

Ch10: 参数估计

# Parameter Estimation

December 21, 2023

# 引言

假设我们已知南京大学男性学生的身高服从正态分布  $\mathcal{N}(\mu, \sigma^2)$ , 但不知道参数  $\mu$  和  $\sigma$  具体的取值, 这时候我们可以利用抽样得到样本均值来推断总体的均值  $\mu$ .

这类已知总体分布形式, 但不知其具体参数, 用样本统计量来估计总体的参数的问题称为参数估计问题. 参数估计是统计推断的核心问题之一, 方法大体上有两类: 点估计与区间估计.

# 点估计

**定义 0.84** 设  $X_1, X_2, \dots, X_n$  是来自总体的一个样本, 用于估计未知参数  $\theta$  的统计量  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  称为  $\theta$  的估计量, 或称为  $\theta$  的 **点估计**.

Remarks: 点估计的本质就是用样本统计量直接作为总体参数的估计值

- 这里的参数是总体的属性, 而统计量是针对样本的计算.
- 在这里如何构造  $\hat{\theta}$  没有明确的规定, 1900 年 K. 皮尔逊提出了一个替换原理, 给出了构造  $\hat{\theta}$  的一种方法, 后来人们称此法为 **矩估计法**.
- 而 1922 年费希尔提出的最大似然法给出了另外一种构造  $\hat{\theta}$  的方法, 称为 **极大似然估计**.

# 矩估计法

替换原理具体为:

- 用样本矩去替换总体矩 (这里的矩可以是原点矩也可以是中心距).
  - 使用原点矩
    - 总体  $X$  的  $k$  阶原点矩:  $a_k = \mathbb{E}[X^k]$
    - 样本的  $k$  阶原点矩:  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
  - 使用中心矩
    - 总体  $X$  的  $k$  阶中心矩:  $b_k = \mathbb{E}[(X - \mathbb{E}(X))^k]$
    - 样本的  $k$  阶中心矩:  $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$
- 用样本矩的函数去替换相应的总体矩的函数.

## 矩估计法 -- 适用场景

根据这个替换原理, 在总体分布形式未知的情况下也可以对参数做出估计, 譬如:

- 用样本均值  $\bar{X}$  估计总体均值  $\mathbb{E}(X)$
- 用样本方差  $S^2$  估计总体方差  $\text{VAR}(X)$ 
  - 注意: 若没有特殊说明, 样本方差采用无偏方差
- 用事件  $A$  出现的频率估计事件  $A$  发生的概率.

## 矩估计法 -- 计算步骤

总体  $X$  的分布函数  $F$  包含  $m$  个未知参数  $\theta_1, \theta_2, \dots, \theta_m$

- 计算总体  $X$  的  $k$  阶矩:  $a_k = a_k(\theta_1, \theta_2, \dots, \theta_m) = \mathbb{E}[X^k]$ ,  $k \in [m]$ 
  - $a_k$  一般为  $\theta_1, \theta_2, \dots, \theta_m$  的函数
- 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:

$$A_k = a_k = a_k(\theta_1, \theta_2, \dots, \theta_m), \quad k = [m]$$

得到  $m$  个关于  $\theta_1, \theta_2, \dots, \theta_m$  的方程组

- 求解方程组得到估计量  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$

## 矩估计：例 0.130

例 0.130 设总体  $X$  的概率密度函数为

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

设  $X_1, X_2, \dots, X_n$  是来自总体的样本, 求参数  $\alpha$  的矩估计.

## 解答：例 0.130

题目：设总体  $X$  的概率密度函数为

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

设  $X_1, X_2, \dots, X_n$  是来自总体的样本, 求参数  $\alpha$  的矩估计.

解答:

- 首先计算总体  $X$  的 1 阶矩:

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

以及样本的 1 阶矩:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

- 根据矩估计方法有

$$\mathbb{E}[X] = \frac{\alpha + 1}{\alpha + 2} = \bar{X}$$

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

## 矩估计：例 0.131

**例 0.131** 设  $X_1, X_2, \dots, X_n$  是来自总体的样本, 且总体  $X$  的概率密度函数为

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

其中  $\theta > 0$ , 求参数  $\mu$  和  $\theta$  的矩估计.

