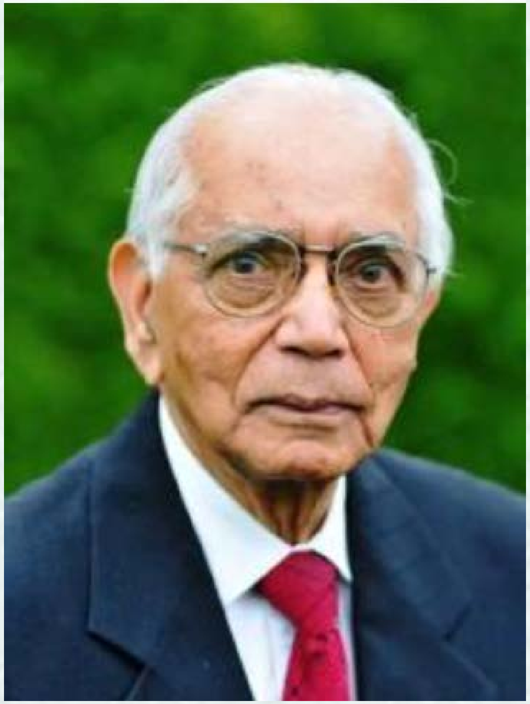
第三节　克拉默-拉奥（Cramer-Rao）不等式

6.3.2　Cramer-Rao 不等式

设$\xi_1,\xi_2,\cdots,\xi_n$为取自具有概率函数$f(x;\theta),\theta\in\Theta=\{\theta:a<\theta<b\}$的母体的一个子样，其中$a,b$为已知常数，且可设$a=-\infty,b=+\infty$.



C.R.Rao (1920-2023)

统计学大师、印度裔美国数学家，师从现代统计学的奠基人罗纳德·费希尔（**Ronald Aylmer Fisher**）

在终极的分析中，一切知识都是历史；

在抽象的意义下，一切科学都是数学；

在理性的世界里，所有的判断都是统计学。

设$\eta = u(\xi_1, \xi_2, \cdots, \xi_n)$是$g(\theta)$的一个无偏估计,且满足 正则条件:

(1)集合$\{x : f(x;\theta) > 0\}$与θ无关;

(2)$g'(\theta)$与$\dfrac{\partial f(x;\theta)}{\partial \theta}$存在,且对一切$\theta \in \Theta$,

$$\frac{\partial}{\partial \theta} \int f(x;\theta) dx = \int \frac{\partial f(x;\theta)}{\partial \theta} dx$$

$$\frac{\partial}{\partial \theta} \int \cdots \int u(x_1, x_2, \cdots, x_n) f(x_1;\theta) \cdots f(x_n;\theta) dx_1 \cdots dx_n$$

$$= \int \cdots \int u(x_1, x_2, \cdots, x_n) \frac{\partial}{\partial \theta} [\prod_{i=1}^{n} f(x_i;\theta)] dx_1 \cdots dx_n$$

(3) 令: $I(\boldsymbol{\theta})=E_{\theta}\left(\dfrac{\partial \log f(\xi;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^{2}>\mathbf{0}$ ---信息量

则有: $D_{\theta}\eta \geq \dfrac{[g'(\theta)]^2}{nI(\theta)}$　　当$g(\theta)=\theta$时,即为: $D_{\theta}\eta \geq \dfrac{1}{nI(\theta)}$

此不等式证明了在某些条件下,无偏估计量$\hat{\theta}$的方差具有一个正的下界.

