

中国科学技术大学

2020—2021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 $0 < P(B) < 1$, 下列为假命题的是().
 - (A) 若 $P(A|B) = P(A)$, 则 $P(A|\bar{B}) = P(A)$
 - (B) 若 $P(A|B) > P(A)$, 则 $P(\bar{A}|\bar{B}) > P(\bar{A})$
 - (C) 若 $P(A|B) > P(A|\bar{B})$, 则 $P(A|B) > P(A)$
 - (D) 若 $P(A|A \cup B) > P(\bar{A}|A \cup B)$, 则 $P(A) > P(B)$
- (2) 设平面上有 n 个点, 编号分别为 $1, 2, \dots, n$. 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 $P(Y > y) =$ ().
 - (A) $e^{-y/p}$ (B) pe^{-y} (C) $p/(y + p)$ (D) $p/(e^y - 1 + p)$
- (4) 若随机变量 X 和 Y 满足 $P(X^2 + Y^2 = 2) = 1$, 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) $P(X + Y = 0) = 1/2$
 - (C) $EX = EY = 0$
 - (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 $P(\max\{X, Y\} \geq 0) =$ _____.
- (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 $E[X|X + Y = n] =$ _____.
- (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 $\text{Cov}(X, XY^2) =$ _____.
- (8) 设 $X_1, X_2, \dots, X_n, X_{n+1}$ 是来自总体 $N(\mu, \sigma^2)$ 的简单随机样本, 且 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $m_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. 已知 $T = c(X_{n+1} - \bar{X})/\sqrt{m_2}$ 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, \dots, X_{16} 是来自总体 $N(\mu, 4)$ 的简单随机样本, \bar{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 $W = \{\bar{X} \geq 11\}$, 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) $1 - \Phi(2)$

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 $\text{Corr}(X, Y)$;

(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, \dots, 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 $\lim_{n \rightarrow \infty} P(\sqrt{n}(\hat{g} - g(\theta))/b \leq x) = \Phi(x)$ 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 $P(X = k) = p(1 - p)^{k-1}$, $k = 1, 2, \dots$. 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 $k \geq 7$ 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: $t_{22}(0.025) = 2.074$, $t_{22}(0.05) = 1.717$, $t_{23}(0.025) = 2.069$, $t_{23}(0.05) = 1.714$

$F_{11,11}(0.025) = 3.474$, $F_{11,11}(0.05) = 2.818$, $F_{12,12}(0.025) = 3.277$, $F_{12,12}(0.05) = 2.687$

$\chi_5^2(0.05) = 11.071$, $\chi_5^2(0.95) = 1.145$, $\chi_6^2(0.05) = 12.592$, $\chi_6^2(0.95) = 1.635$

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

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

再由全概率公式可知, 第二次抽取的球仍为红色的概率为

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

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

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

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

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

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

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 $EX = \frac{\theta+1}{\theta+2}$, 解方程 $\frac{\theta+1}{\theta+2} = \bar{X}$ 可知 θ 的矩估计量 $\hat{\theta} = \frac{2\bar{X}-1}{1-\bar{X}}$.

- (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 $\frac{dl(\theta)}{d\theta} = 0$, 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 $Y_i = -\ln X_i, i = 1, 2, \dots, n$, 则易知 $\{Y_i, 1 \leq i \leq n\}$ 独立同分布于参数为 $\theta+1$ 的指数分布, 由此即知 $E\hat{g} = \frac{1}{\theta+1}$. 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 $\{Y_i, 1 \leq i \leq n\}$ 的平均, 故由经典场合下的中心极限定理可知, 常数 $b = \sqrt{\text{Var}(Y_1)} = \frac{1}{\theta+1}$.

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

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

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.

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

由

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

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

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

(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.