

# 中国科学技术大学

## 2021—2022学年第一学期考试试卷

考试科目 概率论与数理统计 得分 \_\_\_\_\_

所在院系 \_\_\_\_\_ 姓名 \_\_\_\_\_ 学号 \_\_\_\_\_

考试时间: 2022 年 1 月 12 日上午 8:30–10:30; 可使用简单计算器

### 一、(30分, 每小题3分) 填空题或单选题, 答案可以直接写在试卷上.

- (1) 设将ABC三个字母之一输入某信道, 独立地输出结果为原字母的概率是 0.8, 而输出为其它一字母的概率都是 0.1. 现等可能地将字母串AAAA, BBBB和CCCC之一输入该信道, 若已知输出结果为ABCA, 则输入的结果为AAAA的概率是\_\_\_\_\_.
- (2) 设  $A, B, C$  三个事件两两独立, 则它们相互独立的充分必要条件是( )  
 (A)  $A$  与  $B \cap C$  独立 (B)  $A \cap B$  与  $A \cup C$  独立  
 (C)  $A \cap B$  与  $A \cap C$  独立 (D)  $A \cup B$  与  $A \cup C$  独立
- (3) 设随机变量  $X \sim N(0, 1)$ ,  $Y$  服从参数为 0.5 的 Bernoulli 分布, 则  $Z = XY$  的分布函数的间断点个数为( )  
 (A) 0 (B) 1 (C) 2 (D) 3
- (4) 设随机变量  $X$  和  $Y$  相互独立且都服从区间  $(0, 1)$  上的均匀分布, 则对任一正整数  $n$ , 概率  $P(X^n + Y > 1) =$  \_\_\_\_\_.
- (5) 设在单位正方形内部随机取一点, 然后以该点为圆心画一个单位圆, 若以  $X$  表示落在该圆内正方形顶点的个数, 则  $EX =$  \_\_\_\_\_.
- (6) 设  $X_1, X_2, \dots, X_n$  是来自参数  $\lambda = 3$  的指数分布总体的一组简单随机样本, 若对任一  $\varepsilon > 0$ , 都有  $\lim_{n \rightarrow \infty} P(|\frac{1}{n} \sum_{i=1}^n X_i^2 - a| \geq \varepsilon) = 0$  成立, 则常数  $a =$  ( )  
 (A)  $\frac{1}{9}$  (B)  $\frac{2}{9}$  (C)  $\frac{1}{3}$  (D)  $\frac{4}{9}$
- (7) 设  $X_1, X_2, \dots, X_{10}$  是来自标准正态总体的一组简单随机样本, 且记统计量  $Y = \frac{1}{2} \sum_{i=1}^{10} X_i^2 + \sum_{i=1}^5 X_{2i-1} X_{2i}$ , 则  $Y$  的分布为( )  
 (A)  $\chi_4^2$  (B)  $\chi_5^2$  (C)  $\chi_9^2$  (D)  $\chi_{10}^2$
- (8) 设  $X_1, X_2, \dots, X_n$  是来自均匀总体  $U(\theta, 2\theta)$  的一组简单随机样本, 其中  $\theta > 0$  为一未知参数, 则  $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$  为  $\theta$  的( )  
 (A) 矩估计 (B) 极大似然估计 (C) 无偏估计 (D) 相合估计
- (9) 设  $X_1, X_2, \dots, X_{36}$  是来自正态总体  $N(\mu, 8)$  的一组简单随机样本, 且记  $\bar{X}$  为样本均值. 若以区间  $[\bar{X} - 1, \bar{X} + 1]$  作为  $\mu$  的置信区间, 则其置信水平为\_\_\_\_\_.
- (10) 设  $X_1, X_2, X_3$  是来自总体  $X \sim \begin{pmatrix} 1 & 2 & 3 \\ \theta^2 & 2\theta(1-\theta) & (1-\theta)^2 \end{pmatrix}$  的一组简单随机样本, 若假设检验  $H_0: \theta = 0.1 \leftrightarrow H_0: \theta = \theta_1$  的拒绝域为  $\{X_1 = X_2 = X_3 = 1\}$ , 其中  $0.5 < \theta_1 < 1$  是一个给定的常数, 则此检验犯第二类错误的概率为\_\_\_\_\_.

二、(20分) 设随机变量  $X \sim N(0, 1)$ , 而对任一实数  $x$ , 在  $X = x$  条件下,  $Y \sim N(x, 1)$ .

(1) 试求随机变量  $Y$  的密度函数  $f_Y(y)$ , 并指出  $Y$  服从何种分布.

