第 9 章　参数估计

设总体 X 的分布函数为 $F(X,\theta)$, 其中 θ 为一个未知的参数、或未知的向量. 现从总体中抽取一个样本 $X_1,X_2,\cdots,X_n$, 如何依据该样本来估计 θ, 或 θ 的函数 $g(\theta)$, 此类问题称为参数估计问题. 主要内容包括点估计, 估计量标准和区间估计.

9.1 点估计

9.1.1 矩估计法

设总体 X 的 k 阶矩 $a_k=E[X^k]$, 以及一个样本 $X_1,X_2,\cdots,X_n$ 的样本 k 阶矩 $A_k=\sum_{i=1}^n X_i^k/n$. 利用样本矩去估计总体矩, 进而求解参数 θ 的方法称为 **矩估计法**.

矩估计法的理论基础是大数定理, 设独立同分布的随机变量 $X_1,X_2,\cdots,X_n$, 若期望 $a_k=E[X^k]$ 存在, 则有 $A_k\xrightarrow{P}a_k$, 即

$$\frac{1}{n}\sum_{i=1}^n X_i^k\xrightarrow{P}E[X^k].$$

矩估计法还包括利用样本的中心矩估计总体的中心矩, 设样本 $X_1,X_2,\cdots,X_n$ 的样本 k 阶中心矩 $B_k=\sum_{i=1}^n(X_i-\bar{X})^k/n$, 以及总体 X 的 k 阶中心矩 $b_k=E[(X-E(X))^k]$, 利用样本的中心矩 B_k 去估计总体的中心矩 b_k.

设总体 X 的分布函数 F 包含 m 个未知参数 $\theta_1,\theta_2,\cdots,\theta_m$, 矩估计方法的一般步骤包括:

- 求总体 X 的 k 阶矩: $a_k=E[X^k]=a_k(\theta_1,\theta_2,\cdots,\theta_m),\quad k=1,2,\cdots,m$;
- 计算样本的 k 阶矩: $A_k=\frac{1}{n}\sum_{i=1}^n X_i^k$;
- 令样本矩等于总体矩 $A_k=a_k=a_k(\theta_1,\theta_2,\cdots,\theta_m)$ 得到 m 个方程;
- 求解方程组得到估计量 $\widehat{\theta}_1,\widehat{\theta}_2,\cdots,\widehat{\theta}_m$.

下面来看两个例子:

例 9.1　设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 以及总体 X 的概率密度函数

$$f(x)=\begin{cases}(\alpha+1)x^\alpha & x\in(0,1)\\ 0 & \text{其它},\end{cases}$$

求参数 α 的矩估计.

解 首先计算总体 X 的期望

$$E[X]=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^1 x(\alpha+1)x^{\alpha+1}dx=\frac{\alpha+1}{\alpha+2}.$$

样本 X 的均值 $\bar{X}=\sum_{i=1}^n X_i/n$, 根据样本矩等于总体矩有

$$E(X)=\frac{\alpha+1}{\alpha+2}=\bar{X},$$

求解可得 $\alpha=(2\bar{X}-1)/(1-\bar{X})$.

例 9.2 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 以及总体 X 的密度函数为

$$f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x-\mu}{\theta}} & x\geqslant\mu\\ 0 & \text{其它},\end{cases}$$

其中 $\theta>0$, 求 μ 和 θ 的矩估计.

解 设随机变量 $Y=X-\mu$, 则 Y 服从参数为 $1/\theta$ 的指数分布, 有

$$E(Y)=\theta \quad \text{和} \quad \mathrm{Var}(Y)=\theta^2.$$

由此可得 $E(X)=\mu+\theta$ 和 $\mathrm{Var}(X)=\theta^2$. 计算对应的样本矩

$$A_1=\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i,\qquad B_2=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2.$$

求解方程组

$$\mu+\theta=A_1 \quad \text{和} \quad \theta^2=B_2,$$

解得 $\mu=\bar{X}-\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2/n}$ 和 $\theta=\sqrt{\sum_{i=1}^n(X_i-\bar{X})^2/n}$.

9.1.2 最大似然估计法

若总体 X 为离散型随机变量, 其分布列为 $P(X=x)=P(X=x;\theta)$, 则样本 $X_1,X_2,\cdots,X_n$ 的分布列为

$$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^n P(x_i;\theta).$$

这里 $L(\theta)$ 表示样本 $X_1=x_1,X_2=x_2,\cdots,X_n=x_n$ 发生的概率.