## 解答：例 0.131

题目：如上所述.

解答：

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

$$\mathbb{E}[Y] = 0 \quad \text{和} \quad \sigma(Y) = \theta^2.$$

由此可得  $\mathbb{E}[X] = \mu + \theta$  和  $\sigma(X) = \theta^2$ .

- 计算对应的样本矩

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

求解方程组

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

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

## 极大似然估计法 —— 例子

为了叙述极大似然估计的直观想法, 先看下面这个例子:

**例 0.132** 设有两个外形相同的箱子中各有 100 只球, 其中甲箱中有 99 个白球、1 个黑球, 乙箱中有 1 个白球、99 个黑球. 今随机抽取一箱并从中抽取一球, 结果取得白球, 问这个白球是从哪个箱子中取出?

## 极大似然估计法 -- 例子

为了叙述极大似然估计的直观想法, 先看下面这个例子:

**例 0.133** 设有两个外形相同的箱子中各有 100 只球, 其中甲箱中有 99 个白球、1 个黑球, 乙箱中有 1 个白球、99 个黑球. 今随机抽取一箱并从中抽取一球, 结果取得白球, 问这个白球是从哪个箱子中取出?

解答:  $A$  表示事件“从甲箱中取出白球”,  $B$  表示事件“从乙箱中取出白球”, 又

$$P(A) = 0.99 > P(B) = 0.01$$

因此, 按照可以推断白球“最可能”是从甲箱中取出的.

这个推断符合人们的经验事实, 这里的“最可能”就是“极大似然”之意, 这种想法常称为“极大似然原理”. 即, 已经得到了样本, 然后通过样本倒推, 找到能够使的该样本出现的最大概率的条件.

# 极大似然估计法

**定义 0.85** 设总体的概率函数为  $p(X; \theta)$ ,  $\theta \in \Theta$ , 其中  $\theta$  是一个未知参数或几个未知参数组成的参数向量,  $\Theta$  是参数空间.  $X_1, X_2, \dots, X_n$  是来自总体的样本, 将样本的联合概率函数看成  $\theta$  的函数, 用  $L(\theta, X_1, X_2, \dots, X_n)$  表示, 简记  $L(\theta)$ ,

$$L(\theta) = L(\theta, X_1, X_2, \dots, X_n) = p(X_1; \theta)p(X_2; \theta) \dots p(X_n; \theta),$$

$L(\theta)$  称为样本的似然函数. 若某个统计量  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  满足,

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta),$$

则称  $\hat{\theta}$  是  $\theta$  的极大似然估计, 简记为 MLE (MaXimum Likelihood Estimation).

## 极大似然估计法 -- 计算步骤

求  $L(\theta) = p(X_1; \theta)p(X_2; \theta) \dots p(X_n; \theta)$  的最大值可以通过下列步骤:

- 列出  $L(\theta) = p(X_1; \theta)p(X_2; \theta) \dots p(X_n; \theta)$
- 对等式两边取对数, 求关于  $\theta$  求一阶偏导令其为零
- 求解方程组得到极大似然估计  $\hat{\theta}$ .

## 极大似然估计法 -- 计算步骤

求  $L(\theta) = p(X_1; \theta)p(X_2; \theta) \dots p(X_n; \theta)$  的最大值可以通过下列步骤:

- 列出  $L(\theta) = p(X_1; \theta)p(X_2; \theta) \dots p(X_n; \theta)$
- 对等式两边取对数, 求关于  $\theta$  求一阶偏导令其为零
- 求解方程组得到极大似然估计  $\hat{\theta}$ .

如何构造似然函数?

- 核心: 条件概率公式

## 极大似然估计：例 0.134

例 0.134 设  $X_1, X_2, \dots, X_n$  是取自总体  $X \sim B(1, p)$  的样本, 求参数  $p$  的极大似然估计.

