概率论与数理统计

第七章：参数估计 (parameter estimation)

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

内容大纲 I

- 点估计
  - 矩估计
    1. 求总体矩
    2. 样本矩代替总体矩
    3. 求矩估计量
  - 最大似然估计
    1. $L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$
    2. $\log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta).$
    3. $\dfrac{d \log L}{d\theta} = 0.$
- 评价标准
  - 无偏性 $\mathbb{E}\hat{\theta} = \theta.$
  - 有效性 $\mathrm{Var}(\hat{\theta}_1) \leq \mathrm{Var}(\hat{\theta}_2).$
  - 相合性 $\lim\limits_{n\to\infty} \mathsf{P}\Big(|\hat{\theta}_n - \theta| \geq \varepsilon\Big) = 0.$

内容大纲 II

- 区间估计
  - 双侧:
    - 单正态总体 $\mathcal{N}(\mu, \sigma^2)$: $\left(\bar{X} \pm \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right)$ (σ^2 已知), $\left(\bar{X} \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right)$ (σ^2 未知)
    - σ^2: $\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right)$.
  - 双正态总体 $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mathcal{N}(\mu_2, \sigma_2^2)$.
  - 单侧置信区间.

综述

数理统计问题

如何**选取样本**来**对总体**的种种统计特征**作出判断**。

参数估计问题

知道随机变量（总体）的**分布类型**，但**确切的形式不知道，根据样本来估计总体的参数**，这类问题称为**参数估计**（parametric estimation）

参数估计的类型

点估计、区间估计

参数 θ 的估计量

设总体的分布函数为 $F(x;\theta)$ (θ 未知), $X_1,\dots,X_n$ 为样本，构造一个统计量 $\hat\theta=\hat\theta(X_1,\dots,X_n)$ 来估计参数 θ，则称 $\hat\theta(X_1,\dots,X_n)$ 为**参数 θ 的估计量.**
将样本观测值 $x_1,\dots,x_n$ 代入 $\hat\theta(X_1,\dots,X_n)$，得到的值 $\hat\theta(x_1,\dots,x_n)$ 称为**参数 θ 的估计值**

点估计 (point estimation))

如果构造一个统计量 $\hat\theta(X_1,\dots,X_n)$ 来作为参数 θ 的估计量, 则称为 **参数 θ 的点估计.**

区间估计 (interval estimation)

如果构造两个统计量 $\hat\theta_1(X_1,\dots,X_n)$, $\hat\theta_2(X_1,\dots,X_n)$, 而用 $(\hat\theta_1,\hat\theta_2)$ 来作为参数 θ 可能取值范围的估计, 称为**参数 θ 的区间估计.**

实例

某工厂生产了一大批产品, 从中随机抽检了 n 件产品, 发现有 k 件次品. 如何估计整批产品的次品率 p?

分析

从该批产品中任取一件，令

$$X = \begin{cases} 1, & \text{该产品为次品}, \\ 0, & \text{该产品为好品}. \end{cases}$$

则 $X \sim \mathrm{Bin}(1,p)$ 为总体, 按题设, 从总体 X 抽取了一个样本 $X_1, \ldots, X_n$. 现要根据抽检结果, 对未知参数 p 的大小进行推断.

由大数律有

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{k}{n} \to p, \quad n \to \infty.$$

因此当 n 较大时, $\bar{X}$ 与 p 的"差别"应该较小. 故可用 $\hat{p} = \bar{X}$ 作为 p 的估计.

