第二节　估计量的评价标准

- 一、问题的提出
- 二、无偏性
- 三、有效性
- 四、相合性

一、问题的提出

从前一节可以看到, 对于同一个参数, 用不同的估计方法求出的估计量可能不相同. 而且, 很明显, 原则上任何统计量都可以作为未知参数的估计量.

问题

(1) 对于同一个参数究竟采用哪一个估计量好?

(2) 评价估计量的标准是什么?

下面介绍几个常用标准.

二、无偏性

若 $X_1, X_2, \cdots, X_n$ 为总体 X 的一个样本，

$\theta \in \Theta$ 是包含在总体 X 的分布中的待估参数,

(Θ是 θ 的取值范围)

定义6.2 设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是参数 θ 的估计量,

若 $$E(\hat{\theta}) = \theta,$$

则称 $\hat{\theta}$ 是 θ 的 无偏估计(量).

若 $$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta,$$

则称 $\hat{\theta}$ 是 θ 的渐近无偏估计(量).

例 证明：$S_n^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$是$\sigma^2=D(X)$

的渐近无偏估计量； S_n^{*2}是σ^2的无偏估计量.

证 $S_n^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\bar{X}^2$

$$\because\quad E(S_n^2)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)-E(\bar{X}^2)=E(X^2)-E(\bar{X}^2)$$

$$=[D(X)+(EX)^2]-[D(\bar{X})+(E\bar{X})^2]$$

$$=[D(X)+\cancel{(EX)^2}]-[\frac{1}{n}D(X)+\cancel{(EX)^2}]$$

$$E(S_n^2)=\frac{n-1}{n}D(X)=\frac{n-1}{n}\sigma^2$$

$\because\quad \lim\limits_{n\to\infty}E(S_n^2)=\sigma^2$

$\therefore\quad S_n^2$是$\sigma^2=D(X)$的渐近无偏估计量

又$\because\quad E(S_n^{*2})=E\left(\frac{n}{n-1}S_n^2\right)=\frac{n}{n-1}E(S_n^2)$

$$=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2.$$

$\therefore\quad S_n^{*2}$是σ^2的无偏估计量.

例3 设总体X的方差$D(X)$存在，且 $D(X)>0$, $(X_1, X_2, \cdots, X_n)$ 为来自总体X的样本，试选择适当的常数C，使得

$$C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$$

为$D(X)$的无偏估计.

分析 需选择C，使

$$E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]=D(X)$$

解 $\because \quad E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]=C\sum_{i=1}^{n-1}E(X_{i+1}-X_i)^2$

$$=C\sum_{i=1}^{n-1}\{D(X_{i+1}-X_i)+[E(X_{i+1}-X_i)]^2\}$$

而$X_1, X_2, \cdots, X_n$相互独立，且与X 同分布

$\therefore \quad E(X_i)=E(X), \quad D(X_i)=D(X) \quad (i=1,2,\cdots,n)$

$$D(X_{i+1}-X_i)=D(X_{i+1})+D(X_i)=2D(X)$$

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

$$\therefore \quad E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]$$

$$=C\sum_{i=1}^{n-1}\{D(X_{i+1}-X_i)+[E(X_{i+1}-X_i)]^2\}$$

$$=C\sum_{i=1}^{n-1}2D(X)=C\cdot 2(n-1)D(X)$$

依题意，要求： $E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]=D(X)$

即 $C\cdot 2(n-1)D(X)=D(X)$

$\because \quad D(X)>0 \quad \therefore \quad C=\dfrac{1}{2(n-1)}.$

注 一般地，一个参数θ的无偏估计量不唯一.

如：设样本$(X_1, X_2, \cdots, X_n)$来自总体X，$E(X)=\mu$,

则$\overline{X}$是μ的无偏估计. 此外，

$$\sum_{i=1}^{n} C_i X_i \qquad (\sum_{i=1}^{n} C_i = 1)$$

也均是μ的无偏估计.

问题： 对于同一个参数的多个无偏估计量，

如何评价它们的优劣？

三、有效性

比较参数θ的两个无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$, 如果在样本容量n相同的情况下, $\hat{\theta}_1$的观察值较$\hat{\theta}_2$更密集在真值θ的附近, 则认为$\hat{\theta}_1$ 较 $\hat{\theta}_2$ 理想.

