第七章　　参数估计

参数估计的知识架构

```
参数估计
├── 点估计
│    ├── 矩估计
│    ├── 最大似然估计
│    └── 估计量的性质
│         ├── 无偏性
│         ├── 有效性
│         └──相合性
│
└── 区间估计
     ├── 置信区间的定义
     ├── 枢轴量思想
     ├── 常见参数的区间估计
     └── 置信水平与区间长度
```

1️⃣ 点估计（Point Estimation）

目标：

用一个**统计量（单个数值）**来估计未知参数。

（1）基本思想

- 设总体分布含未知参数 θ
- 从样本 $X_1, \ldots, X_n$ 构造统计量 $\hat{\theta} = g(X_1, \ldots, X_n)$
- 用 $\hat{\theta}$ 作为 θ 的估计值

（2）常见点估计方法（重点）

1. **矩估计（Method of Moments）**
    - 用样本矩代替理论矩
    - 思想直观，计算简单
2. **最大似然估计（Maximum Likelihood Estimation, MLE）**
    - 选择使样本出现概率最大的参数值
    - 理论地位最高、应用最广

2️⃣ 区间估计（Interval Estimation）

目标：
给出一个**区间**，并说明参数落在该区间内的**可信程度。**

（1）置信区间的基本形式

$$P\big(\theta \in [L(X), U(X)]\big) = 1 - \alpha$$

- $[L, U]$：置信区间
- $1 - \alpha$：置信水平（如 95%、99%）

（2）常见置信区间类型

- **均值的置信区间**
  - 正态总体（σ 已知 / 未知）
  - 大样本近似（CLT）
- **方差的置信区间**
  - 基于 χ^2 分布
- **比例参数的置信区间**
- **参数函数的置信区间**

（3）构造方法（核心思想）

- 枢轴量法（pivot）
- 利用已知分布（正态 / t / χ^2 / F）

> 👉 区间估计回答的是：
> **“我们对这个估计有多大把握？”**

第一节　　点估计

一、点估计问题的提法

二、估计量的求法

三、小结

一、点估计问题的提法

设总体 X 的分布函数形式已知，但它的一个或多个参数为未知，借助于总体 X 的一个样本来估计总体未知参数的值的问题称为**点估计问题**.

例1 在某炸药制造厂，一天中发生着火现象的次数 X 是一个随机变量，假设它服从以 $\lambda > 0$ 为参数的泊松分布，参数 λ 为未知，设有以下的样本值，试估计参数 λ.

着火次数 k	0	1	2	3	4	5	6	
发生 k 次着火的天数 n_k	75	90	54	22	6	2	1	$\Sigma = 250$

解 因为 $X \sim \pi(\lambda)$, 所以 $\lambda = E(X)$.

用样本均值来估计总体的均值 ***E***(***X***).

$$\bar{x} = \frac{\sum_{k=0}^{6} k n_k}{\sum_{k=0}^{6} n_k} = \frac{1}{250}(0 \times 75 + 1 \times 90 + 2 \times 54 + 3 \times 22 + 4 \times 6 + 5 \times 2 + 6 \times 1) = 1.22.$$

故 $E(X) = \lambda$ 的估计为1.22 .

点估计问题的一般提法（定义）

设总体X的分布函数$F(x;\theta)$的形式为已知, θ是待估参数. $X_1,X_2,\cdots,X_n$是X的一个样本, $x_1,x_2,\cdots,x_n$为相应的一个样本值.

点估计问题就是要构造一个适当的统计量$\hat\theta(X_1,X_2,\cdots,X_n)$, 用它的观察值$\hat\theta(x_1,x_2,\cdots,x_n)$来估计未知参数$\theta$.

$\hat\theta(X_1,X_2,\cdots,X_n)$称为$\theta$的估计量.

$\hat\theta(x_1,x_2,\cdots,x_n)$称为$\theta$的估计值.

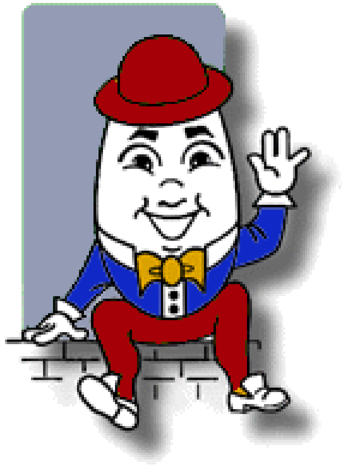
通称估计, 简记为$\hat\theta$.

二、估计量的求法

由于估计量是样本的函数，是随机变量，故对不同的样本值，得到的参数值往往不同，如何求估计量是关键问题.

