数理统计第六章测验题

考试时间2023年5月14日，答卷时间120分钟，总分100分，出题人-李绍文

1．（10 分）设总体X的密度函数$p(x;\lambda,\theta)=\lambda e^{-\lambda(x-\theta)}I_{x>\theta}$，其中$\lambda>0$与$\theta$为参数，$X_1,X_2,\cdots,X_n$为样本，求参数$\lambda,\theta$的最大似然估计。

解：似然函数

$$L(\lambda,\theta)=\prod_{i=1}^{n}\lambda e^{-\lambda(x_i-\theta)}I_{x_i>\theta}=\lambda^n e^{-\lambda\left(\sum_{i=1}^{n}x_i-n\theta\right)}I_{x_1,x_2,\cdots,x_n>\theta},$$

当$x_1,x_2,\cdots,x_n>\theta$时，

$$\ln L(\lambda,\theta)=n\ln\lambda-\lambda\left(\sum_{i=1}^{n}x_i-n\theta\right)。$$

令

$$\frac{\partial\ln L(\lambda,\theta)}{\partial\lambda}=\frac{n}{\lambda}-\left(\sum_{i=1}^{n}x_i-n\theta\right)=0,$$

解得

$$\lambda=\frac{n}{\sum_{i=1}^{n}x_i-n\theta}=\frac{1}{\bar{x}-\theta}。$$

观察似然函数。对于$\lambda^n e^{-\lambda\left(\sum_{i=1}^{n}x_i-n\theta\right)}$，当$\theta$越大时，其值越大；对于$I_{x_1,x_2,\cdots,x_n>\theta}$，当$x_1,x_2,\cdots,x_n>\theta$时，即$x_{(1)}>\theta$，其值才不为 0。可见当$\theta=x_{(1)}$时，似然函数$L(\lambda,\theta)$达到最大值。

故θ的最大似然估计为$\hat{\theta}=X_{(1)}$，λ的最大似然估计为$\hat{\lambda}=\dfrac{1}{\bar{X}-X_{(1)}}$。

2．（10 分）设总体X期望为μ，方差为σ^2，$X_1,X_2,\cdots,X_n$为X的样本。若$Y=c\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$是总体方差$\sigma^2$的无偏估计，求常数$c$。

解：因

$$E(X_{i+1}-X_i)=E(X_{i+1})-E(X_i)=0,$$

$$\operatorname{Var}(X_{i+1}-X_i)=\operatorname{Var}(X_{i+1})+\operatorname{Var}(X_i)=2\sigma^2,$$

有

$$E[(X_{i+1}-X_i)^2]=\operatorname{Var}(X_{i+1}-X_i)+[E(X_{i+1}-X_i)]^2=2\sigma^2,$$

则

$$E(Y)=c\sum_{i=1}^{n-1}E[(X_{i+1}-X_i)^2]=c\cdot(n-1)\cdot 2\sigma^2,$$

故当$c=\dfrac{1}{2(n-1)}$时，$Y=c\sum\limits_{i=1}^{n-1}(X_{i+1}-X_i)^2$是总体方差$\sigma^2$的无偏估计。

3. （10分）某厂产品重量X（克）服从正态分布$N(\mu,\sigma^2)$，标准差$\sigma=20$克。问：

（1）样本容量n取多大时，才能保证μ的95%置信区间长度不超过10；

（2）抽取容量为100的样本，样本均值$\bar{x}=972$克，求μ的95%置信区间。

解：单个正态总体，已知σ，估计μ。μ的置信区间为$\left[\bar{X}\pm u_{1-\frac{\alpha}{2}}\cdot\dfrac{\sigma}{\sqrt{n}}\right]$。

（1）置信区间长度$2u_{1-\frac{\alpha}{2}}\cdot\dfrac{\sigma}{\sqrt{n}}\le 10$。即

$$2\times1.96\times\frac{20}{\sqrt{n}}\le 10,$$

故$n\ge 61.47$，即n至少为62。

（2）μ的95%置信区间为

$$\left[972\pm1.96\times\frac{20}{10}\right]=[968.08, 975.92]。$$

4. （10分）设A、B两台机床生产的金属部件重量（克）各自服从正态分布。分别抽取8件和9件产品，测量后经计算得$\bar{x}=145$，$s_1^2=6^2$，$\bar{y}=130$，$s_2^2=7^2$，求：

