

中国科学技术大学
2022—2023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: 2023 年 2 月 22 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) 假设婴儿性别为男或女的概率相同, 且出生在任一季节的概率也相同. 在一个两孩家庭中, 已知有在春季出生的男孩, 则该家庭两个孩子都是男孩的概率为_____.
- (2) 设随机变量 X 和 Y 的概率密度函数分别为 $f_1(x)$ 和 $f_2(x)$, 分布函数分别为 $F_1(x)$ 和 $F_2(x)$, 则下列说法中一定正确的是()
(A) $f_1(x) + f_2(x)$ 为概率密度函数 (B) $f_1(x)f_2(x)$ 为概率密度函数
(C) $F_1(x) + F_2(x)$ 为分布函数 (D) $F_1(x)F_2(x)$ 为分布函数
- (3) 平面上有 10 个不同的点, 且任意三点不共线. 设它们任意两点之间独立地以概率 $1/2$ 连有一条边, 则以这些点为顶点的三角形个数的期望为_____.
- (4) 设随机变量 X 和 Y 均服从标准正态分布, 则下列说法错误的是()
(A) 若它们相互独立, 则 (X, Y) 服从二元正态分布
(B) 若它们相互独立, 则 X/Y 服从 Cauchy 分布
(C) 若协方差 $\text{Cov}(X, Y) = 0$, 则它们相互独立
(D) 若它们相互独立, 则 $X + Y$ 与 $X - Y$ 也相互独立
- (5) 设随机变量 X 服从参数为 λ 的指数分布, 则它的熵 $H(X) =$ _____.
- (6) 设随机变量 X 服从参数为 3 的 Poisson 分布, Y 服从正态分布 $N(3, 1)$, 且它们相互独立, 则由切比雪夫不等式可得概率 $P(X - 3 < Y < X + 3)$ 满足()
(A) $\leq \frac{1}{4}$ (B) $\leq \frac{4}{9}$ (C) $\geq \frac{3}{4}$ (D) $\geq \frac{5}{9}$
- (7) 设 X_1, X_2, \dots, X_n ($n \geq 2$) 是来自正态总体 $N(\mu, 1)$ 的简单随机样本, 且记 \bar{X} 为样本均值, 则下列中不服从 χ^2 分布的是()
(A) $\sum_{i=1}^n (X_i - \mu)^2$ (B) $2(X_n - X_1)^2$ (C) $\sum_{i=1}^n (X_i - \bar{X})^2$ (D) $n(\bar{X} - \mu)^2$
- (8) 在假设检验中, 下列关于显著性水平 α 和 p 值的说法中正确的是()
(A) 若 $p < \alpha$, 则拒绝原假设 H_0 (B) p 值可随 α 的改变而改变
(C) p 值大小与原假设 H_0 无关 (D) p 值大小与备择假设 H_1 无关
- (9) 已知 X_1, X_2, \dots, X_{16} 是来自正态总体 $N(\mu, 4)$ 的一组简单随机样本, 且记 \bar{X} 为样本均值. 若假设检验 $H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$ 的拒绝域为 $\{|\bar{X} - \mu_0| > c\}$, 则在显著性水平 $\alpha = 0.05$ 下, 临界值 $c =$ _____.
- (10) 假设总体 X 的分布律为 $\begin{pmatrix} -1 & 0 & 1 \\ p & 0.5 & 0.5 - p \end{pmatrix}$, 其中 $0 < p < 0.5$ 为参数. 当使用拟合优度检验来检验一组简单随机样本是否来自总体 X 时, 则检验统计量在原假设成立条件下的渐近 χ^2 分布的自由度为_____.

二、(16分) 设二维随机向量 (X, Y) 服从平面上以 $(0, 0)$, $(0, 1)$ 和 $(1, 1)$ 为顶点的三角形区域上的均匀分布.

(1) 求随机向量 (X, Y) 的概率密度函数 $f(x, y)$.

(2) 设随机变量 $U, V \sim U(0, 1)$ 且相互独立, 以记号 $\stackrel{d}{=}$ 表示同分布, 证明:

$$X \stackrel{d}{=} \min\{U, V\}, \quad Y \stackrel{d}{=} \max\{U, V\}.$$

(3) 对任一 $0 < x < 1$, 在条件 $X = x$ 下, Y 是否服从均匀分布? 请通过计算说明.

