估计量是样本的函数, 它是一个随机变量, 由不同的方法得到的估计量可能相同也可能不同. 而即使使用同一种方法也可能得到不同的统计量. 故对同一参数的多个估计量来说, 需要给出判断好坏的标准.

第二节　估计量的评选标准

一、　无偏性

二、　有效性

三、　一致性

一、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.

无偏估计的统计意义：在大量重复试验下，由估计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 得到的估计值的平均恰好是θ.

----无系统误差.

例7.2.1 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 $E(X_i^k) = E(X^k) = \mu_k$, $i = 1, 2, \cdots, n.$

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.

特别地:

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

而 $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是有偏的(即不是无偏估计).

证 因为 $E(S^2)=\sigma^2$, 故 S^2是σ^2 的无偏估计,

$$\text{而}\hat{\sigma}^2=\frac{n-1}{n}S^2,$$

$$\text{所以 } E(\hat{\sigma}^2)=\frac{n-1}{n}E(S^2)=\frac{n-1}{n}\sigma^2\neq\sigma^2,$$

所以$\hat{\sigma}^2$ 是有偏的. 故通常取S^2作σ^2的估计量.

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).

例7.2.3 设总体X的均值为 μ，X_1, X_2, X_3 是总体 X 的样本，证明下列三个估计量

$$\hat{\mu}_1 = \frac{1}{2}(X_1 + X_2),\ \hat{\mu}_2 = \frac{1}{3}(X_1 + X_2 + X_3),\quad \hat{\mu}_3 = \frac{1}{2}X_1 + \frac{1}{6}X_2 + \frac{1}{3}X_3$$

都是 μ的无偏估计.

证 由于

$$E(\hat{\mu}_1) = \frac{1}{2}[E(X_1) + E(X_2)] = E(X) = \mu,$$

$$E(\hat{\mu}_2) = \frac{1}{3}[E(X_1) + E(X_2) + E(X_3)] = E(X) = \mu,$$

$$E(\hat{\mu}_3) = \frac{1}{2}E(X_1) + \frac{1}{6}E(X_2) + \frac{1}{3}E(X_3)$$

$$= \frac{1}{2}\mu + \frac{1}{6}\mu + \frac{1}{3}\mu = \mu.$$

所以$\hat{\mu}_1, \hat{\mu}_2$与 $\hat{\mu}_3$ 都是 μ 的无偏估计.

注 只需 $k_1 + k_2 + \ldots + k_n = 1$，则$\hat{\mu} = k_1 X_1 + k_2 X_2 + \ldots + k_n X_n$ 就是 μ的无偏估计.

例7.2.4 设总体 X 的均值为μ,方差为σ^2.分别独立地从总体 X中抽取样本$X_1,X_2,\cdots,X_m$ 及样本$X_1',X_2',\cdots,X_n'$,样本均值分别为$\overline{X}$及$\overline{X'}$,令$\hat{\mu}=k\overline{X}+k'\overline{X'}$.问:

(1)当k和k'满足什么条件时, $\hat{\mu}$是μ的无偏估计?

(2)当k和k'为何值时,(1)中μ的无偏估计量$\hat{\mu}$的方差最小?

解(1) 因为 $E(\hat{\mu})=E(k\overline{X}+k'\overline{X'})=kE(\overline{X})+k'E(\overline{X'})=(k+k')\mu$,

所以当$k+k'=1$时,$\hat{\mu}$为μ的无偏估计。

(2)由题设可知,$\overline{X}$与$\overline{X'}$相互独立,因此

$$D(\hat{\mu})=D(k\overline{X}+k'\overline{X'})=k^2D(\overline{X})+k'^2D(\overline{X'})=\left[\frac{k^2}{m}+\frac{k'^2}{n}\right]\sigma^2$$

$$=\left[\frac{k^2}{m}+\frac{(1-k)^2}{n}\right]\sigma^2.$$

令$\dfrac{dD(\hat{\mu})}{dk}=\dfrac{d}{dk}\left[\dfrac{k^2}{m}+\dfrac{(1-k)^2}{n}\right]\sigma^2=\left[\dfrac{2k}{m}-\dfrac{2(1-k)}{n}\right]\sigma^2=2\sigma^2\cdot\dfrac{nk-m+mk}{mn}=0,$

解得$k=\dfrac{m}{m+n},\ k'=1-k=\dfrac{n}{m+n}.$

例7.2.5 设总体X服从$[0,\theta]$上的均匀分布，θ未知。$X_1, X_2, \ldots, X_n$是取自总体X的样本.

