西安電子科技大學
XIDIAN UNIVERSITY

第7章　参数估计

7.1 点估计

设有一个统计总体，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) .

现从该总体抽样，得样本 $X_1, X_2, \ldots, X_n$

要依据该样本对参数θ 作出估计, 或估计 θ 某个已知函数$g(\theta)$.

这类问题称为参数估计.

参数估计两种方法：
- 矩方法
- 最大似然估计

一、矩估计

原理：用样本矩代替相应的总体矩，（大数定律）

回忆矩的记法：

	原点矩	中心矩
样本k阶矩	$a_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \quad (k=1,2,\ldots)$	$m_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k \quad (k=1,2,\ldots)$
总体k阶矩	$\mu_k=E(X^k)$	$d_k=E\left[X-E(X)\right]^k$

在k阶矩存在的情况下，大数定律 $a_k \xrightarrow{P} \mu_k$ ，$m_k \xrightarrow{P} d_k$ 。

$$\overline{X}\sim E(X) \qquad S^2\sim D(X)$$

一、矩估计

设 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ $E(X)=\int_{-\infty}^{+\infty} xf(x,\theta)dx=g_1(\theta_1,\theta_2,\cdots,\theta_k)$

即： $\overline{X}\simeq g_1(\theta_1,\theta_2,\cdots,\theta_k)$

对于k个未知数， $E(X^j)=\int_{-\infty}^{+\infty} x^j f(x,\theta)dx=g_j(\theta_1,\theta_2,\cdots,\theta_k)$

$$\Rightarrow g_j(\theta_1,\theta_2,\cdots,\theta_k)=a_j \quad j=1,2,\cdots,k$$

从这 k 个方程中解出 $\theta_1,\theta_2,\cdots,\theta_k$ 的估计值： $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$

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

θ_1　真正的总体参数，未知常数，非随机。

$\hat{\theta}_1$　由样本对参数做的一个估计，是统计量，与样本有关，随机性。

注意：未知常数 ≠ 随机变量

总体 $X \sim F(x,\theta)$，$E(X)=\mu$；估 $\theta_1=\mu \quad \theta_2=\sigma^2$

$\hat{\mu}=\overline{X}$ 用样本一阶矩代替总体一阶矩。

估 σ^2 ： $D(X)=E\left[X-E(X)\right]^2$ 总体二阶中心距需用样本二阶中心距代替

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

矩估计：
- 优点：简单，不需要事先知道什么分布。
- 缺点：矩估计量不唯一。

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1： $X \sim U(0,\theta)$ 估 θ

解 算一阶矩： $E(X)=\int_{-\infty}^{+\infty} x\frac{1}{\theta}I_{(0,\theta)}(x)dx=\frac{1}{\theta}\int_0^{\theta} xdx=\frac{\theta}{2}$

$\overline{X}$ 用样本距代替

$$\Rightarrow \hat{\theta}=2\overline{X}$$

例2： $X \sim U(\theta_1,\theta_2)$，算一阶矩和二阶距。

解

$$E(X)=\frac{\theta_1+\theta_2}{2} \qquad D(X)=\frac{(\theta_2-\theta_1)^2}{12}$$

$$\begin{cases} \dfrac{\theta_1+\theta_2}{2}=\overline{X} \\ \dfrac{(\theta_2-\theta_1)^2}{12}=m_2 \end{cases} \Rightarrow \begin{cases} \theta_1+\theta_2=2\overline{X} \\ \theta_2-\theta_1=2\sqrt{3m_2} \end{cases} \Rightarrow \theta_1,\theta_2=\overline{X}\pm\sqrt{\frac{3}{n}\sum_n(X_i-\overline{X})^2}$$