$\underline{\underline{pr}}$:主要应用 $Cauchy-Schwarz$ 不等式(P_{157})

$$[E(\zeta-E\zeta)(\eta-E\eta)]^2\le D\zeta\cdot D\eta$$

$\eta=u(\xi_1,\xi_2,\cdots,\xi_n)$ 是 $g(\theta)$ 的一个无偏估计,即 $E\eta=g(\theta)$

且由数学期望的定义:

联合概率密度

$$g(\theta)=E\eta=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}u(x_1,\cdots,x_n)\underline{f(x_1;\theta)\cdots f(x_n;\theta)}dx_1\cdots dx_n$$

又设 $\zeta=\sum_{i=1}^{n}\dfrac{\partial\log f(\xi_i;\theta)}{\partial\theta}$

为了使用 $Cauchy-Schwarz$ 不等式,要计算出

$E\zeta, D\zeta$ 和 $E(\zeta-E\zeta)(\eta-E\eta)$

$$E[\frac{\partial \log f(\xi_i;\theta)}{\partial \theta}] = \int_{-\infty}^{\infty} \frac{\partial \log f(x_i;\theta)}{\partial \theta} f(x_i;\theta) dx_i = \int_{-\infty}^{\infty} \frac{\partial f(x_i;\theta)}{\partial \theta} dx_i$$

$$= \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(x_i;\theta) dx_i = 0$$

$$E\zeta = \sum_{i=1}^{n} E[\frac{\partial \log f(\xi_i;\theta)}{\partial \theta}] \qquad = 0$$

$$D\zeta = \sum_{i=1}^{n} D[\frac{\partial \log f(\xi_i;\theta)}{\partial \theta}] = nD[\frac{\partial \log f(\xi;\theta)}{\partial \theta}]$$

$$= n[E(\frac{\partial \log f(\xi;\theta)}{\partial \theta})^2 - (E\frac{\partial \log f(\xi;\theta)}{\partial \theta})^2]$$

而$E[\frac{\partial \log f(\xi;\theta)}{\partial \theta}] = 0$

$$= nE(\frac{\partial \log f(\xi;\theta)}{\partial \theta})^2 = nI(\theta)$$

下面考虑 $E(\zeta - E\zeta)(\eta - E\eta)$.

注意到 $E(\zeta - E\zeta)(\eta - E\eta) = E\zeta(\eta - E\eta) = E\zeta\eta$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta} \right] u(x_1, \cdots, x_n)$$

$$\cdot f(x_1;\theta) \cdots \cdot f(x_n;\theta) dx_1 \cdots dx_n$$

$$= g'(\theta)$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}u(x_1,\cdots,x_n)[\sum_{i=1}^{n}\frac{\partial\log f(x_i;\theta)}{\partial\theta}]\cdot f(x_1;\theta)\cdots f(x_n;\theta)dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}u(x_1,\cdots,x_n)[\sum_{i=1}^{n}\frac{1}{f(x_i;\theta)}\frac{\partial f(x_i;\theta)}{\partial\theta}]\cdot f(x_1;\theta)\cdots f(x_n;\theta)dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}u(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}[f(x_1;\theta)\cdots f(x_n;\theta)]dx_1\cdots dx_n$$

$$\overset{(2)}{=}\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}u(x_1,\cdots,x_n)f(x_1;\theta)\cdots f(x_n;\theta)dx_1\cdots dx_n=g'(\theta)$$

于是 $E(\zeta-E\zeta)(\eta-E\eta)=g'(\theta).$

所以,由不等式 $:[g'(\theta)]^2\le D\zeta\cdot D\eta=nI(\theta)\cdot D\eta$ 结论得证.

即 $:D_\theta\eta\ge\dfrac{[g'(\theta)]^2}{nI(\theta)}$

$P\{\zeta - E\zeta = K_\theta(\eta - E\eta)\} = 1$时，

不等式中等号成立(P_{157})

即：$\sum_{i=1}^{n} \frac{\partial \log f(\xi_i;\theta)}{\partial \theta} = K_\theta(\eta - g(\theta))$以概率1成立，不等式中等号成立

注：

1.满足正则条件的估计量称为正规估计.

2.$Rao-Cramer$不等式的下界仅是正规无偏估计类的方差下界.

3.相关定义

定义6.4:若θ的一个无偏估计$\hat{\theta}$使$Cramer-Rao$不等式中等式

$$D(\hat{\theta})=\frac{1}{nE[(\frac{\partial \log f(\xi;\theta)}{\partial \theta})^2]}=\frac{1}{nI(\theta)}$$

成立,则称$\hat{\theta}$为θ的有效估计.