(1)求θ的矩估计量$\hat{\theta}_1$和最大似然估计量$\hat{\theta}_2$;

(2)判断$\hat{\theta}_1$和$\hat{\theta}_2$的无偏性.

解

(1)矩估计:

$$\mu_1 = E(X) = \frac{\theta}{2},$$

令$\mu_1 = A_1$,即令$\frac{\theta}{2} = \overline{X}$,

即得参数θ的矩估计量$\hat{\theta}_1 = 2\overline{X}$.

例7.2.5 设总体X服从$[0,\theta]$上的均匀分布，θ未知。

$X_1,X_2,\ldots,X_n$是取自总体X的样本.

(1)求θ的矩估计量$\hat{\theta}_1$和最大似然估计$\hat{\theta}_2$;

(2)判断$\hat{\theta}_1$和$\hat{\theta}_2$的无偏性.

解

(1)最大似然估计： $f(x)=\begin{cases}\dfrac{1}{\theta}, & 0\le x\le\theta\\ 0, & \text{其它}\end{cases}$

设样本值为$x_1,x_2,\ldots,x_n$,

似然函数$L(\theta)=f(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}f(x_i)$

$$=\begin{cases}\dfrac{1}{\theta^n}, & 0\le x_i\le\theta\\ 0, & \text{其它}\end{cases},i=1,2,\ldots,n$$

$\hat{\theta}_2=\max\{x_1,x_2,\ldots,x_n\}$为$\theta$的最大似然估计值；

$\hat{\theta}_2=\max\{X_1,X_2,\ldots,X_n\}$为$\theta$的最大似然估计量。

$$\hat{\theta}_1 = 2\bar{X} = \frac{2}{n}\sum_{i=1}^{n} X_i;\quad \hat{\theta}_2 = \max\{X_1, X_2, \ldots, X_n\}.$$

(2)判断$\hat{\theta}_1$和$\hat{\theta}_2$的无偏性.

解　即判断$E\hat{\theta}_1$和$E\hat{\theta}_2$是否等于θ。

$$E\hat{\theta}_1 = E(2\bar{X}) = 2E\bar{X} = 2EX = 2\times\frac{\theta}{2} = \theta,$$

故$\hat{\theta}_1$是θ的无偏估计量.

$$E\hat{\theta}_2 = ? = \int_{-\infty}^{+\infty} x \underline{f_{\hat{\theta}_2}(x)}\, dx$$

$$\Uparrow$$

$F_{\hat{\theta}_2}(x)$求导

$$E\hat{\theta}_2 = ? = \int_{-\infty}^{+\infty} x f_{\hat{\theta}_2}(x)dx$$

$$F_{\hat{\theta}_2}(x) = P\left\{\hat{\theta}_2 \le x\right\} = P\left\{\max\{X_1,\ldots,X_n\} \le x\right\}$$

$$= P\left\{X_1 \le x,\ldots,X_n \le x\right\}$$

$$= P\left\{X_1 \le x\right\}\ldots P\{X_n \le x\}$$

$$= F_{X_1}(x)\cdots F_{X_n}(x)$$

$$= \left[F(x)\right]^n = \begin{cases} 0, & x < 0, \\ \dfrac{x^n}{\theta^n}, & 0 \le x \le \theta, \\ 1, & x > \theta. \end{cases}$$

$$F(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x}{\theta}, & 0 \le x \le \theta, \\ 1, & x > \theta. \end{cases}$$

$$F_{\hat{\theta}_2}(x)=\begin{cases}0, & x<0,\\ \dfrac{x^n}{\theta^n}, & 0\le x\le\theta,\\ 1, & x>\theta.\end{cases}$$

$$f_{\hat{\theta}_2}(x)=\frac{d\left[F_{\hat{\theta}_2}(x)\right]}{dx}=\begin{cases}\dfrac{n}{\theta^n}x^{n-1}, & 0\le x\le\theta,\\ 0, & \text{其它}.\end{cases}$$

于是$E\hat{\theta}_2=\int_{-\infty}^{+\infty}xf_{\hat{\theta}_2}(x)dx=\int_0^{\theta}x\cdot\dfrac{n}{\theta^n}\cdot x^{n-1}dx=\dfrac{n\theta}{n+1}$

$\neq\theta$

故$\hat{\theta}_2$不是θ的无偏估计量.