## 解答：例 0.134

题目：设  $X_1, X_2, \dots, X_n$  是取自总体  $X \sim B(1, p)$  的样本，求参数  $p$  的极大似然估计。

解答：

- 首先计算似然函数

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

由而可得对数似然函数

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

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  $p = \sum_{i=1}^n X_i / n = \bar{X}$ .

## 极大似然估计：例 0.135

例 0.135 设  $X_1, X_2, \dots, X_n$  是取自总体  $X \sim U(0, \theta)$  的样本, 求参数  $\theta$  的极大似然估计.

## 解答：例 0.135

题目：设  $X_1, X_2, \dots, X_n$  是取自总体  $X \sim U(0, \theta)$  的样本，求参数  $\theta$  的极大似然估计。

解答：

- 首先计算似然函数

$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}_{\{0 < X_i \leq \theta\}} = \frac{1}{\theta^n} \mathbb{I}_{\{0 < X_i \leq \theta\}},$$

要使  $L(\theta)$  最大，首先是示性函数取值应该为 1，其次是  $1/\theta^n$  尽可能大，由于  $1/\theta^n$  是  $\theta$  的单调递减函数，所以  $\theta$  的取值应尽可能小，但示性函数为 1 决定了  $\theta$  不能小于  $X_{(n)}$ ，由此给出  $\theta$  的极大似然估计为  $X_{(n)}$ 。

- 这个例子说明虽然求导函数是求极大似然估计最常用的方法，但并不是所有场合求导都是有效的。

## 极大似然估计 —— 不可变性

极大似然估计有一个简单而有效的性质：

**定理 0.73** 如果  $\hat{\theta}$  是参数  $\theta$  的极大似然估计, 那么对于任一的函数  $g(\cdot)$ ,  $g(\hat{\theta})$  也是  $g(\theta)$  的极大似然估计.

该性质称为极大似然估计的不变性, 从而使得一些复杂结构的参数的极大似然估计的计算变得容易了.

## 极大似然估计：例 0.136

**例 0.136** 设  $X_1, X_2, \dots, X_n$  是取自总体  $X \sim \mathcal{N}(\mu, \sigma^2)$  的样本, 求参数  $\mu$  和  $\sigma > 0$  的极大似然估计.

## 解答：例 0.136

解答：

- 根据正态分布的密度函数，可知似然函数

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

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  $\mu$  求导计算可得

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

对参数  $\sigma$  求导计算，可得

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

根据极大似然估计的不变性，可知方差  $\sigma^2$  的极大似然估计为

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

## 极大似然估计：例 0.137

例 0.137 设  $X_1, X_2, \dots, X_n$  是来自总体的样本, 且总体  $X$  的概率密度函数为

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

求参数  $\mu$  和  $\theta$  的极大似然估计.

## 解答：例 0.137

题目：设  $X_1, X_2, \dots, X_n$  是来自总体的样本，且总体  $X$  的概率密度函数为

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

求参数  $\mu$  和  $\theta$  的极大似然估计.

解答：

- 列出似然函数

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

其对数似然函数为

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

- 对参数  $\theta$  求导计算可得

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

- 对参数  $\mu$  求导计算可得

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

此时无法求解  $\mu$  和  $\theta$  的极大似然估计。

- 回顾似然函数

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

可以发现  $\mu$  越大似然函数  $L(\theta, \mu)$  越大, 但须满足  $X_i \geq \mu (i \in [n])$ . 由此可得极大似然估计为

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

## 极大似然估计：例 0.138

例 0.138 设总体  $X$  的概率密度函数为

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

设  $X_1, X_2, \dots, X_n$  是来自总体的样本, 求参数  $\alpha$  的矩估计.

## 解答：例 0.138

题目：如上所述.

解答：

- 列出似然函数

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

以及其对数似然函数为  $\ln L(\alpha) = n \ln(\alpha + 1) + \alpha \ln(X_1 X_2 \dots X_n)$ . 求导并令导数为零有

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

求解可得

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

对比例 0.130, 可以看到同一密度函数的矩估计和极大似然估计结果可能不同.

## 估计量的评价标准

不同的估计方法可能得到不同的估计值.

