数理统计

冯伟

数学科学学院

wfeng_323@buaa.edu.cn

- 统计推断
  - 参数统计推断
  - 非参数统计推断

- 统计推断
  - 估计问题
    - 参数估计(主要)
    - 非参数估计
  - 检验问题
    - 参数假设检验
    - 非参数假设检验

上页 下页 返回

第二章 参数估计

一、参数及其估计

二、替换原理（矩估计法）

三、极大似然估计法

上页 下页 返回

设X为一总体, 它的分布族为$\{F_\theta(x),\theta\in\Theta\}$

(有的材料记为统计模型$\{P_\theta,\theta\in\Theta\}$)

- 点估计
  - 矩估计方法
  - 极大似然估计法
  - 顺序统计量法
- 区间估计

上页 下页 返回

第一节 点估计

例1 某灯泡厂灯泡的寿命$X \sim N(\mu,\sigma^2)$，当μ,σ^2未知，试估计μ,σ^2

例2 设总体X的密度函数为

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

试估计其参数λ

例3 已知水文站最高水位X服从$\Gamma(\alpha，\beta)$,即

$$f(x)=\begin{cases}\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0\\ 0, & x\le 0\end{cases}$$

试估计其参数$\alpha，\beta$

一、参数及其估计

设统计模型为$\{P_\theta, \theta \in \Theta\}$，任何与总体有关的待估计的量可以看 成是参数空间$\Theta$上的实值函数$q(\theta)$，通常我们也称其为参数。

但上述有关参数的概念并不仅仅局限于参数统计模型，在非参数统计模型中亦存在，如总体均值$\mathrm{E}_P(X)$和方差$\mathrm{VAR}_P(X)$等总体的

上页 下页 返回

特征数也都是参数。一般地，有关参数和估计量，我们有如下定义。

定义 定义在参数空间Θ上的一个实值函数$g(\theta)$称为**参数（Parameter）**，而用来估计参数 $g(\theta)$ 的实值统计量 $T(X)$称为$g(\theta)$的**估计量（Estimate）**，**简称估计。**

常用记号$\hat{g}$来表示估计，如$\hat{\theta}=\hat{\theta}(X)$来表示$\theta$的估计。

上页 下页 返回

二、频率替换原理

1. 频率替换法

考虑n次独立重复试验，每次试验有k种可能的结果 $v_1,v_2,\cdots,v_k$，每个结果发生的概率p_i是未知的，且$\sum_{i=1}^{k} p_i = 1$. 用n_i表示n次试验中结果v_i发生的次数，则p_i最简单的直观估计是n_i/n，即

$$\hat{p}_i = \frac{n_i}{n}, i = 1,2,\cdots,k.$$

上页 下页 返回

例4 考虑男性总体。设男性的职业类型分为五类，用$v_i = i(i=1,2,3,4,5)$表示，p_i是总体中从事第 i 种职业男性的比例。708个丹麦男性的职业数据如下：

i	1	2	3	4	5	
n_i	23	84	289	217	95	$n=\sum_{i=1}^{5} n_i = 708$
$\hat{p}_i$	0.03	0.12	0.41	0.31	0.13	$\sum_{i=1}^{5} \hat{p}_i = 1$

上页 下页 返回

其次，考虑概率$p_i(i=1,2,\cdots,k)$的连续函数$q(p_1,p_2,\cdots,p_k)$的估计问题。根据频率替换原理，最自然的方法就是用样本频率$\frac{n_1}{n},\frac{n_2}{n},\cdots,\frac{n_k}{n}$代替未知的总体频率$p_1,p_2,\cdots,p_k$，即使用

$$\boxed{T(X_1,X_2,\cdots,X_n)=q\left(\frac{n_1}{n},\frac{n_2}{n},\cdots,\frac{n_k}{n}\right)}$$

来估计$q(p_1,p_2,\cdots,p_k)$.