若总体 X 为连续型随机变量, 其概率密度为 $f(x;\theta)$, 则 $X_1,X_2,\cdots,X_n$ 的联合概率密度为

$$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^n f(x_i;\theta).$$

根据概率密度定义可知 $L(\theta)$ 越大, 样本 $(X_1, X_2, \cdots, X_n)$ 落入 $(x_1, x_2, \cdots, x_n)$ 的邻域内概率越大.

综合离散和连续两种情况, 统称 $L(\theta)$ 为样本 $X_1, X_2, \cdots, X_n$ 的似然函数, 很显然 $L(\theta)$ 是 θ 的函数, 若

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

则称 $\hat{\theta}$ 为 θ 的最大似然估计量. 最大似然估计的本质是使观测值 $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ 出现的概率最大. 求解最大似然估计量的步骤如下:

- 计算对数似然函数 $\log(L(x_1, x_2, \cdots, x_m; \theta)$;
- 求对数似然函数中关于参数 θ 的一阶导数或偏导, 令其等于零;
- 求解方程组得到最大似然估计量 $\hat{\theta}$.

例 9.3 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim \mathrm{Ber}(p)$ 的一个样本, 求参数 p 的最大似然估计.

解 首先计算似然函数

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i},$$

从而得到对数似然函数

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

求一阶偏导并令其为零可得

$$\frac{\partial \ln L(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right) = 0.$$

求解得到参数 p 的最大似然估计 $\hat{p} = \sum_{i=1}^{n} X_i / n = \bar{X}$.

例 9.4 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的一个样本, 其中总体 X 的密度函数为

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha} & x \in (0,1) \\ 0 & \text{其它} \end{cases}$$

求 α 的最大似然估计.

解 根据密度函数得到对数似然函数

$$\ln L(\alpha) = \ln\left((\alpha+1)^n \prod_{i=1}^{n} X_i^{\alpha}\right) = n\ln(\alpha+1) + \alpha \ln(X_1 X_2 \cdots X_n).$$

求一阶导数并令其等于零有

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = \frac{n}{\alpha+1} + \ln(X_1X_2\cdots X_n) = 0,$$

最后求解得

$$\alpha = \frac{-n}{\sum_{i=1}^n \ln(X_i)} - 1 = \frac{-1}{\frac{1}{n}\sum_{i=1}^n \ln(X_i)} - 1.$$

对上例, 矩估计值为 $\alpha = (2\bar{X}-1)/(1-\bar{X})$, 因此矩估计值与最大似然估计值可能不同.

下面研究最大似然估计的 **不变性**.

定理 9.1 设 $\mu = \mu(\theta)$ 为 θ 的函数, 且存在反函数 $\theta = \theta(\mu)$. 若 $\hat{\theta}$ 是 θ 的最大似然估计, 则 $\hat{\mu} = \mu(\hat{\theta})$ 是 μ 的最大似然估计.

下面以正态分布为例来看最大似然估计的不变性.

例 9.5 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的一个样本, 求 μ 和 $\sigma > 0$ 的最大似然估计.

解 根据正态分布 $\mathcal{N}(\mu, \sigma^2)$ 的概率密度可知样本 $X_1, X_2, \cdots, X_n$ 的对数似然函数为

$$\ln(L(\mu,\sigma)) = \ln\left(\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left(-\sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2}\right)\right) = -\frac{n\ln 2\pi}{2} - n\ln\sigma - \sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2}.$$

对参数 μ 求偏导并令其等于零可得

$$\frac{\partial \ln L(\mu,\sigma)}{\partial \mu} = \sum_{i=1}^n (X_i - \mu) = 0 \quad \Longrightarrow \quad \mu = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}.$$

对 σ 求偏导并令其等于零可得

$$\frac{\partial \ln L(\mu,\sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (X_i-\mu)^2 = 0 \quad \Longrightarrow \quad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2}.$$

根据最大似然估计的不变性可知方差 σ^2 的最大似然估计为 $\sigma^2 = \sum_{i=1}^n (X_i-\bar{X})^2/n$.

下面验证正态分布中方差估计的最大似然估计不变性: 设 $X_1, X_2, \cdots, X_n$ 为总体 $X \sim \mathcal{N}(\mu, \nu)$ 的样本, 求 μ 和 ν 的最大似然估计. 根据题意可知样本 $X_1, X_2, \cdots, X_n$ 的对数似然函数为

$$\ln L(\mu,\nu) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\nu - \sum_{i=1}^n \frac{(X_i-\mu)^2}{2\nu}.$$

同理得到最大似然估计 $\mu=\bar{X}$, 对 ν 求偏导并令其等于零可得

$$\frac{\partial \ln L(\mu,\nu)}{\partial \nu}=-\frac{n}{2\nu}+\frac{1}{2\nu^2}\sum_{i=1}^{n}(X_i-\mu)^2=0 \quad\Longrightarrow\quad \nu=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,$$

从而完成验证.

