目 录

第六章　参数估计

教学目的:

1) 让学生理解矩估计和极大似然估计方法.

2) 理解置信区间定义.

3) 掌握常见的总体分布下参数的点估计和置信区间的计算.

设有一个总体, 以$f(x;\theta_1,\cdots,\theta_k)$记其概率密度函数(若总体分布是连续性的), 或其概率函数(若总体分布为离散型的). 为叙述方便我们统一称$f(x;\theta_1,\cdots,\theta_k)$为总体的概率函数. 参数估计问题是利用从总体抽样得到的信息来估计总体的某些参数或者参数的某些函数. 一般假定总体分布形式已知，未知的仅仅是一个或几个参数. 利用从总体$f(x;\theta_1,\cdots,\theta_k)$中抽取的一组样本$X_1,\cdots,X_n$去对参数$\theta_1,\cdots,\theta_k$的未知值作出估计或估计它们的某个已知函数$g(\theta_1,\cdots,\theta_k)$.

§6.1　点估计

设总体X的分布函数形式已知, 但它的一个或多个参数为未知, 例如参数θ未知, 根据样本$X_1,\cdots,X_n$来估计参数θ, 就是要构造适当的统计量$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$. 当有了样本$X_1,\cdots,X_n$的值后,就代入$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$中算出一个值,用来作为$\theta$的估计值. 为这样特定目的而构造的统计量$\hat{\theta}$叫做$\theta$的**估计量**. 由于参数$\theta$是数轴上的一个点, 用$\hat{\theta}$估计$\theta$,等于用一个点去估计另一个点, 所以这样的估计叫做**点估计**.

求点估计的方法有多种, 下面介绍两种点估计方法:

§6.1.1　矩估计方法

矩方法追溯到19世纪的Karl Pearson. 矩方法是基于一种简单的“替换”思想建立起来的一种估计方法. 其基本思想是用样本矩估计总体矩. 由大数律，如果未知参数和总体的某个(些)矩有关系，我们很自然的来构造未知参数的估计。

回忆一下以前关于矩的记法：

$$\text{样本}k\text{阶矩:}\quad a_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \qquad m_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$$

$$\text{总体}k\text{阶矩:}\quad \alpha_k = EX^k \qquad \mu_k = E(X-EX)^2$$

因此在k阶矩存在的情况下，根据大数律有

$$a_k \xrightarrow{p} \alpha_k, \quad m_k \xrightarrow{p} \mu_k$$

从而我们可以使用a_k, m_k分别估计α_k, μ_k。介绍如下: 假设总体X包含k个未知参数$\theta_1, \cdots, \theta_k$, 由方程组

$$\begin{cases} \alpha_1 = f_1(\theta_1, \cdots, \theta_k) \\ \qquad\vdots \\ \alpha_k = f_k(\theta_1, \cdots, \theta_k) \end{cases}$$

反解得到

$$\begin{cases} \theta_1 = g_1(\alpha_1, \cdots, \alpha_k) \\ \qquad\vdots \\ \theta_k = g_k(\alpha_1, \cdots, \alpha_k) \end{cases}$$

将其中的总体矩用相应的样本矩代替，则我们可以得到参数$\theta_1, \cdots, \theta_k$的一个估计:

$$\begin{cases} \hat{\theta}_1 = g_1(a_1, \cdots, a_k) \\ \qquad\vdots \\ \hat{\theta}_k = g_k(a_1, \cdots, a_k) \end{cases}$$

若要估计参数$\theta_1, \cdots, \theta_k$的某函数$g(\theta_1, \cdots, \theta_k)$, 则用$g(\hat{\theta}_1, \cdots, \hat{\theta}_k)$去估计它.

这里我们用的都是原点矩α_k，当然也可以使用中心矩μ_k，或者两个都使用。在这种情况下，只需要把相应的总体矩换成样本矩。我们称这种估计方法为矩估计法，得到的估计量称为矩估计量。**矩估计方法应用的原则是：能用低阶矩处理的就不用高阶矩。**

矩估计法的优点是简单易行,并不需要事先知道总体是什么分布. 缺点是，当总体类型已知时，没有充分利用分布提供的信息. 一般场合下, 矩估计量不具有唯一性.

例 6.1.1. *投掷一枚硬币, 为了解正面出现的概率, 现独立重复的投掷n次, 用$X_1, \cdots, X_n$表示投掷结果. 显然此时总体X的分布为$B(1,p)$, p为感兴趣的量. 而$X_1, \cdots, X_n$为样本, 则求参数p的矩估计量。*

解: 由于$EX = p$，而样本均值$\bar{X}$收敛到总体均值EX, 因此p的一个矩估计量为$\hat{p} = \bar{X}$.

