§2 估计量的评选标准

由上一节的学习可知：对于总体分布中的未知参数，利用矩估计法和最大似然估计法可以获得未知参数的估计量，对于同一个参数两种方法所得到的估计量可能相同，也可能不相同。如果不同，我们就会问：

$\left\{\begin{array}{l}\text{采用哪一个估计量更好？}\\ \text{好坏的评价标准是什么？}\end{array}\right.$

本节将介绍几个常用评估标准：

无偏性　有效性　一致性

一、无偏性

估计量是样本的函数，因而它是一个RV，则由不同的样本值便得到参数的不同估计值。既然无法要求每一次由样本值得到的估计值都与参数的真值相等，但却可以要求参数的这些估计值的均值与参数的真值相等，因此便给出了无偏性的评价标准。

定义　若估计量$\hat{\theta}(X_1, X_2, \cdots, X_n)$的数学期望$E(\hat{\theta})$存在，$\forall \theta \in \Theta$，有$E(\hat{\theta}) = \theta$，则$\hat{\theta}$称是$\theta$的无偏估计(量)。(Unbiased Estimator) 否则，称为有偏估计(量)。

并且，称$E(\hat{\theta}) - \theta$是估计量的偏差，或在科学技术中称为系统误差。

因此，无偏估计的实际意义：无系统误差。

如果$E(\hat{\theta}) \neq \theta$，但是$\lim\limits_{n\to\infty} E(\hat{\theta}) = \theta$，则称$\hat{\theta}$是$\theta$的渐近无偏估计(量)。

讨 论

① 总体均值和总体方差的无偏估计

只要总体 X 的均值 $E(X)=\mu$ 和方差 $D(X)=\sigma^2$存在，则无论总体 X 服从什么分布，都有：

$$E(\bar{X})=\mu \qquad E(S^2)=\sigma^2$$

也就是说，<u>不论总体服从何种分布</u>，都有：

样本均值$\bar{X}$是总体均值的无偏估计；

样本方差S^2是总体方差的无偏估计。

讨 论

② 样本二阶中心矩B_2 不是总体方差的无偏估计。

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

若以$\dfrac{n}{n-1}$乘以B_2,则所得的估计量$\dfrac{n}{n-1}B_2$是无偏的。

$$即\ E\left(\frac{n}{n-1}B_2\right)=E(S^2)=\sigma^2$$

这种变化称为无偏化。

但$\lim\limits_{n\to\infty}E(B_2)=\lim\limits_{n\to\infty}\dfrac{n-1}{n}\sigma^2=\sigma^2$, 故$B_2$是总体方差的渐近无偏估计。

例1 设总体X的k阶矩$\mu_k=E(X^k)$存在，若 $X_1, X_2, \ldots, X_n$是来自X的样本，试证：无论X为何种分布，$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$是$\mu_k$的无偏估计。

证：

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

$$=\frac{1}{n}\cdot nE(X_i^k)=\frac{1}{n}\cdot nE(X^k)=\mu_k$$

所以，样本的k阶矩是总体k阶矩的无偏估计。

例2　设总体$X \sim Exp(\theta)$，其PDF为

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

其中参数$\theta>0$未知，$X_1, X_2, \ldots, X_n$是来自X的样本，

试证：$\bar{X}$和$nZ=n\left(\min\left\{X_1,\ldots,X_n\right\}\right)$都是$\theta$的无偏估计。

证：由于$E\left(\bar{X}\right)=\theta$，故$\bar{X}$是$\theta$的无偏估计。

要求$E\left(nZ\right)=nE\left(Z\right)$，需知$Z=\min\left\{X_1,\ldots,X_n\right\}$的分布

利用极值分布易知：$F_Z\left(z\right)=1-\left[1-F_X\left(z\right)\right]^n$

$$F_Z(z)=1-\left[1-F_X(z)\right]^n$$

$$F_X(z)=\begin{cases}1-e^{-z/\theta}, & z>0\\ 0\ , & else\end{cases}$$

$$故 f_Z(z)=F_Z'(z)=n\left[1-F_X(z)\right]^{n-1}f_X(x)$$

$$=\begin{cases}n\left(e^{-z/\theta}\right)^{n-1}\dfrac{1}{\theta}e^{-z/\theta}, & z>0\\ 0\ , & else\end{cases}$$

$$=\begin{cases}\dfrac{n}{\theta}\left(e^{-nz/\theta}\right), & z>0\\ 0\ , & else\end{cases} \qquad \Rightarrow Z\sim Exp\left(\frac{\theta}{n}\right)$$

