第六章 参数估计

参数估计是统计推断的基本问题之一. 在实际问题中, 总体X的分布函数的形式已知, 但它包含未知参数.

↓

概率分布函数的确切表达式不定

↓

由总体样本的一组观测值估计未知参数

↓

利用样本信息估计总体参数的问题

↓

参数估计问题

↓

点估计　　区间估计

第一节 参数的点估计

本节内容:

一. 矩估计法

二. 极大似然估计法

HIGH EDUCATION PRESS

一. 参数的点估计

定义 设总体$F(x)$的分布函数$F(x;\theta)$形式已知, 其中θ是待估计的参数,点估计问题就是利用样本构造一个统计量$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$来估计$\theta$,我们称

$$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$$

为θ的点估计量,它是一个随机变量.

对于一个给定的样本观察值$(x_1,x_2,\cdots,x_n)$,代入估计量得到一个具体的值$\hat{\theta}(x_1,x_2,\cdots,x_n)$,称其为$\theta$的点估计值.

参数估计

- 制定求估计量的一般方法
- 制定评价估计量优良的准则
- 在某种特定准则下，求最优估计量
- 证明某一估计量在某准则下的最优性

参数点估计的常用方法

- 矩估计法
- 极大似然估计法

二. 矩估计法

总体分布中的参数与其数字特征之间存在密切的关系.

而矩是随机变量最基本的数字特征.

总体分布中的参数与总体矩之间必有联系.

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{\infty} X_i^k - E(X^k)\right| < \varepsilon\right\} = 1$$

用样本矩作为总体矩的估计

矩估计法

- 设$X_1,X_2,\cdots,X_n$是连续总体X的样本,其分布函数为

$$F(x;\theta_1,\theta_2,\cdots,\theta_k)$$

其中$\theta_1,\theta_2,\cdots,\theta_k$是待估参数. 假设总体$X$的前$k$阶矩

$$\mu_l=E(X^l)=\int_{-\infty}^{\infty}x^l f(x;\theta_1,\theta_2,\cdots,\theta_k)dx$$

存在. 它是$\theta_1,\theta_2,\cdots,\theta_k$的函数.

基于样本矩

$$A_l=\frac{1}{n}\sum_{i=1}^{\infty}X_i^l$$

以概率收敛于相应的总体矩μ_l,其中$l=1,2,\cdots,k$.

我们可用样本矩作为相应总体矩的估计量

用样本矩的连续函数作为相应总体的连续函数的估计量

矩估计的具体做法:

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_k) \\ \mu_2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_k) \\ \qquad \vdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_k) \end{cases}$$

求得方程组的解

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \cdots, \mu_k) \\ \theta_2 = \theta_2(\mu_1, \mu_2, \cdots, \mu_k) \\ \qquad \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \cdots, \mu_k) \end{cases}$$

分别用$A_1, A_2, \cdots, A_k$代替上述$\theta_1, \theta_2, \cdots, \theta_k$得

$$\begin{cases} \hat{\theta}_1 = \theta_1(A_1, A_2, \cdots, A_k) \\ \hat{\theta}_2 = \theta_2(A_1, A_2, \cdots, A_k) \\ \qquad \vdots \\ \hat{\theta}_k = \theta_k(A_1, A_2, \cdots, A_k) \end{cases}$$

我们称$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$为$\theta_1, \theta_2, \cdots, \theta_k$的矩估计量.

- 设$X_1, X_2, \cdots, X_n$是离散总体X的样本，其分布律为

$$P\{X=x\}=p(x;\theta_1,\theta_2,\cdots,\theta_k)$$

其中$\theta_1,\theta_2,\cdots,\theta_k$是待估参数. 假设总体$X$的前$k$阶矩

$$\mu_l=E(X^l)=\sum_{x\in R_X} x^l f(x;\theta_1,\theta_2,\cdots,\theta_k)$$

存在. 它是$\theta_1,\theta_2,\cdots,\theta_k$的函数.

