概率论与数理统计

第七章 参数估计

第七章　参数估计

7.1 参数估计概念

7.2 矩估计法和极大似然估计法

7.3 估计量的评选标准

7.4 区间估计

第七章　参数估计

7.1 参数估计概念

7.1 参数估计概念

——根据样本给出参数的估计值

θ 是$F(x,\theta)$中的未知参数 $\theta \in \Theta$

样本$(X_1,X_2,\ldots,X_n)$

估计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$

估计值$\hat{\theta}(x_1,x_2,\cdots,x_n)$

θ

点估计

估计区间 $\hat{\theta}_1\left(X_1,\cdots,X_n\right),\hat{\theta}_2\left(X_1,\cdots,X_n\right)$

$\theta \in \left(\hat{\theta}_1,\ \hat{\theta}_2\right).$

区间估计

第七章　参数估计

7.2 矩估计法和极大似然估计法

7.2.1 矩估计法

1 原理：大数定理，样本矩在一定程度上逼近总体矩。

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

总体X的分布函数$F(x;\ \theta_1, \ldots, \theta_k)$,其中$\theta_1,\ldots,\theta_k$为待估参数.
$(X_1,\ldots,X_n)$为X的样本(设X的k阶矩存在), 则

总体的j阶原点矩　　$E(X^j) = a_j(\theta_1,\ldots,\theta_k)$

样本的j阶原点矩　　$\frac{1}{n}\sum_{i=1}^{n} X_i^j = A_j$

2 方法

令　$a_j(\theta_1,\ldots,\theta_k) = A_j \qquad j=1,\ldots,k$

求解方程组,得到解$\hat{\theta}_1,\cdots,\hat{\theta}_k$作为参数$\theta_1,\cdots,\theta_k$的矩估计量.

7.2.1 矩估计法

例 试求总体期望$\theta_1=EX$和方差$\theta_2=DX$的矩估计。

解 令 $\begin{cases} EX = \bar{X} \\ EX^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \end{cases} \Leftrightarrow \begin{cases} \theta_1 = \bar{X} \\ \theta_2 + \theta_1^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \end{cases}$

$$\Leftrightarrow \begin{cases} \theta_2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - (\bar{X})^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 = \tilde{S}^2 \\ \theta_1 = \bar{X} \end{cases} \quad \therefore \hat{\theta}_1 = \bar{X}, \quad \hat{\theta}_2 = \tilde{S}^2.$$

注1 此结论对期望和方差存在的总体都适用，即

$$\hat{E}(X) = \bar{X}, \quad \hat{D}(X) = \tilde{S}^2.$$

注2 估计不唯一，如对总体$P(\lambda)$有 $\hat{\lambda} = \bar{X}$，$\hat{\lambda} = \tilde{S}^2$.

注3 可用 $B_k = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^k$ 估计 $\beta_k = E(X - EX)^k$.

7.2.1 矩估计法

例 设总体$X \sim U[a,b]$, 试由样本$(X_1,X_2,\dots,X_n)$, 求未知参数a,b 的矩估计量。

解 (1) $\begin{cases} \alpha_1 = E(X) = \dfrac{a+b}{2} \\ \beta_2 = D(X) = \dfrac{(b-a)^2}{12} \end{cases} \Rightarrow \begin{cases} a = \alpha_1 - \sqrt{3\beta_2} \\ b = \alpha_1 + \sqrt{3\beta_2} \end{cases}$

(2) $\begin{cases} \hat{a} = A_1 - \sqrt{3B_2} = \bar{X} - \sqrt{3\tilde{S}^2} \\ \hat{b} = A_1 + \sqrt{3B_2} = \bar{X} + \sqrt{3\tilde{S}^2} \end{cases}$

[缺点] 如 $\frac{1}{2},\frac{1}{3},\frac{1}{4},\cdots,\frac{1}{11}$ 为来自$X \sim U[a,b]$的样本观察值，则a, b的估计值为 $\hat{a} = -0.01$，$\hat{b} = 0.414$. 注意到： $x_1 = 0.5 > \hat{b}$

一般 $P(\hat{a} < \min X_i) > 0,\ P(\hat{b} > \max X_i) > 0$

7.2.2极大似然估计法

引例 设$X\sim B(1,p)$, p未知. 设想我们事先知道 p只有两种可能:

$$p=0.8 \quad 或 \quad p=0.1$$

抽样结果: 1, 1, 0, 1.

问: 应如何估计参数 p?

