中 国 科 学 技 术 大 学

2017—2018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: 2018年1月10日上午8:30–10:30;　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) 设随机事件A和B相互独立, A和C相互独立, 且B和C互斥. 若$\mathrm{P}(A) = \mathrm{P}(B) = 1/2$, $\mathrm{P}(AC|AB \cup C) = 1/4$, 则$\mathrm{P}(C) =$ ______.

(2) 一只蚂蚁从等边三角形$\triangle ABC$的顶点A出发开始沿着边爬行, 设它每次爬行到一个顶点后, 会休憩片刻再随机选择一条边继续爬行, 则第n次爬行是往A爬的概率为________.

(3) 设连续型随机变量X的密度函数$f(x)$满足$f(1+x) = f(1-x)$, 且$\int_0^2 f(x)\mathrm{d}x = 0.4$, 则$\mathrm{P}(X < 0) =$ (　)

(A) 0.2　　(B) 0.3　　(C) 0.4　　(D) 0.5

(4) 设随机变量X的分布函数为$F(x) = 0.5\Phi(x) + 0.5\Phi(\frac{x-4}{2})$, 其中$\Phi(x)$为标准正态分布函数, 则$X$的数学期望$\mathrm{E}X =$ ______.

(5) 设随机变量X与Y相互独立, X的概率分布为$\mathrm{P}(X = 1) = \mathrm{P}(X = -1) = 1/2$, Y服从参数为$\lambda > 0$的Poisson分布. 若记$Z = XY$, 则$\mathrm{Cov}(X, Z) =$ ______.

(6) 设将1米长的木棒随机截成两段, 其中一段的长度记为X, 另一段长度的1/3记为Y, 则X与Y的相关系数为(　)

(A) 1　　(B) -1　　(C) $-1/3$　　(D) $1/3$

(7) 设$X_1, X_2, \cdots, X_9$是来自总体$X \sim N(0,1)$的一组简单随机样本, 则下列统计量中服从F分布的是(　)

(A) $\frac{X_1^2+X_2^2+X_3^2}{X_4^2+X_5^2+\cdots+X_9^2}$　(B) $\frac{X_1^2+X_2^2+X_3^2+X_4^2+X_5^2}{X_5^2+X_6^2+X_7^2+X_8^2+X_9^2}$　(C) $\frac{X_1^2+X_2^2+X_3^2}{2(X_4^2+X_5^2+\cdots+X_9^2)}$　(D) $\frac{2(X_1^2+X_2^2+X_3^2)}{X_4^2+X_5^2+\cdots+X_9^2}$

(8) 设$X_1, X_2, \cdots, X_n$是来自总体X的一组简单随机样本, 以$\overline{X}$和S^2分别表示样本均值和样本方差. 若记$\mathrm{Var}(X) = \sigma^2$, 则(　)

(A) S是σ的无偏估计量　　(B) S是σ的极大似然估计量

(C) S与$\overline{X}$相互独立　　(D) 以上均不对

(9) 设来自总体$X \sim N(\mu, 0.9^2)$一组容量为9的简单随机样本, 其样本均值$\overline{X} = 5$, 则未知参数μ的置信度为0.95的置信区间为____________(保留到小数点后三位).

(10) 设$X_1, X_2, \cdots, X_n$是来自正态总体$N(\mu, \sigma^2)$的一组简单随机样本, 据此样本做假设检验$H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$, 其中$\mu_0$是给定的已知常数, 则(　)

(A) 如果在检验水平$\alpha = 0.05$下接受H_0, 那么在检验水平$\alpha = 0.01$下必接受H_0

(B) 如果在检验水平$\alpha = 0.05$下拒绝H_0, 那么在检验水平$\alpha = 0.01$下必接受H_0

(C) 如果在检验水平$\alpha = 0.05$下接受H_0, 那么在检验水平$\alpha = 0.01$下必拒绝H_0

(D) 如果在检验水平$\alpha = 0.05$下拒绝H_0, 那么在检验水平$\alpha = 0.01$下必拒绝H_0

二. (16分)设二维随机向量(X,Y)的联合密度函数为

$$f(x,y)=C\mathrm{e}^{-2x^2+2xy-y^2},\quad -\infty<x,y<\infty.$$

(1) 求常数C的值;

(2) 在$X=x$的条件下, 求Y的条件密度$f_{Y|X}(y|x)$.

三. (16分)设二维随机向量(X,Y)服从二元正态分布$N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 其中$\mu_1=\mu_2=1$, $\sigma_1^2=\sigma_2^2=0.5$, $\rho=0.5$. 记

$$Z=|X-Y|,\quad U=\max(X,Y),\quad V=\min(X,Y).$$

(1) 求Z的密度函数$f_Z(z)$;

(2) 求数学期望$\mathrm{E}(U+V)$;

(3) 分别求数学期望$\mathrm{E}U$和$\mathrm{E}V$.

四. (18分)设总体X的密度函数为

$$f(x)=\frac{2x}{a^2},\quad 0\le x\le a,$$

其中$a>0$为未知参数, 而$X_1,X_2,\cdots,X_n$是来自该总体的一组简单随机样本.

(1) 求a的矩估计量$\hat{a}_1$和极大似然估计量$\hat{a}_2$;

(2) 求$p=\mathrm{P}(0<X<\sqrt{a})$的极大似然估计量$\hat{p}$;

(3) 问$\hat{a}_1$和$\hat{a}_2$是否为无偏估计? 若是, 请证明你的结论; 若不是, 请修正之.

五. (10分) 为了检验某种体育锻炼对减肥的效果, 随机抽取了10名减肥者进行测试. 在进行体育锻炼前后这些减肥者的体重(单位:千克)数据列表如下, 问该体育锻炼方法对降低体重是否具有显著性(设人的体重服从正态分布, 取显著性水平α=0.05)?

锻炼前体重	70	65	67	58	69	72	74	61	63	67
锻炼后体重	68	60	68	58	67	70	70	60	60	65

六. (10分)上海证券综合指数简称“上证指数”, 反映了上海证券交易所上市股票价格的变动情况. 自上证指数诞生的二十七年(1991年1月至2017年12月)以来, 所有月份上涨或下跌的情况如下:

月份	一	二	三	四	五	六	七	八	九	十	十一	十二
上涨月数	14	21	16	15	14	14	13	15	11	13	18	13
下跌月数	13	6	11	12	13	13	14	12	16	14	9	14

结合你所学的知识, 我们能否认为上证指数的涨跌与月份有关?

附录: 上分位数表

$u_{0.025}=1.96,\ u_{0.05}=1.645;$

$t_8(0.025)=2.306,\ t_8(0.05)=1.86,\ t_9(0.025)=2.262,\ t_9(0.05)=1.833;$

$\chi^2_{11}(0.05)=19.675.$

参考答案

一. (每小题3分)

$\frac{1}{4}$; $\frac{1}{3}[1-(-\frac{1}{2})^{n-1}]$; B; 2; λ; B; D; C; $[4.412, 5.588]$; A.

二. (1) (8分) 由

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-2x^2+2xy-y^2}\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\mathrm{d}x \int_{-\infty}^{\infty} \mathrm{e}^{-(x-y)^2}\mathrm{d}y = \pi$$

可知$C=\frac{1}{\pi}$;

(2) (8分) 由于X的边缘密度函数为

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y = \frac{1}{\sqrt{\pi}}\mathrm{e}^{-x^2},$$

从而,

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

三. (1) (6分) 由$\mathrm{E}(X-Y)=0$,

$$\begin{aligned}\mathrm{Var}(X-Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)\\ &= 0.5+0.5-2\times 0.25 = 0.5,\end{aligned}$$

及二元正态分布的性质可知$X-Y\sim N(0,0.5)$, 从而$Z=|X-Y|$的密度函数为

$$f_Z(z) = \frac{2}{\sqrt{\pi}}\mathrm{e}^{-z^2}, \quad z>0.$$

(2) (4分) 易知, $\mathrm{E}(U+V)=\mathrm{E}(X+Y)=2$.