自然地, 我们希望知道采用哪一种估计量更好, 或更好的标准是什么呢? 统计学上, 给出了无偏性、有效性、一致性等评价标准.

- 无偏性:  $\hat{\theta}$  与参数真值  $\theta$  之间的偏差的平均值为 0
- 有效性:  $\hat{\theta}$  围绕参数真值  $\theta$  的方差越小越好
- 一致性: 随着样本量的不断增大,  $\hat{\theta}$  能够有效逼近参数真值  $\theta$

## 无偏性

**定义 0.86** 设  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  是  $\theta$  的一个估计,  $\theta$  的参数空间为  $\Theta$ , 若对任意的  $\theta \in \Theta$ , 有

$$\mathbb{E}_{\theta}(\hat{\theta}) = \theta,$$

则称  $\hat{\theta}$  是  $\theta$  的 **无偏估计**, 否则称为 **有偏估计**.

### Remarks:

- (原点矩) 样本  $k$  阶原点矩为总体  $k$  阶原点矩的无偏估计
- (中心矩) 设  $X_1, X_2, \dots, X_n$  来自总体  $X$  的样本, 期望  $\mu$ , 方差  $\sigma^2$ , 则:
  - $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  是  $\sigma^2$  的有偏估计
  - $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  是  $\sigma^2$  的无偏估计
- 若  $\hat{\theta}$  是  $\theta$  的一个无偏估计,  $g(\hat{\theta})$  不一定也是  $g(\theta)$  的无偏估计.

## 无偏估计：例 0.139

例 0.139 设  $X_1, X_2, \dots, X_n$  来自总体  $X$  的样本, 且总体  $X$  的概率密度函数为

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

证明: 统计量

$$\bar{X} = \sum_{i=1}^n X_i/n \quad \text{和} \quad n \cdot \min\{X_1, X_2, \dots, X_n\}$$

均是  $\theta$  的无偏估计.

## 解答：例 0.139

题目：如上所述.

解答：

- 根据期望和指数分布的性质, 有

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

由此可知,  $\bar{X}$  是  $\theta$  的无偏估计 (原点矩).

- 设随机变量  $Z = \min\{X_1, X_2, \dots, X_n\}$ , 则有

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

于是当  $z \geq 0$  时, 有

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

根据期望的性质, 有

$$\mathbb{E}[Z] = \int_0^{+\infty} e^{-nz/\theta} dz = \frac{\theta}{n},$$

于是有  $\theta = \mathbb{E}[nZ]$ .

# 有效性

例子 0.139 说明: 参数可能存在多个无偏估计.

- 若  $\hat{\theta}_1$  和  $\hat{\theta}_2$  都是  $\theta$  的无偏估计, 如何在多个无偏估计中进行选择?
- 直观的想法是,  $\hat{\theta}$  围绕参数真值  $\theta$  的方差越小越好, 即有效性.

**定义 0.87** 设  $\hat{\theta}_1$  和  $\hat{\theta}_2$  分别是  $\theta$  的两个无偏估计, 如果对任意的  $\theta \in \Theta$  都有

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

且至少有一个  $\theta \in \Theta$  使得上述不等式严格成立, 则称  $\hat{\theta}_1$  比  $\hat{\theta}_2$  有效.

Remarks: 有效性是针对无偏估计而言的, 因此判断有效性之前必须先确认估计量的无偏性.

## 有效性：例 0.140

例 0.140 设  $X_1, X_2, \dots, X_n$  来自总体  $X$  的样本, 且总体  $X$  的概率密度函数为

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

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  $nZ$  更有效.

## 解答：例 0.140

题目：如上所述。

解答：

- 根据样本的独立性有

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

又根据例0.139可知随机变量  $Z$  的密度函数为

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

从而得到

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

因此当  $n \geq 1$  时有  $\sigma(\bar{X}) \leq \sigma(nZ)$  成立, 故估计量  $\bar{X}$  比  $nZ$  更有效.