$\therefore E(nZ)=nE(Z)=n\times\dfrac{\theta}{n}=\theta,$ 故 nZ 是 θ 的无偏估计。

说 明:

例2中，$\bar{X}$和nZ都是θ的无偏估计。

此外，由于$X\sim Exp(\theta)$，则$E(X_i)=\theta$， 所以样本 $X_1, X_2, \dots, X_n$中的任何一个都是θ的无偏估计。

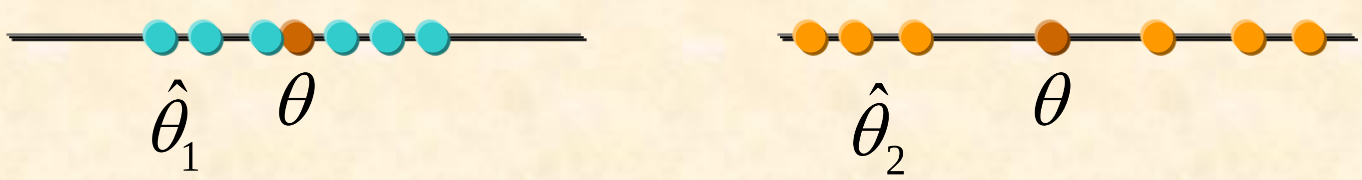
由此可知: **对总体分布中的同一个未知参数，可以有多个无偏估计量。**

因此，仅用无偏性来评估量是不够的，下面来讨论其它评价标准。

二、有效性

$\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别是$\boldsymbol{\theta}$的两个估计量，如何区分哪个估计量较好？

可以对比 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别与$\boldsymbol{\theta}$的均方误差，即 $E\left[\left(\hat{\theta}-\theta\right)^2\right]$，当然是均方误差较小者更优。



针对总体分布中的未知参数θ，若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是θ的无偏估计，在样本容量 n 相同的情况下，如果 $\hat{\theta}_1$ 的观察值较 $\hat{\theta}_2$ 更密集于真值θ附近，即 $\hat{\theta}_1$ 的均方误差 $E\left[\left(\hat{\theta}_1-\theta\right)^2\right]=E\left[\left(\hat{\theta}_1-E\left(\hat{\theta}_1\right)\right)^2\right]$ 较小，我们自然认为 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 理想。

考虑到方差是衡量RV取值与其数学期望的偏离程度，所以在众多的无偏估计中，可以认为方差小者为好，这就引出了有效性的评价标准。

定义　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, $\forall\theta\in\Theta$, 有

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

则称$\hat{\theta}_1$比$\hat{\theta}_2$的有效。

若存在θ的一个无偏估计量$\hat{\theta}_0$, 且对于θ的任一无偏估计量$\hat{\theta}$都有

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

则称$\hat{\theta}_0$是θ的最小方差无偏估计(MVUE)。

(Minimum Variance Unbiased Estimator)

例3（例2续）设总体$X\sim Exp(\theta)$，其中参数$\theta>0$未知，$X_1, X_2, ..., X_n$是来自X的样本，$nZ=n\left(\min\left\{X_1,\ldots,X_n\right\}\right)$和$\bar{X}$都是$\theta$的无偏估计，试证：当$n>1$时，$\bar{X}$较$nZ$更有效.

思路：比较二者的方差

可知$Z\sim Exp\left(\dfrac{\theta}{n}\right)$，$D(nZ)=n^2D(Z)=n^2\left(\dfrac{\theta}{n}\right)^2=\theta^2$

$$D\left(\bar{X}\right)=\frac{D(X)}{n}=\frac{\theta^2}{n}$$

当$n>1$时，$D\left(\bar{X}\right)<D(nZ)$，$\bar{X}$较$nZ$更有效.

