23 数理统计期末残卷

$\mathscr{NULIOUS}$

1. 一个移动通讯公司随机抽取了其 900 个包月客户，计算得知他们一个月平均使用时间是 220 分钟，样本标准差是 90 分钟．假设使用时间服从正态分布 $N\left(\mu,\sigma^2\right)$ ．
（1）求包月客户平均使用时间和标准差的 95% 置信区间。
（2）如果要求客户平均使用时间的 95% 置信区间的长度不超过 5 分钟，应至少抽取多少个客户？该公司的抽样规模是否满足要求？
（3）假设总体标准差为 90 分钟，取 μ 的先验分布为无信息先验，求 μ 的后验分布并据此给出 μ 的可信系数为 95% 的可信区间。
（4）解释（1）和（3）中关于平均使用时间所得区间的含义与区别．

解：（1）μ,σ^2 均未知
利用

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S}\to t_{n-1}$$

估计 μ ，进而有置信区间

$$\left[\overline{X}-t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}},\ \overline{X}+t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}\right]$$

代值有

$$[214.12,225.885]$$

这里因为 $t_n\to N(0,1)$ ，用 $u_{\frac{\alpha}{2}}$ 代替 $t_{n-1}\left(\frac{\alpha}{2}\right)$
再利用

$$\frac{(n-1)S^2}{\sigma^2}\to\chi^2_{(n-1)}$$

σ^2 有置信区间

$$\left[\frac{(n-1)S^2}{\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)},\frac{(n-1)S^2}{\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)}\right]$$

开方有 σ 的置信区间

$$\left[\sqrt{\frac{(n-1)S^2}{\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)}},\sqrt{\frac{(n-1)S^2}{\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)}}\right]$$

代值为

$$[88.15,92.02]$$

（2）即需要

$$2\times u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\leq 5$$

这里同样因为 $t_n\to N(0,1)$ ，用 $u_{\frac{\alpha}{2}}$ 代替 $t_{n-1}\left(\frac{\alpha}{2}\right)$ 解得

$$n\geq 4979$$

不满足要求

（3）位置参数的无信息先验为

$$\pi(\mu)=1$$

样本联合密度为

$$\begin{aligned}
f(x_1,\cdots,x_n;\mu)&=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}\\
&=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2+n(\overline{X}-\mu)^2\right\}
\end{aligned}$$

则

$$\pi(\mu\mid\vec{X})\propto\exp\left\{-\frac{n}{2\sigma^2}\left(\overline{X}-\mu\right)^2\right\}$$

也就是

$$\mu\mid\vec{X}\sim N\left(\overline{X},\frac{\sigma^2}{n}\right)$$

则可信区间为

$$\left[\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right]$$

代入数值为

$$[214.12,225.88]$$

（4）置信区间：经过多次重复实验，μ 落在置信区间的**频率**趋于 95%，参数是一个真实的固定值

可信区间：相当于把 μ 视为**随机变量**，一次试验后 μ 落在可信区间的**概率**为 95%

2. 设 $X_1,\ldots,X_n$ i.i.d.$\sim N(\theta,1),Y_1,\ldots,Y_n$ i.i.d.$\sim N(2\theta,1),\theta$ 为未知参数. 又设合样本 $X_1,\ldots,X_n,Y_1,\ldots,Y_n$ 独立. 试

（1）求 θ 的 UMVUE.

（2）如果要求 θ 的置信系数为 $1-\alpha$ 的置信区间宽不超过指定的 d，则样本量 n 应该至少多大？

解：（1）样本联合密度为

$$\begin{aligned}
f\left(X_1,\cdots,X_n,Y_1,\cdots,Y_n;\theta\right)&=\left(\frac{1}{\sqrt{2\pi}}\right)^{-2n}\cdot e^{\frac{-\sum_{i=1}^{n}\left(X_i-\theta\right)^2-\sum_{i=1}^{n}\left(Y_i-2\theta\right)^2}{2}}\\
&=\left(\frac{1}{2\pi}\right)^{n}\cdot e^{-\frac{5n\theta^2}{2}}\cdot e^{\theta\cdot\left(\sum_{i=1}^{n}\left(X_i+2Y_i\right)\right)}\cdot e^{-\frac{1}{2}\sum_{i=1}^{n}\left(X_i^2+Y_i^2\right)}
\end{aligned}$$

显然自然参数空间有内点，则 $T(\boldsymbol{X})=\sum\limits_{i=1}^{n}\left(X_i+2Y_i\right)$ 为充分完全统计量

