

清华大学试题专用纸

考试课程：概率论与数理统计 考试时间：2019 年 6 月 23 日 19:00—21:00

姓名_____学号 **20** 班级_____.

一 填空题（每空 3 分，共 30 分；答案均写在试卷上，注意标清题号）

1. 设 X 服从几何分布， $P(X=3|X>2)=0.4$ ，则 $P(X=3)=$ _____， $E(X|2<X<5)=$ _____。

$$P(X=3|X>2)=0.4 \Rightarrow p=0.4, \quad P(X=3)=0.6^2 \cdot 0.4=0.144,$$

$$E(X|2<X<5)=\frac{0.6^2 \cdot 0.4 \cdot 3 + 0.6^3 \cdot 0.4 \cdot 4}{0.6^2 \cdot 0.4 + 0.6^3 \cdot 0.4} = \frac{3 + 2.4}{1 + 0.6} = \frac{27}{8}$$

2. 二维随机变量 $(X, Y) \sim N(0, 0, 1, 4, 0)$ ，则 $E(X^2|2X+Y=2)=$ _____。

$$\text{Cov}(2X+Y, 2X-Y)=0 \Rightarrow E(X^2|2X+Y=2)=E\left(\frac{[(2X+Y)+(2X-Y)]^2}{16} \middle| 2X+Y=2\right)$$

$$=\frac{1}{16}E((2X+Y)^2+2(2X+Y)(2X-Y)+(2X-Y)^2|2X+Y=2)=\frac{4+0+8}{16}=\frac{3}{4}$$

3. 随机变量 X 服从二项分布 $b(100, 0.6)$ ， c 为任意实数，则 $E((X-c)^2)$ 的最小值为_____。

$$\min E((X-c)^2) = \text{Var}(X) = npq = 24$$

4. 设随机变量 $X \sim N(0, 0.25)$ ，则随机变量 $Y=|X|$ 的概率密度函数 $p_Y(y)=$ _____。

$$\frac{4}{\sqrt{2\pi}} e^{-2y^2} I_{y>0}$$

5. 利用切比雪夫不等式，估计一个有方差的随机变量落在与其期望左右不超过 3 个标准差的概率至少为_____。

$$E\xi = \mu, \quad \text{Var}(\xi) = \sigma^2, \quad P(|\xi - \mu| \leq 3\sigma) \geq 1 - \frac{\sigma^2}{(3\sigma)^2} = \frac{8}{9}$$

6. 总体分布服从 $N(\mu, \sigma^2)$ ，样本容量为 n ，写出方差 σ^2 的 0.95 置信区间表达式_____。

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1), \Rightarrow \sigma^2 \text{ 的 } 1-\alpha \text{ 置信区间为:}$$

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right] = \left[\frac{(n-1)S^2}{\chi^2_{0.975}(n-1)}, \frac{(n-1)S^2}{\chi^2_{0.025}(n-1)} \right]$$

7. 对正态总体 $N(\mu, 1)$ 的参数 μ 做假设检验 $H_0: \mu=10, H_1: \mu>10$, 显著性水平 $\alpha=0.1$, 样本容量 $n=16$, 检验

统计量为样本均值 \bar{X} , 则拒绝域为_____, 现得到 \bar{X} 的观测值为 11, 则其对应的 p 值 = _____。

$$u = \frac{\bar{x} - \mu}{\frac{1}{\sqrt{n}}} \sim N(0, 1) \Rightarrow \mu \text{ 的拒绝域为: } \{\bar{x}: \bar{x} > 10.32\};$$

$$u_0 = \frac{11 - 10}{\frac{1}{\sqrt{16}}} = 4; p\text{值} = 1 - \Phi(u_0) = 1 - \Phi(4)$$

8. 已知期望为 $1/\lambda$ 的指数分布随机变量的方差为 $1/\lambda^2$, 利用此结果计算积分 $\int_0^{+\infty} 7x^2 e^{-5x} dx =$ _____。

$$\text{原式} = \frac{7}{5} \int_0^{\infty} x^2 \cdot 5e^{-5x} dx = \frac{7}{5} EX^2 = \frac{7}{5} (\text{Var}(X) + (EX)^2) = \frac{14}{125}$$

