第二章　参数估计

- 统计推断
  - 估计理论
    - 参数估计
      - 点估计
        - 矩估计
        - 极大似然估计
        - 最小二乘估计
        - 优劣标准
          - 无偏性
          - 有效性
          - 相合性
      - 区间估计
        - 双侧区间
        - 单侧区间
    - 非参数估计
  - 假设检验
    - 参数检验
      - 正态总体
        - 单个正态总体
        - 两个正态总体
      - 非正态总体
      - 双侧检验
      - 单侧检验
    - 非参数检验

§2.1　点估计方法

设总体X的分布函数$F(x;\theta)$的形式已知，$\theta=(\theta_1,\theta_2,\cdots,\theta_k)$是待估未知参数. $X_1,X_2,\cdots,X_n$是X的一个样本，$x_1,x_2,\cdots,x_n$为相应的一个样本值.

点估计问题就是要构造一个适当的统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$,用它的观察值$\hat{\theta}(x_1,x_2,\cdots,x_n)$来估计未知参数$\theta$.

由于估计量是样本的函数，是随机变量，故对不同的样本值, 得到的参数值往往不同，如何求估计量是关键问题.

一、矩估计法

是英国统计学家*K*.皮尔逊最早提出的，其基本思想是用样本矩估计总体矩. 这种方法的理论依据是大数定律. 记

$$\mu_i = E(X^i) = g_i(\theta_1,\cdots,\theta_k), \qquad A_i = \frac{1}{n}\sum_{j=1}^{n} X_j^k$$

分别表示总体和样本的i 阶矩. 用样本 i 阶矩代替总体的 i 阶矩，建立如下方程(组)：

$$\begin{cases} g_1(\theta_1,\cdots,\theta_k)=A_1 \\ g_2(\theta_1,\cdots,\theta_k)=A_2 \\ \cdots\cdots\cdots\cdots\cdots \\ g_k(\theta_1,\cdots,\theta_k)=A_k \end{cases}$$

用样本矩估计
总体矩

该方程组的解$\hat{\theta}_j = h_j(X_1, \cdots, X_k)$，$j = 1, 2, \cdots, k$.称为$\theta_1, \cdots, \theta_k$的矩估计量, $\hat{\theta}_j = h_j(x_1, \cdots, x_k)$称为估计值.

不论总体服从什么分布，若总体期望μ 与方差σ^2存在，则它们的矩估计量分别为

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S_n^2.$$

事实上，按矩法原理，令

$$\begin{cases} \mu = E(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i = \bar{X} \\ \mu^2 + \sigma^2 = E(X^2) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 = \bar{X}^2 + S_n^2 \end{cases}$$

其解为 $\hat{\mu} = \bar{X}$，$\hat{\sigma}^2 = S_n^2$. 得证.

例1 设总体$X \sim N(\mu,\sigma^2)$, $X_1, X_2, \cdots, X_n$为总体的样本,求μ,σ^2的矩法估计量.

解 $\hat{\mu}_{矩} = \bar{X}$, $\hat{\sigma}^2_{矩} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = S_n^2$.

例2 设总体$X \sim E(\lambda)$, $X_1, X_2, \cdots, X_n$为总体的样本. 求λ 的矩法估计量.

解 由于$E(X) = 1/\lambda$, 因此 $1/\lambda = \bar{X}$,

所以 $\hat{\lambda}_{矩} = 1/\bar{X}$.

例3 设从某灯泡厂某天生产的灯泡中随机抽取10只灯泡，测得其寿命为(单位:小时).

1050, 1100, 1080, 1120, 1200, 1250, 1040, 1130, 1300, 1200. 试用矩法估计该厂这天生产的灯泡的平均寿命及寿命分布的方差.

解 $\hat{\mu} = \bar{x} = \dfrac{1}{10}\sum_{i=1}^{10} x_i = 1147(h);$

$$\hat{\sigma}^2 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 - \bar{x}^2 = 6821(h^2).$$

例4 设总体$X \sim U(a,b)$，a,b未知,求参数a,b的矩法估计量.

解 由于 $E(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$.

$$E(X^2)=D(X)+E^2(X)=\frac{(b-a)^2}{12}+(\frac{a+b}{2})^2$$

令 $$\begin{cases}\dfrac{a+b}{2}=\bar{X}\\[2ex] \dfrac{(b-a)^2}{12}+\left(\dfrac{a+b}{2}\right)^2=A_2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2\end{cases}$$

解得

$$\hat{a}_{矩}=\bar{X}-\sqrt{3(A_2-\bar{X}^2)}=\bar{X}-\sqrt{3}S_n,$$

$$\hat{b}_{矩}=\bar{X}+\sqrt{3(A_2-\bar{X}^2)}=\bar{X}+\sqrt{3}S_n.$$

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. 其主要原因在于建立矩法方程时, 选取那些总体矩用相应样本矩代替带有一定的随意性 (例如：对于泊松分布，参数λ 既是期望也是方差, 因此矩估计$\bar{X}, S_n^2$都可看作是参数λ 的矩估计).