实例

从某厂生产的一批器件中随机抽取 10 件，测得其寿命值 (小时) 分别为

$$1010,\ 980,\ 975,\ 1050,\ 1100,\ 990,\ 1020,\ 1150,\ 1210,\ 960.$$

试问怎样估计该批器件的平均寿命？

分析

一般地，整批产品寿命 $X \sim \mathcal{N}(\mu, \sigma^2)$. 按题设，从总体 X 抽取了一个容量为 10 的样本. 现要根据抽检结果，对未知参数 μ 的大小进行推断.

$$\mathbb{E}\bar{X} = \mu, \quad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

因此，X 与 μ 的 "差别" 应该较小.
故器件的平均寿命估计值为

$$\hat{\mu} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1044.5.$$

点估计问题的一般提法

设总体 $X \sim F(x;\theta_1,\theta_2,\dots,\theta_m)$, 其中 F 的函数形式为已知, $\theta_1,\theta_2,\dots,\theta_m$ 为未知参数, $X_1,\dots,X_n$ 为来自总体 X 的样本. 若记 $\theta=(\theta_1,\theta_2,\dots,\theta_m)$, 则总体分布可记为 $X \sim F(x;\theta)$.

θ 的取值范围称为**参数空间**, 记为 Θ.

参数空间的例子

- 设总体 $X \sim \pi(\lambda)$, 则参数空间为 $\Theta=\{\lambda \mid \lambda>0\}$.
- 设总体 $X \sim \mathcal{N}(\mu,\sigma^2)$, 则参数空间为 $\Theta=\{(\mu,\sigma^2) \mid \mu\in\mathbb{R},\ \sigma>0\}$ (**形参空间**)
- 设某课程的考试成绩总体 $X \sim \mathcal{N}(\mu,\sigma^2)$, 则参数空间为 $\Theta=\{(\mu,\sigma^2) \mid \mu\in[0,100],\ \sigma\in(0,100)\}$. (**实参空间**)

参数推断问题:F 的函数形式已知，推断未知参数 θ.

θ 的点估计

构造一个统计量 $\hat\theta(X_1,\dots,X_n)$, 用统计量观察值 $\hat\theta(x_1,\dots,x_n)$ 作为未知参数 θ 的估计值. **二重性**: $\hat\theta(X_1,\dots,X_n)$ 为 θ 的**估计量**, $\hat\theta(x_1,\dots,x_n)$ 为 θ 的**估计值**.

用 $\hat{\theta}(X_1, \ldots, X_n)$ 估计 θ 的直观要求

- 估计的**误差**应较小
- 当 n 较大时，估计的**精度**应较高

对 “误差”、“精度” 不同的解释，有不同的估计方法.

常用的点估计方法

- 矩估计法 (Method of moments)
- 最大似然估计法 (Maximum likelihood)
- 最小二乘估计 (Least squares)

设总体 $X \sim F(x;\theta_1,\ldots,\theta_m)$, 其中 $\theta_1,\ldots,\theta_m$ 为未知参数, $X_1,\ldots,X_n$ 为来自总体 X 的样本. 设下列总体矩 $\alpha_k = \mathbb{E}X^k$, $k=1,2,\ldots,m$ 都存在.

由大数律 $A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k \to \mathbb{E}X^k = \alpha_k$. 因此

$$A_k \approx \alpha_k = \mathbb{E}X^k = \int_{-\infty}^{\infty} x^k dF(x;\theta_1,\ldots,\theta_m) := \alpha_k(\theta_1,\ldots,\theta_m).$$

求解方程组

$$\begin{cases} \alpha_1(\theta_1,\ldots,\theta_m) = A_1, \\ \alpha_2(\theta_1,\ldots,\theta_m) = A_2, \\ \qquad\cdots \\ \alpha_m(\theta_1,\ldots,\theta_m) = A_m \end{cases} \Rightarrow \begin{cases} \hat\theta_1 = \hat\theta_1(A_1,\ldots,A_m), \\ \hat\theta_2 = \hat\theta_2(A_1,\ldots,A_m), \\ \cdots \\ \hat\theta_m = \hat\theta_m(A_1,\ldots,A_m). \end{cases}$$

称 $(\hat\theta_1,\ldots,\hat\theta_m)$ 为 $(\theta_1,\ldots,\theta_m)$ 的**矩估计量 (矩估计)**.

上面的 $=$ 号其实是 $\approx$! 我们这里做了一个近似.

例: 均匀分布

设总体 X 在 $[a,b]$ 上服从均匀分布, a,b 未知. $X_1,\dots,X_n$ 是来自 X 的样本. 试求 a,b 的矩估计量.

