“概率论与数理统计”第五次习题课答案

题1 设总体X在区间$[0,\theta]$上服从均匀分布，其中θ为位置参数，$X_1,X_2,...,X_n$是来自总体X的简单随机样本。

1. 求θ的矩估计量$\hat{\theta}_1$；
2. 求θ的最大似然估计量$\hat{\theta}_2$；
3. 讨论$\hat{\theta}_1,\hat{\theta}_2$的无偏性；
4. 求常数C_1,C_2，使得$\eta_1=C_1\hat{\theta}_1$和$\eta_2=C_2\hat{\theta}_2$ 均为θ的无偏估计；
5. 上述两个无偏估计量η_1,η_2，那个更有效；
6. 讨论$\hat{\theta}_1,\hat{\theta}_2$的相合性。

解：1). $\hat{E}(X)=\frac{\hat{\theta}_1}{2}=\overline{X} \Rightarrow \hat{\theta}_1=2\overline{X}=\frac{2}{n}\sum_{i=1}^{n}X_i$.

2). $L(\theta)=p(x_1\cdots x_n,\theta)=\frac{1}{\theta^n}\mathbb{1}\{x_1\cdots x_n\in[0,\theta]\}$.

$\theta<X_{(n)}$时，$L(\theta)=0$；　$\theta\geqslant X_{(n)}$时，$L(\theta)=\frac{1}{\theta^n}$为$\theta$的单调减函数.

$\Rightarrow \hat{\theta}_2=X_{(n)}$.

3). $E(\hat{\theta}_1)=E\left(\frac{2}{n}\sum_{i=1}^{n}X_i\right)=\frac{2}{n}\cdot n\cdot\frac{\theta}{2}=\theta$.　故$\hat{\theta}_1$无偏.

$E(\hat{\theta}_2)=E(X_{(n)})=\frac{n}{n+1}\theta\neq\theta$.　故$\hat{\theta}_2$有偏.

4). $E(\eta_1)=E(C_1\hat{\theta}_1)=C_1\theta=\theta \Rightarrow C_1=1$.

$E(\eta_2)=E(C_2\hat{\theta}_2)=C_2\frac{n}{n+1}\theta=\theta \Rightarrow C_2=\frac{n+1}{n}$.

5). $Var(\eta_1)=Var(\hat{\theta}_1)=Var\left(\frac{2}{n}\sum_{i=1}^{n}X_i\right)=\frac{4}{n^2}\cdot n\cdot Var(X)=\frac{4}{n}\cdot\frac{\theta^2}{12}=\frac{\theta^2}{3n}$.

$Var(\eta_2)=Var\left(\frac{n+1}{n}\hat{\theta}_2\right)=\frac{(n+1)^2}{n^2}Var(X_{(n)})=\frac{(n+1)^2}{n^2}\cdot\frac{n}{(n+1)^2(n+2)}\theta^2=\frac{\theta^2}{n(n+2)}$.

故当$n>1$时，$Var(\eta_1)>Var(\eta_2)$. 即η_2比η_1有效.

6). 由于$E(\hat{\theta}_1)=\theta$. 由切比雪夫不等式：

$P\left(\left|\hat{\theta}_{1,n}-\theta\right|>\varepsilon\right)\leq\frac{Var(\hat{\theta}_1)}{\varepsilon^2}=\frac{Var(\eta_1)}{\varepsilon^2}=\frac{\theta^2}{\varepsilon^2\cdot 3n}\to 0.\quad n\to\infty$.

$\Rightarrow \hat{\theta}_1\xrightarrow{P}\theta$. 即$\hat{\theta}_1$相合.

同理可证，$\eta_2\xrightarrow{p}\theta$，又$\hat{\theta}_2=\frac{n}{n+1}\eta_2$，$\frac{n}{n+1}\xrightarrow{p}1$，由P.209定理4.1.1知，$\hat{\theta}_2\xrightarrow{p}\theta*1=\theta$，即$\hat{\theta}_2$为相合估计。

题2 设总体X具有概率密度：

$$p(x)=\begin{cases}\frac{1}{\theta^2}xe^{-\frac{x}{\theta}}, & x>0;\\ 0, & x\le 0.\end{cases}$$

其中$\theta>0$为未知参数，$X_1,X_2,...,X_n$是来自X的样本，$x_1,x_2,...,x_n$是相应的样本观察值。

1. 求θ的最大似然估计量；
2. 求θ的矩估计量；
3. 问求得的估计量是否是无偏估计量。

解：1). $x_1\cdots x_n>0$ 时，$L(\theta)=\frac{1}{\theta^{2n}}\prod_{i=1}^{n}X_i\cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n}X_i}$.

