Lec4: 点估计(一)

张伟平

September 24, 2009

§1 引言

数理统计的任务是用样本去推断总体. 参数估计是统计推断的一种重要形式. 设有参数分布族$\mathscr{F}=\{F_\theta,\ \theta\in\Theta\}$,其中$\Theta$是参数空间, F_θ的分布形式已知, 但其分布与未知参数θ有关. $X_1,\cdots,X_n$是从总体F_θ中抽出的简单随机样本. 我们的任务是要利用样本对未知参数θ或其函数$g(\theta)$作出估计. 所以在参数分布族的场合, 把参数可理解为定义在参数空间Θ上的实值函数$g(\theta)$, 一个重要的特例是$g(\theta)=\theta$. 例如, $X_1,\cdots,X_n$ i.i.d. $\sim N(a,\sigma^2)$,记$\theta=(\mu,\sigma^2)$,我们希望利用样本对μ和σ^2 或其函数$g(\theta)=\mu/\sigma^2$的值作出估计, 这就是参数估计问题.

有时样本分布族$\mathscr{F}=\{F\}$是非参数分布族, 其中F的分布形式未知, 但其均值、方差等都是刻画总体某方面性质的量, 也都是参数. 因此在非参数分布族场合, 把参数理解为分布族$\mathscr{F}$上的泛函$g(F)$.我们希望利用样本对$g(F)$,例如$g(F)$为总体均值、方差或中位数等, 作出估计, 这也属于参数估计问题. 例如, 我们从某城市居民中抽取一部分, 对其年收入作调查, 获得样本$X_1,\cdots,X_n$, 要对该城市居民的年人均收入作出估计, 就属于这类问题.

参数估计(Parameter estimation)问题常有两类: 点估计和区间估计. *点估计*就是用样本函数的一个具体数值去估计一个未知参数. *区间估计*就是用样本函数的两个值构成的区间去估计未知参数的取值范围.

定义 1. *设$\mathbf{X}=(X_1,\cdots,X_n)$ 为从某总体中抽取的样本, $\hat{g}(\mathbf{X})=\hat{g}(X_1,\cdots,X_n)$是样本的函数, 用$\hat{g}(\mathbf{X})$ 作为$g(\theta)$的估计, 称为点估计(Point estimation).*

本章我们将首先讨论点估计的方法. 对于同一个未知参数θ (为方便计, 此处以$g(\theta)=\theta$为例)的估计量可以有很多. 例如, 设$X_1,\cdots,X_n$是取自某总体$F\in\mathscr{F}$ 的一组简单样本. 对此总体的均值$\theta=E_F(X)$可以给出几个估计量:

$$\begin{aligned}\hat{\theta}_1&=\bar{X}=\frac{1}{n}(X_1+\cdots+X_n),\\ \hat{\theta}_2&=\frac{1}{2}\left(X_{(1)}+X_{(n)}\right),\\ \hat{\theta}_3&=m_{1/2}.\end{aligned}$$

其中$X_{(1)}$和$X_{(n)}$为样本最小和最大次序统计量, $m_{1/2}$ 为样本中位数. 还可以给出其它的估计量. 这就产生一个问题: 我们采用哪一个估计量作为θ的点估计较好呢? 这就涉及到评价一个估计量优劣的标准问题. 标准不同, 回答也不同. 在经典估计理论中, 用来评价估计量好坏的标准有: 无偏性、有效性、相合性和渐近正态性等, 以及在某种标准下寻求最好的估计.

§2　判断估计量的优良性标准

1. 无偏性

我们在评价估计量好坏时, 一般总希望估计量$\hat{\theta}(\mathbf{X})$ 的平均值与θ越接近越好, 即$E(\hat{\theta}(\mathbf{X})-\theta)$越小越好. 由于$\hat{\theta}(\mathbf{X})$是随机变量, $\hat{\theta}(\mathbf{X})$的值有时比θ的值大, 有时比θ的值小, 我们希望$\hat{\theta}(\mathbf{X})$在大量重复使用时, 在平均意义下$\hat{\theta}(\mathbf{X})$与θ的偏差很小. 期望值$E(\hat{\theta}(\mathbf{X})-\theta)=0$, 时就得到无偏性的概念. 将其一般化,用$g(\theta)$代替$\theta$, 用$\hat{g}(\mathbf{X})$代替$\hat{\theta}(\mathbf{X})$,得到如下定义:

定义 1. *设$\mathbf{X}=(X_1,\cdots,X_n)$为从总体$\{F_\theta,\ \theta\in\Theta\}$中抽取的样本, $g(\theta)$是定义于参数空间Θ上的已知函数. $\hat{g}(\mathbf{X})=\hat{g}(X_1,\cdots,X_n)$是$g(\theta)$ 的一个估计量, 如果*

$$E_\theta(\hat{g}(\mathbf{X}))=g(\theta),\quad \text{对任何 } \theta\in\Theta$$

则称$\hat{g}(\mathbf{X})$为$g(\theta)$的一个无偏估计量 (Unbiased estimation). 记$\hat{g}(\mathbf{X})=\hat{g}_n(\mathbf{X})$,若

$$\lim_{n\to\infty}E_\theta(\hat{g}_n(\mathbf{X}))=g(\theta),\quad \text{对任何 } \theta\in\Theta$$

则称$\hat{g}_n(\mathbf{X})$为$g(\theta)$的渐近无偏估计 (Asymptotically unbiased estimation).

