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：2022 年 6 月 13 日 9：00—11：00

姓名______________学号 **20**____________班级_________.

1. 设事件 A,B 满足 $P(B)=0.4$，$P(\bar{A}\mid B)=0.7$，$P(\bar{A}\mid\bar{B})=0.3$，则 $P(B\mid A)=$______。

$$P(\bar{A})=P(B)P(\bar{A}\mid B)+P(\bar{B})P(\bar{A}\mid\bar{B})=0.4\times0.7+0.6\times0.3=0.46,$$

$$P(A)=1-P(\bar{A})=0.54,\quad P(B\mid A)=\frac{P(B)P(A\mid B)}{P(A)}=\frac{0.4\times0.3}{0.54}=\frac{2}{9}$$ 或 0.22

2. 设随机变量 X 的分布律为 $P\left(X=k\right)=c\cdot\frac{3^k}{k!},k=1,2,\cdots$。则常数 $c=$______。

解：$$1=\sum_{k=1}^{+\infty}P\left(X=k\right)=\sum_{k=1}^{+\infty}c\cdot\frac{\lambda^k}{k!}=c\left(\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}-1\right)=ce^{\lambda}\left(\sum_{k=0}^{+\infty}e^{-\lambda}\frac{\lambda^k}{k!}-e^{-\lambda}\right)$$

$$1=ce^{\lambda}\left(1-e^{-\lambda}\right)\Rightarrow c=\frac{1}{e^{\lambda}-1}\Rightarrow c=\frac{1}{e^{2}-1}$$

3. 设 $X\sim N\left(1,4\right)$，$P(X<a)=\Phi(-1)$。则 $a=$______________。

【解析】若 $X\sim N\left(1,4\right)$，$\frac{X-1}{2}\sim N(0,1)$ 进行标准化，

$$P(X<a)=P\left(\frac{X-1}{2}<\frac{a-1}{2}\right)=\Phi\left(\frac{a-1}{2}\right)=\Phi\left(-1\right)\Rightarrow a=-1。$$

4. 设 $X\sim N\left(0,1\right)$，$Y=\left|X\right|$，$F_Y\left(y\right)$ 是 Y 的分布函数，则 $F_Y\left(2\right)=$___________。

$$2\Phi\left(y\right)-1=2\Phi\left(2\right)-1=2\cdot0.98-1=0.96$$

$$F_Y\left(y\right)=P\left(Y\le y\right)=P\left(\left|X\right|\le y\right)=P\left(-y\le X\le y\right)=\Phi\left(y\right)-\Phi\left(-y\right)=\Phi\left(y\right)-\left(1-\Phi\left(y\right)\right)=2\Phi\left(y\right)-1$$

5. 随机变量 X 服从负二项分布 $Nb\left(8,0.4\right)$，则 $\min\limits_{c\in R}E\left(\left(X-c\right)^2\right)=$ ___________。

期望的最小二乘性质，$\min E\left(\left(Y-c\right)^2\right)=Var\left(Y\right)=n\cdot3.75=30$

6. 设随机变量 X 的密度函数为 $p(x)=\frac{1}{2}e^{\frac{-1}{2}x}I_{x\geq 0}$，则 $E\left(X^2\cdot e^{\frac{X}{3E(X)}}\right)=(\quad)$

$$E\left(X^2e^{\frac{X}{3E(X)}}\right)=E\left(X^2e^{\frac{X}{6}}\right)=\int_0^{\infty}x^2e^{\frac{x}{6}}\frac{1}{2}e^{-\frac{x}{2}}dx=\frac{1}{2}\int_0^{\infty}x^2e^{-\frac{x}{3}}dx=\frac{1}{2}\cdot 3\cdot\int_0^{\infty}x^2\frac{1}{3}e^{-\frac{x}{3}}dx=\frac{3}{2}E\left(X^2\right)=\frac{3}{2}\left(3^2+3^2\right)=27$$

7. 将一枚均匀的硬币独立地抛掷 n 次，利用中心极限定理估计，若使正面次数的比例在 0.45 次到 0.55 之间的概率不低于 0.96，n 至少要达到____________。

