

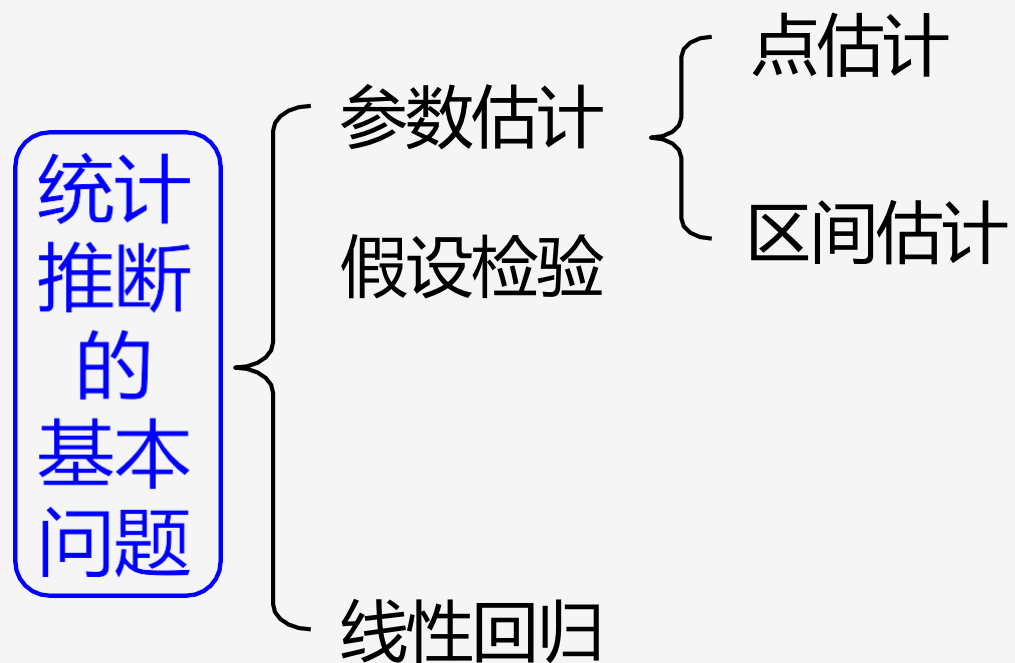


西安交通大学
XI'AN JIAOTONG UNIVERSITY

第七章 参数估计

言涿 特聘研究员
网络空间安全学院

2025年4月





西安交通大学
XI'AN JIAOTONG UNIVERSITY

7.1 点估计

什么是参数估计

参数通常是刻画总体某些概率特征的数量。

例如，正态分布 $N(\mu, \sigma^2)$ 中的参数 μ 就是该分布的均值，参数 σ^2 是该分布的方差。

当该参数未知时，从总体中抽取一个样本，用某种方法对该未知参数进行估计，这就是参数估计。

7.1 点估计

例如，假设总体 $X \sim N(\mu, \sigma^2)$ ，若参数 μ 与 σ^2 未知。

先从该总体中抽样得到样本 X_1, X_2, \dots, X_n ，然后构造样本函数，求出未知参数 μ 与 σ^2 的估计值或取值范围，这就是参数估计。

点估计

区间估计

7.1 点估计

假设总体 $X \sim F(x; \theta_1, \theta_2, \dots, \theta_m)$, 其中分布函数 F 的表达式已知, 但参数 $\theta_1, \theta_2, \dots, \theta_m$ 未知.

若记 $\theta = (\theta_1, \theta_2, \dots, \theta_m)$, 则总体分布可记为:

$$X \sim F(x; \theta)$$

参数 θ 的取值范围称为参数空间, 记为 Θ .

