

中国科学技术大学

2021—2022学年第二学期考试试卷

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 6 月 13 日下午 14:30–16:30; 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) 抛掷一枚均匀的硬币 n 次, 已知出现正面的次数为 k , 则第 1 次抛掷结果为正面的概率是 .

(2) 已知随机变量 X 和 Y 满足

$$P(X = 1) = p_1, \quad P(\max\{X, Y\} = 1) = p_2, \quad P(\min\{X, Y\} = 1) = p_3,$$

则 $P(Y = 1) =$.

(3) 设随机向量 (X, Y) 服从二维正态分布 $N(0, 0; 1, 1; 0)$, 则 $E[\sqrt{X^2 + Y^2}] =$.

(4) 设 X 和 Y 为随机变量, $Y \sim N(2, 2)$, 且满足 $E[X|Y] = Y^2$, 则 $EX =$.

(5) 在任一三角形 $\triangle ABC$ 内部随机取一点记为 P , 然后在边 BC 上随机取一点记为 Q , 则直线 PQ 与边 AB 相交的概率为()

(A) $\frac{1}{2}$ (B) $\frac{AB}{AB+AC}$ (C) $\frac{AB+\frac{1}{2}BC}{AB+AC+BC}$ (D) $\frac{AB^2}{AB^2+AC^2}$

(6) 设 X 和 Y 为两个独立的标准正态随机变量, 则 $(X+Y)^2/(X-Y)^2$ 的分布为()

(A) χ_1^2 分布 (B) $F_{1,1}$ 分布 (C) $F_{2,2}$ 分布 (D) 以上均不对

(7) 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(0, \sigma^2)$ 的一组简单随机样本, 其中 σ^2 已知, 且以 S^2 表示样本方差, 则()

(A) $\sum_{i=1}^n X_i^2 \sim \chi_n^2$ (B) $\left(\frac{X_i}{\sigma}\right)^2 + \frac{(n-1)S^2}{\sigma^2} \sim \chi_n^2$
(C) $\frac{1}{n} \sum_{i=1}^n \left(\frac{X_i}{\sigma}\right)^2 + \frac{(n-1)S^2}{\sigma^2} \sim \chi_n^2$ (D) $\frac{1}{n} \left(\sum_{i=1}^n \frac{X_i}{\sigma}\right)^2 + \frac{(n-1)S^2}{\sigma^2} \sim \chi_n^2$

(8) 设总体 X 的期望为 μ , 而 $\hat{\mu}_1$ 和 $\hat{\mu}_2$ 分别为 μ 的两个无偏估计量, 它们的方差分别为 1 和 2, 相关系数为 $\frac{\sqrt{2}}{4}$. 若使 $\alpha\hat{\mu}_1 + \beta\hat{\mu}_2$ 也为 μ 的一个无偏估计且方差尽可能小, 则两常数之积 $\alpha\beta =$.

(9) 设 X_1, X_2, \dots, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的一组简单随机样本, 其中方差 σ^2 未知. 在置信水平 α 下, 设 μ 的置信区间为 $[\hat{\mu}_1, \hat{\mu}_2]$, 单侧置信下限和上限分别为 $\hat{\mu}_1^*$ 和 $\hat{\mu}_2^*$, 则下列关系中错误的是()

(A) $\hat{\mu}_1 < \hat{\mu}_1^*$ (B) $\hat{\mu}_2 > \hat{\mu}_2^*$ (C) $\hat{\mu}_2 - \hat{\mu}_1 = \hat{\mu}_2^* - \hat{\mu}_1^*$ (D) $\hat{\mu}_2 - \hat{\mu}_2^* = \hat{\mu}_1^* - \hat{\mu}_1$

(10) 现对正态总体 $N(\mu, \sigma^2)$ 中的参数 μ 进行假设检验, 若在显著性水平 $\alpha = 0.05$ 下接受了原假设 $H_0: \mu = \mu_0$, 则在显著性水平 $\alpha = 0.01$ 下, 下列说法正确的是()

(A) 依然接受 H_0 (B) 拒绝 H_0
(C) 可能接受或拒绝 H_0 (D) 犯第一类错误概率变大

二、(16分) 设随机向量 (X, Y) 的联合概率密度函数为

$$f(x, y) = cx(y - x)e^{-y}, \quad 0 \leq x \leq y < \infty.$$

- (1) 求常数 c .
- (2) 分别求条件概率密度函数 $f_{X|Y}(x|y)$ 和 $f_{Y|X}(y|x)$.
- (3) 求期望 $E[XY]$.
- (4) 求 X 和 Y 的相关系数 $\text{Corr}(X, Y)$.

