第八章 参数估计

问题的提出

人们经常遇到的问题是如何选取样本以及根据样本来对总体的种种统计特征作出判断。实际工作中碰到的随机变量(总体)往往是分布类型大致知道, 但确切的形式并不知道, 亦即总体的参数未知. 要求出总体的分布函数$F(x)$(或密度函数$\varphi(x)$), 就等于要根据样本来估计出总体的参数. 这类问题称为参数估计.

8.1 估计量的优劣标准

估计量的优劣标准

设θ为总体中要被估计的一个未知参数, 例如期望值或方差等,

$\hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的估计量.它是容量为$n$的样本的函数,例如样本平均数$\overline{X}$及样本方差$S^2$等等.人们总希望估计量能够代表真实参数,根据不同要求.有三种常用的标准.

(一) 一致估计

一般情况$\hat{\theta} \neq \theta$,但希望当$n \to \infty$时,

$\hat{\theta} \xrightarrow{P} \theta$.这就是说,希望当样本容量$n$无限增大时,估计值$\hat{\theta}$在真值附近的概率趋近于1.

定义8.1 如果当$n \to \infty$时,$\hat{\theta}$依概率收敛于θ,即任给$\varepsilon > 0$, $\lim\limits_{n\to\infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$,

则称$\hat{\theta}$为参数θ的一致估计.

(二) 无偏估计

根据样本推得的估计值与真值可能不同, 然而, 如果有一系列抽样构成各个估计, 很合理地会要求这些估计的期望值与未知参数的真值相等, 它的直观意义是样本估计量的数值在参数的真值周围摆动, 而无系统误差.

定义 8.2 如果$E\hat{\theta}=\theta$成立,则称估计$\hat{\theta}$为参数θ的无偏估计.

例1　从总体ξ中取一样本（$X_1, \cdots, X_n$），$E\xi = \mu$，$D\xi = \sigma^2$，试证样本平均数$\overline{X}$及样本方差S^2分别是μ及σ^2的无偏估计。

证 $\because \quad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i, \quad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$

$$E(\overline{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}EX_i=\frac{1}{n}n\mu=\mu$$

$E(\overline{X})=\mu \quad \therefore$ 样本均值$\overline{X}$是μ的无偏估计。

怎样证明？

$$ES^2=E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2]=\sigma^2.$$

$$D\bar{X} = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n}\sigma^2$$

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

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

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

$$D\bar{X} = \frac{1}{n}\sigma^2$$

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

$$= \frac{1}{n-1}n\sigma^2 - \frac{n}{n-1}\frac{\sigma^2}{n} = \sigma^2$$

$\therefore S^2$是σ^2的无偏估计

(三) 有效估计

定义 8.3 设$\hat{\theta}$和$\hat{\theta}'$都是θ的无偏估计，若样本容量为n,$\hat{\theta}$的方差小于$\hat{\theta}'$的方差,则称$\hat{\theta}$是比$\hat{\theta}'$有效的估计量. 如果在θ的一切无偏估计量中,$\hat{\theta}$的方差达到最小,则$\hat{\theta}$称为θ的有效估计量.

实际上,样本平均数$\overline{X}$是总体期望值的有效估计量.

例2　比较总体期望值μ的两个无偏估计的有效性。

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad X'=\frac{\sum_{i=1}^{n}a_iX_i}{\sum_{i=1}^{n}a_i} \qquad \left(\sum_{i=1}^{n}a_i\neq 0\right)$$

解：

$$E(\bar{X})=\mu \qquad D\bar{X}=\frac{1}{n}\sigma^2$$

$$E\bar{X}'=\frac{\sum_{i=1}^{n}a_iEX_i}{\sum_{i=1}^{n}a_i}=\mu$$

μ

$$D\bar{X} = \frac{1}{n}\sigma^2 \qquad DX' = \frac{\sum_{i=1}^{n} a_i^2 DX}{\left(\sum_{i=1}^{n} a_i\right)^2} = \frac{\sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i\right)^2}\sigma^2$$

利用不等式 $\because \ (a_i - a_j)^2 \geq 0 \qquad 2a_i a_j \leq a_i^2 + a_j^2$

$$\left(\sum_{i=1}^{n} a_i\right)^2 = \sum_{i=1}^{n} a_i^2 + \sum_{i \prec j} 2a_i a_j \leq \sum_{i=1}^{n} a_i^2 + \sum_{i \prec j}(a_i^2 + a_j^2)$$

$$= n\sum_{i=1}^{n} a_i^2$$

$$\frac{1}{n} \leq \frac{\sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i\right)^2}$$

$$D\bar{X} = \frac{1}{n}\sigma^2 \qquad DX' = \frac{\sum_{i=1}^{n} a_i^2 DX}{\left(\sum_{i=1}^{n} a_i\right)^2} = \frac{\sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i\right)^2}\sigma^2$$

$$\frac{1}{n} \le \frac{\sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i\right)^2}$$

$$\therefore DX' \ge \frac{\sum_{i=1}^{n} a_i^2}{n\sum_{i=1}^{n} a_i^2}\sigma^2 = \frac{1}{n}\sigma^2 = D\bar{X}$$

故$\bar{X}$比X'有效

8.2 获得估计量的方法——点估计

(一）矩估计法（简称“矩法”）

矩法是求估计量的最古老的方法。

具体的做法是：以样本矩作为相应的总体矩的估计，

以样本矩的函数作为相应的总体矩的同一函数的估计。

常用的是用样本平均数估计总体期望值 。

例1　某灯泡厂某天生产了一大批灯泡，从中抽取了10个进行寿命实验，得数据如下（单位：小时）问该天生产的灯泡平均寿命是多少?