常用构造估计量的方法：(两种)

矩估计法和最大似然估计法.



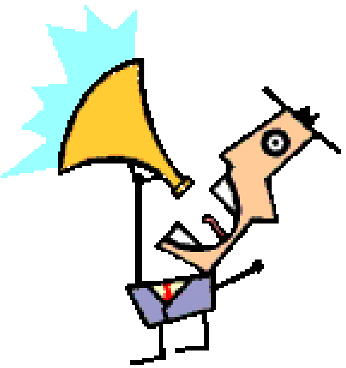
1. 矩估计法

设 X 为连续型随机变量,其概率密度为 $f(x;\theta_1,\theta_2,\cdots,\theta_k)$, 或 X 为离散型随机变量,其分布律为 $P\{X=x\}=p(x;\theta_1,\theta_2,\cdots,\theta_k)$, 其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数,

若 $X_1,X_2,\cdots,X_n$ 为来自 X 的样本,

假设总体 X 的前 k 阶矩存在,

且均为 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数, 即



$$\mu_l = E(X^l) = \int_{-\infty}^{+\infty} x^l f(x;\theta_1,\theta_2,\cdots,\theta_k)\mathrm{d}x \quad (X\text{为连续型})$$

$$\text{或 } \mu_l = E(X^l) = \sum_{x\in R_X} x^l p(x;\theta_1,\theta_2,\cdots,\theta_k), \ (X\text{为离散型})$$

其中R_X是x可能取值的范围，$l = 1,2,\cdots,k$

因为样本矩 $A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l$ 依概率收敛于相应的

总体矩 μ_l $(l = 1, 2,\cdots, k)$,

样本矩的连续函数依概率收敛于相应的总体矩

的连续函数.

矩估计法的定义

用样本矩来估计总体矩,用样本矩的连续函数来估计总体矩的连续函数,这种估计法称为**矩估计法**.

矩估计法的具体做法: 令 $\mu_l = A_l,\quad l = 1,2,\cdots,k.$

这是一个包含 k 个未知参数 $\theta_1,\theta_2,\cdots,\theta_k$ 的方程组,解出其中 $\theta_1,\theta_2,\cdots,\theta_k$.

用方程组的解 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 分别作为 $\theta_1,\theta_2,\cdots,\theta_k$ 的估计量,这个估计量称为矩估计量.

矩估计量的观察值称为矩估计值.

例3 设总体 X 在 $[0,\theta]$ 上服从均匀分布,其中 θ $(\theta>0)$ 未知,$(X_1,X_2,\cdots,X_n)$ 是来自总体 X 的样本,求 θ 的估计量.

解 因为 $\mu_1 = E(X) = \dfrac{\theta}{2}$,

根据矩估计法,令 $\dfrac{\hat{\theta}}{2} = A_1 = \overline{X}$,

所以 $\hat{\theta} = 2\overline{X}$ 为所求 θ 的估计量.

例4 设总体 X 在$[a,b]$上服从均匀分布,其中a, b未知,$(X_1, X_2,\cdots, X_n)$ 是来自总体X的样本,求a, b的估计量.

解 $\mu_1 = E(X)= \dfrac{a+b}{2}$,

$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \frac{(a-b)^2}{12} + \frac{(a+b)^2}{4},$$

令 $\dfrac{a+b}{2} = A_1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,

$$\frac{(a-b)^2}{12} + \frac{(a+b)^2}{4} = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2},$$

即 $\begin{cases} a+b=2A_1, \\ b-a=\sqrt{12(A_2-A_1^2)}. \end{cases}$

解方程组得到*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5 设总体 X 服从几何分布, 即有分布律

$$P\{X=k\}=p(1-p)^{k-1} \quad (k=1,2,\cdots),$$

其中 $p\ (0<p<1)$ 未知, $(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本, 求 p 的估计量.

解 $\mu_1 = E(X) = \sum_{k=1}^{\infty} k p(1-p)^{k-1} = \dfrac{1}{p},$

令 $\dfrac{1}{\hat{p}} = A_1 = \overline{X},$

所以 $\hat{p} = \dfrac{1}{\overline{X}}$ 为所求 p 的估计量.

例6 设总体X的均值μ和方差σ^2都存在,且有$\sigma^2>0$,但μ和σ^2均为未知,又设$X_1,X_2,\cdots,X_n$是一个样本,求μ和σ^2的矩估计量.

解 $\mu_1=E(X)=\mu,$

$\mu_2=E(X^2)=D(X)+[E(X)]^2=\sigma^2+\mu^2,$

令 $\begin{cases}\hat{\mu}=A_1,\\ \hat{\sigma}^2+\hat{\mu}^2=A_2.\end{cases}$

解方程组得到矩估计量分别为 $\hat{\mu}=A_1=\bar{X},$

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

上例表明:

总体均值与方差的矩估计量的表达式不因不同的总体分布而异.