例2： $X \sim P(\lambda)$

解 $E(X)=\lambda \Rightarrow \hat{\lambda}=\overline{X}$

$D(X)=\lambda \Rightarrow \hat{\lambda}=m_2$

不唯一时，选择哪个估计量好？

原则：从低到高，能用低阶矩处理的就不用高阶矩。

矩估计的步骤：

① 计算总体前k阶矩，找出这些矩与参数 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$的关系式；

② 用样本矩代替总体矩，解这k个方程组。

二、最大似然估计（MLE）

什么是似然函数？

设总体X为连续型随机变量，其分布律为 $f(x,\theta)$；

设总体X为离散型随机变量，其分布律为: $P(X=x)=p_i(\theta)$ 记为 $p(x_i,\theta)$ 或 $p(x,\theta)$

$$x=x_i,i=1,2,\cdots,n$$

$X_1,X_2,\ldots,X_n$是来自总体X的样本，$(X_1\ldots X_n)$服从：

$$\prod_{i=1}^{n} f(x_i,\theta) \text{ 或 } \prod_{i=1}^{n} p(x_i,\theta)$$

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

给定θ，L称为$(X_1\ldots X_n)$的密度函数 联合密度函数；

给定 x，L作为 $(\theta_1,\theta_2,\cdots,\theta_k)$的函数，称为似然函数。

密度函数和似然函数的区别

二、最大似然估计（MLE）

最大似然法的基本思想：

假定一个随机试验E有若干个可能的结果$A_1, A_2, \cdots A_n$ 如果只进行了一次试验，而A_k出现了，那么有理由认为试验的条件对结果A_k的出现最有利，即试验E出现结果中的A_k概率最大。

已知：$x=(x_1,\cdots,x_n)$,求 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 使 $L(x,\ \theta)=\max$ 。

二、最大似然估计（MLE）

已知：$x=(x_1,\cdots,x_n)$,求 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 使 $L(x,\ \theta)=\max$ 。

求似然函数$L(\theta)$ 的最大值点，可以应用微积分中的技巧。由于ln(x)是 x 的增函数, ln$L(\theta)$与$L(\theta)$在θ 的同一值处达到它的最大值，假定θ是一实数，且ln$L(\theta)$是 θ的一个可微函数。通过求解方程：

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

可以得到θ 的最大似然估计。

最大似然估计的步骤：

① 写出似然函数$L(\theta)$;

② 取自然对数ln$L(\theta)$;

③ 令：$\dfrac{\partial \ln L(\theta)}{\partial \theta_i}=0$ ，求 $\hat{\theta}_i$ 。

例1： 设总体 $X \sim N(\mu,\sigma^2)$ ，μ,σ^2 未知。$x_1,\ldots,x_n$ 是来自 X 的样本值，试求 μ,σ^2 的最大似然估计量。

解 X 的概率密度为 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$， $-\infty<x<\infty$

似然函数为 $L(x,\theta)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=(2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2]$

于是 $LnL(x,\theta)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

$$LnL(x,\theta)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\theta_2-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

$$\frac{\partial LnL(x,\theta)}{\partial\theta_1}=\frac{1}{\theta_2}(\sum_{i=1}^{n}x_i-n\theta_1)=0$$

$$\frac{\partial LnL(x,\theta)}{\partial\theta_2}=-\frac{n}{2\theta_2}+\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2=0$$

解得 $\theta_1=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$ $\qquad\theta_2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

μ,σ^2 的最大似然估计量为：

$$\mu=\bar{X},$$

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

例： $X \sim U(0,\theta)$ 求 $\hat{\theta}$

解 $X \sim f(x,\theta)=\frac{1}{\theta}I_{(0,\theta)}(x)$

$L(x,\theta)=\frac{1}{\theta}\prod_{i=1}^{n}I_{(0,\theta)}(x_i)$ 似然函数，θ 未知，x已知，固定 x 求 θ 。

$0<x_1<\theta,0<x_1<\theta,\cdots$

$$=\frac{1}{\theta^n}\prod_{i=1}^{n}I_{(x_{(n)},\infty)}(\theta)$$

而 $\frac{1}{\theta^n}$ 是 θ 的减函数，即：θ=$x_{(n)}$ 时最大。

$$\Rightarrow \hat{\theta}=X_{(n)}=\max\left\{X_1,X_2,\cdots,X_n\right\}$$

对于离散型随机变量（以二项分布为例）

$$P(X=k)=C_N^k p^k(1-p)^{N-k}\quad k=0,1,2,\cdots$$

用大写字母以与样本数做区分。

x_i ， $x_i=0,1,2,\cdots,n$ 样本数

似然函数 $L(x,\theta)=\prod_{i=1}^{n}P(X=x_i)=\prod_{i=1}^{n}\left\{C_N^{x_i}p^{x_i}(1-p)^{N-x_i}\right\}=(\prod_{i=1}^{n}C_N^{x_i})p^{\sum_n x_i}(1-p)^{Nn-\sum_n x_i}$

