9 参数估计

设总体X的分布函数$F(X,\theta)$, 其中θ为未知参数(也可为向量). 现从总体中抽取一样本$X_1,X_2,\cdots,X_n$, 如何依据样本估计参数θ, 或θ的函数$g(\theta)$, 此类问题称为参数估计问题. 内容包括: 点估计, 估计量标准, 区间估计.

9.1 点估计

9.1.1 矩估计法

总体X的k阶矩: $a_k=E[X^k]$

总体X的k阶中心矩: $b_k=E[(X-E(X))^k]$

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

样本k 阶中心矩: $B_k=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^k$

用相应的样本矩去估计总体矩, 从而求解参数θ的方法称为矩估计法. 矩估计法的理论基础是大数定理: $X_1,X_2,\cdots,X_n$为i.i.d.的随机变量, 若$E(X)=\mu$, 则当$n\to\infty$时, 有

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

推论: 若$E[X^k]=a_k$存在, 则当$n\to\infty$时, 有

$$\frac{1}{n}\sum_{i=1}^n X_i^k \xrightarrow{P} a_k=E[X^k].$$

矩估计方法: 总体X的分布函数F包含m个未知参数$\theta_1,\theta_2,\cdots,\theta_m$:

1) 求总体X的k阶矩: $a_k=a_k(\theta_1,\theta_2,\cdots,\theta_m)=E[X^k],\quad k\in[m]$. ($a^k$一般为$\theta_1,\theta_2,\cdots,\theta_m$ 的函数)
2) 计算样本的k阶矩: $A_k=\frac{1}{n}\sum_{i=1}^n X_i^k$.
3) 令样本矩等于总体矩$A^k=a^k=a^k(\theta_1,\theta_2,\cdots,\theta_m)\quad(k=1,2,\cdots,m)$, 得到$m$个关于$\theta_1,\theta_2,\cdots,\theta_m$的方程组.
4) 求解方程组得到估计量$\widehat{\theta}_1,\widehat{\theta}_2,\cdots,\widehat{\theta}_n$.

例9.1. 设总体X的概率密度函数

$$f(x)=\begin{cases}(\alpha+1)x^\alpha & x\in(0,1)\\ 0 & \text{其它}\end{cases}$$

$X_1,X_2,\cdots,X_n$取自总体X的样本, 求参数α的矩估计.

解. 根据矩估计化计算总体X的期望

$$E[X]=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^1 x(\alpha+1)x^{\alpha+1}dx=\frac{\alpha+1}{\alpha+2}.$$

样本X的均值$\bar{X}=\sum_{i=1}^{n}X_i/n$. 令样本矩代替总体矩,

$$E(X)=\frac{\alpha+1}{\alpha+2}=\bar{X},$$

求解出$\alpha=(2\bar{X}-1)/(1-\bar{X})$. □

例9.2. *设$X_1,X_2,\cdots,X_n$是取自总体X的一样本, 总体X的密度函数为*

$$f(x)=\begin{cases}\theta e^{-(x-\mu)\theta} & x\geq\mu\\ 0 & \text{其它}\end{cases}$$

其中$\theta>0$, 求μ,θ的矩估计.

解. 令随机变量$Y=X-\mu$, 由题意可知Y服从参数为θ的指数分布, 则总体

$$E[Y]=1/\theta, \mathrm{Var}Y=1/\theta^2,$$

由此得到$E(X)=\mu+1/\theta$和$Var(X)=1/\theta^2$. 计算对应的样本矩:

$$A_1=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i, \qquad B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

求解方程组

$$\mu+\frac{1}{\theta}=A_1, \qquad \frac{1}{\theta^2}=B_2,$$

解得$\mu=\bar{X}-\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/n}$和$\theta=1/\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/n}$. □

习题. *1) 求正态总体$\mathcal{N}(\mu,\sigma^2)$的$\mu,\sigma^2$的矩估计法; 2) 求总体$X\sim\mathcal{U}(a,b)$中$a,b$的矩估计法.*

9.1.2 极大似然估计法

如果X为离散型随机变量, 其分布列为$\Pr(X=x)=\Pr(X;\theta)$, 则样本$X_1,X_2,\cdots,X_n$的分布列为

$$L(\theta)=L(\theta;X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}\Pr(X_i;\theta).$$

$L(\theta)$为样本$X_1,X_2,\cdots,X_n$发生的概率.

若X为连续型随机变量, 概率密度为$f(x;\theta)$, 则$X_1=x_1,X_2=x_2,\cdots,X_n=x_n$的联合概率密度为$\prod_{i=1}^{n}f(x_i;\theta)$, 记

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

由概率密度定义可知: $L(\theta)$越大, $(X_1,X_2,\cdots,X_n)$落入$(x_1,x_2,\cdots,x_n)$的领域的概率越大.

针对上述离散和连续型随机变量, $L(\theta)$是θ的函数, 统称$L(\theta)$为样本$X_1,X_2,\cdots,X_n$的似然函数. 若

$$\widehat{\theta}=\arg\max_{\theta}L(x_1,x_2,\cdots,x_m;\theta),$$

称$\widehat{\theta}$为θ的极大似然估计量. 直觉而言, 极大似然估计量$\widehat{\theta}$是使观测值$X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$出现的概率最大.

求解方法: 最大化似然函数, 步骤如下:

1) 写出对数似然函数$\log(L(x_1, x_2, \cdots, x_m; \theta)$

2) 对对数似然函数中的θ一个/多个变量求一阶导数, 并令一阶导数等于0.

4) 求解方程组得到极大似然估计量$\widehat{\theta}$.