二. (16 分) 一份考卷共有 20 题, 均为 4 个选项的选择题, 每题 5 分。假设对于参加考试的某考生, 考卷中 60% 的题目涉及的知识已基本掌握, 此时答对的概率是 0.8; 其他题目则在 4 个选项中随机地任选一个, 试计算

- (1) 任选一道考题, 该考生回答正确的概率;
- (2) 对于某一道该考生答对的题目, 计算该考题涉及的知识是这名学生已基本掌握的概率;
- (3) 该考生考试成绩的期望值;
- (4) 该考生成绩在 70 分以上的概率 (利用中心极限定理估计)。

解: (1)、设任选一道题为已基本掌握的题目为事件 A; 考生答对该题为事件 B, 则:

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 0.8 * (12/20) + 1/4 * (8/20) = 0.58$$

$$(2)、P(A|B) = \frac{P(A)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} = \frac{24}{29},$$

(3)、由第 (1) 问知: 考生答对题数目 $X \sim b(20, 0.58)$, 所以 $EX = 20 * 0.58$, 知,

成绩的期望 = $5 * EX = 5 * 20 * 0.58 = 58$ 。

(4)、成绩在 70 以上, 也即答对题目数目在 14 题以上,

$$\text{Var}(X) = 20 * 0.58 * (1 - 0.58) = 4.872; \quad \frac{X - EX}{\sqrt{\text{Var}(X)}} \sim N(0, 1);$$

$$P(X > 14) = P(X - EX \geq 14 - 11.6) = P\left(\frac{X - EX}{\sqrt{\text{Var}(X)}} \geq 2.4 / \sqrt{\text{Var}(X)}\right) = 1 - \Phi\left(\frac{2.4}{\sqrt{4.872}}\right)$$

三. (8 分) 随机变量 $X \sim N(0,1)$, $Y \sim \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$,

(1) 求随机变量 $W = XY$ 的分布函数; (2) 判断 X 与 W 是否独立, 并说明理由。

解: (1) 利用全概率公式:

$$\begin{aligned} F_W(w) &= P(W \leq w) = P(W \leq w | Y = 1) \cdot P(Y = 1) + P(W \leq w | Y = -1) \cdot P(Y = -1) \\ &= P(X \leq w) \cdot \frac{1}{2} + P(X \geq -w) \cdot \frac{1}{2} = \Phi(w) \end{aligned}$$

(2) 不独立。

$$E(XW) = E(XW | Y = 1) \cdot P(Y = 1) + E(XW | Y = -1) \cdot P(Y = -1) = E(X^2) \cdot \frac{1}{2} + E(-X^2) \cdot \frac{1}{2} = 0$$

举一反例: 因为 $\rho_{XW}(1,0) = 0, \rho_X(1) > 0, \rho_W(0) > 0$, 所以 X 与 W 不独立。获取定 $P\left(X \in \left(0, \frac{1}{2}\right), W \in (0, 2)\right) = 0$, 但是 $P\left(X \in \left(0, \frac{1}{2}\right)\right) > 0, P(Y \in (0, 2) > 0)$, 所以不独立

四. (12 分) 随机变量 $X_1 \sim U(0,1)$, $X_2 \sim U(0,2)$, X_1 和 X_2 相互独立, $Y_1 = X_1 - 2X_2$, $Y_2 = X_1 + 2X_2$ 。

(1) 求 Y_1 的密度函数; (2) 计算协方差 $Cov(Y_1, Y_2)$ 。

解: (1) 当 $0 < Y \leq 1$ 时, $F(y) = P(Y \leq y) = P(2X_1 + X_2 \leq y) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{y}{2} \cdot y = \frac{y^2}{8}$

当 $1 < Y \leq 4$ 时, $F(y) = P(2X_1 + X_2 \leq Y) = \frac{1}{8} + \frac{1}{2} \left(\frac{y}{2} - \frac{1}{2} \right) = \frac{y}{4} - \frac{1}{8}$

