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：2020 年 1 月 5 日 8：00—10：00

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. 设 X 服从几何分布，$E(X|X>2)=5$，则 $P(X=3)=$__________，$Var(X|1<X<4)=$__________。

$$E(X|X>2)=2+E(X)=5\Rightarrow E(X)=3, p=\frac{1}{3},\quad P(X=3)=p(1-p)^2=\frac{4}{27};$$

$$E(X|1<X<4)=2\cdot\frac{3}{5}+3\cdot\frac{2}{5}=\frac{12}{5},\quad E(X^2|1<X<4)=4\cdot\frac{3}{5}+9\cdot\frac{2}{5}=\frac{30}{5}$$

$$Var(X|1<X<4)=\frac{30}{5}-\left(\frac{12}{5}\right)^2=\frac{6}{25}。$$

2. (X,Y) 的联合密度函数 $p(x,y)=\begin{cases}\frac{1}{4}, & 0<x,y<2\\ 0, & 其他\end{cases}$，$Z=\begin{cases}Y, & 若\ X\ge Y\\ X, & 若\ X<Y\end{cases}$，则 $E(Z^2)=$__________。

$$E(Z^2)=2\int_0^2 dx\int_0^x y^2\frac{1}{4}dy=\frac{2}{3}$$

3. 随机变量 X 服从二项分布 $b(n,0.8)$，c 为任意实数，若 $E\left((X-c)^2\right)$ 的最小值为 12，则 $n=$______。

$$Var(X)=n\cdot 0.8\cdot 0.2=0.16n=12,\quad n=75$$

4. 如果 $0<Var(X)<+\infty$，则 $Var\left(\frac{X-Var(X)}{\sqrt{Var(X)}}\right)=$__________________。

$$Var\left(\frac{X-Var(X)}{\sqrt{Var(X)}}\right)=1$$

5. 已知随机变量 $X\sim U(0,2)$，则 $Var\left(\min\{X^2,1\}\right)=$ __________________。

$$E\left(\min\left\{X^2,1\right\}\right)=\int_0^1 x^2\cdot\frac{1}{2}dx+\int_1^2 1\cdot\frac{1}{2}dx=\frac{1}{6}+\frac{1}{2}=\frac{2}{3}$$

$$E\left[\left(\min\left\{X^2,1\right\}\right)^2\right]=\int_0^1 x^4\cdot\frac{1}{2}dx+\int_1^2 1\cdot\frac{1}{2}dx=\frac{1}{10}+\frac{1}{2}=\frac{3}{5}$$

$$Var\left(\min\left\{X^2,1\right\}\right)=E\left[\left(\min\left\{X^2,1\right\}\right)^2\right]-E\left(\min\left\{X^2,1\right\}\right)=\frac{3}{5}-\left(\frac{2}{3}\right)^2=\frac{7}{45}$$

6. 设随机变量 X 的密度函数为 $p(x)=3e^{-3x}I_{x\ge 0}$，则 $E\left(X^2\cdot e^{\frac{X}{3E(X)}}\right)=$____________________。

$$E(X)=\frac{1}{3},\quad E\left(X^2\cdot e^{\frac{X}{3E(X)}}\right)=E\left(X^2\cdot e^X\right)=\int_0^{+\infty}x^2e^x3e^{-3x}dx=\frac{3}{2}\int_0^{+\infty}x^2 2e^{-2x}dx=\frac{3}{2}\left(\frac{1}{4}+\frac{1}{4}\right)=\frac{3}{4}$$

7. 将一枚均匀的硬币独立地抛掷 100 次，则正面次数，用中心极限定理和切比雪夫不等式估计正面次数在 45 至 55 之间的概率，结果分别为________________________；_________________________。

$$X\sim b\left(100,\frac{1}{2}\right),\quad E(X)=100,\ Var(X)=25,\quad X\dot{\sim}N(50,25)$$

$$P(44.5<X<55.5)\approx P\left(\frac{44.5-50}{5}<\frac{X-50}{5}<\frac{55.5-50}{5}\right)=2\Phi(1.1)-1$$

或 $$P(45<X<55)\approx P\left(\frac{45-50}{5}<\frac{X-50}{5}<\frac{55-50}{5}\right)=2\Phi(1)-1=0.68$$

$$P(45<X<55)=P(|X-50|<5)>1-\frac{Var(X)}{5^2}=0,\ P(45<X<55)>0$$

8. 设二维随机变量 (X,Y) 的密度函数为 $p(x,y)=\begin{cases}1, & 0<|x|<y<1\\ 0, & ohterwise\end{cases}$，则 $E\left(X^2|Y\right)=$__________________。