二、极大似然估计法

是在总体类型已知条件下使用的一种参数估计方法. 它首先是由德国数学家Gauss在1821年提出的，Fisher在1922年重新发现了这一方法，并首先研究了这种方法的一些性质.

基本思想：一次试验就出现的事件有较大的概率.

例如：有两外形相同的箱子，各装100个球.

甲箱	99个白球	1个红球
乙箱	1个白球	99个红球

现从两箱中任取一箱,并从箱中任取一球, 结果所取得的球是白球. 问这个球是从哪个箱子中取出的？

若取到的是甲箱，从中取出白球的概率为0.99，若取到的是乙箱，从中取出白球的概率为0.01. 现在一次试验白球出现了，人们的直觉就是：

"此白球最像从甲箱取出的". $\boxed{\text{"极大似然"之意}}$

分析

对离散型总体 $P(X=x_k)=P(x_k;\theta)$, $k=1,2,\ldots$

抽样结果：事件$\{X_1=x_1,\cdots,X_n=x_n\}$发生的概率为：

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

对连续型总体 密度函数为$f(x;\theta)$

抽样结果：事件$\{X_1=x_1,\cdots,X_n=x_n\}$可理解成事件

$$\{X_1 \in (x_1, x_1+dx_1), \cdots, X_n \in (x_n, x_n+dx_n)\},$$

其发生的概率为:

$$f(x_1, x_2, \cdots, x_n; \theta) \approx \prod_{i=1}^{n} f(x_i; \theta) dx_i, \quad \theta \in \Theta.$$

极大似然估计的思想是选取$\theta \in \Theta$，使得上述概率达到最大.

极大似然估计原理:

设$X_1, X_2, \cdots, X_n$是取自总体X的一个样本，样本的联合密度(连续型)为.

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

或联合概率函数(离散型)

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

(其中Θ是参数空间——θ可能的取值范围).

当给定样本值$x_1, x_2, \cdots, x_n$时，定义似然函数为

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

$L(\theta)$看作参数θ的函数, 它可作为将以多大可能产生样本值$x_1, x_2, \cdots, x_n$的一种度量.

极大似然估计法就是用使$L(\theta)$达到最大值的$\hat{\theta}$去估计θ. 即 $L(\hat{\theta}) = \max\limits_{\theta \in \Theta} L(\theta)$. 其解

$$\hat{\theta} = g(x_1, x_2, \cdots, x_n)$$

为参数θ的极大似然估计值, 简记为$\hat{\theta}_{mle}$.称统计量

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

为参数θ的极大似然估计量. 简记为$\hat{\theta}_{MLE}$.

求极大似然估计的步骤

(一) 写出似然函数

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

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

(二) 取对数（$L(\theta)$与$\ln L(\theta)$具有相同的极值点）

$$\ln L(\theta)=\sum_{i=1}^{n}\ln p(x_i;\theta) \quad \text{或} \quad \ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i;\theta);$$

对数似然函数

(三) 求 $\ln L(\theta)$ 的驻点，即令 $\boxed{\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0,}$

对数似然方程

(四) 验证是否是极大值点. 即得极大似然估计值 $\hat{\theta}$.

当总体分布中含有多个未知参数时, 此时需令

$$\frac{\partial}{\partial \theta_i}\ln L=0,\quad i=1,2,\cdots,k$$

对数似然方程组

求解上述方程组, 即可得极大似然估计值 $\hat{\theta}_1,\cdots,\hat{\theta}_k$.

注：(1) 当样本来自不同的总体时, 似然函数表达式仍然成立, 此时各样本的密度函数 $f(x_i;\theta)$ 或概率函数 $p(x_i;\theta)$ 不一样. (练习题2).

(2) 当似然函数是未知参数的单调函数时, 此时不能利用导数为零的方法求解, 只能从定义出发求解(见例题8).

(3) 待估参数的MLE可能不存在, 也可能不唯一.

例5 设$X \sim B(1, p)$, $X_1, X_2, \cdots, X_n$是来自X的一个样本,求 p的极大似然估计量.