例 6.1.2. *为考察某种考试成绩分布情况, 使用正态分布$N(a, \sigma^2)$来作为总体X的分布. 现在从中随机调查n个人, 即样本为$X_1, \cdots, X_n$. 试求参数a, σ^2的矩估计量。*

解: 由于

$$EX = a, \quad Var(X) = \sigma^2$$

所以a, σ^2的一个矩估计量为

$$\hat{a} = \bar{X}, \quad \hat{\sigma}^2 = m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

我们知道$ES^2 = \sigma^2$，因此，σ^2的另一个矩估计量为$\hat{\sigma}^2 = S^2$.

§6.1.2 极大似然估计方法

极大似然方法到目前为止应用最广的的点估计方法. 这种方法是基于如下的看法:

定义 6.1.1. *设总体X 有概率函数$f(x;\theta) = f(x;\theta_1,\cdots,\theta_k)$，这里参数$\theta = (\theta_1,\cdots,\theta_k) \in \Theta$，而当固定$x$时把$f(x;\theta)$看成为$\theta$的函数，称为似然函数, 常记为$L(x;\theta)$或$L(\theta)$.*

当固定参数θ时，$f(x;\theta)$可以看成是得到样本观察值x的可能性，这样，当把参数θ 看成变动时，也就得到"在不同的θ值下能观察到x的可能性大小, 即$L(x;\theta)$"；由于我们已经观察到了x，所以我们要寻求在哪一个θ的值下，使得能观察到x的可能性$L(x;\theta)$最大。这个θ的值即称为θ 极大似然估计值(看上去最有可能的)。我们先看一个例子:

例 6.1.3. *从鱼池里随机捕捞500条鱼, 做好记号后重新放入鱼池中, 待充分混合后再捕捞1000条鱼, 结果发现其中有72条带有记号. 试问鱼池中可能有多少条鱼.*

解: 先将问题一般化. 设池中有N条鱼, 其中r条做好记号. 随机捕捞s条, 发现x条有记号. 用上述信息来估计N.

用X表示捕捞的s条鱼中带记号鱼的数目, 则

$$P(X = x) = \frac{C_{N-r}^{s-x} C_r^x}{C_N^s}.$$

目前发现在捕捞的s条鱼中有记号的鱼x条, 要寻求N取何值时, 使得观察到这个事件$\{X = x\}$的可能性最大. 即x是固定的, N是变化的, 记$p(x;N) = P(X = x)$. 因为

$$g(N) := \frac{p(x;N)}{p(x;N-1)} = \frac{(N-s)(N-r)}{N(N-r-s+x)} = \frac{N^2 - N(s+r) + rs}{N^2 - N(r+s) + Nx},$$

当$rs > Nx$时, $g(N) > 1$; $rs < Nx$时, $g(N) < 1$. 所以$P(X = x)$在$N = \frac{rs}{x}$附近达到最大, 注意到N 只能取正整数, 故N的最可能的估计即极大似然估计为

$$\hat{N} = \left\lceil \frac{rs}{x} \right\rceil.$$

其中$\lceil \quad \rceil$表示下取整, 即小于该值的最大整数. 将题目中的数字代入,

$$\hat{N} = \left\lceil \frac{500 \times 1000}{72} \right\rceil = 6944.$$

即鱼池中的总的鱼数为6694条.

现给出极大似然估计的一般性定义:

定义 6.1.2. *设$X = (X_1, \cdots, X_n)$为从具有概率函数f的总体中抽取的样本, θ为未知参数或者参数向量. $x = (x_1, \cdots, x_n)$为样本的观察值。若在给定x时, 值$\hat{\theta} = \hat{\theta}(x)$满足下式*

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(x; \theta)$$

则称$\hat{\theta}$为参数θ的极大似然估计值, 而$\hat{\theta}(X)$称为参数θ的极大似然估计量。若待估参数为θ的函数$g(\theta)$, 则$g(\theta)$的极大似然估计量为$g(\hat{\theta})$。

求极大似然估计值相当于求似然函数的最大值。在简单样本的情况下,

$$L(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

而把似然函数的对数$l(\theta) = \log L(\theta)$称为对数似然函数(这是由于在一些情况下，处理对数似然函数更方便)