注意独立性，有

$$T(\boldsymbol{X})\sim N(5n\theta,5n)$$

则有无偏估计

$$\frac{T(\boldsymbol{X})}{5n}=\frac{\overline{X}+2\overline{Y}}{5}$$

它也是充分完全统计量的函数，则其为 UMVUE

（2）取

$$\frac{\frac{\overline{X}+2\overline{Y}}{5}-\theta}{\frac{1}{\sqrt{5n}}}\sim N(0,1)$$

作为枢轴变量

则区间长度

$$\frac{2u_{\alpha/2}}{\sqrt{5n}} \leq d$$

得到

$$n \geq \left\lceil \frac{4u_{\alpha/2}^2}{5d^2} \right\rceil$$

3. 随机地从一批铁钉中抽取 16 枚，测得它们的平均长度（单位：cm)$\bar{x}=3.035$ ．已知铁钉长度 X 服从正态分布 $N\left(\mu, 0.1^2\right)$ ．考虑假设检验问题

$$H_0: \mu \leq 3 \longleftrightarrow H_1: \mu>3 \tag{$\star$}$$

（1）在 $\alpha=0.05$ 水平下对检验问题 $(\star)$ 进行检验，并给出检验的 p 值。

（2）如果行动 $a=0$ 和 $a=1$ 分别表示接受 H_0 和拒绝 H_0, μ 的先验分布为 $N\left(3,0.1^2\right)$ ，损失函数取为

$$L(a, \mu)= \begin{cases}11, & a=1, \mu \leq 3, \\ 1, & a=0, \mu>3, \\ 0, & \text{其他.}\end{cases}$$

基于后验风险最小原则给出检验问题 $(\star)$ 的最优决策行动，并与（1）中检验结论进行对比。

解：（1）给出检验统计量

$$U(\boldsymbol{X})=\frac{\bar{X}-\mu_0}{\sigma / \sqrt{n}} \stackrel{H_0}{\sim} N(0,1)$$

由检验问题的形式知拒绝域为

$$D=\left\{\boldsymbol{X}: U(\boldsymbol{X})>u_\alpha\right\}$$

现有 $\bar{x}=3.035, \mu_0=3, \sigma=0.1, n=16$ ，查表得 $u_{0.05}=1.645$ ，因此

$$u(\boldsymbol{x})=\frac{3.035-3}{0.1 / \sqrt{16}}=1.4<1.645$$

因此不能拒绝原假设.

检验的 p 值为

$$p=P(U(\boldsymbol{X})>u(\boldsymbol{x}))=1-\Phi(1.4)=0.0808$$

（2）在先验分布 $\mu \sim N\left(\mu_0, \tau^2\right)$ 下，正确计算出 μ 的后验分布

$$\mu \,\Big|\, \boldsymbol{x} \sim N\left(\frac{n \tau^2 \bar{x}+\sigma^2 \mu_0}{n \tau^2+\sigma^2}, \frac{\tau^2 \sigma^2}{n \tau^2+\sigma^2}\right) \stackrel{\text{代入数据}}{=} N\left(\frac{1289}{425}, \frac{1}{1700}\right).$$

后验风险

$$R\left(a_0, \mu\right)=\mathbb{P}(\mu>3 \mid \boldsymbol{x})=0.9131 \quad R\left(a_1, \mu\right)=11 \mathbb{P}(\mu \leq 3 \mid \boldsymbol{x})=0.9559$$

结论：按后验风险最小原则，应采取行动 a_0 ，即接受原假设 H_0 。

这与（1）中检验结果一致，这是因为我们对拒绝 H_0 赋予了较大的惩罚。

4. 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\theta, & 0<x<1,\\ 1-\theta, & 1\le x<2,\\ 0, & \text{其他},\end{cases}$$

其中 $0<\theta<1$ 为未知参数．现从此总体中抽出一样本量为 n 的样本 $X_1,\ldots,X_n$ ．

（1）试求参数 θ 的充分统计量，并说明它是否为完全统计量？

（2）求假设检验问题 $H_0:\theta=\theta_0\leftrightarrow H_1:\theta\neq\theta_0$ 的水平 α 似然比检验，其中 $0<\theta_0,\alpha<1$ 已知。

（3）上述假设是否存在 UMPT？为什么？

解：（1）样本联合密度

$$f(x_1,\cdots,x_n;\theta)=\theta^{n_1}\cdot(1-\theta)^{n_2}\mathbb{I}_{\left\{0<X_{(1)}<X_{(n)}<2\right\}}$$

其中

$$n_1=\sum_{i=1}^n\mathbb{I}_{\{0<X_i<1\}}\quad n_2=n-n_1$$

则

$$n_1\sim\mathrm{B}(n,\theta)$$

将联合密度改写成指数族形式

$$\begin{aligned}f(x_1,\cdots,x_n;\theta)&=e^{n_1\ln\theta+n_2\ln(1-\theta)}\mathbb{I}_{\left\{0<X_{(1)}<X_{(n)}<2\right\}}\\&=(1-\theta)^n\cdot e^{n_1\ln\frac{\theta}{1-\theta}}\cdot h(\boldsymbol{X})\end{aligned}$$