7.1 点估计

例如 $X \sim B(1, p)$, p 为未知参数, 则参数空间为:

$$\Theta = \{ p \mid 0 < p < 1 \}.$$

又如 $X \sim N(\mu, \sigma^2)$, μ, σ^2 为未知参数, 则参数空间为:

$$\Theta = \{ (\mu, \sigma^2) \mid -\infty < \mu < \infty, \sigma^2 > 0 \}$$

点估计的思想

X_1, X_2, \dots, X_n 是来自总体 $X \sim F(x; \theta_1, \dots, \theta_m)$ 的一个样本, $\theta_1, \dots, \theta_m$ 是未知参数。

构造 m 个:

$$\text{随机变量} \left\{ \begin{array}{l} \hat{\theta}_1(X_1, X_2, \dots, X_n) \\ \hat{\theta}_2(X_1, X_2, \dots, X_n) \\ \vdots \\ \hat{\theta}_m(X_1, X_2, \dots, X_n) \end{array} \right.$$

7.1 点估计

当把样本观测值 x_1, x_2, \dots, x_n 代入上述统计量里, 就得到 m 个数值:

$$\text{数值} \left\{ \begin{array}{l} \hat{\theta}_1(x_1, x_2, \dots, x_n) \\ \hat{\theta}_2(x_1, x_2, \dots, x_n) \\ \vdots \\ \hat{\theta}_m(x_1, x_2, \dots, x_n) \end{array} \right.$$

称 $\hat{\theta}_k(X_1, X_2, \dots, X_n)$ 为 θ_k 的**估计量** ($k=1, 2, \dots, m$);

称 $\hat{\theta}_k(x_1, x_2, \dots, x_n)$ 为 θ_k 的**估计值** ($k=1, 2, \dots, m$)。

- **问题** 1) 如何构造统计量?
2) 如何评价统计量?

- **常用的点估计方法**

矩估计法

极大似然估计法

最小二乘估计法

贝叶斯方法



西安交通大学
XI'AN JIAOTONG UNIVERSITY

7.1.1 矩估计

矩估计的思想

假设总体 $X \sim F(x; \theta_1, \theta_2, \dots, \theta_m)$, 参数 $\theta_1, \theta_2, \dots, \theta_m$ 未知。
且总体的 m 阶矩存在:

$$\mu_k(\theta_1, \theta_2, \dots, \theta_m) = E(X^k) \quad (k = 1, 2, \dots, m)$$

设 X_1, X_2, \dots, X_n 是来自总体 X 的一个样本, 则由辛钦大数定律, 有:

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mu_k(\theta_1, \theta_2, \dots, \theta_m), \quad n \rightarrow \infty.$$

7.1.1 矩估计

因此当 n 较大时有:

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \approx \mu_k(\theta_1, \theta_2, \dots, \theta_m). \quad (k = 1, 2, \dots, m)$$

令

$$\begin{aligned} \mu_1(\theta_1, \theta_2, \dots, \theta_m) &= A_1 \\ \mu_2(\theta_1, \theta_2, \dots, \theta_m) &= A_2 \\ &\vdots \\ \mu_m(\theta_1, \theta_2, \dots, \theta_m) &= A_m \end{aligned}$$

用样本矩 k 阶矩
作为
总体矩 k 阶矩
的估计量

其解 $\hat{\theta}_k(X_1, X_2, \dots, X_n)$ 称为 θ_k 的矩估计量
($k = 1, 2, \dots, m$).

7.1.1 矩估计

例 X_1, X_2, \dots, X_n 是来自总体 $X \sim P(\lambda)$ ($\lambda > 0$) 的样本, 求未知参数 λ 的矩估计量。

解 因为总体 $X \sim P(\lambda)$, 所以有

$$E(X) = \lambda$$

由矩估计原理, 用样本一阶矩, 即样本均值 \bar{X} 代替总体均值 $E(X)$, 得到

$$\bar{X} = \lambda \Rightarrow \hat{\lambda} = \bar{X}$$

7.1.1 矩估计

命题 不论总体 X 服从什么分布, 若其期望 μ 和方差 σ^2 的存在, 则 μ 和 σ^2 的矩估计量分别为:

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2 \triangleq \tilde{S}^2$$



修正的样本方差

7.1.1 矩估计

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, 求未知参数 μ, σ^2 的矩估计量。

解 因为正态分布 $N(\mu, \sigma^2)$ 的期望是 μ , 方差是 σ^2 , 所以由刚才的命题得到 μ, σ^2 的矩估计量分别为:

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2 \triangleq \tilde{S}^2$$

7.1.1 矩估计

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim B(m, p)$ 的样本, 求未知参数 p 的矩估计量.

