第三节　估计量的评选标准

一、问题的提出

二、无偏性

三、有效性

四、相合性

五、小结

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, 如第一节的例4和例10. 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)

若估计量 $\hat{\theta} = \theta(X_1, X_2, \cdots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在,且对于任意 $\theta \in \Theta$ 有 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 是 θ 的无偏估计量.

无偏估计的实际意义:无系统误差.

例1 设总体 X 的 k 阶矩 $\mu_k = E(X^k)\ (k \geq 1)$ 存在，又设 $X_1, X_2, \cdots, X_n$ 是 X 的一个样本，试证明不论总体服从什么分布，k 阶样本矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 k 阶总体矩 μ_k 的无偏估计.

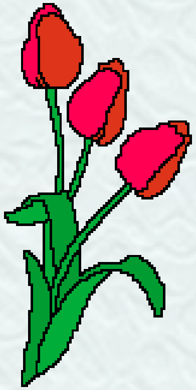
证 因为 $X_1, X_2, \cdots, X_n$ 与 X 同分布，

故有 $E(X_i^k) = E(X^k) = \mu_k,\quad i = 1, 2, \cdots, n.$

即 $E(A_k) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \mu_k.$

故 k 阶样本矩 A_k 是 k 阶总体矩 μ_k 的无偏估计.

特别的:



不论总体 X 服从什么分布,

只要它的数学期望存在,

$\overline{X}$ 总是总体 X 的数学期望 $\mu_1 = E(X)$ 的无偏估计量.

例2 对于均值 μ，方差 $\sigma^2>0$ 都存在的总体，若 $\mu,\ \sigma^2$ 均为未知，则 σ^2 的估计量 $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是有偏的(即不是无偏估计).

证 $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=A_2-\overline{X}^2,$

因为 $E(A_2)=\mu_2=\sigma^2+\mu^2,$

又因为 $E(\overline{X}^2)=D(\overline{X})+[E(\overline{X})]^2=\dfrac{\sigma^2}{n}+\mu^2,$

所以 $E(\hat{\sigma}^2)=E(A_2-\overline{X}^2)=E(A_2)-E(\overline{X}^2)$

$$=\frac{n-1}{n}\sigma^2 \neq \sigma^2,$$ 所以 $\hat{\sigma}^2$ 是有偏的.

若以 $\dfrac{n}{n-1}$ 乘 $\hat{\sigma}^2$,所得到的估计量就是无 偏的.

(这种方法称为**无偏化**).

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

因为 $\dfrac{n}{n-1}\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}^2),$

即 S^2是σ^2 的无偏估计,故通常取S^2作σ^2的估计量.

例3 设 X_1, X_2, L, X_n 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，试求 σ 的无偏估计量.

解 由第六章第二节定理二知 $\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$,

$$E\left(\frac{\sqrt{n-1}S}{\sigma}\right) = \int_0^{+\infty} \sqrt{x}\,\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\mathrm{e}^{-\frac{x}{2}}x^{\frac{n-1}{2}-1}\mathrm{d}x$$

$$= \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\int_0^{+\infty} \mathrm{e}^{-\frac{x}{2}}x^{\frac{n}{2}-1}\mathrm{d}x = \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

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

故 S 不是 σ 的无偏估计量,

$$\sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}S \text{ 是 } \sigma \text{ 的无偏估计量.}$$

例4 设总体 X 在 $[0,\theta]$ 上服从均匀分布，参数 $\theta > 0$, $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，试证明 $2\overline{X}$ 和 $\dfrac{n}{n+1}\max(X_1, X_2, \cdots, X_n)$ 都是 θ 的无偏估计.

证 因为 $E(2\overline{X}) = 2E(\overline{X}) = 2E(X) = 2\times\dfrac{\theta}{2} = \theta,$

所以 $2\overline{X}$ 是 θ 的无偏估计量.

因为 $X_h = \max(X_1, X_2, \cdots, X_n)$ 的概率密度为

$$f(x) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n}, & 0 \le x \le \theta, \\ 0, & \text{其他} \end{cases}$$

所以 $E(X_h)=\int_0^{\theta} x\cdot\frac{nx^{n-1}}{\theta^n}\mathrm{d}x$

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

故有 $E\left(\frac{n+1}{n}X_h\right)=\theta,$

故 $\frac{n}{n+1}\max(X_1,X_2,\cdots,X_n)$ 也是θ的无偏估计量.

例5 设总体 X 服从参数为 θ 的指数分布，概率密度

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

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 $nZ=n[\min(X_1,X_2,\cdots,X_n)]$ 都是 θ 的无偏估计.

证明 因为 $E(\overline{X})=E(X)=\theta,$

所以 $\overline{X}$ 是 θ 的无偏估计量.

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

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

故知 $E(Z)=\dfrac{\theta}{n}$, $E(nZ)=\theta$,

所以 nZ 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.

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，$\hat{\theta}_1$的观察值在真值θ的附近较$\hat{\theta}_2$更密集，则认为$\hat{\theta}_1$较$\hat{\theta}_2$有效.

由于方差是随机变量取值与其数学期望的偏离程度，所以无偏估计以方差小者为好.

设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$与$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$都是$\theta$的无偏估计量，若有$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$，则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

例6　(续例5)

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

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

又因为 $D(Z)=\dfrac{\theta^2}{n^2}$,　故有 $D(nZ)=\theta^2$,

当$n>1$时,　$D(nZ)>D(\overline{X})$,

故θ的无偏估计量$\overline{X}$较nZ有效.

例7 (续例4) 在例4中已证明 $\hat{\theta}_1 = 2\overline{X}$ 和 $\hat{\theta}_2 = \dfrac{n+1}{n}\max\{X_1, X_2, \cdots, X_n\}$ 都是 θ 的无偏估计量，现证当 $n \geq 2$ 时，$\hat{\theta}_2$ 较 $\hat{\theta}_1$ 有效.

证明 由于 $D(\hat{\theta}_1) = 4D(\overline{X}) = \dfrac{4}{n}D(X) = \dfrac{\theta^2}{3n}$,

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

又因为 $E(X_h) = \dfrac{n+1}{n}\theta$,

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

$$D(X_h) = E(X_h{}^2) - [E(X_h)]^2$$

$$= \frac{n}{(n+1)^2(n+2)}\theta^2,$$

故 $D(\hat{\theta}_2) = \dfrac{1}{n(n+2)}\theta^2,$

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

四、相合性

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.

例如 由第六章第二节知，样本$k(k\geq 1)$阶矩是总体X的k阶矩$\mu_k=E(X^k)$的相合估计量，进而若待估参数$\theta=g(\mu_1,\mu_2,\cdots,\mu_n)$，其中$g$为连续函数，则$\theta$的矩估计量$\hat{\theta}=g(\hat{\mu}_1,\hat{\mu}_2,\cdots,\hat{\mu}_n)=g(A_1,A_2,\cdots,A_n)$是$\theta$的相合估计量.

例8 试证：样本均值 $\overline{X}$ 是总体均值 μ 的相合估计量，样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 及样本的二阶中心矩 $B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 都是总体方差 σ^2 的相合估计量.

证明 由大数定律知，

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

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

又 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)$

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

(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 $B_2 = A_2 - \overline{X}^2$

依概率收敛于 $E(X^2) - [E(X)]^2 = \sigma^2$,

所以 B_2 是 σ^2 的相合估计量.

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

所以 $S^2 = \dfrac{n}{n-1} B_2$ 也是 σ^2 的相合估计量.

五、小结

估计量的评选的三个标准 { 无偏性 / 有效性 / 相合性 }

- **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.