三、(18分) 设随机变量 X 服从区间 $(0, 1)$ 上的均匀分布, Y 服从均值为 2 的指数分布, 且它们相互独立.

- (1) 写出随机向量 (X, Y) 的联合概率密度函数 $f(x, y)$.
- (2) 求随机变量 $Z = X + Y$ 的概率密度函数 $p(z)$.
- (3) 求关于 t 的一元二次方程 $t^2 + 2Xt + Y = 0$ 有实根的概率 (精确到小数点后 3 位).

四、(24分) 已知总体 X 服从 Pareto 分布, 即其概率密度函数为

$$f(x) = \frac{\alpha c^\alpha}{x^{\alpha+1}}, \quad x \geq c,$$

其中 $\alpha > 2, c > 0$ 为未知参数. 设 X_1, X_2, \dots, X_n 是来自该总体的一组简单随机样本, 且分别以 \bar{X} 和 S^2 表示样本均值和样本方差.

- (1) 分别求 α 和 c 的矩估计 $\hat{\alpha}_1$ 和 \hat{c}_1 .
- (2) 分别求 α 和 c 的极大似然估计 $\hat{\alpha}_2$ 和 \hat{c}_2 .
- (3) 估计量 \hat{c}_2 是否为 c 的一个无偏估计? 若是, 证明之; 若否, 修正之.

五、(12分) 随机选取 8 个成人, 分别测量了他们在早晨起床时和晚上就寝时的身高 (单位: 厘米), 得到如下数据:

序号	1	2	3	4	5	6	7	8
早晨(x_i)	172	168	180	181	160	163	165	177
晚上(y_i)	172	167	177	179	159	161	166	175

设成人身高服从正态分布, 能否认为成人早晨的身高显著高于晚上的身高 (显著性水平 $\alpha = 0.05$)?

附录 标准正态分布函数: $\Phi(1) = 0.8413$, $\Phi(1.645) = 0.95$, $\Phi(1.96) = 0.975$

上分位数: $t_7(0.025) = 2.365$, $t_7(0.05) = 1.895$, $t_{14}(0.025) = 2.145$, $t_{14}(0.05) = 1.761$

(完)

参考答案

一、 每小题 3 分.

$\frac{k}{n}$; $p_2 + p_3 - p_1$; $\sqrt{\frac{\pi}{2}}$ (或 $\frac{1}{2}\sqrt{2\pi}$); 6; A; B; D; $\frac{3}{16}$; C; A.

二、 每小题 4 分.

(1) 由边缘概率密度函数 $f_X(x) = cxe^{-x}, x > 0$ 或 $f_Y(y) = \frac{c}{6}y^3e^{-y}, y > 0$, 即知 $c = 1$.

(2) 所求条件概率密度函数为[没写变量取值范围, 每处扣 1 分]

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = 6x(y - x)y^{-3}, \quad 0 \leq x \leq y;$$

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = (y - x)e^{x-y}, \quad x \leq y < \infty.$$

(3) 可由

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy = \int_0^{\infty} ye^{-y}dy \int_0^y x^2(y - x)dx$$

及 Γ 函数的性质直接计算得出 $E[XY] = 10$. 但这里提供另一种方法, 无需二重积分. 首先, 由第 1 小题中的边缘密度和 Γ 函数的性质直接可得

$$EX = 2, \quad \text{Var}(X) = 2; \quad EY = 4, \quad \text{Var}(Y) = 4.$$

另一方面, 由 (2) 中结论可知

$$E[Y|X = x] = \int_x^{\infty} y(y - x)e^{x-y}dy = x + 2,$$

亦即 $E[Y|X] = X + 2$. 由此可得,

$$E[XY] = E[E(XY|X)] = E[XE(Y|X)] = E[X^2] + 2EX = 10.$$

(4) 综上所述, $\text{Corr}(X, Y) = (10 - 2 \times 4)/\sqrt{2 \times 4} = \frac{\sqrt{2}}{2}$.