解 因为总体 $X \sim B(m, p)$ 的一阶矩为:

$$E(X) = mp$$

令
$$mp = \bar{X}$$

求得 p 的矩估计量:
$$\hat{p} = \frac{1}{m} \bar{X}$$

 **问题** 若 m, p 都未知, 如何求 m, p 的矩估计?

7.1.1 矩估计

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim U(a, b)$ 的样本, 求未知参数 a, b 的矩估计量。

$$\hat{a} = \bar{X} - \sqrt{3}\tilde{S}, \quad \hat{b} = \bar{X} + \sqrt{3}\tilde{S}$$

7.1.1 矩估计

注 随机产生 $U(0, 1)$ 的随机数40个:

0.4387, 0.3816, 0.7655, 0.7952, 0.1869, 0.4898, 0.4456, 0.6463, 0.7094, 0.7547,
0.2760, 0.6797, 0.6551, 0.1626, 0.1190, 0.4984, 0.9597, 0.3404, 0.5853, 0.2238,
0.7513, 0.2551, 0.5060, 0.6991, 0.8909, 0.9593, 0.5472, 0.1386, 0.1493, 0.2575
0.8407, 0.2543, 0.8143, 0.2435, 0.9293, 0.3500, 0.1966, 0.2511, 0.6160, 0.4733

算得: $\bar{x} = 0.5059275$, $\tilde{s} = 0.2573$

计算得到 a, b 的矩估计值:

$$\hat{a} = \bar{x} - \sqrt{3}\tilde{s} = 0.0602, \hat{b} = \bar{x} + \sqrt{3}\tilde{s} = 0.9516$$

矩估计法小结

- 1) 原理直观;
- 2) 只用到总体矩, 方法简单, 若总体矩不存在, 则无法使用矩估计法;
- 3) 矩估计基于大数定律, 所以通常在大样本情况下, 才有较好的效果.

7.1.1 矩估计

例 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, 总体 X 服从参数为 θ 的 Cauchy 分布, 其密度函数为:

$$f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty$$

则 θ 的矩估计不存在。



西安交通大学
XI'AN JIAOTONG UNIVERSITY


7.1.2 最大似然估计

Fisher的极大似然思想

随机试验有多个可能结果，但在一次试验中，有且只有一个结果会出现。

如果在某次试验中，结果 ω 出现了，则认为该结果（事件 $\{\omega\}$ ）发生的概率 $P\{\omega\}$ 最大。

例如：字“𪛗”读什么音？

 **问题** 如何将Fisher的极大似然思想应用于参数估计?

假设总体 X 是离散型随机变量, 其分布律为:

$$P\{X = a_k\} = p_k(\theta) \quad (k = 1, 2, \dots)$$

其中 θ ($\theta \in \Theta$) 是未知参数.

7.1.2 最大似然估计

X_1, X_2, \dots, X_n 是来自总体 X 的样本。

x_1, x_2, \dots, x_n 是样本观测值。 ?

即事件 $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生了。

由 Fisher 的极大似然思想可以得到：

概率 $P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 最大。

7.1.2 最大似然估计

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$$

$$= P\{X_1 = x_1\}P\{X_2 = x_2\} \dots P\{X_n = x_n\}$$

$$= P\{X = x_1\}P\{X = x_2\} \dots P\{X = x_n\} = \mathbf{L(\theta)}$$



$$P\{X = a_k\} = p_k(\theta) \quad (k = 1, 2, \dots)$$

7.1.2 最大似然估计

定义1 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, x_1, x_2, \dots, x_n 是样本观测值。

1) 若 X 是离散型总体, 其分布律为:

$$P\{X = a_k\} = p_k(\theta) \quad (k = 1, 2, \dots)$$

$$\text{令 } L(\theta) = L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n P\{X_i = x_i\}, \theta \in \Theta$$

2) 若 X 是连续型总体, 其密度为 $f(x; \theta)$ 。

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

称 $L(\theta)$ 为似然函数。

7.1.2 最大似然估计

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim B(1, p)$ 的样本, x_1, x_2, \dots, x_n 是样本观测值。试写出似然函数。

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, x_1, x_2, \dots, x_n 是样本观测值。试写出似然函数。

7.1.2 最大似然估计

定义2 设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, x_1, x_2, \dots, x_n 是样本观测值。

$L(\theta)$ ($\theta \in \Theta$) 是似然函数。若存在统计量

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$$

使得:

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

则称 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 为 θ 的极大似然估计量, 简记为**MLE** (**M**aximum **L**ikelihood **E**stimate).