1050	1100	1080	1120	1200
1250	1040	1130	1300	1200

解 计算出$\bar{X}=1147$，以此作为总体期望值μ的估计。

矩法比较直观，求估计量有时也比较直接，但它产生的估计量往往不够理想。

最大似然估计法

引例 某同学与一位猎人一起去打猎,一只野兔从前方窜过,只听一声枪响,野兔应声倒下,试猜测是谁打中的?

由于只发一枪便打中,而猎人命中的概率一般大于这位同学命中的概率,故一般会猜测这一枪是猎人射中的.

最大似然估计法的思想:在已得到试验结果的情况下,应寻找使这个结果出现的可能性最大的那个θ值作为θ 的估计 $\hat{\theta}$.

(二) 最大似然估计法

Maximum likelyhood estimation

现在要根据从总体ξ中抽到的样本$(X_1, X_2, \ldots, X_n)$，对总体分布中的未知参数θ进行估计. 最大似然法是要选取这样的估计值, 当它作为θ的估计值时, 使观察结果出现的可能性最大.

对于离散型的随机变量就是估计概率函数中的参数θ, 对于连续型的随机变量就是估计概率密度中的θ.

设ξ为连续型随机变量, 它的分布函数是$F(x;\theta)$, 概率密度是$\varphi(x;\theta)$, 其中θ是未知参数, 可以是一个值, 也可以是一个向量, 由于样本的独立性, 则样本$(X_1,X_2,\ldots,X_n)$的联合概率密度是

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

对每一取定的样本值$x_1,x_2,\ldots,x_n$是常数, L是参数θ的函数, 称L为样本的似然函数(如果θ是一个向量, 则L是多元函数)

设ξ为离散型随机变量, 有概率函数 $P(\xi=x_i)=p(x_i;\theta)$, 则似然函数

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

定义8.4 如果$L(x_1,x_2,\cdots,x_n;\theta)$在$\hat{\theta}$处达到最大值,则称$\hat{\theta}$是$\theta$的最大似然估计

3、求极大似然估计的步骤

(1) 做似然函数

$$L(\theta)=L(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n}\varphi(x_i;\theta)$$ 整理合并

(2) 做对数似然函数

$$\ln L(\theta)=\ln L(x_1,\cdots,x_n;\theta)$$

(3) 求导数，列似然方程

$$\frac{d[\ln L(\theta)]}{d\theta}=0$$

(4) 解似然方程

若该方程有解，则其解就是θ的最大似然估计。

例2 已知

$$\xi \sim \varphi(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0, \theta > 0 \\ 0 & \text{其它} \end{cases}$$

$x_1, x_2, \ldots, x_n$为ξ的一组样本观察值, 求θ的最大似然估计.

解 似然函数

$$L(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$$

$$\ln L = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n} x_i$$

令 $\dfrac{d \ln L}{d\theta} = -\dfrac{n}{\theta} + \dfrac{1}{\theta^2} \sum_{i=1}^{n} x_i = 0$

解得 $\hat{\theta} = \dfrac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$

$\overline{x}$就是θ的最大似然估计.

例3 某电子管的使用寿命(从开始使用到初次失效为止)服从指数分布(概率密度见例2), 今抽取一组样本, 其具体数据如下:

16, 29, 50, 68, 100, 130, 140, 270, 280, 340, 410, 450, 520, 620, 190, 210, 800, 1100

问如何估计θ?

解 根据例2的结果, 参数θ用样本平均数估计.

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{18}(16+29+\cdots+800+1100)\approx 318(\text{小时})$$

以此为θ的估计值.

例4 已知ξ服从正态分布$N(\mu,\sigma^2)$, $(x_1,x_2,\ldots,x_n)$为ξ的一组观察值, 用最大似然估计法估计μ,σ^2的值.

解

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

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

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

$$\frac{\partial L}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

解似然方程组

$$\begin{cases} \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0 \\ -\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0 \end{cases}$$

得

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$

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

例 求0-1分布的总体的对参数p的最大似然估计.

解 已知总体ξ的概率函数为

$P(\xi=k)=p^k(1-p)^{1-k}$,$(k=0,1)$, 假设获得了n个样本值$(x_1,x_2,\ldots,x_n)$, 当然这些值不是0就是1.

$$L=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

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

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

令 $\dfrac{d\ln L}{dp}=\dfrac{1}{p}\sum_{i=1}^{n}x_i-\dfrac{1}{1-p}\sum_{i=1}^{n}(1-x_i)=0$

即 $\left(\dfrac{1}{p}+\dfrac{1}{1-p}\right)\sum_{i=1}^{n}x_i=\dfrac{n}{1-p}$,

解得 $\hat{p}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$

2. 设为从总体ξ中取出的一组样本观察值, 试用最大似然法估计ξ的概率密度φ(x)中的未知参数θ, 若

$$\varphi(x)=\begin{cases}\theta x^{\theta-1} & \text{当}\,0<x<1\ (\theta>0)\\ 0 & \text{其它}\end{cases}$$

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

令 $\dfrac{dL}{d\theta}=\dfrac{n}{\theta}+\sum_{i=1}^{n}\ln x_i=0$, 得 $\hat{\theta}=-\dfrac{n}{\sum_{i=1}^{n}\ln x_i}$

例 求普哇松分布中参数λ的最大似然估计.

解 已知总体ξ的概率函数为

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0,1,2,\cdots$$

获得的n个样本值

$x_1, x_2, \cdots, x_n$也是正整数,

则$$L = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}$$

因此

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

令 $\dfrac{d\ln L}{d\lambda} = -n + \dfrac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$

得 $\hat{\lambda} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$