例如在例4中已知第4和第5种职业相于蓝领，

上页 下页 返回

第2和第3种职业相于白领，我们感兴趣的是蓝领工人和白领工人比率的差，即

$$q(p_1,p_2,\cdots,p_5)=(p_4+p_5)-(p_2+p_3).$$

使用频率替换原理，有

$$T(X_1,X_2,\cdots,X_n)=\left(\frac{n_4}{n}+\frac{n_5}{n}\right)-\left(\frac{n_2}{n}+\frac{n_3}{n}\right)$$

$$=0.44-0.53$$

$$=-0.09$$

上页 下页 返回

在实际问题中，常见的情形是：每次试验结果发生的概率$p_1, p_2, \cdots, p_k$不是独立变化的，而是依赖于m维参数$\theta = (\theta_1, \theta_2, \cdots, \theta_m)$的连续函数，

$$\begin{cases} p_1 = h_1(\theta_1, \theta_2, \cdots, \theta_m) \\ \qquad \cdots \\ p_k = h_k(\theta_1, \theta_2, \cdots, \theta_m) \end{cases},$$

我们需要估计的是θ的部分分量或函数$q(\theta)$，

通过反解即可将θ表示成p_i的函数，假如有表达式$q(\theta)=g(p_1,p_2,\cdots,p_k)$，且$g$在区域

$$D_k=\left\{(p_1,p_2,\cdots,p_k): p_i\ge 0,\sum_{i=1}^{k}p_i=1\right\}$$

上有定义和连续，则由频率替换原理可得

$$\boxed{T(X_1,X_2,\cdots,X_n)=g\left(\frac{n_1}{n},\frac{n_2}{n},\cdots,\frac{n_k}{n}\right),}$$

即为$q(\theta)$的频率替换估计。

需要注意的是在上述估计过程中可能得到的估计是不唯一的，举例说明如下。

例5 考虑具有两个等位基因的单个基因的遗传平衡问题。如果三种不同的基因型是可辨识的，这就导致假设个体三种基因型发生的频率为

$$p_1 = \theta^2, p_2 = 2\theta(1-\theta), p_3 = (1-\theta)^2,$$

其中$0 < \theta < 1$, 即著名的Hardy-Weinberg模型。

上页 下页 返回

如果n_i表示n次试验中基因型为i的人数，则(n_1, n_2, n_3)服从参数为(p_1, p_2, p_3)的多项分布，其中p_1, p_2, p_3的表示如上所示。现估计同位基因之一发生的频率θ，反解可得

$$\theta = \sqrt{p_1}, \theta = 1 - \sqrt{p_3}, \theta = p_1 + \frac{p_2}{2}$$

由替换原理可得θ的三个不同的估计

$$\hat{\theta}_1 = \sqrt{\frac{n_1}{n}}, \hat{\theta}_2 = 1 - \sqrt{\frac{n_3}{n}}, \hat{\theta}_3 = \frac{n_1}{n} + \frac{n_2}{2n}.$$

关于这三个估计优劣留以后讨论。

上页 下页 返回

2. 矩估计法

矩估计法的主要思想是基于替换原理。

设总体的前r个原点矩存在，即

$$m_j(\theta)=E_\theta(X^j),\ j=1,2,\cdots,r,\theta=(\theta_1,\theta_2,\cdots,\theta_m)$$

相应的样本的前r个原点矩为

$$\hat{m}_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j,\ j=1,2,\cdots,r.$$

假设需要估计$q(\theta)$，先将其表示为前r个原点矩的函数，即$q(\theta)=g(m_1(\theta),m_2(\theta),\cdots,m_r(\theta))$.

由替换原理可得$q(\theta)$的估计为

$$\boxed{T(X_1, X_2, \cdots, X_n) = g(\hat{m}_1, \hat{m}_2, \cdots, \hat{m}_r).}$$

这种方法称为矩估计法，所得的估计量称为矩估计量（**Moment Estimate**）。

例6 求总体均值μ和方差σ^2的矩估计。

例7 设总体X服从$[\theta_1, \theta_2]$上的均匀分布，

$X_1, X_2, \cdots, X_n$是来自总体的样本，求θ_1和θ_2的矩估计。

解　因为均匀分布的均值和方差为

$$E(X)=\frac{\theta_1+\theta_2}{2}, Var(X)=\frac{(\theta_2-\theta_1)^2}{12}.$$

令

$$\begin{cases} \dfrac{\theta_1+\theta_2}{2}=\overline{X} \\ \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\dfrac{(\theta_2-\theta_1)^2}{12}, \end{cases}$$