（1）总体方差相等时，均值差$\mu_1-\mu_2$的95%置信区间；

（2）方差比$\dfrac{\sigma_1^2}{\sigma_2^2}$的95%置信区间。

解：（1）两个正态总体，未知σ_1^2, σ_2^2，估计$\mu_1-\mu_2$。$\mu_1-\mu_2$的置信区间为

$$\left[\bar{X}-\bar{Y}\pm t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]。$$

故均值差$\mu_1-\mu_2$的 95%置信区间为

$$\left[145-130\pm 2.1314\times 6.5523\times\sqrt{\frac{1}{8}+\frac{1}{9}}\right]=[8.2139, 21.7861]。$$

（2）两个正态总体，估计$\frac{\sigma_1^2}{\sigma_2^2}$。$\frac{\sigma_1^2}{\sigma_2^2}$的置信区间为

$$\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)}, \frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{\frac{\alpha}{2}}(n_1-1, n_2-1)}\right]。$$

故方差比$\frac{\sigma_1^2}{\sigma_2^2}$的 95%置信区间

$$\left[\frac{36}{49}\times\frac{1}{4.53}, \frac{36}{49}\times 4.9\right]=[0.1622, 3.6]。$$

5．（10 分）设事件A在一次试验中发生的概率为p。进行 146 次独立重复试验，事件A发生了 58 次，求概率p的 95%置信区间。

解：估计概率p。概率p的修正置信区间为$\left[\bar{X}*\pm u_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\bar{X}*(1-\bar{X}*)}{n+4}}\right]$。其中修正频率$\bar{X}*=\frac{58+2}{146+4}=0.4$。故概率$p$的 95%修正置信区间为

$$\left[0.4\pm 1.96\times\sqrt{\frac{0.4\times 0.6}{150}}\right]=[0.3216, 0.4784]。$$

注：近似法的结果$[0.3179, 0.4766]$，方程法的结果$[0.3215, 0.4783]$。

6．（15 分）设总体X服从泊松分布$P(\lambda)$，参数λ的先验分布是指数分布$Exp(\theta)$，θ已知，$X_1, X_2, \cdots, X_n$为样本，求λ的贝叶斯估计$\hat{\lambda}_B$。

解：参数λ的先验分布$\pi(\lambda)=\theta e^{-\theta\lambda}I_{\lambda>0}$。

总体条件分布 $p(x\mid\lambda)=\dfrac{\lambda^x}{x!}e^{-\lambda},\quad x=0,1,2,\cdots$。样本条件分布

$$p(x_1,x_2,\cdots,x_n\mid\lambda)=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda},\quad x_1,x_2,\cdots,x_n=0,1,2,\cdots。$$

分子

$$h(x_1,x_2,\cdots,x_n,\lambda)=\frac{\lambda^{n\bar{x}}}{x_1!x_2!\cdots x_n!}\theta e^{-\lambda(\theta+n)}I_{\lambda>0},\quad x_1,x_2,\cdots,x_n=0,1,2,\cdots;$$

分母

$$m(x_1,x_2,\cdots,x_n)=\int_0^{+\infty}\frac{\lambda^{n\bar{x}}}{x_1!x_2!\cdots x_n!}\theta e^{-\lambda(\theta+n)}d\lambda 。$$

$$=\frac{\theta}{x_1!x_2!\cdots x_n!}\cdot\frac{\Gamma(n\bar{x}+1)}{(\theta+n)^{n\bar{x}+1}},\quad x_1,x_2,\cdots,x_n=0,1,2,\cdots。$$

则参数 λ 的后验分布为

$$\pi(\lambda\mid x_1,x_2,\cdots,x_n)=\frac{(\theta+n)^{n\bar{x}+1}}{\Gamma(n\bar{x}+1)}\lambda^{n\bar{x}}e^{-\lambda(\theta+n)}I_{\lambda>0}。$$

这是伽玛分布 $Ga(n\bar{x}+1,\theta+n)$，故 λ 的贝叶斯估计为 $\hat{\lambda}_B=\dfrac{n\bar{x}+1}{\theta+n}$。

7. （15 分）设 $X_1,X_2,\cdots,X_n$ 是来自伽玛分布 $Ga(\alpha,\theta)$ 的样本，已知 $\alpha>0$，试证明，$\dfrac{\bar{X}}{\alpha}$ 是 $g(\theta)=\dfrac{1}{\theta}$ 的有效估计，从而也是 UMVUE。

