第二节　参数的点估计

一、矩估计

二、最（极）大似然估计

7.2.1 矩估计

总体中常含有未知参数，如何利用抽取的样本来对总体的未知参 数进行估计，是统计推断的重要内容。

7.2.1 矩估计

定义7. 2. 1　设总体X的分布中含有未知参数θ, 且 $X_1, \cdots, X_n$是来自于总体X的样本, $x_1, \cdots, x_n$是相应的一个样本值. 若构造一个适当的统计量$\hat{\theta}$（$X_1,X_2,\ldots,X_n$）, 用其观测值 $\hat{\theta}$（$x_1, x_2,\ldots, x_n$）作为θ的近似值, 则称 $\hat{\theta}$（$X_1,X_2,\ldots,X_n$）是θ的一个估计量, 并称$\hat{\theta}$（$x_1, x_2,\ldots, x_n$）是θ的一个估计值.

θ的估计量与估计值统称估计, 都简记为$\hat{\theta}$.

7.2.1 矩估计

矩估计最早由英国统计学家皮尔逊于1894年提出，被称为**矩法**或**矩估计法**，其基本思想是：

1）用样本矩来估计总体矩；

2）用样本矩的连续函数来估计总体矩的连续函数,从而得出参数估计.

矩估计法就是将**样本的 k 阶矩作为总体的对应的 k 阶矩的估计量, 进而对总体中的未知参数进行估计的常用方法。**

理论依据：辛钦大数定律

7.2.1 矩估计

样本**k**阶原点矩 $A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$ ← 是统计量，不包含未知参数

总体**k**阶原点矩 $m_k = E(X^k)$ ← 是数字特征，包含未知参数

7.2.1 矩估计

辛钦大数定律

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X)\right| < \varepsilon\right) = 1$$

$$\Leftrightarrow \lim_{n\to+\infty} P\left(|A_1 - m_1| < \varepsilon\right) = 1.$$

大数定律的启示

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^k - E(X^k)\right| < \varepsilon\right) = 1$$

$$\Leftrightarrow \lim_{n\to+\infty} P\left(|A_k - m_k| < \varepsilon\right) = 1.$$

7.2.1 矩估计

大数定律的启示

$$\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^k - E(X^k)\right| < \varepsilon\right) = 1$$

$$\Leftrightarrow \lim_{n\to+\infty} P\left(|A_k - m_k| < \varepsilon\right) = 1.$$

矩方程(组)

$$A_1 = m_1$$

$$A_2 = m_2$$

$$\cdots\cdots$$

矩方程(组)的选取原则

(1) k 从 1 开始由小往大依次选取.

(2) 有几个待估计参数，则应建立由几个矩方程形成的方程组.

例1 设总体X服从二项分布B(l,p), l已知, p未知, $X_1,\cdots,X_n$是来自于 X 的样本, 求 p 的矩估计量.

解

$$A_1 = m_1$$

$$\Rightarrow \overline{X} = E(X)$$

$$\Rightarrow \overline{X} = lp \qquad \Rightarrow \hat{p} = \frac{\overline{X}}{l}$$

矩方程(组)

$$A_1 = m_1$$

$$A_2 = m_2$$

$$\cdots\cdots$$

7.2.1 矩估计

例2 设总体X服从指数分布$e(\lambda)$，$X_1,\cdots,X_n$是来自于X的样本,求 λ 的矩估计量.

矩方程(组)

$$A_1 = m_1$$

$$A_2 = m_2$$

$$\cdots\cdots$$

解

$$A_1 = m_1$$

$$\overline{X} = E(X)$$

$$\overline{X} = \frac{1}{\lambda} \Rightarrow \hat{\lambda} = \frac{1}{\overline{X}}$$

7.2.1 矩估计

例3 设总体X的数学期望和方差分别为μ和σ^2，但均未知，$X_1,\cdots,X_n$是来自于 X 的样本，求μ和σ^2的矩估计量.

矩方程(组)

$$A_1=m_1$$
$$A_2=m_2$$
$$\cdots\cdots$$

解 $\begin{cases} A_1=m_1 \\ A_2=m_2 \end{cases} \Rightarrow \begin{cases} A_1=m_1 \\ A_2-A_1^2=m_2-m_1^2 \end{cases}$

$$\Rightarrow \begin{cases} \overline{X}=E(X)=\mu \\ S_n^2=D(X)=\sigma^2 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\mu}=\overline{X} \\ \hat{\sigma}^2=S_n^2 \end{cases}$$

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=A_2-A_1^2.$$

$$m_2-m_1^2=E(X^2)-(E(X))^2=D(X).$$

7.2.1 矩估计

例4. 设总体 X 的密度函数为 $f(x)=\frac{1}{2\theta}e^{-\frac{|x|}{\theta}},(x\in\mathrm{R},\theta>0)$,

$X_1,\cdots,X_n$是来自于 X 的样本, 求θ 的矩估计量.