无偏性的含意有两个: 第一个含意是无系统偏差. 由于样本的随机性, $\hat{g}(\mathbf{X})$是样本的函数，因此它是一随机变量, 用估计量$\hat{g}(\mathbf{X})$去估计$g(\theta)$,对某些样本,$\hat{g}(\mathbf{X})$与$g(\theta)$相比, 时而偏低; 对另一些样本, $\hat{g}(\mathbf{X})$时而偏高; 无偏性表示, 把这些正负偏差在概率上平均起来, 其值为0. 如用一杆秤去称东西, 误差的来源有二: (i)这杆秤自身结构上有问题. 用它称东西总是倾向于偏高或总是倾向于偏低, 这属于系统误差. (ii)另一种误差是随机误差, 由不可控制的因素产生, 如温度、湿度和工作人员心理波动等影响造成的, 这属于随机误差. 无偏性相当于要求无系统误差. 随机误差总是存在的, 大量重复使用无偏估计, 误差有时为正, 有时为负, 但随机误差可以正负相抵消.

无偏性的另一个含意是: 要求估计量大量重复使用, 在多次重复使用下给出接近真值$g(\theta)$的估计. 设想这样一种情况: 每天抽样对$g(\theta)$进行估计, 第i天的样本为$\mathbf{X}^{(i)}=(X_1^{(i)},\cdots,X_n^{(i)})$,估计值为$\hat{g}\left(\mathbf{X}^{(i)}\right)$,一共作了$n$天. 设$\mathbf{X}^{(1)},\cdots,\mathbf{X}^{(n)}$是独立同分布的样本, 如$\hat{g}(\mathbf{X})$有无偏性, 按大数定律有

$$P\Big(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\hat{g}\big(\mathbf{X}^{(i)}\big)=g(\theta)\Big)=1$$

就是说, 尽管一次估计结果$\hat{g}\left(\mathbf{X}^{(i)}\right)$不一定恰好等于$g(\theta)$, 但在大量重复使用时, 多次估计的算术平均值, 可以任意接近$g(\theta)$. 如果这一估计量$\hat{g}(\mathbf{X})$只使用一次, 无偏性这个概念就失去意义.

在点估计理论中, 目前无偏性仍占有重要的地位. 除了历史因素外, 还有两个原因. 一是无偏性的要求只涉及一阶矩(均值), 在数学处理时较方便. 二是在没有其它合理准则可循时, 人们心理上觉得: 一个具有无偏性的估计, 总比没有这种性质的估计要好些.

定义 2. *设$X_1,\cdots,X_n$是取自期望为μ,方差为σ^2的总体的一个样本. 显然样本均值$\bar{X}$是μ的无偏估计. 证明样本方差$S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 是σ^2的无偏估计.*

证 显然

$$E(S^2)=\frac{1}{n-1}\Big[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)\Big]=\frac{n}{n-1}\left[E(X_1^2)-E(\bar{X}^2)\right]$$

$$= \frac{n}{n-1}\Big[\big(\sigma^2+\mu^2\big)-\big(\sigma^2/n+\mu^2\big)\Big]=\sigma^2,$$

故样本方差S^2是σ^2的无偏估计.

2. 有效性

在应用中, 同一个参数的无偏估计常常不止一个, 那么选用哪一个无偏估计更好呢? 为了解决好这一问题, 就要讨论估计量的有效性(efficiency). 设$\widehat{\theta}_1$和$\widehat{\theta}_2$为θ的两个无偏估计, 由无偏性可知它们的一阶矩相等, 我们比较它们的二阶中心矩—方差, 方差越小越好.

定义 3. *设$\hat{g}_1(\mathbf{X})=\hat{g}_1(X_1,\cdots,X_n)$和$\hat{g}_2(\mathbf{X})=\hat{g}_2(X_1,\cdots,X_n)$为$g(\theta)$ 的两个不同无偏估计量, 若*

$$D_\theta(\hat{g}_1(\mathbf{X}))\le D_\theta(\hat{g}_2(\mathbf{X})),\qquad \text{对一切 } \theta\in\Theta,$$

且至少存在一个$\theta\in\Theta$,使得严格不等号成立, 则称估计量$\hat{g}_1(\mathbf{X})$比$\hat{g}_2(\mathbf{X})$有效.