求解可得

$$\begin{cases} \hat{\theta}_1=\overline{X}-\sqrt{\dfrac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \\ \hat{\theta}_2=\overline{X}+\sqrt{\dfrac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \end{cases}.$$

矩估计

•问题　设总体X的分布族为$\{F_\theta(x):\theta\in R^k\}$, EX^k存在,

且未知参数θ可表示为总体原点矩的函数。

•理论依据　样本矩依概率收敛于总体矩。

•矩方程

$$a_l(\theta_1,\theta_2,\cdots,\theta_k)=A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l, l=1,2,\cdots,k.\quad（1）$$

这矩方程确定了含k个未知参数的方程组，其中a_l是总体的l阶原点矩。

•矩估计量　上面方程(1)的解称为参数θ的矩估计量，记为$\hat{\theta}$。并把这种求估计量的方法称为矩方法。

上页　下页　返回

例8 设$X \sim \Pi(\lambda)$,求均值与方差的矩估计。

- 矩估计的优点
  - 不依赖总体的分布，简便易行
  - 只要n充分大，精确度也很高。
- 矩估计的缺点
  - 矩估计的精度较差；
  - 要求总体的某个k阶矩存在；
  - 要求未知参数能写为总体的原点矩的函数形式

注意：

1. 总体不一定存在适当阶的矩。

例 考虑Cauchy分布，其密度函数为

$$f(x,\theta)=\frac{1}{\pi(1+(x-\theta)^2)},-\infty<x<+\infty,$$

其各阶矩均不存在。

2. 对相同的参数 $q(\theta)$，存在多个矩估计。

例如，考虑总体是参数为λ的Poisson分布，

λ既是总体的均值，又是总体的方差。

上页 下页 返回

三、极大似然估计

- 问题 设总体X的密度函数为$f(x, \theta)$, θ是未知参数。
- 基本思想 就是选取θ，使得样本落在取定的样本值的邻域内的概率最大。
- 似然函数 $L(x_1,x_2,\ldots,x_n, \theta)$
- 极大似然估计 如果L在$\hat{\theta}$达到最大值，则称$\hat{\theta}$是θ的极大似然估计。这种求估计量的方法称为极大似然法。
- 理论依据 只要n足够大，极大似然估计和未知参数的真值可以任意接近。

- **极大似然估计的优点** 不仅吸收了提供的样本参数的信息，也利用了分布函数形式已知的有利条件，因此得到的估计量的精度一般较高。
- **极大似然估计的缺点** 要求必须知道总体的分布函数形式

例9 设总体X服从参数为λ的指数分布，即有概率密度

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

又$x_1,x_2,\ldots x_n$为来自于总体的样本值，试求λ的极大似然估计

上页 下页 返回

解 似然函数为

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

于是 $\ln L = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$

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

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

经验证，$\ln L(\lambda)$ 在 $\lambda = \hat{\lambda} = \dfrac{1}{\bar{x}}$ 处达到最大，

所以 $\hat{\lambda} = \dfrac{1}{\bar{x}}$ 是λ的极大似然估计。

上页 下页 返回

现来看离散型总体的极大似然估计

一般地，若总体**X**是离散型的随机变量，有分布律(分布列)

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_k & \cdots \\ p(a_1;\theta) & p(a_2;\theta) & \cdots & p(a_k;\theta) & \cdots \end{pmatrix}$$

$\boldsymbol{\theta}$是未知参数 $(\theta \in \Theta)$

设$\boldsymbol{x_1, x_2, \ldots x_n}$为来自于总体**X**的样本值, 其中

$x_i \in \{a_1, a_2, \cdots, a_k, \cdots\}, i = 1, 2, \cdots, n$

则似然函数为 $L(\theta) = L(x_1, x_2, \cdots, x_n; \theta)$

$$= p(x_1;\theta)\, p(x_2;\theta) \cdots p(x_n;\theta)$$

上页 下页 返回

如果有一个统计量$\hat{\theta}(x_1,x_2,\cdots,x_n)$ 使

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

则称$\hat{\theta}(x_1,x_2,\cdots,x_n)$ 是θ的极大似然估计量。

例10　设总体X 服从参数为λ的泊松分布，

即X有分布列（分布律）

$$p(k;\lambda) = P\{X = k\} = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0,1,2,\cdots \quad \lambda > 0$$