解

由

$$\begin{cases} \mu_1 = \mathbb{E}X = \dfrac{a+b}{2}, \\ \mu_2 = \mathbb{E}X^2 = \dfrac{(b-a)^2}{12} + \dfrac{(a+b)^2}{4}, \end{cases}$$

解得 $a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}$, $b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)}$.
分别以 A_1, A_2 代替 μ_1, μ_2, 得到 a,b 的矩估计量分别为

$$\hat{a} = A_1 - \sqrt{3(A_2 - A_1^2)} = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

$$\hat{b} = A_1 + \sqrt{3(A_2 - A_1^2)} = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

例: 均值与方差

设总体 X 的均值 $\mu = \mathbb{E}X$, 方差 $\sigma^2 = \mathrm{Var}(X)$ 都存在. $X_1, \ldots, X_n$ 为总体样本, 求未知参数 μ, σ^2 的矩估计.

解

由

$$\begin{cases} \mu_1 = \mathbb{E}X = \mu, \\ \mu_2 = \mathbb{E}X^2 = \mu^2 + \sigma^2. \end{cases}$$

得矩估计

$$\hat{\mu} = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2 := \tilde{S}^2.$$

称 $\hat{S}^2$ 为**修正的样本方差**.

当 X 服从正态分布时, $\hat{\mu}$, $\hat{\sigma}^2$ 服从什么分布?

例题

设总体有均值 μ 与方差 σ^2, 今有样本观察值

$$-1.20,\ 0.82,\ 0.12,\ 0.45,\ -0.85,\ -0.30$$

求 μ 与 σ^2 的矩估计值.

解

$$\hat{\mu} = -0.16, \quad \hat{\sigma}^2 \approx 0.5.$$

例: 泊松分布

设 $X_1,\ldots,X_n$ 为总体 $X\sim\pi(\lambda)$, $\lambda>0$ 的样本. 求未知参数 λ 的矩估计.

总体一阶矩和样本一阶矩分别为 $\mathbb{E}X=\lambda, \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. 因此 λ 的矩估计为 $\hat{\lambda}=\bar{X}$.

例: 指数分布

设 $X_1,\ldots,X_n$ 为总体 $X\sim\frac{1}{\theta}e^{-x/\theta}\mathbb{1}_{x>0}$, $\theta>0$ 的样本. 求未知参数 θ 的矩估计.

总体一阶矩和样本一阶矩分别为 $\mathbb{E}X=\theta$, $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. 令 $\bar{X}=\theta$, 求得 θ 的矩估计为 $\hat{\theta}=\bar{X}$.

样本均值 $\bar{X}$ 是总体均值 $\mu:=\mathbb{E}X$ 的矩估计.

例

设有一批灯管，其寿命为 $X \sim \mathrm{Exp}(\lambda)$. 从中随机抽取 11 只，其寿命数据为

$$110, 184, 145, 122, 165, 143, 78, 129, 62, 130, 168.$$

用矩估计法估计 λ 的值.

解

由 $\mu_1 = \mathbb{E}X = \dfrac{1}{\lambda}$, 得 $\lambda = \dfrac{1}{\mu_1}$. 用 $A_1 = \bar{X}$ 代替 μ_1, 得参数 λ 的矩估计量为 $\hat{\lambda} = \dfrac{1}{\bar{X}}$, 而

$$\bar{X} = \frac{110 + 184 + 145 + 122 + 165 + 143 + 78 + 129 + 62 + 130 + 168}{11} \approx 130.55.$$

故 $\hat{\lambda} = 1/130.55 \approx 0.0077$.

例

设 $X_1, \ldots, X_n$ 为来自总体 $X \sim \text{Bin}(m, p)$ 的样本，$p \in (0,1)$. 求未知参数 p 的矩估计.

