数理统计

冯伟

数学科学学院

wfeng_323@buaa.edu.cn

上页 下页 返回

第二节 点估计量的优良性(续)

有效性

定义1 设$T_1(x_1,x_2,\ldots,x_n)$和$T_2(x_1,x_2,\ldots,x_n)$是参数θ的两个无偏估计量，如果对一切θ∈Θ都有

$$Var_\theta(T_1(x)) \le Var_\theta(T_2(x))$$

则称估计量T_1比T_2有效。

上页 下页 返回

例1 设总体$X \sim N(\mu,\sigma^2)$，$(X_1,X_2,\ldots,X_n)$是总体的容量为n的样本，μ已知。试证σ^2的MLE是较样本方差S^2更有效的估计量。

例2 设$(X_1,X_2,\ldots,X_n)$是任一总体的容量为n的样本，如果总体的方差存在，则样本均值与X_1都是总体均值μ的无偏估计，但样本均值是较X_1有效的μ的无偏估计。

上页 下页 返回

一致最小方差无偏估计

设统计模型为$\{P_\theta,\theta\in\Theta\}$，$q(\theta)$是可估参数，$U_q$是$q(\theta)$的无偏估计类，$q(\theta)$的一致最小方差无偏估计定义如下。

定义2 如果存在无偏估计$T(x)\in U_q$，使得对任意的$S(x)\in U_q$，有

$$Var_\theta(T(x))\le Var_\theta(S(x))$$

对所有的$\theta\in\Theta$都成立，则称$T(x)$为$q(\theta)$的

上页 下页 返回

一致最小方差无偏估计， 简称为UMVUE。

(Uniformly Minimum Variance Unbiased Estimate)

- 定义
- 存在问题
- 唯一性问题
- 如何判断
- 具体的求法

上页 下页 返回

一致最小方差线性无偏估计

- 限制：要求估计量是样本的线性函数
- 解法：转化为求条件极值问题求解

例3 估计总体X的均值的一致最小方差线性无偏估计

上页 下页 返回

零无偏定理

记$U=\{T:E_{\theta}(T)=\theta,E_{\theta}(T^2)<\infty$,对一切$\theta\in I\}$,

$U_0=\{V:E_{\theta}(V)=0,E_{\theta}(V^2)<\infty$,对一切$\theta\in I\}$.

定理1 设U和U_0分别如上定义,且非空,则$T_0\in U$是θ的一个一致最小方差无偏估计的充要条件是$E_{\theta}(VT_0)=0$,对一切$\theta\in I$和$V\in U_0$.

上页 下页 返回

例4 设$X \sim N(\mu,\sigma^2)$， μ,σ^2未知，求参数μ,σ^2的一致最小方差无偏估计。

解 设$V \in U_0$,则有

$$\int \ldots \int V \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}dx_1\cdots dx_n = 0$$

通过对μ求一次微分,可得样本均值是μ的一致最小方差无偏估计,再结合对μ的二次微分和对σ²的一次微分,可得样本方差是总体方差的一致最小方差无偏估计.

上页 下页 返回

信息不等式

在上面，我们知道如果UMVUE存在，则它在无偏估计类中是最好的，且其方差不可能是零，因为参数$q(\theta)$的方差为零的平凡估计不是无偏估计。那么，现在的问题是：

对$q(\theta)$的无偏估计类 U_q，在一定的条件下，

（1） 既然无偏估计的方差不是零，则必存在一个下界，这个下界到底是多少？

上页 下页 返回

（2）若UMVUE存在，那么它的方差是否可以达到这个下界？

问题（1）已由Cramer-Rao不等式（信息不等式）揭示；问题（2）不一定成立，我们举例予以阐述。

上页 下页 返回

为了使问题简化，我们仅讨单参数和连续总体情况。 对多参数及离散总体也有相应结论，可参看《高等数理统计学》（茆诗松），或《线性统计推断及应用》（C.R.Rao）。

上页 下页 返回

设分布族为$\{P_\theta, \theta \in \Theta\}$，密度函数为$p(x,\theta)$，$\Theta$为直线上的一个开区间。满足下述条件的分布族$\{P_\theta, \theta \in \Theta\}$称为 **Cramer-Rao正则族：**

（1）支撑$A = \{x : p(x,\theta) > 0\}$与$\theta$无关，且对任一$x \in A, \theta \in \Theta$，偏导数$\frac{\partial}{\partial \theta} \ln p(x,\theta)$存在。

（2）如果对所有$\theta \in \Theta$，$T(x)$是满足$E_\theta |T| < \infty$的任一统计量，则对$T(x)L(x,\theta)$，积分和微分可交换次序，即

上页 下页 返回

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x)L(x,\theta)dx_1\cdots dx_n$$

$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x)\frac{\partial}{\partial\theta}L(x,\theta)dx_1\cdots dx_n$$

这里的$L(x,\theta)=L(x_1,x_2,\ldots,x_n,\theta)$是联合概率密度.