从这个定义出发可以看出, 在均值相等的条件下, 方差越小的估计量越有效. 例如, $X_1,\cdots,X_n$是取自总体F的一个简单样本, 设总体均值μ和总体方差σ^2都存在, 则$\hat{\theta}_1=X_1$ 和$\hat{\theta}_2=\bar{X}$都是总体均值μ的无偏估计量, 它们的方差分别是

$$D(\hat{\theta}_1)=\sigma^2,\qquad D(\hat{\theta}_2)=\frac{1}{n}\sigma^2.$$

后者更小, 可见$\bar{X}$比X_1更有效, 且n越大, $\bar{X}$对μ的估计就越有效. 这就是在物体的称重问题中, 为什么我们要将物体称n次, 用其平均值作为物重的理由.

3. 相合性

大量实践表明, 随着样本容量n的增加, 估计量$\hat{g}(\mathbf{X})=\hat{g}(X_1,\cdots,X_n)$ 与被估计参数$g(\theta)$的偏差越来越小, 这是一个良好估计量应具有的性质. 试想, 若不然, 无论作多少次试验, 也不能把$g(\theta)$估计到任意指定的精确程度,这样估计量显然是不可取的.

定义 4. *设对每个自然数n, $\hat{g}_n(\mathbf{X})=\hat{g}_n(X_1,\cdots,X_n)$是$g(\theta)$一个估计量, 若$\hat{g}_n(\mathbf{X})$依概率收敛到$g(\theta)$,即对任何$\theta\in\Theta$及$\varepsilon>0$有*

$$\lim_{n\to\infty}P_\theta(|\hat{g}_n(\mathbf{X})-g(\theta)|\ge\varepsilon)=0,$$

则称$\hat{g}_n(\mathbf{X})$为$g(\theta)$的弱相合估计 (Weakly consistent estimation).若对任何$\theta\in\Theta$有

$$P_\theta\left(\lim_{n\to\infty}\hat{g}_n(\mathbf{X})=g(\theta)\right)=1,$$

则称$\hat{g}_n(\mathbf{X})$为$g(\theta)$的强相合估计 (Strongly consistent estimation).若$r>0$ 和对任何$\theta\in\Theta$,有

$$\lim_{n\to\infty}E_\theta|\hat{g}_n(\mathbf{X})-g(\theta)|^r=0,$$

称$\hat{g}_n(\mathbf{X})$为$g(\theta)$的r阶矩相合估计 (Consistent estimation in r'th mean).当$r=2$时称为均方相合估计 (Consistent estimation in quadratic mean).

估计量的相合性是对大样本问题提出的要求, 是估计量的一种大样本性质.

由概率论中关于这几种收敛性的关系, 可知上述三种相合性有如下关系: 强相合$\Longrightarrow$ 弱相合,反之不必对; 对任何$r>0$有:r阶矩相合 $\Longrightarrow$ 弱相合, 反之不必对. 又强相合与r阶矩相合之间没有包含关系.

估计量的大样本性质,还有渐近正态性,我们将在本章后面有关的小节中给出定义.

§3 矩估计

一、矩法和矩估计量

设$X_1,\cdots,X_n$ 是从总体F中抽取的简单随机样本. 这时, 样本矩可用来估计F的相应的总体矩. 即总体k阶原点矩$\alpha_k=E(X^k)$ 的矩估计量是相应的样本 k阶原点矩

$$a_{nk}=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\qquad k=1,2,\cdots, \tag{3.1}$$

特别总体均值$\alpha_1=E(X)$的矩估计量是样本均值$a_{n1}=\bar{X}$.

总体k阶中心矩$\mu_k=E(X-E(X))^k$的矩估计量是相应的样本 k阶中心矩

$$m_{nk}=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^k,\qquad k=1,2,\cdots, \tag{3.2}$$

特别总体方差$\mu_2=E(X-EX)^2$的矩估计量量是$m_{n2}=S_n^2=\frac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)^2$, 它与样本方差$S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 只差一个常数因子.

用a_{nk}, m_{nk}分别估计α_k和 μ_k是一种基于直观的方法, 它的依据是: a_{nk}是α_k的无偏估计, 即

$$Ea_{nk}=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=\frac{1}{n}\sum_{i=1}^{n}\alpha_k=\alpha_k. \tag{3.3}$$

但用m_{nk}估计μ_k,一般不是无偏的, 当样本大小n较大时, 偏差不显著, 且必要时可作一些修正, 使之成为无偏估计. 请看下例:

例1. 设$\mu_2=\sigma^2$是总体方差, 则$S_n^2=m_{n2}$不是σ^2的无偏估计.

证 由例可知

$$\begin{aligned}E(S_n^2) &= E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)=\frac{n-1}{n}E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)\\ &=\frac{n-1}{n}E(S^2)=\frac{n-1}{n}\sigma^2.\end{aligned} \tag{3.4}$$

因而m_{n2}不是σ^2的无偏估计, 且是系统地偏低. 将其修正, 只须用

$$m_{n2}^*=\frac{n}{n-1}m_{n2}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=S^2 \tag{3.5}$$

代替m_{n2},就得到$E(m_{n2}^*)=E(S^2)=\sigma^2$,即$S^2$为总体方差的无偏估计. 这就是我们用$S^2$作为样本方差的定义, 而不用$m_{n2}$的理由所在. 但当$k\geq 4$,就不能通过这样简单的修正得出$\mu_k$的无偏估计.

