北京师范大学 432 应用统计专业课模拟卷（2）

一、选择题，共 10 题，每题 3 分。

1. 设 A,B,C 都是事件，又 A 和 B 独立，B 和 C 独立，A 和 C 互不相容。$P(A)=1/2$,$P(B)=1/4$,$P(C)=1/8$ 。则概率 $P(A\cup B\cup C)$ 为（　　）。

A. 23/33　　B. 23/32

C. 11/16　　D. 2/3

【答案】B

【解析】由题意， $P(AB)=P(A)P(B)$,$P(BC)=P(B)P(C)$,$P(AC)=0$ ，所以 $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)=P(A)+P(B)+P(C)-P(A)P(B)-P(B)P(C)=23/32$ ，所以选择 B 项。

2. 参数估计中，其他条件不变的前提下，置信区间越宽则置信度（　　）。

A. 越低　　B. 越高

C. 不变　　D. 不确定

【答案】B

【解析】置信区间是指在某一置信度时，总体参数所在的区域距离或区域长度。置信度又称置信水平，是指总体参数值落在样本统计值某一区间内的概率。由此可见，在其他条件不变的前提下，置信区间越宽，置信水平越高，则置信度越高。

3. 最不受极端值影响的描述性统计量是（　　）。

A. 众数　　B. 方差

C. 标准差　　D. 均值

【答案】A

【解析】众数是一组数据中出现次数最多的变量值，用 M_0 表示。众数是一个位置代表值，不受极端值的影响，抗干扰性强。而方差、标准差、均值均利用了全部数据的信息，因此容易受到极端值的影响。

4. 设 X 和 Y 都服从标准正态分布，则（　　）。

A. $X+Y$ 服从正态分布　　B. X^2+Y^2 服从卡方分布

C. X^2 和 Y^2 都服从卡方分布　　D. X^2/Y^2 服从 F 分布

【答案】C

【解析】 A 项，正态分布具有可加性的前提是随机变量 X,Y 相互独立，题目未说明 X,Y 相互独立，所以 $X+Y$ 不一定服从正态分布；B 项，卡方分布要求 X,Y 相互独立且同分布于标准正态分布，题目未说明 X,Y 相互独立；D 项，F 分布要求 X^2,Y^2 是相互独立的卡方分布，题目未说明 X,Y 相互独立。

5. 线性回归预测过程中，在自变量取值和置信度为一定的条件下，总是（　　）。

A. 个别值的估计区间小于平均值的估计区间

B. 个别值的估计区间大于平均值的估计区间

C. 个别值的估计区间等于平均值的估计区间

D. 个别值的估计区间与平均值的估计区间没有关系

【答案】B

【解析】利用估计的回归方程，对于 x 的一个特定值 x_0 ，求出 y 的一个估计值的区间为估计区间。区间估计有两种类型：一是置信区间的估计，它是对 x 的一个给定值 x_0 ，求出 y 的平均值的估计区间，这一区间称为平均值置信区间；二是预测区间估计，它是对 x 的一个给定值 x_0 ，求出 y 的一个个别值的估计区间，这一区间称为个别值预测区间。由置信区间和预测区间的计算公式可以得出，个别值的预测区间要比平均值的置信区间宽一些。

显著性水平 α 下 $E\left(y_f\right)$ 的置信区间为：

$$\left(\hat{y}_f - t_{\alpha/2} \times \sqrt{\left(\frac{1}{n} + \frac{x_f - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\hat{\sigma}_n^2}, \hat{y}_f + t_{\alpha/2} \times \sqrt{\left(\frac{1}{n} + \frac{x_f - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\hat{\sigma}_n^2}\right)$$

显著性水平 α 下 y_f 的置信区间为：

$$\left(\hat{y}_f - t_{\alpha/2} \times \sqrt{\left(1 + \frac{1}{n} + \frac{x_f - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\hat{\sigma}_n^2}, \hat{y}_f + t_{\alpha/2} \times \sqrt{\left(1 + \frac{1}{n} + \frac{x_f - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\hat{\sigma}_n^2}\right)$$

