

2012-2013 第一学期概率论期末考试试卷

一. 判断选择题 (每题 3 分, 答题请写在试卷上):

1. 设 A, B, C 是三个随机事件, 则在下列不正确的是_____.

(A) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

(B) $(A \cup B) \cap C = A \cup (B \cap C)$

(C) $A \cap (B \cap C) = (A \cap B) \cap C$

(D) $A \cap (\overline{B \cap C}) = (A \cap \overline{B}) \cup (A \cap \overline{C})$

2. 设事件 A 与自身独立, 则 A 的概率为_____.

(A) 0

(B) 1

(C) 0 或 1

(D) 1/2

3. 设 $f(x)$ 和 $g(x)$ 为两个概率密度函数, 则下述还是密度函数的是_____.

(A) $f(x)/g(x)$

(B) $f(x)-g(x)$

(C) $(f(x) + g(x))/2$

(D) $(1 + f(x))(1-g(x))$

4. 随机变量 X 和 Y 独立, Y 和 Z 独立, 且都有期望方差, 则必有_____.

(A) X 和 Z 独立

(B) X 和 Z 不相关

(C) X 和 Z 相关

(D) $\text{Cov}(X, Y) = 0$

5. 设 $0 < P(B) < 1$, 则 $P(A|B) = P(A|\overline{B})$ 成立的充分必要条件是_____.

(A) $P(AB) = P(A)P(B)$

(B) $P(A + B) = P(A) + P(B)$

(C) $P(A) = P(B)$

(D) $P(A) = P(\overline{B})$

6. 设 X_1, \dots, X_n 为来自均匀分布 $U(-\theta, \theta)$ 的一组样本, θ 为未知参数, 则下述量为统计量的是_____.

(A) $\overline{X} - \theta$

(B) $\max_{1 \leq i \leq n} (X_i - \theta) - \min_{1 \leq i \leq n} (X_i - \theta)$

(C) $\max_{1 \leq i \leq n} (X_i - \theta)$

(D) $\min_{1 \leq i \leq n} (X_i - \theta)$

7. 假设从两个独立的正态总体中各得到样本量为 10 的两组样本, 若两个总体的方差相同, 则使用两样本 t 检验时 t 分布的自由度为_____.

- (A)9 (B)10 (C)18 (D)20

8. 当原假设 H_0 为真时, 检验 ϕ 有可能_____.

- (A) 犯一类错误 (B) 犯第二类错误
(C) 犯第一类或者第二类错误 (D) 同时犯第一类和第二类错误

9. 假设总体密度为 $f_\theta(x)$, 其中 θ 为参数. 若 X 为来自该总体的样本, 则下述不正确的是_____.

- (A) 固定 x 时 $f_\theta(x)$ 为似然函数 (B) 固定 θ 时 $f_\theta(x)$ 为似然函数
(C) 固定 θ 时 $f_\theta(x)$ 为密度函数 (D) $f_\theta(x)$ 衡量了不同 θ 下观测到 x 的可能性大小

10. 假设总体 X 为取值 $0, 1, 2$ 的离散型随机变量, 且取各值的概率分别为 $P(X = 0) = 0.5, P(X = 1) = p, P(X = 2) = 0.5 - p$, 其中 $0 < p < 0.5$ 为参数. 则当使用拟合优度检验时, 检验统计量的渐近卡方分布的自由度为_____.

- (A)3 (B)2 (C)1 (D)0

二.(15 分) 设昆虫产卵数目服从参数为 1 的 Poisson 分布, 而每个卵孵化为幼虫的概率为 p , 各卵是否孵化相互独立, 试求

- (1) 一个昆虫产生 m 个幼虫的概率。
(2) 若已知某个昆虫产生了 m 个幼虫, 求该昆虫产了 $n(n \geq m)$ 个卵的概率。

三.(15 分) 设随机变量 X, Y 相互独立, 且 X 服从均匀分布 $U(-1, 1), Y$ 服从均值为 $1/2$ 的指数分布, 则

- (1) 求随机变量 $Z = (X + 1)Y$ 和 X 的相关系数。
(2) 求条件概率 $P(Z > 1 | X = 0)$.

四.(15 分) 当 PM2.5 值全天监测平均在 35 微克/立方米以内时, 空气质量属于一级. 现观测到合肥市琥珀山庄过去 10 天的日平均 PM2.5 值分别为 28.24, 31.48, 33.85, 39.34, 37.78, 30.21, 29.92, 31.21, 30.17, 37.84. 若假设琥珀山庄区域日均 PM2.5 值 X 服从正态分布, 各天日均 PM2.5 值相互独立. 则

- (1) 试给出日均 PM2.5 值的 95% 置信上限。
(2) 若感兴趣空气质量为一级的概率 $p = P(X \leq 35)$, 试基于观测的日均数据给出 p 的极大似然估计.

