

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：2021 年 1 月 10 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. 线上内容中提到过的贝特朗奇论 (Bertrand' s paradox) 中，不同理解下得到的概率值分别是_____。

$$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}$$

2. (X, Y) 的联合密度函数 $p(x, y) = e^{-x-y} I_{x>0, y>0}$, $Z = \begin{cases} Y, & \text{若 } X \geq Y \\ 2X, & \text{若 } X < Y \end{cases}$, 则 $E(Z^2) =$ _____。

$$\text{解: } E(Z^2) = 5 \cdot \int_0^{+\infty} y^2 e^{-y} dy \int_y^{+\infty} e^{-x} dx = 5 \cdot \int_0^{+\infty} y^2 e^{-y} e^{-y} dy$$

$$= \frac{5}{2} \cdot \int_0^{+\infty} y^2 2 \cdot e^{-2y} dy = \frac{5}{2} \left(\frac{1}{2^2} + \frac{1}{2^2} \right) = \frac{5}{4}$$

二元随机变量 (X, Y) 的联合密度函数 $f(x, y) = \begin{cases} e^{-x-y}, & x > 0, y > 0 \\ 0, & \text{其他} \end{cases}$, $Z = \begin{cases} Y, & \text{若 } X \geq Y \\ 2X, & \text{若 } X < Y \end{cases}$, 则 $E(Z^2) =$

3. 将一枚均匀的硬币独立地抛掷 16 次，正面出现次数记为 X ，希望对概率 $P(6 \leq X \leq 10)$ 做尽可能准确的估计，用中心极限定理的结果为_____，用切比雪夫不等式的结果为_____。

$$\text{解: } X \sim b(16, 0.5), \quad X \sim N(8, 4)$$

$$P(6 \leq X \leq 10) = P(5.5 \leq X \leq 10.5) \approx P\left(\frac{5.5-8}{2} \leq \frac{X-8}{2} \leq \frac{10.5-8}{2}\right) = 2\Phi(1.25) - 1 = 0.78$$

$$P(6 \leq X \leq 10) = P(5.5 \leq X \leq 10.5) = P(|X - E(X)| \leq 2.5) \geq 1 - \frac{4}{2.5^2} = 0.36$$

$$P(6 \leq X \leq 10) = P(|X - E(X)| \leq 2) \geq 1 - \frac{4}{2^2} = 0$$

$$P(6 \leq X \leq 10) = P(5 < X < 11) = P(|X - E(X)| < 3) \geq 1 - \frac{4}{3^2} = \frac{5}{9}$$

4. 二维随机变量 $(X, Y) \sim N(0, 0, 4, 4, 0.5)$, 则 $E(X^2 | X + Y = 4) =$ _____。

解析: $Cov(X+Y, X-Y)=0$, $X+Y \sim N(0,12)$, $X-Y \sim N(0,4)$

$$\begin{aligned} E(X^2 | X+Y=4) &= E\left(\left(\frac{X+Y+X-Y}{2}\right)^2 | X+Y=4\right) \\ &= E\left(\left(\frac{X+Y}{2}\right)^2 | X+Y=4\right) + E\left(\left(\frac{X-Y}{2}\right)^2 | X+Y=4\right) + E\left(2\left(\frac{X+Y}{2}\right)\left(\frac{X-Y}{2}\right) | X+Y=4\right) \\ &= 4 + E\left(\left(\frac{X-Y}{2}\right)^2\right) = 5 \end{aligned}$$

5. 设 X_1, X_2, \dots, X_8 为相互独立的 $N(0,1)$ 随机变量, 则 $P\left(\frac{X_1 - X_2}{|X_1 + X_2 + \dots + X_8|} \geq \underline{\hspace{2cm}}\right) = 0.05$ 。

解: $X_1 - X_2 \sim N(0,2)$, $X_1 + X_2 + \dots + X_8 \sim N(0,8)$

$Cov(X_1 - X_2, X_1 + X_2) = 0$,

$$\frac{(X_1 - X_2)^2 / 2}{(X_1 + X_2 + \dots + X_8)^2 / 8} \sim F(1,1), \quad P\left(\frac{(X_1 - X_2)^2 / 2}{(X_1 + X_2 + \dots + X_8)^2 / 8} \geq F_{0.9}(1,1)\right) = 0.1$$

