, X_2, \dots, X_n$ 来自总体 $X \sim N(0, \sigma^2)$ $Y_i = X_i - \bar{X}$, 则

① $\text{Cov}(Y_i, Y_n)$

② $E[(Y_i + Y_n)^2] = \sigma^2$

解 $\text{Cov}(Y_i, Y_n) = E(Y_i Y_n) - E(Y_i)E(Y_n)$

$Y_i = X_i - \bar{X} \quad E(Y_i) = 0$

$= E(Y_i Y_n)$

$E(X_i) = 0 \quad E(X_n) = 0 \quad X_i, X_n$ 独立: $E(X_i X_n) = E(X_i)E(X_n) = 0$

$= E[(X_i - \bar{X})(X_n - \bar{X})]$

$X \sim N(0, \sigma^2) \quad \bar{X} \sim N(0, \frac{\sigma^2}{n}) \quad \therefore D(\bar{X}) = \frac{\sigma^2}{n}$

$= E(X_i X_n - X_i \bar{X} - X_n \bar{X} + \bar{X}^2)$

$E(X_i^2) = D(X) + [E(X)]^2 = \sigma^2$

$= E(X_i X_n) - E(X_i \bar{X}) - E(X_n \bar{X}) + E(\bar{X}^2)$

$= -2E(X_i \bar{X}) + D(\bar{X}) + [E(\bar{X})]^2$

$= -\frac{2}{n} E(X_i \sum_{j=1}^n X_j) + \frac{\sigma^2}{n}$

$= -\frac{2}{n} E(X_i^2 + \sum_{j=1, j \neq i}^n X_i X_j) + \frac{\sigma^2}{n}$

$= -\frac{2}{n} E(X_i^2) + \frac{\sigma^2}{n} = -\frac{\sigma^2}{n}$

解 $E[(Y_i + Y_n)^2] = E[D(Y_i + Y_n) + (E(Y_i + Y_n))^2]$

$E(Y_i + Y_n) = E(Y_i) + E(Y_n) = 0$

$= E[D(Y_i + Y_n)]$

$D(Y_i) = D(Y_n) = \sigma^2$

$= E[D Y_i + D Y_n + 2\text{Cov}(Y_i, Y_n)]$

$= E\left(\frac{n-1}{n} + \frac{n-1}{n} - \frac{2}{n}\right) \sigma^2 = \frac{2(n-1)}{n} \sigma^2 = \sigma^2 \quad \therefore C = \frac{n}{2(n-1)}$

$(D(Y_i) = \frac{n-1}{n} \sigma^2)$

例: 若 $X_1, X_2, \dots, X_n \sim N(0, 1)$ 时, 求 $D(\bar{X}^2 - \frac{1}{n}S^2)$

$$\begin{aligned} D(\bar{X}^2 - \frac{1}{n}S^2) &= D(\bar{X}^2) + \frac{1}{n^2} D(S^2) \\ &= \frac{1}{n^2} D(n\bar{X}^2) + \frac{1}{n^2(n-1)^2} D((n-1)S^2) \\ &= \frac{1}{n^2} \cdot 2 + \frac{1}{n^2(n-1)^2} [2(n-1)] \\ &= \frac{2}{n^2} + \frac{2}{n^2(n-1)} \\ &= \frac{2n}{n^2(n-1)} = \frac{2}{n(n-1)} \end{aligned}$$

$$\bar{X} \sim N(0, \frac{1}{n})$$

$$\frac{\bar{X}}{\frac{1}{\sqrt{n}}} = \sqrt{n}\bar{X} \sim N(0, 1) \quad \therefore D(\bar{X}^2) = \frac{2}{n}$$

$$\frac{1}{\sqrt{n}} S^2 \sim \chi^2(n-1) \quad \therefore D(n\bar{X}^2) = 2$$

$$\frac{n-1}{2} S^2 \sim \chi^2(n-1) \quad b=1$$

$$\therefore (n-1)S^2 \sim \chi^2(n-1) \quad D((n-1)S^2) = 2(n-1) = 2n-2$$

例: 用天平对物体称重 n 次, 物体的质量 μ 未知, 设 n 次测量结果 $X_1, X_2, \dots, X_n$ 相互独立, 服从正态分布 $N(\mu, \sigma^2)$. 该样本记录的 n 次测量的绝对误差 $Z_i = |X_i - \mu|$ ($i=1, 2, \dots, n$), 利用 $Z_1, Z_2, \dots, Z_n$ 估计 σ .