由因子分解定理和自然参数空间有内点，$T(\boldsymbol{X})=n_1$ 为充分完全统计量

（2）对全空间 $0<\theta<1$，易得

$$\hat{\theta}_{MLE}=\frac{n_1}{n}$$

则

$$\sup_{0<\theta<1}f(x_1,\cdots,x_n;\theta)=\left(\frac{n_1}{n}\right)^{n_1}\left(1-\frac{n_1}{n}\right)^{n-n_1}\mathbb{I}_{\left\{0<X_{(1)}<X_{(n)}<2\right\}}$$

另一方面

$$\sup_{\theta=\theta_0}f(x_1,\cdots,x_n;\theta)=f(x_1,\cdots,x_n;\theta_0)$$

则似然比为

$$\lambda(\boldsymbol{X})=\frac{\sup\limits_{0<\theta<1}f(x_1,\cdots,x_n;\theta)}{\sup\limits_{\theta=\theta_0}f(x_1,\cdots,x_n;\theta)}=\left(\frac{n_1}{\theta_0 n}\right)^{n_1}\left(\frac{n-n_1}{(1-\theta_0)\,n}\right)^{n-n_1}$$

则

$$2\log\lambda(\boldsymbol{X})\xrightarrow{H_0}\chi^2(1)$$

拒绝域为

$$\left\{(X_1,\cdots,X_n):2\log\lambda(\boldsymbol{X})>\chi_1^2(\alpha)\right\}$$

（3）设 φ_1 为假设检验问题

$$H_0:\theta=\theta_0\leftrightarrow H_1':\theta=\theta_1$$

的 UMPT，这里 $\theta_1 > \theta_0$
则由 N-P 引理

$$\varphi_1(\boldsymbol{x}) = \begin{cases} 1 & f(x,\theta_1)/f(x,\theta_0) > c_1 \\ r_1 & f(x,\theta_1)/f(x,\theta_0) = c_1 \\ 0 & f(x,\theta_1)/f(x,\theta_0) < c_1 \end{cases}$$

同理对假设检验问题

$$H_0 : \theta = \theta_0 \leftrightarrow H_1'' : \theta = \theta_2$$

这里 $\theta_2 < \theta_0$
有 UMPT

$$\varphi_2(\boldsymbol{x}) = \begin{cases} 1 & f(x,\theta_2)/f(x,\theta_0) > c_2 \\ r_2 & f(x,\theta_2)/f(x,\theta_0) = c_2 \\ 0 & f(x,\theta_2)/f(x,\theta_0) < c_2 \end{cases}$$

另一方面

$$\begin{aligned} f(\vec{x},\theta_1)/f(\vec{x},\theta_0) &= \left(\frac{\theta_1}{\theta_0}\right)^{n_1} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-n_1} \\ &= \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n} \quad \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{n_1} \end{aligned}$$

关于 n_1 单调递增
则 $\varphi_1(\boldsymbol{x})$ 可以改写为

$$\varphi_1(\boldsymbol{x}) = \begin{cases} 1 & n_1 > c_1 \\ r_1 & n_1 = c_1 \\ 0 & n_1 < c_1 \end{cases}$$

同理

$$f(\vec{x},\theta_2)/f(\vec{x},\theta_0) = \left(\frac{1-\theta_2}{1-\theta_0}\right)^{n} \cdot \left(\frac{\theta_2-\theta_2\theta_0}{\theta_0-\theta_0\theta_2}\right)^{n_1}$$

关于 n_1 单调递减
则 $\varphi_1(\boldsymbol{x})$ 可以改写为

$$\varphi_2(\boldsymbol{x}) = \begin{cases} 1 & n_1 < c_2 \\ r_2 & n_1 = c_2 \\ 0 & n_1 > c_2 \end{cases}$$

如果假设检验问题

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$$

的 UMPT 存在，记为 $\varphi_0(\boldsymbol{x})$
则由 N-P 引理的唯一性

$$\begin{aligned} \theta_1 > \theta_0 \text{ 时} \quad \varphi_1 = \varphi_0 \quad a.e. \\ \theta_2 < \theta_0 \text{ 时} \quad \varphi_2 = \varphi_0 \quad a.e. \end{aligned}$$

但这与 φ_1, φ_2 的形式矛盾，也就不存在假设检验问题 $H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0$ 的 UMPT

附表：$u_{0.025} = 1.960, u_{0.05} = 1.645, \chi^2_{899}(0.025) = 984, \chi^2_{899}(0.975) = 817.8$.