$$P\left(\frac{(X_1 - X_2) / \sqrt{2}}{|X_1 + X_2 + X_3 + X_4| / 2\sqrt{2}} \geq \sqrt{F_{0.9}(1,1)}\right) = 0.05 \Rightarrow P\left(\frac{X_1 - X_2}{|X_1 + X_2 + X_3 + X_4|} \geq \frac{\sqrt{F_{0.9}(1,1)}}{2}\right) = 0.05$$

$$P\left(\frac{X_1 - X_2}{|X_1 + X_2 + X_3 + X_4|} \geq \sqrt{10}\right) = 0.05$$

$$\frac{(X_1 - X_2) / \sqrt{2}}{\sqrt{(X_1 + X_2 + \dots + X_8)^2 / 8}} \sim t(1), \quad P\left(\frac{(X_1 - X_2) / \sqrt{2}}{|X_1 + X_2 + \dots + X_8| / \sqrt{8}} \geq t_{0.95}(1)\right) = 0.05$$

$$P\left(\frac{X_1 - X_2}{|X_1 + X_2 + X_3 + X_4|} \geq \frac{t_{0.95}(1)}{2}\right) = 0.05, \quad P\left(\frac{X_1 - X_2}{|X_1 + X_2 + X_3 + X_4|} \geq 3.16\right) = 0.05$$

6. 对均匀总体 $U(0, \theta)$ 做假设检验, 原假设与备择假设分别为 $H_0: \theta = 5$, $H_1: \theta < 5$, 以 $x_{(n)} = \max\{x_1, \dots, x_n\}$ 为检验统计量, 显著性水平 $\alpha = 0.064$, 若样本容量 $n = 3$, 则拒绝域为_____。

$$\text{解: } P(x_{(n)} < c) = \left(\frac{c}{5}\right)^n = \alpha \Rightarrow \left(\frac{c}{5}\right)^3 = 0.064 \Rightarrow c = 2, \quad \{(x_1, x_2, x_3): x_{(3)} < 2\}$$

7. 设 X_1, \dots, X_{100} 是来自总体 $N(\mu, 4)$ 的样本。对期望进行假设检验, $H_0: \mu = 0$, $H_1: \mu > 0$, 若取拒绝域为

$\{(x_1, \dots, x_{100}): \bar{x} > 0.2\}$, 则此检验犯第一类错误的概率为_____，当 $\mu = 0.6$ 时，

此检验犯第二类错误的概率为_____， $\bar{x} = 0.2$ 的 p 值为_____。

解：第一类错误 $P_{\mu=0}(\bar{X} > 0.2) = P_{\mu=0}\left(\frac{\bar{X}}{\sqrt{\frac{4}{100}}} > 1\right) = 1 - \Phi(1) = 0.16,$

第二类错误 $P_{\mu=0.6}(\bar{X} \leq 0.2) = P_{\mu=0.6}\left(\frac{\bar{X} - 0.6}{\sqrt{\frac{4}{100}}} \leq \frac{0.2 - 0.6}{0.2}\right) = \Phi(-2) = 1 - \Phi(2) = 0.02$

$p\text{值} = P_{\mu=0}(\bar{X} > 0.2) = P_{\mu=0}\left(\frac{\bar{X}}{\sqrt{\frac{4}{100}}} > \frac{0.2}{0.2}\right) = 1 - \Phi(1) = 0.16$

二. (10 分) 学生在做有 4 个选项的选择题时, 如果他知道正确答案就一定可以答对, 如果不知道问题的正确答案, 就随机猜 1 个选项. 假设学生知道正确答案的概率是 0.6, 不知道正确答案的概率是 0.4。

(1) 求答对的概率, (2) 现从卷面上看题是答对了, 求学生知道正确答案的概率。

解: 设事件 A 为知道正确答案, 事件 B 为答对

(1) 由全概率公式 $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 1 \times 0.6 + 0.25 \times (1 - 0.6) = 0.7$

(2) 由贝叶斯公式 $P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1 \times 0.6}{0.7} = \frac{6}{7}.$

三. (10 分) 设随机变量 $X \sim N(0, 2)$, $Y = |X|$, 求随机变量 Y 的分布函数、密度函数、期望和方差。

解: 当 $y > 0$ 是, 分布函数

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = P\left(\frac{-y}{\sqrt{2}} \leq \frac{X}{\sqrt{2}} \leq \frac{y}{\sqrt{2}}\right) \\ &= \Phi\left(\frac{y}{\sqrt{2}}\right) - \Phi\left(\frac{-y}{\sqrt{2}}\right) = 2\Phi\left(\frac{y}{\sqrt{2}}\right) - 1 \end{aligned}$$