① 求 Z_1 的概率密度

② 求 σ 的估计量

③ 求 σ 的极大似然估计量

① $X_1 \sim N(\mu, \sigma^2)$

$$\begin{aligned} Z \geq 0 \text{ 时 } F(Z, \sigma) &= P(|X_1 - \mu| \leq Z) = P(-Z \leq X_1 - \mu \leq Z) \\ &= P(-\frac{Z}{\sigma} \leq \frac{X_1 - \mu}{\sigma} \leq \frac{Z}{\sigma}) \\ &= 2\Phi(\frac{Z}{\sigma}) - 1 \end{aligned}$$

$$Z < 0 \text{ 时 } F_Z(Z) = 0$$

$$\therefore f_Z(Z) = \begin{cases} \frac{2}{\sigma} \cdot \varphi(\frac{Z}{\sigma}) & (Z \geq 0) \\ 0 & (Z < 0) \end{cases}$$

$$\textcircled{2} L(\theta) = \prod_{i=1}^n f(Z_i, \theta) = \begin{cases} \frac{2^n}{\sigma^n} \prod_{i=1}^n \varphi(\frac{Z_i}{\sigma}) & \\ 0 & \end{cases} = \frac{2^n}{\sigma^n} \left(\frac{1}{\sqrt{\pi}}\right)^n e^{-\frac{\sum_{i=1}^n Z_i^2}{2\sigma^2}}$$

$$\ln L(\theta) = n \ln 2 - n \ln \sigma - \frac{n}{2} (\ln \pi) - \frac{\sum_{i=1}^n Z_i^2}{2\sigma^2}$$

$$\frac{d \ln L(\theta)}{d \sigma} = -n \frac{1}{\sigma} - \frac{\sum_{i=1}^n Z_i^2}{\sigma^3} = 0 \quad \therefore n\sigma^2 = \sum_{i=1}^n Z_i^2$$

$$\therefore \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}$$

② $E(Z) = \int_{-\infty}^{+\infty} z f_Z(z) dz$

$$= \int_0^{+\infty} \frac{2z}{\sigma} \varphi(\frac{z}{\sigma}) dz$$

$$= 2\sigma \int_0^{+\infty} \frac{z}{\sigma} \varphi(\frac{z}{\sigma}) d(\frac{z}{\sigma})$$

$$= 2\sigma \int_0^{+\infty} t \varphi(t) dt$$

$$= 2\sigma \int_0^{+\infty} \frac{t}{\sqrt{\pi}} e^{-\frac{t^2}{2}} dt$$

$$= \frac{2\sigma}{\sqrt{\pi}} \int_0^{+\infty} e^{-\frac{t^2}{2}} d(\frac{t^2}{2})$$

$$= \sqrt{\frac{2}{\pi}} \sigma$$

例: 设总体 X 的概率密度为 $f(x; \theta) = \theta x^{\theta-1}$ ($0 < x < 1$), 其中未知参数 $\theta > 0$, $X_1, X_2, \dots, X_n$ 是来自 X 的简单随机样本, 试求:

(1) θ 的矩估计量 $\hat{\theta}_1$.

(2) θ 的最大似然估计量 $\hat{\theta}_2$.

$$\textcircled{1} E(X) = \int_0^1 x \theta x^{\theta-1} dx = \int_0^1 \theta x^{\theta} dx = \frac{\theta}{\theta+1}$$

$$E(X) = \bar{X} = \frac{\theta}{\theta+1}$$

$$\therefore \theta \text{ 的矩估计量 } \hat{\theta} = \frac{\bar{X}}{1-\bar{X}}$$

$$\textcircled{2} L(\theta) = \prod_{i=1}^n f(x_i) = \int_0^1 \theta (x_1 x_2 \dots x_n)^{\theta-1} \quad 0 < x_1, x_2, \dots, x_n < 1$$

