一、简单题（共 40 分，共 4 题，每题 10 分）

1. 做一系列独立试验，假设每次试验成功的概率为 p（0<p<1），请给出（1）首次成功时，试验次数 X 的分布律；（2）n 次成功之前已经失败次数 Y 的分布律；（3）在 n 次试验中成功次数 Z 的分布律。

答：X 的分布律为：$P\{X=k\}=(1-p)^{k-1}p,\ k=1,2,\cdots$ （3 分）

Y 的分布律为：$P\{Y=k\}=C_{n+k-1}^{k}(1-p)^{k}p^{n},\ k=0,1,2,\cdots$ （3 分）

Z 的分布律为：$P\{X=k\}=C_{n}^{k}(1-p)^{n-k}p^{k},\ k=0,1,2,\cdots,n$ （4 分）

2. 假设总体服从正态分布 $N(\mu,\sigma^2)$，抽样后构造统计量对 μ 进行估计，下面哪些估计量是无偏，哪个估计量最有效？为什么？

$(1)\ \overline{X}\quad (2)\ X_n\quad (3)\ \frac{1}{2}(X_1+X_2)\quad (4)\ X_1+X_2$

答：因为 $E(\overline{X})=\mu,\ E(X_n)=\mu,\quad E\left(\frac{1}{2}(X_1+X_2)\right)=\mu,\quad E(X_1+X_2)=2\mu$ （3 分）

所以前三个统计量是无偏估计量 （1 分）

又因为 $D(\overline{X})=\frac{\sigma^2}{n},\ D(X_n)=\sigma^2,\quad D\left(\frac{1}{2}(X_1+X_2)\right)=\frac{\sigma^2}{2}$ （3 分）

所以当 n=2 时，（1）和（3）等同，是最有效的估计量 （2 分）

当 n>2 时，（1）是最有效的估计量 （1 分）

3. 设二维正态随机变量 $(X,Y)\sim N\left(1,3^2;0,4^2;-\frac{1}{2}\right)$，设 $Z=2X+bY$，b 为实数，问 X 与 Z 在什么条件下相关？什么条件下不相关？

答：$E(X)=1,\ E(Z)=2$

$$\begin{aligned}E(X\ Z)&=E(X(2X+bY))=2E(X^2)+bE(XY)\\&=2\left[E(X)^2+D(X)\right]+b\left[\rho_{XY}\sqrt{D(X)D(Y)}+E(X)E(Y)\right]\\&=2(1+9)+b[(-0.5)\times 3\times 4+1\times 0]=20-6b\end{aligned}$$ （6 分）

(注：用到相关系数定义)

$$\operatorname{cov}(X,Z)=E(XZ)-E(X)E(Z)=18-6b$$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Z)}{\sqrt{D(X)D(Z)}}$$

则当 b=3 时，$\rho_{XY}=0$，即 X 与 Z 不相关；否则相关 （4 分）

4. 设 $X_1,X_2,\cdots,X_{n+m}$ 是来自总体 $X\sim N(0,\sigma^2)$ 的容量为 n+m 的简单随机样本，问 Y_1 及

Y_2服从什么分布？为什么？

$$Y_1=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}},\quad Y_2=\frac{m\sum_{i=1}^{n}X_i^2}{n\sum_{i=n+1}^{n+m}X_i^2}$$

答：由于 $X_i\sim N\left(0,\sigma^2\right)$，则 $\sum_{i=1}^{n}X_i\sim N\left(0,n\sigma^2\right)$，从而 $\frac{\sum_{i=1}^{n}X_i}{\sqrt{n}\,\sigma}\sim N(0,1)$

又由于 $\frac{\sum_{i=n+1}^{n+m}X_i^2}{\sigma^2}=\sum_{i=n+1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2\sim\chi^2(m)$，故 $Y_1=\frac{\sqrt{m}\sum_{i=1}^{n}X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m}X_i^2}}\sim t(m)$　　（5 分）

由于 $\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2\sim\chi^2(n)$，$\sum_{i=n+1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2\sim\chi^2(m)$

故 $Y_2=\frac{m\sum_{i=1}^{n}X_i^2}{n\sum_{i=n+1}^{n+m}X_i^2}=\frac{\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2\Big/n}{\sum_{i=n+1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2\Big/m}\sim F(n,m)$　　（5 分）