当 $4 < Y \leq 5$ 时, $F(y) = P(2X_1 + X_2 \leq y) = 1 - \frac{1}{2} \cdot \left(2 - \frac{y-1}{2} \right)^2 = 1 - \frac{(5-y)^2}{8}$

$$F(y) = \begin{cases} 0, & y \leq 0 \\ \frac{y^2}{8}, & 0 < y \leq 1 \\ \frac{y}{4} - \frac{1}{8}, & 1 < y \leq 4 \\ 1 - \frac{(5-y)^2}{8}, & 4 < y \leq 5 \\ 1, & y > 5 \end{cases}$$

$$(2) E(X_1)=1, \quad Var(X_1)=\frac{1}{3}, \quad E(X_2)=\frac{1}{2}, \quad Var(X_2)=\frac{1}{12}$$

$$(3) Cov(X_1, Y) = Cov(X_1, 2X_1 + X_2) = 2Var(X_1) = \frac{2}{3}$$

$$Var(Y) = Var(2X_1 + X_2) = 4Var(X_1) + Var(X_2) = \frac{17}{12} \quad Corr(X_1, Y) = \frac{Cov(X_1, Y)}{\sqrt{Var(X_1)Var(Y)}} = \frac{4}{\sqrt{17}}.$$

五. (10 分) X_1, \dots, X_n 为参数 λ 的泊松分布总体的一个样本, 定义 $S_n = \sum_{k=1}^n X_k$ 及 $\theta_0 = \begin{cases} 1, & X_1 = 0 \\ 0, & \text{其他} \end{cases}$ 。

(1) 求 $E(\theta_0)$; (2) 求 $E(\theta_0 | S_n)$ 。

解: (1) $E(\theta_0) = P(X_1 = 0) = e^{-\lambda}$;

$$(2) S_n = \sum_{k=1}^n X_k \sim P(n\lambda), \quad \sum_{k=2}^n X_k \sim P((n-1)\lambda)$$

$$E(\theta_0 | S_n = k) = P(X_1 = 0 | S_n = k) = \frac{P(X_1 = 0, S_n = k)}{P(S_n = k)} = \frac{P(X_1 = 0)P(X_2 + \dots + X_n = k)}{P(S_n = k)}$$

$$= \frac{e^{-\lambda} e^{-(n-1)\lambda} \frac{((n-1)\lambda)^k}{k!}}{e^{-n\lambda} \frac{(n\lambda)^k}{k!}} = \left(1 - \frac{1}{n}\right)^k$$

$$E(\theta_0 | S_n) = \left(1 - \frac{1}{n}\right)^{S_n}.$$

六. (16 分) 设总体 $X \sim U(0, \theta)$, X_1, X_2, \dots, X_n 是来自 X 的简单随机样本, $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ 为样本均

值, $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$ 为样本方差, $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ 为 X_1, X_2, \dots, X_n 的次序统计量

(1) 求 $X_{(1)}$ 和 $X_{(n)}$ 各自的分布函数、期望和方差, 以及 $X_{(1)}$ 和 $X_{(n)}$ 的协方差与相关系数;

(2) 判断并解释 $X_{(1)} + X_{(n)}$ 、 $X_{(n)}$ 、 $2\bar{X}$ 与 $2\sqrt{3}S$ 是否为参数 θ 的无偏估计量, 如果不是, 是否可以做无偏矫正。