且

$$\ln L(\theta) = n \ln \theta + (\theta-1) \sum_{i=1}^n \ln x_i$$

$$\frac{d \ln L(\theta)}{d \theta} = \frac{n}{\theta} + \sum_{i=1}^n \ln x_i = 0$$

$$\therefore \hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln x_i}$$

第八讲 参数的区间估计与假设检验

① 从总体 X 中取得样本观测值 $\rightarrow$ 点估计或参数的估计值但仅靠点估计无法判断日与日误差大小 $\rightarrow$ 区间估计② 置信度: 即可信度, 置信度为 $1-\alpha$, 一般约定置信度为95%, 即 $\alpha=0.05$, $1-\alpha=0.95$ (区间估计研究的总体 X 正态总体分布, 即正态分布, χ^2 分布, t 分布, F 分布之一)③ 置信区间: 设总体 X 的分布函数为 $F(x; \theta)$, 其中 θ 为未知参数, $X_1, X_2, \dots, X_n$ 为来自总体 X 的简单随机样本对于给定的 $\alpha \in (0, 1)$, 如果由样本确定的统计量 T_1, T_2 满足 $P(T_1 \leq \theta \leq T_2) = 1-\alpha$, 则称随机区间 $[T_1, T_2]$ 是参数 θ 的置信度为 $1-\alpha$ 的置信区间。若统计量 T_1 满足 $P(T_1 \leq \theta) = 1-\alpha$, 则 T_1 为单侧置信下限 $P(\theta \leq T_2) = 1-\alpha$, 则 T_2 为单侧置信上限

④ 求置信区间的方法

① 仅考虑总体参数的区间估计

三种类

① 未知 μ , 已知 σ^2 , 求 μ 的置信区间 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ ② 未知 μ , 未知 σ^2 , 求 μ 的置信区间 $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ ③ 未知 σ^2 , 求 σ^2 的置信区间 $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$

② 选用合适的四大分布

(X 55 是样本的
n 56 是总体的)③ 满足 $P(T_1 \leq Y < T_2) = 1-\alpha$ ④ 解出 θ $P(T_1 \leq \theta \leq T_2) = 1-\alpha$ ⑤ $[T_1, T_2]$ 即为 θ 的置信度为 $1-\alpha$ 的置信区间例: 某地区磁铁矿强度 $X \sim N(\mu, 20)$, 现从该地区抽36点, 测得样本均值 $\bar{X}=61.6$, 在0.95置信度, 求该地磁铁矿强度的置信区间。① μ 未知, σ^2 已知, 求 μ 的置信区间 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = P\left(61.6 - 1.96 \times \frac{\sqrt{20}}{6} < \mu < 61.6 + 1.96 \times \frac{\sqrt{20}}{6}\right)$$
$$= P(54.6 < \mu < 67.6)$$

$$z_{0.025} = 1.96$$

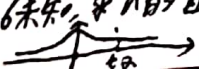
$$z_{0.975} = -z_{0.025} = -1.96$$

 $\therefore$ 磁铁矿强度 μ 的置信区间 $[54.6, 67.6]$

例: 已知灯泡寿命服从正态分布, 现从一批灯泡中随机抽取15只作为样本, 测得其平均寿命为1476h, 样本标准差为25.4h,

则在0.95置信度下, 求灯泡平均使用寿命 μ 的置信区间。① μ 未知, σ^2 未知, 求 μ 的置信区间 $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

$$P\left(-t_{\frac{\alpha}{2}}(n-1) < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}}(n-1)\right) = 1-\alpha \quad \therefore \mu \text{ 的置信区间}$$
$$P\left(\bar{X} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = 0.95 \quad [1476.46, 1502.54]$$



$$t_{0.025}(15) = 2.1315$$
$$t_{0.975}(15) = -2.1315$$

例: 设车间甲的滚珠直径 $X \sim N(\mu, \sigma^2)$, 现从某日生产的滚珠中抽取 25, 测得样本方差 $s^2 = 0.25$, 则在 0.95 置信度下求总体方差 σ^2 的置信区间。

