概率论与数理统计第20讲

第六章　参数估计

在实际问题中, 当所研究的总体分布类型已知, 但分布中含有一个或多个未知参数时, 如何根据样本来估计未知参数, 这就是参数估计问题.

参数估计问题分为点估计问题与区间估计问题两类. 所谓点估计就是用某一个函数值作为总体未知参数的估计量; 区间估计就是对于未知参数给出一个范围, 并且在一定可靠度下使这个范围包含未知参数的真值.

参数估计问题的一般提法

设有一个总体X, 总体的分布函数为$F(x,\theta)$, 其中θ为未知参数(θ可以是向量). 现从该总体中随机地抽样, 得到一个样本

$$X_1, X_2, \ldots, X_n,$$

再依据该样本对参数θ作出估计, 或估计参数θ的某已知函数$g(\theta)$.

§6.1 点估计问题概述

一, 点估计的概念

设 $X_1,X_2,\ldots,X_n$ 是取自总体 X 的一个样本，$x_1,x_2,\ldots,x_n$ 是相应的一个样本值. θ是总体分布中的未知参数，为估计未知参数θ，需构造一个适当的统计量

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

然后用其观察值 $\hat{\theta}(x_1,x_2,\ldots,x_n)$

来估计θ的值. 称$\hat{\theta}(X_1,X_2,\ldots,X_n)$为$\theta$的**估计量**，称$\hat{\theta}(x_1,x_2,\ldots,x_n)$为$\theta$的**估计值**，它们统称为**点估计**，简称为**估计**，并简记为$\hat{\theta}$

注：估计量$\hat{\theta}(X_1, X_2, \ldots, X_n)$是一个随机变量，是样本的函数，即是一个统计量，对于不同的样本值，θ的估计值$\hat{\theta}$一般是不同的.

二, 评价估计量的标准

估计量的评价一般有三条标准:

(1) 无偏性;

(2) 有效性;

(3) 相合性(一致性)

1. 无偏性

定义 1 设$\hat{\theta}(X_1,\ldots,X_n)$是未知参数$\theta$的估计量，若

$$E(\hat{\theta}) = \theta,$$

则称$\hat{\theta}$为θ的无偏估计量.

定理1 设$X_1,\ldots,X_n$为取自总体X的样本，总体X的均值为μ，方差为σ^2，则

(1) 样本均值$\bar{X}$是μ的无偏估计量;

(2) 样本方差S^2是σ^2的无偏估计量;

(3) 样本二阶中心矩$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$是$\sigma^2$的有偏估计量(但却是**渐近无偏估计量**).

2. 有效性

定义 2 设$\hat{\theta}_1=\hat{\theta}_1(X_1,\ldots,X_n)$和

$\hat{\theta}_2=\hat{\theta}_2(X_1,\ldots,X_n)$都是参数$\theta$的无偏估计量,

若$D(\hat{\theta}_1)<D(\hat{\theta}_2)$, 则称$\hat{\theta}_1$较$\hat{\theta}_2$**有效**.

注: 最小方差无偏估计的定义如下:

设 $X_1,\ldots,X_n$ 是取自总体 X 的一个样本,

$\hat{\theta}(X_1,\ldots,X_n)$是未知参数$\theta$的一个估计量,若

$\hat{\theta}$满足:

① $E(\hat{\theta})=\theta$, 即$\hat{\theta}$为θ的无偏估计;

② $D(\hat{\theta})\le D(\hat{\theta}^*)$, $\hat{\theta}^*$是θ的任一无偏估计.

则称$\hat{\theta}$为θ的**最小方差无偏估计**(也称**最佳无偏估计**).

3. 相合性(一致性)

定义 3 设$\hat{\theta}=\hat{\theta}(X_1,\ldots,X_n)$为未知参数$\theta$的估计量，若$\hat{\theta}$依概率收敛于$\theta$，即对任意$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\hat{\theta}-\theta|<\varepsilon\}=1,$$

或

$$\lim_{n\to\infty} P\{|\hat{\theta}-\theta|\geq\varepsilon\}=0,$$

则称$\hat{\theta}$为θ的**相合估计量(一致估计量)**.