解：正面次数 $X\sim b(n,0.5)$， 所以 $X\dot{\sim}N\left(\frac{n}{2},\frac{n}{4}\right)$， $\frac{X}{n}\dot{\sim}N\left(\frac{1}{2},\frac{1}{4n}\right)$

$$2\sqrt{n}\left(\frac{X}{n}-\frac{1}{2}\right)\dot{\sim}N(0,1),\quad P\left(0.4\leq\frac{X}{n}\leq 0.6\right)=P\left(-0.1\sqrt{n}\leq 2\sqrt{n}\left(\frac{X}{n}-\frac{1}{2}\right)\leq 0.1\sqrt{n}\right)\geq 0.96$$

$$0.1\sqrt{n}\geq u_{0.98}\quad\Rightarrow\quad n\geq 400$$

8. 设总体 $X\sim N\left(\mu,0.5^2\right)$，$X_1,X_2,\cdots,X_n$ 为样本，要使 μ 的置信系数 96%的双侧置信区间长度不超过 0.4，则样本容量 n 至少要达到____________。

解：$X\sim N\left(\mu,0.5^2\right)\Rightarrow\bar{X}\sim N\left(\mu,\frac{0.5^2}{n}\right)\Rightarrow\frac{\bar{X}-\mu}{0.5/\sqrt{n}}\sim N(0,1)$.

$$P\left\{\left|\frac{\bar{X}-\mu}{0.5/\sqrt{n}}\right|\leq u_{0.98}\right\}=0.96\ \Rightarrow\ P\left(\bar{X}-\frac{u_{0.98}}{2\sqrt{n}}\leq\mu\leq\bar{X}+\frac{u_{0.98}}{2\sqrt{n}}\right)$$

置信区间长度为 $2\cdot\frac{u_{0.98}}{2\sqrt{n}}=\frac{2}{\sqrt{n}}$，令 $\frac{2}{\sqrt{n}}\leq 0.4$，得到 $n\geq 25$

9. 设总体 X 服从期望为 θ 的指数分布，$X_1,X_2,\cdots,X_n$ 是来自 X 的简单随机样本，$Y=\min\left(X_1,X_2,\cdots,X_n\right)$，则利用 Y 得到的参数 θ 的无偏统计量是____________。

nY $\qquad Y=\min\left(X_1,X_2,\cdots,X_n\right)\sim Exp(n\lambda)$，$E(Y)=\frac{1}{n\lambda}$

10. 设一批零件的长度服从正态分布 $N\left(\mu,\sigma^2\right)$，其中 μ 与 σ^2 均未知，先从中随机抽取 4 个零件，测得样本均值 $\bar{x}=20$ cm，样本标准差 $s=1$ cm，则 μ 的置信水平 0.9 的置信区间______________。

$$\left[20-\frac{t_{0.95}(3)}{2},20+\frac{t_{0.95}(3)}{2}\right]=[20-1.18,20+1.18]=[18.82,21.18],$$

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 $y\le 0$ 时，$F_Y(y)=0$；当 $y\ge 8$ 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 $F_Y(y)=\begin{cases}0, & y\le 0\\ \dfrac{\sqrt[3]{y}}{2}, & 0<y<8\\ 1, & y\ge 8\end{cases}$，

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．(10 分)随机变量 X_1 以等可能取值为 0 和 1，X_2 以等可能取值为 $0,1,2$，X_1 和 X_2 相互独立

（1）$Y_1 = X_1 - 2X_2$，$Y_2 = X_1 + 2X_2$， 求 Y_1, Y_2 的联合分布；

（2）计算相关系数 $\rho(Y_1, Y_2)$。

解：（1）$(X_1 = 0, X_2 = 0) \to (Y_1 = 0, Y_2 = 0), (X_1 = 0, X_2 = 1) \to (Y_1 = -2, Y_2 = 2)$

$(X_1 = 0, X_2 = 2) \to (Y_1 = -4, Y_2 = 4), (X_1 = 1, X_2 = 0) \to (Y_1 = 1, Y_2 = 1)$

$(X_1 = 1, X_2 = 1) \to (Y_1 = -1, Y_2 = 3), (X_1 = 1, X_2 = 2) \to (Y_1 = -3, Y_2 = 5)$

