第三节　估计量的评选标准

- 无偏性
- 有效性
- 相合性
- 小结　布置作业

$$X \sim N(\mu, \sigma^2)$$

样本均值是否是 μ 的一个好的估计量?

样本方差是否是 σ^2 的一个好的估计量?

这就需要讨论以下几个问题:

(1) 我们希望一个“好的”估计量具有什么特性?

(2) 怎样决定一个估计量是否比另一个估计量“好”?

(3) 如何求得合理的估计量?

例如，要估计某院19级本科生的高数平均成绩

方案一：设计一个抽样方案，取200个同学的高数成绩，计算出他们的平均成绩，作为真实成绩的估计；

方案二：随便取一个同学的成绩作为真实成绩的估计。

估计量的评选标准

在介绍估计量的评选标准之前，我们必须强调指出：

评价一个估计量的好坏，不能仅仅依据一次试验的结果，而必须由多次试验结果（某种整体性能）来衡量．

这是因为估计量是样本的函数，是随机变量．因此，由不同的观测结果，就会求得不同的参数估计值．因此一个好的估计，应在多次试验中体现出优良性．

常用的几条标准是:

1. 无偏性

2. 有效性

3. 相合性

这里我们重点介绍前面两个标准 .

一、无偏性

估计量是随机变量，对于不同的样本值会得到不同的估计值．我们希望估计值在未知参数真值附近摆动，而它的期望值等于未知参数的真值. 这就导致无偏性这个标准．

设 $\hat{\theta}(X_1,\cdots,X_n)$ 是未知参数θ 的估计量，若

$$E(\hat{\theta})=\theta$$

则称 $\hat{\theta}$为θ 的无偏估计．

无偏性是对估计量的一个常见而重要的要求．

$E(\hat{\theta})-\theta$称为$\hat{\theta}$作为θ的估计的系统误差。

无偏性的实际意义是指没有系统性的偏差．

例如，用样本均值作为总体均值的估计时，虽无法说明一次估计所产生的偏差，但这种偏差随机地在0的周围波动，对同一统计问题大量重复使用不会产生系统偏差．

不论服从什么分布，样本均值X是总体均值μ的无偏估计，样本方差$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$是总体方差$\sigma^2$的无偏估计，而估计量$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$却不是$\sigma^2$的无偏估计。

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

简证: $E(A_k) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^k)$

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

- 对于中心距此结果则不成立！

【证明】 $E(S^2)=\sigma^2, E(B_2)=\dfrac{n-1}{n}\sigma^2,$

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

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

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

$$=\frac{1}{n-1}\{\sum_{i=1}^{n}E(X_i^2)-nE[(\overline{X})^2]\}$$

$$\because E(X_i^2)=D(X_i)+(EX_i)^2=\sigma^2+\mu^2$$

$$E[(\overline{X})^2]=D(\overline{X})+[E(\overline{X})]^2=\frac{\sigma^2}{n}+\mu^2$$

$$\therefore E(S^2)=\frac{1}{n-1}[(n\sigma^2+n\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)]=\sigma^2$$

考察 $B_2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$

由于 $E\left(B_2\right)=E\left\{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right\}$

$$=E\left\{\frac{n-1}{n}\cdot\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right\}$$

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

因此，$B_2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ 是总体方差 σ^2 的有偏估计.

- 同一参数可能有多个无偏估计

如果未知参数θ有两个不同的无偏估计$\hat{\theta}_1$与$\hat{\theta}_2$，则θ一定有无穷多个无偏估计.

这是因为，对任意的实数α，

$$\alpha\hat{\theta}_1+(1-\alpha)\hat{\theta}_2$$

一定是未知参数θ的无偏估计.

- 无偏估计可能不存在
- 某些有偏估计可以修正为无偏估计
- 无偏性在函数变换下不一定有不变性
- 无偏估计有可能是不合理的

例1 设$X_1,\cdots,X_n \overset{iid}{\sim} U(0,\theta)$,

考察θ的矩估计和极大似然估计的无偏性

解: θ的矩估计和极大似然估计分别为

$$\hat{\theta}_M = 2\overline{X} \quad ,\hat{\theta}_{MLE} = \max(X_i)$$

$$E(\hat{\theta}_M) = 2E(\overline{X}) = 2\times E(X) = 2\times\frac{\theta}{2} = \theta$$

θ的矩估计是无偏的.记 $Z = \hat{\theta}_{MLE} = \max(X_i)$

$$f_Z(z) = \begin{cases} \dfrac{nz^{n-1}}{\theta^n}, 0 < z < \theta \\ 0, \qquad 其它 \end{cases}$$

$$E\hat{\theta}_{MLE}=\int_0^\theta z\frac{nz^{n-1}}{\theta^n}dz=\frac{n}{n+1}\theta\neq\theta$$

故θ的极大似然估计不是无偏的.

注:取

$$\hat{\theta}^*=\frac{n+1}{n}\hat{\theta}_{MLE}$$

则 $\hat{\theta}^*$是θ的无偏估计.

例1　设总体 X 服从参数为 θ 的指数分布，其概率密度为

$$f(x)=\begin{cases}\frac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & \text{其它},\end{cases}$$

其中 $\theta>0$ 为未知，$X_1, X_2, \ldots X_n$ 是取自总体的一个样本，试证 $\bar{X}$ 和 $nZ=n[\min(X_1,\ldots,X_n)]$ 都是参数 θ 的无偏估计量.