(2) 试求条件期望  $E[XY|X = x]$ .

(3) 试求  $X$  和  $Y$  的相关系数.

(4) 试求常数  $a$ , 使得随机变量  $aX + Y$  和  $aX - Y$  相互独立.

三、(15分) 设随机变量  $X_1, X_2, X_3$  相互独立且都服从区间  $(0, 1)$  上的均匀分布.

(1) 若随机变量  $Y = -a \ln X_1$ , 其中  $a > 0$  为一给定常数, 试求  $Y$  的概率密度函数.

(2) 试求随机变量  $Z = X_2/X_1$  的分布函数.

(3) 试求随机变量  $U = 1/(X_1 X_2 X_3)$  的概率密度函数.

四、(20分) 设一系列随机变量  $Y_1, Y_2, \dots, Y_n$  满足

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

其中  $x_1, x_2, \dots, x_n$  为给定非负常数且不全相等,  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  相互独立且均服从正态分布  $N(0, \sigma^2)$ , 而  $\beta$  和  $\sigma^2$  为两个未知参数.

(1) 根据  $Y_1, Y_2, \dots, Y_n$  的分布, 请写出似然函数  $L(\beta, \sigma^2)$ .

(2) 试求  $\beta$  的极大似然估计量  $\hat{\beta}$ , 并证明  $\hat{\beta}$  为  $\beta$  的一个无偏估计.

(3) 证明  $\hat{\beta}^* = \sum_{i=1}^n Y_i / \sum_{i=1}^n x_i$  也为  $\beta$  的一个无偏估计, 并比较  $\hat{\beta}$  和  $\hat{\beta}^*$  哪个更有效.

五、(15分) 在 2021 年日本东京举行的第 32 届夏季奥运会中, 我国运动员杨倩和杨皓然获得了射击混合双人团体 10 米气步枪金牌. 决赛中 15 轮射击结果如下(单位: 环):

杨倩	10.5	9.7	10.0	10.5	10.3	10.2	10.1	10.6	10.4	9.9	10.8	10.8	10.4	10.4	10.4
杨皓然	10.3	10.4	10.9	10.2	10.5	10.4	10.2	10.8	10.6	10.0	10.5	10.7	10.4	10.1	10.7

设两位运动员的每次射击结果相互独立, 且均服从正态分布. 利用你所学的统计知识并结合上述决赛数据, 回答如下问题 (显著性水平  $\alpha = 0.05$ ):

(1) 两位运动员在比赛中射击成绩的方差是否可以认为是相等的?

(2) 假设“杨皓然平均射击水平高于杨倩”是否显著成立?

## 附录

标准正态分布函数:  $\Phi(1.645) = 0.95$ ,  $\Phi(1.96) = 0.975$ ,  $\Phi(2.121) = 0.983$

上分位数:  $t_{28}(0.025) = 2.048$ ,  $t_{28}(0.05) = 1.701$ ,  $F_{14,14}(0.025) = 2.979$ ,  $F_{14,14}(0.05) = 2.484$

(完)

## 参考答案

一、 每小题 3 分.

$\frac{4}{5}$ ; A; B;  $\frac{1}{n+1}$ ;  $\pi$ ; B; B; D; 0.966;  $1 - \theta_1^6$ .

二、 每小题 5 分.

(1) 由

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}},$$

可知随机向量  $(X, Y)$  的联合密度为

$$f(x, y) = f_X(x) f_{Y|X}(y|x) = \frac{1}{2\pi} e^{-\frac{x^2 + (y-x)^2}{2}}.$$

从而, 随机变量  $Y$  的密度函数为

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \frac{1}{2\sqrt{\pi}} e^{-\frac{y^2}{4}},$$

即  $Y$  服从正态分布  $N(0, 2)$ . [没指明具体分布扣 1 分.]

(2) 由题目条件,  $E[XY|X=x] = xE[Y|X=x] = x^2$ .

(3) 由 (1) 知  $\text{Var}(Y) = 2$ , 而由 (2) 知  $E[XY] = E[E(XY|X)] = E[X^2] = 1$ . 从而,

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{E[XY]}{\sqrt{2}} = \frac{\sqrt{2}}{2}.$$

(4) 易知,  $(aX + Y, aX - Y)$  服从二维正态分布, 且  $\text{Cov}(aX + Y, aX - Y) = a^2 - 2$ .

由于二维正态随机向量的不相关性和独立性等价, 故所求常数  $a = \pm\sqrt{2}$ .

三、 每小题 5 分.