换句话说，对参数θ的无偏估计量$\hat{\theta}$关于θ的波动越小，即方差

$$D(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 \qquad (E(\hat{\theta}) = \theta)$$

$$= E(\hat{\theta} - \theta)^2$$

越小越好.

1. $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效

定义6.3 设 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$

均是θ的无偏估计量，若

$$D(\hat{\theta}_1)<D(\hat{\theta}_2),$$

则称 $\hat{\theta}_1$比 $\hat{\theta}_2$有效.

例4 设$E(X)=\mu,D(X)=\sigma^2>0$存在，(X_1,X_2)是来自总体X的样本，问：下列三个对μ的无偏估计量哪一个最有效？

$$\hat{\mu}_1 = \frac{3}{4}X_1 + \frac{1}{4}X_2,$$

$$\hat{\mu}_2 = \frac{1}{2}X_1 + \frac{1}{2}X_2,$$

$$\hat{\mu}_3 = \frac{2}{3}X_1 + \frac{1}{3}X_2.$$

注 一般地，在μ的无偏估计量

$$\sum_{i=1}^{n} C_i X_i \quad (\sum_{i=1}^{n} C_i = 1)$$

中，$\overline{X}$最有效.

解 $D(\hat{\mu}_1) = (\frac{9}{16} + \frac{1}{16})\sigma^2 = \frac{5}{8}\sigma^2,$

$D(\hat{\mu}_2) = \frac{1}{2}\sigma^2, D(\hat{\mu}_3) = \frac{5}{9}\sigma^2,$

可用求条件极值的拉格朗日乘数法证明

$\because \quad D(\hat{\mu}_2) < D(\hat{\mu}_3) < D(\hat{\mu}_1) \qquad \therefore \ \hat{\mu}_2$最有效.

例5 设总体$X \sim U[0,\theta]$，参数$\theta > 0$，$(X_1, X_2, \cdots, X_n)$是来自总体X样本，

(1) 试证明：θ的矩估计量$\hat{\theta}_1 = 2\overline{X}$和修正的最大似然估计量$\hat{\theta}_2 = \dfrac{n+1}{n} X_{(n)}$均是$\theta$的无偏估计；

(2) 问：$\hat{\theta}_1$和$\hat{\theta}_2$哪一个更有效？

(1) 证 $\because$ $E(\hat{\theta}_1) = E(2\overline{X}) = 2E(\overline{X}) = 2E(X)$

$$= 2 \times \frac{\theta}{2} = \theta,$$

$\therefore$ $2\overline{X}$是θ的无偏估计量.

$$\because \quad X \sim U[0,\theta] \qquad \therefore \quad p(x)=\begin{cases}\dfrac{1}{\theta}, & x\in[0,\theta]\\ 0, & \text{其它}\end{cases}$$

$$F(x)=\int_{-\infty}^{x} p(t)\,\mathrm{d}t=\begin{cases}0, & x<0\\ \dfrac{x}{\theta}, & 0\le x\le\theta\\ 1, & x>\theta\end{cases}$$

$\therefore\ X_{(n)}=\max(X_1,X_2,\cdots,X_n)$的概率密度为

$$p_{X_{(n)}}(x)=F'_{X_{(n)}}(x)=nF^{n-1}(x)\cdot p(x)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n}, & 0\le x\le\theta\\ 0, & \text{其它}\end{cases}$$

$$E(\hat{\theta}_2)=\frac{n+1}{n}E(X_{(n)})\overset{?}{=}\theta$$

$$\because \quad E(X_{(n)})=\int_{-\infty}^{+\infty} xp_{X_{(n)}}(x)\,\mathrm{d}x=\int_0^{\theta} x\cdot\frac{nx^{n-1}}{\theta^n}\,\mathrm{d}x$$

$$=\frac{n}{n+1}\theta,$$

$$\therefore \quad E\left(\frac{n+1}{n}X_{(n)}\right)=\theta,$$

即$\hat{\theta}_2=\dfrac{n}{n+1}X_{(n)}$也是$\theta$的无偏估计量.