解

总体一阶矩为 $\mathbb{E}X = mp$, 样本一阶矩为 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. 令 $mp = \bar{X}$, 求得 p 的矩估计为

$$\hat{p} = \frac{\bar{X}}{m}.$$

小结

不管总体 X 服从何种分布，总体期望和方差的矩估计量分别为样本平均和样本 2 阶中心距：

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

约定

若 $\hat{\theta}$ 是未知参数 θ 的矩估计，则 $g(\theta)$ 的矩估计为 $g(\hat{\theta})$.

最大似然思想

实际问题

有两个射手，一人的命中率为 0.9，另一人的命中率为 0.1，现在他们中的一个向目标射击了一发，结果命中了，估计是谁射击的？

最大似然思想

一般说，事件 A 发生的概率与参数 $\theta \in \Theta$ 有关，θ 取值不同，则 $\mathrm{P}(A)$ 也不同. 因而应记事件 A 发生的概率为 $\mathrm{P}(A \mid \theta)$. 若 A 发生了，则认为此时的 θ 值应是在 Θ 中使 $\mathrm{P}(A \mid \theta)$ 达到最大的那一个. 这就是**最大似然思想**.

Fisher 的最大似然思想

一个随机试验有很多可能结果，如果在一次试验中，某结果发生了，则认为该结果 (事件) 发生的可能性最大.

例

老战士与一新同学一同进行射击训练，每人打了一枪，结果有一枪中靶. 试问这一枪是谁打中的？

分析

按照 Fisher 的最大似然思想，应该认为是老战士打中的较合理.

例

一袋中有红、白两颜色的球若干，只知道两种球的比例为 4:1，但不知道哪种颜色的球占多. 现从中任取一球，结果为白色. 问袋中哪种颜色的球较多？

分析

按照 Fisher 的最大似然思想，应该认为袋中白球较多.

离散型随机变量模型

设总体 X 为离散型随机变量，它的分布律为

$$\mathsf{P}(X=x)=p(x;\theta),\quad \theta\in\Theta.$$

现有样本观察值 $x_1,\dots,x_n$.

最大似然估计

考虑样本的**似然函数**

$$L(\theta)=\mathsf{P}(X_1=x_1,\dots,X_n=x_n)=\prod_{i=1}^{n}p(x_i;\theta).$$

求得 $L(\theta)$ 的最大值点 $\hat{\theta}(x_1,\dots,x_n)$，称为 θ 的**最大似然估计值**，称 $\hat{\theta}(X_1,\dots,X_n)$ 为 θ 的**最大似然估计量**.

连续型情形

用密度函数代替 $p(x;\theta)$.

最大似然估计

设 $X_1, \ldots, X_n$ 是总体 $X \sim f(x;\theta)$ 的样本 (f 为频率函数或密度函数). 令

$$L(\theta) = L(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} f(X_i;\theta),$$

称 $L(\theta)$ 为**似然函数**. 若存在统计量 $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ 使得

$$L(\hat{\theta}) = \max_{\theta\in\Theta} L(\theta; X_1, \ldots, X_n),$$

则称 $\hat{\theta}(X_1, \ldots, X_n)$ 为 θ 的**最大似然估计 (Maximum Likelihood Estimation, MLE)**, 称 $\hat{\theta}(x_1, \ldots, x_n)$ 为**最大似然估计值**.

最大似然估计法的一般步骤

(1) 构造似然函数 $L(\theta) = f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i;\theta)$. p : 频率函数或密度函数.

(2) 取自然对数: $\log L(\theta) = \sum_{i=1}^{n} \ln p(x_i;\theta)$.

(3) 令 $\dfrac{d \log L}{d\theta} = 0$ 求最大值.

例: 两点分布

设 $X \sim \mathrm{Bin}(1,p)$. $X_1,\ldots,X_n$ 是来自 X 的一个样本. 试求参数 p 的最大似然估计量.