三、相合性/一致性

无偏性和有效性都是在样本容量 n 固定的前提下提出的，除了这两个评价标准以外，我们自然希望当样本容量 n 充分大时，估计量的观察值稳定于待估参数的真值，这就引出下面相合性的评价标准。

定义　设 $\hat{\theta}=\hat{\theta}(X_1, X_2, \cdots, X_n)$ 是参数 θ 的估计量，$\forall\theta\in\Theta$，有 $\hat{\theta}\xrightarrow{P}\theta$，则称 $\hat{\theta}$ 为 θ 的**相合估计(量)**，或**一致估计(量)**。

说明：当 n 充分大时，估计量 $\hat{\theta}$ 的值将趋于参数 θ 的真值。

判断一个估计量是不是相合估计，除利用相合估计定义外，还可以利用下面的重要定理来判断。

定理 设$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$是参数$\theta$的估计量，若

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

则$\hat{\theta}_n$是θ的**相合估计**。

根据**依概率收敛的性质**，还可以得到下面的定理。

定理 如果$\hat{\theta}_n$是参数θ的相合估计，$g(x)$在$x=\theta$连续，则$g(\hat{\theta}_n)$也是$g(\theta)$的相合估计。

相合估计$\hat{\theta} \xrightarrow{P} \theta$

说 明

① 相合性是对一个估计量的**基本要求**。

从相合性的定义可知当n充分大时，信息量越来越大，对θ的估计应该越来越精确，则估计量的值应该与θ相差无几才合理。如果相合性不满足，即不管样本容量取多大，都不能将θ估计的很准确，显然这样的估计量是不合理的，自然也就不用考虑估计量的其它评价标准了。

② 样本k阶矩是总体k阶矩的相合估计。$A_l \xrightarrow{P} \mu_l$

相合估计 $\hat{\theta} \xrightarrow{P} \theta$

说 明

③ 参数的矩估计往往也是相合估计。

$$\hat{\theta}_l(A_1, A_2, \cdots, A_k) \xrightarrow{P} \theta_l(\mu_1, \mu_2, \cdots, \mu_k)$$ 真值

④ 在一定条件下，最大似然估计也是相合估计。

⑤ 判断一个估计量具有相合性的方法：

定义法 判断 $\hat{\theta} \xrightarrow{P} \theta$（当估计量是样本矩或样本矩的函数时）

定理法 判断 $\lim\limits_{n\to\infty} E(\hat{\theta}_n) = \theta$ 且 $\lim\limits_{n\to\infty} D(\hat{\theta}_n) = 0$

结 论

1. 无论总体服从何种分布，样本均值和样本方差分别是总体均值和总体方差的无偏估计。

$$E\left(\bar{X}\right)=E\left(X\right) \qquad E\left(S^2\right)=D\left(X\right)$$

2. 样本**k** 阶矩是总体**k** 阶矩的无偏估计和相合估计。

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^k\right)=E\left(X^k\right) \qquad \frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E\left(X^k\right)$$

结 论

3. 样本二阶中心矩B_2不是总体方差的无偏估计，而是总体方差的渐近无偏估计。

4. 样本标准差S 不是总体标准差的无偏估计，但用S估计总体标准差总是系统偏低的。

$$E\left(S^2\right)=D(X)$$

$$E(S)=\sqrt{E\left(S^2\right)-D(S)}=\sqrt{D(X)-D(S)}\underset{\uparrow \atop D(S)>0}{<}\sqrt{D(X)}$$

$$\Rightarrow E(S)-\sqrt{D(X)}<0$$

结 论

5. 无偏估计的函数未必是相应函数的无偏估计，这一点与矩估计、MLE的不变性不同。

6. 矩估计和最大似然估计不一定都是无偏的，但一般都是相合估计。

7. 经验分布函数是总体分布函数的无偏估计和相合估计。

第六章 $E\left[F_n(x)\right]=F(x), \qquad F_n(x)\xrightarrow{P}F(x)$

8. 总体均值$E(X)=\mu$的线性无偏估计$\sum_{i=1}^{n} a_i X_i$（常数a_i满足$\sum_{i=1}^{n} a_i = 1$）中最有效估计是$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$。

证：无偏性 $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i E(X) = \mu \sum_{i=1}^{n} a_i = \mu$

有效性 $D\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 D(X_i) = D(X)\sum_{i=1}^{n} a_i^2 = \sigma^2 \sum_{i=1}^{n} a_i^2$