(3) (6分) 由$\mathrm{E}U-\mathrm{E}V=\mathrm{E}Z=\frac{1}{\sqrt{\pi}}$, 可知$\mathrm{E}U=1+\frac{1}{2\sqrt{\pi}}$, $\mathrm{E}V=1-\frac{1}{2\sqrt{\pi}}$.

四. (1) (6分) 矩估计量$\hat{a}_1=\frac{3}{2}\overline{X}$, 极大似然估计量$\hat{a}_2=X_{(n)}$;

(2) (4分) 由$p=\frac{1}{a}$知其极大似然估计量为$\hat{p}=1/X_{(n)}$;

(3) (8分) 矩估计$\hat{a}_1$是无偏的, 因$\mathrm{E}(\hat{a}_1)=\frac{3}{2}\mathrm{E}(\overline{X})=\frac{3}{2}\mathrm{E}(X)=a$; 而由$X_{(n)}$的密度函数为

$$h(x) = n[F(x)]^{n-1}f(x) = \frac{2n}{a^{2n}}x^{2n-1}, \quad 0<x<a,$$

知$\mathrm{E}(\hat{a}_2)=\frac{2n}{2n+1}a$. 故$\hat{a}_2$不是无偏估计, 可修正为$\hat{a}_2^*=\frac{2n+1}{2n}X_{(n)}$.

五. (10分) 成对数据. 首先可算得相减之后, 有$\overline{X}=2, S^2=28/9$. 故由

$$t=\frac{\sqrt{n}\overline{X}}{S} = 3.59 > t_9(0.05) = 1.833,$$

可拒绝原假设(H_0: 锻炼前后体重无显著变化), 即认为该体育锻炼方法对降低体重具有显著性.

六. (10分) 列联表齐一性检验. 两行的和分别为177和147, 每列之和均为27. 由此可算得χ^2统计量的值为$11.394<\chi^2_{11}(0.05)=19.675$, 故可认为“无充分证据表明上证指数的涨跌与月份有关”或“上证指数的涨跌与月份无关”.

中 国 科 学 技 术 大 学

2018—2019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: 2019 年1月9日上午8:30–10:30; 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) 已知 10 台洗衣机中有 7 台一等品和 3 台二等品, 现已售出一台, 在余下的 9 台中随机抽取 2 台后发现均为一等品, 则原先售出的一台为二等品的概率为_____.

(2) 设随机变量 X 的概率密度 $f(x) = A\mathrm{e}^{-x^2+x}$, $-\infty < x < \infty$, 则常数 $A =$ _____.

(3) 设随机变量 X 与 Y 相互独立, 且它们的取值范围分别是 $\{1,2\}$ 和 $\{1,2,3\}$. 已知

$$\mathrm{P}(Y=1) = \frac{1}{6}, \quad \mathrm{P}(X=1, Y=2) = \mathrm{P}(Y=1, X=2) = \frac{1}{8},$$

则 $\mathrm{P}(Y=3) =$ ______.

(4) 设随机变向量 (X,Y) 的分布函数为 $\Phi(2x)\Phi(y-1)$, 其中 $\Phi(x)$ 为标准正态分布函数, 则 (X,Y) 服从二元正态分布( )

(A) $N(0,1;\frac{1}{4},1;0)$ (B) $N(0,-1;\frac{1}{4},1;0)$ (C) $N(0,1;4,1;0)$ (D) $N(0,-1;4,1;0)$

(5) 设随机变量 X 与 Y 相互独立, 且 X 服从参数为 2 的泊松分布, Y 服从区间 $[-3,3]$ 上的均匀分布, 则它们的乘积的方差 $\mathrm{Var}(XY) =$ ______.

(6) 设 X_1, X_2, X_3, X_4 是来自标准正态总体的简单随机样本, $a>0$为某个常数. 若已知

$$Y = a\Big(\frac{1}{2}X_1^2 + \frac{1}{2}X_2^2 + \frac{1}{2}X_3^2 + \frac{1}{2}X_4^2 + X_1X_2 + X_3X_4\Big)$$

服从 χ_n^2 分布, 则 $n + a =$ ____.

(7) 已知随机变量 X 服从 $F_{3,4}$ 分布. 设对给定的 $\alpha(0<\alpha<1)$, 实数 $F_{3,4}(\alpha)$ 满足 $\mathrm{P}(X > F_{3,4}(\alpha)) = \alpha$. 若有 $\mathrm{P}(X \le x) = 1-\alpha$, 则 x 等于( )

(A) $\frac{1}{F_{4,3}(1-\alpha)}$ (B) $\frac{1}{F_{3,4}(1-\alpha)}$ (C) $F_{4,3}(\alpha)$ (D) $F_{4,3}(1-\alpha)$

(8) 设 $X_1, X_2, \cdots, X_n$ 是来自均匀总体 $U[-\theta,\theta]$ 的简单随机样本, 则参数 θ 的极大似然估计量 $\hat{\theta}$ 为( )

(A) $\max\limits_{1\le i\le n} X_i$ (B) $\max\limits_{1\le i\le n} |X_i|$ (C) $-\min\limits_{1\le i\le n} X_i$ (D) $-\min\limits_{1\le i\le n} |X_i|$

(9) 设 $X_1, X_2, \cdots, X_n$ $(n>2)$ 是来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本, 且 $\overline{X}$ 为样本均值. 若统计量 $T = c(X_1 + X_n - 2\overline{X})^2$ 为 σ^2 的无偏估计, 则常数 $c =$ ____.

(10) 已知两个正态总体 X_1 和 X_2 分别为 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$, 为了检验总体 X_1 的均值大于 X_2 的均值, 则应作检验的假设为( )

(A) $H_0: \mu_1 > \mu_2 \leftrightarrow H_1: \mu_1 \le \mu_2$ (B) $H_0: \mu_1 \ge \mu_2 \leftrightarrow H_1: \mu_1 < \mu_2$

(C) $H_0: \mu_1 < \mu_2 \leftrightarrow H_1: \mu_1 \ge \mu_2$ (D) $H_0: \mu_1 \le \mu_2 \leftrightarrow H_1: \mu_1 > \mu_2$

二、 (24分) 设二维随机向量 (X,Y) 的联合密度函数为

$$f(x,y)=\frac{1}{5}(2x+y),\quad 0\le x\le 2, 0\le y\le 1.$$

(1) 分别求 X 和 Y 的边缘密度函数 $f_X(x)$ 和 $f_Y(y)$;

(2) 在给定 $X=1$ 的条件下, 求 Y 在点 $y=0.5$ 处的概率密度 $f_{Y|X}(0.5\,|\,1)$;

(3) 求 X 和 Y 的协方差 $\mathrm{Cov}(X,Y)$;

(4) 求随机变量 $Z=\max\{X,Y\}$ 的密度函数 $f_Z(z)$.

三、 (12分) 某药厂试制了一种新药, 声称对贫血患者的治疗有效率达到 80%. 医药监管部门随机抽取 200 个贫血患者进行此药的临床试验, 若至少有 152 人用药有效, 就批准此药的生产. 试利用中心极限定理, 求解如下问题:

(1) 若该药的有效率确实达到 80%, 此药被批准生产的概率大约是多少?