解: (1) 先求 $x_{(1)}$ 和 $x_{(n)}$ 的分布函数, 当 $0 < X < \theta$ 时

$$F_{x_{(1)}}(x) = P(x_{(1)} \leq x) = 1 - P(x_{(1)} > x) = 1 - P(\min(x_1, x_2, \dots, x_n) > x) = 1 - \prod_{k=1}^n P(x_k > x) = 1 - \left(1 - \frac{x}{\theta}\right)^n,$$

$$F_{x_{(n)}}(x) = P(x_{(n)} \leq x) = P(\max(x_1, x_2, \dots, x_n) \leq x) = \prod_{k=1}^n P(x_k \leq x) = \left(\frac{x}{\theta}\right)^n,$$

$$F_{x_{(1)}}(x) = \begin{cases} 0, & x \leq 0 \\ 1 - \left(1 - \frac{x}{\theta}\right)^n, & 0 < x \leq \theta \\ 1, & x > \theta \end{cases}, \quad F_{x_{(n)}}(x) = \begin{cases} 0, & x \leq 0 \\ \left(\frac{x}{\theta}\right)^n, & 0 < x \leq \theta \\ 1, & x > \theta \end{cases}.$$

$$p_{x_{(1)}}(x) = \begin{cases} \frac{n}{\theta} \left(1 - \frac{x}{\theta}\right)^{n-1}, & 0 \leq x \leq \theta \\ 0, & \text{其他} \end{cases}, \quad p_{x_{(n)}}(x) = \begin{cases} \frac{n}{\theta} \left(\frac{x}{\theta}\right)^{n-1}, & 0 \leq x \leq \theta \\ 0, & \text{其他} \end{cases}$$

$$E(x_{(1)}) = \int_{-\infty}^{+\infty} x p_{x_{(1)}}(x) dx = \int_0^{\theta} x \cdot \frac{n}{\theta} \left(1 - \frac{x}{\theta}\right)^{n-1} dx = \int_0^{\theta} \frac{nx}{\theta^n} (\theta - x)^{n-1} dx = \int_{\theta}^0 \frac{n(\theta - y)}{\theta^n} y^{n-1} dy = \frac{\theta}{n+1}$$

$$E(x_{(n)}) = \int_{-\infty}^{+\infty} x p_{x_{(n)}}(x) dx = \int_0^{\theta} x \cdot \frac{n}{\theta} \left(\frac{x}{\theta}\right)^{n-1} dx = \int_0^{\theta} \frac{n}{\theta^n} x^n dx = \frac{n}{n+1} \theta$$

$$\begin{aligned} E(x_{(1)}^2) &= \int_{-\infty}^{+\infty} x^2 p_{x_{(1)}}(x) dx = \int_0^{\theta} x^2 \cdot \frac{n}{\theta} \left(1 - \frac{x}{\theta}\right)^{n-1} dx \\ &= \int_0^{\theta} \frac{nx^2}{\theta^n} (\theta - x)^{n-1} dx = \int_0^{\theta} \frac{n(\theta - y)^2}{\theta^n} y^{n-1} dy = \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

$$E(x_{(n)}^2) = \int_{-\infty}^{+\infty} x^2 p_{x_{(n)}}(x) dx = \int_0^{\theta} x^2 \cdot \frac{n}{\theta} \left(\frac{x}{\theta}\right)^{n-1} dx = \int_0^{\theta} \frac{n}{\theta^n} x^{n+1} dx = \frac{n}{n+2} \theta^2$$

$$Var(x_{(1)}) = E(x_{(1)}^2) - E(x_{(1)})^2 = \frac{2\theta^2}{(n+1)(n+2)} - \left(\frac{\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$$

$$Var(x_{(n)}) = E(x_{(n)}^2) - E(x_{(n)})^2 = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$$

计算 $E(x_{(1)} \cdot x_{(n)})$

当 $0 \leq u \leq v \leq \theta$ 时,

$$\begin{aligned} F_{x_{(1)}, x_{(n)}}(u, v) &= P(x_{(1)} \leq u, x_{(n)} \leq v) = P(x_{(n)} \leq v) - P(x_{(1)} > u, x_{(n)} \leq v) \\ &= \left(\frac{v}{\theta}\right)^n - P(u < x_1, x_2, \dots, x_n \leq v) = \left(\frac{v}{\theta}\right)^n - \left(\frac{v-u}{\theta}\right)^n \end{aligned}$$

$$\text{当 } 0 \leq u \leq v \leq \theta \text{ 时, } \frac{\partial^2 F_{x_{(1)}, x_{(n)}}(u, v)}{\partial u \partial v} = \frac{n \cdot (n-1) \cdot (v-u)^{n-2}}{\theta^n}$$