## 有效性：例 0.141

**例 0.141** 设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本, 且  $\mathbb{E}(X) = \mu$  及  $\text{VAR}(X) = \sigma^2$ . 设常数  $c_1, c_2, \dots, c_n \geq 0$ , 满足  $\sum_{i=1}^n c_i = 1, c_i \neq \frac{1}{n}$ . 求证:  $\bar{X}$  比  $\sum_{i=1}^n c_i X_i$  有效.

## 解答：例 0.141

题目：设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本，且  $\mathbb{E}(X) = \mu$  及  $\text{VAR}(X) = \sigma^2$ 。设常数  $c_1, c_2, \dots, c_n \geq 0$ ，满足  $\sum_{i=1}^n c_i = 1, c_i \neq \frac{1}{n}$ 。求证： $\bar{X}$  比  $\sum_{i=1}^n c_i X_i$  有效。

解答：

- 根据样本的独立同分布的性质，有

$$\mathbb{E}(\bar{X}) = \mu \quad \text{和} \quad \text{VAR}(\bar{X}) = \frac{\sigma^2}{n}.$$

根据期望的性质，有  $\mathbb{E}[\sum_{i=1}^n c_i X_i] = \mu$ ，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  $\sum_{i=1}^n c_i^2/n \geq (\sum_{i=1}^n c_i/n)^2$ ，所以有  $\text{VAR}(\sum_{i=1}^n c_i X_i) \geq \text{VAR}(\bar{X})$ 。

## Rao-Crammer 不等式

有效性希望  $\hat{\theta}$  围绕参数真值  $\theta$  的方差越小越好, 那么这个方差能小到什么程度? 有无下界? 若有的话, 如何去求? Rao-Crammer 不等式回答了这些问题.

**定理 0.74** 随机变量  $X$  的概率密度为  $f(X; \theta)$  或分布函数为  $F(X; \theta)$ , 令

$$\text{VAR}_0(\theta) = \frac{1}{n\mathbb{E} \left[ \left( \frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]} \quad \text{或} \quad \text{VAR}_0(\theta) = \frac{1}{n\mathbb{E} \left[ \left( \frac{\partial \ln F(X; \theta)}{\partial \theta} \right)^2 \right]}$$

对任意的无偏估计量  $\hat{\theta}$ , 有

$$\text{VAR}(\hat{\theta}) \geq \text{VAR}_0(\theta),$$

称  $\text{VAR}_0(\theta)$  为估计量  $\hat{\theta}$  方差的下界. 当  $\text{VAR}(\hat{\theta}) = \text{VAR}_0(\theta)$  时, 称  $\hat{\theta}$  为达到方差下界的无偏估计量, 此时  $\hat{\theta}$  为最有效估计量, 简称 **有效估计量**.

## 有效性：例 0.142

例 0.142 设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本, 且总体  $X$  的概率密度函数为

$$f(X) = \begin{cases} \frac{1}{\theta} e^{-\frac{X}{\theta}}, & X > 0 \\ 0, & X \leq 0 \end{cases}$$

证明:  $\theta$  的极大似然估计为有效估计量.

## 解答：例 0.142

题目：如上所述.

解答：

- 根据定理 0.74, 首先计算  $\sigma_0(\theta)$ . 又

$$\ln f(X; \theta) = -\ln \theta - \frac{X}{\theta}, \quad \frac{\partial \ln f(X; \theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{X}{\theta^2}$$

所以

$$\text{VAR}_0(\theta) = \frac{1}{n\mathbb{E}\left[\left(\frac{\partial \ln f(X; \theta)}{\partial \theta}\right)^2\right]} = \frac{1}{n\mathbb{E}\left[\left(-\frac{1}{\theta} + \frac{X}{\theta^2}\right)^2\right]} = \frac{1}{\frac{n}{\theta^4}\mathbb{E}[(X - \mathbb{E}[X])^2]} = \frac{\theta^2}{n}$$

- 计算对数似然函数, 有

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

进一步得到极大似然估计  $\hat{\theta}$  的方差  $\text{VAR}(\hat{\theta}) = \frac{\theta^2}{n}$ , 因此  $\theta$  的极大似然估计为有效估计量.