一般, 样本的k阶中心矩可以用样本的原点矩表出(令 $a_{n0}=1$):

$$\begin{aligned}m_{nk}&=\frac{1}{n}\sum_{i=1}^{n}(X_i-a_{n1})^k=\frac{1}{n}\sum_{i=1}^{n}\sum_{r=0}^{k}\binom{k}{r}(-1)^{k-r}X_i^r a_{n1}^{k-r}\\ &=\sum_{r=0}^{k}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^r\right)(-1)^{k-r}\binom{k}{r}a_{n1}^{k-r}\end{aligned}$$

$$= \sum_{r=0}^{k} (-1)^{k-r} \binom{k}{r} a_{nr}\, a_{n1}^{k-r}, \tag{3.6}$$

下面给出矩法及矩估计量的定义.

定义 1. 设有总体分布族$\{F_\theta, \theta \in \Theta\}$, Θ是参数空间, $g(\theta)$是定义在Θ 上的参数θ的函数, 它可以表为总体分布的某些矩的函数, 即

$$g(\theta) = G(\alpha_1, \cdots, \alpha_k; \mu_1, \cdots, \mu_S). \tag{3.7}$$

设$\mathbf{X} = (X_1, \cdots, X_n)$是从上述分布族中抽取的简单样本, 用$a_{ni}$和 m_{nj}分别代替*(3.7)*式中的α_i和 μ_j得

$$\hat{g}(\mathbf{X}) = G(a_{n1}, \cdots, a_{nk}; m_{n1}, \cdots, m_{ns}), \tag{3.8}$$

其中a_{ni}是α_i的矩估计量, m_{nj}是μ_j的矩估计量, 则$\hat{g}(\mathbf{X})$作为$g(\theta)$的估计量, 称为$g(\theta)$ 的矩估计量*(Moment estimate)*. 这种求矩估计量的方法称为矩法 *(Moment method of estimation)*.

二、若干例子

例2. 设$X_1, \cdots, X_n$是从具有成功概率θ的两点分布总体$b(1, \theta)$中抽取的简单样本, 求θ和$g(\theta) = \theta(1-\theta)$的矩估计.

解 设$X \sim b(1, \theta)$,即有$P(X = x) = \theta^x (1-\theta)^{1-x}$. 由于$E(X) = \theta$,因此$\theta$的矩估计就是$\hat{\theta}(X_1, \cdots, X_n) = \bar{X}$, 而$g(\theta)$的矩估计是

$$\hat{g}(X_1, \cdots, X_n) = \bar{X}(1 - \bar{X}).$$

例3. 设$X_1, \cdots, X_n$ *i.i.d.* $\sim$ 均匀分布$U(\theta_1, \theta_2)$, 参数$\theta = (\theta_1, \theta_2)$,其中$-\infty < \theta_1 < \theta_2 < +\infty$. 求$\theta_1$和$\theta_2$的矩估计量.

解 设$X \sim U(\theta_1, \theta_2)$, 由均匀分布的性质可知

$$E(X) = \alpha_1 = (\theta_1 + \theta_2)/\,2,$$
$$D(X) = \mu_2 = (\theta_2 - \theta_1)^2/12.$$

解此方程组得

$$\theta_1 = \alpha_1 - \sqrt{3\mu_2}, \quad \theta_2 = \alpha_1 + \sqrt{3\mu_2}.$$

将上式中α_1和 μ_2 分别用$\bar{X}$ 和 m_{n2}代入得

$$\hat{\theta}_1(X_1, \cdots, X_n) = \bar{X} - \sqrt{3m_{n2}} = \bar{X} - \sqrt{3}S_n,$$
$$\hat{\theta}_2(X_1, \cdots, X_n) = \bar{X} + \sqrt{3}S_n,$$

其中S_n^2由(3.4)给出(亦可用S代替S_n,此处S^2由公式(3.5)给出).

例4. 设总体分布有概率密度

$$f_\theta(x) = \begin{cases} 2\sqrt{\frac{\theta}{\pi}}\; exp\{-\theta x^2\} & \text{当 } x > 0 \\ 0 & \text{当 } x \le 0, \end{cases}$$

其中$\theta > 0$为未知参数. 这个分布称为*Maxwell*分布, 在气体分子动力学中有应用. 设$X_1, \cdots, X_n$为抽自此总体的简单随机样本, 求$g(\theta) = 1/\theta$的矩估计量.