例 $X \sim N(\mu,\sigma^2)$, μ,σ^2 未知,即得 μ,σ^2 的矩估计量

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

一般地,

用样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 作为总体 X 的均值的矩估计,

用样本二阶中心矩 $B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 作为总体

X 的方差的矩估计.

2. 最大似然估计法

(1) 设总体 X 属离散型

似然函数的定义

设分布律 $P\{X=k\}=p(x;\theta)$, θ 为待估参数, $\theta\in\Theta$,

(其中 Θ 是 θ 可能的取值范围)

$X_1,X_2,\cdots,X_n$是来自总体 X的样本,

则 $X_1,X_2,\cdots,X_n$ 的联合分布律为 $\prod_{i=1}^{n} p(x_i;\theta)$.

又设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值.

则样本 $X_1, X_2, \cdots, X_n$ 取到观察值 $x_1, x_2, \cdots, x_n$ 的概率，

即事件 $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ 发生的概率为

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta), \quad \theta \in \Theta,$$

$L(\theta)$ 称为样本似然函数.

最大似然估计法

得到样本值 $x_1,x_2,\cdots,x_n$ 时,选取使似然函数 $L(\theta)$

取得最大值的 $\hat{\theta}$ 作为未知参数 θ 的估计值,

即 $L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max\limits_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$.

(其中 Θ 是 θ 可能的取值范围)

这样得到的 $\hat{\theta}$ 与样本值 $x_1,x_2,\cdots,x_n$ 有关,记为

$\boxed{\hat{\theta}(x_1,x_2,\cdots,x_n)}$, 参数 θ 的最大似然估计值,

$\boxed{\hat{\theta}(X_1,X_2,\cdots,X_n)}$ 参数 θ 的最大似然估计量.

(2) 设总体 X 属连续型

似然函数的定义

设概率密度为 $f(x;\theta)$, θ 为待估参数， $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本,

则 $X_1, X_2, \cdots, X_n$ 的联合密度为 $\prod_{i=1}^{n} f(x_i;\theta)$.

又设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值.

则随机点$(X_1,X_2,\cdots,X_n)$落在点$(x_1,x_2,\cdots,x_n)$的邻域(边长分别为$\mathrm{d}x_1,\mathrm{d}x_2,\cdots,\mathrm{d}x_n$的$n$维立方体)内的概率近似地为 $\prod_{i=1}^{n} f(x_i;\theta)\mathrm{d}x_i$,

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

$L(\theta)$称为样本的似然函数.

若 $L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta).$

$\boxed{\hat{\theta}(x_1,x_2,\cdots,x_n)}$ 参数θ的最大似然估计值,

$\boxed{\hat{\theta}(X_1,X_2,\cdots,X_n)}$ 参数θ的最大似然估计量.

最大似然估计法是由费舍尔引进的.

求最大似然估计量的步骤:

(一) 写出似然函数

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

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

(二) 取对数

$$\ln L(\theta)=\sum_{i=1}^{n}\ln p(x_i;\theta)$$ 或 $$\ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i;\theta);$$

(三) 对 θ 求导 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}$， 并令 $\boxed{\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0}$, 对数似然方程

解方程即得未知参数 θ 的最大似然估计值 $\hat{\theta}$.

最大似然估计法也适用于分布中含有多个未知参数的情况. 此时只需令

$$\boxed{\frac{\partial}{\partial\theta_i}\ln L=0,\qquad i=1,2,\cdots,k.}$$ 对数似然方程组

解出由 k 个方程组成的方程组, 即可得各未知参数 $\theta_i\ (i=1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

例7 设 $X \sim B(1,p)$, $X_1, X_2, \cdots, X_n$是来自X的一个样本,求p的最大似然估计量.

解 设$x_1, x_2, \cdots, x_n$为相应于样本$X_1, X_2, \cdots, X_n$的一个样本值,

X的分布律为 $P\{X = x\} = p^x (1-p)^{1-x}, x = 0,1,$

似然函数 $L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i},$$

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

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

解得 p 的最大似然估计值 $$p=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

p 的最大似然估计量为 $$\hat{p}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}.$$

这一估计量与矩估计量是相同的.

例8 设 X 服从参数为 $\lambda\,(\lambda>0)$ 的泊松分布，$X_1,X_2,\cdots,X_n$ 是来自 X 的一个样本，求 λ 的最大似然估计量.