当 $y \leq 0$ 是, 分布函数 $F_Y(y) = 0$

密度函数 $p_Y(y) = \frac{d\left(2\Phi\left(\frac{y}{\sqrt{2}}\right) - 1\right)}{dy} = \sqrt{2}\varphi\left(\frac{y}{\sqrt{2}}\right)I_{y>0},$ 或 $p_Y(y) = \frac{1}{\sqrt{\pi}}e^{-\frac{y^2}{4}}I_{y>0}$

$$E(Y) = E(|X|) = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{x^2}{4}} dx = 2 \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{x^2}{4}} dx = 2 \int_0^{+\infty} \frac{1}{\sqrt{\pi}} e^{-\frac{x^2}{4}} d\frac{x^2}{4} = \frac{2}{\sqrt{\pi}}$$

$$E(Y^2) = E(X^2) = 2,$$

$$Var(Y) = E(Y^2) - E(Y)^2 = 2 - \frac{4}{\pi}.$$

四 (10 分) X, Y 均服从参数为 1 的指数分布, 且相互独立。令 $U = \max(X, Y)$, $V = \min(X, Y)$

(1) 分别求 U, V 的密度函数, (2) 计算 U, V 的相关系数。

$$\text{解: (1) } F_U(u) = P(U \leq u) = P(\max(X, Y) \leq u) = P(X \leq u, Y \leq u)$$

$$F_U(u) = P(U \leq u) = P(\max(X, Y) \leq u) = P(X \leq u, Y \leq u) = (1 - e^{-u})^2$$

$$p_U(u) = P(U \leq u) = P(\max(X, Y) \leq u) = P(X \leq u, Y \leq u) = 2e^{-u}(1 - e^{-u}) = 2(e^{-u} - e^{-2u}) I_{u>0}$$

$$F_V(v) = P(V \leq v) = P(\min(X, Y) \leq v) = 1 - P(\min(X, Y) > v) = 1 - P(X > v, Y > v) = 1 - e^{-2v}$$

$$p_V(v) = 2e^{-2v} I_{v>0}$$

$$(2) E(U) = \int_0^{+\infty} u \cdot 2(e^{-u} - e^{-2u}) du = 2 - \frac{1}{2} = \frac{3}{2}$$

$$E(U^2) = \int_0^{+\infty} u^2 \cdot 2(e^{-u} - e^{-2u}) du = 2 \cdot 2 - \left(\frac{2}{2^2} + \frac{2}{2^2}\right) = \frac{7}{2}, \quad Var(U) = \frac{7}{2} - \frac{9}{4} = \frac{5}{4}$$

$$V \sim \text{Exp}(2), \quad E(V) = \frac{1}{2}, \quad Var(V) = \frac{1}{4}$$

$$Cov(U, V) = E(UV) - E(U)E(V) = E(XY) - E(U)E(V)$$

$$= E(X)E(Y) - E(U)E(V) = 1 - \frac{3}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\text{Corr}(U, V) = \frac{Cov(U, V)}{\sqrt{Var(U)Var(V)}} = \frac{\frac{1}{4}}{\sqrt{\frac{5}{4} \cdot \frac{1}{4}}} = \frac{1}{\sqrt{5}}.$$

五. (10 分) 泊松总体 $X \sim P(\lambda)$, X_1, \dots, X_n 为来自该总体的样本,

(1) 用最大似然估计法给出总体期望的估计量,

(2) 试给出参数 λ^2 的无偏估计量。

解: (1) 似然函数 $L(\lambda; x_1, x_2, \dots, x_n) = \prod_{k=1}^n \frac{\lambda^{x_k}}{x_k!} \cdot e^{-\lambda} = \frac{1}{x_1! x_2! \dots x_n!} \lambda^{x_1+x_2+\dots+x_n} \cdot e^{-n\lambda}$

对数似然函数 $\ln L(\lambda) = C + (x_1 + x_2 + \dots + x_n) \ln \lambda - n\lambda$

$\max \ln L(\lambda) \Rightarrow \frac{d \ln L(\lambda)}{d\lambda} = \frac{1}{\lambda} (x_1 + x_2 + \dots + x_n) - n = 0$

解得 $\lambda = \frac{x_1 + x_2 + \dots + x_n}{n}$, 所以参数 λ 的最大似然估计量 $\hat{\lambda} = \bar{x}$ 。