例9.3. *设$X_1, X_2, \cdots, X_n$是取自总体$X \sim B(1, p)$的一个样本, 求参数p的极大似然估计.*

解. 似然函数

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i},$$

$$\ln L(p) = \sum_{i=1}^{n} X_i \ln p + (n - \sum_{i=1}^{n}) \ln(1-p),$$

设

$$0 = \frac{d \ln L(p)}{dp} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}(n - \sum_{i=1}^{n} X_i).$$

所以$p = \sum_{i=1}^n X_i / n = \bar{X}$. [验证矩估计法] □

例9.4. *设总体$X \sim \mathcal{N}(\mu, \sigma^2)$, μ, σ^2未知, $X_1, X_2, \cdots, X_n$为X的样本, 求μ, σ^2的极大似然估计.*

解. 由正太分布可知X的概率密度函数为$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. 则样本$X_1, X_2, \cdots, X_n$的似然函数为

$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2}\right).$$

由此得到对数似然函数为$\ln L(\mu, \sigma^2) = -n \ln(2\pi)^{\frac{1}{2}} - n \ln \sigma - \sum_{i=1}^n (X_i - \mu)^2 / 2\sigma^2$. 对参数$\mu$求导计算可得

$$\frac{d \ln L(\mu, \sigma^2)}{d\mu} = \sum_{i=1}^{n} (X_i - \mu) = 0 \quad \Longrightarrow \quad \mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

对σ求导计算可得

$$\frac{d \ln L(\mu, \sigma^2)}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2 = 0 \quad \Longrightarrow \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

□

前面两例的矩估计和极大似然估计完全一样.

例9.5. *设总体X的密度函数*

$$f(x) = \begin{cases} (\alpha+1)x^\alpha & x \in (0,1) \\ 0 & \text{其它} \end{cases}$$

设$X_1, X_2, \cdots, X_n$是总体X的样本, 求α的极大似然估计.

解. 首先得到对数似然函数

$$L(\alpha) = (\alpha+1)^n \prod_{i=1}^{n} X_i^{\alpha} = (\alpha+1)^n (X_1 X_2 \cdots X_n)^{\alpha},$$

以及其对数似然函数$\ln L(\alpha) = n\ln(\alpha+1) + \alpha\ln(X_1X_2\cdots X_n)$. 求导并令其为0有

$$\frac{d\ln L(\alpha)}{d\alpha} = \frac{n}{\alpha+1} + \ln(X_1X_2\cdots X_n) = 0,$$

求解得

$$\alpha = \frac{n}{\sum_{i=1}^{n}\ln(\frac{1}{X_i})} - 1 = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\ln(\frac{1}{X_i})} - 1.$$

□

对于上例, 其矩估计值为$\alpha = (2\bar{X}-1)/(1-\bar{X})$, 因此矩估计值与极大似然估计值可能不同.

例9.6. 总体$X \sim U(a,b)$, 求a,b的极大似然估计.

解. 首先得到X的密度函数为$f(x) = 1/(b-a)$ $(x \in [a,b])$, 其它为零, 因此似然函数为

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n} & a \le X_1, X_2, \cdots, X_n \le b \\ 0 & \text{其它} \end{cases}$$

直接求导无法解出a与b的估计, 此时可以从极大似然定义出发, 最大化$L(a,b)$, 应使得b尽可能小、a尽可能大, 但需满足$a \le X_1, X_2, \cdots, X_n \le b$, 因此极大似然估计量为:

$$b = \max\{X_1, X_2, \cdots, X_n\} \qquad a = \min\{X_1, X_2, \cdots, X_n\}.$$

□

极大似然估计不可变性: 设$\widehat{\theta}$为θ的极大似然估计, $\mu(\theta)$为θ的函数, 且具有单值反函数$\theta = \theta(\mu)$, 则$\widehat{\mu} = \mu(\widehat{\theta})$是$\mu$的极大似然估计.

例9.7. $X_1, X_2, \cdots, X_n$是总体X的一个样本, 总体X的概率密度为

$$f(x) = \begin{cases} \theta e^{-(x-\mu)\theta} & x \ge \mu \\ 0 & \text{其它} \end{cases}$$

求μ, θ的极大似然估计.

解. 首先得到似然函数为

$$L(\theta,\mu) = \begin{cases} \theta^n e^{-\theta\sum_{i=1}^{n}(X_i-\mu)} & X_i \ge \mu \\ 0 & \text{其它} \end{cases}$$

进一步得到对数似然函数为

$$\ln L(\theta,\mu) = n\ln\theta - \theta\sum_{i=1}^{n}(X_i-\mu).$$

求导并设导数为零可得

$$\frac{\ln L(\theta,\mu)}{d\theta}=\frac{n}{\theta}-\sum_{i=1}^{n}(X_i-\mu)=0 \quad\Rightarrow\quad \theta=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)},$$

另一方程为

$$\frac{\ln L(\theta,\mu)}{d\mu}=\theta\sum_{i=1}^{n}X_i=0 \quad\Rightarrow\quad \theta=0,$$

此时无解求解θ,μ的极大似然估计. 回到似然函数的定义

$$L(\theta,\mu)=\begin{cases}\theta^n e^{-\theta\sum_{i=1}^{n}(X_i-\mu)} & X_1,X_2,\cdots,X_n\geq\mu\\ 0 & \text{其它}\end{cases}$$

可以发现μ越大, 似然函数$L(\theta,\mu)$越大, 但须满足$X_i\leq\mu$ $(i\in[n])$. 由此可得极大似然估计

$$\hat{\mu}=\min\{X_1,X_2,\cdots,X_n\},$$

以及

$$\hat{\theta}=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})}.$$

□