(4) 求随机变量 X 和 Y 的相关系数 $\rho_{X,Y}$.

三、(18分) 设 X, Y, U_1, U_2, \dots 为一列相互独立的随机变量, 且 X 和 Y 的分布律为

$$P(X = n) = (e - 1)e^{-n}, \quad P(Y = n) = \frac{1}{(e - 1)n!}, \quad n = 1, 2, \dots,$$

而 $U_i \sim U(0, 1)$, $i = 1, 2, \dots$. 记随机变量 $M = \max\{U_1, U_2, \dots, U_Y\}$ 及 $Z = X - M$.

(1) 对任一正整数 n 和 $0 \leq x \leq 1$, 求条件概率 $P(M \leq x | Y = n)$.

(2) 求随机变量 M 的概率密度函数 $f_M(x)$.

(3) 对任一实数 $z > 0$, 求概率 $P(Z > z)$ 并依此具体指出 Z 服从的分布.

四、(18分) 设 X_1, X_2, \dots, X_m 和 Y_1, Y_2, \dots, Y_n 为分别来自正态总体 $N(\mu, \sigma^2)$ 和 $N(\mu, 2\sigma^2)$ 的两组独立简单随机样本, 其中参数 μ 已知而 σ^2 未知.

(1) 求 σ^2 的最大似然估计量 $\hat{\sigma}^2$.

(2) 上述估计量是否为无偏估计? 若是, 请证明你的结论; 若否, 请修正之.

(3) 求枢轴变量 $(m + n)\hat{\sigma}^2/\sigma^2$ 的分布, 并以此构造 σ^2 的置信系数为 0.95 的置信区间.

五、(18分) 为了检验成年人的体脂率是否存在性别差异, 随机从某地区抽取了一些志愿者进行了测量, 分组数据如下(单位: %):

男	13.3	6.0	20.0	8.0	14.0	19.0	18.0	25.0	16.0	24.0	15.0	1.0	15.0
女	22.0	16.0	21.7	21.0	30.0	26.0	12.0	23.2	28.0	23.0			

已知成年男性和成年女性的体脂率均服从正态分布. 在显著性水平 $\alpha = 0.05$ 下,

(1) 是否可以认为体脂率的方差没有性别差异?

(2) 是否可以认为成年女性的体脂率显著地大于成年男性的体脂率?

附录

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

上分位数: $t_{21}(0.025) = 2.08$, $t_{21}(0.05) = 1.72$,

$F_{12,9}(0.025) = 3.44$, $F_{9,12}(0.025) = 3.87$, $F_{12,9}(0.05) = 2.8$, $F_{9,12}(0.05) = 3.07$

(完)

参考答案

一、 每小题 3 分.

[1-5] $\frac{7}{15}$; D; 15; C; $1 - \ln \lambda$;

[6-10] D; B; A; 0.98; 1.

二、 每小题 4 分.

(1) 所求为 $f(x, y) = 2, 0 < x < y < 1$. (取值范围出错扣 2 分.)

(2) 由 (1) 可知, 随机变量 X 和 Y 的边缘密度函数分别为

$$f_X(x) = 2(1-x)I_{(0,1)}(x), \quad f_Y(y) = 2yI_{(0,1)}(y).$$

再通过分别计算 U 和 V 的概率密度函数, 即可得结论.

(3) 服从. 由 $f_{Y|X}(y|x) = \frac{1}{1-x}I_{(x,1)}(y)$ 可知 Y 的条件分布为 $U(x, 1)$.

(4) 简单计算知

$$EX = \frac{1}{3}, \quad EY = \frac{2}{3}, \quad \text{Var}(X) = \text{Var}(Y) = \frac{1}{18}, \quad E(XY) = \frac{1}{4},$$

故由定义可知

$$\rho_{X,Y} = \frac{\frac{1}{4} - \frac{1}{3} \cdot \frac{2}{3}}{\frac{1}{18}} = \frac{1}{2}.$$

三、 每小题 6 分.

(1) $P(M \leq x|Y = n) = P(U_i \leq x, i = 1, \dots, n) = x^n$.