7.1.2 最大似然估计

极大似然估计求解的一般过程

1) 根据总体分布的表达式, 写出似然函数:

$$L(\theta_1, \theta_2, \dots, \theta_m) \quad (\theta = (\theta_1, \theta_2, \dots, \theta_m) \in \Theta)$$

2) 因为 $L(\theta_1, \theta_2, \dots, \theta_m)$ 与 $\ln L(\theta_1, \theta_2, \dots, \theta_m)$ 有相同的极值点, 称 $\ln L(\theta_1, \theta_2, \dots, \theta_m)$ 为对数似然函数, 记为 $l(\theta_1, \theta_2, \dots, \theta_m)$ 。求出 $l(\theta_1, \theta_2, \dots, \theta_m)$ 。

3) 求出 $l(\theta_1, \theta_2, \dots, \theta_m)$ 的极大值点, $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ 即为 $\theta_1, \theta_2, \dots, \theta_m$ 的 MLE.

7.1.2 最大似然估计

● 关于 3) 的说明

若 $l(\theta_1, \theta_2, \dots, \theta_m)$ 关于 θ_i ($i = 1, 2, \dots, m$) 可导, 则称

$$\left\{ \begin{array}{l} \frac{\partial l(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_1} = 0 \\ \frac{\partial l(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_2} = 0 \\ \vdots \\ \frac{\partial l(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_m} = 0 \end{array} \right.$$

为对数似然方程组。

7.1.2 最大似然估计

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim B(1, p)$ 的样本, x_1, x_2, \dots, x_n 是样本观测值。试求未知参数 p 的极大似然估计。

7.1.2 最大似然估计

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的样本, x_1, x_2, \dots, x_n 是样本观测值。试求未知参数 μ, σ^2 的极大似然估计。

7.1.2 最大似然估计

例 设 x_1, x_2, \dots, x_n 是来自总体 $X \sim f(x; \theta, c)$ 的样本观测值, 其中

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

试求未知参数 θ, c 的极大似然估计。

7.1.2 最大似然估计

解 似然函数为

$$\begin{aligned} L(\theta, c) &= \prod_{i=1}^n f(x_i; \theta, c) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{1}{\theta}(x_i - c)} \\ &= \theta^{-n} e^{-\frac{1}{\theta}n(\bar{x} - c)}, \quad c \leq x_{(1)}. \end{aligned}$$

对数似然函数为

$$l(\theta, c) = \ln L(\theta, c) = -n \ln \theta - \frac{n}{\theta}(\bar{x} - c), \quad c \leq x_{(1)}.$$

7.1.2 最大似然估计

对数似然方程组为

$$\begin{cases} \frac{\partial l(\theta, c)}{\partial \theta} = -\frac{n}{\theta} + \frac{n}{\theta^2} (\bar{x} - c) = 0 \\ \frac{\partial l(\theta, c)}{\partial c} = \frac{n}{\theta} = 0 \end{cases}$$

● **问题** 怎么解上对数似然方程组?

$$\frac{\partial l(\theta, c)}{\partial c} = \frac{n}{\theta} > 0 \implies l(\theta, c) \text{ 关于 } c \text{ 严格单调增加}$$

7.1.2 最大似然估计

似然函数

$$L(\theta, c) = \theta^{-n} e^{\frac{-1}{\theta} n(\bar{x} - c)}, c \leq x_{(1)}$$

⇒ c 的极大似然估计为 $\hat{c} = x_{(1)}$

将 c 的极大似然估计代入对数似然方程

$$\frac{\partial l(\theta, c)}{\partial \theta} = -\frac{n}{\theta} + \frac{n}{\theta^2} (\bar{x} - c) = 0$$

得到 θ 的极大似然估计为 $\hat{\theta} = \bar{x} - x_{(1)}$

● 问题 未知参数的极大似然估计唯一吗？

例 设 X_1, X_2, \dots, X_n 是来自总体 $X \sim U(\theta - 1, \theta + 1)$ 的样本, x_1, x_2, \dots, x_n 是样本观测值。试求 θ 的极大似然估计。