定义6.5: $\boldsymbol{e}=\dfrac{\dfrac{1}{\boldsymbol{nI(\theta)}}}{\boldsymbol{D(\hat{\theta}_1)}}$为无偏估计$\hat{\theta}_1$的有效率.

定义6.6 当$\boldsymbol{n}\to\infty$,若$\hat{\theta}$的$\boldsymbol{e}\to 1$,则称$\hat{\theta}$为θ的渐近有效估计.

极大似然估计的渐近正态性

定理6.1（书P277）

在总体分布满足一定条件的情况下，存在具有相合性和渐近正态性的极大似然估计量$\hat{\theta}_n$，且

$$\hat{\theta}_n \sim N\left(\theta, \frac{1}{nI(\theta)}\right), \quad 当 n \to \infty 时,$$

即$\hat{\theta}_n$亦是θ的渐近无偏估计和渐近有效估计.

6.3.3 简单应用

例6. 3. 1: 假设$\xi \sim f(x;p)=\begin{cases} p^x(1-p)^{1-x} & x=0,1 \\ 0 & 其他 \end{cases}$ $0<p<1$,

试证: $\overline{\xi}$是p的有效估计.

证明: 显然, $\bar{\xi}$是p的无偏估计。

$$\frac{\partial \ln f(x;p)}{\partial p}=\frac{\partial}{\partial p}[x\ln p+(1-x)\ln(1-p)]=\frac{x}{p}-\frac{1-x}{1-p}$$

$$I(p)=E\left(\frac{\partial \ln f(\xi;p)}{\partial p}\right)^2=\sum_{x=0,1}\left(\frac{x}{p}-\frac{1-x}{1-p}\right)^2 p^x(1-p)$$

$$=\frac{1}{p(1-p)}$$

$$D(\hat{p}) = D(\bar{\xi}) = \frac{D\xi}{n} = \frac{p(1-p)}{n} = \frac{1}{nI(p)}.$$

故$\bar{\xi}$是p的有效估计。

例6.3.2:

设 $\xi_1,\xi_2,\cdots,\xi_n$ 是取自具有下列泊松分布的一个子样

$$f(x;\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!}, & x=0,1,2,\cdots \\ 0 & \text{其他}\end{cases}$$

,试证:$\bar{\xi}=\dfrac{1}{n}\sum_{i=1}^{n}\xi_i$ 是 λ 的有效估计.

证明:显然,$\bar{\xi}$是λ的无偏估计。

在$x=0,1,2,\cdots$ 上,

$$\frac{\partial \ln f(x;\lambda)}{\partial\lambda}=\frac{\partial}{\partial\lambda}[-\lambda+x\ln\lambda-\ln x!]=-1+\frac{\bar{x}}{\lambda}$$

$$I(\lambda)=E\left(\frac{\partial \ln f(\xi;\lambda)}{\partial\lambda}\right)^2=E\left(-1+\frac{\xi}{\lambda}\right)^2=\frac{1}{\lambda}$$

则

$$D(\hat{\lambda})=D(\bar{\xi})=\frac{\lambda}{n}=D(\bar{\xi})=\frac{1}{nI(\lambda)}.$$

例6.3.3:

设 $\xi_1,\xi_2,\cdots,\xi_n$ 是取自正态母体N(μ,σ^2)的一个子样，

试证:$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 是 μ的有效估计.

证明:因为$f(x;\mu)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad(-\infty<x<\infty)$

$$\ln f(x;\mu)=\ln[\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}]$$

$$=-\ln(\sigma\sqrt{2\pi})-\frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\mu}\ln f(x;\mu)=\frac{(x-\mu)}{\sigma^2}$$

$$I(\mu)=E\left(\frac{(\xi-\mu)^2}{\sigma^4}\right)=\frac{1}{\sigma^4}E(\xi-\mu)^2=\frac{\sigma^2}{\sigma^4}=\frac{1}{\sigma^2}$$

若$\hat{\mu}=\overline{\xi}$,则$D\overline{\xi}=\dfrac{\sigma^2}{n}=\dfrac{1}{nI(\mu)}$

所以,$\hat{\mu}=\overline{\xi}$是μ的有效估计.