解: $P(X_1=1,X_2=1,X_3=0,X_4=1)=p^3(1-p).$

参数p	观测值出现的概率
0.8	0.1024
0.1	0.0009

故参数p的极大似然估计值是**0.8**

7.2.2极大似然估计法

1. 原理

极大似然原则：已发生的事件，其概率应该最大。

极大似然估计（Maximum Likelihood Estimation, MLE）

总体X的分布形式已知，未知的仅仅是参数$\theta\in\Theta$

$$x_1,\ldots,x_n$$

参数θ的选择应有利于样本观测值的发生，即让这组数据发生的概率达到最大。

7.2.2极大似然估计法

似然函数：

$\theta=(\theta_1,\ldots,\theta_k)$为待估未知参数，

$(x_1,\cdots,x_n)$为样本$(X_1,\cdots,X_n)$的一组观测值。

$$L(x_1,x_2,\cdots,x_n;\boldsymbol{\theta})=\begin{cases}\prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) & \text{其中}f(x;\boldsymbol{\theta})\text{为}X\text{的密度函数}\\ \prod_{i=1}^{n} P_{\boldsymbol{\theta}}(X=x_i) & \text{其中}P_{\boldsymbol{\theta}}(X=x)\text{为}X\text{的分布律}\end{cases}$$

θ 的**极大似然估计值** $\hat{\theta}(x_1,x_2,\cdots,x_n)$：

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

θ 的**极大似然估计量** $\hat{\theta}(X_1,X_2,\cdots,X_n)$

7.2.2极大似然估计法

2. 一般步骤

选$\hat{\theta}$使$L(\theta)$最大

(1) 写出似然函数 $L(\theta)$

(2) 建立似然方程(组) $\dfrac{\partial L(\theta)}{\partial \theta_i} = 0$或$\dfrac{\partial \ln L(\theta)}{\partial \theta_i} = 0.$

(3) 解方程(组)得极大似然估计 $\hat{\theta}$

7.2.2极大似然估计法

例 $X \sim P(\lambda)$，求λ的极大似然估计。

解：设$(x_1,\ldots,x_n)$为样本的一组观测值，

似然函数 $$L(\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1!\cdots x_n!}e^{-n\lambda}$$

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

似然方程 $$\frac{d\ln L}{d\lambda}=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i=0$$

解方程

λ的极大似然估计为： $$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

比如，样本观测值为：10，13，65，18，79，42，65，77，88，123，n=10。则 $\hat{\lambda}=\bar{x}=58$。

7.2.2极大似然估计法

例 设总体$X\sim N(\mu,\sigma^2)$和样本观测值$x_1,\ldots,x_n$，求μ和σ^2的极大似然估计。

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

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

$$\begin{cases}\dfrac{\partial}{\partial\mu}\ln L=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=\dfrac{1}{\sigma^2}(n\bar{x}-n\mu)=0\\[2ex]\dfrac{\partial}{\partial\sigma^2}\ln L=-\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2}+\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$$

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

7.2.2极大似然估计法

例 设总体 $X \sim U[a,b]$和样本观测值$x_1,\dots,x_n$，试求a 和 b 的极大似然估计。

解

$$L(a,b)=\prod_{i=1}^{n} f(x_i;a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a\le x_1,x_2,\cdots,x_n\le b\\ 0, & others\end{cases}$$

$$\because\ a\le x_1,x_2,\cdots x_n\le b,$$

$$a\uparrow, b\downarrow\ \Rightarrow (b-a)\downarrow \Rightarrow L(a,b)\uparrow$$

$$\therefore\ a=\min_{1\le i\le n}\{x_i\}, b=\max_{1\le i\le n}\{x_i\},\quad L(a,b)=\max.$$

即a,b的极大似然估计为 $\hat{a}=X_1^*,\ \hat{b}=X_n^*.$

3. 性质

若 $\hat{\theta}$ 为θ 的极大似然估计，$u=u(\theta)$有唯一反函数

$\theta=\theta(u)$，则 $\hat{u}=u(\hat{\theta})$ 为$u=u(\theta)$的极大似然估计。

例 设总体$X\sim N(\mu,\sigma^2)$，求σ 的极大似然估计。

解 $\sigma>0$，$\sigma=\sqrt{\sigma^2}$有反函数。由上例，$\hat{\sigma}^2=\tilde{s}^2$

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

第七章　参数估计

7.3 估计量的评选原则

7.3.1 无偏性

1. 定义 设 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 为参数θ 的估计量，若

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

则称$\hat{\theta}$是θ的无偏估计量.