解

设 $x_1,\ldots,x_n$ 是对应于 $X_1,\ldots,X_n$ 的一个样本值. 因为 X 的分布律是 $\mathsf{P}(X=x)=p^x(1-p)^{1-x}$, $x=0,1$, 所以似然函数及对数似然函数为

$$L(p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i},\quad \log L(p)=\ln p\cdot\sum_{i=1}^{n}x_i+(n-\sum_{i=1}^{n}x_i)\ln(1-p).$$

由

$$\frac{d}{dp}\ln L(p)=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{n-\sum_{i=1}^{n}x_i}{1-p}=0,$$

得最大似然估计值及最大似然估计量为

$$\hat{p}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},\quad \hat{p}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X},$$

例: 指数分布

设总体服从指数分布 $f(x;\theta)=\dfrac{1}{\theta}e^{-x/\theta}\cdot\mathbb{1}_{\{x>0\}}$. $X_1,\ldots,X_n$ 是来自于总体的样本, 求 θ 的最大似然估计.

解

似然函数为

$$L(\theta)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{X_i}{\theta}}=\theta^{-n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}X_i}.$$

由对数似然方程

$$\frac{d\ln L(\theta)}{d\theta}=-\frac{n}{\theta}+\frac{n\bar{X}}{\theta^2}=0$$

得 θ 的最大似然估计 $\hat{\theta}=\bar{X}$.

例: 正态分布

设 $X \sim \mathcal{N}(\mu, \sigma^2)$, μ, σ^2 为未知参数, $x_1, \ldots, x_n$ 是样本 X 的一个样本值. 求 μ, σ^2 的最大似然估计值.

解

X 的概率密度为 $f(x; \mu, \sigma^2) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. 则似然函数为

$$L(\mu, \sigma^2) = (2\pi)^{-n/2} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}.$$

对数似然函数为 $\ln L = -\dfrac{n}{2} \ln(2\pi) - \dfrac{n}{2} \ln \sigma^2 - \dfrac{1}{2\sigma^2} \displaystyle\sum_{i=1}^n (x_i - \mu)^2$. 解对数似然方程得

$$\begin{cases} \dfrac{\partial}{\partial \mu} \ln L = \dfrac{1}{\sigma^2} \left[\displaystyle\sum_{i=1}^n x_i - n\mu \right] = 0, \\ \dfrac{\partial}{\partial \sigma^2} \ln L = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2} \displaystyle\sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \dfrac{1}{n} \displaystyle\sum_{i=1}^n x_i = \bar{x}, \\ \hat{\sigma}^2 = \dfrac{1}{n} \displaystyle\sum_{i=1}^n (x_i - \bar{x})^2. \end{cases}$$

例：均匀分布

设 $X_1, \ldots, X_n$ 为来自均匀分布总体 $X \sim \mathrm{U}[a, b]$, $a < b$ 的样本. 求参数 a, b 的最大似然估计.

解

似然函数为

$$L(a, b) = \frac{1}{(b-a)^n}, \quad a \le X_{\min}, X_{\max} \le b.$$

显然, 对于任意满足上述条件的 a, b, 有

$$L(a, b) = \frac{1}{(b-a)^n} \le \frac{1}{(X_{\max} - X_{\min})^n}.$$

因此 a, b 的最大似然估计就是 $\hat{a} = \min\limits_{1 \le i \le n} X_i, \hat{b} = \max\limits_{1 \le i \le n} X_i$.

例题

设 X 的概率密度函数为

$$f(x;\theta) = \mathbb{1}_{(0,1)}(x) \cdot \sqrt{\theta} x^{\sqrt{\theta}-1}$$

$X_1, X_2, \ldots, X_n$ 是样本，求 θ 的极大似然估计量.

解

似然函数及对数似然函数为

$$L(\theta) = \theta^{\frac{n}{2}} \left(\prod_{i=1}^{n} x_i \right)^{\sqrt{\theta}-1}, \quad \ln L(\theta) = \frac{n}{2} \ln \theta + (\sqrt{\theta} - 1) \sum_{i=1}^{n} \ln x_i.$$

求解对数似然方程得

$$\frac{d \ln L(\theta)}{d\theta} = \frac{n}{2} \frac{1}{\theta} + \frac{1}{2\sqrt{\theta}} \sum_{i=1}^{n} \ln x_i = 0 \quad \Rightarrow \quad \sqrt{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln x_i}.$$

因此 θ 的极大似然估计量为 $\hat{\theta} = \dfrac{n^2}{\left[\sum_{i=1}^{n} \ln X_i \right]^2}$.

