中 国 科 学 技 术 大 学

2023—2024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: 2024 年 1 月 17 日上午 8:30–10:30; 可使用简单计算器

装 …………… 订 …………… 线

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 $\mathrm{P}(A)=0.7, \mathrm{P}(B)=0.4, \mathrm{P}(A|B)=0.5$, 则 $\mathrm{P}(B|A\cup\bar{B})=$__________.

2. 下述表述正确的是( )

(A) 分布函数连续的随机变量即为连续型随机变量

(B) 将一个随机变量加上一个常数则熵会增加

(C) 一个参数的 95% 置信区间为 [0.1,0.3], 则该参数落入该区间的概率是 0.95

(D) 一个假设检验问题中得到的 p 值为 0.02, 则犯第一类错误的概率至多为 0.02

3. 设随机变量 X, Y 的密度函数分别为 $f(x), g(x)$ 且均连续, 而对应的分布函数分别为 $F(x), G(x)$, 则下列中一定为密度函数的是( )

(A) $f(x)g(x)$ (B) $2f(x)G(x)$ (C) $F(x)g(x)$ (D) $f(x)G(x)+F(x)g(x)$

4. 设随机变量 X, Y 相互独立且均服从参数为 λ 的 Poisson 分布, 若记 $S=X+Y$, 则 $\mathrm{Var}[\mathrm{E}(S|X)]=$__________.

5. 如果随机变量序列$\{X_n\}$依分布收敛到随机变量X, 则下述表述正确的是( )

(A) $\lim\limits_{n\to\infty} X_n = X$ (B) $\lim\limits_{n\to\infty} \mathrm{P}(X_n\le x)=\mathrm{P}(X\le x)$, $\forall x\in\mathbb{R}$

(C) $X_n+\frac{1}{n}$ 依分布收敛到 X (D) $\lim\limits_{n\to\infty}\mathrm{P}(|X_n-X|\le\varepsilon)=1$, $\forall\varepsilon>0$

6. 下述表述正确的是( )

(A) 两个正态分布随机变量之和服从正态分布

(B) t_{30} 分布可近似为标准正态分布

(C) 标准正态分布的尾部比 t 分布的尾部高

(D) 标准正态分布密度的峰比 t 分布密度的峰要低

7. 下述表述错误的是( )

(A) 矩估计量一般不唯一 (B) 无偏估计总是优于有偏估计

(C) 相合性是一个估计量应具有的性质 (D) 最大似然估计可以不存在

8. 设 $X_1, X_2, \cdots, X_n$ 为来自均匀总体 $U(0,\theta)$ 的简单随机样本, 其中 $\theta>0$ 为未知参数. 记 $X_{(n)}=\max\{X_1,X_2,\cdots,X_n\}$, 若使损失函数 $h(c)=\mathrm{E}[(cX_{(n)}-\theta)^2]$ 最小, 则 $c=$__________.

9. 设正态总体 $N(\mu,\sigma^2)$ 的方差 σ^2 已知, 若样本容量 n 和置信水平不变, 则对不同的观测样本, 参数 μ 的置信区间长度______________ (填 "保持不变"或"会改变").

10. 设 $X_1,\ldots,X_n$ 为来自正态总体 $N(\mu,1)$ 的简单样本, 考虑假设检验问题 $H_0:\mu=0 \leftrightarrow H_1:\mu=0.5$. 如果要求检验犯第一类和第二类错误的概率均不超过 0.05, 则样本量 n 至少为__________.

二、 (20分) 设 $n\ (n \geq 3)$ 维随机向量 $(X_1, X_2, \cdots, X_n)$ 的密度函数为

$$f(x_1, x_2, \cdots, x_n) = 1 + \prod_{i=1}^{n} x_i, \quad -0.5 \leq x_i \leq 0.5, \quad i = 1, 2, \cdots, n.$$

(1) 对任意 $1 \leq k \leq n$, 试求 X_k 的边缘分布.

(2) 试求概率 $\mathrm{P}(X_1 > 0, X_2 > 0, \cdots, X_n > 0)$.

(3) 对任一整数 $2 \leq m < n$, 证明: $X_1, \cdots, X_m$ 相互独立, 但 $X_1, \cdots, X_n$ 不相互独立.

(4) 设随机向量 $\boldsymbol{X}^{(1)} = (X_1, \cdots, X_m), \boldsymbol{X}^{(2)} = (X_{m+1}, \cdots, X_n), 1 \leq m < n$. 对给定的常数 $-0.5 \leq x_{m+1}, \cdots, x_n \leq 0.5$, 证明在条件 $\boldsymbol{X}^{(2)} = (x_{m+1}, \cdots, x_n)$ 下, $\boldsymbol{X}^{(1)}$ 的条件密度函数 $f(x_1, \cdots, x_m | x_{m+1}, \cdots, x_n)$ 与 $f(x_1, x_2, \cdots, x_n)$ 具有相同表达式.