解 设$x_1, x_2, \cdots, x_n$为相应于样本$X_1, X_2, \cdots, X_n$的一个样本值, X的分布律为

$$P\{X = x\} = p^x (1-p)^{1-x}, \qquad x = 0, \ 1.$$

似然函数 $L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}$,

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

令

$$\frac{\mathrm{d}}{\mathrm{d}p}\ln L(p)=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{n-\sum_{i=1}^{n}x_i}{1-p}=0,$$

解得 p 的极大似然估计值

$$\hat{p}_{mle}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

p的极大似然估计量为

$$\hat{p}_{MLE}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}.$$

与矩估计相同

例6 设X服从参数为$\lambda(\lambda>0)$的泊松分布, $X_1, X_2, \cdots, X_n$是来自X的一个样本, 求λ的极大似然估计量.

解 X的分布律为

$$P\{X=x\}=\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}, \quad (x=0,1,2,\cdots)$$

似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\left(\frac{\lambda^{x_i}}{x_i!}\mathrm{e}^{-\lambda}\right)=\mathrm{e}^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}\left(\prod_{i=1}^{n}(x_i!)\right)^{-1},$$

$$\ln L(\lambda)=-n\lambda+\left(\sum_{i=1}^{n}x_i\right)\ln\lambda-\sum_{i=1}^{n}(x_i!),$$

令 $$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda)=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i=0,$$

解得 p 的极大似然估计值 $\hat{p}_{mle}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$

p的极大似然估计量为 $\hat{p}_{MLE}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}.$

与矩估计相同

例7 设总体$X\sim N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$为总体的样本. 求μ ,σ^2的极大似然估计.

解 似然函数为

$$L(x_1,x_2,\cdots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

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

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

似然方程组为 $\begin{cases} \dfrac{\partial}{\partial \mu}\ln L = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0 \\ \dfrac{\partial}{\partial(\sigma^2)}\ln L = \dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2 - \dfrac{n}{2\sigma^2}=0 \end{cases}$

解得

$$\hat{\mu}_{mle} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}, \qquad \hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

μ，σ^2的极大似然估计量分别为

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = S_n^2.$$

与矩估计相同

例8 设$X \sim U(a,b)$，$x_1, x_2, \cdots, x_n$是X的一个样本值．求a, b的极大似然估计值与极大似然估计量.

解 X的密度函数为

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

似然函数为

$$L(x_1,x_2,\cdots,x_n;a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_i \le b,\\ & i=1,2,\cdots,n\\ 0, & \text{其它}.\end{cases}$$

似然函数L是a,b的单调函数, a越大,b越小,L越大.

显然a的最大可能取值为$x_{\min}$，b的最小可能取值为$x_{\max}$．因此 a,b 的极大似然估计值分别为$x_{\min}, x_{\max}$．估计量为

$$\hat{a} = X_{\min}, \quad \hat{b} = X_{\max}.$$

与矩估计不同

极大似然估计的性质（不变性）

设$\hat{\theta}$是θ 的极大似然估计值，$u(\theta)(\theta \in \Theta)$是$\theta$ 的函数,且有单值反函数 $\theta = \theta(u)$，$u \in U$． 则

$$\hat{u} = u(\hat{\theta})$$

是$u(\theta)$的极大似然估计值$(\theta \in \Theta)$.

如正态总体$N(\mu, \sigma^2)$中, σ^2的极大似然估计值为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

$\sigma = \sqrt{\sigma^2}$是σ^2的单值函数，且具有单值反函数，故σ的极大似然估计值为

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

练习题1 设X_1,X_2,X_3分别是来自参数为λ，2λ，3λ的指数分布的样本，它们相互独立，试求参数λ的极大似然估计量.

练习题2 设某种电子器件的寿命X服从双参数的指数分布，$X_1,\cdots,X_n$是来自X的简单样本，X的概率密度为

$$f(x;\lambda,\theta)=\begin{cases}\lambda e^{-\lambda(x-\theta)}, & x\geq\theta,\\ 0, & \text{其他}.\end{cases}$$

其中$\lambda,\theta\ (\lambda,\theta>0)$为未知参数.

(1)求λ与θ的矩估计量,

(2)求λ与θ的极大似然估计量.

练习题1答案

似然函数为

$$L(x_1,x_2,x_3;\lambda)=f(x_1;\lambda)f(x_2;\lambda)f(x_3;\lambda)$$

$$=\begin{cases}6\lambda^3 e^{-\lambda(x_1+2x_2+3x_3)}, & x_1,x_2,x_3\ge 0,\\ 0, & \text{其他}.\end{cases}$$

极大似然估计 $\hat{\lambda}=\dfrac{3}{X_1+2X_2+3X_3}$.

练习题2答案

（1）矩估计 $\hat{\lambda}_1=1/S_n$，$\hat{\theta}_1=\bar{X}-S_n$,;

（2）极大似然估计 $\hat{\lambda}_2=\dfrac{1}{\bar{X}-X_{(1)}}$，$\hat{\theta}_2=X_{(1)}$.

二者不同