① μ 未知, σ^2 未知, 求 σ^2 的置信区间 $\bar{X} = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$

$$P(\chi_{1-\frac{\alpha}{2}}^2(n-1) \leq \frac{n-1}{\sigma^2} S^2 \leq \chi_{\frac{\alpha}{2}}^2(n-1)) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2(n-1)}\right) = 1-\alpha$$

$$n=9 \quad S^2=0.25 \quad \chi_{0.025}^2(8)=17.535 \quad \chi_{0.975}^2(8)=2.180$$

$\therefore \sigma^2$ 的置信度为 0.95 的置信区间 $[0.03, 0.23]$

(又作假设检验中假设检验步骤是若检验 X 的 $\mu_1 > Y$ 的 μ_2 , 则检验假设 $H_0: \mu_1 \leq \mu_2$ (B), $H_1: \mu_1 > \mu_2$ (B))

⑤ 假设检验: 当总体分布未知时, 可通过区间估计得到置信区间 (原理: 小概率事件不发生)

当总体分布未知时, 可根据已有经验或专业知识假设一个分布, 再通过假设检验判断是否正确。

例: 一台自动包装机包装葡萄糖, 额定标准重量 0.5kg, 设包装机包装每袋重量 $X \sim N(\mu, 0.015^2)$, 每箱 10 袋。

称得重量分别为 0.497, 0.506, 0.518, 0.524, 0.498, 0.511, 0.520, 0.513, 0.512 判断包装机是否工作正常?

已知总体方差 $\sigma^2 = 0.015^2$, μ 未知, 假设包装机工作正常, 即 $\mu = \mu_0 = 0.5$, 记为原假设 H_0 。

“假设英文”Hypothesis

假设包装机工作不正常, 即 $\mu \neq 0.5$, 记为备择假设 H_1 。

若 H_0 成立, 则样本均值 $\bar{X}$ 与 μ 偏差不会太大。设 α 为一个小的正数, 通常取 0.05, 称为显著性水平。

$$\text{则 } P(|\bar{X} - \mu| > \alpha) = \alpha \quad \text{即 } P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| > Z_{\frac{\alpha}{2}}\right) = \alpha$$

$$\bar{X} = 0.511 \quad \sigma = 0.015 \quad n = 9 \quad Z_{0.025} = 1.96 \quad \mu_0 = 0.5 \quad (\text{只需判断 } \left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| \text{ 与 } Z_{\frac{\alpha}{2}} \text{ 大小即可})$$

$$\therefore \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = 2.2 > 1.96 \quad \therefore \text{拒绝 } H_0, \text{ 接受 } H_1, \text{ 包装机工作不正常。}$$

认为这个概率 α 为小概率, 在这个事件下不发生。

例: 根据过去经验, 某商品使用寿命符合 $X \sim N(1020, 100^2)$, 现随机抽 16 件, 测得寿命分别为 1080, 1040, 1060, 1030, 1070, 1050, 1020, 1010, 1090, 1000, 1060, 1040, 1070, 1030, 1050, 1020, 则在 0.05 的显著性水平下, 这批产品的使用寿命是否有显著提高?

假设使用寿命没有显著提高, 即 $\mu \leq 1020$, 记为 H_0 , 假设使用寿命显著提高, 即 $\mu > 1020$, 记为 H_1 。

$$\text{已知方差 } \sigma^2 = 100, \text{ 未知 } \mu, \therefore Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{得 } P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > Z_{\frac{\alpha}{2}}\right) = \alpha$$

$$\bar{X} = 1080 \quad \sigma = 100 \quad n = 16 \quad Z_{0.05} = 1.645 \quad \mu_0 = 1020$$

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1080 - 1020}{100/\sqrt{16}} = 2.4 > 1.645 = Z_{0.05}$$

$\therefore$ 拒绝 H_0 , 接受 H_1 , 在 0.05 显著性水平下, 寿命显著提高。

例: 某手机生产商声称手机待机平均时间至少为71.5h, 抽查其中一部, 得到待机时间数据为
69, 68, 72, 70, 66, 75, 假设手机待机时间 $X \sim N(\mu, \sigma^2)$, 则在显著性水平0.05下, 商家是否欺骗消费者?