二、（共 15 分）

甲箱中有 5 个正品和 3 个次品，乙箱中有 4 个正品和 3 个次品。现从甲箱中取两个产品放入乙箱，再从乙箱中任取一个产品，求：（1）从乙箱中取出为正品的概率；（2）若从乙箱中取得的为次品，求原来从甲箱中取出的都是正品的概率。

解：设 Ai 表示“甲箱中取出的两件产品中有 i 件正品”，i=0,1,2；　B 表示“从乙箱中取出的为正品”，则

（1）$P(B)=\sum_{i=0}^{2}P(A_i)P(B\mid A_i)=\frac{C_3^2}{C_8^2}\times\frac{C_4^1}{C_9^1}+\frac{C_5^1\times C_3^1}{C_8^2}\times\frac{C_5^1}{C_9^1}+\frac{C_5^2}{C_8^2}\times\frac{C_6^1}{C_9^1}=\frac{7}{12}\approx 0.5833$

（7 分）

（2）$P\left(A_2\mid\overline{B}\right)=\frac{P(A_2)P\left(\overline{B}\mid A_2\right)}{P\left(\overline{B}\right)}=\frac{\frac{C_5^2}{C_8^2}\times\frac{C_3^1}{C_9^1}}{1-P(B)}=\frac{2}{7}\approx 0.2857$　　（8 分）

三、（共 15 分）

学生完成一道作业的时间 X 是一个随机变量，单位为小时，它的密度

$$f(x)=\begin{cases}c\,x^2+x, & 0\le x\le 0.5\\ 0, & \text{其他}\end{cases}$$

（1）确定常数 c；（2）写出 X 的分布函数；（3）试求学生在 10min 以上 20min 以内完成一道作业的概率。

解：（1）由密度函数的归一性知

$$1=\int_{-\infty}^{\infty} f(x)dx=\int_0^{0.5}\left(c\,x^2+x\right)dx=\frac{c}{24}+\frac{1}{8} \qquad \text{（4 分）}$$

求解得到 c=21

（2）$F(x)=P\{X\le x\}=\int_{-\infty}^{x} f(u)du$

当 $x<0$ 时，$F(x)=\int_{-\infty}^{x} 0du=0$

当 $x\ge 0.5$ 时，$F(x)=1$

当 $0\le x<0.5$ 时，$F(x)=\int_0^x\left(21u^2+u\right)du=7x^3+\frac{1}{2}x^2$

X 的分布函数为

$$F(x)=\begin{cases}0, & x<0\\ 7x^3+\frac{1}{2}x^2 & 0\le x<0.5\\ 1, & x\ge 0.5\end{cases} \qquad \text{（6 分）}$$

（3）学生在 10min 以上 20min 以内完成一道作业的概率为

$$P\left\{\frac{1}{6}\le X\le\frac{1}{3}\right\}=F\left(\frac{1}{3}\right)-F\left(\frac{1}{6}\right)=\frac{17}{54}-\frac{5}{108}=\frac{29}{108} \qquad \text{（5 分）}$$

四、(共 10 分)

已知总体 X 服从几何分布 $P\{X=x\}=(1-p)^{x-1}p,\ x=1,2,\cdots$，其中 p 为未知参数，$0<p<1$ 。设 $X_1,X_2,\cdots,X_n$ 为总体的一组样本，求参数 p 的极大似然估计。

解： 设 $x_1,x_2,\cdots,x_n$ 是总体的一组样本观测值，则似然函数为

$$L(x_1,x_2,\cdots,x_n;p)=\prod_{i=1}^{n}p(1-p)^{x_i-1}=p^n(1-p)^{\sum_{i=1}^{n}x_i-n} \qquad \text{（3 分）}$$

$$\ln L(x_1,x_2,\cdots,x_n;p)=n\ln p+\left(\sum_{i=1}^{n}x_i-n\right)\ln(1-p) \qquad \text{（3 分）}$$

$$\frac{d\ln L}{dp}=\frac{n}{p}-\frac{1}{1-p}\left(\sum_{i=1}^{n}x_i-n\right)=0 \qquad \text{（2 分）}$$

