中国科学技术大学

2016—2017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: 2017年1月3日上午8:30–10:30;　使用简单计算器

一. (30分, 每小题均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为随机事件, $P(A) = 0.4, P(B) = 0.3, P(A|B) = 0.4$, 则( ).
   (A) A与B相互独立　(B) A与B互斥
   (C) $A \supset B$　(D) $P(A \cup B) = P(A) + P(B)$

2. 甲乙二人进行网球比赛, 每回合胜者得 1 分, 且每回合甲胜的概率为 $p(0 < p < 1)$, 乙胜的概率为 $1 - p$, 比赛进行到有一人比另外一个人多 2 分就终止, 多 2 分者最终获胜, 则甲最终获胜的概率为________.

3. 设随机变量X服从参数为λ的Poisson分布, 且已知$P(X = 1) = P(X = 2)$, 则X取值为 3 的概率为________.

4. 设随机变量X和Y均服从正态分布$N(0, \sigma^2)$, 且已知$P(X \le 2,\ Y \le -2) = 0.25$, 则$P(X > 2, Y > -2) =$ ________.

5. 设X和Y相互独立且分别服从均值为 1 和 1/4 的指数分布, 则$P(X < Y) =$ ______.

6. 设随机变量X_1和X_2相互独立, 方差均存在, 且概率密度函数分别为$f_1(x)$和$f_2(x)$, 若Y_1的概率密度函数为$[f_1(y) + f_2(y)]/2$, 而$Y_2 = (X_1 + X_2)/2$, 则( ).
   (A) $E[Y_1] > E[Y_2], Var[Y_1] > Var[Y_2]$　(B) $E[Y_1] = E[Y_2], Var[Y_1] = Var[Y_2]$
   (C) $E[Y_1] = E[Y_2], Var[Y_1] < Var[Y_2]$　(D) $E[Y_1] = E[Y_2], Var[Y_1] > Var[Y_2]$

7. 在假设检验中, 下列关于拒绝域和接受域说法错误的是( ).
   (A) 与显著性水平 α 有关　(B) 与所构造的统计量的分布有关
   (C) 随样本观测值的不同而改变　(D) 互不相交

8. 设X_1, X_2, X_3, X_4为来自总体$N(1, \sigma^2)$的一组简单随机样本, 则$T = \frac{(X_1 - X_2)^2}{(X_3 + X_4 - 2)^2}$的分布为( ).
   (A) t_1　(B) $F_{1,1}$　(C) $F_{2,2}$　(D) 以上皆不正确.

9. 设$X_1, X_2, \cdots, X_n$为来自二项总体$B(n, p)$的一组简单随机样本, 且记 $\overline{X}$ 和 S^2 为样本均值和样本方差. 若 $\overline{X} + kS^2$ 为 np^2 的一个无偏估计, 则常数 $k =$ ________

10. 设正态总体$X \sim N(\mu, \sigma^2)$, 其中参数 μ 和 σ^2 均未知. 若样本容量 n 和置信系数 $1 - \alpha$ 均保持不变, 对于不同的样本观测值, 则总体均值 μ 的置信区间长度( ).
    (A) 始终保持不变　(B) 与 μ 的真值有关
    (C) 与样本均值有关　(D) 不固定.

考试科目 概率论与数理统计　　　　得分 ________　　80↑

所在系 ______　　姓名 ________　　学号 ________

考试时间：2016年1月13日上午8:30—10:30；　使用简单计算器

$EN=6\times\frac{1}{6}=1$

一. 单项选择填空题(每题3分,共30分,答题请写在试卷上)：

1 考虑 $\{1,2,3,4,5,6\}$ 的一个随机排列，与原来位置序号相同的数字个数记为N（比如排列513246与原来位置序号只有数字3和6是相同的，所以$N=2$），则 $EN=$________.

2 设$X\sim N(1,9)$, $Y\sim N(1,16)$, 且X与Y的相关系数为$\rho_{XY}=-1/2$，设$Z=\frac{X}{2}+\frac{Y}{3}$，则$X$和$Z$的相关系数$\rho_{XZ}=$ $\frac{5}{\sqrt{73}}$ ________.