(2) 若监管部门的方案是 200 个人中要有 160 人用药有效才批准, 这对药厂是否公平? 需说明理由.

四、 (18分) 已知总体 X 的密度函数为

$$f(x;\theta)=\begin{cases} -\theta^x\ln\theta, & x>0;\\ 0, & x\le 0,\end{cases}$$

其中 $0<\theta<1$ 为未知参数. 设 $X_1,X_2,\cdots,X_n$ 为来自该总体的一组简单随机样本.

(1) 求 θ 的矩估计量 $\widehat{\theta}$;

(2) 求 $h(\theta)=(\ln\theta)^{-1}$ 的极大似然估计量 $\widehat{h}_\theta$;

(3) 试求实数 a, 使得 $\widehat{h}_\theta$依概率收敛到 a, 即对任何$\varepsilon>0$, 都有 $\lim\limits_{n\to\infty}\mathrm{P}\left(\left|\widehat{h}_\theta-a\right|\ge\varepsilon\right)=0$.

五、 (8分) 为比较A和B两种型号步枪子弹的枪口速度, 随机地抽取A型子弹 10 发, 得到枪口速度的平均值为 $\overline{x}=500(\mathrm{m/s})$, 样本标准差 $s_1=1.10(\mathrm{m/s})$; 随机地抽取B型子弹 20 发, 得到枪口速度的平均值为 $\overline{y}=496(\mathrm{m/s})$, 样本标准差 $s_2=1.20(\mathrm{m/s})$. 假设A和B型号子弹的枪口速度分别近似服从方差相等的正态分布 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$. 试在显著性水平 $\alpha=0.05$下检验假设 $H_0:\mu_1-\mu_2=5\leftrightarrow H_1:\mu_1-\mu_2\ne 5$.

六、 (8分) 某机构为了研究鼻咽癌是否与血型有关, 随机调查了一些患者和健康人, 得到的数据如下:

	A	B	O	AB
患者	64	86	130	20
健康人	125	138	210	26

请你根据所学的统计知识给出适当的结论(显著性水平设为 $\alpha=0.05$).

附录: 上分位数表

$u_{0.025}=1.96,\ u_{0.05}=1.645,\ \Phi(1.414)=0.9214;$

$t_{28}(0.025)=2.0484,\ t_{28}(0.05)=1.7011,\ t_{29}(0.025)=2.0452,\ t_{29}(0.05)=1.6991;$

$\chi^2_3(0.95)=0.3518,\ \chi^2_3(0.05)=7.8147.$

参考答案

一. (每小题3分)

3/8; $\frac{1}{\sqrt{\pi}}\mathrm{e}^{-\frac{1}{4}}$; 1/3; A; 18; 3; A; B; $\frac{n}{2(n-2)}$; D.

二. (1) (6分) 由边缘密度和联合密度的关系可知,

$$\begin{aligned} f_X(x) &= \int_0^1 f(x,y)dy = \frac{2}{5}x + \frac{1}{10}, \quad 0 \le x \le 2; \\ f_Y(y) &= \int_0^2 f(x,y)dx = \frac{2}{5}y + \frac{4}{5}, \quad 0 \le y \le 1. \end{aligned}$$

(2) (6分) 由边缘密度、条件密度和联合密度的关系可知,

$$f_{Y|X}(0.5\,|\,1) = \frac{f(1,0.5)}{f_X(1)} = 1.$$

(3) (6分) 由

$$\mathrm{E}[XY] = \int_0^1\int_0^2 xyf(x,y)dxdy = \frac{2}{3},$$

及 $\mathrm{E}X = \frac{19}{15}$ 和 $\mathrm{E}Y = \frac{8}{15}$, 可知 $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}X\mathrm{E}Y = -\frac{2}{225}$.

(4) (6分) 由

$$F_Z(z) = \mathrm{P}(Z \le z) = \iint\limits_{x,y\le z} f(x,y)dxdy$$

可知,

$$F_Z(z) = \begin{cases} 0, & z < 0; \\ \frac{3}{10}z^3, & 0 \le z < 1; \\ \frac{1}{5}z^2 + \frac{1}{10}z, & 1 \le z < 2; \\ 1, & z \ge 2. \end{cases}$$

从而, 所求密度函数为

$$f_Z(z) = \begin{cases} \frac{9}{10}z^2, & 0 \le z < 1; \\ \frac{2}{5}z + \frac{1}{10}, & 1 \le z < 2. \end{cases}$$

三. (1) (6分) 所求概率为 $\Phi(\sqrt{2}) = 0.9214$.

(2) (6分) 不公平. 对药厂而言, 在治疗有效率达到80%的情况下被批准的概率大约为 $\Phi(0) = 0.5$, 这相当于用掷硬币的方式来决定是否得到批准.

四. (1) (6分) 矩估计量 $\widehat{\theta} = \exp\{-\frac{1}{\overline{X}}\}$, 其中 $\overline{X}$ 为样本均值;

(2) (6分) 参数 θ 的极大似然估计量同样为 $\exp\{-\frac{1}{\overline{X}}\}$, 从而 $h(\theta)$ 的极大似然估计量为 $\widehat{h}_\theta = -\overline{X}$;

(3) (6分) 由弱大数律可知, 所求的实数 $a = -\mathrm{E}X = \frac{1}{\ln\theta}$.

五. (8分) 两样本 t 检验, 其检验统计量为

$$T = \frac{\overline{X} - \overline{Y} - \mu_0}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

代入数据计算可知, $s_w = 1.169, t = -2.209$. 由于 $|t| > t_{28}(0.025) = 2.0484$, 在显著性水平 $\alpha = 0.05$下我们应该拒绝原假设 H_0.

六. (8分) 拟合优度联列表检验. 原假设为鼻咽癌与血型无关, 而其检验统计量为

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

代入数据计算可知, $\chi^2 = 1.921 < \chi_3^2(0.05) = 7.8147$. 故在显著性水平 $\alpha = 0.05$下我们不能拒绝原假设, 即可以认为鼻咽癌与血型无关.

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

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 $\mathrm{P}(A)=\mathrm{P}(B)=0.4$, 且 $\mathrm{P}(B|A)+\mathrm{P}(\overline{B}|\overline{A})=1$, 则 $\mathrm{P}(AB)=$ ______.

(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

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

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 $\{(x,y):x^2+y^2\le 1\}$ 上随机取两个点, 以随机变量 X 表示它们之间的距离, 则 $\mathrm{E}(X^2)=$ ______.

(5) 设 $X_1,X_2,\cdots,X_n$ 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) $\lim\limits_{n\to\infty}\mathrm{P}\left(\frac{\sqrt{n}}{\lambda}(\overline{X}-\lambda)\le x\right)=\Phi(x)$　(B) $\lim\limits_{n\to\infty}\mathrm{P}\left(\sqrt{\frac{n}{\lambda}}(\overline{X}-\lambda)\le x\right)=\Phi(x)$

(C) $\lim\limits_{n\to\infty}\mathrm{P}\left(\sqrt{n}(\lambda\overline{X}-1)\le x\right)=\Phi(x)$　(D) $\lim\limits_{n\to\infty}\mathrm{P}\left(\sqrt{n\lambda}(\overline{X}-\frac{1}{\lambda})\le x\right)=\Phi(x)$

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

(7) 设 $X_1,X_2,\cdots,X_n$ 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 $X_1, X_2, \cdots, X_n$ 为一列独立的随机变量, 且均服从 $U(0,1)$ 分布. 记

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

(1) 试证明: 对任意常数 $0 < y, z < 1$, 有

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

(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 $\mathrm{Cov}(Y,Z)$.