证明：无偏性：

$$E\left(\frac{\bar{X}}{\alpha}\right)=\frac{1}{\alpha}E(X)=\frac{1}{\alpha}\cdot\frac{\alpha}{\theta}=\frac{1}{\theta},$$

即 $\dfrac{\bar{X}}{\alpha}$ 是 $g(\theta)=\dfrac{1}{\theta}$ 的无偏估计。

有效性：

$$\mathrm{Var}\left(\frac{\bar{X}}{\alpha}\right)=\frac{1}{n\alpha^2}\mathrm{Var}(X)=\frac{1}{n\alpha^2}\cdot\frac{\alpha}{\theta^2}=\frac{1}{n\alpha\theta^2}。$$

因 $Ga(\alpha,\theta)$ 的密度函数为

$$p(x;\alpha,\theta)=\frac{\theta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\theta x}I_{x>0},$$

当 $x>0$ 时，

$$\ln p(x;\alpha,\theta)=\alpha\ln\theta+(\alpha-1)\ln x-\theta x-\ln\Gamma(\alpha),$$

$$\frac{\partial\ln p(x;\alpha,\theta)}{\partial\theta}=\frac{\alpha}{\theta}-x,$$

可得 θ 的 Fisher 信息量为

$$I(\theta)=E\left[\left(\frac{\alpha}{\theta}-X\right)^2\right]=\mathrm{Var}(X)=\frac{\alpha}{\theta^2},$$

则 $g(\theta)=\dfrac{1}{\theta}$ 的 CR 下界为

$$\frac{[g'(\theta)]^2}{nI(\theta)}=\frac{\left(-\dfrac{1}{\theta^2}\right)^2}{n\cdot\dfrac{\alpha}{\theta^2}}=\frac{1}{n\alpha\theta^2}=\mathrm{Var}\left(\frac{\bar{X}}{\alpha}\right)。$$

故 $\dfrac{\bar{X}}{\alpha}$ 是 $g(\theta)=\dfrac{1}{\theta}$ 的有效估计。

8．（20 分）总体 X 服从泊松分布 $P(\lambda)$， $X_1,X_2,\cdots,X_n$ 为样本。

（1）参数 λ 的点估计 $\hat{\lambda}=\bar{X}$，由此猜测 $g(\lambda)=\lambda^2$ 的点估计为 $\bar{X}^2$。判断 $\bar{X}^2$ 是否 $g(\lambda)=\lambda^2$ 的无偏估计？如果不是，请根据 $E(\bar{X}^2)$ 的结果及 $E(\bar{X})=\lambda$ 修偏得到 $\hat{g}$，使得 $\hat{g}$ 是 $g(\lambda)=\lambda^2$ 的无偏估计，即 $E(\hat{g})=\lambda^2$；

（2）写出样本联合密度函数 $p(x_1,x_2,\cdots,x_n;\lambda)$，证明 $\hat{g}$ 是 $g(\lambda)=\lambda^2$ 的 UMVUE；

（3）求出 λ 的 Fisher 信息量 $I(\lambda)$ 及 $g(\lambda)=\lambda^2$ 的 C-R 下界；

（4）设 Y 服从泊松分布 $P(\theta)$，可知 $\mathrm{Var}(Y^2-Y)=4\theta^3+2\theta^2$，根据此结论以及泊松分布的可加性求出 $\mathrm{Var}(\hat{g})$，并判断 $\hat{g}$ 是否 $g(\lambda)=\lambda^2$ 的有效估计。

解：（1）因

$$E(\bar{X}^2)=\mathrm{Var}(\bar{X})+[E(\bar{X})]^2=\frac{1}{n}\mathrm{Var}(X)+[E(X)]^2=\frac{\lambda}{n}+\lambda^2\neq\lambda^2,$$

故 $\bar{X}^2$ 不是 $g(\lambda)=\lambda^2$ 的无偏估计。而

$$E\left(\bar{X}^2-\frac{\bar{X}}{n}\right)=E(\bar{X}^2)-\frac{1}{n}E(\bar{X})=\frac{\lambda}{n}+\lambda^2-\frac{\lambda}{n}=\lambda^2,$$

