

中国科学技术大学

2019—2020学年第一学期考试试卷

考试科目 概率论与数理统计(B) 得分 _____

所在系 _____ 姓名 _____ 学号 _____

考试时间: 2020年1月13日上午8:30-10:30; 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 $P(A) = P(B) = 0.4$, 且 $P(B|A) + P(\overline{B}|\overline{A}) = 1$, 则 $P(AB) =$ _____.
- (2) 甲乙二人抛掷一枚均匀的硬币, 甲抛了101次, 乙抛了100次, 则甲抛出的正面次数比乙多的概率是_____.
- (3) 设随机变量 X 的密度函数为 $f(x) = \frac{2}{\pi}\sqrt{1-x^2}$, $x \in (-1, 1)$. 对任意 $x \in (-1, 1)$, 若在条件 $X = x$ 下, 随机变量 Y 的条件分布律为

$$P(Y = -\sqrt{1-x^2}) = P(Y = \sqrt{1-x^2}) = 1/2,$$

则 Y _____ 连续型随机变量, (X, Y) _____ 连续型随机向量. ()

(A) 是, 是 (B) 是, 不是 (C) 不是, 是 (D) 不是, 不是

- (4) 在单位圆盘 $\{(x, y) : x^2 + y^2 \leq 1\}$ 上随机取两个点, 以随机变量 X 表示它们之间的距离, 则 $E(X^2) =$ _____.

- (5) 设 X_1, X_2, \dots, X_n 为一列独立同分布的随机变量, 且均服从参数为 $\lambda > 0$ 的指数分布. 记 $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 且 $\Phi(x)$ 为标准正态分布函数, 则对任意 $x \in \mathbb{R}$, 有()

(A) $\lim_{n \rightarrow \infty} P\left(\frac{\sqrt{n}}{\lambda}(\overline{X} - \lambda) \leq x\right) = \Phi(x)$ (B) $\lim_{n \rightarrow \infty} P\left(\sqrt{\frac{n}{\lambda}}(\overline{X} - \lambda) \leq x\right) = \Phi(x)$

(C) $\lim_{n \rightarrow \infty} P\left(\sqrt{n}(\lambda\overline{X} - 1) \leq x\right) = \Phi(x)$ (D) $\lim_{n \rightarrow \infty} P\left(\sqrt{n\lambda}(\overline{X} - \frac{1}{\lambda}) \leq x\right) = \Phi(x)$

- (6) 设 X_1, X_2, \dots, X_n 是来自标准正态总体的简单随机样本, 且 $1 \leq m < n$, 则当常数 $c =$ _____ 时, 统计量 $c(\sum_{i=1}^m X_i)^2 / \sum_{i=m+1}^n X_i^2$ 服从 F 分布.

- (7) 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, 其中 μ 为已知常数, 记 \overline{X} 和 S^2 分别为样本均值和样本方差, 则下列统计量中与 \overline{X} 不独立的是()

(A) 样本标准差 S (B) $\frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$ (C) $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ (D) $X_1 - X_2$

- (8) 设 X_1, X_2, X_3 是来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, 则下列统计量中, () 为 μ 的无偏估计且方差最小.

(A) $\frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3$ (B) $\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$

(C) $\frac{1}{5}X_1 + \frac{2}{5}X_2 + \frac{2}{5}X_3$ (D) $\frac{1}{7}X_1 + \frac{2}{7}X_2 + \frac{3}{7}X_3$

- (9) 对一正态总体 $N(\mu, 100)$ 的均值 μ 求置信水平为 95% 的置信区间, 若要求其区间长度不大于 4, 则样本容量 n 至少应取_____.

- (10) 假设检验中, 在显著性水平 $\alpha = 0.05$ 下若原假设 H_0 被接受, 这说明()

(A) 有充分的理由表明 H_0 是正确的 (B) 没有充分的理由表明 H_0 是错误的

(C) 有充分的理由表明 H_1 是错误的 (D) 没有充分的理由表明 H_1 是正确的

二、(20分) 设 X_1, X_2, \dots, X_n 为一列独立的随机变量, 且均服从 $U(0, 1)$ 分布. 记

$$Y = \min\{X_1, X_2, \dots, X_n\}, \quad Z = \max\{X_1, X_2, \dots, X_n\}.$$

(1) 试证明: 对任意常数 $0 < y, z < 1$, 有

$$P(Y \leq y, Z \leq z) = \begin{cases} z^n - (z - y)^n, & y < z; \\ z^n, & y \geq z. \end{cases}$$

(2) 利用上述结果, 试求随机变量 Y 和 Z 的联合密度函数 $f(y, z)$.

(3) 在 $Y = y$ 条件下 ($0 < y < 1$), 试求 Z 的条件密度函数 $f_{Z|Y}(z|y)$.

(4) 若 $n = 2$, 试求 Y 和 Z 的协方差 $\text{Cov}(Y, Z)$.