解得极大似然估计值为 $\hat{p}=\frac{1}{\bar{x}}$，极大似然估计量为 $\hat{p}=\frac{1}{\bar{X}}$ （2 分）

五、(共 10 分)

已知某电子元件的长度服从正态分布，且方差为 0.01。从一批次的产品中任取 10 个，测得 $s^2=0.012$，则在显著性水平 $\alpha=0.05$ 下，这批次电子元件的精度（即标准差）是否正常？（$\chi^2_{0.975}(10)=3.247,\ \chi^2_{0.025}(10)=20.483,\ \chi^2_{0.975}(9)=2.700,\ \chi^2_{0.025}(9)=19.023$）

解：设电子元件的长度$X\sim N\left(\mu,\sigma^2\right)$，由题意检验

$$H_0:\sigma^2=0.01,\quad H_1:\sigma^2\neq 0.01 \qquad （2 分）$$

由于总体期望未知，在原假设成立条件下，检验统计量

$$\chi^2=(n-1)\frac{S^2}{\sigma^2}\sim\chi^2(n-1) \qquad （3 分）$$

原假设 H0 的拒绝域为$(n-1)\frac{S^2}{\sigma^2}>\chi^2_{\alpha/2}(n-1)$或$(n-1)\frac{S^2}{\sigma^2}<\chi^2_{1-\alpha/2}(n-1)$ （2 分）

因为$(n-1)\frac{s^2}{\sigma^2}=9\times\frac{0.012}{0.01}=10.8$，

$$\chi^2_{\alpha/2}(n-1)=\chi^2_{0.975}(9)=2.700,\qquad \chi^2_{1-\alpha/2}(n-1)=\chi^2_{0.025}(9)=19.023$$

$$2.700\le 10.8\le 19.023$$

因此接受原假设 H0，即在显著性水平$\alpha=0.05$下，这批次电子元件的精度正常。

（3 分）

六、（共 10 分）

在考察硝酸钠的可溶程度时，对一系列不同温度观察它在 100 毫升的水中溶解的硝酸钠的重量，得到一组观测值$(x_i,y_i),i=1,2,\cdots,9$，计算得

$$\sum_{i=1}^{9}x_i=234,\ \sum_{i=1}^{9}y_i=811.3,\ \sum_{i=1}^{9}x_i^2=10144,\ \sum_{i=1}^{9}y_i^2=76218.1339,\ \sum_{i=1}^{9}x_iy_i=24628.6$$

从理论上推测，温度x_i与溶解的硝酸钠的重量Y_i之间有关系式：

$$Y_i=a+bx_i+\varepsilon_i,i=1,2,\cdots,9$$

式中$\varepsilon_i,i=1,2,\cdots,9$相互独立，均服从正态分布$N\left(0,\sigma^2\right)$

（1）求未知参数a,b的最小二乘估计值和σ^2的无偏估计值；（2）检验线性回归是否显著（$\alpha=0.01$）？（结果保留四位有效数字）

α \ 自由度	6	7
0.01	0.834	0.798
0.005	0.870	0.836

解：$l_{xx}=\sum_{i=1}^{9}(x_i-\bar{x})^2=\sum_{i=1}^{9}x_i^2-n\,\bar{x}^2=4060$

$$l_{xy}=\sum_{i=1}^{9}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{9}x_iy_i-n\,\bar{x}\,\bar{y}=3534.8$$

$$l_{yy}=\sum_{i=1}^{9}\left(y_i-\bar{y}\right)^2=\sum_{i=1}^{9}y_i^2-n\,\bar{y}^2=3083.9461$$ （3 分）

（1） $\hat{b}=\dfrac{l_{xy}}{l_{xx}}\approx 0.8706,\quad \hat{a}=\bar{y}-\hat{b}\,\bar{x}=67.5088,\quad \hat{\sigma}^2=\dfrac{1}{n-2}\left(l_{yy}-\hat{b}^2\,l_{xx}\right)=0.9560$

（4 分）

（2） $R=\dfrac{l_{xy}}{\sqrt{l_{xx}l_{yy}}}\approx 0.9990>R_{0.01}(7)=0.798$，线性关系显著 （3 分）

【注意题目如果要求写出回归方程则必须写出方程式;这个题目只要求计算参数,可以不写】