最有效估计是要寻找方差最小的无偏估计，即：

寻找 $a_i\left(满足\sum_{i=1}^{n} a_i = 1\right)$使得$\sum_{i=1}^{n} a_i^2$达到最小。

条件极值问题

目标 Min $\sum_{i=1}^{n} a_i^2$

约束 $\sum_{i=1}^{n} a_i = 1$

引入拉格朗日函数：$L(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} a_i^2 + \lambda\left(\sum_{i=1}^{n} a_i - 1\right)$

令 $\begin{cases} \dfrac{\partial L}{\partial a_i} = 2a_i + \lambda = 0, i = 1, \cdots, n \\ \dfrac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} a_i - 1 = 0 \end{cases}$ 解得 $\begin{cases} \lambda = -\dfrac{2}{n} \\ a_i = \dfrac{1}{n}, i = 1, \cdots, n \end{cases}$

即当$\sum_{i=1}^{n} a_i X_i = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$是最有效的线性无偏估计, 得证。

例4 设总体X的PDF为

$$f(x;\sigma)=\frac{1}{2\sigma}e^{-|x|/\sigma},\quad -\infty<x<+\infty,$$

$\sigma>0$为未知参数, $X_1,\ldots,X_n$是来自X的样本。

(1)求参数σ的矩估计$\hat{\sigma}_{\mathrm{M}}$; $\hat{\sigma}_{\mathrm{M}}$是否为相合估计?

(2)求参数σ的最大似然估计$\hat{\sigma}_{\mathrm{MLE}}$;

(3)判断$\hat{\sigma}_{\mathrm{MLE}}$是否为无偏估计和相合估计?

解: (1) $E(X)=\int_{-\infty}^{+\infty}x\frac{1}{2\sigma}e^{-|x|/\sigma}\mathrm{d}x\ =0$

由于$E(X)=0$, 不能用其来估计σ。

$$E\left(X^2\right)=\int_{-\infty}^{+\infty} x^2 \frac{1}{2\sigma} e^{-|x|/\sigma} \mathrm{d}x = 2\int_0^{+\infty} x^2 \frac{1}{2\sigma} e^{-x/\sigma} \mathrm{d}x = \int_0^{+\infty} x^2 \frac{1}{\sigma} e^{-x/\sigma} \mathrm{d}x$$

$$=-\int_0^{+\infty} x^2 \mathrm{d}e^{-x/\sigma} = -x^2 e^{-x/\sigma}\Big|_0^{\infty} + \int_0^{+\infty} e^{-x/\sigma} \mathrm{d}x^2 = 2\int_0^{+\infty} x e^{-x/\sigma} \mathrm{d}x$$

法1: $=-2\sigma\int_0^{+\infty} x\mathrm{d}e^{-x/\sigma} = -2\sigma x e^{-x/\sigma}\Big|_0^{\infty} + 2\sigma\int_0^{+\infty} e^{-x/\sigma}\mathrm{d}x = -2\sigma^2 e^{-x/\sigma}\Big|_0^{\infty} = 2\sigma^2$

法2: $=2\sigma\int_0^{+\infty} x \boxed{\frac{1}{\sigma} e^{-x/\sigma}} \mathrm{d}x \overset{Y\sim\exp(\sigma)}{=} 2\sigma E(Y) = 2\sigma^2$

求矩估计：令$E\left(X^2\right)=A_2=\frac{1}{n}\sum_{i=1}^{n} X_i^2$，解得$\hat{\sigma}_{\mathrm{M}}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}$

相合性(定义法)：$A_2=\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E\left(X^2\right)=2\sigma^2$

$\therefore \sqrt{A_2/2} \xrightarrow{P} \sqrt{2\sigma^2/2}$，即$\hat{\sigma}_{\mathrm{M}} \xrightarrow{P} \sigma$ $\therefore \hat{\sigma}_{\mathrm{M}}$是$\sigma$的相合估计

(2) 求MLE：写似然函数

$$L(\sigma)=\prod_{i=1}^{n} f(x_i;\sigma)=\prod_{i=1}^{n}\frac{1}{2\sigma}e^{-|x_i|/\sigma}=\frac{1}{2^n\sigma^n}e^{-\sum_{i=1}^{n}|x_i|/\sigma}$$

$$\ln L(\sigma)=-n\ln 2-n\ln\sigma-\sum_{i=1}^{n}|x_i|/\sigma$$

$$\frac{d\ln L(\sigma)}{d\sigma}=\frac{-n}{\sigma}+\sum_{i=1}^{n}|x_i|/\sigma^2=0$$