矩方程(组)

$$A_1=m_1$$
$$A_2=m_2$$
$$\cdots\cdots$$

解 方法一

由于$m_1=$E(X)= 0, 与 θ 无关, 故考虑使用 m_2建立矩方程。

$$m_2=E(X^2)=\int_{-\infty}^{+\infty}x^2\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}\mathrm{d}x=2\theta^2.$$

$$A_2=m_2\Rightarrow A_2=E(X^2)\Rightarrow A_2=2\theta^2\Rightarrow\hat{\theta}=\sqrt{\frac{A_2}{2}}.$$

例4. 设总体 X 的密度函数为 $f(x)=\dfrac{1}{2\theta}\mathrm{e}^{-\frac{|x|}{\theta}},(x\in\mathrm{R},\theta>0)$,

$X_1,\cdots,X_n$是来自于 X 的样本, 求θ 的矩估计量.

矩方程(组)

$$A_1=m_1$$

$$A_2=m_2$$

$$\cdots\cdots$$

解 方法二

将 $Y=|X|$ 视为总体, 则样本为 $|X_1|,\cdots,|X_n|$.

$$m_1=E(Y)=E(|X|)$$

$$=\int_{-\infty}^{+\infty}|x|\frac{1}{2\theta}\mathrm{e}^{-\frac{|x|}{\theta}}\,\mathrm{d}x$$

$$=\theta.$$

$$A_1=m_1\Rightarrow\overline{Y}=E(Y)$$

$$\Rightarrow\hat{\theta}=\overline{Y}=\frac{1}{n}\sum_{i=1}^{n}|X_i|.$$

7.2.1 矩估计

矩估计方法优缺点

优点： 简单.

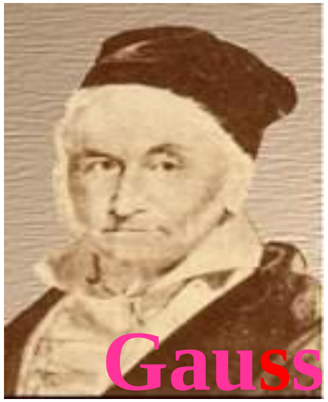
缺点： (1) 对同一个待估计参数, 选取不同的矩方程所得的估计结果可能不同.

(2) 总体的矩若不存在, 则无法使用该方法.

7.2.2 最（极）大似然估计

极大似然法是在**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高斯在1821年提出的．



然而，这个方法常归功于英国统计学家费歇尔，主要是由于其在1912年重新提出并证明了这个方法的一些性质。

基本原理:

在一次试验就出现的事件在所有可能的结果中理应具有较大的发生概率.

7.2.2 最（极）大似然估计

最大似然估计方法的基本原理

一般地，设随机次试验有$A, B, C, \dots$.若干个可能出现的结果，若 在一次试验中结果 A 已出现, 则一般说来当时的试验条件应最有利于 结果 A 的出现, 从而使结果 A 出现的概率为最大, 称之为最大似然原理.

由此, 当未知参数θ有多个可供作为估计值的选择时, 自然应选择使结果 A 出现的概率为最大的那一个$\hat{\theta}$ 作为 θ 估计值, 这就是最大似然估计法选择未知参数估计值的基本思想.

离散型总体的最大似然估计方法

设离散型总体 X 的样本为 $X_1,\cdots,X_n$，其观测值为 $x_1,\cdots,x_n$，总体X的分布律形式为

$$P(X=x_i)=p(x_i;\theta_1,\cdots,\theta_k),$$

其中 $\theta_1,\cdots,\theta_k\in\Theta$ 为未知参数.

7.2.2 最（极）大似然估计

离散型总体的最大似然估计方法

样本的联合分布： $L(\theta_1,\cdots,\theta_k)=P(X_1=x_1,\cdots,X_n=x_n)$

$$=\prod_{i=1}^{n}P(X_i=x_1)=\prod_{i=1}^{n}p(x_i;\theta_1,\theta_2,\ldots,\theta_k)$$

称$L(\theta_1,\cdots,\theta_k)$为$(\theta_1,\cdots,\theta_k)$的似然函数。

若存在 $\hat{\theta}_1,\cdots,\hat{\theta}_k$， 使得 $L(\hat{\theta}_1,\cdots,\hat{\theta}_k)=\max\limits_{\theta_1,\cdots,\theta_k\in\Theta}L(\theta_1,\cdots,\theta_k)$,

则称 $\hat{\theta}_1,\cdots,\hat{\theta}_k$ 为 $\theta_1,\cdots,\theta_k$ 的最大似然估计.