当 $0<y<1$ 时，$X|Y=y\sim U(-y,y)$，　$p(x|Y=y)=\frac{1}{2y}I_{-y<x<y}$

$$E\left(X^2|Y=y\right)=\int_{-y}^{y}x^2\frac{1}{2y}dx=\frac{y^2}{3},\quad 所以\ E\left(X^2|Y\right)=\frac{Y^2}{3}。$$

二. （10 分） 某厂有甲、乙、丙三车间生产同一种产品，产量分别占总产量的 60%，30%和 10%。各车间的次品率分别是 2%，5%，6%。试求

（1）在该厂产品中任取一件，恰为次品的概率；

（2）若发现一件产品为次品，该次品来自甲车间的概率？

解：设事件 A 为次品，产品由甲、乙、丙三车间生产分别设为事件 B_1,B_2,B_3

（1） $P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)$

$$=0.6\times0.02+0.3\times0.05+0.1\times0.06=0.033$$

（2） $P(B_1|A)=\dfrac{P(B_1)P(A|B_1)}{P(A)}=\dfrac{0.6\times0.02}{0.033}=\dfrac{4}{11}$。

三. (8 分) 设随机变量 $X\sim U(0,2)$，$Y=X^3$，求随机变量 Y 的分布函数、密度函数、期望和方差。

解：$p_X(x)=\dfrac{1}{2}I_{0<x<2}$，

当 $y\le0$ 时，$F_Y(y)=0$；当 $y\ge8$ 时，$F_Y(y)=1$；

当 $0<y<8$ 时，$F_Y(y)=P(Y<y)=P(X^3<y)=P(X<\sqrt[3]{y})=\dfrac{\sqrt[3]{y}}{2}$；

Y 的分布函数 $F_Y(y)=\begin{cases}0, & y\le0\\ \dfrac{\sqrt[3]{y}}{2}, & 0<y<8\\ 1, & y\ge8\end{cases}$，

Y 的密度函数 $p_Y(y)=\dfrac{dF_Y(y)}{dy}=\dfrac{1}{6}y^{-\frac{2}{3}}I_{0<y<8}$；

$$E(Y)=\int_{-\infty}^{+\infty}yp_Y(y)dy=\int_0^8 y\frac{1}{6}y^{-\frac{2}{3}}dy=\frac{1}{6}\cdot\frac{3}{4}\cdot y^{\frac{4}{3}}\Big|_0^8=2$$

$$E(Y^2)=\int_{-\infty}^{+\infty}y^2p_Y(y)dy=\int_0^8 y^2\frac{1}{6}y^{-\frac{2}{3}}dy=\frac{1}{6}\cdot\frac{3}{7}\cdot y^{\frac{7}{3}}\Big|_0^8=\frac{64}{7}$$

$$Var(Y)=E(Y^2)-E(Y)^2=\frac{64}{7}-4=\frac{36}{7}。$$

四. (12 分) 随机变量 (X_1,X_2) 的密度函数为 $p(x,y)=\dfrac{1}{2\sqrt{3}\,\pi}e^{\frac{-2}{3}\left[(x-1)^2-\frac{(x-1)(y-2)}{2}+\frac{(y-2)^2}{4}\right]}$，$Y=2X_1+X_2$。

（1）求Y的分布；（2）计算X_1和Y的相关系数$Corr(X_1,Y)$；（3）计算$E(X_1|Y=1)$。

解：（1）(X_1,X_2)服从二维正态分布$N(1,2,1,4,-0.5)$，$Y=2X_1+X_2$服从正态分布

$$E(Y)=E(2X_1+X_2)=2E(X_1)+E(X_2)=4$$

$$Cov(X_1,X_2)=\rho\cdot\sigma_1\cdot\sigma_2=-0.5\cdot 2=-1$$

$$Var(Y)=Var(2X_1+X_2)=4Var(X_1)+Var(X_2)+4Cov(X_1,X_2)$$

$$=4\cdot 1+4-4\cdot 1=4$$

$Y\sim N(4,2^2)$。

（2）$Cov(X_1,Y)=Cov(X_1,2X_1+X_2)=2Var(X_1)+Cov(X_1,X_2)=2-1=1$

$$Corr(X_1,Y)=\frac{Cov(X_1,Y)}{\sqrt{Var(X_1)\cdot Var(Y)}}=\frac{1}{1\cdot 2}=0.5$$