记$b_n=E\hat{\theta}-\theta$ ——估计量$\hat{\theta}$的偏差

若$b_n\neq 0$ ——$\hat{\theta}$是θ的有偏估计

若$\lim\limits_{n\to\infty}b_n=0$ ——$\hat{\theta}$是θ的渐进无偏估计

例1 证明样本均值$\overline{X}$为总体期望$\mu=E(X)$的无偏估计。

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

一般 $$E(A_K)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=E(X^k)=\alpha_k$$

7.3.1 无偏性

例2 证明样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是总体方差 $D(X)=\sigma^2$ 的无偏估计。

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

$$=\frac{1}{n-1}[nE(X_1^2)-nE(\overline{X}^2)]$$

$$\xrightarrow[E(X^2)=D(X)+(EX)^2]{D(X)=E(X^2)-(EX)^2}=\frac{1}{n-1}[n(\sigma^2+\mu^2)-n(D(\overline{X})+\mu^2)]$$

$$=\frac{1}{n-1}[n\sigma^2+n\mu^2-\sigma^2-n\mu^2)]$$

$$=\sigma^2$$

7.3.1 无偏性

注1 例2说明 $\widetilde{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 不是σ^2 的无偏估计。

$$\because\quad E(\widetilde{S}^2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2\xrightarrow[n\to\infty]{}\sigma^2$$

~ 渐近无偏估计

注2 $\hat{\theta}$为θ的无偏估计，但$u(\hat{\theta})$不一定是$u(\theta)$的无偏估计。

例如：若$D(X)>0$，则

$$E[(\hat{\mu})^2]=E(\overline{X}^2)=D(\overline{X})+[E(\overline{X})]^2>[E(\overline{X})]^2=\mu^2$$

注3 无偏估计不唯一，如X_1和 $\overline{X}$ 均为$\mu=E(X)$的无偏估计。

事实上对任何 $c_1, c_2, \cdots, c_n$，当$c_1+c_2+\cdots+c_n=1$ 时，

$$E(\sum_{i=1}^{n}c_iX_i)=\mu$$

7.3.2 有效性

1. 定义 设$\hat{\theta}_1$，$\hat{\theta}_2$都是参数θ的无偏估计量，若

$$D(\hat{\theta}_1) < D(\hat{\theta}_2)$$

则称$\hat{\theta}_1$比$\hat{\theta}_2$ 有效.

比如 $E(X_1) = E(\overline{X}) = \mu$，但 $D(X_1) = \sigma^2$，$D(\overline{X}) = \dfrac{1}{n}\sigma^2$

故 $\hat{\mu} = \overline{X}$ 比 $\hat{\mu}_1 = X_1$ 有效。

设$\hat{\theta}_0$是参数θ的无偏估计量，若对θ的任一无偏估计$\hat{\theta}$，有

$$D(\hat{\theta}_0) \le D(\hat{\theta})$$

则称 $\hat{\theta}_0$ 是θ 的最小方差无偏估计.

7.3.2 有效性

例3 在总体期望$\mu = E(X)$的线性无偏估计类

$$\overline{U} = \left\{ \hat{\mu} = \sum_{i=1}^{n} c_i X_i \;\middle|\; \sum_{i=1}^{n} c_i = 1 \right\}$$ **中求μ 最小方差无偏估计。**

解 $$E(\hat{\mu}) = \sum_{i=1}^{n} c_i E(X_i) = \sum_{i=1}^{n} c_i \mu = \mu, \quad \hat{\mu} \in \overline{U}$$

由Cauchy-Schwarz不等式 $$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

$$D(\hat{\mu}) = \sum_{i=1}^{n} c_i^2 D(X) = \frac{1}{n}\left(\sum_{i=1}^{n} 1^2\right)\left(\sum_{i=1}^{n} c_i^2\right) D(X)$$

$$\ge \frac{1}{n}\left(\sum_{i=1}^{n} 1 \times c_i\right)^2 D(X) = \frac{1}{n} D(X)$$

而 $D(\overline{X}) = \dfrac{1}{n} D(X)$, **故** $\hat{\mu}_0 = \overline{X}$ **是$\boldsymbol{\mu}$ 的最小方差线性无偏估计。**

7.3.2 一致性

1.定义 设$\hat{\theta}$是参数θ的估计量，若对任意$\varepsilon > 0$,有

$$\lim_{n\to\infty} P(|\hat{\theta}-\theta|<\varepsilon)=1 \quad 即\ \hat{\theta}\xrightarrow{P}\theta.$$

则称$\hat{\theta}$为θ的一致估计量.

例　由大数定律$\lim_{n\to\infty} P(|\bar{X}-\mu|<\varepsilon)=1$

知样本均值$\bar{X}$是总体均值μ的一致估计。

7.3.2 一致性

例 证明正态总体的样本方差S^2是σ^2的一致估计。

证 $\because \dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1) \quad E(S^2)=\sigma^2,$

$$DS^2 = D\left(\frac{\sigma^2}{n-1}\cdot\frac{(n-1)}{\sigma^2}S^2\right) = \frac{\sigma^4}{(n-1)^2}D\left(\frac{(n-1)}{\sigma^2}S^2\right) = \frac{2\sigma^4}{(n-1)}.$$

由切比雪夫不等式

$$P(\left|S^2-\sigma^2\right|<\varepsilon) \geq 1-\frac{2\sigma^4}{(n-1)\varepsilon^2} \xrightarrow[n\to\infty]{} 1.$$