清华大学试题专用纸

考试课程：概率论与数理统计　考试时间：2021 年 1 月 10 日 19：00—21：00

姓名____________学号 **20**____________班级________.

一 填空题（每空 3 分，共 30 分；答案均写在试卷上，注意标清题号）

1. 线上内容中提到过的贝特朗奇论(Bertrand’s paradox)中，不同理解下得到的概率值分别是__________。

$$\frac{1}{2},\frac{1}{3},\frac{1}{4}$$

2. (X,Y) 的联合密度函数 $p(x,y)=e^{-x-y}I_{x>0,y>0}$， $Z=\begin{cases}Y, & 若X\ge Y\\ 2X, & 若X<Y\end{cases}$，则 $E(Z^2)=$________________。

解：$E(Z^2)=5\cdot\int_0^{+\infty}y^2e^{-y}dy\int_y^{+\infty}e^{-x}dx=5\cdot\int_0^{+\infty}y^2e^{-y}e^{-y}dy$

$$=\frac{5}{2}\cdot\int_0^{+\infty}y^2 2\cdot e^{-2y}dy=\frac{5}{2}\left(\frac{1}{2^2}+\frac{1}{2^2}\right)=\frac{5}{4}$$

3. 将一枚均匀的硬币独立地抛掷 16 次，正面出现次数记为 X，希望对概率 $P(6\le X\le 10)$ 做尽可能准确的估计，用中心极限定理的结果为____________________，用切比雪夫不等式的结果为________________________。

解：$X\sim b(16,0.5)$， $X\dot\sim N(8,4)$

$$P(6\le X\le 10)=P(5.5\le X\le 10.5)\approx P\left(\frac{5.5-8}{2}\le\frac{X-8}{2}\le\frac{10.5-8}{2}\right)=2\Phi(1.25)-1=0.78$$

$$P(6\le X\le 10)=P(5.5\le X\le 10.5)=P\left(\left|X-E(X)\right|\le 2.5\right)\ge 1-\frac{4}{2.5^2}=0.36$$

$$P(6\le X\le 10)=P\left(\left|X-E(X)\right|\le 2\right)\ge 1-\frac{4}{2^2}=0$$

$$P(6\le X\le 10)=P(5<X<11)=P\left(\left|X-E(X)\right|<3\right)\ge 1-\frac{4}{3^2}=\frac{5}{9}$$

4. 二维随机变量 $(X,Y)\sim N(0,0,4,4,0.5)$，则 $E(X^2\mid X+Y=4)=$______________________________。

解析： $Cov(X+Y,X-Y)=0$， $X+Y\sim N(0,12)$， $X-Y\sim N(0,4)$

$$E(X^2\mid X+Y=4)=E\left(\left(\frac{X+Y+X-Y}{2}\right)^2\mid X+Y=4\right)$$

$$=E\left(\left(\frac{X+Y}{2}\right)^2 \mid X+Y=4\right)+E\left(\left(\frac{X-Y}{2}\right)^2 \mid X+Y=4\right)+E\left(2\left(\frac{X+Y}{2}\right)\left(\frac{X-Y}{2}\right)\mid X+Y=4\right)$$

$$=4+E\left(\left(\frac{X-Y}{2}\right)^2\right)=5$$

5. 设 $X_1,X_2,\ldots,X_8$ 为相互独立的 $N(0,1)$ 随机变量，则 $P\left(\frac{X_1-X_2}{|X_1+X_2+\cdots+X_8|}\geq \_\_\_\_\_\_\_\_\right)=0.05$。

解：$X_1-X_2\sim N(0,2)$，$X_1+X_2+\ldots+X_8\sim N(0,8)$

$Cov(X_1-X_2,X_1+X_2)=0$，

$$\frac{(X_1-X_2)^2/2}{(X_1+X_2+\ldots+X_8)^2/8}\sim F(1,1),\quad P\left(\frac{(X_1-X_2)^2/2}{(X_1+X_2+\ldots+X_8)^2/8}\geq F_{0.9}(1,1)\right)=0.1$$

$$P\left(\frac{(X_1-X_2)/\sqrt{2}}{|X_1+X_2+X_3+X_4|/2\sqrt{2}}\geq\sqrt{F_{0.9}(1,1)}\right)=0.05\Rightarrow P\left(\frac{X_1-X_2}{|X_1+X_2+X_3+X_4|}\geq\frac{\sqrt{F_{0.9}(1,1)}}{2}\right)=0.05$$