6. 假设一个袋子中有黑色、白色和红色三种颜色的球，它们的比例为 3:4:3，现每次有放回地从袋子随机摸出一个球，记下被摸出球的颜色，如此反复，则白球比黑球先被摸出来的概率为(　　)。

A. 3/7　　　　B. 4/7

C. 4/10　　　　D. 3/10

【答案】B

【解析】与每次取到的球是红色无关，所以此问题等价于袋中有黑色球：白色球=3:4，求第一次摸球摸到白色球的概率。

7. 假设 $X_1, \cdots, X_n$ 为从正态总体 $N(\mu, 1)$ 中抽取的简单随机样本，其中 μ 未知，如果要求 μ 的 95%置信区间的长度不超过 0.6，则样本量 n 至少需要等于（　　）。

A. 42　　　　B. 43

C. 43　　　　D. 45

【答案】B

【解析】因为总体方差 $\sigma^2 = 1$ 已知，故采用 z 统计量。要求 μ 的 95%置信区间的长度不超过 0.6，即 $2z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq 0.6$ ，解得 $n \geq 42.68$ ，所以样本量 n 至少需要等于 43。

8. 设随机变量 X 和 Y 独立同分布，其分布为正态分布 $N(\mu, \sigma^2)$ ，则 $(X + Y - 2\mu)/(X - Y)^2$ 的分布为（　　）。

A. 自由度为（1,1）的 F 分布　　　　B. 自由度为（1,2）的 F 分布

C. 自由度为（2,1）的 F 分布　　　　D. 自由度为（2,2）的 F 分布

【答案】A
【解析】随机变量 X 和 Y 独立同分布， $X \sim N(\mu,\sigma^2)$, $Y \sim N(\mu,\sigma^2)$ ，则有

$$X+Y-2\mu \sim N(0,2\sigma^2)\,,X-Y \sim N(0,2\sigma^2)$$

即：

$$\left(\frac{X+Y-2\mu}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)\,,\left(\frac{X-Y}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(1)$$

因此，

$$(X+Y-2\mu)^2/(X-Y)^2 \sim F(1,1)$$

9. 将一枚骰子先后投掷 2 次，事件 A 表示两次出现的点数之和为 8，事件 B 表示第 1 次出现点数大于第 2 次出现的点数，则下列命题中，正确的是（　　）。

① $P(A) > P(A|B)$　　② $P(A) < P(A|B)$

③ $P(B) > P(B|A)$　　④ $P(B) < P(B|A)$

A. ①③　　B. ①④　　C. ②③　　D. ②④

【答案】A
【解析】令 (x,y) 表示先后投掷出的点数，则 $A=\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$, $AB=\{(5,3),(6,2)\}$ ，而事件 B 包含的基本事件个数等于从 6 个数中不计次序地选 2 个数的选法总数，为 C_6^2 。分别计算 $P(A)$, $P(B)$, $P(AB)$ ：

$$P(AB)=\frac{2}{6\times 6}=\frac{1}{18}\,,P(A)=\frac{5}{6\times 6}=\frac{5}{36}\,,P(B)=\frac{C_6^2}{6\times 6}=\frac{5}{12}$$

于是，

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{2}{15}\,,P(B|A)=\frac{P(AB)}{P(A)}=\frac{2}{5}$$

因此， $P(A) > P(A|B)$, $P(B) > P(B|A)$ ，命题①③正确，应选 A。

10. 设总体 X 服从正态分布 $N(\mu,\sigma^2)(\sigma>0)$, X_1 , X_2 , $\cdots$, $X_{2n}(n\ge 2)$ 为来自该总体的简单随机样本，其样本均值为 $\bar{X}=\frac{1}{2n}\sum_{i=1}^{2n}X_i$ 。记统计量 $Y_1=\sum_{i=1}^{2n}(X_i-\bar{X})^2$, $Y_2=\sum_{i=1}^{n}(X_i-X_{n+i})^2$, $Y_3=\sum_{i=1}^{n}(X_i+X_{n+i}-2\bar{X})^2$ ，则这 3 个统计量的数学期望 $E(Y_1)$, $E(Y_2)$, $E(Y_3)$ 的大小关系为（　　）。