三、 每小题 6 分.

(1) 所求答案为 $f(x, y) = \frac{1}{2}e^{-y/2}, 0 < x < 1, y > 0$. [没写变量取值范围扣 2 分.]

(2) [此小题方法有多种选择, 如直接利用卷积公式, 但必须注意变量的取值范围满足 $z - 1 \leq y \leq z, y > 0$, 故需对 z 进行分段讨论.] 所求答案为

$$p(z) = \begin{cases} 1 - e^{-z/2}, & 0 < z < 1; \\ (\sqrt{e} - 1)e^{-z/2}, & z \geq 1. \end{cases}$$

(3) 即求概率 $P(Y \leq X^2)$. 设 G 为由曲线 $y = x^2, x = 1$ 和 $y = 0$ 所围成的区域, 则

$$\begin{aligned} P(Y \leq X^2) &= \iint_G f(x, y)dxdy = \int_0^1 dx \int_0^{x^2} \frac{1}{2}e^{-y/2}dy \\ &= \int_0^1 \left(1 - e^{-\frac{x^2}{2}}\right)dx = 1 - \sqrt{2\pi}[\Phi(1) - \Phi(0)] \\ &= 0.145. \end{aligned}$$

四、 每小题 8 分.

(1) 首先, 可计算得

$$EX = \frac{\alpha c}{\alpha - 1}, \quad \text{Var}(X) = \frac{\alpha c^2}{(\alpha - 1)^2(\alpha - 2)}.$$

然后通过联立方程 $EX = \bar{X}$, $\text{Var}(X) = S^2$, 可得 α 和 c 的矩估计分别为

$$\hat{\alpha}_1 = 1 + \sqrt{1 + \bar{X}^2/S^2}, \quad \hat{c}_1 = \frac{\bar{X}\sqrt{1 + \bar{X}^2/S^2}}{1 + \sqrt{1 + \bar{X}^2/S^2}} \text{ 或 } \frac{\bar{X}\sqrt{\bar{X}^2 + S^2}}{S + \sqrt{\bar{X}^2 + S^2}}.$$

注: 在本小题中, 可将样本方差 S^2 替换成样本中心二阶矩 $m_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

(2) 首先, 可得似然函数为

$$L(\alpha, c) = \prod_{i=1}^n \frac{\alpha c^\alpha}{x_i^{\alpha+1}}, \quad c \leq x_1, x_2, \dots, x_n,$$

且由此可知参数 c 的极大似然估计为 $\hat{c}_2 = X_{(1)}$, 即 $\min\{X_1, X_2, \dots, X_n\}$. 然后, 可求得对数似然函数为

$$l(\alpha, c) = n \ln \alpha + n\alpha \ln c - (\alpha + 1) \sum_{i=1}^n \ln x_i.$$

最后, 通过令 $\frac{\partial l(\alpha, c)}{\partial \alpha} = 0$, 可得 α 的极大似然估计为 $\hat{\alpha}_2 = n / \sum_{i=1}^n \ln(X_i / X_{(1)})$.

(3) 否. 通过计算可知估计量 $\hat{c}_2 = X_{(1)}$ 服从参数为 $n\alpha$ 和 c 的 Pareto 分布. 故由第 1 小问可知 $E[\hat{c}_2] = \frac{n\alpha c}{n\alpha - 1}$, 从而 \hat{c}_2 不是 c 的无偏估计, 但可修正为 $\hat{c}_2^* = \frac{n\alpha - 1}{n\alpha} X_{(1)}$.

五、 成对数据检验, 且可认为成对数据之差来自正态总体 $N(\mu, \sigma^2)$. 本题要求在显著性水平 $\alpha = 0.05$ 下检验 $H_0: \mu \leq 0 \leftrightarrow H_1: \mu > 0$. 对成对数据, 样本容量 $n = 8$, $\bar{x} - \bar{y} = 1.25$, $s = 1.2817$, 则由一样本 t 检验的检验统计量为

$$t = \frac{\bar{x} - \bar{y}}{s/\sqrt{n}} = 2.758 > t_7(0.05) = 1.895,$$

故拒绝原假设, 即认为“成人早晨的身高比晚上的身高要高”具有显著性.