§2.4　有效估计量

瑞典统计学家克拉美和印度统计学家劳分别于1945年和1946年对单参数正则分布族证明了一个重要不等式，这个不等式给出了可估参数g(θ)的无偏估计方差的下界．这个下界可用来评价一个无偏估计的优劣.

定理2.3.1 (Cramer – Rao不等式)　设总体X的分布密度为$f(x;\theta),\theta\in\Theta, X_1, X_2,\cdots, X_n$是来自总体$X$的样本, $T=T(X_1,\cdots,X_n)$是参数$g(\theta)$的一个无偏估计量，且满足条件：

(1)参数空间Θ是直线上的某个开区间;

(2)支撑集合$\{x\colon\ f(x;\theta)>0\}$不依赖于$\theta$;

(3)对一切$\theta\in\Theta, \dfrac{\partial f(x;\theta)}{\partial\theta}$存在，且有

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int_{-\infty}^{+\infty} f(x;\theta)\,\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{\partial f(x;\theta)}{\partial\theta}\,\mathrm{d}x$$

(4)下列数学期望存在，且

$$0 < I(\theta) \triangleq E[\frac{\partial}{\partial\theta}\ln f(X;\theta)]^2 < +\infty$$

$I(\theta)$称为**Fisher信息量.**

满足上述的分布称为**C－R正则分布族**。

(5)如果$g'(\theta)$存在,且

$$\frac{\partial}{\partial\theta}\int\cdots\int T(x_1,\cdots,x_2)f(x_1;\theta)\cdots f(x_n;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n$$

$$=\int\cdots\int T(x_1,\cdots,x_2)\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n}f(x_i;\theta)\right]\mathrm{d}x_1\cdots\mathrm{d}x_n$$

则

$$D(T) \geq \frac{[g'(\theta)]^2}{nI(\theta)}.$$

特别$g(\theta)=\theta$ 时, $D(T)\geq 1/nI(\theta)$. 等号成立的充分必要条件是存在一个函数$K(\theta)$, 使得

$$\sum_{i=1}^{n}\frac{\partial \ln f(x_i;\theta)}{\partial\theta}=K(\theta)\left[T-g(\theta)\right]$$

以概率1成立.

注：(1) 定理对离散型总体也适用；

(2) 不等式的右端称为克拉美－罗下界；

(3) 如果 $\dfrac{\partial}{\partial\theta}\int\dfrac{\partial}{\partial\theta}f(x;\theta)\mathrm{d}x=\int\dfrac{\partial^2 f(x;\theta)}{\partial\theta^2}\mathrm{d}x$，则

$$I(\theta)=-E\left(\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right).$$

事实上 $\dfrac{\partial^2}{\partial\theta^2}\ln f(X;\theta)=\dfrac{\partial}{\partial\theta}\left(\dfrac{f'_\theta(X;\theta)}{f(X;\theta)}\right)$

$$=\frac{f''_\theta(X;\theta)f(X;\theta)-\left[f'_\theta(X;\theta)\right]^2}{f^2(X;\theta)}=\frac{f''_\theta(X;\theta)}{f(X;\theta)}-\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2$$

两边取数学期望即可.

证明： 不妨设$D(T)<+\infty$. 因为.

$$1=\int_{-\infty}^{+\infty} f(x_i;\theta)dx_i,$$

$$g(\theta)=\int\cdots\int T(x_1,\cdots,x_2)f(x_1;\theta)\cdots f(x_n;\theta)dx_1\cdots dx_n$$

上式两边对θ求导可得

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

$$=\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}\left[\ln f(x_i;\theta)\right]f(x_i;\theta)dx_i$$

$$=E\left[\frac{\partial}{\partial\theta}\left[\ln f(X_i;\theta)\right]\right]$$

$$g'(\theta)=\int\cdots\int T(x_1,\cdots,x_2)\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n}f(x_i;\theta)\right]\mathrm{d}x_1\cdots\mathrm{d}x_n$$

$$=\int\cdots\int T(x_1,\cdots,x_2)\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right.$$

$$\left.\prod_{i=1}^{n}f(x_i;\theta)\right]\mathrm{d}x_1\cdots\mathrm{d}x_n$$

$$=E(TY).$$

其中$Y=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(X_i;\theta),E\left[Y\right]=0.\ \ D(Y)=nI(\theta).$