参数点估计

矩估计三步法

1. 求总体矩.
2. 样本矩代替总体矩.
3. 求出矩估计量（矩估计值）.

最大似然估计法三步法

1. 求（对数）似然函数.
2. 列出（对数）似然方程组.
3. 求（对数）似然函数的最大值点.

作业

- P218: 5 (a) (b) (c)
- 补充题
  1. 设总体 X 具有密度函数
$$f(x;\theta)=\begin{cases}\dfrac{2}{\theta^2}(\theta-x), & 0<x<\theta,\\ 0, & \text{其它}.\end{cases}$$
$X_1,\ldots,X_n$ 是样本, 求 θ 的矩估计.
  2. 设总体 X 的密度函数 (或频率函数) 如下, $X_1,\ldots,X_n$ 为样本, 求下列情况下 θ 的最大似然估计.
     (1) $p(x;\theta)=\dfrac{\theta^x}{x!}e^{-\theta}$, $x=0,1,2,\ldots$ $(\theta>0)$.
     (2) $f(x;\theta)=\theta\alpha x^{\alpha-1}e^{-\theta x^{\alpha}}$, $x>0$ (α 已知).
  3. 设总体 X 具有密度函数
$$f(x;\theta)=\begin{cases}\theta(1-x)^{\theta-1}, & 0\le x\le 1,\\ 0, & \text{其它}.\end{cases}$$
$X_1,\ldots,X_n$ 是其样本. 求 θ 的矩估计及最大似然估计.

例

设 $X_1,\ldots,X_n$ 为均匀分布总体 $X\sim \mathrm{U}(a,b)$ 的样本，按矩估计法，求得 a,b 的点估计分别为

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

按最大似然估计法，求得 a,b 的点估计分别为

$$\hat{a}=\min_{1\le i\le n} X_i,\quad \hat{b}=\max_{1\le i\le n} X_i.$$

例

设 $X_1,\ldots,X_n$ 为来自 Poisson 分布总体 $X\sim\pi(\lambda)$ 样本，因为

$$\mathbb{E}X=\mathrm{Var}(X)=\lambda,$$

所以 $\hat{\lambda}_1=\bar{X}$, $\hat{\lambda}_2=\tilde{S}^2$ 都可以作为未知参数 λ 的矩估计.

问题

用什么标准来评价和选择同一参数的不同的点估计量？**无偏性，有效性，相合性**

设总体 $X \sim F(x;\theta)$, $\theta \in \Theta$, Θ 为参数空间. 设 $X_1,\dots,X_n$ 为总体 X 的样本, $\hat{\theta} = \hat{\theta}(X_1,\dots,X_n)$ 为未知参数 θ 的点估计.
$\hat{\theta} = \hat{\theta}(X_1,\dots,X_n)$ **也是一个随机变量, 会在真值附近 "波动"**.

定义

若估计量 $\hat{\theta} = \hat{\theta}(X_1,\dots,X_n)$ 的数学期望存在, 且 $\forall \theta \in \Theta$ 有

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

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

称 $b_n(\hat{\theta}) = \mathbb{E}_\theta \hat{\theta} - \theta$ 为估计量 $\hat{\theta}$ 的**偏差 (偏)**.

- $b_n(\hat{\theta}) \neq 0$, 则 $\hat{\theta}$ 是 θ 的**有偏估计**.
- 若 $\lim\limits_{n\to\infty} b_n(\hat{\theta}) = 0$, 则称 $\hat{\theta}$ 为 θ 的**渐近无偏估计**.

期望和方差

例

无论 X 服从什么分布，若 $\mu := \mathbb{E}X$, $\sigma^2 := \mathrm{Var}(X)$ 都存在，则 $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = S^2$ 分别是 μ, σ^2 的无偏估计.