（3）因为 $Cov(X_1,Y)=Cov(X_1,2X_1+X_2)=1$，$Var(Y)=4$

所以 $Cov\left(X_1-\frac{Y}{4},Y\right)=0$，$\left(X_1-\frac{Y}{4},Y\right)$服从二位正态分布，所以相互独立

$$E(X_1|Y=1)=E\left(X_1-\frac{Y}{4}+\frac{Y}{4}\middle|Y=1\right)=E\left(X_1-\frac{Y}{4}\right)+\frac{1}{4}=\frac{1}{4}。$$

五 (6分)抛掷一枚6面的色子，出现1点至6点的概率均为$1/6$，抛掷过程是相互独立的。直至首次连续出现1点6点停止。例如：3, 2, 3, 5, 1, 1, 6停止，抛掷7次。试求抛掷次数的期望。

解：设抛掷次数为随机变量X，第k次投掷得到m点记为$Y_k=m$，

$$E(X)=E(X|Y_1=1)P(Y_1=1)+E(X|Y_1\neq 1)P(Y_1\neq 1)=E(X|Y_1=1)\frac{1}{6}+(1+E(X))\frac{5}{6}$$

即 $E(X)-E(X|Y_1=1)=5$

$$E(X|Y_1=1)=E(X|Y_1=1,Y_2=1)P(Y_2=1)+E(X|Y_1=1,Y_2=6)P(Y_2=6)$$

$$+E\left(X\left|Y_1=1,Y_2=2,3,4,5\right.\right)P\left(Y_2=2,3,4,5\right)$$

$$=\left(1+E\left(X\left|Y_1=1\right.\right)\right)\frac{1}{6}+2\cdot\frac{1}{6}+\left(2+E\left(X\right)\right)\frac{4}{6}$$

即 $5E\left(X\left|Y_1=1\right.\right)-4E\left(X\right)=11$

解得 $E\left(X\right)=36$。

六．(8 分) 样本 $X_1,X_2,\cdots,X_n$ 来自总体 X， $E\left(X\right)=\mu$， $Var\left(X\right)=\sigma^2$，证明 $E\left(S^2\right)=\sigma^2$。

证明： $E\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=E\sum_{i=1}^{n}\left[\left(x_i-E\left(x_i\right)\right)-\left(\bar{x}-E\left(\bar{x}\right)\right)\right]^2$

$$=\sum_{i=1}^{n}E\left(x_i-E\left(x_i\right)\right)^2+\sum_{i=1}^{n}E\left(\bar{x}-E\left(\bar{x}\right)\right)^2-2\cdot E\sum_{i=1}^{n}\left[\left(x_i-E\left(x_i\right)\right)\cdot\left(\bar{x}-E\left(\bar{x}\right)\right)\right]$$

$$=\sum_{i=1}^{n}Var\left(x_i\right)+\sum_{i=1}^{n}Var\left(\bar{x}\right)-2\cdot E\left[\left(\sum_{i=1}^{n}x_i-nE\left(x_i\right)\right)\cdot\left(\bar{x}-E\left(\bar{x}\right)\right)\right]$$

$$=n\sigma^2+n\cdot Var\left(\bar{x}\right)-2\cdot E\left[\left(n\bar{x}-nE\left(\bar{x}\right)\right)\cdot\left(\bar{x}-E\left(\bar{x}\right)\right)\right]$$

$$=n\sigma^2+n\cdot Var\left(\bar{x}\right)-2n\cdot Var\left(\bar{x}\right)=n\sigma^2\text{-}n\cdot Var\left(\bar{x}\right)=\left(n-1\right)\sigma^2$$

所以 $E\left(S^2\right)=E\left(\frac{1}{n-1}\sum_{k=1}^{n}\left(X_k-\bar{X}\right)^2\right)=\sigma^2$。

七．(10 分) $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim Exp\left(\lambda\right)$ 的样本。

(1) 指定常数 $T>0$，设样本中取值小于 T 的样品数为 r，利用比例 $\frac{r}{n}$ 给出参数 λ 的估计量；

(2) 求参数 $\frac{1}{\lambda}$ 的极大似然估计量。

解 (1) $P\left(X\le T\right)=\int_0^T\lambda e^{-\lambda x}dx=1-e^{-\lambda T}$，设样本中取值小于 T 的样品比例为 $\frac{r}{n}$，

令 $1-e^{-\lambda T}=\frac{r}{n}$，解得估计量为 $\hat{\lambda}=\dfrac{\ln\left(1-\frac{r}{n}\right)}{-T}$。

(2) 指数分布的密度函数为 $p\left(x\right)=\lambda e^{-\lambda x}I_{x\ge 0}$

似然函数 $L(x_1,\cdots,x_n;\lambda)=\prod_{k=1}^{n}\lambda e^{-\lambda x_k}=\lambda^n e^{-\lambda(x_1+\cdots+x_n)}$，令 $t=\frac{1}{\lambda}$，则