$$P\left(\frac{X_1-X_2}{|X_1+X_2+X_3+X_4|}\geq\sqrt{10}\right)=0.05$$

$$\frac{(X_1-X_2)/\sqrt{2}}{\sqrt{(X_1+X_2+\ldots+X_8)^2/8}}\sim t(1),\quad P\left(\frac{(X_1-X_2)/\sqrt{2}}{|X_1+X_2+\ldots+X_8|/\sqrt{8}}\geq t_{0.95}(1)\right)=0.05$$

$$P\left(\frac{X_1-X_2}{|X_1+X_2+X_3+X_4|}\geq\frac{t_{0.95}(1)}{2}\right)=0.05,\quad P\left(\frac{X_1-X_2}{|X_1+X_2+X_3+X_4|}\geq 3.16\right)=0.05$$

6. 对均匀总体 $U(0,\theta)$ 做假设检验，原假设与备择假设分别为 $H_0:\theta=5,\ H_1:\theta<5$，以 $x_{(n)}=\max\{x_1,\cdots,x_n\}$ 为检验统计量，显著性水平 $\alpha=0.064$，若样本容量 $n=3$，则拒绝域为________________。

解：$P\left(x_{(n)}<c\right)=\left(\frac{c}{5}\right)^n=\alpha\Rightarrow\left(\frac{c}{5}\right)^3=0.064\Rightarrow c=2$，　$\{(x_1,x_2,x_3):x_{(3)}<2\}$

7. 设 $X_1,\ldots,X_{100}$ 是来自总体 $N(\mu,4)$ 的样本。对期望进行假设检验，$H_0:\mu=0,\ H_1:\mu>0$，若取拒绝域为 $\{(x_1,\ldots,x_{100}):\bar{x}>0.2\}$，则此检验犯第一类错误的概率为________________，当 $\mu=0.6$ 时，此检验犯第二类错误的概率为________________，$\bar{x}=0.2$ 的 p 值为________________。

解：第一类错误 $P_{\mu=0}(\bar{X}>0.2)=P_{\mu=0}\left(\dfrac{\bar{X}}{\sqrt{\frac{4}{100}}}>1\right)=1-\Phi(1)=0.16$，

第二类错误 $P_{\mu=0.6}\left(\bar{X}\le 0.2\right)=P_{\mu=0.6}\left(\dfrac{\bar{X}-0.6}{\sqrt{\frac{4}{100}}}\le\dfrac{0.2-0.6}{0.2}\right)=\Phi(-2)=1-\Phi(2)=0.02$

p值 $=P_{\mu=0}(\bar{X}>0.2)=P_{\mu=0}\left(\dfrac{\bar{X}}{\sqrt{\frac{4}{100}}}>\dfrac{0.2}{0.2}\right)=1-\Phi(1)=0.16$

二.（10 分）学生在做有 4 个选项的选择题时，如果他知道正确答案就一定可以答对，如果不知道问题的正确答案，就随机猜 1 个选项。假设学生知道正确答案的概率是 0.6，不知道正确答案的概率是 0.4。

（1）求答对的概率，（2）现从卷面上看题是答对了，求学生知道正确答案的概率。

解：设事件 A 为知道正确答案，事件 B 为答对

（1）由全概率公式 $P(B)=P(B|A)P(A)+P(B|\bar{A})P(\bar{A})=1\times0.6+0.25\times(1-0.6)=0.7$

（2） 由贝叶斯公式 $P(A|B)=\dfrac{P(B|A)P(A)}{P(B)}=\dfrac{1\times0.6}{0.7}=\dfrac{6}{7}$。

三.（10 分）设随机变量 $X\sim N(0,2)$，$Y=|X|$，求随机变量 Y 的分布函数、密度函数、期望和方差。

解：当 $y>0$ 是，分布函数

$$F_Y(y)=P(Y\le y)=P(|X|\le y)=P(-y\le X\le y)=P\left(\frac{-y}{2}\le\frac{X}{2}\le\frac{y}{2}\right)$$

$$=\Phi\left(\frac{y}{2}\right)-\Phi\left(\frac{-y}{2}\right)=2\Phi\left(\frac{y}{2}\right)-1$$

当 $y\le 0$ 是，分布函数 $F_Y(y)=0$

密度函数 $p_Y(y)=\dfrac{d\left(2\Phi\left(\frac{y}{2}\right)-1\right)}{dy}=\varphi\left(\dfrac{y}{2}\right)I_{y>0}$， 或 $p_Y(y)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{8}}I_{y>0}$

$$E(Y)=E(|X|)=\int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{x^2}{4}}dx=2\int_0^{+\infty}x\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{x^2}{4}}dx=2\int_0^{+\infty}\frac{1}{\sqrt{\pi}}e^{-\frac{x^2}{4}}d\frac{x^2}{4}=\frac{2}{\sqrt{\pi}}$$