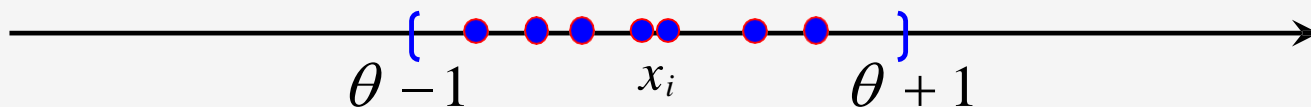
7.1.2 最大似然估计

解 因为总体的密度函数为

$$f(x; \theta) = \begin{cases} \frac{1}{2}, & \text{if } \theta - 1 < x < \theta + 1, \\ 0, & \text{otherwise.} \end{cases}$$

似然函数为

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = 2^{-n} \quad \theta \in ?$$



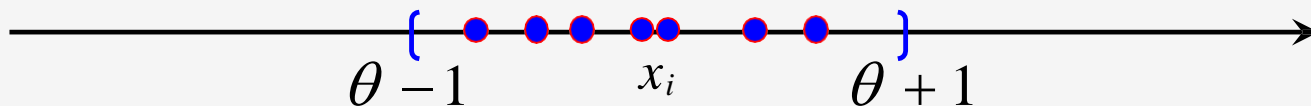
7.1.2 最大似然估计

解 因为总体的密度函数为

$$f(x; \theta) = \begin{cases} \frac{1}{2}, & \text{if } \theta - 1 < x < \theta + 1, \\ 0, & \text{otherwise.} \end{cases}$$

似然函数为

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = 2^{-n} \quad x_{(n)} - 1 < \theta < x_{(1)} + 1$$



7.1.2 最大似然估计

即当 $x_{(n)} - 1 < \theta < x_{(1)} + 1$ 时, 似然函数 $L(\theta)$ 取得最大值 2^{-n} 。

所以区间 $(x_{(n)} - 1, x_{(1)} + 1)$ 内任一点都是 θ 的极大似然估计。

极大似然估计的不变性

设 $\hat{\theta}$ 是 θ 的极大似然估计, $u = u(\theta)$ 是 θ 的函数,
且有单值反函数:

$$\theta = \theta(u)$$

则 $u(\hat{\theta})$ 是 u 的极大似然估计。

7.1.2 最大似然估计

- 例** 假设袋中有黑球和白球，其中白球所占比例为 $p(0 < p < 1)$ 未知。每次有放回地从袋中随机摸取 1 个球出来观测其颜色后放回，共摸了 m 个球，其中白球个数记为 X 。共重复了 n 次这样的试验，得到样本观察值为 x_1, x_2, \dots, x_n ，试求
- (1) p 的极大似然估计；
 - (2) 袋中白球和黑球数之比 R 的极大似然估计。

7.1.2 最大似然估计

解 (1) 先求 p 的极大似然估计

因为总体 $X \sim B(m, p)$, 所以似然函数为

$$\begin{aligned} L(p) &= \prod_{i=1}^n P\{X_i = x_i\} = \prod_{i=1}^n \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \\ &= p^{n\bar{x}} (1-p)^{n(m-\bar{x})} \prod_{i=1}^n \binom{m}{x_i} \end{aligned}$$

7.1.2 最大似然估计

对数似然函数为

$$l(p) = \ln L(p) = n\bar{x} \ln p + n(m - \bar{x}) \ln(1 - p) + \ln \prod_{i=1}^n \binom{m}{x_i}$$

对数似然函数方程

$$\frac{dl(p)}{dp} = \frac{n\bar{x}}{p} - \frac{n(m - \bar{x})}{1 - p} = 0$$

解得未知参数 p 的极大似然估计为 $\hat{p} = \frac{\bar{x}}{m}$

7.1.2 最大似然估计

(2) 求白球和黑球数之比 R 的极大似然估计

因为白球和黑球数之比

$$R = \frac{p}{1 - p}$$

所以由极大似然估计的不变性, 有

$$\hat{R} = \frac{\hat{p}}{1 - \hat{p}} = \frac{\bar{x}}{m - \bar{x}}$$

● **问题** 矩估计有不变性吗?