【证明】 指数分布$E(X)=\theta$

$$E(\overline{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu$$

故 $E\left(\overline{X}\right)=\theta$

所以$\overline{X}$是参数θ的无偏估计量 .

$Z=\min(X_1,\ldots,X_n)$ （详见3.5节）

当$X_1,\ldots,X_n$相互独立且具有相同分布函数$F(x)$时，有

$$F_{\min}\left(z\right)=1-[1-F\left(z\right)]^n$$

[参阅3.5节P82 例5的结论]

若X~$E(\alpha)$， Y~$E(\beta)$，且相互独立，即：

$$f_X(x)=\begin{cases}\alpha e^{-\alpha x}, & x>0,\\ 0, & x\le 0,\end{cases}\quad f_Y(y)=\begin{cases}\beta e^{-\beta y}, & y>0,\\ 0, & y\le 0,\end{cases}$$

则 $Z=\min(X,Y)$ 的概率密度为

$$f_{\min}(z)=F'_{\min}(z)=\begin{cases}(\alpha+\beta)e^{-(\alpha+\beta)z}, & z>0,\\ 0, & z\le 0,\end{cases}$$

因此, $Z=\min(X_1,\ldots,X_n)$具有概率密度

$$f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-nx/\theta}, & x>0,\\ 0, & \text{其它},\end{cases}$$

【证明】

$Z=\min(X_1,\ldots,X_n)$ 具有概率密度

$$f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-nx/\theta}, & x>0,\\ 0, & \text{其它},\end{cases}$$

即 $Z \sim E(n/\theta)$

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

即 nZ 也是参数 θ 的无偏估计量.

一个参数往往有不止一个无偏估计, 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是参数 θ 的无偏估计量, 我们可以比较 $E(\hat{\theta}_1-\theta)^2$ 和 $E(\hat{\theta}_2-\theta)^2$ 的大小来决定二者谁更优 .

由于

$$D(\hat{\theta}_1)=E(\hat{\theta}_1-\theta)^2$$

$$D(\hat{\theta}_2)=E(\hat{\theta}_2-\theta)^2$$

所以无偏估计以方差小者为好, 这就引进了有效性这一概念 .

二、有效性

设 $\hat{\theta}_1=\hat{\theta}_1(X_1,\cdots,X_n)$ 和 $\hat{\theta}_2=\hat{\theta}_2(X_1,\cdots,X_n)$ 都是参数 θ 的无偏估计量，若对任意 $\theta\in\Theta$，

$$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$$

且至少对于某个 $\theta\in\Theta$ 上式中的不等号成立，

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

- 有效性是在无偏性的前提下才考虑的！

例2 (续例1) 试证 当 $n>1$ 时 θ 的无偏估计量

$\bar{X}$ 较 $nZ=n\times\min(X_1,\dots,X_n)$ 有效．

【证明】指数分布 $D(X)=\theta^2$，

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

又 $Z\sim E(n/\theta)$

故 $D(Z)=\left(\frac{\theta}{n}\right)^2=\frac{\theta^2}{n^2}$， 故有 $D(nZ)=\theta^2$．

当 $n>1$ 时， $D(nZ)>D(\bar{X})$， 故 $\bar{X}$ 较 nZ 有效．

三、相合性（一致性）

设 $\hat{\theta}(X_1,\ldots,X_n)$ 是参数 θ 的估计量，若对于任意 $\theta\in\Theta$，当 $n\to\infty$ 时 $\hat{\theta}(X_1,\ldots,X_n)$依概率收敛于 θ，则称 $\hat{\theta}$ 为 θ 的相合估计量.

$\hat{\theta}$ 为 θ 的相合估计量

$\Leftrightarrow$ 对于任意 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\hat{\theta}-\theta|<\varepsilon\}=1,\ \theta\in\Theta$$

弱大数定理(辛钦大数定理) 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布且具有数学期望 $E(X_k)=\mu\ (k=1,2,\cdots)$，则序列 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$ 依概率收敛于 μ，即 $\overline{X} \xrightarrow{P} \mu$.

由辛钦定理

若总体 X 的数学期望 $E(X)=\mu$ 有限，

则有

$$A_k=\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k)=\mu_k\ (k=1,2,\cdots)$$

$$g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \cdots, \mu_k)$$

其中 g 为连续函数.

故

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$$ 为 $E(X^k)=\mu_k\ (k=1,2,\cdots)$ 的相合估计量.

若 g 为连续函数，则有

$g(A_1,A_2,\cdots,A_k)$ 为 $g(\mu_1,\mu_2,\cdots,\mu_k)$ 的相合估计量.

例. 设 $X_1,\cdots,X_n \overset{iid}{\sim} B(m,p)$，m已知,0<p<1,
讨论p的极大似然估计量的相合性。

解: $$\hat{p}_{MLE} = \frac{1}{m}\overline{X}.$$

由辛钦大数定理,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E(X) = mp$$

$$\therefore \frac{1}{m}\overline{X} \xrightarrow{p} p$$

故p的极大似然估计量是相合性估计量.

四、小结

对于一个未知参数可以提出不同的估计量，因此自然提出比较估计量的好坏的问题，这就需要给出评定估计量好坏的标准.

在本节中,介绍了评定估计量好坏的三个标准:无偏性、有效性、和相合性.

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.