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 $F(t) = \begin{cases} 1-\exp\{-(\frac{t}{\theta})^m\}, & t \ge 0; \\ 0, & t < 0, \end{cases}$ 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, 测得它们的寿命分别为 $T_1, T_2, \cdots, T_n$. 记 $g(\theta) = \theta^m$.

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

(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:

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

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

$\chi^2_5(0.95) = 1.145,\ \chi^2_5(0.05) = 11.071,\ \chi^2_{15}(0.975) = 6.262,\ \chi^2_{15}(0.025) = 27.488.$

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 $X_1,\cdots,X_n$ 独立同分布可知

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

再由 $X_1\sim U(0,1)$ 即知

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

(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 $\mathrm{P}(Y\le y,Z\le z)$ 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

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

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

$$\mathrm{E}Y=\int_0^1 2y(1-y)\mathrm{d}y=\frac{1}{3}.$$

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 $\mathrm{E}Z=\frac{2}{3}$. 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

$$\mathrm{E}[YZ]=\int_0^1\int_y^1 2yz\mathrm{d}z\mathrm{d}y=\frac{1}{4}.$$

所以,

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

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 $f(x,y,z)=\mathrm{e}^{-(x+y+z)},\ x,y,z>0$, 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

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

(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

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

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

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

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)

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

解之即可得所求极大似然估计量为

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

(2) 由

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

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

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

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

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. $H_0: \sigma^2 = 6^2, \quad \longleftrightarrow \quad H_1: \sigma^2 \neq 6^2$. 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 $\chi^2_{15}(0.975) = 6.262 < \chi^2 < \chi^2_{15}(0.025) = 27.488$, 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

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

代入数据计算可知, $\chi^2 = 3.1922 < \chi^2_5(0.05) = 11.071$, 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.

中 国 科 学 技 术 大 学

2020—2021学年第一学期考试试卷

考试科目 <u>概率论与数理统计(B)</u>　　　　得分 ____________

所在院系 _______　姓名 ____________　学号 ____________

考试时间: 2021 年 3 月 6 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 A, B 为随机事件, 且 $0 < \mathrm{P}(B) < 1$, 下列为假命题的是(　)

(A) 若 $\mathrm{P}(A|B) = \mathrm{P}(A)$, 则 $\mathrm{P}(A|\overline{B}) = \mathrm{P}(A)$

(B) 若 $\mathrm{P}(A|B) > \mathrm{P}(A)$, 则 $\mathrm{P}(\overline{A}|\overline{B}) > \mathrm{P}(\overline{A})$

(C) 若 $\mathrm{P}(A|B) > \mathrm{P}(A|\overline{B})$, 则 $\mathrm{P}(A|B) > \mathrm{P}(A)$

(D) 若 $\mathrm{P}(A|A\cup B) > \mathrm{P}(\overline{A}|A\cup B)$, 则 $\mathrm{P}(A) > \mathrm{P}(B)$

(2) 设平面上有 n 个点, 编号分别为 $1, 2, \cdots, n$. 现一质点在此点集上做随机游动, 每次它在一点上停留片刻后就会在其余各点中随机地选择一个并移动到该点上. 已知其初始位置为点 1, 则它在第一次返回点 1 之前访问过点 2 的概率是_______.

(3) 设随机变量 X 服从参数为 $0 < p < 1$ 的几何分布, 且在条件 $X = k$ 下, Y 服从参数为 k 的指数分布. 对任一实数 $y > 0$, 则 $\mathrm{P}(Y > y) =$(　).

(A) $\mathrm{e}^{-y/p}$　　(B) $p\mathrm{e}^{-y}$　　(C) $p/(y+p)$　　(D) $p/(\mathrm{e}^{y} - 1 + p)$

(4) 若随机变量 X 和 Y 满足 $\mathrm{P}(X^2 + Y^2 = 2) = 1$, 则下列说法中一定不成立的是(　)

(A) (X, Y) 为连续型随机向量　　(B) $\mathrm{P}(X + Y = 0) = 1/2$

(C) $\mathrm{E}X = \mathrm{E}Y = 0$　　(D) X 和 Y 相互独立

(5) 设 X 和 Y 为相互独立的标准正态随机变量, 则 $\mathrm{P}(\max\{X, Y\} \geq 0) =$_______.

(6) 设随机变量 X 和 Y 分别服从参数为 λ 和 μ 的 Poisson 分布, 且相互独立. 对任一非负整数 n, 则条件期望 $\mathrm{E}[X|X + Y = n] =$_______.

(7) 设随机向量 (X, Y) 服从二维正态分布 $N(a, a; \sigma^2, \sigma^2; 0)$, 则 $\mathrm{Cov}(X, XY^2) =$_______.

(8) 设 $X_1, X_2, \cdots, X_n, X_{n+1}$ 是来自总体 $N(\mu, \sigma^2)$ 的简单随机样本, 且 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. 已知 $T = c(X_{n+1} - \overline{X})/\sqrt{m_2}$ 服从 t 分布, 则 $c =$_______.

(9) 设正态总体 $X \sim N(\mu, \sigma^2)$, 其中 σ^2 未知, 若样本容量 n 和置信水平 $1 - \alpha$ 均保持不变, 对不同的样本观察值, 则总体均值 μ 的置信区间长度(　)

(A) 与样本均值有关　(B) 与 μ 本身有关　(C) 保持不变　(D) 不确定

(10) 设 $X_1, X_2, \cdots, X_{16}$ 是来自总体 $N(\mu, 4)$ 的简单随机样本, $\overline{X}$ 为其样本均值, $\Phi(x)$ 为标准正态分布函数. 考虑假设检验问题: $H_0 : \mu \leq 10 \leftrightarrow H_1 : \mu > 10$. 若其拒绝域为 $W = \{\overline{X} \geq 11\}$, 则当 $\mu = 11.5$ 时, 该检验犯第二类错误的概率为(　)

(A) $1 - \Phi(0.5)$　(B) $1 - \Phi(1)$　(C) $1 - \Phi(1.5)$　(D) $1 - \Phi(2)$

装　订　线

二、 (10分) 设有两个罐子, 一罐中有 m 个红球和 n 个黑球, 另一罐中有 n 个红球和 m 个黑球, 且 $m > n$. 某人随机选取一个罐子并从中随机抽取一球, 发现为红球. 现将该球放回原罐后并摇匀, 然后再次在此罐中随机抽取一球, 则它仍为红色的概率是否比 $1/2$ 大? 通过计算事件的概率来证明你的结论.

三、 (20分) 将区间 $(0, 2)$ 随机截成两段, 记较短一段的长度为 X, 较长一段的长度为 Y.

(1) 求 X 和 Y 的相关系数 $\mathrm{Corr}(X, Y)$;

(2) 求 X 的概率密度函数 $f(x)$;

(3) 求 $Z = \frac{Y}{X}$ 的概率密度函数 $g(z)$;

(4) 设随机变量 X^* 与 X 独立同分布, 试求 $V = 2|X - X^*|$ 的概率密度函数 $h(v)$.

四、 (16分) 已知总体 X 的概率密度函数为 $f(x) = (\theta + 1)x^\theta$, $0 < x < 1$, 其中 $\theta > -1$ 为一未知参数. 设 $X_1, X_2, \cdots, X_n$ 是来自该总体的一组简单随机样本.