证明

我们有

$$\mathbb{E}\hat{\mu} = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n} X_i = \mu,$$

$$\mathbb{E}\hat{\sigma}^2 = \mathbb{E}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sigma^2$$

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2, \quad \mathbb{E}\tilde{S}^2 = \frac{n-1}{n}\mathbb{E}S^2 = \frac{n-1}{n}\sigma^2 \to \sigma^2.$$

因此修正样本方差 $\hat{S}^2$ 是 σ^2 的**渐近无偏估计**.

正态分布的方差和期望估计

设 $X_1, \ldots, X_n$ 为来自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本，则 μ 的矩估计和最大似然估计 $\hat{\mu} = \bar{X}$ 是 μ 的无偏估计，而 σ^2 的矩估计和最大似然估计

$$\hat{\sigma}^2 = \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

是 σ^2 的有偏估计或渐近无偏估计.

σ^2 的无偏估计是

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

例

设总体 X 的 k 阶矩 $\mu_k = \mathbb{E}X^k$, $k \geq 1$ 存在, 又设 $X_1, \ldots, X_n$ 是 X 的一个样本. 试证明不论总体服从什么分布, k 阶样本矩 $A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 k 阶总体矩 μ_k 的无偏估计量.

证明

$X_1, \ldots, X_n$ 与 X 同分布, 因此

$$\mathbb{E}X_i^k = \mathbb{E}X^k = \mu_k, \quad k = 1, 2, \ldots, n.$$

由期望的线性性, 有

$$\mathbb{E}A_k = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i^k = \mu_k.$$

例题

设 $X_1,\ldots,X_n$ 为总体 $X\sim \mathrm{U}(0,\theta)$ 的样本，试讨论 θ 的矩估计 $\hat{\theta}_M$ 和最大似然估计 $\hat{\theta}_L$ 的无偏性.

解

由 $\mathbb{E}X=\dfrac{\theta}{2}$，令 $\bar{X}=\dfrac{\theta}{2}$，得 θ 的矩估计为 $\hat{\theta}_M=2\bar{X}$. 于是

$$\mathbb{E}\hat{\theta}_M=2\mathbb{E}\bar{X}=2\mathbb{E}X=2\cdot\frac{\theta}{2}=\theta,$$

因此 $\hat{\theta}_M$ 是 θ 的无偏估计.

θ 的最大似然估计是 $\hat{\theta}_L=\max\limits_{1\le i\le n}X_i$，其密度函数为

$f_{\max}(z)=nf(z)[F(z)]^{n-1}=\dfrac{n}{\theta}\cdot\left(\dfrac{z}{\theta}\right)^{n-1}$，$0<z<\theta$. 因此

$$\mathbb{E}\hat{\theta}_L=\int_0^{\theta}z\cdot\frac{n}{\theta}\cdot\left(\frac{z}{\theta}\right)^{n-1}dz=\frac{n}{n+1}\theta<\theta.$$

所以 $\hat{\theta}_L=X_{(n))}$ 是 θ 的有偏估计 (渐近无偏估计). 经修正后

$\hat{\theta}=\dfrac{n+1}{n}\hat{\theta}_L=\dfrac{n+1}{n}X_{(n)}$ 是 θ 的无偏估计.

例

设总体 X 服从指数分布, 其概率密度为

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & \text{其它}.\end{cases}$$

其中参数 $\theta>0$ 为未知, 又设 $X_1,\ldots,X_n$ 是来自 X 的样本, 试证 $\bar{X}$ 和 $nZ=n\min(X_1,\ldots,X_n)$ 都是 θ 的无偏估计量.

证

因为 $\mathbb{E}\bar{X}=\mathbb{E}X=\theta$, 所以 $\bar{X}=\theta$ 的无偏估计量. 而 $Z=\min(X_1,\ldots,X_n)$ 具有概率密度

$$f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-nx/\theta}, & x>0,\\ 0, & \text{其它}.\end{cases}$$

故知 $\mathbb{E}Z=\dfrac{\theta}{n}$, $\mathbb{E}(nZ)=\theta$. 即 nZ 也是参数 θ 的无偏估计量.