λ是未知参数，试求其极大似然估计。

解

上页 下页 返回

例11 考虑n次独立重复试验，每次试验有k种可能的结果$v_1,v_2,\cdots,v_k$，每个结果发生的概率p_i是未知的，且$\sum_{i=1}^{k} p_i = 1$. 用n_i表示n次试验中结果v_i发生的次数，试证明

$$\hat{p}_i = \frac{n_i}{n}(i = 1,2,\cdots,k)$$

是p_i的MLE。

上页 下页 返回

例12 设总体X在$[a,b]$上服从均匀分布，其中a和b是未知参数，试求它们的MLE。

解　设样本为$x_1, x_2, \cdots, x_n$,则似然函数为

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

从这个表达式可知，$b-a$越小，$L(a,b;x)$就越大，因此a应尽可能的大，而b应尽可能

上页　下页　返回

的小。又$a \le x_{(1)} \le x_{(n)} \le b$，故有

$$\hat{a} = x_{(1)} = \min\{x_i\}, \hat{b} = x_{(n)} = \max\{x_i\}.$$

注：常常可以证明矩估计或MLE存在，但无法获得解的解析表达式（显式解），这时可用迭代法等求数值解。

上页 下页 返回

第二节 点估计量的优良性

- 无偏性
- 有效性
- 相合性

上页 下页 返回

均方误差准则

假设用$T(x)$作为参数$q(\theta)$的估计量，评价估计优劣的一个自然准则可定义如下：

$$MSE_\theta(T)=R(\theta,T)=\boxed{E(T(x)-q(\theta))^2}$$

称上式为**均方误差**，简记为**MSE**。

(Mean Squared Error)

如果$R(\theta,T)<+\infty$，则MSE由T的均值和方差确定，即

$$\boxed{R(\theta,T)=Var_\theta(T(x))+b^2(\theta,T),}$$

上页 下页 返回

$$\boxed{R(\theta,T)=Var_\theta(T(x))+b^2(\theta,T),}$$

其中 $$\boxed{b(\theta,T)=E_\theta(T(x)-q(\theta)),}$$

称$b(\theta,T)$为用$T(x)$估计$q(\theta)$产生的偏差。（bias）

例1 求正态总体$N(\mu,\sigma^2)$均值μ和方差σ^2的MLE的均方误差。

$$b(\theta,\overline{X})=E(\overline{X})-\mu=0,$$

$$R(\theta,\overline{X})=Var(\overline{X})=\frac{\sigma^2}{n},$$

上页 下页 返回

$$b(\theta,\hat{\sigma}^2)=E(\hat{\sigma}^2)-\sigma^2=-\frac{\sigma^2}{n},$$

$$R(\theta,\hat{\sigma}^2)=Var(\hat{\sigma}^2)+b^2(\theta,\hat{\sigma}^2)=\frac{\sigma^4(2n-1)}{n^2}.$$

设$S(x)$和$T(x)$是参数$q(\theta)$的两个估计，如果对所有$\theta\in\Theta$不等式

$$R(\theta,T)\le R(\theta,S)$$

成立，且对某些θ严格不等式成立，则称T比S好，也说S是非容许的。

(Inadmissible)

上页 下页 返回

从均方误差可知，我们自然希望估计的MSE越小越好。

用G_q表示$q(\theta)$所有可能估计组成的类，如果在G_q中存在一个元T^*使得对任一$T\in G_q$，有

$$R(\theta,T^*)\le R(\theta,T)$$

对所有的$\theta\in\Theta$成立，则$T^*(x)$应是$q(\theta)$的最好估计。

遗憾的是，这样的估计 T^* 并不存在。因为倘若这样的估计 $T^*(x)$ 存在，那么对任一 $\theta_0 \in \Theta$，令 $S(x) \equiv q(\theta_0)$，则 $R(\theta_0, S) = 0$，这样

$$R(\theta_0, T^*) = E_{\theta_0}(T^*(x) - q(\theta_0))^2 \le R(\theta_0, S) = 0,$$