三、 (15分) 设随机变量 X 和 Y 相互独立且均服从正态分布 $\mathcal{N}(0, \sigma^2)$. 记随机变量 $U = (X^2 + Y^2)/\sigma^2$ 及 $V = |Y|/X$. 试求 (U, V) 的联合密度函数, 指出 U 和 V 各自服从的具体分布, 并证明两者相互独立.

四、 (15分) 设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的简单随机样本, 且 X 的密度函数为 $f(x) = \frac{1}{\sigma} e^{-(x-\theta)/\sigma} I_{[\theta, \infty)}(x)$, 其中 $\sigma > 0$ 为一已知的常数, 而 θ 为未知参数.

(1) 试求 θ 的矩估计 $\hat{\theta}$ 和最大似然估计 $\tilde{\theta}$.

(2) 问 $\hat{\theta}$ 和 $\tilde{\theta}$ 是否为 θ 的无偏估计? 若是, 请证明之; 若否, 请修正之.

(3) 试求常数 b, 使对任意 $x \in \mathbb{R}$, 均有 $\lim\limits_{n\to\infty} \mathrm{P}(\sqrt{n}(\hat{\theta} - \theta)/b \leq x) = \Phi(x)$ 成立, 其中 $\Phi(x)$ 为标准正态分布函数.

五、 (12分) 某种内服药有使病人血压增高的副作用, 且血压增高值的分布为 $N(22, 84.64)$. 现研制出一种新药, 通过测试 10 名服用新药病人的血压, 发现血压增高的样本均值和样本方差分别为 17.9 和 25.4. 在检验水平 $\alpha = 0.05$ 下,

(1) 通过比较均值, 所测数据能否支持"新药的副作用显著变小"这一结论?

(2) 所测数据能否支持"新药的方差显著变小"这一结论?

六、 (8分) 英国女作家Jane Austen(1775–1817)的作品有*Sense and Sensibility*, *Pride and Prejudice* 和 *Emma* 等, 她哥哥在她去世后主持了遗作*Persuasion* 和 *Northanger Abbey* 两部作品出版. 下面表格收集了*Sense and Sensibility*, *Emma* 和 *Persuasion* 三部作品前两章中常用代表词的出现频数, 根据你所学统计知识, 我们能否认为这三部作品在选择这些常用词的习惯没有差异? (检验水平 $\alpha = 0.05$)

单词	*Sense and Sensibility*	*Emma*	*Persuasion*
a	147	186	184
an	25	26	40
this	32	39	30
that	94	105	59

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

上分位数: $t_9(0.025) = 2.2622$, $t_9(0.05) = 1.8331$,

$\chi^2_6(0.05) = 12.592$, $\chi^2_6(0.95) = 1.635$, $\chi^2_9(0.05) = 16.919$, $\chi^2_9(0.95) = 3.325$,

参考答案

一、 每小题 3 分.

[1–5] $\frac{1}{4}$; D; D; λ; C;

[6–10] B;B/C; $\frac{n+2}{n+1}$; 保持不变; 44.

二、 每小题 5 分.

(1) 对任意 $1 \le k \le n$, 由

$$f_k(x_k) = \int_{-1/2}^{1/2} \cdots \int_{-1/2}^{1/2} \Big(1+\prod_{i\neq k} x_i\Big) \mathrm{d}x_1 \cdots \mathrm{d}x_{k-1}\mathrm{d}x_{k+1} \cdots \mathrm{d}x_n = 1, \quad -\frac{1}{2} \le x_k \le \frac{1}{2},$$

知 $X_k \sim U(-1/2, 1/2)$, 即 X_k 服从区间 $(-1/2, 1/2)$ 上的均匀分布.

(2) 由密度函数的基本性质可知

$$\begin{aligned}\mathrm{P}(X_1 > 0, X_2 > 0, \cdots, X_n > 0) &= \int_0^{1/2} \cdots \int_0^{1/2} \Big(1 + \prod_{i=1}^{n} x_i\Big) \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \frac{1}{2^n} + \left(\int_0^{1/2} x\mathrm{d}x\right)^n = \frac{1}{2^n} + \frac{1}{8^n} = \frac{4^n+1}{8^n}.\end{aligned}$$

(3) 对 $2 \le m < n$, 类似 (1) 可知 $X_1, X_2, \cdots, X_m$ 的联合密度函数为

$$f_{1,2,\cdots,m}(x_1, x_2, \cdots, x_m) = 1, \quad -\frac{1}{2} \le x_1, x_2, \cdots, x_m \le \frac{1}{2}.$$

再由 (1) 可知 $X_1, X_2, \cdots, X_m$ 相互独立. 而当 $x_1, x_2, \cdots, x_n$ 均不为 0 时, 联合密度函数 $f(x_1, x_2, \cdots, x_n) \neq 1$, 从而知 $X_1, X_2, \cdots, X_n$ 不相互独立.