（3）如果对所有$\theta\in\Theta$, $\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}p(x,\theta)dx=\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}p(x,\theta)dx$

当仅有（1）成立时，我们可以定义所谓的

Fisher 信息量（Fisher Information Number）

$$\boxed{I(\theta)=E\left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)^2}\quad (0\le I(\theta)\le\infty)$$

上页 下页 返回

如果$X_1, X_2, \cdots, X_n$是来自总体的样本，可以证明$I_1(\theta) = nI(\theta)$，其中$I_1(\theta) = E(\frac{\partial}{\partial\theta}\ln L(X,\theta))^2$.

例5 设总体分布是Poisson分布族，即

$$p(x,\theta) = \frac{\theta^x}{x!}e^{-\theta}, x = 0,1,\cdots.$$

则
$$\frac{\partial}{\partial\theta}\ln p(x,\theta) = \frac{x}{\theta} - 1,$$

因而
$$I(\theta) = E(\frac{x}{\theta} - 1)^2 = Var(\frac{x}{\theta}) = \frac{1}{\theta}.$$

故
$$I_1(\theta) = nI(\theta) = \frac{n}{\theta}$$

上页 下页 返回

定理2(Cramer-Rao or Information Inequality)

设$T(X)$是对所有$\theta\in\Theta$满足$Var_\theta(T(X))<+\infty$的统计量，记$\varphi(\theta)=E_\theta(T(X))$。如果分布族是Cramer-Rao正则族，且$0<I(\theta)<\infty$，则对所有的$\theta\in\Theta$，$\varphi(\theta)$是可微的，且

$$\boxed{Var_\theta(T(X))\geq\frac{(\varphi'(\theta))^2}{nI(\theta)}.}$$

上页 下页 返回

证明　由于对所有$\theta \in \Theta$，有

$$\varphi(\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} T(x) L(x,\theta) dx_1 \cdots dx_n$$

等式两边对求导可得

$$\varphi'(\theta) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} T(x) \frac{\partial}{\partial \theta} L(x,\theta) dx_1 \cdots dx_n$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} T(x) \frac{\partial}{\partial \theta} (\ln L(x,\theta)) L(x,\theta) dx_1 \cdots dx_n$$

$$= E\left(T(x) \frac{\partial}{\partial \theta} \ln L(x,\theta) \right).$$

上页　下页　返回

又因为对所有的 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, 有

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}L(x,\theta)dx_1\cdots dx_n=1$$

等式两边对求导可得

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}L(x,\theta)dx_1\cdots dx_n=0.$$

即就是 $$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial\ln L(x,\theta)}{\partial\theta}L(x,\theta)dx_1\cdots dx_n=0.$$

这样就有 $$E\left(\frac{\partial\ln L(x,\theta)}{\partial\theta}\right)=0.$$

上页 下页 返回

从而有 $$\varphi'(\theta) = E\left(T(x)\frac{\partial}{\partial\theta}\ln L(x,\theta) \right)$$

$$= Cov\left(T(x), \frac{\partial}{\partial\theta}\ln L(x,\theta) \right).$$

由协方差的性质，得

$$|\varphi'(\theta)| = \left| Cov\left(T(x), \frac{\partial}{\partial\theta}\ln L(x,\theta) \right) \right|$$

$$\le \sqrt{Var(T(X))}\sqrt{Var\left(\frac{\partial}{\partial\theta}\ln L(x,\theta) \right)}$$

而

$$Var\left(\frac{\partial \ln L(x,\theta)}{\partial\theta}\right)=E\left(\frac{\partial \ln L(x,\theta)}{\partial\theta}\right)^2=nI(\theta)$$