解 由方程

$$\alpha_1 = E(X) = 2\sqrt{\frac{\theta}{\pi}}\int_0^\infty xe^{-\theta x^2}dx = \frac{1}{\sqrt{\pi\theta}}$$

解得$g(\theta) = 1/\theta = \pi\alpha_1^2$,将$\alpha_1$用$\bar{X}$代替得

$$\widehat{g}_1(X_1, \cdots, X_n) = \pi\bar{X}^2.$$

另一方面, 由另一方程

$$\alpha_2 = E_\theta(X^2) = 2\sqrt{\frac{\theta}{\pi}}\int_0^\infty x^2e^{-\theta x^2}dx = \frac{1}{2\theta}$$

解得$g(\theta) = 1/\theta = 2\alpha_2$, 将$\alpha_2$用$a_{n2} = \frac{1}{n}\sum\limits_{i=1}^n X_i^2$代入,得$g(\theta)$的矩估计

$$\hat{g}_2(X_1, \cdots, X_n) = 2\, a_{n2} = \frac{2}{n}\sum_{i=1}^n X_i^2.$$

由此可见矩估计量不唯一, 但它们同样合理. 这两个矩估计量$\hat{g}_1(\mathbf{X})$ 和$\hat{g}_2(\mathbf{X})$中哪一个更好? 以后我们将证明基于a_{n2}的估计量$\hat{g}_2(\mathbf{X})$,在$g(\theta)$一切无偏估计类中为方差最小者. 基于a_{n1}的估计$\hat{g}_1(\mathbf{X})$不是$g(\theta)$的无偏估计.

例5. 设总体分布有概率密度

$$f_\theta(x) = \begin{cases} \frac{\theta_2}{\Gamma((1+\theta_1)/\theta_2)}x^{\theta_1}\exp\{-x^{\theta_2}\} & x > 0 \\ 0 & x \le 0, \end{cases}$$

参数$\theta = (\theta_1, \theta_2)$的变化范围$-1 < \theta_1 < \infty$, $\theta_2 > 0$.设$X_1, \cdots, X_n$为抽自此总体的简单随机样本, 求θ_1和θ_2的矩估计量.

解 由简单计算得到总体分布的前两阶矩为:

$$\begin{aligned} \alpha_1 &= \Gamma\Big(\frac{2+\theta_1}{\theta_2}\Big)\Big/\Gamma\Big(\frac{1+\theta_1}{\theta_2}\Big), \\ \alpha_2 &= \Gamma\Big(\frac{3+\theta_1}{\theta_2}\Big)\Big/\Gamma\Big(\frac{1+\theta_1}{\theta_2}\Big). \end{aligned} \tag{3.9}$$

按矩估计方法, 用a_{n1}和a_{n2}分别代替(3.9)中的α_1和α_2,用$\hat{\theta}_1$和$\hat{\theta}_2$分别代替θ_1和θ_2, 得到如下的方程组:

$$\begin{aligned} a_{n1} &= \Gamma\Big(\frac{2+\hat{\theta}_1}{\hat{\theta}_2}\Big)\Big/\Gamma\Big(\frac{1+\hat{\theta}_1}{\hat{\theta}_2}\Big), \\ a_{n2} &= \Gamma\Big(\frac{3+\hat{\theta}_1}{\hat{\theta}_2}\Big)\Big/\Gamma\Big(\frac{1+\hat{\theta}_1}{\hat{\theta}_2}\Big). \end{aligned}$$

其解就是θ_1和θ_2的矩估计. 但此处得不出$\hat{\theta}_1$和$\hat{\theta}_2$的解析表达式, 而只能用数值方法. 此例说明不是所有的矩估计都有解析表达式的.

矩估计方法也可以用于多维样本, 请看下例.

例6. 设(X_i, Y_i), $i = 1, 2, \cdots, n$为从一个二维总体中抽取的简单随机样本, 求总体分布的协方差σ_{12}和相关系数和ρ的矩估计.

解 按定义σ_{12}的矩估计量是样本协方差, 即

$$m_{12} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}), \tag{3.10}$$

其中$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$,而$\rho$的矩估计量是样本相关系数, 即

$$r = \frac{\sum\limits_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 \sum\limits_{j=1}^{n}(Y_i - \bar{Y})^2}}\ . \tag{3.11}$$

三、矩估计的无偏性

矩法是由K. Pearson 1894年提出的点估计的古老方法. 它的特点是直观性强, 用此法获得估计量简便、易行, 且不要求事先知道总体的分布, 矩估计量还具有相合性. 它的缺点是: 在参数分布族场合, 没有充分利用其提供的有关参数的信息, 小样本性质不突出. 此外, 矩估计量不具唯一性.

下面我们研究矩估计的下列三方面的性质: 小样本性质有无偏性, 大样本性质有相合性和渐近正态性. 其大样本性质将放在下段考虑.

估计量的无偏性和渐近无偏性的定义在§3.1中已给出, 下面讨论矩估计的无偏性和渐近无偏性.

(1) 样本的k阶原点矩a_{nk}是总体k阶原点矩$\alpha_k (k = 1, 2, \cdots)$的无偏估计, 公式(3.3)已给出了证明.

(2) 对$k \geq 2$,样本的k阶中心矩不是总体k阶中心矩的无偏估计.