3 下列哪些函数不是概率密度或者分布律？________

(A) $f(x)=\frac{1}{2}I\{a-3/2\le x\le a\}+\frac{1}{4}I\{a-1\le x\le a\}, a\in\mathbb{R}$

(B) $f(x)=\frac{5}{6}I\{a-3/2\le x\le a\}-\frac{1}{4}I\{a-1\le x\le a\}, a\in\mathbb{R}$

(C) $P(X=k)=\frac{1}{2^k}, k=0,1,2,\ldots$

(D) $f(x)=\lambda\alpha x^{\alpha-1}e^{-\lambda x^\alpha}I\{x>0\}, \lambda>0, \alpha>0$

4 若两随机事件A,B相互独立，下列哪些说法不正确？________

(A) $P(A|B)=P(A|B^c)$ (B) $P(A|B)=P(A^c|B)$

(C) $P(B|A)=P(B|A^c)$ (D) $P(B^c|A)=P(B^c|A^c)$

5 假设盒子中有一个不知颜色是黑色还是白色的球，现在给盒子里再放一个白球，然后从中随机拿出一个球，发现是白色的。则此时盒中剩余的球是白色的概率为________.

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

6 下述对正态总体均值的置信区间的表述错误的是________.

(A) 置信度愈高，则可靠性愈高

(B) 置信度愈高，则置信区间愈宽

(C) 置信区间的大小与测量次数的平方根成正比

(D) 置信区间的位置取决于测量的平均值

7 设随机变量X和Y都服从标准正态分布，则下述正确的是 C ________.

(A) $X+Y$服从正态分布 (B) X^2+Y^2服从χ^2分布

(C) X^2和Y^2都服从χ^2分布 (D) $\frac{X^2}{Y^2}$服从F分布

8 下述对一个检验方法的第二类错误描述错误的是________.

(A) 在给定样本量下，第二类错误的概率不可能任意小

(B) 在对立假设空间的子集下控制第二类错误的概率，可以用来确定样本量大小

(C) 在有限样本量下，第二类错误是不可以避免的

(D) 在一个检验结果是拒绝零假设时候，我们会有很大的风险犯第二类错误

9 下述检验正态性假设的方法中错误的是 D ________.

(A) 直方图方法　(B) 拟合优度检验方法

(C) 使用偏度系数和峰度系数　(D) t检验

二. (11分) 设二维随机向量 (X,Y) 的概率密度函数为 $f(x,y)=\begin{cases} Ax^2, & 0<|x|<y<1; \\ 0, & 其它. \end{cases}$

试求常数 A 的值及条件概率 $\mathrm{P}(X\le 0.25|Y=0.5)$.

三. (16分) 设随机变量 X 和 Y 相互独立, 且 X 的概率分布为 $\mathrm{P}(X=0)=\mathrm{P}(X=2)=0.5$, Y的概率密度函数为$f(y)=\begin{cases} 2y, & 0<y<1; \\ 0, & 其它. \end{cases}$

1. 求 $\mathrm{P}(Y\le \mathrm{E}Y)$;
2. 求 $Z=X+Y$ 的概率密度函数.

四. (15分) 某工程师为了解一台天平的精度, 用该天平对一物体进行 n 次测量, 该物体的质量 μ 是已知的. 设 n 次测量结果 $X_1,X_2,\cdots,X_n$ 相互独立且服从正态分布$N(\mu,\sigma^2)$, 该工程师记录的是 n 次测量的绝对误差 $Z_i=|X_i-\mu|$, $i=1,2,\cdots,n$. 现利用这些绝对误差来估计标准差 σ.

1. 求 Z_i 的概率密度函数;
2. 利用一阶矩求 σ 的矩估计量;
3. 求 σ 的极大似然估计量.

五. (18分) 在质量管理中, 产品的优良性和稳定性可以分别用均值和方差来体现. 要比较甲乙两厂所生产的电视机的质量是否有差异, 对两厂所生产的各 10 台产品的使用情况进行了追踪调查, 得到这 20 台电视机的寿命(单位: 年)数据如下:

甲	8	7	9	5	12	10	9	10	8	7
乙	10	8	5	7	8	7	11	4	5	6

设电视机的寿命服从正态分布, 利用假设检验的知识判断两厂所生产电视机的寿命是否有显著差异(显著性水平取 $\alpha=0.05$).