$l(\theta)=\ln L(\theta)=-2n\ln\theta+\sum_{i=1}^{n}\ln X_i-\frac{1}{\theta}\sum_{i=1}^{n}X_i$.

令 $\frac{\partial l}{\partial\theta}(\hat{\theta}_1)=-\frac{2n}{\hat{\theta}_1}+\frac{1}{\hat{\theta}_1^2}\sum_{i=1}^{n}X_i=0.\Rightarrow\hat{\theta}_1=\frac{1}{2n}\sum_{i=1}^{n}X_i$.

验证：$\frac{\partial^2 l}{\partial\theta^2}(\hat{\theta}_1)=\frac{2n}{\hat{\theta}_1^2}-\frac{2}{\hat{\theta}_1^3}\sum_{i=1}^{n}X_i=\frac{2}{\hat{\theta}_1^2}\left(n-\frac{\sum_{i=1}^{n}X_i}{\hat{\theta}_1}\right)=\frac{-2n}{\hat{\theta}_1^2}<0$.

这就验证了$\hat{\theta}_1$确为最大似然估计量.

2). $\hat{E}(X)=\int_0^{+\infty}\frac{1}{\hat{\theta}_2^2}x^2e^{-\frac{x}{\hat{\theta}_2}}dx=\hat{\theta}_2\cdot\int_0^{+\infty}\left(\frac{x}{\hat{\theta}_2}\right)^2\cdot e^{-\frac{x}{\hat{\theta}_2}}d\left(\frac{x}{\hat{\theta}_2}\right)=\Gamma(3)\cdot\hat{\theta}_2=2\hat{\theta}_2$

$=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

$\Rightarrow\hat{\theta}_2=\frac{1}{2n}\sum_{i=1}^{n}X_i=\hat{\theta}_1$.

3). 通过与2)中类似的计算可得：$E(X)=2\theta$.

从而 $E(\hat{\theta}_1)=E(\hat{\theta}_2)=E\left(\frac{1}{2n}\sum_{i=1}^{n}X_i\right)=\frac{1}{2n}\sum_{i=1}^{n}E(X_i)=\frac{1}{2n}\cdot n\cdot 2\theta=\theta$.

即$\hat{\theta}_1,\hat{\theta}_2$是无偏估计量.

□.

题3 设$X_1,X_2,...,X_{n+1}$是来自正态总体的简单样本，$X_1,X_2,...,X_n$的样本均值和样本二阶中心矩分别为$\bar{X}$和S_n^2，求$\frac{X_{n+1}-\bar{X}}{S_n}\sqrt{\frac{n-1}{n+1}}$的分布。

解：记总体分布为 $N(\mu,\sigma^2)$.

$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$. 记 $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$.

由定理5.4.1有：

$\bar{X}$与s^2相互独立，$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$，$\frac{(n-1)\cdot s^2}{\sigma^2}\sim\chi^2(n-1)$.

X_{n+1}与$X_1,\cdots X_n$独立 $\Rightarrow$ X_{n+1}与$\bar{X}$独立. 且 $X_{n+1}\sim N(\mu,\sigma^2)$. $-\bar{X}\sim N(-\mu,\frac{\sigma^2}{n})$.

$\Rightarrow X_{n+1}-\bar{X}\sim N(0,\frac{n+1}{n}\sigma^2)$.

$\Rightarrow \frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}\sim N(0,1)$.

$$\frac{X_{n+1}-\bar{X}}{S_n}\sqrt{\frac{n-1}{n+1}}=\frac{\frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}}{\frac{\sqrt{\frac{n+1}{n-1}S_n^2}}{\sqrt{\frac{n+1}{n}\sigma^2}}}=\frac{\frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}}{\sqrt{\frac{nS_n^2}{(n-1)\sigma^2}}}=\frac{\frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}}{\sqrt{\frac{(n-1)\cdot s^2}{\sigma^2}\Big/(n-1)}}$$

X_{n+1}与$X_1\cdots X_n$独立 $\Rightarrow$ X_{n+1}与s^2独立. $\Rightarrow$ $X_{n+1}-\bar{X}$与s^2独立.