解 因为X的分布律为

$$P\{X=x\}=\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda},\quad (x=0,1,2,\cdots,n)$$

所以λ的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\left(\frac{\lambda^{x_i}}{x_i!}\mathrm{e}^{-\lambda}\right)=\mathrm{e}^{-n\lambda}\,\frac{\lambda^{\sum\limits_{i=1}^{n}x_i}}{\prod\limits_{i=1}^{n}(x_i!)},$$

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \sum_{i=1}^{n} \ln(x_i!),$$

令 $$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0,$$

解得 λ 的最大似然估计值 $$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

λ 的最大似然估计量为 $$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

这一估计量与矩估计量是相同的.

例9 设总体 $X \sim N(\mu,\sigma^2)$, μ,σ^2为未知参数, $x_1,x_2,\cdots,x_n$ 是来自X的一个样本值, 求μ和σ^2的最大似然估计量.

解 X的概率密度为

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

X 的似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}},$$

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

$$令\begin{cases} \dfrac{\partial}{\partial\mu}\ln L(\mu,\sigma^2) = 0, \\ \dfrac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2) = 0, \end{cases}$$

$$\begin{cases} \dfrac{1}{\sigma^2}\left[\sum_{i=1}^{n}x_i - n\mu\right] = 0 , \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 , \end{cases}$$

由$\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} x_i - n\mu\right]=0$解得 $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n} x_i=\bar{x},$

由$-\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0$解得

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

故μ和σ^2的最大似然估计量分别为

$\hat{\mu}=\bar{X},\ \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$ 它们与相应的矩估计量相同.

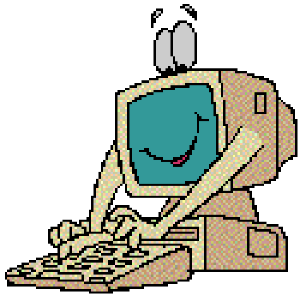
例10 设总体 X 在 $[a,b]$ 上服从均匀分布,其中 a, b 未知, $x_1,x_2,\cdots,x_n$ 是来自总体 X 的一个样本值,求 a,b 的最大似然估计量.

解 记 $x_{(l)}=\min(x_1,x_2,\cdots,x_n)$,

$$x_{(h)}=\max(x_1,x_2,\cdots,x_n),$$

X 的概率密度为

$$f(x;a,b)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & \text{其他}.\end{cases}$$



因为$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_{(l)}, x_{(h)} \le b$,

作为a, b的函数的似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_{(l)},\ b \ge x_{(h)}, \\ 0, & \text{其他}\end{cases}$$

于是对于满足条件 $a \le x_{(l)},\ b \ge x_{(h)}$的任意$a,b$有

$$L(a,b)=\frac{1}{(b-a)^n} \le \frac{1}{(x_{(h)}-x_{(l)})^n},$$

即似然函数$L(a,b)$在$a=x_{(l)}$, $b=x_{(h)}$时

取到最大值$(x_{(h)}-x_{(l)})^{-n}$,

a,b的最大似然估计值

$$\hat{a}=x_{(l)}=\min_{1\le i\le n} x_i, \qquad \hat{b}=x_{(h)}=\max_{1\le i\le n} x_i,$$

a,b的最大似然估计量

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

最大似然估计的性质　　　　　**不变性**

设 θ 的函数 $u=u(\theta), \theta\in\Theta$ 具有单值反函数 $\theta=\theta(u), u\in\mathrm{U}$. 又设 $\hat{\theta}$ 是 X 的概率密度函数 $f(x;\theta)$ (f 形式已知) 中的参数 θ 的最大似然估计, 则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计.

证明　因为$\hat{\theta}$是θ的最大似然估计值,

所以 $L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max\limits_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta)$,

其中$x_1,x_2,\cdots,x_n$是来自总体 X的一个样本值,

由于　$\hat{u}=u(\hat{\theta})$,　$\hat{\theta}=\theta(\hat{u})$,

故 $L(x_1,x_2,\cdots,x_n;\theta(\hat{u}))=\max\limits_{u\in U} L(x_1,x_2,\cdots,x_n;\theta(u))$,

于是 $\hat{u}=u(\hat{\theta})$是$u(\theta)$的最大似然估计.

此性质可以推广到总体分布中含有多个未知参数的情况.

如例9中σ^2 的最大似然估计值为 $\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$,

函数 $u=u(\sigma^2)=\sqrt{\sigma^2}$ 有单值反函数 $\sigma^2=u^2(u\geq 0)$,

故标准差 σ 的最大似然估计值为

$$\hat{\sigma}=\sqrt{\hat{\sigma}^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}.$$

三、小结

两种求点估计的方法: $\begin{cases} \text{矩估计法} \\ \text{最大似然估计法} \end{cases}$

在统计问题中往往先使用最大似然估计法,

在最大似然估计法使用不方便时, 再用矩估计法.