解得$\hat{\sigma}=\frac{1}{n}\sum_{i=1}^{n}|x_i|$ ⇒ $\hat{\sigma}_{\mathrm{MLE}}=\frac{1}{n}\sum_{i=1}^{n}|X_i|$

(3) 判断$\hat{\sigma}_{\text{MLE}} = \dfrac{1}{n}\sum_{i=1}^{n}|X_i|$是否为无偏估计

$$E\left(\hat{\sigma}_{\text{MLE}}\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(|X|\right)=E\left(|X|\right)$$

$$=\int_{-\infty}^{+\infty}|x|\frac{1}{2\sigma}e^{-|x|/\sigma}\mathrm{d}x=2\int_{0}^{+\infty}x\frac{1}{2\sigma}e^{-x/\sigma}\mathrm{d}x$$

$$\overset{Y\sim\exp(\sigma)}{=}E\left(Y\right)=\sigma$$

所以$\hat{\sigma}_{\text{MLE}}=\dfrac{1}{n}\sum_{i=1}^{n}|X_i|$是$\sigma$的无偏估计。

判断$\hat{\sigma}_{\text{MLE}}=\frac{1}{n}\sum_{i=1}^{n}|X_i|$是否为相合估计

$$E\left(X^2\right)=2\sigma^2$$

$$E\left(\hat{\sigma}_{\text{MLE}}\right)=\sigma$$

$$D\left(\hat{\sigma}_{\text{MLE}}\right)=D\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\right)=\frac{1}{n^2}\sum_{i=1}^{n}D\left(|X|\right)=\frac{1}{n}D\left(|X|\right)$$

$$=\frac{1}{n}\left[E\left(|X|^2\right)-E^2\left(|X|\right)\right]=\frac{1}{n}\left[2\sigma^2-\sigma^2\right]=\frac{\sigma^2}{n}$$

$$\lim_{n\to\infty}D\left(\hat{\sigma}_{\text{MLE}}\right)=\lim_{n\to\infty}\frac{\sigma^2}{n}=0$$

由相合性定理知$\hat{\sigma}_{\text{MLE}}=\frac{1}{n}\sum_{i=1}^{n}|X_i|$是$\sigma$的相合估计。

小结

本节讨论了评价估计量优良性的三个评价标准：无偏性，有效性，相合性。它们都是某种意义下用于衡量估计量与参数真值的接近程度，是从某一特定方面来看其最优性的：相合性是在样本容量相当大时才显示出来的一种特性，而无偏性和有效性则是样本容量固定的前提下给出的。

作业

Pages 174, 175：

第 9, 10, 12, 13(2), 14 题

例5 1)试证样本均值$\bar{X}$是总体均值$E(X)$的相合估计.

2) 样本方差$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$及样本二阶中心矩

$B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$都是总体方差$D(X)$的相合估计.

证:1) $\because \bar{X} \xrightarrow{P} E(X)$,

$\therefore$由相合估计定义可知$\bar{X}$是总体均值$E(X)$的相合估计.

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

$$\therefore B_2 = A_2 - \bar{X}^2$$

$$B_2 = A_2 - \bar{X}^2$$

$$\because \left.\begin{array}{l} \bar{X} \xrightarrow{P} E(X) \Rightarrow \bar{X}^2 \xrightarrow{P} E^2(X) \\ \qquad\qquad\qquad A_2 \xrightarrow{P} E(X^2) \end{array}\right\} \Rightarrow A_2 - \bar{X}^2 \xrightarrow{P} E(X^2) - E^2(X)$$

也即$B_2 \xrightarrow{P} D(X)$ $\therefore B_2$是总体方差$D(X)$的相合估计.

又$S^2 = \dfrac{n}{n-1} B_2$, $\lim\limits_{n\to\infty} \dfrac{n}{n-1} = 1$ 故当$n \to \infty$时, $S^2 \xrightarrow{P} D(X)$

$\therefore S^2$是总体方差$D(X)$的相合估计.