无偏性的实际意义

- 在工程技术中：称 $\mathbb{E}\hat{\theta}-\theta$ 为系统误差.
- 在经济活动中：无偏性反映了商业行为的公平性.
- 在竞技评分中：无偏性反映了评分的公正性

问题

在什么情况下，无偏性才有意义？

无偏性只有在大量试验的情况下才有意义.

例

设 $X_1,\dots,X_n$ 为总体 $X\sim\pi(\lambda)$ 的样本

$$\mathbb{E}X=\mathrm{Var}(X)=\lambda,\quad \mathbb{E}\bar{X}=\lambda,\quad \mathbb{E}S^2=\mathrm{Var}(X)=\lambda,$$

故 $\hat{\lambda}_1=\bar{X}$, $\hat{\lambda}_2=S^2$ 都是 λ 的无偏估计.
更进一步, 对任意常数 c, 统计量

$$\hat{\lambda}=c\hat{\lambda}_1+(1-c)\hat{\lambda}_2=c\bar{X}+(1-c)S^2$$

都是 λ 的无偏估计.

定义

设 $X_1,\dots,X_n$ 是总体 $X\sim F(x;\theta)$, $\theta\in\Theta$ 的样本. 假设 $\hat{\theta}_1$, $\hat{\theta}_2$ 都是无偏估计. 即 $\mathbb{E}_\theta\hat{\theta}_1=\mathbb{E}_\theta\hat{\theta}_2=\theta$, $\theta\in\Theta$. 若对任意 $\theta\in\Theta$ 有 $\mathrm{Var}(\hat{\theta}_1)\le\mathrm{Var}(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ **较** $\hat{\theta}_2$ **有效**.

例子

设总体 X 服从指数分布，其概率密度为

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & \text{其它}.\end{cases}$$

其中参数 $\theta>0$ 未知. 又设 $X_1,\ldots,X_n$ 是来自于 X 的样本，则 $\bar{X}$ 和 $nZ=n\min(X_1,\ldots,X_n)$ 都是 θ 的无偏估计量. 试证 $n>1$ 时，$\bar{X}$ 比 nZ 有效.

证明

由于 $\mathrm{Var}(X)=\theta^2$，故有 $\mathrm{Var}(\bar{X})=\dfrac{\theta^2}{n}$. 再者，由于 $\mathrm{Var}(Z)=\dfrac{\theta^2}{n^2}$，$\mathrm{Var}(nZ)=\theta^2$. 当 $n>1$ 时，$\mathrm{Var}(\bar{X})<\mathrm{Var}(nZ)$，故 $\bar{X}$ 比 nZ 有效.

例子

设 $X_1,\ldots,X_n$ 为总体 $X\sim \mathrm{U}(0,\theta)$ $(\theta>0)$ 的样本. 试讨论的两个无偏估计 $\hat\theta_M=2\bar X$, $\hat\theta=\dfrac{n+1}{n}X_{(n)}$ 的有效性.

解

$$\mathrm{Var}_\theta(\hat\theta_M)=4\cdot\mathrm{Var}_\theta(\bar X)=4\cdot\frac{\mathrm{Var}_\theta(X)}{n}=4\cdot\frac{\theta^2}{12n}=\frac{\theta^2}{3n}.$$

$$\begin{aligned}\mathrm{Var}_\theta(\hat\theta)&=\frac{(n+1)^2}{n^2}\mathrm{Var}_\theta(X_{(n)})\\&=\frac{(n+1)^2}{n^2}\int_0^\theta\left(z-\frac{n}{n+1}\theta\right)^2\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1}dz=\frac{\theta^2}{n(n+2)}.\end{aligned}$$

当 $n>1$ 时, $\forall\theta>0$ 有 $\mathrm{Var}_\theta(\hat\theta)<\mathrm{Var}_\theta(\hat\theta_M)$, 故 $\hat\theta$ 较 $\hat\theta_M$ 有效.