$P(Y_1 = 0, Y_2 = 0) = \frac{1}{6}$，$P(Y_1 = -2, Y_2 = 2) = \frac{1}{6}$

$P(Y_1 = -4, Y_2 = 4) = \frac{1}{6}$，$P(Y_1 = 1, Y_2 = 1) = \frac{1}{6}$

$P(Y_1 = -1, Y_2 = 3) = \frac{1}{6}$，$P(Y_1 = -3, Y_2 = 5) = \frac{1}{6}$

（2） $E(X_1) = \frac{1}{2}, E(X_1^2) = \frac{1}{2}$，$E(X_2) = 1, E(X_2^2) = \frac{5}{3}$

$$E(X_1X_2) = \frac{1}{6}(1+2) = \frac{1}{2}$$

$$E(Y_1) = E(X_1 - 2X_2) = -\frac{3}{2},\quad E(Y_2) = E(X_1 + 2X_2) = \frac{5}{2}$$

$$Cov(Y_1, Y_2) = E\left((X_1 - 2X_2)(X_1 + 2X_2)\right) - E(Y_1)E(Y_2)$$

$$= E(X_1^2) - 4E(X_2^2) + \frac{15}{4} = -\frac{29}{12}$$

$$Var(Y_1) = E\left((X_1 - 2X_2)^2\right) - \frac{9}{4} = E(X_1^2) + 4E(X_2^2) - 4E(X_1, X_2) = \frac{43}{6} - 2 - \frac{9}{4} = \frac{35}{12}$$

$$Var(Y_2) = E\left((X_1 + 2X_2)^2\right) - \frac{25}{4} = \frac{43}{6} + 2 - \frac{25}{4} = \frac{35}{12}$$

$$\rho(Y_1, Y_2) = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}} = \frac{-\frac{29}{12}}{\sqrt{\frac{35}{12}\frac{35}{12}}} = -\frac{29}{35}。$$

五．（10 分）已知$(X,Y)\sim N(0,0,2,2,0)$，求(1) $E(X|X+Y)$；（2） $E\left(X^2|X+Y=4\right)$。

解： $Cov(X+Y,X-Y)=0$ ，所以X,Y 相互独立，

（1） $E(X|X+Y)=E\left(\frac{X+Y}{2}+\frac{X-Y}{2}\middle|X+Y\right)$

$$=E\left(\frac{X+Y}{2}\middle|X+Y\right)+E\left(\frac{X-Y}{2}\middle|X+Y\right)=\frac{X+Y}{2}+0=\frac{X+Y}{2}$$

或根据对称性 $E(X|X+Y)=E(Y|X+Y)$，

$E(X|X+Y)+E(Y|X+Y)=E(X+Y|X+Y)=X+Y$

所以$E(X|X+Y)=\frac{X+Y}{2}$

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

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

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

$$=4+0+E\left(\left(\frac{X-Y}{2}\right)^2\right)=4+\frac{Var(X-Y)}{4}=4+\frac{Var(X)+Var(Y)}{4}=5$$

六（10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2点停止。例如：3, 2, 3, 5, 1, 1, 2 停止，抛掷7次。试求抛掷次数的期望和方差。

解：设抛掷次数为随机变量X，第k次投掷得到m点记为$Y_k=m$，

$$E(X)=E(X|Y_1=1)P(Y_1=1)+E(X|Y_1\neq 1)P(Y_1\neq 1)$$

$$=\frac{1}{6}E(X|Y_1=1)+\frac{5}{6}(1+E(X)) \quad \Rightarrow \quad E(X)-E(X|Y_1=1)=5$$

$$E(X|Y_1=1)=E(X|Y_1=1,Y_2=1)P(Y_2=1)+E(X|Y_1=1,Y_2=2)P(Y_2=2)$$

$$+E(X|Y_1=1,Y_2\neq 1,2)P(Y_2\neq 1,2)$$

$$=\frac{1}{6}(1+E(X|Y_1=1))+\frac{1}{6}\cdot 2+\frac{4}{6}(2+E(X)) \quad \Rightarrow \quad 5E(X|Y_1=1)-4E(X)=11$$