(i) 由例3.2.1可知

$$E(m_{n2}) = \frac{n-1}{n}\,\mu_2,$$

将其修正, 得

$$S^2 = \left(\frac{n}{n-1}\right) m_{n2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

是μ_2的无偏估计.

(ii) 经过计算可知样本的3阶中心矩m_{n3}也不是总体3阶中心矩μ_3的无偏估计, 事实上

$$E(m_{n3}) = \frac{(n-1)(n-2)}{n^2}\,\mu_3. \tag{3.12}$$

将其修正, 得

$$m_{n3}^{*} = \frac{n^2}{(n-1)(n-2)}\,m_{n3},$$

它是μ_3的无偏估计.

(iii) 更进一步, 可以证明对$\nu \geq 4$有

$$E(m_{n\nu}) = \mu_\nu + O(1/n), \tag{3.13}$$

因此对$\nu \geq 4$, $m_{n\nu}$也不是总体的ν阶中心矩μ_ν 的无偏估计.

(3) 矩估计一般具有渐近无偏性. 由(2)可见$m_{n\nu}$ ($\nu \geq 2$)是总体ν阶中心矩μ_ν的渐近无偏估计. 如

$$\lim_{n\to\infty} E(m_{n3}) = \lim_{n\to\infty}\frac{(n-1)(n-2)}{n^2}\mu_3 = \mu_3.$$

例7. 设$X_1,\cdots,X_n$ *i.i.d.* $\sim$ 指数分布 $EP(\lambda)$,其密度为$f_\lambda(x)=\lambda e^{-\lambda x}I_{[x>0]}$. 求$\lambda$的矩估计量$\hat{\lambda}$, 并讨论它的无偏性.

解 由于$\alpha_1=\int_0^\infty xf_\lambda(x)dx=1/\lambda$,解得$\lambda=1/\alpha_1$, 将$\alpha_1$用其矩估计量$\bar{X}=a_{n1}$代入, 得到$\lambda$的矩估计量为

$$\hat{\lambda}(X_1,\cdots,X_n)=1\,/\,\bar{X}.$$

由于$Y=\sum\limits_{i=1}^n X_i\sim$ Gamma分布 $G(n,\lambda)$,

$$E(\hat{\lambda})=nE\big(1/Y\big)=n\int_0^\infty\frac{1}{y}\cdot\frac{\lambda^n}{\Gamma(n)}y^{n-1}e^{-\lambda y}dy=n\lambda/(n-1),$$

可见$\hat{\lambda}(X_1,\cdots,X_n)=1/\bar{X}$不是$\lambda$的无偏估计. 因此矩估计量不一定都具有无偏性. 但$\lim\limits_{n\to\infty}E(\hat{\lambda})=\lim\limits_{n\to\infty}[n/(n-1)\cdot\lambda]=\lambda$, 故$\hat{\lambda}(X_1,\cdots,X_n)$是$\lambda$的渐近无偏估计. 对$\hat{\lambda}$略作修正, 可得$\lambda$的一个无偏估计

$$\hat{\lambda}^*(X_1,\cdots,X_n)=(n-1)\hat{\lambda}(X_1,\cdots,X_n)/n=(n-1)/(n\bar{X}).$$

四、矩估计的相合性和渐近正态性

1. 矩估计的相合性

估计量的几种相合性的定义已在§3.1中给出. 一般说来矩估计在较一般的条件下具有相合性. 此处我们给出矩估计的强相合性, 显然相应的矩估计的弱相合性也成立.

(1) 样本k阶原点矩是总体k阶原点矩的强相合估计. 设$X_1,\cdots,X_n$是从总体F中抽取的简单随机样本, a_{nk}为样本的k阶原点矩, α_k为总体的k阶原点矩. 由独立同分布场合的柯尔莫哥洛夫强大数律可知$a_{nk}\xrightarrow{a.s.}\alpha_k$, $k=1,2,\cdots$.即

$$P\big(\lim_{n\to\infty}a_{nk}=\alpha_k\big)=1,\quad k=1,2,\cdots$$

(2) 样本k阶中心矩是总体k阶中心矩的强相合估计. 这一结论的证明用到下列事实(A):

(**A**) 设函数$f(y_1,\cdots,y_k)$在$(c_1,c_2,\cdots,c_k)$处连续, 若$y_{n1}\xrightarrow{a.s.}c_1,\cdots,y_{nk}\xrightarrow{a.s.}c_k$,则$f(y_{n1},\cdots,y_{nk})\xrightarrow{a.s.}f(c_1,c_2,\cdots,c_k)$.

这一事实的证明留作练习. 我们在事实(A)成立的前提下, 证明上述结论: 由于

$$\mu_k=\sum_{r=0}^k(-1)^{k-r}\binom{k}{r}\alpha_k\alpha_1^{k-r}=f(\alpha_1,\alpha_2,\cdots,\alpha_k),$$

显见$f(\cdot)$是其变元的连续函数. 由(1)可知$a_{ni}\xrightarrow{a.s.}\alpha_i$, $i=1,\cdots,k$, 故由上述事实(A)及公式(3.6)立得

$$m_{nk}=f(a_{n1},\cdots,a_{nk})\xrightarrow{a.s.}f(\alpha_1,\cdots,\alpha_k)=\mu_k,\quad k=1,2,\cdots,$$

这就证明了结论.