未知方差 σ^2 , 未知 μ , $\therefore$ 用 $t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$ 的抽样分布。

假设未欺骗消费者, 即 $\mu \geq 71.5$ 记为 H_0

假设其欺骗消费者, 即 $\mu < 71.5$ 记为 H_1

$$\bar{x} = 70 \quad \mu_0 = 71.5 \quad s^2 = \frac{1}{5} \sum_{i=1}^6 (x_i - \bar{x})^2 \quad s = 3.16 \quad t_{0.05}(5) = 2.015$$

$$P(t < t_{-\alpha/2}(n-1)) = \alpha$$

$$t = \frac{70 - 71.5}{3.16/\sqrt{6}} = -1.16 > -t_{0.05}(5) = -2.015$$

$\therefore$ 接受 H_0 , 拒绝 H_1 , 在0.05显著性水平下, 商家没有欺骗消费者。

PS: 双侧检验与单侧检验

① 令 $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ 的假设检验为双侧检验

统计量 当未知 μ , 已知 σ^2 , 判断 μ 时

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad P(|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}| > Z_{\frac{\alpha}{2}}) = \alpha$$

即若 $|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}| > Z_{\frac{\alpha}{2}}$ 则 H_0 不成立, H_1 成立

若 $|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}| < Z_{\frac{\alpha}{2}}$ 则 H_0 成立, H_1 不成立。

当未知 μ , 未知 σ^2 , 判断 μ 时

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \quad P(|\frac{\bar{X} - \mu}{S/\sqrt{n}}| > t_{\frac{\alpha}{2}}(n-1)) = \alpha$$

即若 $|\frac{\bar{X} - \mu}{S/\sqrt{n}}| > t_{\frac{\alpha}{2}}(n-1)$ 则 H_0 不成立, H_1 成立

若 $|\frac{\bar{X} - \mu}{S/\sqrt{n}}| < t_{\frac{\alpha}{2}}(n-1)$ 则 H_0 成立, H_1 不成立。

当未知 μ , 未知 σ^2 , 判断 σ^2 时

$$\chi^2 = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

$$P((\frac{n-1}{\sigma^2} S^2 > \chi^2_{\frac{\alpha}{2}}(n-1)) \cup (\frac{n-1}{\sigma^2} S^2 < \chi^2_{1-\frac{\alpha}{2}}(n-1))) = \alpha$$

即 $\frac{n-1}{\sigma^2} S^2 > \chi^2_{\frac{\alpha}{2}}(n-1)$ 或 $\frac{n-1}{\sigma^2} S^2 < \chi^2_{1-\frac{\alpha}{2}}(n-1)$ 时 H_0 不成立, H_1 成立。

② 令 $H_0: \mu = \mu_0$ (或者 $\mu \leq \mu_0$) $H_1: \mu > \mu_0$ 的假设检验为右侧检验

令 $H_0: \mu = \mu_0$ (或者 $\mu \geq \mu_0$) $H_1: \mu < \mu_0$ 的假设检验为左侧检验

统计量 当未知 μ , 已知 σ^2 , 判断 μ 时

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > Z_{\alpha}) = \alpha \text{ 右侧}$$

$$P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < -Z_{\alpha}) = \alpha \text{ 左侧}$$

例: 某车间滚珠直径 $X \sim N(\mu, \sigma^2)$, 抽 11 件, 测得样本方差 $s^2 = 0.25$, 则在 $\alpha = 0.05$ 下是否认为总体方差 $\sigma^2 = 0.36$?

$$H_0: \sigma^2 = \sigma_0^2 = 0.36$$

$$H_1: \sigma^2 \neq 0.36$$

$$\chi^2 = \frac{n-1}{\sigma^2} s^2 \sim \chi^2(n-1)$$

$$P(\chi^2 > \chi_{0.025}^2(8) \cup \chi^2 < \chi_{0.975}^2(8)) = \alpha \quad | \quad P(\chi_{0.975}^2(8) < \chi^2 < \chi_{0.025}^2(8)) = 1 - \alpha$$

$$\chi^2 = \frac{n-1}{\sigma^2} s^2 = 3.858 \quad \chi_{0.025}^2(8) = 17.535 \quad \chi_{0.975}^2(8) = 2.180$$

$\therefore \chi^2$ 未落在拒绝域内: 在显著性水平 $\alpha = 0.05$ 下认为总体方差 $\sigma^2 = 0.36$.