$\Rightarrow \frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}$与$\frac{(n-1)s^2}{\sigma^2}$独立. $\Rightarrow \frac{X_{n+1}-\bar{X}}{S_n}\sqrt{\frac{n-1}{n+1}}\sim t(n-1)$. □

题4 设$x_1,x_2,...,x_n$为抽自正态总体$N(\mu,\sigma^2)$的简单样本。

1. 如果$\sigma=4$，为使得μ的置信水平为$1-\alpha$的置信区间的长度不大于给定L，试问样本容量至少要多少？
2. 如果σ未知，在已知样本的容量、均值和标准差分别为$n,\bar{x},s$的前提下，试以$1-\alpha$ 把握估计最小的μ的值.

解：(1) 在置信水平为$1-\alpha$下，μ的置信区间为

$$[\bar{x}-4u_{1-\alpha/2}/\sqrt{n},\bar{x}+4u_{1-\alpha/2}/\sqrt{n}],$$

故依题有，$8u_{1-\alpha/2}/\sqrt{n}\le L$，所以$n\ge 64(u_{1-\alpha/2}/L)^2$。

(2)构造t统计量$T_n=\frac{\sqrt{n}(\bar{x}-\mu)}{s}\sim t(n-1)$并以此作为枢轴量，为了构造单侧置信下限，我们考虑

$$P(T_n\le t_{1-\alpha}(n-1))=P(\frac{\sqrt{n}(\bar{x}-\mu)}{s}\le t_{1-\alpha}(n-1))=P(\mu\ge\bar{x}-t_{1-\alpha}(n-1)s/\sqrt{n})=1-\alpha.$$

所以，μ的$1-\alpha$置信下限为$\bar{x}-t_{1-\alpha}(n-1)s/\sqrt{n}$。

题5 设$X_1,...,X_{16}$是来自正态总体$N(\mu,1)$的样本，这里$\mu\ge 0$是未知参数。对以下原假设和备择假设

$$H_0:\mu=0\ vs\ H_1:\mu>0,$$

若取拒绝域为$\{(x_1,...,x_{16}):\bar{x}>0.5\}$，试求(结果可用标准正态分布函数$\phi(x)$表示)

1. 此检验犯的第一类错误的概率；
2. 当$\mu=1$时此检验犯的第二类错误的概率；

解：(1)犯第一类错误的概率：$\alpha=P(\bar{x}>0.5|\mu=0)=P(\frac{\bar{x}-0}{1/4}>2|\mu=0)=1-\phi(2)$。

(2)犯第二类错误的概率：$\beta=P(\bar{x}\le 0.5|\mu=1)=P(\frac{\bar{x}-1}{1/4}\le -2|\mu=1)=\phi(-2)=1-\phi(2)$。

题6 设$x_1, ..., x_{16}$是来自正态总体$N(\mu, 4)$的样本，考虑检验问题

$$H_0 : \mu = 6 \ vs \ H_1 : \mu \neq 6,$$

拒绝域取为$W = \{|\bar{x} - 6| \geq c\}$，试求$c$使得检验的显著性水平为0.05，并求该检验在$\mu = 6.5$处犯第二类错误的概率。

解：在H_0为真的条件下，$\bar{x} \sim N(6, 1/4)$，因而由

$$P(|\bar{x} - 6| \geq c|\mu = 6) = 0.05$$

得，

$$P(\frac{\bar{x} - 6}{0.5} \geq \frac{c}{0.5}) = 1 - \phi(2c) = 0.025,$$

解得，$c = u_{0.975}/2 = 0.98$。

该检验在$\mu = 6.5$处犯第二类错误的概率为

$$\beta = P(|\bar{x} - 6| < 0.98|\mu = 6.5) = P(-2.96 \leq \frac{\bar{x} - 6.5}{0.5} < 0.96) = \phi(0.96) - \phi(-2.96) = 0.83.$$

题7 设需要对某正态总体的均值进行假设检验

$$H_0 : \mu = 15 \ vs \ H_1 : \mu < 15.$$

已知$\sigma^2 = 2.5$，取$\alpha = 0.05$，若要求当H_1中的$\mu \leq 13$时犯的第二类错误的概率不超过0.05，求所需的样本容量。

