第三节 估计量的优良性

一. 无偏性

定义： 设 $\hat{\theta}$ 是 θ 的估计量，若 $E(\hat{\theta})$ 存在且对任意的$\theta \in H$ 有：$E(\hat{\theta}) = \theta$

则称 $\hat{\theta}$ 是 θ 的无偏估计量.

例1. 设总体 X 的均值 μ，方差 $\sigma^2>0$ 都存在，若 μ，σ^2 均未知. $X_1, X_2, \cdots, X_n$ 是总体的一个样本。

判断：σ^2 的两个估计量 $\hat{\sigma}_1^{\,2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$\hat{\sigma}_2^{\,2}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 哪个更好一些.

解：

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

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

$$=n\left(D(X)+\left(E(X)\right)^2\right)-n\left(D(\bar{X})+\left(E(\bar{X})\right)^2\right)$$

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

$$E(\hat{\sigma}_2^{\,2})=\sigma^2$$

$$\therefore E(\hat{\sigma}_1^{\,2})=\frac{n-1}{n}\sigma^2\neq\sigma^2$$

故 $\hat{\sigma}_1^{\,2}$ 是有偏的.

例1. 设总体 X 的均值 μ, 方差 $\sigma^2>0$ 都存在，若 μ, σ^2 均未知. $X_1 X_2,\cdots,X_n$是总体的一个样本。

判断:σ^2 的两个估计量 $\hat{\sigma}_1{}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$\hat{\sigma}_2{}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 哪个更好一些.

解:

$$\boxed{E(\hat{\sigma}_1^2)=\frac{n-1}{n}\sigma^2}$$

$E(\hat{\sigma}_2{}^2)=\sigma^2$ 更好一些.

而 $\hat{\sigma}_2{}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=s^2$

结论：样本方差是总体方差的无偏估计量。

例2 设 $X_1, X_2, \cdots, X_{n+1}$ 是总体X的一个样本

$$EX=\mu,\quad DX=\sigma^2,\quad Y=a\sum_{i=1}^{n}\left(X_{i+1}-X_i\right)^2$$

求 a 使 Y 为 σ^2 的无偏估计量

解 $$EY=E\left(a\sum_{i=1}^{n}\left(X_{i+1}-X_i\right)^2\right)=a\sum_{i=1}^{n}E\left(X_{i+1}-X_i\right)^2$$

$$=a\sum_{i=1}^{n}\left(D\left(X_{i+1}-X_i\right)+\left(E\left(X_{i+1}-X_i\right)\right)^2\right)$$

$$=a\sum_{i=1}^{n}\left(DX_{i+1}+DX_i\right)=a\cdot n\cdot 2\sigma^2\overset{令}{=}\sigma^2$$

$\longrightarrow a=\dfrac{1}{2n}.$ 这时 Y 为 σ^2 的无偏估计量

二. 有效性

注意到：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，则可通过比较 $E(\hat{\theta}_1-\theta)^2$ 和 $E(\hat{\theta}_2-\theta)^2$ 的大小来决定二者谁更优 。又由于 $D(\hat{\theta})=E(\hat{\theta}-\theta)^2$，这就引出有效性这个评选标准。

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 θ 的无偏的估计，且两个样本的容量相等。

若：$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$ 则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效。

注：有效性指的是在同是 θ 的无偏估计量的前提下，希望估计值与真值的偏离程度越小越好。一般称方差愈小的估计量愈有效。

例3 若总体 X 的均值为 μ ,方差 σ^2 ，但均为未知，$X_1, X_2, \cdots, X_n$ 是总体的一个样本。

现有两个 μ 的无偏估计量：

$$\hat{\mu}_1 = \bar{X}, \quad \hat{\mu}_2 = X_1$$

求: $\hat{\mu}_1$ 与 $\hat{\mu}_2$ 哪个作为 μ 的无偏估计更有效?

解: $\because$ $D(\hat{\mu}_1) = D(\bar{X}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sigma^2$

$D(\hat{\mu}_2) = D(X_1) = \sigma^2$

且显然， $\frac{\sigma^2}{n} \le \sigma^2$

$\therefore$ 用 $\hat{\mu}_1 = \bar{X}$ 作为 μ 的估计量更有效。

三. 一致性（相合性）

注意到，前面介绍的无偏性和有效性都是在样本容量 n 固定的前提下提出的。当样本容量 n 增大时，自然希望估计量对未知参数的估计更精确。这就有“一致性”的概念。

定义： 设 $\hat{\theta}_n = \theta_n(X_1, X_2, \cdots, X_n)$ 为参数 θ 的估计量，

当 $n \to \infty$ 时 $\hat{\theta}_n$ 依概率收敛于 θ,

即： $\forall \varepsilon > 0$, $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$

则称 $\hat{\theta}_n$ 为 θ 的一致估计量 (相合估计量)

注：应用大数定律容易证明，样本均值、样本方差和样本*k*阶矩分别是总体均值、总体方差和总体*k*阶矩的一致估计量。

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

证明：

设 $X_1, X_2, \dots, X_n$ 是来自总体 X 的一个样本，

结论：若总体 X 的 k 阶矩 $E(X^k)$存在,则 $A_k \xrightarrow{P} E(X^k)$

所以样本*k*阶矩是总体*k*阶矩的一致估计量。

特别，$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E(X)$

所以样本均值是总体均值的一致估计量。

例题4 总体$X \sim U(\theta, 2\theta)$ 服从均匀分布,其中$\theta > 0$是未知参数

$X_1, \dots, X_n$是取自该总体的样本, $\bar{X}$ 为样本均值.

证明: $\hat{\theta} = \frac{2}{3}\bar{X}$ 是参数θ的无偏估计和相合（一致）估计.

证明(1) $X \sim U(\theta, 2\theta) \Rightarrow E(X) = \frac{\theta + 2\theta}{2} = \frac{3}{2}\theta$

$$E(\hat{\theta}) = E\left(\frac{2}{3}\bar{X}\right) = \frac{2}{3}E(\bar{X}) = \frac{2}{3}\cdot\frac{3\theta}{2} = \theta$$是参数θ的无偏估计

(2) $D(X) = \frac{(2\theta - \theta)^2}{12} = \frac{1}{12}\theta^2$

$$D(\hat{\theta}) = \frac{4}{9}D(\bar{X}) = \frac{4}{9n}D(X) = \frac{4\theta^2}{9\times 12n} = \frac{\theta^2}{27n}$$

由切比雪夫不等式得 $P(|\hat{\theta} - \theta| < \varepsilon) \ge 1 - \frac{1}{\varepsilon^2}D(\hat{\theta})$

≤1

$\lim\limits_{n\to\infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$是参数$\theta$的相合估计

≥1 夹逼