(2) 由 (1) 及全概率公式可知, 对任一实数 $0 \leq x \leq 1$,

$$P(M \leq x) = \sum_{n=1}^{\infty} P(M \leq x|Y = n)P(Y = n) = \sum_{n=1}^{\infty} \frac{x^n}{(e-1)n!} = \frac{e^x - 1}{e - 1}.$$

对上式求导, 即得 M 的概率密度函数

$$f_M(x) = \frac{e^x}{e-1}I_{(0,1)}(x).$$

(取值范围出错扣 1 分.)

(3) 以 $[z]$ 表示不大于 z 的最大整数, 注意到 $0 < M < 1$, 则对任一 $z > 0$,

$$\begin{aligned} P(Z > z) &= P(X \geq [z] + 2) + P(X = [z] + 1, M \leq [z] + 1 - z) \\ &= \frac{(e-1)e^{-[z]-2}}{1-1/e} + (e-1)e^{-[z]-1} \cdot \frac{e^{[z]+1-z}-1}{e-1} \\ &= e^{-z}. \end{aligned}$$

由上可知, 随机变量 Z 的分布函数为 $F(z) = (1 - e^{-z})I_{(0,\infty)}(z)$, 即 Z 服从参数为 1 的指数分布.

四、 每小题 6 分.

(1) 由题意可知, 似然函数为

$$L(\sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^m (2\sqrt{\pi}\sigma)^n} \prod_{i=1}^m \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} \prod_{j=1}^n \exp\left\{-\frac{(y_j - \mu)^2}{4\sigma^2}\right\}.$$

故对数似然函数为

$$\ell(\sigma^2) = C - \frac{m+n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left(\sum_{i=1}^m (x_i - \mu)^2 + \frac{1}{2} \sum_{j=1}^n (y_j - \mu)^2 \right),$$

这里 C 表示一个与 σ^2 无关的常数. 对 $\ell(\sigma^2)$ 求一阶导数, 并令之为 0, 解方程可得 σ^2 的最大似然估计为

$$\hat{\sigma}^2 = \frac{1}{m+n} \left[\sum_{i=1}^m (X_i - \mu)^2 + \frac{1}{2} \sum_{j=1}^n (Y_j - \mu)^2 \right].$$

(2) 估计量 $\hat{\sigma}^2$ 为无偏的. 对任意 $1 \leq i \leq m$ 和 $1 \leq j \leq n$, 由方差的定义可知 $E[(X_i - \mu)^2] = \sigma^2$ 及 $E[(Y_j - \mu)^2] = 2\sigma^2$ 成立, 由此得 $E(\hat{\sigma}^2) = \sigma^2$, 故结论成立.

(3) 由 χ^2 分布的定义, 可知

$$\frac{(m+n)\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^m \left(\frac{x_i - \mu}{\sigma} \right)^2 + \sum_{j=1}^n \left(\frac{y_j - \mu}{\sqrt{2}\sigma} \right)^2 \sim \chi_{m+n}^2.$$

故 σ^2 的置信系数为 0.95 的置信区间为

$$\left[\frac{(m+n)\hat{\sigma}^2}{\chi_{m+n}^2(0.025)}, \frac{(m+n)\hat{\sigma}^2}{\chi_{m+n}^2(0.975)} \right].$$

五、 每小题 9 分. 注意 H_0, H_1 的设置(2 分), 检验统计量的选取(2 分), 数值计算(2 分), 分位数的使用(2 分), 决策结果(1 分)各步骤是否正确.

先做一些计算. 男性总体: 样本均值 $\bar{x} = 14.95$, 样本方差 $s_1^2 = 6.84^2$, 样本容量 $n_1 = 13$; 女性总体: $\bar{y} = 22.29$, $s_2^2 = 5.32^2$, $n_2 = 10$; 混合方差 $s_w^2 = 6.24^2$.

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

$$\frac{1}{3.87} = \frac{1}{F_{9,12}(0.025)} < \frac{s_1^2}{s_2^2} = 1.65 < F_{12,9}(0.025) = 3.44,$$

接受 H_0 , 即可以认为两个总体的方差相等.

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

$$t = \frac{\bar{x} - \bar{y}}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -2.80 < -t_{21}(0.05) = -1.72,$$

拒绝 H_0 , 即可认为成年女性的体脂率显著地大于男性的体脂率.