(2) $E(X) = \lambda, \text{Var}(X) = \lambda$, 于是

$E(X^2) = \text{Var}(X) + (E(X))^2 = \lambda^2 + \lambda, \quad E(X^2 - X) = E(X^2) - E(X) = \lambda^2$

于是, $E\left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}\right) = \lambda^2$, 即 $X_1^2 - X_1$ 或 $\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}$ 是 λ^2 的无偏估计。

六. (10分) 设总体 $X \sim \text{Ge}(p)$, X_1, X_2, \dots, X_n 为来自该总体的样本。定义 $Y = \begin{cases} 1, & X_1 = 1 \\ 0, & X_1 > 1 \end{cases}$

(1) 计算 $E(Y | X_1 + X_2 + \dots + X_n)$

(2) 判断 Y 与 $E(Y | X_1 + X_2 + \dots + X_n)$ 是否为参数 p 的无偏估计量, 若否可否进行无偏校正。

解: (1) $X_1 + X_2 + \dots + X_n \sim \text{Nb}(n, p)$

$P(X_1 + X_2 + \dots + X_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k \geq n$

$$\begin{aligned} P(Y=1 | X_1 + X_2 + \dots + X_n = k) &= \frac{P(X_1=1, X_1 + X_2 + \dots + X_n = k)}{\binom{k-1}{n-1} p^n (1-p)^{k-n}} \\ &= \frac{P(X_1=1) P(X_2 + \dots + X_n = k-1)}{\binom{k-1}{n-1} p^n (1-p)^{k-n}} = \frac{p \cdot \binom{k-2}{n-2} p^{n-1} (1-p)^{k-n}}{\binom{k-1}{n-1} p^n (1-p)^{k-n}} = \frac{n-1}{k-1} \end{aligned}$$

$E(Y | X_1 + X_2 + \dots + X_n = k) = P(Y=1 | X_1 + X_2 + \dots + X_n = k) = \frac{n-1}{k-1}$

所以 $E(Y | X_1 + X_2 + \dots + X_n) = \frac{n-1}{X_1 + X_2 + \dots + X_n - 1}$

(2) $E(Y) = P(Y=1) = P(X_1=1) = p$

$E(E(Y | X_1 + X_2 + \dots + X_n)) = E(Y) = p,$

Y 与 $E(Y|X_1+X_2+\cdots+X_n)$ 均为统计量, 且期望都等于 p , 所以它们都是参数 p 的无偏估计量。

七. (15 分) X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu_1, \sigma^2)$ 的样本, Y_1, Y_2, \dots, Y_m 是来自总体 $Y \sim N(\mu_2, c \cdot \sigma^2)$ 的样本, $n=4, m=6, c=4.5, \bar{x}=52, \bar{y}=47, s_X^2=30, s_Y^2=99$ 。

(1) 给出参数 $\mu_1 - \mu_2$ 的矩估计量, 并计算该估计量的期望和方差;

(2) 若已知 $\sigma^2 = 25$, 给出参数 $\mu_1 - \mu_2$ 的 95% 置信水平的双侧置信区间;

(3) 若 σ^2 未知, 给出参数 $\mu_1 - \mu_2$ 的 90% 置信水平的单侧置信区间的置信下界。

解 (1) $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$, 所以参数 $\mu_1 - \mu_2$ 的矩估计量为 $\bar{X} - \bar{Y}$;

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{n}\right), \quad \bar{Y} \sim N\left(\mu_2, \frac{c \cdot \sigma^2}{m}\right),$$

$$E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2, \quad Var(\bar{X} - \bar{Y}) = \left(\frac{1}{n} + \frac{c}{m}\right) \sigma^2 = \sigma^2.$$

$$(2) \quad \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{c}{m}}} \sim N(0, 1) \text{ 可作为枢轴量, } \sqrt{\frac{1}{n} + \frac{c}{m}} = 1$$

$$P\left(-u_{0.975} \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma} \leq u_{0.975}\right) = 0.9, \text{ 参数 } \mu_1 - \mu_2 \text{ 的 95\% 双侧置信区间为}$$

$$[\bar{X} - \bar{Y} - u_{0.975} \cdot \sigma, \bar{X} - \bar{Y} + u_{0.975} \cdot \sigma], \text{ 即 } [-4.8, 14.8]$$

$$(3) \quad \frac{(n-1)s_X^2}{\sigma^2} \sim \chi^2(n-1), \quad \frac{(m-1)s_Y^2}{c \cdot \sigma^2} \sim \chi^2(m-1), \quad \frac{(n-1)s_X^2}{\sigma^2} + \frac{(m-1)s_Y^2}{c \cdot \sigma^2} \sim \chi^2(n+m-2),$$