所以有 $|\varphi'(\theta)|\le\sqrt{Var(T(X))}\sqrt{nI(\theta)}$

即就是 $$Var_\theta(T(X))\ge\frac{(\varphi'(\theta))^2}{nI(\theta)}.$$

上页 下页 返回

定理应用于参数$q(\theta)$的无偏估计类U_q就有：

对参数$q(\theta)$的任一无偏估计$T(X)\in U_q$，有

$$\boxed{Var_\theta(T(X)) \geq \frac{(q'(\theta))^2}{nI(\theta)}.}$$

特别地，当$q(\theta)=\theta$时，对任一$T(X)\in U_\theta$，有

$$\boxed{Var_\theta(T(X)) \geq \frac{1}{nI(\theta)}.}$$

注意：（1）在以上不等式中

$$I_1(\theta) = nI(\theta)$$

其中$I(\theta) = E(\frac{\partial}{\partial\theta}\ln p(X_1,\theta))^2$，$p(x_1,\theta)$为总体的密度函数或分布率。

通常将$I(\theta)$ 看成一次观察所能获得的关于参数 θ的信息，即一个观测值X_1所含θ 的信息，那么$I_1(\theta)$ 就表示样本$X_1,\cdots,X_n$ 所含 θ的信息。

上页 下页 返回

（2） C－R(Cramer-Rao)定理中两个不等式的右端项是方差的下界。在实用中也常用 $I(\theta)$的另一等价表达式

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

上页 下页 返回

（3） 在将定理2应用于无偏估计类U_q时，一定要注意定理的条件是否满足。**Cramer**在1946年举例说明当定理的条件不满足时，存在这样的无偏估计，其方差小于信息不等式的下界。这个例子为：设$X_1, X_2, \cdots, X_n$是来自总体X的样本，X的密度函数为

$$p(x,\theta)=\begin{cases} e^{-(x-\theta)} & x>\theta \\ 0 & \text{□ □} \end{cases}.$$

取充分统计量$T(X)=X_{(1)}$作为参数θ的估计，通过取其数学期望可获得参数的无偏估计为

$$\hat{\theta}(X)=X_{(1)}-\frac{1}{n},$$

则有 $$Var_{\theta}(\hat{\theta}(X))=\frac{1}{n^2}<\frac{1}{n}=\frac{1}{nI(\theta)}\quad(n>1).$$

其具体证明过程课后自己完成。

上页 下页 返回

对无偏估计类而言，既然信息不等式给出了方差的下界，那么UMVUE方差是否一定取得这个下界？下述例子说明不一定。

例6 设$X_1, X_2, \cdots, X_n$来自正态总体$N(\mu,1)$的一个简单样本。试求参数μ^2的UMVUE，并证明其方差大于信息不等式的下界。

如果参数$q(\theta)$的无偏估计$\hat{q}(X)$的方差取得信息不等式的下界，即

$$Var(\hat{q}(X)) = \frac{(q'(\theta))^2}{I(\theta)},$$

则$\hat{q}(X)$必是参数$q(\theta)$的UMVUE。

例7 设$X_1,X_2,\cdots,X_n$来自正态总体$N(0,\sigma^2)$的一个简单样本。试求参数σ^2的UMVUE。

解 由于

$$I_1(\sigma^2)=-E\left(\frac{\partial^2}{\partial(\sigma^2)^2}\ln p(X_1,\sigma^2)\right)$$

$$=-E\left(\frac{\partial^2}{\partial(\sigma^2)^2}\ln\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right)$$

$$=-E\left(\frac{1}{2(\sigma^2)^2}-\frac{x^2}{(\sigma^2)^3}\right)=\frac{1}{2(\sigma^2)^2},$$

上页 下页 返回

从而 $I(\sigma^2)=\dfrac{n}{2(\sigma^2)^2}$。由信息不等式知，对任一无偏估计 $\hat{\sigma}^2(X)\in U_{\sigma^2}$，有

$$Var(\hat{\sigma}^2(X))\geq\frac{((\sigma^2)')^2}{I(\theta)}=\frac{2(\sigma^2)^2}{n}.$$

若取 $\hat{\sigma}^2(X)=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2$，由 $\dfrac{X_i^2}{\sigma^2}$ 服从 $\chi^2(1)$ 可知

$$\frac{n\hat{\sigma}^2(X)}{\sigma^2}=\sum_{i=1}^{n}\frac{X_i^2}{\sigma^2}\sim\chi^2(n)$$

所以 $E\left(\dfrac{n\hat{\sigma}^2(X)}{\sigma^2}\right)=n,\ Var\left(\dfrac{n\hat{\sigma}^2(X)}{\sigma^2}\right)=2n,$

即 $E(\hat{\sigma}^2(X))=\sigma^2$， $Var(\hat{\sigma}^2(X))=\dfrac{2(\sigma^2)^2}{n}$.