$$p_{x_{(1)}, x_{(n)}}(u, v) = \begin{cases} \frac{n \cdot (n-1) \cdot (v-u)^{n-2}}{\theta^n}, & 0 \leq u \leq v \leq \theta \\ 0, & \text{其他} \end{cases}$$

$$E(x_{(1)} \cdot x_{(n)}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} uv p_{x_{(1)}, x_{(n)}}(u, v) du dv = \frac{1}{\theta^n} \int_0^\theta du \int_u^\theta uv n \cdot (n-1) \cdot (v-u)^{n-2} dv$$

$$= \frac{1}{\theta^n} \int_0^\theta n \cdot u \cdot du \int_u^\theta v \cdot d(v-u)^{n-1} = \frac{1}{\theta^n} \int_0^\theta n \cdot u \cdot du \int_u^\theta v \cdot d(v-u)^{n-1} dv$$

$$= \left(\theta + \frac{1}{2}\right) \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} (-u) \cdot d\left(\theta + \frac{1}{2} - u\right)^n + \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} u \cdot d\frac{\left(\theta + \frac{1}{2} - u\right)^{n+1}}{n+1}$$

$$= \left(\theta + \frac{1}{2}\right) \left(\theta - \frac{1}{2} + \frac{1}{n+1}\right) - \frac{\theta - \frac{1}{2}}{n+1} - \frac{1}{(n+1)(n+2)}$$

七. (8分) 总体 $X \sim N(\mu, 2^2)$, 做假设检验 $H_0: \mu = 10, H_1: \mu = 11$, 显著性水平 $\alpha = 0.05$, 求当样本容量至少达到多少, 可使第二类错误不超过 0.01。

$$\text{解: } \left\{ \bar{x} : \frac{\bar{x} - 10}{2/\sqrt{n}} > u_{1-\alpha} \right\}, \text{ 即 } W = \left\{ \bar{x} > 10 + \frac{2}{n} u_{0.95} \right\}$$

$$\text{第二类错误概率 } \beta = P\left(\bar{x} \leq 10 + \frac{2}{n} u_{0.95} \mid \mu = 11\right)$$

$$P\left(\bar{x} \leq 10 + \frac{2}{\sqrt{n}} 1.65 \mid \mu = 11\right) \stackrel{\bar{x} \sim N\left(11, \frac{2^2}{n}\right)}{=} P\left(\frac{\bar{x} - 11}{2/\sqrt{n}} \leq \frac{10 + \frac{3.3}{\sqrt{n}} - 11}{2/\sqrt{n}}\right) < 0.01,$$

$$\text{则要求 } \frac{10 + \frac{3.3}{\sqrt{n}} - 11}{2/\sqrt{n}} < -2.33, \text{ 解得 } \sqrt{n} > 7.96 \Rightarrow n \geq 64$$

备注 1. 本考卷的样本均为简单随机样本, 样本均值 $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$, 样本方差为 $S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$

备注 2. 正态总体的样本均值和样本方差相互独立, 且 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, 其中 n 为样本容量

备注 3. 解答中标准正态随机变量的分布函数和密度函数分别可用 $\Phi(x)$ 和 $\varphi(x)$ 表示

备注 4. $\Phi(1.28)=0.9$, $\Phi(1.44)=0.925$, $\Phi(1.65)=0.95$, $\Phi(1.96)=0.975$, $\Phi(2.33)=0.99$

备注 5. 正态、 χ^2 、 t 等分布所需取值, 均用 (下侧) 分位数表示, 例如 $X \sim \chi^2(n)$, 则 $P(X < \chi_\alpha^2(n)) = \alpha$

备注 6. $t_{0.75}(1)=1, t_{0.75}(2)=0.79, t_{0.8}(1)=1.38, t_{0.8}(2)=1.06, F_{0.5}(1,1)=1, F_{0.5}(1,2)=0.67, F_{0.75}(1,1)=5.83$