当似然函数对变量θ单调时, 我们可以容易得到其最大值点. 反之当似然函数为非单调函数且对变量θ可微分时，我们可以求其驻点: 令

$$\frac{dl(\theta)}{d\theta} = 0 \quad (\text{或者} \frac{dL(\theta)}{d\theta} = 0)$$

当θ为多维时, 比如$\theta = (\theta_1, \cdots, \theta_k)$时令

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0 \quad (\text{或者} \frac{\partial L(\theta)}{\partial \theta_i} = 0) \; i = 1, \cdots, k$$

然后判断此驻点是否是最大值点。

例 6.1.4. 设$X_1, \cdots, X_n$为从总体$X \sim N(a, \sigma^2)$中抽取的样本，求参数a, σ^2的极大似然估计量。

解: 易得对数似然函数为

$$l(a, \sigma^2) = c - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2 - \frac{n}{2}log(\sigma^2)$$

其中c是与参数无关的常数. 令

$$\begin{cases} \frac{\partial l(a,\sigma^2)}{\partial a} = 0 \\ \frac{\partial l(a,\sigma^2)}{\partial \sigma^2} = 0 \end{cases}$$

得到

$$\begin{cases} a = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \sigma^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - a)^2 \end{cases}$$

容易验证此驻点是唯一的最大值点，因此得到a, σ^2的极大似然估计量:

$$\begin{cases} \hat{a} = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2. \end{cases}$$

有时函数f并不对$\theta_1, \cdots, \theta_k$可导, 甚至$f$本身也不连续, 这时求导就没法用, 必须回到原始定义.

例 6.1.5. 设总体X服从$[a, b]$上的均匀分布, $a < b$, 求参数a, b的极大似然估计.

解: 易得似然函数为

$$L(a, b) = \frac{1}{(b-a)^n}\prod_{j=1}^{n} I(a \le x_j \le b) = \frac{1}{(b-a)^n} I(a \le x_{(1)} \le x_{(n)} \le b).$$

于是对任何满足条件$a \le x_j \le b$的a, b都有

$$L(a, b) = \frac{1}{(b-a)^n} \le \frac{1}{(x_{(n)} - x_{(1)})^n},$$

即似然函数$L(a, b)$在$a = x_{(1)}, b = x_{(n)}$时取到最大值. 于是a, b的极大似然估计量为$\hat{a} = X_{(1)}, \hat{b} = X_{(n)}$.

例 6.1.6. 设$X_1,\cdots,X_n$为从具有如下形式密度的总体中抽取的样本:

$$f(x;a,b)=\begin{cases}\frac{1}{b}\exp\{-\frac{x-a}{b}\} & ,x>a\\ 0 & ,x\le a\end{cases}$$

求参数a,b的极大似然估计量.

解: 易得似然函数为

$$L(a,b)=\prod_{i=1}^{n}f(x_i;a,b)=\frac{1}{b^n}\exp\{-\frac{1}{b}\sum_{i=1}^{n}(x_i-a)\}I(x_{(1)}>a)$$

在固定b时，显然似然函数为a的单调增函数，因此$L(a)$的驻点为$\hat{a}=x_{(1)}$。再令$\frac{\partial L(a,b)}{\partial b}=0$，得到$b=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-x_{(1)})$，容易验证此解是最大值点。从而得到$a,b$的极大似然估计量:

$$\begin{cases}\hat{a}=X_{(1)}\\ \hat{b}=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-X_{(1)}).\end{cases}$$

例 6.1.7. 设$X_1,\cdots,X_n$为从如下分布中抽取的简单样本，求θ的极大似然估计.

$$f(x)=\frac{1}{x!(2-x)!}[\theta^x(1-\theta)^{2-x}+\theta^{2-x}(1-\theta)^x],\quad x=0,1,2;\quad \theta\in(0,\frac{1}{2})$$

解: 由题设知$f(x)$为离散型，其分布律为

X	0	1	2
P	$\frac{1}{2}[(1-\theta)^2+\theta^2]$	$2\theta(1-\theta)$	$\frac{1}{2}[(1-\theta)^2+\theta^2]$

若直接从此分布出发，则不能得到θ的极大似然估计的显式表达。为此，我们重新参数化，记$\eta=2\theta(1-\theta)$. 则由题设知$\eta<1/2$。则

X	0	1	2
P	$\frac{1}{2}(1-\eta)$	η	$\frac{1}{2}(1-\eta)$

再记$n_i=\#\{X_1,\cdots,X_n$中等于i的个数$\}$, $\quad i=0,1,2,$ 则得到似然函数为

$$L(\eta)=(\frac{1}{2}(1-\eta))^{n_0}\eta^{n_1}(\frac{1}{2}(1-\eta))^{n_2}=(\frac{1}{2}(1-\eta))^{n-n_1}\eta^{n_1}$$