三、(15分) 设随机变量 X, Y 和 Z 相互独立, 且均服从参数为 1 的指数分布. 记

$$U = \frac{X}{X+Y}, \quad V = \frac{X+Y}{X+Y+Z}, \quad W = X+Y+Z.$$

(1) 计算随机向量 (U, V, W) 的联合密度函数.

(2) 随机变量 U, V 和 W 是否相互独立? 请证明你的结论.

四、(15分) 设某种元件的使用寿命 T 的分布函数为 $F(t) = \begin{cases} 1 - \exp\{-(\frac{t}{\theta})^m\}, & t \geq 0; \\ 0, & t < 0, \end{cases}$ 其中 $m > 0$ 为已知参数, 而 $\theta > 0$ 为未知参数. 随机取 n 个这种元件, 测得它们的使用寿命分别为 T_1, T_2, \dots, T_n . 记 $g(\theta) = \theta^m$.

(1) 试求 $g(\theta)$ 的极大似然估计 $\hat{g}(T_1, T_2, \dots, T_n)$.

(2) 上述估计是否为无偏估计? 请证明你的结论.

五、(12分) 经大量调查, 已知一般健康成年男子每分钟脉搏的次数服从正态分布 $N(72, 6^2)$. 现测得 16 例成年男子慢性铅中毒患者的脉搏平均 67 次/分钟, 标准差为 7 次/分钟. 问在显著性水平 0.05 下, 这群患者每分钟脉搏的次数(假设也服从正态分布)和正常人有无显著性差异? (要求对均值和方差都进行检验.)

六、(8分) 中国科学技术大学 2019 级本科新生入学考试中, 某学院两个班级的英语科目各档成绩(从低到高)人数如下表所示:

档次	I	II	III	IV	V	VI	合计
一班	8	27	10	6	8	6	65
二班	15	25	8	7	6	4	65

我们能否认为这两个班级的英语水平大致相当? 显著性水平设为 $\alpha = 0.05$.

附录:

$$\Phi(1.645) = 0.95, \quad \Phi(1.96) = 0.975;$$

$$t_{15}(0.025) = 2.131, \quad t_{15}(0.05) = 1.753, \quad t_{16}(0.025) = 2.12, \quad t_{16}(0.05) = 1.746;$$

$$\chi_5^2(0.95) = 1.145, \quad \chi_5^2(0.05) = 11.071, \quad \chi_{15}^2(0.975) = 6.262, \quad \chi_{15}^2(0.025) = 27.488.$$

参考答案

- 一. (1) 0.16 (2) 0.5 (3) B (4) 1 (5) C
(6) $\frac{n-m}{m}$ (7) C (8) B (9) 97 (若答 96 也算对) (10) B.

二. (1) 当 $0 < y, z < 1$ 时, 由 X_1, \dots, X_n 独立同分布可知

$$\begin{aligned} P(Y \leq y, Z \leq z) &= P(Z \leq z) - P(Y > y, Z \leq z) \\ &= P(X_1 \leq z, \dots, X_n \leq z) - P(y < X_1 \leq z, \dots, y < X_n \leq z) \\ &= [P(X_1 \leq z)]^n - [P(y < X_1 \leq z)]^n. \end{aligned}$$

再由 $X_1 \sim U(0, 1)$ 即知

$$P(Y \leq y, Z \leq z) = \begin{cases} z^n - (z - y)^n, & y < z; \\ z^n, & y \geq z. \end{cases}$$

(2) 注意到联合密度函数 $f(y, z)$ 的取值范围是 $0 < y < z < 1$, 将 (1) 中联合分布函数 $P(Y \leq y, Z \leq z)$ 对变量 y 和 z 求一阶偏导数, 即得

$$f(y, z) = n(n-1)(z-y)^{n-2}, \quad 0 < y < z < 1.$$

(如果上述表达式正确, 但没有说明取值范围为 $0 < y < z < 1$, 扣 1-2 分.)

(3) 由 (2) 可知, 随机变量 Y 的密度函数为

$$f_Y(y) = \int_y^1 f(y, z) dz = n(1-y)^{n-1}, \quad 0 < y < 1.$$

从而

$$f_{Z|Y}(z|y) = \frac{f(y, z)}{f_Y(y)} = \frac{(n-1)(z-y)^{n-2}}{(1-y)^{n-1}}, \quad y < z < 1.$$

(如果上述表达式正确, 但没有说明取值范围为 $y < z < 1$, 扣 1-2 分.)