二、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 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$，则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

例7.2.6（7.2.3续）设总体X的均值为μ ,方差为σ^2 ($\sigma>0$) , X_1, X_2, X_3 是总体X 的样本，比较下列三个无偏估计量

$$\hat{\mu}_1=\frac{1}{2}(X_1+X_2),\ \hat{\mu}_2=\frac{1}{3}(X_1+X_2+X_3),\quad \hat{\mu}_3=\frac{1}{2}X_1+\frac{1}{6}X_2+\frac{1}{3}X_3$$

哪个更有效?

证 由于

$$D(\hat{\mu}_1)=\frac{1}{4}[D(X_1)+D(X_2)]=\frac{1}{2}\sigma^2,$$

$$D(\hat{\mu}_2)=\frac{1}{9}[D(X_1)+D(X_2)+D(X_3)]=\frac{1}{3}\sigma^2,$$

$$D(\hat{\mu}_3)=\frac{1}{4}D(X_1)+\frac{1}{36}D(X_2)+\frac{1}{9}D(X_3)=\frac{7}{18}\sigma^2,$$

可见$D(\hat{\mu}_2)<D(\hat{\mu}_3)<D(\hat{\mu}_1)$,

所 以 在 μ的 三 个 无 偏 估 计 量 $\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3$中 ， $\hat{\mu}_2$最 有 效 。

三、相合性（一致性）

定义7.2.3 设$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$ $(n=1,2,\cdots)$是未知参数θ的估计量序列,如果对于任意$\theta\in\Theta$, 当$n\to\infty$时$\hat{\theta}_n$依概率收敛于θ, 即对于任意给定的正数ε, 有$\lim\limits_{n\to\infty}P\{|\hat{\theta}_n-\theta|<\varepsilon\}=1$,

则称$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$为参数$\theta$的一致估计量或相合估计量.

也称以$\hat{\theta}$估计θ具有一致性或相合性。

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通常是不予以考虑的.

三、相合性（一致性）

定义7.2.3 设$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$ $(n=1,2,\cdots)$是未知参数θ的估计量序列,如果对于任意$\theta\in\Theta$, 当$n\to\infty$时$\hat{\theta}_n$依概率收敛于θ, 即对于任意给定的正数ε, 有$\lim\limits_{n\to\infty}P\{|\hat{\theta}_n-\theta|<\varepsilon\}=1$,

则称$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$为参数$\theta$的一致估计量或相合估计量.

也称以$\hat{\theta}$估计θ具有一致性或相合性。

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通常是不予以考虑的.

如果一个估计量在样本容量不断增大时，它都不能把被估参数估计到任意指定的精度，那么这个估计很值得怀疑。

注 根据辛钦大数定律知, 样本均值 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是总体均值 μ 的一致估计量.

样本k阶原点矩$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$是总体$X$的$k$阶原点矩

$\mu_k = E(X^k)\ (k = 1, 2, \cdots)$的一致估计量。

如果参数θ是$\mu_1, \mu_2, \cdots, \mu_l$的连续函数 $\theta = g(\mu_1, \mu_2, \cdots, \mu_l)$,

则θ的矩估计量$\hat{\theta} = g(A_1, A_2, \cdots, A_l)$是 θ的一致估计量。

矩估计一般都具有一致性。

一致性估计量仅在样本容量 n 足够大时, 才显示其优越性.

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7.2.7 设 θ 是总体 X 分布中的未知参数，$X_1, X_2, \cdots, X_n$ 是来自总体的样本，$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是 θ 的无偏估计量，且 $\lim\limits_{n\to\infty} D(\hat{\theta}_n) = 0$，证明 $\hat{\theta}_n$ 为 θ 的一致估计量.

证明 因为 $E(\hat{\theta}_n) = \theta$，根据切比雪夫不等式，对于任意的 $\varepsilon > 0$，有

$$1 - \frac{D(\hat{\theta}_n)}{\varepsilon^2} \le P\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} \le 1.$$

又因为 $\lim\limits_{n\to\infty} D(\hat{\theta}_n) = 0$，所以 $\lim\limits_{n\to\infty} P\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} = 1$，

故 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 为 θ 的一致估计量.