五.(15 分) 设甲乙两家食用盐工厂生产的食盐每袋重量均服从正态分布 (忽略重量不可取负值). 现从这两家工厂产品中各随机抽出 10 件标称为 500 克的袋装食盐, 分别测得抽出各袋食盐的重量 (单位为克) 为

甲厂: 495, 494, 500, 502, 501, 492, 495, 495, 499, 503;

乙厂: 494, 506, 496, 505, 500, 508, 502, 504, 502, 499.

试问甲乙两家工厂生产这种标称为 500 克的袋装盐重量上是否有差异 ($\alpha = 0.05$).

六.(10 分) 为研究人们每天阅读电子书的时间 (T) 长短与购买实体书 (Y) 两者之间的关系, 随机调查了 210 个人, 结果如下

	$t < 1$	$1 < t < 3$	$t > 3$
购买	12	70	20
不购买	40	28	40

试在水平 $\alpha = 0.05$ 下判断每天阅读电子书的时间长短和购买实体书两者之间是否有关? 阅读电子书的时间长短和购买实体书之间呈现何种特点?

一. 判断选择题 (每题 3 分):

1. B 2. C 3. C 4. D 5. A 6. B 7. C 8. A 9. B 10. C

二.(15 分) (1)

$$\begin{aligned} P(X = m) &= \sum_{n=m}^{\infty} P(X = m|Y = n)P(Y = n) = \sum_{n=m}^{\infty} \binom{n}{m} p^m q^{n-m} \frac{1}{n!} e^{-1} \\ &= \frac{p^m}{m!} e^{-p}, \quad m = 0, 1, 2, \dots \end{aligned}$$

(2)

$$\begin{aligned} P(Y = n|X = m) &= \frac{P(X = m|Y = n)P(Y = n)}{P(X = m)} \\ &= \frac{q^{n-m}}{(n-m)!} e^{-q}, \quad n = m, m+1, \dots \end{aligned}$$

三.(15 分) (1) 由于 $EZX = E[X(X+1)]EY = 1/6, Var(Z) = 5/12$, 因此

$$\rho_{Z,X} = [EZ X - EZ \cdot EX] / \sqrt{Var(Z)Var(X)} = 1/\sqrt{5}$$

求随机变量 $Z = (X+1)Y$ 的期望和方差.

$$(2) P(Z > 1|X = 0) = P(Y > 1) = e^{-2}.$$

四.(15 分)(1) 由于 $\bar{x} = 33.004, s = 3.95$, 在题设下易知日均 PM2.5 的 95% 置信上限为 $\bar{x} + \frac{s}{\sqrt{n}} t_{0.05}(n-1)$, 带入样本值得到 35.5.

(2) $p = P(X \leq 35) = P(\frac{x-\mu}{\sigma} \leq \frac{35-\mu}{\sigma}) = \Phi(\frac{35-\mu}{\sigma})$, 而 \bar{x} 和 $\sqrt{(n-1)/ns^2} = \sqrt{9/10}s = 3.75$ 为 μ 和 σ 的似然估计值, 因此 p 的极大似然估计值为 $\hat{p} = \Phi(\frac{\bar{x}-\hat{\mu}}{\hat{\sigma}}) = \Phi(0.53) = 0.70$.

五.(15 分) 记 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 分别表示两家工厂袋装盐的重量分布

(1) 考虑方差是否一致: 对假设 $H_0: \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2$, 由检验统计量 $F = S_x^2/S_y^2 = 14.71/19.6 = 0.75 > F_{0.975}(9, 9) = 1/F_{0.025}(9, 9) = 1/4.03 = 0.25$, 因此在 0.05 水平下不能拒绝零假设.

(2) 考虑均值是否一致: 考虑假设 $H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 \neq \mu_2$, 由 (1) 结果知可以使用两样本 t 检验, 由检验统计量 $T = |\bar{x} - \bar{y}| / \sqrt{(s_x^2 + s_y^2)/10} = 2.16 > t_{0.025}(18) = 2.10$, 因此拒绝零假设. 即在 0.05 水平下拒绝“两家工厂的袋装食盐平均重量一致”这一假设.

六.(10 分) 假设每天阅读电子书时间长短与购买实体书之间无关, 则由 Pearson 卡方检验有 $T = \sum \frac{(O-E)^2}{E} = 39.60 > \chi_{0.05}(2) = 5.99$, 因此在 0.05 水平下拒绝“每天阅读电子书时间长短与购买实体书之间无关”这一假设。注意到在三类阅读时间下, 购买实体书人的比例分别为 0.23, 0.71 和 0.33, 因此每天阅读电子书时间在 1 小时和 3 小时之间的人群购买实体书的比例最高, 而当每天阅读电子书时间长于 3 小时后, 购买实体书的人比例反而下降为 0.33.