(4) 当 $n = 2$ 时, 随机变量 Y 的密度函数为 $f_Y(y) = 2(1-y)$, $0 < y < 1$, 故

$$EY = \int_0^1 2y(1-y) dy = \frac{1}{3}.$$

类似地, 可求得随机变量 Z 的密度函数为 $f_Z(z) = 2z$, $0 < z < 1$, 及 $EZ = \frac{2}{3}$. 此外, 由于此时联合密度函数退化为 $f(y, z) = 2$, $0 < y < z < 1$, 我们有

$$E[YZ] = \int_0^1 \int_y^1 2yz dz dy = \frac{1}{4}.$$

所以,

$$\text{Cov}(Y, Z) = E[YZ] - EY \cdot EZ = \frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{36}.$$

三. (1) 由

$$u = \frac{x}{x+y}, \quad v = \frac{x+y}{x+y+z}, \quad w = x+y+z,$$

可得 $x = uvw$, $y = (1-u)vw$, $z = (1-v)w$, 从而Jacobi行列式

$$J = \begin{vmatrix} vw & uw & uv \\ -vw & (1-u)w & (1-u)w \\ 0 & -w & 1-v \end{vmatrix} = vw^2.$$

由 (X, Y, Z) 的联合密度函数 $f(x, y, z) = e^{-(x+y+z)}$, $x, y, z > 0$, 及密度变换公式可得所求随机向量 (U, V, W) 的联合密度函数为

$$p(u, v, w) = vw^2 e^{-w}, \quad 0 < u, v < 1, w > 0.$$

(如果上述表达式正确, 但没有说明取值范围, 扣 1-2 分.)

(2) 随机变量 U, V 和 W 相互独立. 事实上, 由上述联合密度函数可以分解为

$$p(u, v, w) = p_U(u)p_V(v)p_W(w)$$

即知该结论成立, 其中

$$p_U(u) = 1, \quad 0 < u < 1; \quad p_V(v) = 2v, \quad 0 < v < 1; \quad p_W(w) = \frac{1}{2}w^2 e^{-w}, \quad w > 0.$$

四. (1) 由总体 T 的概率密度函数为 $f(t) = \frac{mt^{m-1}}{\theta^m} \exp\{-(\frac{t}{\theta})^m\}$, $t \geq 0$, 故似然函数为

$$L(\theta) = \frac{m^n}{\theta^{mn}} \prod_{i=1}^n t_i^{m-1} \exp\left\{-\left(\frac{t_i}{\theta}\right)^m\right\}.$$

对其取对数, 得对数似然函数为

$$l(\theta) = \ln L(\theta) = C - n \ln(\theta^m) - \frac{1}{\theta^m} \sum_{i=1}^n t_i^m.$$

将上式对 θ^m 求导数, 并令其等于 0, 可得(也可对 θ 求导, 结果相同)

$$\frac{dl(\theta)}{d(\theta^m)} = -\frac{n}{\theta^m} + \frac{1}{\theta^{2m}} \sum_{i=1}^n t_i^m = 0.$$

解之即可得所求极大似然估计量为

$$\hat{g}(T_1, T_2, \dots, T_n) = \frac{1}{n} \sum_{i=1}^n T_i^m.$$

(2) 由

$$\begin{aligned} E[\hat{g}(T_1, T_2, \dots, T_n)] &= \frac{1}{n} \sum_{i=1}^n E[T_i^m] = E[T^m] \\ &= \int_0^\infty \frac{mt^{2m-1}}{\theta^m} \exp\left\{-\left(\frac{t}{\theta}\right)^m\right\} dt = \theta^m \int_0^\infty x e^{-x} dx = \theta^m, \end{aligned}$$

可知, $\hat{g}(T_1, T_2, \dots, T_n)$ 是 $g(\theta)$ 的一个无偏估计.

五. (1) 均值的检验. $H_0: \mu = 72$, $\longleftrightarrow H_1: \mu \neq 72$. 由 t 检验统计量

$$t = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = \frac{\sqrt{16}(67 - 72)}{7} = -2.857,$$

可知 $|t| > t_{15}(0.025) = 2.131$, 故应拒绝原假设, 即认为患者每分钟脉搏的平均次数与正常人有显著性差异.

(2) 方差的检验. $H_0: \sigma^2 = 6^2$, $\longleftrightarrow H_1: \sigma^2 \neq 6^2$. 由 χ^2 检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{15 \cdot 7^2}{6^2} = 20.417,$$

可知 $\chi_{15}^2(0.975) = 6.262 < \chi^2 < \chi_{15}^2(0.025) = 27.488$, 故应接受原假设, 即认为患者每分钟脉搏次数的方差与正常人相同. 结合均值和方差两个方面, 我们最终可认为患者每分钟脉搏的次数与正常人有显著性差异.

(注: 如果先进行方差的检验, 认定方差可以等于 6^2 , 然后利用一样本 u 检验也算正确答案, 均值的检验结果与上面相同.)

六. 拟合优度联列表齐一性检验. 原假设为两个班级的英语水平相当, 而其检验统计量为

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^6 \frac{(nn_{ij} - n_{i.}n_{.j})^2}{nn_{i.}n_{.j}}.$$

代入数据计算可知, $\chi^2 = 3.1922 < \chi_5^2(0.05) = 11.071$, 故在显著性水平 $\alpha = 0.05$ 下我们不能拒绝原假设, 即可认为两个班级的英语水平相当.