故 $Var(\hat{\sigma}^2(X))=\dfrac{2(\sigma^2)^2}{n}=\dfrac{((\sigma^2)')^2}{I(\theta)}$,

从而$\hat{\sigma}^2(X)=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2$是参数$\sigma^2$的UMVUE。

定义3 设分布族$\{P_\theta,\theta\in\Theta\}$是Cramer－Rao正则族，$q(\theta)$是可估参数，如果存在某无偏估计$\hat{q}(X)\in U_q$，其方差达到信息不等式的下界，即

上页 下页 返回

$$Var(\hat{q}(X)) = \frac{(q'(\theta))^2}{I(\theta)},$$

则称 $\hat{q}(X)$ 为 $q(\theta)$ 的有效估计(Efficient Estimate)。

定义4 对参数 $q(\theta)$ 的任一无偏估计 $\hat{q}(X) \in U_q$,

令 $$e(\hat{q}(X)) = \frac{(q'(\theta))^2}{I(\theta)} \Big/ Var(\hat{q}(X)),$$

则称 $e(\hat{q}(X))$ 为估计 $q(\theta)$ 的有效率(Efficiency)。

显然 $0 \le e(\hat{q}(X)) \le 1,$

因此,有效估计乃是有效率为1的无偏估计。

上页 下页 返回

定义5　设$\{T_n(X)\}$是参数$q(\theta)$估计序列，如果对所有的$\theta\in\Theta$，都有

$$\lim_{n\to\infty}E_\theta(T_n(X))=q(\theta),$$

则称$T_n(X)$为参数$q(\theta)$的渐近无偏估计。（**Asymptotic Unbiased Estimate**）

例如对正态总体$N(\mu,\sigma^2)$，我们知道$\hat{\sigma}_n^2$是总体方差σ^2的有偏估计，且

$$E(\hat{\sigma}_n^2)=\frac{n-1}{n}\sigma^2.$$

上页　下页　返回

这样有 $\lim\limits_{n\to\infty} E(\hat{\sigma}_n^2) = \lim\limits_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2,$

故 $\hat{\sigma}_n^2$ 是总体方差 σ^2 的渐近无偏估计。

定义6 设 $q(\theta)$ 是可估参数，如果存在无偏估计序列 $\hat{q}_n(X) \in U_q$，使得

$$\lim_{n\to\infty} e(\hat{q}(X)) = \lim_{n\to\infty} \frac{(q'(\theta))^2}{I(\theta)} \Big/ Var(\hat{q}(X)) = 1$$

成立，则称 $\hat{q}_n(X)$ 为 $q(\theta)$ 的渐近有效估计。

(Asymptotic Efficient Estimate)

上页 下页 返回

例如$X_1,\cdots,X_n$是来自正态总体$N(0,\sigma^2)$的样本，

由例5知方差σ^2的有效估计是$\frac{1}{n}\sum_{i=1}^{n}X_i^2$。

由于

$$\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1)$$

所以

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right)=n-1,$$

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right)=2(n-1).$$

即 $E(S^2)=\sigma^2$, $Var(S^2)=\frac{2(\sigma^2)^2}{n-1}$.

而Cramer-Rao下界为

$$\frac{1}{I(\sigma^2)}=\frac{2(\sigma^2)^2}{n},$$

这说明无偏估计S^2既不是σ^2的UMVUE，也不是有效估计。但是

$$\lim_{n\to\infty} e(S^2)=\lim_{n\to\infty}\frac{\frac{1}{I(\sigma^2)}}{Var(S^2)}=\lim_{n\to\infty}\frac{\frac{2(\sigma^2)^2}{n}}{\frac{2(\sigma^2)^2}{n-1}}=1,$$

故样本方差S^2是σ^2的渐近有效估计。

需要说明的是当UMVUE的方差较大时，方差小的有偏估计也不失为一个好的估计。

作业: 注意作业会做的 前提下可自行减半

1.设$X \sim N(\mu,\sigma^2)$， μ,σ^2未知，求参数σ^2的一致最小方差无偏估计。（用零无偏定理）

2. 设$X_1, X_2, \cdots, X_n$来自正态总体$N(\mu,1)$的一个简单样本。试求参数 μ^2的UMVUE，并证明其方差大于信息不等式的下界。

3. 教材87页28,29,32,41-43,45

练习（可以不交） 87页38,39