解得 $E(X)=36$

$$E(X^2)=E(X^2|Y_1=1)P(Y_1=1)+E(X^2|Y_1\neq 1)P(Y_1\neq 1)$$

$$=\frac{1}{6}E(X^2|Y_1=1)+\frac{5}{6}(1+2E(X)+E(X^2))$$

$$E(X^2)-E(X^2|Y_1=1)=365$$

$$E(X^2|Y_1=1)=E(X^2|Y_1=1,Y_2=1)P(Y_2=1)+E(X^2|Y_1=1,Y_2=2)P(Y_1=2)$$
$$+E(X^2|Y_1=1,Y_2\neq 1,2)P(Y_2\neq 1,2)$$

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

$$5E(X^2|Y_1=1)-4E(X^2)=659,$$

$$E(X^2)=2484, \quad Var(X)=1188$$

七．（15 分） $X_1, X_2, \cdots, X_m$ 是来自二项分布总体 $X \sim b(n,p)$ 的样本

（1）当 n 已知时，用矩估计法求参数 p^2 的无偏估计量；

（2）当 n 已知时，用极大似然估计法求参数 p 的估计量，并判断无偏性，说明理由；
（3）当 n, p 均未知时，用矩估计法求参数 p 的估计量，并判断是否无偏，说明理由。

五． 解：（1） $X \sim b(n,p)$， 有 $E(X)=np$， $Var(X)=np(1-p)$

对样本均值和样本方差有 $E(\bar{X})=np$， $E(S^2)=np(1-p)$

因为 $E(\bar{X}-S^2)=np^2$， 所以 $\dfrac{\bar{X}-S^2}{n}$ 是参数 p^2 的无偏估计量。

解法二： $E(X)=np$， $E(X^2)=np-np^2+n^2p^2=np+n(n-1)p^2$

$E(X^2-X)=n(n-1)p^2$， 设 $A_2=\dfrac{X_1^2+X_2^2+\cdots+X_n^2}{n}$ 为二阶样本原点矩

可得参数 p^2 的无偏估计量 $\dfrac{A_2-\bar{X}}{n(n-1)}$。

（2）似然函数 $p(x_1,x_2,\cdots,x_m;p)=\prod_{k=1}^{m}\begin{pmatrix} m \\ x_k \end{pmatrix} p^{x_k}(1-p)^{n-x_k}$

对数似然函数 $\ln p(x_1,x_2,\cdots,x_m;p)=\sum_{k=1}^{m}\left\{\ln\begin{pmatrix} n \\ x_k \end{pmatrix}+x_k\ln p+(n-x_k)\ln(1-p)\right\}$

$$\frac{d\ln p(x_1,x_2,\cdots,x_m;p)}{dp}=\sum_{k=1}^{m}\left\{\frac{x_k}{p}-\frac{(n-x_k)}{1-p}\right\}=\sum_{k=1}^{m}\frac{x_k}{p}-\sum_{k=1}^{m}\frac{(n-x_k)}{1-p}=0$$

$\dfrac{m\bar{x}}{p}=\dfrac{mn-m\bar{x}}{1-p}$ 解得 $p=\dfrac{\bar{x}}{n}$， 所以参数 p 的极大似然估计量 $\hat{p}=\dfrac{\bar{X}}{n}$。

（3） $E(X)=np,\qquad Var(X)=np(1-p)$

$$\begin{cases} \bar{X}=np \\ S^2=np(1-p) \end{cases} \Rightarrow \hat{p}=1-\frac{S^2}{\bar{X}},\ \hat{n}=\frac{\bar{X}^2}{\bar{X}-S^2},\quad \text{其中 } \bar{X}=\frac{X_1+X_2+\cdots+X_m}{m},\ S^2=\frac{\sum_{k=1}^{m}X_k}{m-1}$$

一般情况下 $\hat{p}=1-\dfrac{S^2}{\bar{X}}$ 不是参数 p 的无偏估计

$E(\hat{p}\cdot\bar{X})=E(\bar{X}-S^2)=np^2$，所以只有当 $E(\hat{p}\cdot\bar{X})=E(\hat{p})\cdot E(\bar{X})$ 时， $E(\hat{p})=p$，

也就是只有当 $\dfrac{S^2}{\bar{X}}$ 与 $\bar{X}$ 不相关时， $\hat{p}=1-\dfrac{S^2}{\bar{X}}$ 是参数 p 的无偏估计。