A. $E(Y_1)>E(Y_2)>E(Y_3)$　　B. $E(Y_1)>E(Y_3)>E(Y_2)$

C. $E(Y_3)>E(Y_1)>E(Y_2)$　　D. $E(Y_2)>E(Y_1)>E(Y_3)$

【答案】D
【解析】对于 Y_1 ，由于 $X_1,X_2,\cdots,X_{2n}(n\ge 2)$ 是来自正态总体的简单随机样本，其样本方差 $s^2=\frac{1}{2n-1}\sum_{i=1}^{2n}(X_i-\bar{X})^2$ 满足 $\frac{(2n-1)S^2}{\sigma^2}\sim\chi^2(2n-1)$ ，于是：

$$E\left[\frac{(2n-1)S^2}{\sigma^2}\right]=E\left(\frac{Y_1}{\sigma^2}\right)=2n-1\,,E(Y_1)=(2n-1)\sigma^2$$

对于 Y_2 ：

$$E(X_i-X_{n+i})=\mu-\mu=0\,,D(X_i-X_{n+i})=\sigma^2+\sigma^2=2\sigma^2$$

于是，$X_i - X_{n+i} \sim N(0,2\sigma^2)$, $\frac{X_i - X_{n+i}}{\sqrt{2}\sigma} \sim N(0,1)$ ，且 $X_1 - X_{n+1}, X_2 - X_{n+2}, \cdots, X_n - X_{2n}$ 相互独立，

因此 $\sum_{i=1}^{n}\left(\frac{X_i - X_{n+i}}{\sqrt{2}\sigma}\right)^2 \sim \chi^2(n)$ ，

$$E\left[\sum_{i=1}^{n}\left(\frac{X_i - X_{n+i}}{\sqrt{2}\sigma}\right)^2\right] = n \, , E(Y_2) = E\left[\sum_{i=1}^{n}(X_i - X_{n+i})^2\right] = 2n\sigma^2$$

对于 Y_3 ，由于 $X_1, X_2, \cdots, X_{2n} (n \geq 2)$ 是来自正态总体的简单随机样本，故 X_i 与 $X_{n+i} (i = 1,2,\cdots,n)$ 相互独立，

$$E(X_i + X_{n+i}) = \mu + \mu = 2\mu \, , D(X_i + X_{n+i}) = \sigma^2 + \sigma^2 = 2\sigma^2$$

于是，$X_1 + X_{n+1}, X_2 + X_{n+2}, \cdots, X_n + X_{2n}$ 可看作来自正态总体 $N(2\mu, 2\sigma^2)$ 的简单随机样本，其样本均值为 $\frac{1}{n}\sum_{i=1}^{n}(X_i + X_{n+i}) = 2\bar{X}$ ，样本方差为 $\frac{1}{n-1}\sum_{i=1}^{n}(X_i + X_{n+i} - 2\bar{X})^2 = \frac{1}{n-1}Y_3$.

因此，$\frac{Y_3}{2\sigma^2} \sim \chi^2(n-1)$, $E\left(\frac{Y_3}{2\sigma^2}\right) = n - 1$, $E(Y_3) = 2(n-1)\sigma^2$.

综上所述，$E(Y_2) > E(Y_1) > E(Y_3)$ ，应选 D。

二、解答题，共 7 题。

1.（15 分）甲乙两个班级统计学考试成绩资料如下：

甲班的平均分数为 75 分，标准差为 7 分；乙班的考试成绩频数分布表如下：

乙班考试成绩频数分布表

分数（分）	人数（人）
60 以下	4
60～70	8
70～80	13
80～90	7
90 以上	5

要求：（1）计算乙班的平均考试分数。

（2）计算乙班考试分数的方差及标准差。

（3）计算乙班考试分数的离散系数。

（4）比较甲乙两个班级考试分数的离散程度的大小。

【解】

（1）乙班平均考试分数计算过程如下表所示：

乙班平均考试分数的分组数据表

分数（分）	组中值M_1	人数f_i（人）	$M_i f_i$
60 以下	55	4	220
60～70	65	8	520
70～80	75	13	975
80～90	85	7	595

90 以上	95	5	475
合计	——	37	2785

由上表中数据可得：

$$\bar{x}_{乙} = \frac{\sum_{i=1}^{n} M_i f_i}{n} = \frac{2785}{37} = 75.27\text{（分）}$$