$E\left(Y^2\right)=E\left(X^2\right)=2$，

$Var\left(Y\right)=E\left(Y^2\right)-E\left(Y\right)^2=2-\dfrac{4}{\pi}$。

四（10 分）X,Y 均服从参数为 1 的指数分布，且相互独立。令 $U=\max\left(X,Y\right)$，$V=\min\left(X,Y\right)$

（1）分别求 U、V 的密度函数，（2）计算 U、V 的相关系数。

解：（1）$F_U\left(u\right)=P\left(U\le u\right)=P\left(\max\left(X,Y\right)\le u\right)=P\left(X\le u,Y\le u\right)$

$$F_U\left(u\right)=P\left(U\le u\right)=P\left(\max\left(X,Y\right)\le u\right)=P\left(X\le u,Y\le u\right)=\left(1-e^{-u}\right)^2$$

$$p_U\left(u\right)=P\left(U\le u\right)=P\left(\max\left(X,Y\right)\le u\right)=P\left(X\le u,Y\le u\right)=2e^{-u}\left(1-e^{-u}\right)=2\left(e^{-u}-e^{-2u}\right)I_{u>0}$$

$$F_V\left(v\right)=P\left(V\le v\right)=P\left(\min\left(X,Y\right)\le v\right)=1-P\left(\min\left(X,Y\right)>v\right)=1-P\left(X>v,Y>v\right)=1-e^{-2v}$$

$$p_V\left(v\right)=2e^{-2v}I_{v>0}$$

（2）$E\left(U\right)=\int_0^{+\infty}u\cdot 2\left(e^{-u}-e^{-2u}\right)du=2-\dfrac{1}{2}=\dfrac{3}{2}$

$$E\left(U^2\right)=\int_0^{+\infty}u^2\cdot 2\left(e^{-u}-e^{-2u}\right)du=2\cdot 2-\left(\frac{2}{2^2}+\frac{2}{2^2}\right)=\frac{7}{2},\quad Var\left(U\right)=\frac{7}{2}-\frac{9}{4}=\frac{5}{4}$$

$V\sim Exp\left(2\right)$，$E\left(V\right)=\dfrac{1}{2}$，$Var\left(V\right)=\dfrac{1}{4}$

$$Cov\left(U,V\right)=E\left(UV\right)-E\left(U\right)E\left(V\right)=E\left(XY\right)-E\left(U\right)E\left(V\right)$$

$$=E\left(X\right)E\left(Y\right)-E\left(U\right)E\left(V\right)=1-\frac{3}{2}\cdot\frac{1}{2}=\frac{1}{4}$$

$$\mathrm{Corr}\left(U,V\right)=\frac{Cov\left(U,V\right)}{\sqrt{Var\left(U\right)Var\left(V\right)}}=\frac{1/4}{\sqrt{\frac{1}{4}\cdot\frac{5}{4}}}=\frac{1}{\sqrt{5}}$$。

五.（10 分）泊松总体 $X\sim P\left(\lambda\right)$，$X_1,\cdots,X_n$ 为来自该总体的样本，

（1）用最大似然估计法给出总体期望的估计量，

（2）试给出参数 λ^2 的无偏估计量。

解：(1) 似然函数 $L(\lambda;x_1,x_2,\cdots,x_n)=\prod_{k=1}^{n}\frac{\lambda^{x_k}}{x_k!}\cdot e^{-\lambda}=\frac{1}{x_1!x_2!\cdots x_n!}\lambda^{x_1+x_2+\cdots+x_n}\cdot e^{-n\lambda}$

对数似然函数 $\ln L(\lambda)=C+(x_1+x_2+\cdots+x_n)\ln\lambda-n\lambda$

$$\max\ln L(\lambda)\Rightarrow\frac{d\ln L(\lambda)}{d\lambda}=\frac{1}{\lambda}(x_1+x_2+\cdots+x_n)-n=0$$

解得 $\lambda=\frac{x_1+x_2+\cdots+x_n}{n}$，所以参数$\lambda$的最大似然估计量 $\hat{\lambda}=\bar{x}$。

(2) $E(X)=\lambda,Var(X)=\lambda$，于是

$E(X^2)=Var(X)+\left(E(X)\right)^2=\lambda^2+\lambda,\quad E(X^2-X)=E(X^2)-E(X)=\lambda^2$