求解并注意η的上界即得到η的极大似然估计为

$$\hat{\eta}=\min\{\frac{n_1}{n},\frac{1}{2}\}$$

再由$\theta=\frac{1-\sqrt{1-2\eta}}{2}$得到$\theta$的极大似然估计为

$$\hat{\theta}=\frac{1-\sqrt{1-2\hat{\eta}}}{2}$$

§6.1.3 点估计的优良准则

我们看到对同一个参数，有多个不同的估计量，因此，评选不同估计量的优劣性是需要考虑的。

1. 无偏性

设$\hat{g}(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的一个估计量，若

$$E\hat{g}(X_1,\cdots,X_n)=g(\theta)$$

则称$\hat{g}(X_1,\cdots,X_n)$为$g(\theta)$的无偏估计量。无偏性是对一个估计量的最基本的要求, 其实际意义就是无系统误差. 因此在有多个估计量可供选择时，我们优先考虑无偏估计量。

很多时候我们得到的估计量是有偏, 例如正态总体的方差σ^2的极大似然估计量$\hat{\sigma}^2=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$是有偏的, $E\hat{\sigma}^2=\frac{n-1}{n}\sigma^2$. 若以$\frac{n}{n-1}$乘以$\hat{\sigma}^2$, 所得到的估计量就是无偏的. 这种方法称为修正.

若某一参数存在多个无偏估计时, 如何来选择使用哪个估计量? 人们又在无偏性的基础上增加了对方差的要求.

2. 有效性

设$\hat{g}_1(X_1,\cdots,X_n)$和$\hat{g}_2(X_1,\cdots,X_n)$为待估参数函数$g(\theta)$的两个不同的无偏估计量，若对任意的$\theta\in\Theta$,有

$$Var(\hat{g}_1(X_1,\cdots,X_n))\le Var(\hat{g}_2(X_1,\cdots,X_n))$$

而且至少对某个$\theta_0\in\Theta$使得严格不等式成立。则称$\hat{g}_1$较$\hat{g}_2$有效。

3. 相合性

设总体分布依赖于参数$\theta_1,\cdots,\theta_k$, $g(\theta_1,\cdots,\theta_k)$是待估参数函数。设$X_1,\cdots,X_n$为自该总体中抽取的样本，$T(X_1,\cdots,X_n)$为$g(\theta_1,\cdots,\theta_k)$的一个估计量，如果对任意的$\epsilon>0$和$\theta_1,\cdots,\theta_k$的一切可能值都有

$$\lim_{n\to\infty}P_{\theta_1,\cdots,\theta_k}(|T(X_1,\cdots,X_n)-g(\theta_1,\cdots,\theta_k)|\ge\epsilon)=0$$

我们则称$T(X_1,\cdots,X_n)$为$g(\theta_1,\cdots,\theta_k)$的一个(弱)相合估计量。

相合性是对一个估计量的最基本的要求，如果一个估计量没有相合性，那么无论样本大小多大，我们也不能把未知参数估计到任意预定的精度。这种估计量显然是不可取的。

矩估计量是满足相合性的，极大似然估计量在很一般的条件下也是满足相合性的。

4. 渐近正态性

估计量是样本$X_1, \cdots, X_n$的函数，其确切的分布一般不是容易得到。但是，许多形式很复杂的统计量(未必是和)，当n很大时，其分布都渐近于正态分布，这个性质称为统计量的"渐近正态性"。

无偏性和有效性都是对固定的样本大小n而言的，这种性质称为估计量的"小样本性质"，而相合性和渐近正态性都是考虑在样本大小趋于无穷时的性质，这种性质称为"大样本性质"。

例 6.1.8. 设从总体

X	0	1	2	3
P	$\theta/2$	θ	$3\theta/2$	$1-3\theta$

抽取的一个简单样本$X_1, \cdots, X_{10}$的观察值为$(0,3,1,1,0,2,0,0,3,0)$，

(1) 求θ的矩估计量$\hat{\theta}_M$和极大似然估计量$\hat{\theta}_L$，并求出估计值。

(2) 上述估计量是否为无偏的？若不是，请作修正.

(3) 比较修正后的两个估计量，指出那个更有效.

由有效性的定义，我们自然会问在一切可能的无偏估计里，能否找到具有最小方差的无偏估计量？如果存在这样的估计量，我们称其为最小方差无偏估计量, 详细地可以参考课本。