例 9.6 设 $X_1,X_2,\cdots,X_n$ 是总体 $X\sim\mathcal{U}(a,b)$ 的一个样本, 求 a 和 b 的最大似然估计.

解 当 $x\in[a,b]$ 时, 总体 X 的概率密度为 $f(x)=1/(b-a)$, 其它情况为零, 因此似然函数为

$$L(a,b)=\begin{cases}\frac{1}{(b-a)^n} & a\leqslant X_1,X_2,\cdots,X_n\leqslant b\\ 0 & \text{其它}\end{cases}$$

直接求偏导无法解出 a 和 b, 此时可以从最大似然的定义出发, 应使得 b 尽可能小且 a 尽可能大, 但需满足 $a\leqslant X_1,X_2,\cdots,X_n\leqslant b$, 因此最大似然估计量为:

$$b=\max\{X_1,X_2,\cdots,X_n\} \qquad \text{和} \qquad a=\min\{X_1,X_2,\cdots,X_n\}.$$

例 9.7 设 $X_1,X_2,\cdots,X_n$ 是总体 X 的样本, 以及总体 X 的概率密度为

$$f(x)=\begin{cases}\theta e^{-(x-\mu)\theta} & x\geqslant\mu\\ 0 & \text{其它,}\end{cases}$$

求 μ 和 θ 的最大似然估计.

解 首先计算似然函数为

$$L(\theta,\mu)=\begin{cases}\theta^n e^{-\theta\sum_{i=1}^{n}(X_i-\mu)} & X_i\geqslant\mu\\ 0 & \text{其它}\end{cases}$$

进一步得到对数似然函数为

$$\ln L(\theta,\mu)=n\ln\theta-\theta\sum_{i=1}^{n}(X_i-\mu).$$

求偏导、并令偏导等于零有

$$\frac{\partial \ln L(\theta,\mu)}{\partial\theta}=\frac{n}{\theta}-\sum_{i=1}^{n}(X_i-\mu)=0 \quad\Rightarrow\quad \theta=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)},$$

另一方面有

$$\frac{\partial \ln L(\theta,\mu)}{\partial \mu} = n\theta = 0 \quad \Rightarrow \quad \theta = 0,$$

此时无法求解 θ 和 μ 的最大似然估计. 回到似然函数的定义

$$L(\theta,\mu) = \begin{cases} \theta^n e^{-\theta\sum_{i=1}^n (X_i-\mu)} & X_1, X_2, \cdots, X_n \geqslant \mu \\ 0 & \text{其它} \end{cases}$$

可以发现 μ 越大似然函数 $L(\theta,\mu)$ 越大, 但须满足 $X_i \geqslant \mu$ $(i \in [n])$. 由此可得最大似然估计

$$\hat{\mu} = \min\{X_1, X_2, \cdots, X_n\},$$

进一步求解可得

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})}.$$

最大似然估计的思想, 最早始于高斯的误差理论, 到 1912 年由费尔希 (R. A. Fisher) 作为一般的方法正式提出. 很多统计学家对最大似然估计进行了研究, 由于考虑了总体的分布信息, 因而在通常情况下会比矩估计法更好, 但在个别情况下也会出现不理想的结果, 尤其当总体分布信息未知时, 最大似然估计无法使用.

9.2 估计量的评价标准

前面介绍了多种估计方法, 不同的估计方法可能得到不同的估计值, 有必要回答一个问题: 采用哪一种估计量更好, 或更好的标准是什么. 衡量估计量好坏的常用标准: 无偏性, 有效性, 一致性.

9.2.1 无偏性

定义 9.1 设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 以及 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是 θ 的一个估计量, 若

$$E_{X_1,X_2,\ldots,X_n}\left[\hat{\theta}\right] = E_{X_1,X_2,\ldots,X_n}\left[\hat{\theta}(X_1, X_2, \cdots, X_n)\right] = \theta,$$

则称 $\hat{\theta}$ 为 θ 的无偏估计.

无偏性不要求在任意情况下估计值 $\hat{\theta}$ 都等于 θ, 但在平均期望的情形下有 $E(\hat{\theta}) = \theta$ 成立. 其意义在于无系统性偏差, 无偏性是在有限样本情形下一种常见的估计量标准.

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本, 设总体期望 $\mu = E[X]$, 方差 $\sigma^2 = \mathrm{Var}(X)$ 以及 k 阶矩 $a_k = E[X^k]$ 存在, 则有

- 样本原点矩 $A_k = \sum_{i=1}^n X_i^k/n$ 是 a_k 的无偏估计;
- 样本方差 $S_0^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ 是 σ^2 的有偏估计;