(1) 求 θ 的矩估计量 $\hat{\theta}$;

(2) 求 $g(\theta) = \frac{1}{\theta+1}$ 的极大似然估计量 $\hat{g}$;

(3) 问 $\hat{g}$ 是否为 $g(\theta)$ 的一个无偏估计? 证明你的结论.

(4) 求常数 b, 使得对任意实数 x, 都有 $\lim\limits_{n\to\infty} \mathrm{P}\big(\sqrt{n}(\hat{g} - g(\theta))/b \le x\big) = \Phi(x)$ 成立, 其中 $\Phi(x)$ 为标准正态分布函数.

五、 (14分) 在 1970 年代后期, 人们发现酿造啤酒时麦芽干燥的过程中会形成致癌物质亚硝基二甲胺(NDMA). 在 1980 年代初期为此开发了一种新麦芽干燥工艺. 独立地随机抽查了新旧工艺下各一组样本, 得到NDMA含量(以10亿份中的份数计)的结果如下:

旧工艺	6	4	5	5	6	5	5	6	4	6	7	4
新工艺	2	1	2	2	1	0	3	2	1	0	1	3

设旧、新工艺下的两样本均来自正态总体. 在显著性水平 $\alpha = 0.05$ 下,

(1) 是否可以认为两个总体的方差相等?

(2) 是否可以认为旧工艺下NDMA平均含量比新工艺下显著地大 3?

六、 (10分) 某种鸟在起飞前, 双足齐跳的次数 X 服从参数为 p 的几何分布, 即其分布律为 $\mathrm{P}(X = k) = p(1-p)^{k-1}$, $k = 1, 2, \cdots$. 某人观测 130 次后, 获得一组样本如下:

k	1	2	3	4	5	6	7	8	9	10	11	12	≥ 13
频数	48	31	20	9	6	5	4	2	1	1	2	1	0

(1) 求 p 的最大似然估计值(精确到小数点后三位);

(2) 在拟合优度检验中频数一般不能小于 5, 故需将上述所有 $k \ge 7$ 情形下的频数进行合并, 此时请检验假设“X 服从几何分布”是否成立(显著性水平 $\alpha = 0.05$).

附录: $t_{22}(0.025) = 2.074$, $t_{22}(0.05) = 1.717$, $t_{23}(0.025) = 2.069$, $t_{23}(0.05) = 1.714$
$F_{11,11}(0.025) = 3.474$, $F_{11,11}(0.05) = 2.818$, $F_{12,12}(0.025) = 3.277$, $F_{12,12}(0.05) = 2.687$
$\chi^2_5(0.05) = 11.071$, $\chi^2_5(0.95) = 1.145$, $\chi^2_6(0.05) = 12.592$, $\chi^2_6(0.95) = 1.635$

参考答案

一. (1) D (2) $\frac{n}{2(n-1)}$ (3) D (4) A (5) $\frac{3}{4}$ (6) $\frac{n\lambda}{\lambda+\mu}$ (7) $(a^2+\sigma^2)\sigma^2$ (8) $\sqrt{\frac{n-1}{n+1}}$
(9) D (10) B

二. 以 A 表示选取的罐子为甲罐 (m 红 n 黑) 的事件, B 表示第一次取出的球为红球的事件, 则由 Bayes 公式可知

$$\mathrm{P}(A|B)=\frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B|A)\mathrm{P}(A)+\mathrm{P}(B|A^c)\mathrm{P}(A^c)}=\frac{m}{m+n}.$$

再由全概率公式可知, 第二次抽取的球仍为红色的概率为

$$\frac{m}{m+n}\mathrm{P}(A|B)+\frac{n}{m+n}\mathrm{P}(A^c|B)=\frac{m^2+n^2}{(m+n)^2}>\frac{1}{2}.$$

三. (1) 由 $\mathrm{P}(X+Y=2)=1$ 立知 $\mathrm{Corr}(X,Y)=-1$.

(2) 设随机变量 $U\sim U(0,2)$, 而 X 取值范围为 $(0,1)$, 故对任意 $0<x<1$,

$$\mathrm{P}(X\le x)=\mathrm{P}(U\le x)+\mathrm{P}(2-U\le x)=\frac{x}{2}+\frac{x}{2}=x,$$

即 $X\sim U(0,1)$. 故 X 的概率密度函数 $f(x)=1$, $0<x<1$.

(3) 易知 $Z=\frac{Y}{X}=\frac{2-X}{X}$ 及 $X=\frac{2}{Z+1}$, 由 (2) 和密度变换公式可知 $g(z)=\frac{2}{(z+1)^2}$, $z>1$.

(4) 利用密度变换公式或者几何概型, 可知 V 的概率密度函数为

$$h(v)=\begin{cases} v, & 0<v\le 1;\\ 2-v, & 1<v<2.\end{cases}$$

注：若上述密度函数表达式中变量范围缺乏或不正确, 按每处扣分.

四. (1) 由 $\mathrm{E}X=\frac{\theta+1}{\theta+2}$, 解方程 $\frac{\theta+1}{\theta+2}=\overline{X}$ 可知 θ 的矩估计量 $\hat\theta=\frac{2\overline{X}-1}{1-\overline{X}}$.

(2) 由题意, 似然函数 $L(\theta)=(\theta+1)^n\prod_{i=1}^n x_i^{\theta}$, 从而对数似然函数为

$$l(\theta)=\ln L(\theta)=n\ln(\theta+1)+\theta\sum_{i=1}^n \ln x_i.$$

令 $\frac{\mathrm{d}l(\theta)}{\mathrm{d}\theta}=0$, 可得 $\frac{n}{\theta+1}+\sum_{i=1}^n\ln x_i=0$. 由此可知, $g(\theta)=\frac{1}{\theta+1}$ 的极大似然估计量

$$\hat g=-\frac{1}{n}\sum_{i=1}^n \ln X_i.$$

(3) 记 $Y_i=-\ln X_i$, $i=1,2,\cdots,n$, 则易知 $\{Y_i, 1\le i\le n\}$ 独立同分布于参数为 $\theta+1$ 的指数分布, 由此即知 $\mathrm{E}\hat g=\frac{1}{\theta+1}$. 故 $\hat g$ 是 $g(\theta)$ 的一个无偏估计.

(4) 由上可知, $\hat g$ 可表示为一列独立同分布随机变量 $\{Y_i, 1\le i\le n\}$ 的平均, 故由经典场合下的中心极限定理可知, 常数 $b=\sqrt{\mathrm{Var}(Y_1)}=\frac{1}{\theta+1}$.

五. 先计算一些统计量的值. 旧工艺: $n_1 = 12$, $\overline{x} = 5.25$, $(n_1 - 1)S_1^2 = 10.25$; 新工艺: $n_2 = 12$, $\overline{y} = 1.5$, $(n_2 - 1)S_2^2 = 11$; $S_w^2 = 0.983^2$.

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

$$\frac{1}{3.474} = \frac{1}{F_{11,11}(0.025)} < \frac{S_1^2}{S_2^2} = 0.932 < F_{11,11}(0.025) = 3.474,$$

接受 H_0, 即可以认为两个总体的方差相等.

(2) $H_0 : \overline{X} - \overline{Y} \leq 3 \ \leftrightarrow\ H_1 : \overline{X} - \overline{Y} > 3$.
由

$$t = \frac{\overline{X} - \overline{Y} - 3}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 1.867 > t_{22}(0.05) = 1.717,$$

拒绝 H_0, 即可以认为旧工艺下NDMA平均含量比新工艺下显著地大 3.