（2）方差计算过程如下表所示：

分数（分）	组中值M_1	人数f_i（人）	$(M_i-\bar{x})^2$	$(M_i-\bar{x})^2 f_i$
60 以下	55	4	410.873	1643.492
60～70	65	8	105.473	843.783
70～80	75	13	0.073	0.948
80～90	85	7	94.673	662.710
90 以上	95	5	389.273	1946.365
合计	——	37	1000.365	5097.298

由上表中的数据可得：

$$S_{乙}^2 = \frac{\sum_{i=1}^{n}(M_i-\bar{x})^2 f_i}{n} = \frac{5097.298}{37} = 137.76$$

$$s_{乙} = \sqrt{S_{乙}^2} = \sqrt{137.76} = 11.74\text{（分）}$$

（3）$v_{s_{乙}} = \frac{s_{乙}}{\bar{x}_{乙}} = \frac{11.74}{75.27} = 0.156$

（4）$v_{s_{甲}} = \frac{s_{甲}}{\bar{x}_{甲}} = \frac{7}{75} = 0.093$

$v_{s_{甲}} < v_{s_{乙}}$，说明两个班的统计学考试成绩相比较，甲班的成绩较集中，乙班的成绩较分散。

2.（15 分）设有 N 个袋子，每个袋子中装有 a 只黑球，b 只白球，从第一袋中取出一球放入第二袋中，然后从第二袋中取出一球放入第三袋，如此下去，问从最后一个袋中取出一球为黑球的概率是多少？

【解】设 A_1 ={从第 1 袋中取出一球是黑球}，则

$$P(A_1) = \frac{a}{a+b}, P(\bar{A}_1) = \frac{b}{a+b}$$

设 A_2 ={从第 1 袋中取出一球放入第 2 袋中，再从第二袋中取出一球是黑球}，则

$$P(A_2) = \frac{a}{a+b}\cdot\frac{a+1}{a+b+1} + \frac{b}{a+b}\cdot\frac{a}{a+b+1} = \frac{a}{a+b}, P(\bar{A}_2) = \frac{b}{a+b}$$

一般设 A_k ={按上述规则取球放球，并在第 k 袋中取出一球是黑球}，且 $P(A_k) = \frac{a}{a+b}$，则

$$P(A_{k+1}) = P(A_k)P(A_{k+1}|A_k) + P(\bar{A}_k)P(A_{k+1}|\bar{A}_k) = \frac{a}{a+b}$$

由数学归纳法得，$P(A_N) = \frac{a}{a+b}$.

3.（10 分）飞机坠落在 A,B,C 三个区域之一，营救部门判断其概率分别为 0.7,0.2,0.1；用直升机搜索这些区域，若有残骸，被发现的概率分别为 0.3,0.4,0.5，若已用直升机搜索过 A 区域及 B 区域，没有发现残骸，在这种情况下，试计算飞机坠落在 C 区域的概率。

【解】设 A ={飞机坠落在 A 区域}， B ={飞机坠落在 B 区域}， C ={飞机坠落在 C 区域}， S ={已搜索过 A,B 区域，没有发现残骸}，则

$$\begin{aligned}P(S) &= P(A)P(S|A) + P(B)P(S|B) + P(C)P(S|C)\\ &= 0.7\times 0.7 + 0.2\times 0.6 + 0.1 = 0.71\end{aligned}$$

所求概率为

$$P(C|S) = \frac{P(C)P(S|C)}{P(S)} = \frac{0.1}{0.71} = 0.14$$

4.（20 分）若 ξ,η 服从二元正态分布， $E\xi = a, D\xi = 1, E\eta = b, D\eta = 1$，证明： ξ 与 η 的相关系数 $r = \cos q\pi$，其中 $q = P\{(\xi-a)(\eta-b) < 0\}$.