六. (10分) 对截止目前为止一共 44 位美国总统(含候任总统Trump)的星座进行分析, 发现十二星座中天蝎座和水瓶座各有 5 人, 双子座和射手座各有 3 人, 处女座和白羊座各有两人, 而其余星座均有 4 人. 于是有人宣称有些星座擅长当美国总统, 而有些星座则不擅长. 结合你所学的知识, 说明该说法是否有统计学上的依据?(显著性水平取 $\alpha=0.05$)

附录: 上分位数表

$F_{10,10}(0.025)=3.72$, $F_{9,9}(0.025)=4.03$, $F_{10,10}(0.05)=2.98$, $F_{9,9}(0.05)=3.18$,
$t_{20}(0.025)=2.086$, $t_{19}(0.025)=2.093$, $t_{18}(0.025)=2.101$, $\chi^2_{11}(0.05)=24.725$.

10 若一个总体的期望和方差分别为μ和σ^2，设$X_1,\ldots,X_n$为来自该总体的一组简单样本，则总体变异系数$\frac{\sigma}{\mu}$的相合估计为______

二. (15分) 假设$Y\sim U(0,\theta),\theta>1$，若随机变量$X=\begin{cases}Y, & Y\geq 1\\ 0, & Y<1\end{cases}$。试求

(1)X的分布函数。(2) 期望EX。

三. (15分) 设某生产线上组装每件产品的时间服从指数分布，平均需要10分钟,且各产品的组装时间是相互独立的。

(1) 试求组装100件产品需要15小时至20小时的概率。

(2) 保证有95%的可能性，问16小时内最多可以组装多少件产品。

四. (15分) 称随机变量$X\sim Exp(a,b)$，如果X的概率密度函数为$f(x)=\frac{1}{b}e^{-(x-a)/b}I(x>a)$，其中$a\in\mathbb{R},b>0$为参数。现从总体$Exp(a,1)$中抽取简单样本$X_1,\ldots,X_n$，从总体$Exp(a,2)$中抽取简单样本$Y_1,\ldots,Y_m$，且两组样本相互独立。试

(1) 求a的矩估计和最大似然估计。

(2) 是否都为无偏估计？若不是，请修正为无偏估计，并比较修正后的估计何者最优。

五. (15分) StreetInsider.com报道了2002年一些著名公司的每股收益的数据.在2002年之前，财务分析家就预测了这些公司2002年的每股收益.利用下面数据评论实际的和预测的每股收益的差异.

公司	实际每股收益	预测每股收益	公司	实际每股收益	预测每股收益股收益
AT&T	1.29	0.38	埃克森-美孚	2.72	2.19
美国运通	2.01	2.31	通用电气	1.51	1.71
花旗银行	2.59	3.43	强生	2.28	2.18
可口可乐	1.60	1.78	麦当劳	0.77	1.55
杜邦	1.84	2.18	沃尔玛	1.81	1.74

试

(1) 在显著性水平0.05下，检验实际的和预测的每股平均收益之间是否存在差异，你的结论是什么？

(2) 两均值之差的点估计是多少？分析家是低估还是高估了每股的收益？

(3) 给出两均值之差的95%置信区间，并据此对(1)的检验问题作出结论并解释。

六. (10分) 用甲、乙、丙、丁四种棉纱织成坯布，其中用甲纱织成的18匹坯布中17匹为上等品，乙纱织成的15匹坯布中11匹为上等品，丙纱织成的15匹坯布中8匹为上等品，丁纱织成的13匹坯布中11匹为上等品，问这四种棉纱的质量有无显著差异？(显著性水平为0.05)

附录 分布及上分位数: $u_{0.025}=1.960$, $u_{0.05}=1.645$, $\chi^2_1(0.05)=3.841$, $\chi^2_2(0.05)=5.990$, $t_9(0.025)=2.262$, $\chi^2_3(0.05)=7.815$, $t_9(0.05)=1.833$, $\chi^2_4(0.05)=9.488$.

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$\mathrm{P}(X < 0) = (\quad)$

(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".