由柯西—许瓦兹不等式得

$$\left[g'(\theta)\right]^2=\left[E(TY)\right]^2=\left[E(T-g(\theta))Y\right]^2$$

$$\le E\left[\left(T-g(\theta)\right)\right]^2 EY^2 = D(T)D(Y)$$

$$= nI(\theta)D(T)$$

所以

$$D(T) \ge \frac{[g'(\theta)]^2}{nI(\theta)}.$$

等号成立的充要条件是存在一个函数$K(\theta)$,使得

$$\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\theta)}{\partial\theta} = K(\theta)\left[T-g(\theta)\right]$$

以概率1成立. $K(\theta)\ne 0$与样本无关.

若$g(\theta)$的无偏估计量$T=T(X_1,\cdots,X_n)$的方差到达$C-R$不等式的下界，则称T是$g(\theta)$的有效估计量.

注:(1)$C-R$不等式在充分正规的条件下给出了无偏估计方差的下界；在有些场合，这样的下界还是下确界，而且存在估计量，其方差达到了这个下确界. 此时，一定是最小方差无偏估计.

(2)有效估计量(若存在)一定是极大似然估计.

例如：总体$X\sim B(1,p), p\in\Theta=(0,1)$. $X_1,\cdots,X_n$是来自总体X的样本，验证$\bar{X}$是参数p的有效估计量.

我们知道：$E(\bar{X})=p, D(\bar{X})=n^{-1}D(X)=n^{-1}p(1-p)$.

因此，$\bar{X}$是参数p的有效估计量.

例1 设$X_1, X_2, \cdots, X_n$是来自总体$X \sim N(\mu,\sigma^2)$的一个样本，已知$\overline{X}$和S^2分别是μ和σ^2的无偏估计.

试证明(1)$\overline{X}$是μ的有效估计量. S^2不是σ^2的有效估计量. (2)当μ已知时，$\sum_{i=1}^{n}(X_i-\mu)^2/n$是$\sigma^2$的有效估计量.

证明：总体X的分布密度为

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

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

$$\frac{\partial}{\partial\mu}\left[\ln f(x;\mu,\sigma^2)\right]=\frac{x-\mu}{\sigma^2},\ \ I(\mu)=E(\frac{x-\mu}{\sigma^2})^2=\frac{1}{\sigma^2}.$$

由于 $E(\bar{X})=\mu,\qquad D(\bar{X})=\frac{\sigma^2}{n}=\frac{1}{nI(\mu)}$. 故$\bar{X}$是

参数μ的有效估计量.

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

$$\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(x;\mu,\sigma^2)=\frac{1}{2\sigma^4}-\frac{1}{\sigma^4}(\frac{x-\mu}{\sigma})^2$$

$$I(\sigma^2)=-E\left[\frac{\partial^2}{\partial(\sigma^2)^2}\ln f(X;\mu,\sigma^2)\right]=\frac{1}{2\sigma^4}$$

由 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，可得

$$E(S^2)=\sigma^2, \quad D(S^2)=\frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n} = \frac{1}{nI(\sigma^2)}.$$

因此S^2不是参数σ^2的有效估计量(是MVUE).

由 $\displaystyle\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$，可得

$$E\left(\sum_{i=1}^{n}(X_i-\mu)^2/n\right)=\sigma^2,$$

$$D\left(\sum_{i=1}^{n}(X_i-\mu)^2/n\right)=\frac{2\sigma^4}{n}=\frac{1}{nI(\sigma^2)}.$$

故当μ已知时，$\displaystyle\sum_{i=1}^{n}(X_i-\mu)^2/n$是$\sigma^2$的有效估计量.