【解】

由题设得

$$q = \frac{1}{2\pi\sqrt{1-r^2}}\iint_{(x-a)(y-b)<0} \mathrm{e}^{-\frac{1}{2(1-r^2)}[(x-a)^2-2r(x-a)(y-b)+(y-b)^2]}\,dxdy$$

令 $x = a + \rho\cos\theta$, $y = b + \rho\sin\theta$, $|J| = \rho$,

$$q = \frac{1}{\pi\sqrt{1-r^2}}\int_{\frac{\pi}{2}}^{\pi} d\theta \int_0^{+\infty} \rho\,\mathrm{e}^{-\frac{\rho^2}{2(1-r^2)}(1-2r\sin\theta\cos\theta)}\,d\rho$$

$$= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi} \frac{d\theta}{1-2r\,sin\,\theta\,cos\,\theta}$$

$$= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi} \frac{d\theta}{\cos^2\theta\,(\sec^2\theta - 2r\tan\theta)}$$

$$= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi} \frac{d\tan\theta}{1+\tan^2\theta - 2r\tan\theta}$$

$$= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi} \frac{d\,(\tan\theta - r)}{1-r^2+(\tan\theta-r)^2}$$

$$= \frac{\sqrt{1-r^2}}{\pi}\left(\frac{1}{\sqrt{1-r^2}}\arctan\frac{-r}{\sqrt{1-r^2}} + \frac{\pi}{2\sqrt{1-r^2}}\right)$$

$$= \frac{1}{\pi}\arctan\frac{-r}{\sqrt{1-r^2}} + \frac{1}{2}$$

于是得

$$\tan\left(q\pi-\frac{\pi}{2}\right)=-\frac{r}{\sqrt{1-r^2}}$$

$$\cot q\pi=\frac{r}{\sqrt{1-r^2}}$$

由此推得 $r=\cos q\pi$.

5.（20 分）已知随机变量 X 的概率密度为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, 0\le x\le\theta,\\ 0, \quad 其他,\end{cases}$$

其中 θ 为大于 0 的未知参数。 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个简单随机样本。

（1）求 θ 的最大似然估计量 $\hat{\theta}$;

（2）求 $E(\hat{\theta})$;

（3）求 α ，使得 $E(\alpha\hat{\theta}-\theta)^2$ 最小。

【解】

（1）设 $x_1,x_2,\cdots,x_n$ 是相应于样本 $X_1,X_2,\cdots,X_n$ 的一组样本值，则似然函数为

$$L(\theta)=\prod_{i=1}^{n}f(x_i;\theta)=\begin{cases}\frac{1}{\theta^n}, & 0\le x_1,x_2,\cdots,x_n\le\theta\,,\\ 0\ , & 其他.\end{cases}=\begin{cases}\frac{1}{\theta^n}, & \theta\ge\max\limits_{1\le i\le n}x_i\,,\\ 0\ , & 其他.\end{cases}$$

又因为 $\frac{1}{\theta^n}$ 是关于 θ 的单调减少函数，故当 $\theta=\max\limits_{1\le i\le n}x_i$ 时，$L(\theta)$ 取值最大。

因此， θ 的最大似然估计量为 $\hat{\theta}=\max\limits_{1\le i\le n}X_i$.

（2）由于 X 服从 $[0,\theta]$ 上的均匀分布，故 X 的分布函数为

$$F_X(x)=\begin{cases}0\,, & x<0\,,\\ \frac{x}{\theta}, & 0\le x<\theta\,,\\ 1\,, & x\ge\theta\,.\end{cases}$$

设随机变量 $Y=\max\limits_{1\le i\le n}X_i$ ，计算 Y 的分布函数 $F_Y(y)$。

当 $y<0$ 时， $F_Y(y)=0$ 。

当 $0\le y<\theta$ 时，

$$F_Y(y)=P\{Y\le y\}=P\left\{\max_{1\le i\le n}X_i\le y\right\}=P\{X_1\le y,X_2\le y,\cdots,X_n\le y\}$$

$$\xRightarrow{独立性}P\{X_1\le y\}P\{X_2\le y\}\cdots P\{X_n\le y\}=F_X^n(y)=\left(\frac{y}{\theta}\right)^n.$$

当 $y\ge\theta$ 时， $F_Y(y)=1.$

于是，

$$F_Y(y)=\begin{cases}0\ , & y<0\,,\\ \left(\frac{y}{\theta}\right)^n, & 0\le y<\theta\,,\\ 1\ , & y\ge\theta\,.\end{cases}$$