PS: 两类错误 第一类错误 (弃真) $P(\text{拒绝 } H_0 | H_0 \text{ 为真}) = \alpha$ 称为显著性水平

第二类错误 (取伪) $P(\text{接受 } H_0 | H_0 \text{ 为假}) = \beta$

α, β 不可能同时减小, 实际应用为在一定的显著性水平 α 下, 尽量减小 β 损失.

用两类错误求 α, β . 例: 设总体 X 的概率密度只有两种可能, 设

$$H_0: f(x) = \begin{cases} \frac{1}{2} & 0 \leq x \leq 2 \\ 0 & \text{其它} \end{cases} \quad H_1: f(x) = \begin{cases} \frac{x}{2} & 0 \leq x \leq 2 \\ 0 & \text{其它} \end{cases}$$

对于 X 进行一次观测, 得样本 X , 若观测 X 以 α 拒绝 H_0 , 否则就接受 H_0 .

则此检验的 α, β 分别为

$$\alpha = P(\text{拒绝 } H_0 | H_0 \text{ 为真}) = \int_{\frac{1}{2}}^2 \frac{1}{2} dx = \frac{1}{4}$$

$$\beta = P(\text{接受 } H_0 | H_0 \text{ 为假}) = \int_0^{\frac{1}{2}} \frac{x}{2} dx = \frac{9}{16}$$

例: 设 $X_1, X_2, \dots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的简单样本, n, σ^2 未知, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $Q^2 = \sum_{i=1}^n (X_i - \bar{X})^2$, 用何检验 $H_0: \mu = 0$ 的检验, 则使用统计量 $T =$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X}}{S/\sqrt{n}} \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} = \sqrt{\frac{1}{n-1} Q^2}$$

$$\therefore T = \frac{\bar{X} \sqrt{n(n-1)}}{Q}$$

例: 考试成绩正态分布, 抽了36人, 平均成绩66.5, 标准差15, 在显著性水平0.05下, 是否可认为平均成绩为70分

$$H_0: \text{可认为70分} \quad \mu = \mu_0 = 70$$

$$H_1: \text{不可认为} \quad \mu \neq 70$$

$$\mu \text{已知 } \sigma^2 \text{未知} \therefore \text{选用统计量} \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

$$P\left(\left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| \geq t_{0.025}(35)\right) = \alpha$$

$$\bar{X} = 66.5 \quad n = 36 \quad \mu = 70 \quad S = 15$$

$$\therefore \left|\frac{66.5 - 70}{15/\sqrt{36}}\right| = 1.4 \quad t_{0.025}(35) = 2.0301 \quad \text{拒绝域为 } T = \left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| \geq t_{0.025}(35)$$

$$\therefore T = \left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| < t_{0.025}(35) \quad \therefore T \text{未落在拒绝域内}$$

$\therefore$ 接受 H_0 , 拒绝 H_1 , 在显著性水平0.05下可认为平均成绩70分

例: 设 $X_1, X_2, \dots, X_{16}$ 为来自总体 $N(\mu, \sigma^2)$ 的样本, 样本均值 $\bar{X}$ 呈正态分布 $\sigma^2 = 0.05$

检验假设 $H_0: \mu \geq 5, H_1: \mu < 5$ 的拒绝域为 _____ $\Phi(1.645) = 0.95$

$$H_0: \mu \geq 5$$

$$H_1: \mu < 5$$

$$\mu \text{未知, } \sigma^2 \text{已知, 选用统计量} \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$\text{左侧检验} \quad P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{0.95}\right) = \alpha$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 5}{\sqrt{0.05}/\sqrt{16}}$$

$$z_{0.05} = 1.645 \quad z_{0.95} = -1.645$$

$$\therefore \text{拒绝域 } \frac{\bar{X} - 5}{\sqrt{0.05}/\sqrt{16}} < -1.645$$

$$\bar{X} \leq 4.177$$