§6.2 点估计的常用方法

一，矩估计法

根据大数定理，可用样本矩估计总体矩.

一般地，记总体 k 阶矩 $\mu_k=E(X^k)$;

样本 k 阶矩 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$;

总体 k 阶中心矩 $v_k=E[X-E(X)]^k$;

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

定义1 用相应的样本矩去估计总体矩的方法就称为**矩估计法**. 用矩估计法确定的估计量称为**矩估计量**. 相应的估计值称为**矩估计值**. 矩估计量与矩估计值统称为**矩估计**.

求矩估计的方法:

设总体X的分布函数$F(x;\theta_1,\ldots,\theta_k)$中有$k$个未知参数$\theta_1,\ldots,\theta_k$, 则

(1) 求总体X的前k阶矩$\mu_1,\ldots,\mu_k$, 一般都是这k个未知参数的函数, 记为

$$\mu_i=g_i(\theta_1,\ldots,\theta_k),\ i=1,2,\ldots,k \qquad (2.1)$$

(2) 从(1)中解得$\theta_j=h_j(\mu_1,\ldots,\mu_k), j=1,\ldots,k$

(3) 再用μ_i的估计量A_i代替上式中的μ_i, 即可得$\theta_j(j=1,2,\ldots,k)$的矩估计量:

$$\hat{\theta}_j=h_j(A_1,\cdots,A_k),\quad j=1,2,\cdots,k.$$

例3　设总体 X 的均值 μ 和方差 $\sigma^2(>0)$ 都存在，μ,σ^2 未知．$X_1,\ldots,X_n$ 是来自 X 的样本，试求 μ,σ^2 的矩估计量．

解　$\mu_1 = E(X) = \mu$

$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2$$

解得　$\mu = \mu_1$

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

于是 μ,σ^2 的矩估计量为　$\hat{\mu} = A_1 = \bar{X}$

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

二, 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.

1. 最大似然估计法的思想: 在已经得到实验结果的情况下, 应该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

$\hat{\theta}$

注: 最大似然估计法首先由德国数学家高斯于1821年提出, 英国统计学家费歇于1922年重新发现并作了进一步的研究.

下面分别就离散型总体和连续型总体情形作具体讨论.

(1) 离散型总体 的情形: 设总体X的概率分布为 $P\{X=x\}=p(x,\theta)$, (θ为未知参数). 如果$X_1,X_2,\ldots,X_n$是取自总体X的样本, 样本的观察值为$x_1,x_2,\ldots,x_n$,
则样本的联合分布律

$$P\{X_1=x_1,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i,\theta),$$

在给定观察值的情况下, 它是θ的函数, 记为

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

称其为**似然函数**.

(2) 连续型总体 的情形: 设总体X的概率密度为$f(x,\theta)$, 其中θ为未知参数, 此时定义**似然函数**

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

似然函数 $L(\theta)$的值的大小意味着该样本值出现的可能性的大小，在已得到样本值 $x_1,x_2,\dots,x_n$ 的情况下，则应该选择使 $L(\theta)$达到最大值的那个θ作为θ的估计$\hat{\theta}$. 这种求点估计的方法称为**最大似然估计法**.

定义 2 若对任意给定的样本值 $x_1,x_2,\ldots,x_n$,存在

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

使

$$L(\hat{\theta})=\max_{\theta} L(\theta),$$

则称 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为 θ 的**最大似然估计值**. 称相应的统计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 θ 的**最大似然估计量**. 它们统称为 θ 的最大似然估计(MLE).

2. 求最大似然估计的一般方法

求未知参数θ的最大似然估计问题, 归结为求似然函数$L(\theta)$的最大值点的问题. 当似然函数关于未知参数可微时, 可利用微分学中求最大值的方法求之.