对 $F_Y(y)$ 求导，可得

$$f_Y(y)=\begin{cases}\frac{ny^{n-1}}{\theta^n}, & 0<y<\theta\,,\\ 0\ , & \text{其他}.\end{cases}$$

因此，

$$E(\hat\theta)=E(Y)=\int_{-\infty}^{+\infty}y\cdot f_Y(y)\,dy=\int_0^\theta y\cdot\frac{ny^{n-1}}{\theta^n}dy=\frac{n}{\theta^n}\cdot\frac{y^{n+1}}{n+1}\bigg|_0^\theta=\frac{n}{n+1}\theta\,.$$

（3）将 $E(\alpha Y-\theta)^2$ 记为 α 的函数 $g(\alpha)$，则

$$g(\alpha)=\alpha^2\cdot\frac{n}{n+2}\theta^2-2\alpha\cdot\frac{n}{n+1}\theta^2+\theta^2=\left(\frac{n}{n+2}\alpha^2-\frac{2n}{n+1}\alpha+1\right)\theta^2\,.$$

令 $\frac{d[g(\alpha)]}{d\alpha}=\left(\frac{2n}{n+2}\alpha-\frac{2n}{n+1}\right)\theta^2=0$，解得 $\alpha=\frac{n+2}{n+1}$。该点是 $g(\alpha)$ 的唯一驻点。又因为 $g''(\alpha)=\frac{2n}{n+2}\theta^2>0$，所以该唯一驻点是 $g(\alpha)$ 的极小值点，也是最小值点。

因此，当 $\alpha=\frac{n+2}{n+1}$ 时，$E\left(\alpha\hat\theta-\theta\right)^2$ 最小。

6.（20 分）假设回归直线过原点，即一元线性回归模型为 $y_i=\beta x_i+\epsilon_i\,,i=1,2,\cdots,n\,,E(\epsilon_i)=0\,,Var(\epsilon_i)=\sigma^2$，诸观测值相互独立。

（1）写出 β 的最小二乘估计和 σ^2 的无偏估计；

（2）对给定的 x_0，其对应的因变量均值的估计为 $\hat y_0$，求 $Var(\hat y_0)$。

【解】（1）根据最小乘法原理，令 $Q=\sum_{i=1}^n(y_i-\beta x_i)^2$，则正规方程为：

$$\frac{\partial Q}{\partial\beta}\bigg|_{\hat\beta}=-2\sum_{i=1}^n\left(y_i-\hat\beta x_i\right)x_i=0$$

解得最小二乘估计为 $\hat\beta=\frac{\sum_{i=1}^n x_iy_i}{\sum_{i=1}^n x_i^2}$.

因此，

$$E(\hat\beta)=\frac{1}{\sum_{i=1}^n x_i^2}\sum_{i=1}^n x_i^2\,E(y_i)=\beta$$

$$Var(\hat\beta)=\left(\sum_{i=1}^n x_i^2\right)^{-2}\sum_{i=1}^n x_i^2\,Var(y_i)=\frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

于是由

$$S_e=\sum(y_i-\hat{y}_i)^2=\sum\left(\beta_1x_i+\epsilon_i-\hat{\beta}_1x_i\right)^2$$

$$=\sum\left[x_i^2\left(\hat{\beta}_1-\beta_1\right)^2+\epsilon_i^2-2\left(\hat{\beta}_1-\beta_1\right)x_i\epsilon_i\right]$$

有，

$$E(S_e)=\sum x_i^2Var\left(\hat{\beta}_1\right)+nVar(\epsilon)-2\sum x_iE\left(\hat{\beta}_1\epsilon_i\right)$$

将 $\hat{\beta}_1$ 可看作 $y_1,y_2,\cdots,y_n$ 的线性组合，又因为 y_j 与 $\epsilon_i(i\neq j)$ 间的独立性，有

$$E\left(\hat{\beta}_1\epsilon_i\right)=\frac{1}{\sum_{j=1}^n x_j^2}E\left(\epsilon_i\sum_j x_jy_j\right)=\frac{x_i}{\sum_{i=1}^n x_j^2}\sigma^2$$