似然函数 $L(\theta) = L(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta)$

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

随堂练习

9. 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{1}{1-\theta}, & \theta\le x\le 1,\\ 0, & \text{其他},\end{cases}$$

其中 $0<\theta<1$ 为未知参数，$X_1,X_2,\ldots,X_n$ 为来自该总体的简单随机样本。

(1) 求 θ 的矩估计量 $\hat{\theta}_1$；

(2) 求 θ 的最大似然估计量 $\hat{\theta}_2$。

解：

(1) 计算矩估计量：

$$E(X) = \int_\theta^1 x\frac{1}{1-\theta}\,dx = \frac{1-\theta^2}{2(1-\theta)} = \frac{1+\theta}{2}.$$

令 $E(X) = \bar{X}$，得

$$\hat{\theta}_1 = 2\bar{X} - 1.$$

(2) 最大似然估计量：

$$L(\theta) = \prod_{i=1}^n f(x_i;\theta) = \left(\frac{1}{1-\theta}\right)^n \mathbf{1}\{\theta \le x_{(1)}\},$$

其中 $x_{(1)} = \min\{X_1,\ldots,X_n\}$。

在 $\theta \le x_{(1)}$ 上，

$$\ln L(\theta) = -n\ln(1-\theta), \qquad \frac{d}{d\theta}\ln L(\theta) = \frac{n}{1-\theta} > 0,$$

故 $\ln L(\theta)$ 随 θ 单调递增，在可行区间右端取得最大值：

$$\hat{\theta}_2 = x_{(1)}.$$

随堂练习

10. 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{\theta^2}{x^3}e^{-\theta/x}, & x>0,\\ 0, & \text{其他},\end{cases}$$

其中 $\theta>0$ 为未知参数，$X_1,X_2,\ldots,X_n$ 为来自总体 X 的简单随机样本。

(1) 求 θ 的矩估计量；

(2) 求 θ 的最大似然估计量。

解：

(1) 计算矩估计量：

$$E(X) = \int_0^\infty x \frac{\theta^2}{x^3} e^{-\theta/x}\, dx = \theta.$$

因而

$$\hat{\theta}_1 = \bar{X}.$$

(2) 最大似然估计量：

似然函数为

$$L(\theta) = \prod_{i=1}^n \frac{\theta^2}{x_i^3} e^{-\theta/x_i} = \frac{\theta^{2n}}{\prod_{i=1}^n x_i^3} \exp\left(-\theta \sum_{i=1}^n \frac{1}{x_i}\right).$$

对数似然

$$\ln L(\theta) = 2n \ln\theta - 3\sum_{i=1}^n \ln x_i - \theta \sum_{i=1}^n \frac{1}{x_i}.$$

求导并令其为零：

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

解得

$$\hat{\theta}_2 = \frac{2n}{\sum\limits_{i=1}^n \frac{1}{x_i}}.$$

随堂练习

11. 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{\theta}{2}, & 0<x<2,\\ \dfrac{1}{2}-\dfrac{\theta}{2}, & 2\le x\le 4,\\ 0, & \text{其他},\end{cases}$$

其中 $0<\theta<2$ 为未知参数。$X_1,X_2,\ldots,X_n$ 为来自总体 X 的简单随机样本，记 k 为样本值 $x_1,x_2,\ldots,x_n$ 中小于 2 的个数。

(1) 求 θ 的矩估计量；

(2) 求 θ 的最大似然估计量。

解：

(1) 计算矩估计量：

$$\begin{aligned} E(X) &= \int_0^2 x\frac{\theta}{2}\,dx + \int_2^4 x\left(\frac{1}{2}-\frac{\theta}{2}\right)\,dx \\ &= \frac{\theta}{2}\cdot\frac{2^2}{2} + \left(\frac{1}{2}-\frac{\theta}{2}\right)\cdot\frac{4^2-2^2}{2} \\ &= 3-2\theta. \end{aligned}$$

令 $E(X)=\bar{X}$，得

$$\hat{\theta}_1 = \frac{3-\bar{X}}{2}.$$

(2) 最大似然估计量：

若有 k 个样本落在 $(0,2)$ 内，其余 $n-k$ 个落在 $[2,4]$ 内，则

$$L(\theta) = \left(\frac{\theta}{2}\right)^k \left(\frac{1}{2}-\frac{\theta}{2}\right)^{n-k} = \left(\frac{\theta}{2}\right)^k \left(\frac{1-\theta}{2}\right)^{n-k}.$$