令 $\hat{g}=\bar{X}^2-\dfrac{\bar{X}}{n}$，有 $E(\hat{g})=\lambda^2$，故 $\hat{g}=\bar{X}^2-\dfrac{\bar{X}}{n}$ 是 $g(\lambda)=\lambda^2$ 的无偏估计。

（2）因 $E(\hat{g})=\lambda^2$，且

$$p(x_1,x_2,\cdots,x_n;\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i\,!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1\,!x_2\,!\cdots x_n\,!}e^{-n\lambda},\quad x_1,x_2,\cdots,x_n=0,1,2,\cdots\ ,$$

$$e^{n\lambda}p(x_1,x_2,\cdots,x_n;\lambda)=\frac{\lambda^{n\bar{x}}}{x_1\,!x_2\,!\cdots x_n\,!},\quad x_1,x_2,\cdots,x_n=0,1,2,\cdots,$$

则

$$\frac{\partial[e^{n\lambda}p(x_1,x_2,\cdots,x_n;\lambda)]}{\partial\lambda}=\frac{n\bar{x}\lambda^{n\bar{x}-1}}{x_1\,!x_2\,!\cdots x_n\,!}=\frac{n\bar{x}}{\lambda}e^{n\lambda}p(x_1,x_2,\cdots,x_n;\lambda),$$

令统计量 $T=\bar{X}$，$c=e^{n\lambda}$，$a=\dfrac{ne^{n\lambda}}{\lambda}$，$b=0$，即 $\dfrac{\partial(cp)}{\partial\lambda}=(a\bar{x}+b)p$，可知根据 $E(\varphi)=0$ 可得到 $E(\varphi\bar{X})=0$。

取 $\varphi^*=\varphi\bar{X}$，根据 $E(\varphi)=0$ 可得到 $E(\varphi^*)=0$，再根据 $E(\varphi^*)=0$ 及统计量 φ 的任意性，可得到 $E(\varphi^*\bar{X})=0$，即 $E(\varphi\bar{X}^2)=0$。从而根据 $E(\varphi)=0$ 可得到

$$E(\varphi\hat{g})=E\left[\varphi\left(\bar{X}^2-\frac{1}{n}\bar{X}\right)\right]=E(\varphi\bar{X}^2)-\frac{1}{n}E(\varphi\bar{X})=0,$$

故 $\hat{g}$ 是 $g(\lambda)=\lambda^2$ 的 UMVUE。

（3）因泊松分布 $P(\lambda)$ 的质量函数为

$$p(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda},\quad x=0,1,2,\cdots,$$

则

$$\ln p(x;\lambda)=x\ln\lambda-\lambda-\ln(x!),\quad x=0,1,2,\cdots,$$

$$\frac{\partial \ln p(x;\lambda)}{\partial \lambda}=\frac{x}{\lambda}-1=\frac{x-\lambda}{\lambda},$$

故

$$I(\lambda)=E\left[\frac{\partial \ln p(X;\lambda)}{\partial \lambda}\right]^2=E\left(\frac{X-\lambda}{\lambda}\right)^2=\frac{E(X-\lambda)^2}{\lambda^2}=\frac{\mathrm{Var}(X)}{\lambda^2}=\frac{\lambda}{\lambda^2}=\frac{1}{\lambda}。$$

且 $g(\lambda)=\lambda^2$ 的 C-R 下界为

$$\frac{[g'(\lambda)]^2}{nI(\lambda)}=\frac{(2\lambda)^2}{n\cdot\frac{1}{\lambda}}=\frac{4\lambda^3}{n}。$$

（4）因 $E(\hat{g})=\lambda^2$，并根据泊松分布的可加性可知

$$Y=n\bar{X}=\sum_{i=1}^{n}X_i\sim P(n\lambda),$$

则

$$\mathrm{Var}(Y^2-Y)=\mathrm{Var}(n^2\bar{X}^2-n\bar{X})=4n^3\lambda^3+2n^2\lambda^2,$$

$$\mathrm{Var}(\hat{g})=\mathrm{Var}\left(\bar{X}^2-\frac{\bar{X}}{n}\right)=\frac{4n^3\lambda^3+2n^2\lambda^2}{n^4}=\frac{4n\lambda^3+2\lambda^2}{n^2}>\frac{4\lambda^3}{n}=\frac{[g'(\lambda)]^2}{nI(\lambda)},$$

故 $\hat{g}$ 不是 $g(\lambda)=\lambda^2$ 的有效估计。