即有 $\sum x_iE\left(\hat{\beta}_1\epsilon_i\right)=\sigma^2$ ，从而

$$E(S_e)=\sum\frac{x_i^2}{\sum_{i=1}^n x_i^2}\sigma^2+n\sigma^2-2\sigma^2=(n-1)\sigma^2$$

因此， σ^2 的无偏估计为 $S_e/(n-1)$ 。

（2）对给定的 x_0 ，对应均值的估计为 $\hat{y}_0=\hat{\beta}x_0$ ，于是

$$Var(\hat{y}_0)=x_0^2Var\left(\hat{\beta}\right)=\frac{x_0^2\sigma^2}{\sum_{i=1}^n x_i^2}$$

7.（20 分）总体 X 服从如下分布。 $X_1,\cdots,X_4$ 为其样本量为 4 的简单随机样本 $\left(0<\theta<\frac{1}{2}\right)$ 。

X	-1	0	1
P	θ	$1-2\theta$	θ

（1）求 θ 的矩估计和极大似然估计。

（2）令 $T(X_1,\cdots,X_4)=\sum_{i=1}^4 I(X_i=0)$, 其中 I 为示性函数。针对假设 $H_0\colon\theta=1/3\,,H_1\colon\theta=1/4$ 构建拒绝域 $C\{(x_1,x_2,x_3,x_4)\colon T(x_1,x_2,x_3,x_4)>2\}$ 。求此检验的第一类错误概率 α 与第二类错误概率 β 。

【解】

（1）矩估计：

$$\bar{X}=EX=0\,,\overline{X^2}=EX^2=\theta+\theta=2\theta$$

$$\hat{\theta}=\frac{\overline{x^2}}{2}=\frac{1}{2}\cdot\frac{1}{4}\sum_{i=1}^4 x_i^2=\frac{1}{8}\sum_{i=1}^4 x_i^2$$

设 n_1 为 $X=-1$ 的频数， n_2 为 $X=0$ 的频数。

极大似然估计：

$$L(\theta)=\theta^{n_1}\cdot(1-2\theta)^{n_2}\cdot\theta^{4-n_1-n_2}$$

$$\ln L(\theta)=n_1\ln\theta+n_2\ln(1-2\theta)+(4-n_1-n_2)\ln\theta$$

$$=(4-n_2)\ln\theta+n_2\ln(1-2\theta)$$

$$\frac{\partial\ln L(\theta)}{\partial\theta}=\frac{4-n_2}{\theta}-\frac{2n_2}{1-2\theta}=0$$

解得：

$$\theta^* = \frac{4-n_2}{8}$$

故矩估计为 $\hat{\theta} = \frac{1}{8}\sum_{i=1}^{4} x_i^2$ ，极大似然估计为 $\theta^* = \frac{4-n_2}{8}$ 。

（2）犯第一类错误的概率：

$$\alpha = P\left(X \in C \middle| \theta = \frac{1}{3}\right)$$

$$= P\left(T(x_1, x_2, x_3, x_4) > 2 \middle| \theta = \frac{1}{3}\right)$$

$$= P\left(\sum_{i=1}^{4} I(X_i = 0) > 2 \middle| \theta = \frac{1}{3}\right)$$

$$= P\left(\sum_{i=1}^{4} I(X_i = 0) = 3 \middle| \theta = \frac{1}{3}\right) + P\left(\sum_{i=1}^{4} I(X_i = 0) = 4 \middle| \theta = \frac{1}{3}\right)$$

$$= C_4^3 \left(1 - 2 \times \frac{1}{3}\right)^3 \times \frac{2}{3} + \left(1 - 2 \times \frac{1}{3}\right)^4 = \frac{1}{9}$$

犯第二类错误的概率：

$$\beta = P\left(X \notin C \middle| \theta = \frac{1}{4}\right)$$

$$= 1 - P\left(X \in C \middle| \theta = \frac{1}{4}\right)$$

$$= 1 - C_4^3 \left(1 - 2 \times \frac{1}{4}\right)^3 \times \frac{1}{2} - \left(1 - 2 \times \frac{1}{4}\right)^4$$

$$= \frac{11}{16}$$