基于样本矩

$$A_l=\frac{1}{n}\sum_{i=1}^{\infty} X_i^l$$

以概率收敛于相应的总体矩μ_l, 其中$l=1,2,\cdots,k$.

矩估计的具体做法:

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_k) \\ \mu_2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_k) \\ \qquad \vdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_k) \end{cases}$$

求得方程组的解

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \cdots, \mu_k) \\ \theta_2 = \theta_2(\mu_1, \mu_2, \cdots, \mu_k) \\ \qquad \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \cdots, \mu_k) \end{cases}$$

分别用$A_1, A_2, \cdots, A_k$代替上述$\theta_1, \theta_2, \cdots, \theta_k$得

$$
\begin{cases}
\hat{\theta}_1 = \theta_1(A_1, A_2, \cdots, A_k) \\
\hat{\theta}_2 = \theta_2(A_1, A_2, \cdots, A_k) \\
\qquad \vdots \\
\hat{\theta}_k = \theta_k(A_1, A_2, \cdots, A_k)
\end{cases}
$$

我们称$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$为$\theta_1, \theta_2, \cdots, \theta_k$的矩估计量.

例1 设总体X在$[a,b]$上服从均匀分布，且a,b未知. $X_1, X_2, \cdots, X_n$是来自X的样本,试求a,b的矩估计量.

解 首先求解总体X的前2阶矩

$$\begin{cases} \mu_1 = E(X) = \dfrac{a+b}{2} \\ \mu_2 = E(X^2) = D(X) + [E(X)]^2 = \dfrac{(b-a)^2}{12} + \dfrac{(a+b)^2}{4} \end{cases}$$

也就是说$\begin{cases} a+b = 2\mu_1 \\ b-a = \sqrt{12(\mu_2 - \mu_1^2)} \end{cases}$

求得方程组的解

$$a=\mu_1-\sqrt{3(\mu_2-\mu_1^2)} \quad b=\mu_1+\sqrt{3(\mu_2-\mu_1^2)}$$

分别用A_1, A_2代替μ_1, μ_2得

$$\hat{a}=A_1-\sqrt{3(A_2-A_1^2)}=\overline{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

$$\hat{b}=A_1+\sqrt{3(A_2-A_1^2)}=\overline{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

故$\hat{a}, \hat{b}$就是a, b的矩估计量.

例2 设总体X的均值μ和方差σ^2都存在，且有$\sigma^2>0$. 但μ,σ^2都未知. 又设$X_1,X_2,\cdots,X_n$是来自总体X的样本. 试求μ,σ^2的矩估计量.

解 首先求解总体X的前2阶矩

$$\begin{cases}\mu_1=E(X)=\mu \\ \mu_2=E(X^2)=D(X)+[E(X)]^2=\sigma^2+\mu^2\end{cases}$$

求得方程组的解

$$\mu=\mu_1 \qquad \sigma^2=\mu_2-\mu_1^2$$

分别用A_1,A_2代替μ_1,μ_2得

$$\hat{\mu}_1 = A_1 = \overline{X}$$

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{\infty} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{\infty}(X_i - \overline{X})^2$$

注: 总体均值与方差的矩估计量的表达式不因总体分布的变化而变化.

三. 极大似然估计法

- 设$X_1,X_2,\cdots,X_n$是离散总体X的样本　总体X的分布律

$$P\{X=x\}=p(x;\theta),\quad \theta\in\Theta$$

的形式为已知. θ是待估参数, 其中Θ是θ的取值范围.