即 $T^*(x) = q(\theta_0)$. 由 θ_0 的任意性，因此这样 $T^*(x)$ 不存在。

平凡估计

(Trivial Estimate)

由此可见，均方误差一致达到最小的最优估计并不存在，那么应如何评判和寻找优良的估计呢？方法之一是对估计提出一些合理性的要求，将那些诸如不合理的平凡估计排除在外，然后在满足合理性要求的估计类中寻找优良的估计。无偏性便是一种常用的合理性要求。

一 无偏性

定义1 设统计模型为$\{P_\theta, \theta \in \Theta\}$，$q(\theta)$未知参数，$X_1, X_2, \cdots, X_n$是来自总体的样本，$T$是一个统计量，如果对所有的$\theta \in \Theta$有

$$E_\theta(T(X)) = q(\theta)$$

成立，即$b(\theta, T) = 0$，则称$T(X)$是参数$q(\theta)$的**无偏估计量(Unbiased Estimate)**。

上页 下页 返回

例2　设总体X服从正态分布$N(\mu,\sigma^2)$，

其均值μ和方差σ^2的MLE估计为

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

试讨论它们的无偏性。

解　容易验证$\overline{X}$是无偏的。

因为
$$\frac{n\hat{\sigma}^2}{\sigma^2}\sim\chi^2(n-1)$$

上页　下页　返回

且$E(\frac{n\hat{\sigma}^2}{\sigma^2})=n-1$，所以 $E(\hat{\sigma}^2)=\frac{n-1}{n}\sigma^2$.

故$\hat{\sigma}^2$是σ^2有偏估计。然而 $E(\frac{n}{n-1}\hat{\sigma}^2)=\sigma^2$.

这样 σ^2 的无偏估计为

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

例3 设EX^2存在，判断均值的矩估计的无偏性。

上页 下页 返回

例4 设$X \sim \Pi(\lambda)$,判断均值与方差的矩估计的无偏性。

例5 设$X \sim U[0, \theta]$, 求θ的矩估计和极大似然估计， 并判断其无偏性。再进一步比较两者均方误差的大小。

例6 验证当$X \sim N(\mu,\sigma^2)$时，样本均值是μ的无偏估计，但样本均值的平方就不再是μ^2的无偏估计。样本方差S^2是σ^2的无偏估计，但S不是σ的无偏估计。

上页 下页 返回

注意：

（1）无偏估计可能存在。如果参数$q(\theta)$的无偏估计存在，则称$q(\theta)$是可估的。

若无特别声明，均认为$q(\theta)$是可估参数。

（2）对可估参数 $q(\theta)$，无偏估计一般不唯一。

（3）无偏估计不一定是好的估计，即它可能是非容许的。

（4）一般来说，无偏估计量的函数并不是未知参数相应函数的无偏估计量。即

对θ而言，$\hat{\theta}$是无偏的，但$q(\hat{\theta})$可能是$q(\theta)$的有偏估计。

（5）尽管以往一般将无偏性作为估计量的绝对优良性的标准，但有时一个参数可能不存在无偏估计，或者无偏估计有明显的弊病。因此近来对有偏估计的研究也多了起来。

上页 下页 返回

设$q(\theta)$是可估参数，令

$$U_q=\{T(x):E_\theta(T(x))=q(\theta),\text{对所有 }\theta\in\Theta\}$$

即U_q表示参数$q(\theta)$的所有无偏估计组成的类.

由定义1可知无偏估计的均方误差就是它的方差，即

$$R(\theta,T)=Var_\theta(T(X)).$$

上页 下页 返回

在均方误差准则下，既然最好的估计不存在，那么现在的问题是对无偏估计类U_q而言，同样在均方误差（方差）准则下，最好的无偏估计（一致最小方差无偏估计）是否存在？若存在，它是否是唯一的？如何求？这些就是我们下次需要讨论的主题。

上页 下页 返回

作业 **P84 4,6-8,10,12,14,16,18-22,24-27**

上页 下页 返回