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{c}{m}}}}{\frac{1}{\sigma} \sqrt{\left((n-1)s_X^2 + \frac{(m-1)s_Y^2}{c}\right) / (n+m-2)}} = \frac{\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n} + \frac{c}{m}}}}{\sqrt{\left((n-1)s_X^2 + \frac{(m-1)s_Y^2}{c}\right) / (n+m-2)}} \sim t(n+m-2)$$

$$\sqrt{\frac{1}{n} + \frac{c}{m}} = 1, \quad \sqrt{\left((n-1)s_X^2 + \frac{(m-1)s_Y^2}{c}\right) / (n+m-2)} = 5$$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\frac{1}{5}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{5} \sim t(8), \quad P\left(\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{5} \leq t_{0.9}(8)\right) = 0.9$$

参数 $\mu_1 - \mu_2$ 的 90% 置信水平的单侧置信区间的置信下界为 $\bar{X} - \bar{Y} - 5 \cdot t_{0.9}(8) = 5 - 5 \cdot 1.4 = -2$

八. (5 分) 设 y_1, \dots, y_n 是 x_1, \dots, x_n 按照升序排列后的结果, $y_1 < y_2 < \dots < y_n$ 。用随机化快速排序算法对 x_1, \dots, x_n 进行排序, 每一次都是从所有可能的元素中独立且均匀地选取基准元素, 对 $1 \leq i < j \leq n$, 定义随机变量 X_{ij} , 如果在算法过程中 y_i 与 y_j 进行了比较 $X_{ij} = 1$, 否则 $X_{ij} = 0$ 。求 $E(X_{ij})$ 。

解: 考虑 $\{y_i, y_{i+1}, \dots, y_j\}$, 当且仅当 y_i 或 y_j 是集合 $\{y_i, y_{i+1}, \dots, y_j\}$ 中第一个被选作基准的元素时, y_i 与 y_j 进行比较。

$$\text{所以 } P(X_{ij} = 1) = \frac{2}{j-i+1}, \quad E(X_{ij}) = \frac{2}{j-i+1}。$$

备注 1. 本考卷的样本均为简单随机样本, 样本均值 $\bar{X} = \frac{X_1 + \dots + X_n}{n}$, 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. 泊松分布随机变量 $X \sim P(\lambda)$ 的分布列 $P(X=k) = e^{-\lambda} \lambda^k / k!$

备注 4. 指数分布随机变量 $X \sim \text{Exp}(\lambda)$, $F_X(x) = (1 - e^{-\lambda x}) I_{x>0}$, $E(X) = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$

备注 5. 第三题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 6. 分布函数和分位数 (题目解答要严格按照下面给出数值进行计算)

$$\Phi(1.28) = 0.9, \quad \Phi(1.44) = 0.925, \quad \Phi(1.65) = 0.95, \quad \Phi(1.96) = 0.975,$$

$$\Phi(1) = 0.84, \quad \Phi(1.25) = 0.89, \quad \Phi(1.5) = 0.93, \quad \Phi(1.75) = 0.96, \quad \Phi(2) = 0.98, \quad \Phi(3) = 0.999$$

$$P(t(7) > 1.41) = 0.1, \quad P(t(8) > 1.40) = 0.1, \quad P(t(9) > 1.38) = 0.1, \quad P(t(10) > 1.37) = 0.1$$

$$P(t(7) > 1.89) = 0.05, \quad P(t(8) > 1.86) = 0.05, \quad P(t(9) > 1.83) = 0.05, \quad P(t(10) > 1.81) = 0.05$$

$$P(F(1,1) > 39.9) = 0.1, \quad P(F(1,2) > 8.53) = 0.1, \quad P(F(1,5) > 4.06) = 0.1, \quad P(F(2,1) > 49.5) = 0.1, \quad P(F(5,1) > 57.2) = 0.1$$

$$P(F(1,1) > 161) = 0.05, \quad P(F(1,2) > 18.5) = 0.05, \quad P(F(1,5) > 6.61) = 0.05, \quad P(F(2,1) > 200) = 0.05, \quad P(F(5,1) > 230) = 0.05$$