六. (1) 对几何分布总体, 易知其参数 p 的极大似然估计值为

$$\hat{p} = \frac{1}{\overline{X}} = \frac{130}{363} = 0.358.$$

(2) 合并后的数据为

k	1	2	3	4	5	6	≥ 7
频数	48	31	20	9	6	5	11

从而检验统计量

$$\begin{aligned}\chi^2 &= \frac{(48 - 130 \times \hat{p})^2}{130 \times \hat{p}} + \frac{[31 - 130 \times \hat{p}(1 - \hat{p})]^2}{130 \times \hat{p}(1 - \hat{p})} + \cdots + \frac{[11 - 130 \times (1 - \hat{p})^6]^2}{130 \times (1 - \hat{p})^6} \\ &= 1.868 < \chi_5^2(0.05) = 11.071.\end{aligned}$$

故接受原假设.

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

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, 且相互独立, 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 $\mathrm{P}(X^n + Y > 1) =$ ______.

(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 $\mathrm{E}X =$ ______.

(6) 设 $X_1, X_2, \cdots, X_n$ 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 $\lim\limits_{n\to\infty} \mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i^2 - a\right| \geq \varepsilon\right) = 0$ 成立, 则常数 $a =$ (　)

(A) $\frac{1}{9}$　(B) $\frac{2}{9}$　(C) $\frac{1}{3}$　(D) $\frac{4}{9}$

(7) 设 $X_1, X_2, \cdots, X_{10}$ 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) χ^2_4　(B) χ^2_5　(C) χ^2_9　(D) χ^2_{10}

(8) 设 $X_1, X_2, \cdots, X_n$ 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 $X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}$ 为 θ 的(　)

(A) 矩估计　(B) 极大似然估计　(C) 无偏估计　(D) 相合估计

(9) 设 $X_1, X_2, \cdots, X_{36}$ 是来自正态总体 $N(\mu, 8)$ 的一组简单随机样本, 且记 $\overline{X}$ 为样本均值. 若以区间 $[\overline{X}-1, \overline{X}+1]$ 作为 μ 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 $H_0: \theta = 0.1 \leftrightarrow H_1: \theta = \theta_1$ 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 $\mathrm{E}[XY|X=x]$.

(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 $U=1/(X_1X_2X_3)$ 的概率密度函数.

四、 (20分) 设一列随机变量 $Y_1,Y_2,\cdots,Y_n$ 满足

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

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

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

(2) 试求 β 的极大似然估计量 $\hat{\beta}$, 并证明 $\hat{\beta}$ 为 β 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 β 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; $\frac{1}{n+1}$; π; B; B; D; 0.966; $1-\theta_1^6$.

二、 每小题 5 分.

(1) 由

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

可知随机向量 (X,Y) 的联合密度为

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

从而, 随机变量 Y 的密度函数为

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

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, $\mathrm{E}[XY|X=x]=x\mathrm{E}[Y|X=x]=x^2$.

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

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

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

$$\mathrm{P}(Y\le y)=\mathrm{P}(-a\ln X_1\le y)=\mathrm{P}(X_1\ge \mathrm{e}^{-y/a})=1-\mathrm{e}^{-y/a},$$

故 Y 服从参数为 $1/a$ 的指数分布, 从而其概率密度函数为

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

(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 $0\le z<1$ 时,

$$\mathrm{P}(Z\le z)=\mathrm{P}(X_2\le zX_1)=z/2;$$

而当 $z\ge 1$ 时,

$$\mathrm{P}(Z\le z)=\mathrm{P}(X_2\le zX_1)=1-1/(2z).$$

故 Z 的分布函数为

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

[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

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

再由 $U = \mathrm{e}^V$ 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 $Y_i \sim N(\beta x_i, \sigma^2), i = 1, 2, \cdots, n$ 及它们相互独立可知,

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

(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 β 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

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

从而 $\hat\beta$ 为 β 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

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

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

五、 先做一些计算. 杨倩: $n_1 = 15$, $\overline{x} = 10.333$, $(n_1 - 1)S_1^2 = 1.353$; 杨皓然: $n_2 = 15$, $\overline{y} = 10.447$, $(n_2 - 1)S_2^2 = 0.957$; $S_w^2 = 0.287^2$. [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

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

故接受 H_0, 即不能认为杨皓然的比赛成绩显著高于杨倩.

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

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) 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

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

则 $\mathrm{P}(Y=1)=$ ______.

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

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

(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^2_1 分布　(B) $F_{1,1}$ 分布　(C) $F_{2,2}$ 分布　(D) 以上均不对

(7) 设 $X_1,X_2,\cdots,X_n$ 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) $\sum_{i=1}^n X_i^2\sim\chi^2_n$　　(B) $\left(\frac{X_i}{\sigma}\right)^2+\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_n$

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

(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 $\frac{\sqrt2}{4}$. 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 $X_1,X_2,\cdots,X_n$ 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

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

(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 $\mathrm{E}[XY]$.

(4) 求 X 和 Y 的相关系数 $\mathrm{Corr}(X,Y)$.

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

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

其中 $\alpha>2, c>0$ 为未知参数. 设 $X_1,X_2,\cdots,X_n$ 是来自该总体的一组简单随机样本, 且分别以 $\overline{X}$ 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 $f_X(x)=cx\mathrm{e}^{-x}, x>0$ 或 $f_Y(y)=\frac{c}{6}y^3\mathrm{e}^{-y}, y>0$, 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]

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=6x(y-x)y^{-3},\quad 0\le x\le y;$$
$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=(y-x)\mathrm{e}^{x-y},\quad x\le y<\infty.$$

(3) 可由

$$\mathrm{E}[XY]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)\mathrm{d}x\mathrm{d}y=\int_0^{\infty}y\mathrm{e}^{-y}\mathrm{d}y\int_0^{y}x^2(y-x)\mathrm{d}x$$

及 Γ 函数的性质直接计算得出 $\mathrm{E}[XY]=10$. 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

$$\mathrm{E}X=2,\quad \mathrm{Var}(X)=2;\quad \mathrm{E}Y=4,\quad \mathrm{Var}(Y)=4.$$

另一方面, 由 (2) 中结论可知

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

亦即 $\mathrm{E}[Y|X]=X+2$. 由此可得,

$$\mathrm{E}[XY]=\mathrm{E}[\mathrm{E}(XY|X)]=\mathrm{E}[X\mathrm{E}(Y|X)]=\mathrm{E}[X^2]+2\mathrm{E}X=10.$$

(4) 综上可知, $\mathrm{Corr}(X,Y)=(10-2\times4)/\sqrt{2\times4}=\frac{\sqrt{2}}{2}$.

三、 每小题 6 分.

(1) 所求答案为 $f(x,y)=\frac{1}{2}\mathrm{e}^{-y/2}$, $0<x<1, y>0$. [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 $z-1\le y\le z, y>0$, 故需对 z 进行分段讨论.] 所求答案为

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

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

$$\begin{aligned}\mathrm{P}(Y\le X^2)&=\iint\limits_G f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^1\mathrm{d}x\int_0^{x^2}\frac{1}{2}\mathrm{e}^{-y/2}\mathrm{d}y\\&=\int_0^1\left(1-\mathrm{e}^{-\frac{x^2}{2}}\right)\mathrm{d}x=1-\sqrt{2\pi}[\Phi(1)-\Phi(0)]\\&=0.145.\end{aligned}$$

四、 每小题 8 分.