对数似然

$$\ln L(\theta) = k\ln\theta + (n-k)\ln(1-\theta) + \text{常数}.$$

求导并令其为零：

$$\frac{k}{\theta} - \frac{n-k}{1-\theta} = 0 \quad\Rightarrow\quad \hat{\theta}_2 = \frac{k}{n}.$$

由 $0\le k\le n$ 可知 $0\le\hat{\theta}_2\le 1<2$，满足参数空间 $0<\theta<2$。

第三节　　估计量的评选标准

一、问题的提出

二、无偏性

三、有效性

四、相合性

五、小结

一、问题的提出

从前一节可以看到, 对于同一个参数, 用不同的估计方法求出的估计量可能不相同, 如第一节的例4和例10. 而且, 很明显, 原则上任何统计量都可以作为未知参数的估计量.

问题

(1)对于同一个参数究竟采用哪一个估计量好?

(2)评价估计量的标准是什么?

下面介绍几个常用标准.

二、无偏性

若$X_1,X_2,\cdots,X_n$为总体X的一个样本，

$\theta\in\Theta$是包含在总体X的分布中的待估参数，

(Θ是θ的取值范围)

若估计量$\hat{\theta}=\theta(X_1,X_2,\cdots,X_n)$的数学期望

$E(\hat{\theta})$存在，且对于任意$\theta\in\Theta$有$E(\hat{\theta})=\theta$，则称

$\hat{\theta}$是θ的无偏估计量.

无偏估计的实际意义：无系统误差.

例1 设总体 X 的 k 阶矩 $\mu_k = E(X^k)\ (k \geq 1)$ 存在，又设 $X_1, X_2, \cdots, X_n$ 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 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 k 阶总体矩 μ_k 的无偏估计.

证 因为 $X_1, X_2, \cdots, X_n$ 与 X 同分布，

故有 $E(X_i^k) = E(X^k) = \mu_k, \quad i = 1, 2, \cdots, n.$

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

故 k 阶样本矩 A_k 是 k 阶总体矩 μ_k 的无偏估计.

特别的:

不论总体 X 服从什么分布,

只要它的数学期望存在,

$\overline{X}$ 总是总体 X 的数学期望 $\mu_1 = E(X)$ 的无偏估计量.

例2 对于均值 μ, 方差 $\sigma^2 > 0$ 都存在的总体, 若 μ, σ^2 均为未知, 则 σ^2 的估计量 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是有偏的(即不是无偏估计).

证 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = A_2 - \overline{X}^2$,

因为 $E(A_2) = \mu_2 = \sigma^2 + \mu^2$,

又因为 $E(\overline{X}^2) = D(\overline{X}) + [E(\overline{X})]^2 = \frac{\sigma^2}{n} + \mu^2$,

所以 $E(\hat{\sigma}^2) = E(A_2 - \overline{X}^2) = E(A_2) - E(\overline{X}^2)$

$$=\frac{n-1}{n}\sigma^2 \neq \sigma^2,$$ 所以$\hat{\sigma}^2$是有偏的.

若以$\dfrac{n}{n-1}$乘$\hat{\sigma}^2$,所得到的估计量就是无偏的.

(这种方法称为无偏化).

$$E\left(\frac{n}{n-1}\hat{\sigma}^2\right)=\frac{n}{n-1}E(\hat{\sigma}^2)=\sigma^2.$$

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

即S^2是σ^2的无偏估计,故通常取S^2作σ^2的估计量.

例3 设总体X在 $[0,\theta]$上服从均匀分布, 参数$\theta > 0$, $X_1, X_2, \cdots, X_n$是来自总体X的样本，试证明$2\bar{X}$和$\dfrac{n+1}{n}\max(X_1, X_2, \cdots, X_n)$ 都是θ的无偏估计.

证 因为 $E(2\overline{X}) = 2E(\overline{X}) = 2E(X) = 2\times\dfrac{\theta}{2} = \theta,$

所以 $2\overline{X}$ 是 θ 的无偏估计量.

因为 $X_h = \max(X_1, X_2, \cdots, X_n)$的概率密度为

$$f(x) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n}, & 0 \le x \le \theta, \\ 0, & \text{其他} \end{cases}$$

所以 $$E(X_h)=\int_0^\theta x\cdot\frac{nx^{n-1}}{\theta^n}\mathrm{d}x$$

$$=\frac{n}{n+1}\theta,$$

故有 $$E\left(\frac{n+1}{n}X_h\right)=\theta,$$

故$\dfrac{n+1}{n}\max(X_1,X_2,\cdots,X_n)$ 也是θ的无偏估计量.

