9 参数估计

设总体 X 的分布函数为 $F(X,\theta)$, 其中 θ 为未知参数(也可为向量). 现从总体中抽取一样本 $X_1, X_2, \cdots, X_n$, 如何依据样本估计参数 θ, 或 θ 的函数 $g(\theta)$, 此类问题称为参数估计问题. 内容包括: 点估计, 估计量标准, 区间估计.

9.1 点估计

9.1.1 矩估计法

总体 X 的 k 阶矩: $a_k = E[X^k]$ 样本 k 阶矩: $A_k = \frac{1}{n}\sum_{i=1}^n X_i^k$

用相应的样本矩去估计总体矩, 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: $X_1, X_2, \cdots, X_n$ 为 i.i.d. 的随机变量, 若 $E(X) = \mu$, 则当 $n \to \infty$ 时有

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

推论: 若 $E[X^k] = a_k$ 存在, 则当 $n \to \infty$ 时有

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

还可利用中心距进行估计:
总体 X 的 k 阶中心矩: $b_k = E[(X - E(X))^k]$ 样本 k 阶中心矩: $B_k = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^k$

矩估计方法: 总体 X 的分布函数 F 包含 m 个未知参数 $\theta_1, \theta_2, \cdots, \theta_m$,

1) 求总体 X 的 k 阶矩: $a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m) = E[X^k],\ k \in [m]$ (a_k 一般为 $\theta_1, \theta_2, \cdots, \theta_m$ 的函数).
2) 计算样本的 k 阶矩: $A_k = \frac{1}{n}\sum_{i=1}^n X_i^k$.
3) 令样本矩等于总体矩 $A^k = a^k = a^k(\theta_1, \theta_2, \cdots, \theta_m)$ $(k = 1, 2, \cdots, m)$, 得到 m 个关于 $\theta_1, \theta_2, \cdots, \theta_m$ 的方程组.
4) 求解方程组得到估计量 $\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_n$.

例9.1. 设总体 X 的概率密度函数

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

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 求参数 α 的矩估计.

解. 首先计算总体 X 的期望

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

样本 X 的均值 $\bar{X}=\sum_{i=1}^n X_i/n$. 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

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

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

课堂习题:

- 求正态总体 $\mathcal{N}(\mu,\sigma^2)$ 的 μ,σ^2 的矩估计法.
- 求总体 $X\sim\mathcal{U}(a,b)$ 中 a,b 的矩估计法.

9.1.2 最大似然估计法

设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的一个样本. 若总体 X 为离散型随机变量, 其分布列为 $\Pr(X=x)=\Pr(X=x;\theta)$, 则样本 $X_1,X_2,\cdots,X_n$ 的分布列为

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

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 $X_1=x_1,X_2=x_2,\cdots,X_n=x_n$ 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.

综合上述离散和连续两种随机变量, 统称 $L(\theta)$ 为样本 $X_1, X_2, \cdots, X_n$ 的似然函数, 可以发现 $L(\theta)$ 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. 直觉而言: 最大似然估计量 $\hat{\theta}$ 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:

i) 计算对数似然函数 $\log(L(x_1, x_2, \cdots, x_m; \theta)$;

ii) 求对数似然函数中参数 θ 的一阶偏导, 令其等于零;

iii) 求解方程组得到最大似然估计量 $\hat{\theta}$.

例9.3. 设 $X_1, X_2, \cdots, X_n$ 是取自总体 $X \sim B(1, p)$ 的样本, 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

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

由此求解 $p = \sum_{i=1}^{n} X_i / n = \bar{X}$. [验证矩估计法] □

下面讨论 **最大似然估计不可变性**

性质9.1. 设 $\mu(\theta)$ 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.

例9.4. 设 $X_1, X_2, \cdots, X_n$ 为总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 求 μ 和 $\sigma > 0$ 的最大似然估计.

解. 根据高斯分布知 X 的概率密度为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. 样本 $X_1, X_2, \cdots, X_n$ 的似然函数为

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

其对数似然函数为 $\ln L(\mu, \sigma) = -n \ln(2\pi)^{1/2} - n \ln \sigma - \sum_{i=1}^{n} (X_i - \mu)^2 / 2\sigma^2$. 对参数 μ 求导计算可得

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

对 σ 求导计算可得

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

□

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 $\sigma^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2/n$. 下面进行验证最大似然估计的不变性: 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

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

对参数 μ 求偏导计算其最大似然估计 $\mu = \sum_{i=1}^{n} X_i/n = \bar{X}$, 对 ν 求偏导计算可得

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

从而完成验证.

例9.5. 设总体 X 的密度函数为

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

设 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本, 求 α 的最大似然估计.

解. 首先得到似然函数为

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

以及其对数似然函数 $\ln L(\alpha) = n\ln(\alpha+1) + \alpha\ln(X_1 X_2 \cdots X_n)$. 求导并令偏导为零有

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = \frac{n}{\alpha+1} + \ln(X_1 X_2 \cdots X_n) = 0,$$

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.

例9.6. 设 $X_1, X_2, \cdots, X_n$ 是总体 $X \sim \mathcal{U}(a,b)$ 的样本, 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

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

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 $a \leq X_1, X_2, \cdots, X_n \leq b$, 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

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

求 μ 和 θ 的最大似然估计.

解. 首先计算似然函数为

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

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

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

可以发现 μ 越大似然函数 $L(\theta, \mu)$ 越大, 但须满足 $X_i \geq \mu$ $(i \in [n])$. 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

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

□

9.2 估计量的评价标准

前一节已经讲过不同的点估计方法, 不同的估计方法可能得到不同的估计值, 自然涉及到一个问题: 采用哪一种估计量更好, 或更好的标准是什么呢? 估计量的常用标准: 无偏性, 有效性, 一致性.