(2) 问：$\hat{\theta}_1 = 2\overline{X}$ 和 $\hat{\theta}_2 = \dfrac{n+1}{n} X_{(n)}$ 哪一个更有效？

解

由于 $$D(\hat{\theta}_1) = 4D(\bar{X}) = 4D\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

$$= \frac{4}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{4}{n^2}\sum_{i=1}^{n} D(X)$$

$$= \frac{4}{n} D(X) = \frac{\theta^2}{3n},$$

$$D(\hat{\theta}_2) = D\left(\frac{n+1}{n} X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 D(X_{(n)}),$$

又因为 $$E(X_{(n)}) = \frac{n}{n+1}\theta,$$

$$E(X_{(n)}^2) = \int_{-\infty}^{+\infty} x^2 p_{X_{(n)}}(x)\mathrm{d}x$$

$$= \int_0^\theta \frac{n}{\theta^n} x^{n+1}\mathrm{d}x = \frac{n}{n+2}\theta^2,$$

$$\boxed{D(\hat{\theta}_1) = \frac{\theta^2}{3n}}$$

$$D(X_{(n)}) = E(X_{(n)}^2) - [E(X_{(n)})]^2 = \frac{n}{(n+1)^2(n+2)}\theta^2,$$

$$\therefore \quad D(\hat{\theta}_2) = D\left(\frac{n+1}{n} X_{(n)}\right)$$

$$= \left(\frac{n+1}{n}\right)^2 D(X_{(n)}) = \frac{1}{n(n+2)}\theta^2,$$

又$n \geq 2$, 所以$D(\hat{\theta}_2) < D(\hat{\theta}_1)$, $\hat{\theta}_2$较$\hat{\theta}_1$有效.

2. 最小方差无偏估计量

定义 如果存在θ的一个无偏估计量$\hat{\theta}_0$,使得对于θ的任一方差存在的无偏估计量$\hat{\theta}$,都有

$$D(\hat{\theta}_0) \le D(\hat{\theta})$$

则称$\hat{\theta}_0$是θ的最小方差无偏估计(量),缩写为 $MVUE$.

注 最小方差无偏估计是一种最优估计.

四、相合性

定义6.6 若$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为参数$\theta$的估计量，若对于任意$\theta\in\Theta$，当$n\to\infty$时，$\hat{\theta}(X_1,X_2,\cdots,X_n)$依概率收敛于$\theta$，则称$\hat{\theta}$为$\theta$的 **相合估计量(或一致估计量)**.

例如 由第五章第一节知样本$k(k\geq 1)$阶矩是总体X的k阶矩$\alpha_k=E(X^k)$的相合估计量

进而若待估参数$\theta=g(\alpha_1,\alpha_2,\cdots,\alpha_n)$，其中$g$为连续函数，则$\theta$的矩估计量$\hat{\theta}=g(\hat{\alpha}_1,\hat{\alpha}_2,\cdots,\hat{\alpha}_n)=g(A_1,A_2,\cdots,A_n)$是$\theta$的相合估计量.

例6 试证:(1) 样本均值 $\overline{X}$ 是总体均值 μ 的相合估计量; (2) 修正样本方差 $S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 及样本方差 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 都是总体方差 σ^2 的相合估计量.

证 (1) 由大数定律知,

$$\forall \varepsilon > 0, \quad 有 \quad \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right\} = 1,$$

所以 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是 μ 的相合估计量.

(2) 又 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^{\ 2} - 2X_i\overline{X} + \overline{X}^2)$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^{\ 2} - \overline{X}^2 = A_2 - \overline{X}^2,$$

(A_2是样本二阶原点矩)

由大数定律知,

$A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$依概率收敛于$E(X^2)$,

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$依概率收敛于$E(X)$,

故 $S_n^2 = A_2 - \overline{X}^2$

依概率收敛于$E(X^2)-[E(X)]^2=\sigma^2$,

所以S_n^2是σ^2的相合估计量.

又 $\lim\limits_{n\to\infty}\dfrac{n}{n-1}=1,$

所以 $S_n^{*2}=\dfrac{n}{n-1}S_n^2$ 也是σ^2的相合估计量.