(1) 首先, 可计算得

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

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

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

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

(2) 首先, 可得似然函数为

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

且由此可知参数 c 的极大似然估计为 $\hat{c}_2=X_{(1)}$, 即 $\min\{X_1,X_2,\cdots,X_n\}$. 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 $\mathrm{E}[\hat{c}_2]=\frac{n\alpha c}{n\alpha-1}$, 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 $H_0:\mu\le 0\leftrightarrow H_1:\mu>0$. 对成对数据, 样本容量 $n=8$, $\overline{x}-\overline{y}=1.25, s=1.2817$, 则由一样本 t 检验的检验统计量为

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

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.

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

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) 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 $\mathrm{Cov}(X,Y)=0$, 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 $\mathrm{P}(X-3<Y<X+3)$ 满足(　)

(A) $\le \frac{1}{4}$　(B) $\le \frac{4}{9}$　(C) $\ge \frac{3}{4}$　(D) $\ge \frac{5}{9}$

(7) 设$X_1,X_2,\cdots,X_n$ $(n\ge 2)$ 是来自正态总体 $N(\mu,1)$ 的简单随机样本, 且记 $\overline{X}$ 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) $\sum_{i=1}^n (X_i-\overline{X})^2$　(D) $n(\overline{X}-\mu)^2$

(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 $X_1,X_2,\cdots,X_{16}$ 是来自正态总体 $N(\mu,4)$ 的一组简单随机样本, 且记 $\overline{X}$ 为样本均值. 若假设检验 $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\ne\mu_0$ 的拒绝域为 $\{|\overline{X}-\mu_0|>c\}$, 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 $\overset{d}{=}$ 表示同分布, 证明:

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

(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

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

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 $0\le x\le 1$, 求条件概率 $\mathrm{P}(M\le x|Y=n)$.

(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 $\mathrm{P}(Z>z)$ 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

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

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) $\mathrm{P}(M\le x|Y=n)=\mathrm{P}(U_i\le x,i=1,\dots,n)=x^n$.

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

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

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 $\lfloor z\rfloor$ 表示不大于 z 的最大整数, 注意到 $0<M<1$, 则对任一 $z>0$,

$$\begin{aligned}\mathrm{P}(Z>z)&=\mathrm{P}(X\ge\lfloor z\rfloor+2)+\mathrm{P}(X=\lfloor z\rfloor+1,M\le\lfloor z\rfloor+1-z)\\&=\frac{(e-1)e^{-\lfloor z\rfloor-2}}{1-1/e}+(e-1)e^{-\lfloor z\rfloor-1}\cdot\frac{e^{\lfloor z\rfloor+1-z}-1}{e-1}\\&=e^{-z}.\end{aligned}$$

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

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

故对数似然函数为

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

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

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

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

(3) 由 χ^2 分布的定义, 可知

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

故 σ^2 的置信系数为 0.95 的置信区间为

$$\Big[\frac{(m+n)\hat{\sigma}^2}{\chi^2_{m+n}(0.025)},\ \frac{(m+n)\hat{\sigma}^2}{\chi^2_{m+n}(0.925)}\Big].$$

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 $\overline{x}=14.95$, 样本方差 $s_1^2=6.84^2$, 样本容量 $n_1=13$; 女性总体: $\overline{y}=22.29$, $s_2^2=5.32^2$, $n_2=10$; 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) $H_0:\mu_1\ge\mu_2\ \leftrightarrow\ H_1:\mu_1<\mu_2$. 由

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

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.

中国科学技术大学

2022—2023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: 2023 年 6 月 27 日下午 14:30–16:30; 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 A 和 B 是一随机试验的两个互斥结果, 且 $\mathrm{P}(A)=p, \mathrm{P}(B)=q, 0<p+q\le 1$. 现独立重复这一试验时, 则结果 A 比 B 先出现的概率为______.

(2) 设随机变量 X 的概率密度函数 $f(x)$ 满足 $f(2+x)=f(2-x)$ 及 $\int_0^2 f(x)\mathrm{d}x=0.3$, 则 $\mathrm{P}(X\le 0)=(\quad)$

(A) 0.1 (B) 0.2 (C) 0.3 (D) 0.4

(3) 设 X 和 Y 分别为离散型和连续型随机变量且相互独立. 则 $X+Y$ 的类型为 ( ).

(A) 连续型 (B) 离散型 (C) 非离散型也非连续型 (D) 以上皆有可能

(4) 掷一枚均匀的骰子, 欲使 1 至 6 点均出现至少一次, 则平均需要掷______次.

(5) 在单位圆盘上独立地随机取两点 A 和 B, 则以 A 为圆心, 线段 AB 为半径的圆完全落在该单位圆盘内的概率是( )

(A) $\frac{1}{2}$ (B) $\frac{1}{3}$ (C) $\frac{1}{4}$ (D) $\frac{1}{6}$

(6) 设 $X_1, X_2, \cdots, X_n$ 是来自指数总体 $\mathrm{Exp}(2)$ 的一组简单随机样本, 则当 $n\to\infty$ 时, 统计量 $Y_n=\sum_{i=1}^n X_i^2/n$ 依概率收敛到 ( )

(A) $\frac{1}{2}$ (B) $\frac{1}{4}$ (C) 2 (D) 4

(7) 设 X_1, X_2, X_3, X_4, X_5 是来自正态总体 $N(0,\sigma^2)$ 的一组简单随机样本, 若统计量 $T=a(X_1+X_2)/\sqrt{X_3^2+X_4^2+X_5^2}$ 服从 t 分布, 则常数 $a=(\quad)$

(A) $\frac{1}{2}$ (B) $\frac{\sqrt{2}}{2}$ (C) $\frac{\sqrt{6}}{2}$ (D) $\frac{\sqrt{6}}{6}$

(8) 已知一批零件的长度 X (单位: cm)服从正态分布, 从中随机抽取 25 个零件, 得到样本的平均长度为 36 cm, 样本方差为 $4\,\mathrm{cm}^2$, 则总体均值 $\mathrm{E}X$ 的置信水平为 0.95 的置信区间为____________ (精确到小数点后三位).

(9) 从电信公司用户每月话费账单中随机抽取 25 张, 算得平均消费为 32.8 元, 标准差为 20.8 元. 设用户每月话费服从均值为 μ 的正态分布, 对假设检验 $H_0: \mu\le 30 \leftrightarrow H_1: \mu>30$, 则在显著性水平 $\alpha=0.5$ 下我们____拒绝原假设 (填“可以”或“不能”).

(10) 设一总体分布函数为 $F(x;\theta)$, 其中 θ 为一未知参数, 而 θ_0 为一已知常数. 若一假设检验 $H_0: \theta=\theta_0 \leftrightarrow H_1: \theta\ne\theta_0$ 在显著性水平 α 下的拒绝域为 $|\overline{X}-\theta_0|>1$, 这里 $\overline{X}$ 为样本均值, 则 θ 的置信系数为 $1-\alpha$ 的置信区间为________.

二、 (20分) 设 $X_1, X_2, \cdots, X_n$ 为一列独立的随机变量, 且同分布于参数为 λ 的指数分布 $\mathrm{Exp}(\lambda)$. 记 $Y = \min(X_1, X_2, \cdots, X_n)$.

(1) 求 Y 的分布.

(2) 求概率 $\mathrm{P}(X_1 = Y)$.

(3) 求随机向量 (X_1, Y) 的联合分布函数 $F(x, y)$.

(4) 对任一实数 $y > 0$, 求条件期望 $\mathrm{E}[X_1 | Y = y]$.

三、 (14分) 设随机变量 X, Y, Z 为相互独立的标准正态随机变量, 且记

$$U = \frac{X - Y}{\sqrt{2}}, \quad V = \frac{X + Y - 2Z}{\sqrt{6}}, \quad W = \frac{X + Y + Z}{\sqrt{3}}.$$

(1) 试求随机向量 (U, V, W) 的联合概率密度函数 $f(u, v, w)$.