八．(10 分)设某工厂生产一种产品，其质量指标的参数服从正态分布 $N\left(\mu,3^2\right)$， $\mu\le 10$ 为优级，$X_1,X_2,\cdots,X_{36}$ 为来自该正态总体的样本。做假设检验 H_0： $\mu\le 10$ VS H_1： $\mu>10$。

（1）给出显著性水平为 $\alpha=0.05$ 的拒绝域范围，以及 $\mu=11$ 时的错误类型和犯错概率；

（2）若希望 $\mu=11$ 时犯错误的概率低于 0.01，至少需要多大的样本容量。

六． 解：（1） $\bar{X}\sim N\left(\mu,\frac{9}{n}\right)$， 当 $\mu=10$，$n=36$ 时， $\bar{X}\sim N\left(10,\left(\frac{1}{2}\right)^2\right)$

$2\left(\bar{X}-10\right)\sim N\left(0,1\right)$，$P\left(2\left(\bar{X}-10\right)>u_{0.95}\right)=0.05$，

拒绝域为 $\left\{\bar{X}\middle|\bar{X}>10+\frac{u_{0.95}}{2}\right\}$，$u_{0.95}=1.65$，所以拒绝域为 $\left\{\bar{X}\middle|\bar{X}>10.825\right\}$。

$\mu=11$ 时为备择假设成立，所以所犯错误未第二类错误， 此时 $\bar{X}\sim N\left(11,\left(\frac{1}{2}\right)^2\right)$

第二类错误的概率 $P=P\left(\bar{X}\le 10.82\middle|\mu=11\right)=P\left(2\left(\bar{X}-11\right)\le 2\left(10.82-11\right)\right)=\Phi\left(-0.36\right)$

（2）样本容量为 n 时，原假设成立，即 $\mu=10$ 时，0.05 显著性水平的拒绝域为

满足 $\frac{\sqrt{n}\left(\bar{X}-10\right)}{3}>u_{0.95}$，即 $\left\{\bar{X}\middle|\bar{X}>10+\frac{3u_{0.95}}{\sqrt{n}}\right\}$

$\mu=11$ 时， $\bar{X}\sim N\left(11,\frac{9}{n}\right)$， $\frac{\sqrt{n}\left(\bar{X}-11\right)}{3}\sim N\left(0,1\right)$

第二类错误的概率 $P=P\left(\bar{X}\le 10+\frac{3u_{0.95}}{\sqrt{n}}\middle|\mu=11\right)=P\left(\frac{\sqrt{n}\left(\bar{X}-11\right)}{3}\le\frac{\sqrt{n}\left(10+\frac{3u_{0.95}}{\sqrt{n}}-11\right)}{3}\right)=0.01$

$\frac{\sqrt{n}\left(10+\frac{3u_{0.95}}{\sqrt{n}}-11\right)}{3}=u_{0.95}-\frac{\sqrt{n}}{3}<u_{0.01}$， $\sqrt{n}>3\left(u_{0.95}-u_{0.01}\right)=12.2$， $n>9\left(u_{0.95}+u_{0.99}\right)^2$，或 140 左右

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

备注 2. 参数为 p 的几何分布的方差为 $\frac{1-p}{p^2}$，泊松分布 $X\sim P\left(\lambda\right)$ 的分布列 $P\left(X=k\right)=e^{-\lambda}\lambda^k/k!$

备注 3. 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）

$\Phi(1.28)=0.9,\quad \Phi(1.44)=0.925,\quad \Phi(1.65)=0.95,\quad \Phi(1.96)=0.975,$

$\Phi(1)=0.84,\quad \Phi(1.25)=0.89,\quad \Phi(1.5)=0.93,\quad \Phi(1.75)=0.96,\quad \Phi(2)=0.98,\quad \Phi(3)=0.999$

$P\left(t\left(2\right)>2.92\right)=0.05, P\left(t\left(3\right)>2.36\right)=0.05, P\left(t\left(4\right)>2.13\right)=0.05, P\left(t\left(5\right)>2.02\right)=0.05$

$P\left(t\left(2\right)>1.89\right)=0.1, P\left(t\left(3\right)>1.64\right)=0.1, P\left(t\left(4\right)>1.53\right)=0.1, P\left(t\left(5\right)>1.48\right)=0.1$