例4 设总体 X 服从参数为 θ 的指数分布, 概率密度

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

其中参数 $\theta>0$, 又设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 试证 $\overline{X}$ 和 $nZ=n[\min(X_1,X_2,\cdots,X_n)]$ 都是 θ 的无偏估计.

证明 因为 $E(\overline{X})=E(X)=\theta$,

所以 $\overline{X}$ 是 θ 的无偏估计量.

而 $Z=\min(X_1,X_2,\cdots,X_n)$ 服从参数为 $\frac{\theta}{n}$ 的指数分布,

概率密度 $f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}\mathrm{e}^{-\frac{nx}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$

故知 $E(Z)=\dfrac{\theta}{n}$, $E(nZ)=\theta$,

所以 nZ 也是 θ 的无偏估计量.

由以上两例可知,一个参数可以有不同的无偏估计量.

$$Z \sim \mathrm{Exp}\left(\frac{\theta}{n}\right), \quad f_Z(z) = \frac{n}{\theta} e^{-nz/\theta},\ z > 0$$

设 $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Exp}(\theta)$，

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta},\ x > 0; \qquad F(x) = P(X \le x) = 1 - e^{-x/\theta},\ x > 0.$$

令 $Z = \min(X_1, \ldots, X_n)$。对任意 $z > 0$，有

$$\begin{aligned} P(Z > z) &= P\big(\min(X_1, \ldots, X_n) > z\big) \\ &= P(X_1 > z, \ldots, X_n > z) \\ &= \prod_{i=1}^{n} P(X_i > z) \quad (\text{独立}) \\ &= \big(P(X > z)\big)^n \\ &= \big(1 - F(z)\big)^n \\ &= \big(e^{-z/\theta}\big)^n = e^{-nz/\theta}. \end{aligned}$$

因此

$$F_Z(z) = P(Z \le z) = 1 - P(Z > z) = 1 - e^{-nz/\theta}, \quad z > 0,$$

对 $z > 0$ 求导得密度

$$f_Z(z) = F_Z'(z) = \frac{n}{\theta} e^{-nz/\theta}, \quad z > 0.$$

这表明

$$Z \sim \mathrm{Exp}\left(\frac{\theta}{n}\right), \qquad E(Z) = \frac{\theta}{n} \ \Rightarrow\ E(nZ) = \theta.$$

三、有效性

比较参数θ的两个无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，如果在样本容量n相同的情况下，$\hat{\theta}_1$的观察值在真值θ的附近较$\hat{\theta}_2$更密集，则认为$\hat{\theta}_1$较$\hat{\theta}_2$有效.

由于方差是随机变量取值与其数学期望的偏离程度，所以无偏估计以方差小者为好.

设$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$与$\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$都是$\theta$的无偏估计量，若有 $D(\hat{\theta}_1) \le D(\hat{\theta}_2)$，则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

例5　(续例4)

试证当$n>1$时，θ的无偏估计量$\overline{X}$较nZ有效.

证明　由于 $D(X)=\theta^2$，　故有 $D(\overline{X})=\dfrac{\theta^2}{n}$，

又因为 $D(Z)=\dfrac{\theta^2}{n^2}$，　故有 $D(nZ)=\theta^2$，

当$n>1$时，　$D(nZ)>D(\overline{X})$，

故θ的无偏估计量$\overline{X}$较nZ有效.

例6 (续例3) 在例3中已证明 $\hat{\theta}_1 = 2\overline{X}$和

$\hat{\theta}_2 = \dfrac{n+1}{n}\max\{X_1, X_2, \cdots, X_n\}$ 都是 θ 的无偏估

计量, 现证当 $n \geq 2$ 时, $\hat{\theta}_2$ 较 $\hat{\theta}_1$ 有效.

证明 由于 $D(\hat{\theta}_1) = 4D(\overline{X}) = \dfrac{4}{n}D(X) = \dfrac{\theta^2}{3n}$,

$$D(\hat{\theta}_2) = D\left(\frac{n+1}{n}X_h\right) = \left(\frac{n+1}{n}\right)^2 D(X_h),$$

又因为 $E(X_h) = \dfrac{n+1}{n}\theta$,

$$E(X_h^2) = \int_0^\theta \frac{n}{\theta^n} x^{n+1} \mathrm{d}x = \frac{n}{n+2}\theta^2,$$

$$D(X_h) = E(X_h^2) - [E(X_h)]^2$$

$$= \frac{n}{(n+1)^2(n+2)}\theta^2,$$

故 $D(\hat{\theta}_2) = \dfrac{1}{n(n+2)}\theta^2,$

又$n \geq 2$, 所以$D(\hat{\theta}_2) < D(\hat{\theta}_1)$, $\hat{\theta}_2$ 较 $\hat{\theta}_1$ 有效.