似然函数 $L(x_1,\cdots,x_n;\lambda)=t^{-n}e^{-\frac{(x_1+\cdots+x_n)}{t}}$

求对数似然函数的最大值，$\max\ln L(t)\Rightarrow\max\left(-n\ln t-\frac{(x_1+x_2+\cdots+x_n)}{t}\right)$

$$\frac{d\ln L(t)}{dt}=\frac{-n}{t}+\frac{(x_1+x_2+\cdots+x_n)}{t^2}=0\Rightarrow\hat{t}=\frac{(x_1+x_2+\cdots+x_n)}{n}=\bar{X}$$

参数 $\frac{1}{\lambda}$ 的极大似然估计量 $\widehat{\frac{1}{\lambda}}=\bar{X}$。

八.（16分）设某工厂生产一种产品，它的一个指标参数服从正态分布 $N(\mu,3^2)$，$\mu\le 10$ 为优级。某一天从产品中随机抽取 36 个样品，测得指标值为 $x_1,\cdots,x_{36}$。对参数 μ 做假设检验，$H_0:\mu\le 10$ VS $H_1:\mu>10$，显著性水平 $\alpha=0.1$。

（1）写出拒绝域的范围；（2）计算 $\bar{x}=11$ 的 p 值；

（3）$\mu=11.5$ 时，若出错是第几类错误，并计算发生这类错误的概率；

（4）样本容量增大到多少时能够保证当 $\mu>10.1$ 时，第二类错误不超过 0.001。

（1）$\mu=10$ 时，检验统计量 $\bar{x}\sim N\left(10,\left(\frac{1}{2}\right)^2\right)$

$P\left(\frac{\bar{x}-10}{\frac{1}{2}}>u_{0.9}\right)=0.1$，所以拒绝域为 $\left\{\bar{x}:\bar{x}-10>\frac{1}{2}u_{0.9}\right\}$

即 $\{\bar{x}:\bar{x}>10.64\}$ 或 $\left\{\bar{x}:\bar{x}>10+\frac{u_{0.9}}{2}\right\}$；

（2）$p=P(\bar{x}-10\ge 1)=P\left(\frac{\bar{x}-10}{\frac{1}{2}}\ge 2\right)=1-\Phi(2)\approx 0.025$；

（3）$\mu=11.5$ 时，若出错是第二类错误，出错概率为

$$\beta = P\left(\bar{x} < 10.64 \middle| \mu = 11.5\right) = P\left(\frac{\bar{x} - 11.5}{1/2} < \frac{10.64 - 11.5}{1/2}\right) = \Phi(-1.72) \approx 0.04$$

（4）设样本容量为n，$\bar{x} \sim N\left(10, \frac{3^2}{n}\right)$，$P\left(\frac{\sqrt{n}\left(\bar{x} - 10\right)}{3} > u_{0.9}\right) = 0.1$

显著性水平$\alpha = 0.1$的拒绝域为$\left\{\bar{x} : \bar{x} > 10 + \frac{3u_{0.9}}{\sqrt{n}}\right\}$

$\mu = 10.1$时， 出第二类错误的概率为

$$\beta = P\left(\bar{x} \le 10 + \frac{3u_{0.9}}{\sqrt{n}} \middle| \mu = 10.1\right) = P\left(\frac{\bar{x} - 10.1}{3/\sqrt{n}} < \frac{10 + \frac{3u_{0.9}}{\sqrt{n}} - 10.1}{3/\sqrt{n}}\right) = \Phi\left(1.28 - \frac{0.1\sqrt{n}}{3}\right) < 0.001$$

$$\beta = P\left(\bar{x} \le 10 + \frac{3u_{0.9}}{\sqrt{n}} \middle| \mu = 10.1\right) = P\left(\frac{\bar{x} - 10.1}{3/\sqrt{n}} < \frac{10 + \frac{3u_{0.9}}{\sqrt{n}} - 10.1}{3/\sqrt{n}}\right) = \Phi\left(1.28 - \frac{0.1\sqrt{n}}{3}\right) < 0.001$$

$$1.28 - \frac{0.1\sqrt{n}}{3} \le u_{0.001} \Rightarrow \sqrt{n} \ge 30\left(1.28 - u_{0.001}\right) \Rightarrow n \ge 30^2\left(1.28 - u_{0.001}\right)^2$$

$$u_{0.001} \approx -3,\ n \ge 30^2\left(1.28 - u_{0.001}\right)^2 \approx 16500$$