

10.1.2 最大似然估计法

设 X_1, X_2, \dots, 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, \dots, X_n 的分布列为

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

这里 $L(\theta)$ 表示样本 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 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 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 的联合概率密度为

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

根据概率密度定义可知 $L(\theta)$ 越大, 样本 (X_1, X_2, \dots, X_n) 落入 (x_1, x_2, \dots, 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, \dots, X_n 的似然函数, 可以发现 $L(\theta)$ 是 θ 的函数, 若

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

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 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 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 $\log(L(x_1, x_2, \dots, x_n; \theta))$;
- 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 10.3 设 X_1, X_2, \dots, 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 10.1 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 10.4 设 X_1, X_2, \dots, 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, \dots, 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

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

对 σ 求导计算可得

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

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, \dots, 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, \dots, 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

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

从而完成验证.

例 10.5 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, \dots, 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 10.6 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

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

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:

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

例 10.7 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

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

另一方面有

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

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

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

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

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

进一步求解可得

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

10.2 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.

10.2.1 无偏性

定义 10.1 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, 令 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 是 θ 的一个估计量, 若

$$E_{X_1, X_2, \dots, X_n} [\hat{\theta}] = E_{X_1, X_2, \dots, X_n} [\hat{\theta}(X_1, X_2, \dots, X_n)] = \theta$$

则称 $\hat{\theta}$ 为 θ 的无偏估计.

无偏估计不要求估计值 $\hat{\theta}$ 在任意情况下都等于 θ , 但在期望的情形下有 $E(\hat{\theta}) = \theta$ 成立. 其意义在于无系统性偏差, 无偏性是一种对估计量常见而且重要的标准.

首先看看如下例子:

例 10.8 (样本 k 阶原点矩为总体 k 阶原点矩的无偏估计) 设 X_1, X_2, \dots, X_n 是总体 X 的样本, 若 $E[X^k]$ 存在, 则 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ 是总体 $a_k = E[X^k]$ 的无偏估计.

例 10.9 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, 其期望为 μ , 方差为 σ^2 , 则: 1) $S_0^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n$ 是 σ^2 的有偏估计; 2) $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ 是 σ^2 的无偏估计.

注意 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 是 θ 的无偏估计, 但并不一定有 $g(\hat{\theta})$ 是 $g(\theta)$ 的无偏估计, 这是因为 $E[\hat{\theta}] = \theta$ 并不能推导出 $E[g(\hat{\theta})] = g(\theta)$. 例如

$$E[\bar{X}] = E[X] = \mu \quad \text{但} \quad E[(\bar{X})^2] \neq \mu^2.$$

例 10.10 设 X_1, X_2, \dots, X_n 是总体 X 的样本, 以及总体 X 的概率密度为

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

证明: $\bar{X} = \sum_{i=1}^n X_i/n$ 和 $n \min\{X_1, X_2, \dots, X_n\}$ 均是 θ 的无偏估计.

证明 根据期望和指数分布的性质有

$$E[\bar{X}] = E[X] = \theta,$$

由此可知 \bar{X} 是 $E[X]$ 的无偏估计. 设随机变量 $Z = \min\{X_1, X_2, \dots, X_n\}$, 则有

$$\begin{aligned} F_Z(z) &= \Pr[Z \leq z] = 1 - \Pr[Z > z] \\ &= 1 - \Pr[X_1 > z] \Pr[X_2 > z] \cdots \Pr[X_n > z] \\ &= 1 - \prod_{i=1}^n (1 - \Pr[X_i \leq z]) = \begin{cases} 0 & z < 0 \\ 1 - e^{-nz/\theta} & z \geq 0. \end{cases} \end{aligned}$$

于是当 $z \geq 0$ 时有

$$\Pr[Z > z] = 1 - F_Z(z) = e^{-nz/\theta}.$$

根据期望的性质有

$$E[Z] = \int_0^{+\infty} \Pr[Z > z] dz = \int_0^{+\infty} e^{-nz/\theta} dz = \frac{\theta}{n}.$$

于是有 $\theta = E[nZ]$ 成立.

10.2.1.1 有效性

参数可能存在多个无偏估计, 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是 θ 的无偏估计, 则可以比较方差

$$\text{Var}(\hat{\theta}_1) = E[(\hat{\theta}_1 - \theta)^2] \quad \text{和} \quad \text{Var}(\hat{\theta}_2) = E[(\hat{\theta}_2 - \theta)^2].$$

一般而言: 方差越小, 无偏估计越好.

定义 10.2 设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \dots, X_n)$ 是 θ 的两个无偏估计, 若

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2),$$

则称 θ_1 比 θ_2 有效.

例 10.11 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, 且 X 的概率密度为

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

令 $Z = \min\{X_1, X_2, \dots, X_n\}$, 证明: 当 $n > 1$ 时 $\bar{X} = \sum_{i=1}^n X_i/n$ 比 nZ 有效.

证明 根据独立性有

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\theta^2}{n}.$$

根据例 10.10 可知随机变量 Z 的概率密度为

$$f(z) = \begin{cases} 0 & z < 0 \\ \frac{n}{\theta} e^{-\frac{nz}{\theta}} & z \geq 0 \end{cases}$$

从而得到

$$\text{Var}(nZ) = n^2 \text{Var}(Z) = n^2 \frac{\theta^2}{n^2} = \theta^2,$$

因此当 $n \geq 1$ 时有 $\text{Var}(\bar{X}) \leq \text{Var}(nZ)$ 成立, 故 \bar{X} 比 nZ 有效.

例 10.12 设 X_1, X_2, \dots, X_n 是总体 X 的样本, 且 $E(X) = \mu$ 和 $\text{Var}(X) = \sigma^2$. 设常数 $c_1, c_2, \dots, c_n \geq 0$ 满足 $\sum_{i=1}^n c_i = 1, c_i \neq 1/n$, 求证: \bar{X} 比 $\sum_{i=1}^n c_i X_i$ 有效.

证明 根据样本的独立同分布条件有

$$E[\bar{X}] = \mu \quad \text{和} \quad \text{Var}(\bar{X}) = \sigma^2/n.$$

根据期望的性质有 $E[\sum_{i=1}^n c_i X_i] = \mu$, 进一步有

$$\text{Var}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 \text{Var}(X_i) = \sigma^2 \sum_{i=1}^n c_i^2 \geq \frac{\sigma^2}{n}$$

这里利用不等式 $\sum_{i=1}^n c_i^2/n \geq (\sum_{i=1}^n c_i/n)^2 = 1/n^2$, 所以有 $\text{Var}(\sum_{i=1}^n c_i X_i) \geq \text{Var}(\bar{X})$.

下面定义有效统计量:

定理 10.1 (Rao-Crammer 不等式) 设随机变量 X 的概率密度为 $f(x; \theta)$ 或分布函数为 $F(x; \theta)$, 令

$$\text{Var}_0(\theta) = \frac{1}{nE\left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta}\right)^2\right]} \quad \text{或} \quad \text{Var}_0(\theta) = \frac{1}{nE\left[\left(\frac{\partial \ln F(X; \theta)}{\partial \theta}\right)^2\right]},$$