样本$X_1,X_2,\cdots,X_n$的联合分布律为

$$\prod_{i=1}^{\infty}p(x_i;\theta)$$

构造函数

$$L(\theta)=L(\underbrace{x_1,x_2,\cdots,x_n}_{\text{常数}};\underbrace{\theta}_{\text{待估参数}})=\prod_{i=1}^{\infty}p(x_i;\theta)$$

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{\infty} p(x_i; \theta)$$

$L(\theta)$样本$X_1, X_2, \cdots, X_n$的似然函数

当$\theta = \theta_0 \in \Theta$时, $L(\theta)$的取值很大,
而其余$\theta \in \Theta$使得$L(\theta)$取值较小

取θ_0作为未知参数θ的估计值较为合理

极大似然估计思想

对任一固定的样本值$x_1,x_2,\cdots,x_n$，在θ取值的可能范围Θ内挑选使得似然函数

$$L(x_1,x_2,\cdots,x_n;\theta)$$

达到最大的参数值$\hat{\theta}$作为θ的估计值. 即取$\hat{\theta}$使

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$$

这样的$\hat{\theta}$是样本值$x_1,x_2,\cdots,x_n$的函数，记作$\hat{\theta}(x_1,x_2,\cdots,x_n)$，并称为参数$\theta$的极大似然估计值. 而相应的统计量

$$\hat{\theta}(X_1,X_2,\cdots,X_n)$$

称为参数θ的极大似然估计量.

- 设$X_1,X_2,\cdots,X_n$是连续总体X的样本 总体X的概率密度

$f(x;\theta)$的形式已知,其中$\theta\in\Theta$是待估参数.

样本$X_1,X_2,\cdots,X_n$的联合概率密度为

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

若$x_1,x_2,\cdots,x_n$是样本的一个样本值, 则$(X_1,X_2,\cdots,X_n)$落在点$(x_1,x_2,\cdots,x_n)$的邻域内的概率近似为

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

其值随θ的变化而变化.

我们取θ的估计值$\hat{\theta}$使概率

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

取到最大值，但因子$\prod_{i=1}^{n} dx_i$不受θ的影响. 因此只考虑

$$L(\theta) = L(x_1, x_2, \cdots, x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

的最大值.此时将$L(\theta)$称为样本的似然函数.

我们称满足

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$$

的 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为 θ 极大似然估计值. 而相应的统计量

$$\hat{\theta}(X_1,X_2,\cdots,X_n)$$

称为参数 θ 的极大似然估计量.

例3 设总体X服从正态分布$N(\mu,\sigma^2)$,其中μ和σ^2为未知参数.$X_1,X_2,\cdots,X_n$是来自X的一个样本.试求μ和σ^2的极大似然估计量.

解 总体X的分布密度为

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

构造似然函数

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$=(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

对上式两边取对数

$$\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

则 $$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0\\[2ex]\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$$

解上述方程组得

$$\begin{cases} \hat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \end{cases}$$

故μ和σ^2的极大似然估计量为

$$\begin{cases} \hat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 \end{cases}$$

与例2的结论比较

例4 设总体X在$[a,b]$服从均匀分布，a,b未知，$x_1,x_2,\cdots,x_n$是一个样本值. 试求a,b的极大似然估计量.

解 记 $x_{(1)}=\min\{x_1,x_2,\cdots,x_n\}$，$x_{(2)}=\max\{x_1,x_2,\cdots,x_n\}$，

已知总体X的概率密度是

$$f(x;a,b)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{其他}\end{cases}$$

似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n} & a\le x_1,x_2,\cdots,x_n\le b\\ 0 & \text{其他}\end{cases}$$

由于$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_{(1)}, x_{(n)} \le b$.

故似然函数可改写为

$$L(a,b) = \begin{cases} \dfrac{1}{(b-a)^n} & a \le x_{(1)}, x_{(n)} \le b \\ 0 & \text{其他} \end{cases}$$

于是对满足条件$a \le x_{(1)}, x_{(2)} \le b$的任意$a, b$有

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

即$L(a,b)$在$a = x_{(1)}, b = x_{(n)}$时取到最大值. 故a, b的极大似然估计值为

$$\hat{a} = x_{(1)} = \min_{1 \le i \le n} x_i \qquad \hat{b} = x_{(n)} = \max_{1 \le i \le n} x_i$$

进而a,b的极大似然估计量为

$$\hat{a} = X_{(1)} = \min_{1 \le i \le n} X_i \qquad \hat{b} = X_{(n)} = \max_{1 \le i \le n} X_i$$

Homework

P135: 2+3(双号)+5+6+10