(1) 对任一  $y > 0$ , 由于

$$P(Y \leq y) = P(-a \ln X_1 \leq y) = P(X_1 \geq e^{-y/a}) = 1 - e^{-y/a},$$

故  $Y$  服从参数为  $1/a$  的指数分布, 从而其概率密度函数为

$$f_Y(y) = \frac{1}{a} e^{-\frac{y}{a}}, \quad y > 0. \quad [\text{缺少或写错取值范围扣 2 分.}]$$

(2) 将  $(X_1, X_2)$  视为单位正方形上的均匀分布, 利用几何概型可知, 当  $0 \leq z < 1$  时,

$$P(Z \leq z) = P(X_2 \leq zX_1) = z/2;$$

而当  $z \geq 1$  时,

$$P(Z \leq z) = P(X_2 \leq zX_1) = 1 - 1/(2z).$$

故  $Z$  的分布函数为

$$F_Z(z) = \begin{cases} 0, & z < 0; \\ z/2, & 0 \leq z < 1; \\ 1 - 1/(2z), & z \geq 1. \end{cases}$$

[漏了  $z < 0$  的部分扣 1 分.]

- (3) (此小题也可用其它方法计算, 但稍繁) 由  $V := \ln U = -\sum_{i=1}^3 \ln X_i$  及 (1) 可知,  $V$  为 3 个独立  $\text{Exp}(1)$  随机变量之和, 故  $V$  服从  $\Gamma(1, 3)$  分布, 即其概率密度函数为

$$f_V(v) = \frac{1}{2}v^2e^{-v}, \quad v > 0.$$

再由  $U = e^V$  及密度函数变换公式可知,  $U$  的概率密度函数为

$$f_U(u) = \frac{\ln^2 u}{2u^2}, \quad u > 1. \quad [\text{缺少或写错取值范围扣 2 分.}]$$

四、第 1 小题 4 分, 后面两个小题都涉及两个结论, 各 8 分.

- (1) 由  $Y_i \sim N(\beta x_i, \sigma^2), i = 1, 2, \dots, n$  及它们相互独立可知,

$$L(\beta, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2 \right\}.$$

- (2) 记对数似然函数  $l(\beta, \sigma^2) = \ln L(\beta, \sigma^2)$ , 并令  $\frac{\partial l}{\partial \beta} = 0$  可得,  $\sum_{i=1}^n (y_i - \beta x_i)x_i = 0$ , 即  $\beta$  的极大似然估计量为

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

由  $\hat{\beta}$  为一列独立正态随机变量的线性组合, 故  $\hat{\beta}$  也服从正态分布, 且其期望和方差为

$$E(\hat{\beta}) = \frac{\sum_{i=1}^n \beta x_i^2}{\sum_{i=1}^n x_i^2} = \beta, \quad \text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

从而  $\hat{\beta}$  为  $\beta$  的一个无偏估计.

- (3) 与  $\hat{\beta}$  类似, 估计量  $\hat{\beta}^*$  也是一列独立正态随机变量的线性组合, 且

$$E(\hat{\beta}^*) = \frac{\sum_{i=1}^n \beta x_i}{\sum_{i=1}^n x_i} = \beta, \quad \text{Var}(\hat{\beta}^*) = \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2}.$$

故  $\hat{\beta}^*$  也为  $\beta$  的一个无偏估计, 且由 Cauchy-Schwarz 不等式可知, 当  $x_1, x_2, \dots, x_n$  不全相等时,  $\text{Var}(\hat{\beta}) < \text{Var}(\hat{\beta}^*)$ , 从而  $\hat{\beta}$  更有效.

五、先做一些计算. 杨倩:  $n_1 = 15, \bar{x} = 10.333, (n_1 - 1)S_1^2 = 1.353$ ; 杨皓然:  $n_2 = 15, \bar{y} = 10.447, (n_2 - 1)S_2^2 = 0.957; S_w^2 = 0.287^2$ . [上面的计算 3 分, 后面每小题各 6 分.]

- (1)  $H_0: \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2$ .

由

$$0.336 = \frac{1}{F_{14,14}(0.025)} < \frac{S_1^2}{S_2^2} = 1.414 < F_{14,14}(0.025) = 2.979,$$

接受  $H_0$ , 即可以认为他们的发挥稳定性相同.

- (2)  $H_0: \mu_1 \geq \mu_2 \leftrightarrow H_1: \mu_1 < \mu_2$ .

由于

$$t = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -1.08 > -t_{28}(0.05) = -1.701,$$

故接受  $H_0$ , 即不能认为杨皓然的比赛成绩显著高于杨倩.