(2) 随机变量 U, V, W 是否相互独立? 需说明理由.

四、 (21分) 设某种电子器件的寿命 (单位: 小时) T 服从双参数指数分布, 即其概率密度函数为

$$f(t) = \frac{1}{\theta} \exp\Big\{ -\frac{t - c}{\theta} \Big\} I_{(c,\infty)}(t),$$

其中参数 $\theta, c > 0$. 现自一批这种器件中随机地取 n 件进行寿命试验, 记录它们的失效时间依次为 $t_1 < t_2 < \cdots < t_n$.

(1) 求 θ 和 c 的矩估计值.

(2) 求 θ 和 c 的极大似然估计值.

(3) 若已知参数 $\theta = \frac{1}{2}$, 则 c 的极大似然估计是否无偏? 若是, 证明之; 若否, 修正之.

五、 (15分) 某篮球爱好者每天进行 5 组投篮练习, 每组练习均包含 4 次投篮, 完成后他记录下每组投中的次数. 经过 80 天后, 得到如下数据:

投中数	0	1	2	3	4
频数	25	118	139	93	25

结合你所学的统计知识, 我们能否认为该篮球爱好者每组投中的次数服从二项分布 (显著性水平 $\alpha = 0.05$)?

附录 上分位数: $t_{24}(0.025) = 2.064$, $t_{24}(0.05) = 1.711$

$\chi_3^2(0.05) = 7.815$, $\chi_3^2(0.95) = 0.352$, $\chi_4^2(0.05) = 9.488$, $\chi_4^2(0.95) = 0.711$

(完)

参考答案

一、 每小题 3 分.

$\frac{p}{p+q}$; B; A; 14.7; D; A; C; $(35.174, 36.826)$; 不能; $[\overline{X}-1, \overline{X}+1]$.

二、 每小题 5 分.

(1) 对任一 $y>0$, 由 $\mathrm{P}(Y>y)=[\mathrm{P}(X_1>y)]^n=\mathrm{e}^{-n\lambda y}$, 可知 $Y\sim \mathrm{Exp}(n\lambda)$, 即 Y 服从参数为 $n\lambda$ 的指数分布.

(2) 由对称性可知 $\mathrm{P}(X_1=Y)=\cdots=\mathrm{P}(X_n=Y)$, 且各项之和为 1. 从而所求概率为

$$\mathrm{P}(X_1=Y)=\frac{1}{n}.$$

(3) 当 $0<y<x$ 时, 由独立性可知

$$\begin{aligned}\mathrm{P}(X_1\le x, Y>y)&=\mathrm{P}(y<X_1\le x, X_2>y,\cdots,X_n>y)\\&=(\mathrm{e}^{-\lambda y}-\mathrm{e}^{-\lambda x})\mathrm{e}^{-(n-1)\lambda y}\\&=\mathrm{e}^{-n\lambda y}-\mathrm{e}^{-\lambda[x+(n-1)\lambda y]}.\end{aligned}$$

从而,

$$\begin{aligned}\mathrm{P}(X_1\le x, Y\le y)&=\mathrm{P}(X_1\le x)-\mathrm{P}(X_1\le x, Y>y)\\&=1-\mathrm{e}^{-\lambda x}-\mathrm{e}^{-n\lambda y}+\mathrm{e}^{-\lambda[x+(n-1)\lambda y]}.\end{aligned}$$

而当 $0<x\le y$ 时, $\mathrm{P}(X_1\le x, Y\le y)=\mathrm{P}(X_1\le x)=1-\mathrm{e}^{-\lambda x}$. 综上可知, 随机向量 (X_1, Y) 的联合分布函数为

$$F(x,y)=\begin{cases}1-\mathrm{e}^{-\lambda x}-\mathrm{e}^{-n\lambda y}+\mathrm{e}^{-\lambda[x+(n-1)\lambda y]}, & 0<y<x;\\ 1-\mathrm{e}^{-\lambda x}, & 0<x\le y;\\ 0, & \text{其它}.\end{cases}$$

(4) 由 (2) 及指数分布的无记忆性可知

$$\mathrm{E}[X_1|Y=y]=y\cdot\frac{1}{n}+\Big(y+\frac{1}{\lambda}\Big)\frac{n-1}{n}=y+\frac{n-1}{n\lambda}.$$

三、 每小题 7 分.

(1) 由计算可知 Jacob 行列式绝对值为 1, 且 $u^2+v^2+w^2=x^2+y^2+z^2$, 故由密度变换公式可知随机向量 (U,V,W) 的概率密度函数为

$$f(u,v,w)=\frac{1}{(2\pi)^{3/2}}\exp\Big\{-\frac{u^2+v^2+w^2}{2}\Big\}.$$

(2) 由上式中的概率密度函数 $f(u,v,w)$ 可分离变量即知随机变量 U,V,W 相互独立.

四、 每小题 7 分.

(1) 首先, 可计算得

$$\mathrm{E}T = \theta + c, \quad \mathrm{Var}(T) = \theta^2.$$

然后通过联立方程

$$\mathrm{E}T = \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i, \quad \mathrm{Var}(T) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{t})^2,$$

可得 α 和 c 的矩估计值分别为$\hat{\theta}_1 = s, \quad \hat{c}_1 = \bar{t} - s.$

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

(2) 可得似然函数为

$$L(\theta, c) = \frac{1}{\theta^n}\exp\Big\{-\frac{1}{\theta}\sum_{i=1}^{n}(t_i - c)\Big\}, \quad c \le t_1 < t_2 < \cdots < t_n.$$

由于该函数关于 c 单调递增, 可知参数 c 的极大似然估计值为 $\hat{c}_2 = t_1$. 然后, 可求得对数似然函数为

$$l(\alpha, c) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n}(t_i - c).$$

故通过令 $\frac{\partial l(\theta,c)}{\partial\theta} = 0$, 可得 θ 的极大似然估计值为 $\hat{\theta}_2 = \bar{t} - t_1$.

(3) 否. 若 $\theta = 1/2$, 则可求得估计量 $\hat{c}_2 = T_1$ 的概率密度函数为

$$g(t) = 2n\mathrm{e}^{-2n(t-c)} I_{(c,\infty)}(t).$$

由此可求得 $\mathrm{E}[\hat{c}_2] = c + \frac{1}{2n}$, 从而 $\hat{c}_2$ 不是 c 的无偏估计, 但可修正为 $\hat{c}_2^* = t_1 - \frac{1}{2n}$.

五、 拟合优度检验. 原假设为每组投中的次数服从二项分布 $B(4, p)$, 其中 $0 < p < 1$ 为某个常数. 我们先估计参数 p, 利用极大似然估计或者矩估计, 均有

$$\hat{p} = \frac{1}{80 \times 5 \times 4}(25 + 118 \times 1 + 139 \times 2 + 93 \times 3 + 25 \times 4) = \frac{775}{1600} = 0.4844.$$

由此可计算得出投中次数0, 1, 2, 3, 4的理论频数分别为 $28.3, 106.2, 149.7, 93.8, 22.0$. 从而, 由检验统计量

$$\chi^2 = \frac{(25 - 28.3)^2}{28.3} + \cdots + \frac{(25 - 22.0)^2}{22.0} = 2.877 < \chi_3^2(0.05) = 7.815,$$

故我们应接受每组投中次数服从二项分布的说法.

步骤分: 估计 p, 5 分; 计算出检验统计量的值, 5 分; 临界值, 3分; 统计决策, 2 分.