(3) 设$g(\theta)$有(**??**)形式, 其矩估计为(**??**), 关于此类矩估计的强相合性有下述定理.

定理 1. 设$\mathbf{X}=(X_1,\cdots,X_n)$为从总体$F$中抽取的简单随机样本, 待估函数$g(\theta)=G(\alpha_1,\cdots,\alpha_k,\mu_2,\cdots,\mu_s)$, 其矩估计量为$\hat{g}_n(\mathbf{X})=G(a_{n1},\cdots,a_{nk},m_{n2},\cdots,m_{ns})$,且$G$为其变元的连续函数, 则$\hat{g}_n(\mathbf{X})$为$g(\theta)$之强相合估计.

证 由(1)可知$a_{ni} \xrightarrow{a.s.} \alpha_i$, $i=1,2,\cdots$.由(2)可知$m_{ni} \xrightarrow{a.s.} \mu_i$, $i=2,3,\cdots$.再由G是其变元的连续函数, 利用事实(A), 立即可得$\hat{g}_n(\mathbf{X}) \xrightarrow{a.s.} g(\theta)$.

由这一定理可得出一些常见估计的相合性. 例如正态总体$N(a,\sigma^2)$中,样本均值$\bar{X}$和样本方差S^2分别是a和σ^2的强相合估计. 也不难证明S^2是σ^2的均方相合估计. 其实对任何$r>0$, S^2是σ^2的r阶矩相合估计. 例3.2.5中定义的偏度、峰度和变异系数的矩估计都是强相合的.

2. 矩估计的渐近正态性

本段我们将在很一般的条件下, 给出矩估计是相合渐近正态估计. 下面首先给出定义

定义 2. *设$\mathbf{X}=(X_1,\cdots,X_n)$是从总体$\{F_\theta,\ \theta\in\Theta\}$中抽取的简单样本, $\hat{g}_n(\mathbf{X})$是$g(\theta)$的矩估计量. 若存在与样本大小n有关的, 定义于参数空间Θ上的函数$A_n(\theta)$和$B_n(\theta)$, 其中$B_n(\theta)$在Θ上处处大于0,使当$n\to\infty$时*

$$(\hat{g}_n(\mathbf{X}))-A_n(\theta))\,/B_n(\theta) \xrightarrow{\mathscr{L}} N(0,1)$$

且$\hat{g}(\mathbf{X})$为$g(\theta)$的弱相合估计, 则称$\hat{g}(\mathbf{X})$为$g(\theta)$的相合渐近正态估计 (Consistent Asymptotic Normal Estimation, 简记为CAN估计).

就是说CAN估计是既相合, 其分布又渐近服从正态分布的那种估计. 本段我们提出两个重要结果,即在很一般的条件下,矩估计为CAN估计.

> **Delta 方法** 假设$\phi: D\subset R^k \longmapsto R^m$ 为一定义在R^k的子集D上的映射, 且在$\theta(\theta\in R^k)$处可微. 假设T_n为一列取值于D上的随机向量, 且存在正的趋于无穷的常数数列r_n使得$r_n(T_n-\theta)\overset{d}{\longmapsto} T$, 则
>
> $$r_n[\phi(T_n)-\phi(\theta)] \overset{d}{\longmapsto} \phi'(\theta)T.$$
>
> 其中$\phi'(\theta)$为函数ϕ在θ处的$m\times k$的导数矩阵.

特别, 若$T\sim N(0,A)$, 则可以得出$\phi'(\theta)T\sim N(0,\phi'(\theta)A[\phi'(\theta)]^T)$. 使用此方法可以证明如下结论.

(1) 设样本$X_1,\cdots,X_n$为从总体$\{F_\theta,\ \theta\in\Theta\}$ 中抽取的简单样本, $g(\theta)$是定义在参数空间Θ上的实函数, 它可以表为形式: $g(\theta)=G(\alpha_1,\cdots,\alpha_k)$ (若G 是$\alpha_1,\cdots,\alpha_k,\mu_1,\cdots,\mu_s$的函数, 不妨令$s\le k$,可将$\mu_j$用$\alpha_1,\alpha_2,\cdots,\alpha_s$表出, 则$G$仍可表为$\alpha_1,\cdots,\alpha_k$的函数), 而$\hat{g}(\mathbf{X})=\hat{g}(X_1,\cdots,X_n)=G(a_{n1},\cdots,a_{nk})$为$g(\theta)$的矩估计. 再设总体的$2k$阶原点矩存在, 且$G$ 对其各变元的一阶偏导数存在、连续, 令

$$\begin{aligned}
&b_{ij}=\alpha_{i+j}-\alpha_i\alpha_j,\quad i,j=1,2,\cdots,k;\qquad B=\big(b_{ij}\big)\ \text{为}\ k\times k\ \text{的方阵},\\
&d_i=\frac{\partial G(\alpha_1,\cdots,\alpha_k)}{\partial\alpha_i},\quad i=1,2,\cdots,k;\quad d=(d_1,\cdots,d_k)'\\
&b^2=d'Bd.
\end{aligned}$$