7.2.2 最（极）大似然估计

例7.2.7 **设总体X服从$\{1,2,\ldots,N\}$上的均匀分布，（$x_1,\cdots,x_n$）**为样本的一个观测值，试求$\widehat{N}_{MLE}$.

解 **设（$X_1,\cdots,X_n$）为取自总体X的样本, 样本出现观测值（$x_1,\cdots,x_n$）的概率为**

$$L(\theta_1,\cdots,\theta_k)=P(X_1=x_1,\cdots,X_n=x_n)$$

$$=\prod_{i=1}^{n}P(X_i=x_i)=\prod_{i=1}^{n}p(x_i;\theta_1,\theta_2,\ldots,\theta_k)$$

$$L(N)=\prod_{i=1}^{n}P(X_i=x_i)=\frac{1}{N^n}$$

欲使似然函数$L(N)$取最大值，则N取最小值。因为$N\geq$max$\{x_1,\cdots,x_n\}=x_{(n)}$

所以N的最大似然估计量$\widehat{N}_{MLE}=X_{(n)}$.

7.2.2 最（极）大似然估计

考虑到对数函数的单调性，以及似然函数的本质为概率，有时会将求 $L(\theta_1,\cdots,\theta_k)$ 最值的问题转化为求 $\ln L(\theta_1,\cdots,\theta_k)$ (称为对数似然函数)最值的问题.

7.2.2 最（极）大似然估计

例5 **设有一大批产品, 其次品率 p $(0 < p < 1)$ 未知. 今从中随意抽取 100 个, 发现有 5 个次品, 试求 p 的最大似然估计值.**

解 设总体为X, 则$X \sim b(1, p)$.

次品记为“1”,合格品记为“0”,观测值实为5个“1”和95个“0”.

似然函数 $L(p) = p^5(1-p)^{95}$.

取对数 $\ln L(p) = 5\ln p + 95\ln(1-p)$.

令 $$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p} = \frac{5}{p} - \frac{95}{1-p} = 0,$$

$$\Rightarrow \hat{p} = 0.05.$$

7.2.2 最（极）大似然估计

连续型总体的最大似然估计方法

设连续型总体 X 的样本为 $X_1, \cdots, X_n$，其观测值为 $x_1, \cdots, x_n$，总体的密度函数形式为 $f(x;\theta_1, \cdots, \theta_k)$，

其中$\theta_1, \cdots, \theta_k \in \Theta$为未知参数.

对于连续型总体有 $P(X = x_i;\ \theta_1, \cdots, \theta_k) \equiv 0.$

所以取似然函数为 $L(\theta_1, \cdots, \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \cdots, \theta_k)$.

结论：

对于离散型或连续型总体，只要知道其概率分布或密度函数，总可以得到一个关于参数θ的函数$L(\theta)$,称之为似然估计。

7.2.2 最（极）大似然估计

例6 设总体X服从指数分布$e(\lambda)$,$X_1,\cdots,X_n$是来自于X的样本，$x_1,\cdots,x_n$为其观测值，求λ的最大似然估计量.

解 X的密度函数为 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}$

似然函数

$$L(\lambda)=\prod_{i=1}^{n}f(x_i;\lambda)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}=\lambda^n e^{-\lambda n\bar{x}}.\ (x_i>0,i=1,\cdots,n).$$

取对数 $\ln L(\lambda)=n\ln\lambda-\lambda n\bar{x}.$

令 $\dfrac{d\ln L(\lambda)}{d\lambda}=\dfrac{n}{\lambda}-n\bar{x}=0,\ \Rightarrow\hat{\lambda}=\dfrac{1}{\overline{X}}.$

7.2.2 最（极）大似然估计

例7.2.10 设总体$X\sim U[0,\theta]$, $X_1,\cdots,X_n$是来自于X的样本, $x_1,\cdots,x_n$为其观测值, 求 θ 的最大似然估计量.

解 X的密度函数为 $f(x)=\begin{cases}\dfrac{1}{\theta}, & 0\le x\le\theta,\\ 0, & \text{其他}.\end{cases}$

求导数并不是最大似然估计的必要过程.

似然函数 $L(\theta)=\prod_{i=1}^{n}f(x_i)=\prod_{i=1}^{n}\dfrac{1}{\theta}=\dfrac{1}{\theta^n},(0\le x_i\le\theta,i=1,\cdots,n).$

$x_i\le\theta,$

$$\Rightarrow \theta\ge\max\{x_i\}\Rightarrow\frac{1}{\theta^n}\le\frac{1}{(\max\{x_i\})^n}$$

$$\therefore \hat{\theta}=\max\{X_i\}.$$

小结

两种求点估计的方法: $\begin{cases} \text{矩估计法} \\ \text{最大似然估计法} \end{cases}$

在统计问题中往往先使用最大似然估计法，在最大似然估计法使用不方便时，再用矩估计法.

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