于是，$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}\right)=\lambda^2$，即$X_1^2-X_1$或$\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}$是$\lambda^2$的无偏估计。

六.（10 分）设总体 $X\sim Ge(p)$，$X_1,X_2,\cdots,X_n$ 为来自该总体的样本。定义 $Y=\begin{cases}1, & X_1=1\\0, & X_1>1\end{cases}$

(1) 计算 $E\left(Y\middle|X_1+X_2+\cdots+X_n\right)$

(2) 判断 Y 与 $E\left(Y\middle|X_1+X_2+\cdots+X_n\right)$ 是否为参数 p 的无偏估计量，若否可否进行无偏校正。

解：(1) $X_1+X_2+\cdots+X_n\sim Nb(n,p)$

$$P(X_1+X_2+\cdots+X_n=k)=\binom{k-1}{n-1}p^n(1-p)^{k-n},\quad k\ge n$$

$$P\left(Y=1\middle|X_1+X_2+\cdots+X_n=k\right)=\frac{P(X_1=1,X_1+X_2+\cdots+X_n=k)}{\binom{k-1}{n-1}p^n(1-p)^{k-n}}$$

$$=\frac{P(X_1=1)P(X_2+\cdots+X_n=k-1)}{\binom{k-1}{n-1}p^n(1-p)^{k-n}}=\frac{p\cdot\binom{k-2}{n-2}p^{n-1}(1-p)^{k-n}}{\binom{k-1}{n-1}p^n(1-p)^{k-n}}=\frac{n-1}{k-1}$$

$$E\left(Y\middle|X_1+X_2+\cdots+X_n=k\right)=P\left(Y=1\middle|X_1+X_2+\cdots+X_n=k\right)=\frac{n-1}{k-1}$$

所以 $E\left(Y\middle|X_1+X_2+\cdots+X_n\right)=\frac{n-1}{X_1+X_2+\cdots+X_n-1}$

(2) $E(Y)=P(Y=1)=P(X_1=1)=p$

$$E\left(E\left(Y\middle|X_1+X_2+\cdots+X_n\right)\right)=E(Y)=p,$$

Y与$E\left(Y\middle|X_1+X_2+\cdots+X_n\right)$均为统计量，且期望都等于$p$，所以它们都是参数$p$的无偏估计量。

七.（15分）$X_1,X_2,\cdots,X_n$是来自总体$X\sim N\left(\mu_1,\sigma^2\right)$的样本，$Y_1,Y_2,\cdots,Y_m$是来自总体$Y\sim N\left(\mu_2,c\cdot\sigma^2\right)$的样本，$n=4,m=6,c=4.5,\bar{x}=52,\bar{y}=47,s_X^2=30,s_Y^2=99$。

（1）给出参数$\mu_1-\mu_2$的矩估计量，并计算该估计量的期望和方差；

（2）若已知$\sigma^2=25$，给出参数$\mu_1-\mu_2$的95%置信水平的双侧置信区间；

（3）若σ^2未知，给出参数$\mu_1-\mu_2$的90%置信水平的单侧置信区间的置信下界。

解 （1）$E\left(\bar{X}-\bar{Y}\right)=\mu_1-\mu_2$，所以参数$\mu_1-\mu_2$的矩估计量为$\bar{X}-\bar{Y}$；

$$\bar{X}\sim N\left(\mu_1,\frac{\sigma^2}{n}\right),\ \bar{Y}\sim N\left(\mu_2,\frac{c\cdot\sigma^2}{m}\right),。$$

$$E\left(\bar{X}-\bar{Y}\right)=\mu_1-\mu_2\ ,\quad Var\left(\bar{X}-\bar{Y}\right)=\left(\frac{1}{n}+\frac{c}{m}\right)\sigma^2=\sigma^2。$$

（2） $\dfrac{\bar{X}-\bar{Y}-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{c}{m}}}\sim N\left(0,1\right)$可作为枢轴量，$\sqrt{\dfrac{1}{n}+\dfrac{c}{m}}=1$

$P\left(-u_{0.975}\le\dfrac{\bar{X}-\bar{Y}-\left(\mu_1-\mu_2\right)}{\sigma}\le u_{0.975}\right)=0.9$，参数$\mu_1-\mu_2$的95%双侧置信区间为

$\left[\bar{X}-\bar{Y}-u_{0.975}\cdot\sigma,\bar{X}-\bar{Y}+u_{0.975}\cdot\sigma\right]$，即$\left[-4.8,14.8\right]$