定理 2. *在上述记号和条件下,$\hat{g}_n(\mathbf{X})$为$g(\theta)=G(\alpha_1,\cdots,\alpha_k)$的CAN 估计. 即$\hat{g}_n(\mathbf{X})$为$g(\theta)$的弱相合估计,且有*

$$\sqrt{n}\big(\hat{g}_n(\mathbf{X})-G(\alpha_1,\cdots,\alpha_k)\big) \xrightarrow{\mathscr{L}} N(0,b^2),\quad \text{当}n\to\infty\ \text{时}$$

证 由定理3.2.1可知$\hat{g}_n(\mathbf{X})$ 为$g(\theta)=G(\alpha_1,\cdots,\alpha_k)$ 的强相合估计. 其渐近正态性的证明见参考文献[2] P_{82} 定理2.6.

(2) 在一些情况下, $g(\theta)$ 可表为一、两个中心矩的函数, 还可能包含总体均值α_1, $g(\theta)$ 的表达式较简单. 若将中心矩用原点矩表出, $g(\theta)$ 的表达式则显得复杂, 因此有必要给出这种情形下渐近正态性的结果. 一般, 将$g(\theta)$ 表达成如下形式

$$g(\theta)=H(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r}), \tag{3.14}$$

其矩估计量为$H(\bar{X},m_{nt_2},\cdots,m_{nt_r})$,使用与定理3.2.2类似证明方法, 可得如下结果:

定理 3. *设(3.14)式中的函数H 在点$(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r})$ 的邻域内有一阶偏导数, 且此偏导数在点$(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r})$ 处连续, 则有*

$$\sqrt{n}\left(H(\bar{X},m_{nt_2},\cdots,m_{nt_r})-H(\alpha_1,\mu_{t_2},\cdots,\mu_{t_r})\right)\xrightarrow{\mathscr{L}}N(0,b^2). \tag{3.15}$$

此处

$$b^2=\sum_{i=1}^{r}\sum_{j=1}^{r}\sigma_{ij}H_iH_j, \tag{3.16}$$

其中

$$\begin{aligned}
&H_1=\frac{\partial H}{\partial\alpha_1},\quad H_i=\frac{\partial H}{\partial\mu_{t_i}},\quad i=2,3,\cdots,r;\\
&\sigma_{11}=\mu_2,\quad \sigma_{1i}=\sigma_{i1}=\mu_{t_i+1}-t_i\mu_{t_i-1}\mu_2,\quad i=2,\cdots,r;\\
&\sigma_{ij}=\mu_{t_i+t_j}-t_i\mu_{t_i-1}\mu_{t_j+1}-t_j\mu_{t_i+1}\mu_{t_j-1}-\mu_{t_i}\mu_{t_j}\\
&\qquad+t_it_j\mu_2\mu_{t_i-1}\mu_{t_j-1},\quad i,j=2,3,\cdots,r.
\end{aligned}$$

如果$g(\theta)$ 有$H(\mu_{t_2},\cdots,\mu_{t_r})$ 的形状, 即与α_1 无关, 则(3.15)仍成立, 只需把(3.16)式所确定的b^2 改为$\sum\limits_{i=2}^{r}\sum\limits_{j=2}^{r}\sigma_{ij}H_iH_j$ 即可.

例8. *继续考虑例3.2.5.被估计量$g(\theta)$ 为偏度β_1,峰度β_2 和变异系数V,其矩估计量由例3.2.6可知为*

$$\hat{\beta}_1=\frac{m_{n3}}{m_{n2}^{3/2}},\quad \hat{\beta}_2=\frac{m_{n4}}{m_{n2}^2},\quad \hat{V}=\sqrt{m_{n2}}/\,\bar{X}.$$

讨论它们的渐近正态性.

解 按(3.15) 和(3.16)式, 对这三个矩估计量分别算得b^2 之值为

$$b^2(\beta_1)=6,\quad b^2(\beta_2)=24,\quad b^2(V)=V^2/2+V^4.$$

于是根据定理3.2.3有

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_1-\beta_1)\xrightarrow{\mathscr{L}}N(0,6),\\
&\sqrt{n}(\hat{\beta}_2-\beta_2)\xrightarrow{\mathscr{L}}N(0,24),\\
&\sqrt{n}(\hat{V}-V)\xrightarrow{\mathscr{L}}N\big(0,V^2/2+V^4\big).
\end{aligned}$$

值得注意的是$\hat{\beta}_1$, $\hat{\beta}_2$ 的极限分布的方差与被估计的参数值无关, 这点对β_1, β_2 的大样本推断有用.