主要步骤:

(1) 写出似然函数 $L(\theta)=L(x_1,x_2,\ldots,x_n;\theta)$;

(2) 令 $\dfrac{\mathrm{d}L(\theta)}{\mathrm{d}\theta}=0$ 或 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0$, 求出驻点;

(3) 判断并求出最大值点, 在最大值点的表达式中, 用样本值代入就得参数的最大似然估计值.

注: 因函数$\ln L$是L的单调增加函数, 且函数$\ln L(\theta)$与函数$L(\theta)$有相同的极值点, 故常转化为求函数$\ln L(\theta)$的最大值点较方便.

注: ①当似然函数关于未知函数不可微时, 只能按最大似然估计法的基本思想求出最大值点.

②上述方法易推广至多个未知参数的情形.

例5 设$X\sim b(1,p)$, $X_1,X_2,\ldots,X_n$是取自总体X的一个样本. 试求参数p的最大似然估计.

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

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

故似然函数为

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

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

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

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

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

$$\frac{n}{p}\overline{x}-\frac{n}{1-p}(1-\overline{x})=0$$

$$(1-p)\overline{x}-p(1-\overline{x})=0\Rightarrow \overline{x}-p=0,\hat{p}=\overline{x}$$

例6 设总体X服从$[0,\theta]$上的均匀分布, θ未知. $X_1,\ldots,X_n$为X的样本, $x_1,\ldots,x_n$为样本值. 试求θ的最大似然估计.

解 似然函数为

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 0\le x_1,\cdots,x_n\le\theta\\ 0, & \text{其它},\end{cases}$$

显然无法从$L'(\theta)=0$得到最大似然估计. 我们考虑直接按最大似然法的基本思想来确定.

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 0\le x_1,\cdots,x_n\le\theta\\ 0, & \text{其它},\end{cases}$$

欲使$L(\theta)$最大, θ应尽量小但又不能太小, 它必须同时满足

$$\theta\ge x_i,\quad i=1,\ldots,n,$$

否则$L(\theta)=0$, 而0不可能是$L(\theta)$的最大值. 因此

$$\hat{\theta}=\max\{x_1,\cdots,x_n\}\text{和}\hat{\theta}=\max\{X_1,\cdots,X_n\}$$

是最大似然估计值和估计量.

例7 设总体X服从指数分布, 概率密度为

$$f(x,\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}\quad(\lambda>0)$$

λ是未知参数. $x_1,x_2,\ldots,x_n$是来自总体X的一个样本观察值, 求参数λ的最大似然估计值.

解 似然函数

$$L(x_1,\cdots,x_n;\lambda)=\begin{cases}\prod_{i=1}^{n}\lambda e^{-\lambda x_i}, & x_i>0\\ 0, & 其它\end{cases}$$

则L的最大值点一定是

$$L_1(x_1,\cdots,x_n;\lambda)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}$$

的最大值点.

$$L_1(x_1,\cdots,x_n;\lambda)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}$$

对其取对数得

$$\ln L_1(x_1,\cdots,x_n;\lambda)=\sum_{i=1}^{n}(\ln\lambda-\lambda x_i)$$

由 $$\frac{\mathrm{d}L_1(x_1,\cdots,x_n;\lambda)}{\mathrm{d}\lambda}=\frac{n}{\lambda}-n\overline{x}=0$$

得 $$\hat{\lambda}=\frac{1}{\overline{x}}$$

例8 设$x_1,x_2,\ldots,x_n$是正态总体$N(\mu,\sigma^2)$的样本观察值, 其中μ,σ^2是未知参数, 试求μ和σ^2的最大似然估计值.

解 似然函数

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

$$\ln L(x_1,\cdots,x_n;\mu,\sigma^2)=$$

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

$$\ln L(x_1,\cdots,x_n;\mu,\sigma^2)=$$

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

由

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

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

得μ和σ^2的最大似然估计值为

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

最大似然估计量为

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

作业 习题6-2 第161页开始
第1,2,6题