（3） $\dfrac{(n-1)s_X{}^2}{\sigma^2}\sim\chi^2\left(n-1\right)$， $\dfrac{(m-1)s_Y{}^2}{c\cdot\sigma^2}\sim\chi^2\left(m-1\right)$ ， $\dfrac{(n-1)s_X{}^2}{\sigma^2}+\dfrac{(m-1)s_Y{}^2}{c\cdot\sigma^2}\sim\chi^2\left(n+m-2\right)$，

$$\frac{\dfrac{\bar{X}-\bar{Y}-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{c}{m}}}}{\dfrac{1}{\sigma}\sqrt{\left((n-1)s_X{}^2+\dfrac{(m-1)s_Y{}^2}{c}\right)\Big/(n+m-2)}}=\frac{\dfrac{\bar{X}-\bar{Y}-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{1}{n}+\dfrac{c}{m}}}}{\sqrt{\left((n-1)s_X{}^2+\dfrac{(m-1)s_Y{}^2}{c}\right)\Big/(n+m-2)}}\sim t\left(n+m-2\right)$$

$$\sqrt{\frac{1}{n}+\frac{c}{m}}=1\ ,\quad\sqrt{\left((n-1)s_X{}^2+\frac{(m-1)s_Y{}^2}{c}\right)\Big/(n+m-2)}=5$$

$$\frac{\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{1}}{5}=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{5}\sim t(8),\quad P\left(\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{5}\le t_{0.9}(8)\right)=0.9$$

参数$\mu_1-\mu_2$的90%置信水平的单侧置信区间的置信下界为$\bar{X}-\bar{Y}-5\cdot t_{0.9}(8)=5-5\cdot 1.4=-2$

八. (5分) 设$y_1,\cdots,y_n$是$x_1,\cdots,x_n$按照升序排列后的结果，$y_1<y_2<\cdots<y_n$。用随机化快速排序算法对$x_1,\cdots,x_n$进行排序，每一次都是从所有可能的元素中独立且均匀地选取基准元素，对$1\le i<j\le n$，定义随机变量X_{ij}，如果在算法过程中y_i与y_j进行了比较$X_{ij}=1$，否则$X_{ij}=0$。求$E(X_{ij})$.

解：考虑$\{y_i,y_{i+1},\cdots,y_j\}$，当且仅当$y_i$或$y_j$是集合$\{y_i,y_{i+1},\cdots,y_j\}$中第一个被选作基准的元素时，$y_i$与$y_j$进行比较。

所以$P(X_{ij}=1)=\dfrac{2}{j-i+1}$，　$E(X_{ij})=\dfrac{2}{j-i+1}$。

备注 1. 本考卷的样本均为简单随机样本，样本均值$\bar{X}=\dfrac{X_1+\cdots+X_n}{n}$，样本方差为$S^2=\dfrac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2$

备注 2. 正态总体的样本均值和样本方差相互独立，且$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，其中$n$为样本容量

备注 3. 泊松分布随机变量$X\sim P(\lambda)$的分布列 $P(X=k)=e^{-\lambda}\lambda^k/k!$

备注 4. 指数分布随机变量$X\sim Exp(\lambda)$，$F_X(x)=(1-e^{-\lambda x})I_{x>0}$，$E(X)=1/\lambda, Var(X)=1/\lambda^2$

备注 5. 第三题解答中标准正态随机变量的分布函数和密度函数分别可用$\Phi(x)$和$\varphi(x)$表示

备注 6. 分布函数和分位数（题目解答要严格按照下面给出数值进行计算）

$\Phi(1.28)=0.9,\quad \Phi(1.44)=0.925,\quad \Phi(1.65)=0.95,\quad \Phi(1.96)=0.975,$
$\Phi(1)=0.84,\quad \Phi(1.25)=0.89,\quad \Phi(1.5)=0.93,\quad \Phi(1.75)=0.96,\quad \Phi(2)=0.98,\quad \Phi(3)=0.999$

$P(t(7)>1.41)=0.1, P(t(8)>1.40)=0.1, P(t(9)>1.38)=0.1, P(t(10)>1.37)=0.1$

$P(t(7)>1.89)=0.05, P(t(8)>1.86)=0.05, P(t(9)>1.83)=0.05, P(t(10)>1.81)=0.05$

$P(F(1,1)>39.9)=0.1, P(F(1,2)>8.53)=0.1, P(F(1,5)>4.06)=0.1, P(F(2,1)>49.5)=0.1, P(F(5,1)>57.2)=0.1$

$P(F(1,1)>161)=0.05, P(F(1,2)>18.5)=0.05, P(F(1,5)>6.61)=0.05, P(F(2,1)>200)=0.05, P(F(5,1)>230)=0.05$