给定 $x=(x_1,x_2,\cdots,x_n)$，求 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 使 $L(x_1,\cdots,x_n;\ \theta_1,\cdots,\theta_k)=\max$

$$\ln L(x,p)=C+(\sum x_i)\ln p+(Nn-\sum_n x_i)\ln(1-p)$$

$$\frac{d\ln L}{dp}=\frac{1}{p}(\sum x_i)-\frac{(Nn-\sum x_i)}{1-p}=0$$

$$\frac{1}{p}n\overline{x}=\frac{Nn-n\overline{x}}{1-p}\Rightarrow p=\frac{\overline{x}}{N}$$

$$\therefore \hat{p}=\frac{\overline{X}}{N}$$

7.2 点估计的优良性准则

一、无偏性

估计量是随机变量，对于不同的样本值会得到不同的估计值。我们希望估计值在未知参数真值附近摆动，而它的期望值等于未知参数的真值。这就导致无偏性这个标准。

设 $\hat{\theta}(X_1,\cdots,X_n)$ 是未知参数 θ 的估计量，若 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 为 θ 的无偏估计。

例1： $\hat{\mu}=\bar{X}$ 是否无偏？

解

$$E(\bar{X})=E(\sum_{i=1}^{n}X_i/n)=\frac{1}{n}nE(X)=\mu$$

所以 $\bar{X}$ 是总体均值的无偏估计。

总体均值还有无别的无偏估计？

$$\hat{\mu}_1 = \sum_{i=1}^{n} W_i X_i \qquad W_i \geq 0,\ \sum W_i = 1$$

$$E(\hat{\mu}_1) = \sum_{i=1}^{n} W_i E(X_i) = \mu \sum W_i = \mu$$

所以加权平均也是无偏估计，即无偏估计不唯一。

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \hat{\sigma}^2 = m_2$$ 那么 m_2 是否无偏？

$$\sum (X_i - \overline{X})^2 = \sum \left[(X_i - \mu) + (\overline{X} - \mu) \right]^2 = \sum (X_i - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)\sum_{n}(X_i - \mu)$$

$$E\left[\sum (X_i - \mu)^2 \right] = n\sigma^2 \qquad E(\overline{X} - \mu)^2 = \frac{1}{n}\sigma^2 \quad \overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\sum (X_i - \mu) = \sum X_i - n\mu = n\overline{X} - n\mu = n(\overline{X} - \mu)$$

$$\therefore E(m_2) = \frac{1}{n}(n\sigma^2 + \sigma^2 - 2\sigma^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$$ 系统偏小

对 m_2 做修正：$S^2=\dfrac{n}{n-1}m_2=\dfrac{1}{n-1}\sum(X_i-\overline{X})^2$

$E(S^2)=\sigma^2$ 即样本方差是 m_2 的无偏估计。

注意：不是“S是 σ 的无偏估计”。 $E(S)\neq\sigma$

例2： $X\sim U(0,\theta)$ $\hat{\theta}_M=2\overline{X}$ $\hat{\theta}_L=X_{(n)}$

解 $E(\hat{\theta}_M)=2E(\overline{X})=2\cdot\dfrac{\theta}{2}=\theta$ 无偏

计算 $E(X_{(n)})$ 要先计算 $X_1,X_2,\cdots,X_n$ 最大值的分布。按定义求最大值的分布函数：

$$P(X_{(n)}\le x)=P(X_1\le x,X_2\le x,\cdots,X_n\le x)=F^n(x)$$

由于 $X_1,X_2,\cdots,X_n$ 独立同分布，$F(x)=\dfrac{x}{\theta}I_{(0,\theta)}(x)$

$$f_{X_{(n)}}(x)=nF^{n-1}(x)f(x)=n\frac{x^{n-1}}{\theta^{n-1}}\frac{1}{\theta}I_{(0,\theta)}(x)$$

$$\therefore E(X_{(n)})=\int_{-\infty}^{\infty}xf_{X_{(n)}}(x)dx=\int_0^{\theta}x\frac{nx^{n-1}}{\theta^n}dx=\frac{n}{\theta^n}\int_0^{\theta}x^n dx=\frac{n}{n+1}\theta$$ 系统偏小

对 $\hat{\theta}_L$ 做修正，$\dfrac{n+1}{n}X_{(n)}$ 是无偏估计。