解：此题为单边假设，考虑拒绝域$\{u \leq u_\alpha\}$，其中$u = \dfrac{\bar{x} - 15}{\sqrt{2.5/n}}$。查表知，$u_\alpha = u_{0.05} = -1.65$。则第二类错误概率

$$\beta(\mu) = P(u > -1.65|\mu \leq 13) = P(\frac{\bar{x} - \mu}{\sqrt{2.5/n}} > \frac{15 - \mu}{\sqrt{2.5/n}} - 1.65|\mu \leq 13) = 1 - \phi(\frac{15 - \mu}{\sqrt{2.5/n}} - 1.65) \leq 0.05.$$

因为$\beta(\mu)$关于μ是递增函数，又$\mu \leq 13$，故只需$\beta(13) \leq 0.05$成立即可，由此解得$n \geq 7$。

题8 假定考生成绩服从正态分布，在某地一次数学考试中，随机抽取了36位考生的成绩，算得平均成绩为66.5分，标准差为15分，现要对“这次考试的平均成绩是否为70分”做出判断，

1. 构造相应假设检验，并求该检验的p值；
2. 在显著性水平为0.05下，是否可以认为这次考试的平均成绩为70分？

解：(1)设考生成绩成绩服从$N(\mu, \sigma^2)$，构造假设检验

$$H_0 : \mu = 70 \ vs \ H_1 : \mu \neq 70.$$

由于σ未知，故用t检验法，拒绝域为$\left\{|t| \geq t_{1-\alpha/2}(35)\right\}$，其中$\alpha$为给定的显著水平，$t = \dfrac{\bar{x} - 70}{s/6} \sim t(35)$，由题知，$\bar{x} = 66.5, s = 15$。所以$p = P(|t| \geq \left|\frac{66.5-70}{15/6}\right|) = P(|t| \geq 1.4) = 2(1 - P(t \leq 1.4)) = 0.1703$。

(2)因为$p > 0.05$，所以接受原假设，即认为这次考试的平均成绩为70分。

题9 设 $x_1,\ldots,x_n$ 是来自正态总体$N(\mu,\sigma^2)$的一个样本,对σ^2考虑如下三个估计

$$\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,\hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,\hat{\sigma}_3^2=\frac{1}{n+1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

(1)哪一个是σ^2的无偏估计?

(2)哪一个均方误差最小?

解(1)由于$\frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\sim\chi^2(n-1)$,故有$E[\sum\limits_{i-1}^{n}(x_i-\bar{x})^2]=(n-1)\sigma^2$,从而

$$E(\hat{\sigma}_1^2)=\sigma^2,E(\hat{\sigma}_2^2)=\frac{n-1}{n}\sigma^2,E(\hat{\sigma}_3^2)=\frac{n-1}{n+1}\sigma^2.$$

这说明仅有$\hat{\sigma}_1^2$ 是σ^2的无偏估计,而$\hat{\sigma}_2^2$与$\hat{\sigma}_3^2$是σ^2的有偏估计.

(2)我们知道,估计的均方误差是估计的方差加上偏差的平方，即

$$E(\hat{\sigma}^2-\sigma^2)^2=Var(\hat{\sigma}^2)+(E(\hat{\sigma}^2)-\sigma^2)^2.$$

而$Var(\sum\limits_{i=1}^{n}(x_i-\bar{x})^2)=2(n-1)\sigma^4$,这给出

$$Var(\hat{\sigma}_1^2)=\frac{2\sigma^4}{n-1},Var(\hat{\sigma}_2^2)=\frac{2(n-1)\sigma^4}{n^2},Var(\hat{\sigma}_3^2)=\frac{2(n-1)\sigma^4}{(n+1)^2}.$$

于是

$$MSE(\hat{\sigma}_1^2)=Var(\hat{\sigma}_1^2)=\frac{2\sigma^4}{n-1},$$

$$MSE(\hat{\sigma}_2^2)=\frac{2(n-1)\sigma^4}{n^2}+(\frac{n-1}{n}\sigma^2-\sigma^2)^2=\frac{2n-1}{n^2}\sigma^4,$$

$$MSE(\hat{\sigma}_3^2)=\frac{2(n-1)\sigma^4}{(n+1)^2}+(\frac{n-1}{n+1}\sigma^2-\sigma^2)^2=\frac{2}{(n+1)}\sigma^4.$$

显然$\frac{2}{n-1}>\frac{2}{n+1},\frac{2n-1}{n^2}>\frac{2}{n+1}(n>1)$,所以$\hat{\sigma}_3^2$的均方误差最小.