(4) 设 $f_{m+1,\cdots,n}(x_{m+1}, \cdots, x_n)$ 为 $\boldsymbol{X}^{(2)}$ 的边缘密度函数, 则类似于 (3) 中结论可知 $f_{m+1,\cdots,n}(x_{m+1}, \cdots, x_n) = 1$, $-1/2 \le x_{m+1}, \cdots, x_n \le 1/2$. 故由条件密度基本公式

$$f(x_1, \cdots, x_m | x_{m+1}, \cdots, x_n) = \frac{f(x_1, x_2, \cdots, x_n)}{f_{m+1,\cdots,n}(x_{m+1}, \cdots, x_n)},$$

可知结论成立.

三、 记 $Z = |Y|$, 函数 $u = \frac{1}{\sigma^2}(x^2 + z^2), v = z/x$. 注意到 $(x, z) \longmapsto (u, v)$ 为一一映射, 且 Jacobi 行列式为

$$J^{-1} = \left|\frac{\partial(u, v)}{\partial(x, z)}\right| = \begin{vmatrix} 2x/\sigma^2 & 2z/\sigma^2 \\ -z/x^2 & 1/x \end{vmatrix} = \frac{2}{\sigma^2}\Big(1 + \frac{z^2}{x^2}\Big) = \frac{2(1+v^2)}{\sigma^2}.$$

由 (X, Z) 的联合密度函数为 $f(x, z) = \frac{1}{\pi\sigma^2} e^{-(x^2+z^2)/(2\sigma^2)}$, $-\infty < x < \infty, z > 0$ 及密度变换公式可知 (U, V) 的联合密度为

$$g(u, v) = f(x, z)|J| = \frac{1}{2\pi(1+v^2)} e^{-u/2}, \quad u > 0, -\infty < v < \infty.$$

(注: 变量取值范围不写或写错扣 1 分.) 由 $g(u,v)$ 可分离变量知, U 服从参数为 1/2 的指数分布 (或自由度为 2 的 χ^2 分布), V 服从 Cauchy 分布, 且两者相互独立.

四、 每小题 5 分.

(1) 由 $\mathrm{E}X = \theta + \sigma$ 可知所求矩估计 $\hat{\theta} = \overline{X} - \sigma$, 其中 $\overline{X}$ 为样本均值. 再由似然函数

$$L(\theta) = \frac{1}{\sigma^n} \exp\Big\{ -\sum_{i=1}^{n} \frac{x_i - \theta}{\sigma} \Big\} I_{[\theta,\infty)}(x_1, x_2, \cdots, x_n)$$

可知最大似然估计 $\tilde{\theta} = X_{(1)} = \min\{X_1, \cdots, X_n\}$.

(2) 通过计算知 $\mathrm{E}(\hat{\theta}) = \theta$ 及 $\mathrm{E}(\tilde{\theta}) = \theta + \sigma/n$, 故 $\hat{\theta}$ 是 θ 的无偏估计, 而 $\tilde{\theta}$ 则不是, 可修正为 $\tilde{\theta}^* = X_{(1)} - \sigma/n$.

(3) 由 $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \sigma)$ 及独立同分布场合下的中心极限定理可知

$$b/\sqrt{n} = \sqrt{\mathrm{Var}(\hat{\theta})} = \sigma/\sqrt{n},$$

故 $b = \sigma$.

五、 每小题 6 分. 注意 H_0, H_1 的设置(1 分), 检验统计量的选取(1 分), 数值计算(1 分), 分位数的使用(1 分), 决策结果(2 分)各步骤是否正确.

(1) $H_0: \mu \geq \mu_0 = 22$ (或 $\mu = \mu_0$) $\leftrightarrow$ $H_1: \mu < \mu_0$. 由检验统计量

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{17.9 - 22}{\sqrt{25.4/10}} = -2.57 < -t_9(0.05) = -1.8331,$$

故拒绝 H_0, 即可以认为新药的副作用显著变小.

(2) $H_0: \sigma^2 \geq \sigma_0^2 = 84.64$ (或 $\sigma^2 = \sigma_0^2$) $\leftrightarrow$ $H_1: \sigma^2 < \sigma_0^2$. 由检验统计量

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \times 25.4}{84.64} = 2.7 < \chi_9^2(0.95) = 3.325,$$

故拒绝 H_0, 即可以认为新药的方差显著变小.

六、 齐次性检验. 原假设 H_0: 这三部作品在选择这些常用词的习惯没有差异. 检验统计量为

$$\chi^2 = \sum_{i=1}^{4}\sum_{j=1}^{3} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n}.$$

在原假设成立条件下, 该统计量的极限分布是自由度为 $(4-1)\times(3-1) = 6$ 的卡方分布. 代入数据计算可得

$$\chi^2 = 19.722 > \chi^2_{0.05}(6) = 12.592,$$

故在 $\alpha = 0.05$下, 我们可以拒绝原假设, 即认为这三部作品在选择常用词的习惯上有显著差异.