# 一致性

**定义 0.88** 设  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  是  $\theta$  一个估计量. 当  $n \rightarrow \infty$  时, 有  $\hat{\theta} \xrightarrow{P} \theta$  成立, 即对任意  $\epsilon > 0$ , 有

$$\lim_{n \rightarrow \infty} P \left[ |\hat{\theta} - \theta| > \epsilon \right] = 0,$$

则称  $\hat{\theta}$  为  $\theta$  的 **一致估计量**.

Remarks: 一致性被认为是对估计的一个最基本要求.

- 如果一个估计量, 在样本不断增多时都不能有效的靠近被估参数的真实值, 那么这个估计是很值得怀疑的.
- 通常, 不满足一致性的估计都不予考虑.

# 一致性

在判断或计算参数的一致估测量时, 下述两个定理是很有用的.

**定理 0.75** (一致性的充分条件) 设  $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  是  $\theta$  的一个估测量, 若满足以下两个条件:

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}_n] = \theta, \quad \lim_{n \rightarrow \infty} \text{VAR} [\hat{\theta}_n] = 0$$

则  $\hat{\theta}$  为  $\theta$  的一致估测量.

**定理 0.76** (一致性的函数不变性) 设  $\hat{\theta}_{n1}, \hat{\theta}_{n2}, \dots, \hat{\theta}_{nk}$  分别是  $\theta_1, \theta_2, \dots, \theta_k$  的一致性估计,  $G = g(\theta_1, \theta_2, \dots, \theta_k)$  是  $\theta_1, \theta_2, \dots, \theta_k$  的连续函数, 则

$$\hat{G} = g(\hat{\theta}_{n1}, \hat{\theta}_{n2}, \dots, \hat{\theta}_{nk})$$

是  $G$  的一致性估计.

## 一致性：例 0.143

**例 0.143** 设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本, 且总体  $X$  的概率密度函数为

$$f(X) = \begin{cases} \frac{1}{\theta} e^{-\frac{X}{\theta}}, & X > 0 \\ 0, & X \leq 0 \end{cases}$$

证明: 样本均值  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$  是  $\theta$  的一致估计量.

## 解答：例 0.143

题目：设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本，且总体  $X$  的概率密度函数为

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

证明：样本均值  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$  是  $\theta$  的一致估计量。

解答：

- 根据定理 0.75, 有

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}] = \theta, \quad \lim_{n \rightarrow \infty} \text{VAR}[\hat{\theta}] = \lim_{n \rightarrow \infty} \frac{\theta^2}{n} = 0.$$

证毕.

## 一致性：例 0.144

**例 0.144** 设  $X_1, X_2, \dots, X_n$  是取自总体  $X \sim U(0, \theta)$  的样本. 证明: 参数  $\theta$  的极大似然估计是一致估计量.

## 解答：例 0.144

- 根据前面的例题, 可知  $\theta$  的极大似然估计是  $\hat{\theta} = x_{(n)}$ . 设随机变量  $Z = x_{(n)}$ , 则  $Z$  的分布函数为

$$F_Z(z) = P(Z \leq z) = P(x_{(n)} \leq z) = \prod_{i=1}^n P(x_i \leq z) = \begin{cases} 1, & z > \theta \\ (\frac{z}{\theta})^n, & z \in [0, \theta] \\ 0, & z < 0 \end{cases}$$

由此可得当  $z \in [0, \theta]$  时随机变量  $Z$  的密度函数为  $f_Z(z) = nz^{n-1}/\theta^n$ .

- 进一步有

$$\mathbb{E}[\hat{\theta}] = \int_0^\theta \frac{nz^n}{\theta^n} dz = \frac{n}{n+1}\theta \Rightarrow \lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}] = \theta$$

又  $\mathbb{E}[Z^2] = \int_0^\theta \frac{nz^{n+1}}{\theta^n} dz = \frac{n}{n+2}\theta^2$ , 因此有

$$\text{VAR}[\hat{\theta}] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 \Rightarrow \lim_{n \rightarrow \infty} \text{VAR}[\hat{\theta}] = 0$$

由此,  $\hat{\theta}$  是  $\theta$  的有偏、但一致估计量.