2.性质(P_{283})

若 $\frac{\partial}{\partial\theta}\int\frac{\partial f(x;\theta)}{\partial\theta}dx=\int\frac{\partial^2 f(x;\theta)}{\partial\theta^2}dx$,则$I(\theta)=-E[\frac{\partial^2 \log f(\xi;\theta)}{\partial\theta^2}]$

证明见PDF文件

备用 设$\xi_1,\xi_2,\cdots,\xi_n$为取自正态母体$N(\mu,\sigma^2)$的一个子样.试证：

(1)$\hat{\mu}=\overline{\xi}$是μ的一个有效估计;

(2)若μ已知,则$S_\mu^2=\dfrac{1}{n}\sum_{i=1}^{n}(\xi_i-\mu)^2$是$\sigma^2$的有效估计;

若μ未知,则$S_n^{*2}=\dfrac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\overline{\xi})^2$不是$\sigma^2$的有效估计.

(2)若μ已知，$S_\mu^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \mu)^2$

$\because \xi_i \sim N(\mu, \sigma^2) \therefore \frac{\xi_i - \mu}{\sigma} \sim N(0,1)$

$\sum_{i=1}^{n}\left(\frac{\xi_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$　　即$\frac{nS_\mu^2}{\sigma^2} \sim \chi^2(n)$

所以，　$ES_\mu^2 = \sigma^2$　　　$D\left(\frac{nS_\mu^2}{\sigma^2}\right) = 2n$

则$D(S_\mu^2) = \frac{2\sigma^4}{n}$

同样有, $\ln f(x;\sigma^2) = -\ln(\sqrt{2\pi\sigma^2}) - \dfrac{(x-\mu)^2}{2\sigma^2}$

则 $\dfrac{\partial}{\partial\sigma^2}\ln f(x;\sigma^2) = -\dfrac{1}{2\sigma^2} + \dfrac{(x-\mu)^2}{2\sigma^4}$

则 $\dfrac{\partial^2}{\partial(\sigma^2)^2}\ln f(x;\sigma^2) = \dfrac{1}{2\sigma^4} - \dfrac{(x-\mu)^2}{\sigma^6}$

由性质, σ^2的信息量: $I(\sigma^2) = -E\left[\dfrac{\partial^2 \ln f(\xi;\sigma^2)}{\partial(\sigma^2)^2}\right]$

$$= -E\left[\frac{1}{2\sigma^4} - \frac{(\xi-\mu)^2}{\sigma^6}\right]$$

$$= -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}E(\xi-\mu)^2 = \frac{1}{2\sigma^4}$$

所以，$DS_\mu^2 = \dfrac{1}{nI(\sigma^2)} = \dfrac{2\sigma^4}{n}$

若μ未知，则可知S_n^{*2}是σ^2的一个无偏估计.

由定理，$\dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$

$$D\left(\frac{(n-1)S_n^{*2}}{\sigma^2}\right) = 2(n-1)$$

则$D(S_n^{*2}) = \dfrac{2\sigma^4}{n-1} \neq \dfrac{2\sigma^4}{n}$ 所以，S_n^{*2}不是σ^2的有效估计.

$e = \dfrac{\dfrac{1}{nI(\sigma^2)}}{D(\hat{\sigma}^2)} = \dfrac{\dfrac{2\sigma^4}{n}}{\dfrac{2\sigma^4}{n-1}} = \dfrac{n-1}{n} \xrightarrow{n\to\infty} 1$ 所以，S_n^{*2}是σ^2的渐近有效估计.

作业： p.309 6.38 (1)，6.39（6.37）.

思考：

设 $\xi_1,\xi_2,\cdots,\xi_n$ 是取自具有下列指数分布的一个子样

$$f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x\geq 0\\ 0 & 其他\end{cases},试证:\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i 是 \theta 的有效估计.$$