通过此例题,我们看到,要证明一个估计量具有相合性,必须证明它依概率收敛,这有时很麻烦.因此,我们下面我们不加证明的给出一个相合性的判定定理.

定理6.2　设$\hat{\theta}_n$是θ的一个估计量,若

$$\lim_{n\to\infty} E(\hat{\theta}_n)=\theta, \quad 且 \lim_{n\to\infty} D(\hat{\theta}_n)=0$$

则$\hat{\theta}_n$是θ的相合估计(或一致估计).

利用定理6.2再证例6.

由于　$E(\overline{X})=E(X)=\mu$

$$\lim_{n\to\infty} D(\overline{X})=\lim_{n\to\infty}\frac{D(X)}{n}=\lim_{n\to\infty}\frac{\sigma^2}{n}=0$$

故,由定理6.2, $\overline{X}$是X的相合估计.

卡方分布的方差是?

同样 $\lim\limits_{n\to\infty} E(S_n^2) = \lim\limits_{n\to\infty} \dfrac{n-1}{n}\sigma^2 = \sigma^2,$

$\because \quad \dfrac{n}{\sigma^2}S_n^2 \sim \chi^2(n-1),$ $\therefore \quad D(\dfrac{n}{\sigma^2}S_n^2) = 2(n-1),$

从而 $\lim\limits_{n\to\infty} D(S_n^2) = \lim\limits_{n\to\infty} \dfrac{2(n-1)}{n^2}\sigma^4 = 0,$

故由定理6.2, S_n^2是σ^2的相合估计.

类似的可证, S_n^{*2}也是σ^2的相合估计.

六、小结

估计量的评选的三个标准
- 无偏性
- 有效性
- 相合性

相合性是对估计量的一个基本要求, 不具备相合性的估计量是不予以考虑的.

由最大似然估计法得到的估计量, 在一定条件下也具有相合性. 估计量的相合性只有当样本容量相当大时,才能显示出优越性, 这在实际中往往难以做到,因此,在工程中往往使用无偏性和有效性这两个标准.

例5 设总体 X 服从参数为 θ 的指数分布,概率密度 $p(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0\\ 0, & \text{其它}\end{cases}$,其中参数 $\theta>0$,又设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，试证 $\overline{X}$ 和 $nX_{(1)}=n[\min(X_1,X_2,\cdots,X_n)]$ 都是 θ 的无偏估计.

证 因为 $E(\overline{X})=E(X)=\theta$,

所以 $\overline{X}$ 是 θ 的无偏估计量.

而 $X_{(1)}=\min(X_1,X_2,\cdots,X_n)$ 服从参数为 $\dfrac{\theta}{n}$ 的指数分布,

概率密度 $p_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-\frac{nx}{\theta}}, & x>0\\ 0, & 其它\end{cases}$

故知 $E(X_{(1)})=\dfrac{\theta}{n}$, $E(nX_{(1)})=\theta$,

所以 $nX_{(1)}$ 也是 θ 的无偏估计量.

由以上两例可知,一个参数可以有不同的无偏估计量.

例6　(续例5)

试证当$n>1$时,θ的无偏估计量$\overline{X}$较$nX_{(1)}$有效.

证明　由于 $D(X)=\theta^2$,　故有 $D(\overline{X})=\dfrac{\theta^2}{n}$,

又因为 $D(X_{(1)})=\dfrac{\theta^2}{n^2}$,　故有 $D(nX_{(1)})=\theta^2$,

当$n>1$时,　$D(nX_{(1)})>D(\overline{X})$,

故θ的无偏估计量$\overline{X}$较$nX_{(1)}$有效.

X的分布函数为：

$$F(x)=\begin{cases}1-\exp(-\dfrac{x}{\theta}) & x\geq 0\\ 0 & x<0\end{cases}$$

最小次序统计量$X_{(1)}$的分布密度为

$$p_{X_{(1)}}(x)=n[1-F(x)]^{n-1}p(x)$$

$$=\begin{cases}\dfrac{n}{\theta}\exp(-\dfrac{n}{\theta}x) & x\geq 0\\ 0 & x<0\end{cases}$$