四、相合性

若$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为参数$\theta$的估计量,

若对于任意$\theta\in\Theta$,当$n\to\infty$时,$\hat{\theta}(X_1,X_2,\cdots,X_n)$

依概率收敛于θ,则称$\hat{\theta}$为θ的相合估计量.

例如 由第六章第三节知, 样本$\boldsymbol{k}(\boldsymbol{k}\geq 1)$阶矩是总体$\boldsymbol{X}$的$\boldsymbol{k}$阶矩 $\mu_k=E(X^k)$的相合估计量,

进而若待估参数$\theta=g(\mu_1,\mu_2,\cdots,\mu_n)$,其中$g$为连续函数,则$\theta$的矩估计量$\hat{\theta}=g(\hat{\mu}_1,\hat{\mu}_2,\cdots,\hat{\mu}_n)=g(A_1,A_2,\cdots,A_n)$是$\theta$的相合估计量.

例7 试证：样本均值 $\overline{X}$ 是总体均值 μ 的相合估计量，样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 及样本的二阶中心矩 $B_2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 都是总体方差 σ^2 的相合估计量.

证明 由大数定律知，

$$\forall \varepsilon>0,\ \text{有}\ \lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1,$$

所以 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ 是 μ 的相合估计量.

又 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}(X_i^{\,2}-2X_i\overline{X}+\overline{X}^2)$

$$=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2}-\overline{X}^2=A_2-\overline{X}^2,$$

(A_2是样本二阶原点矩)

由大数定律知,

$$A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2\text{依概率收敛于}E(X^2),$$

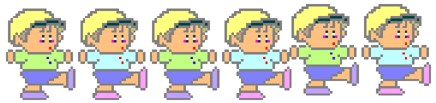
$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\text{依概率收敛于}E(X),$$

故 $B_2 = A_2 - \overline{X}^2$

依概率收敛于 $E(X^2) - [E(X)]^2 = \sigma^2$,

所以 B_2 是 σ^2 的相合估计量.

又 $\lim\limits_{n\to\infty} \dfrac{n}{n-1} = 1$,



所以 $S^2 = \dfrac{n}{n-1} B_2$ 也是 σ^2 的相合估计量.

五、小结

估计量的评选的三个标准 $\begin{cases}\text{无偏性}\\ \text{有效性}\\ \text{相合性}\end{cases}$

相合性是对估计量的一个基本要求，不具备相合性的估计量是不予以考虑的.

由最大似然估计法得到的估计量，在一定条件下也具有相合性.估计量的相合性只有当样本容量相当大时,才能显示出优越性，这在实际中往往难以做到,因此,在工程中往往使用无偏性和有效性这两个标准.

课后作业第10题

设 $X_1,X_2,\cdots,X_n$ 为来自总体 X 的一个样本，$\overline{X},S^2$ 分别为样本均值和样本方差，$E(X)=\mu, D(X)=\sigma^2$。

(1)确定常数 c，使 $c\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$ 为 σ^2 的无偏估计；

(2)确定常数 c，使 $(\overline{X})^2-cS^2$ 为 μ^2 的无偏估计。

解： (1) $E\left[c\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2\right]=c\sum_{i=1}^{n-1}\left[E\left(X_{i+1}^2\right)-2E\left(X_{i+1}\right)E\left(X_i\right)+E\left(X_i^2\right)\right]$

$=c\sum_{i=1}^{n-1}\left[\left(\sigma^2+\mu^2\right)-2\mu^2+\left(\sigma^2+\mu^2\right)\right]=2c(n-1)\sigma^2=\sigma^2$，得 $c=\dfrac{1}{2(n-1)}$。

(2) $E\left[\left(\overline{X}\right)^2-cS^2\right]=E\left(\overline{X}\right)^2-cE\left(S^2\right)=D\left(\overline{X}\right)+E^2\left(\overline{X}\right)-cE\left(S^2\right)$

$=\dfrac{\sigma^2}{n}+\mu^2-c\sigma^2=\mu^2$，得 $c=\dfrac{1}{n}$。

4. 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\dfrac{1}{2\theta}, & 0<x<\theta,\\ \dfrac{1}{2(1-\theta)}, & \theta<x<1,\\ 0, & \text{其他},\end{cases}$$

其中 $0<\theta<1$。

设 $X_1,X_2,\ldots,X_n$ 为来自总体 X 的简单随机样本，$\bar{X}$ 为样本均值。

(1) 求参数 θ 的矩估计量 $\hat{\theta}$;

(2) 判断 $4\bar{X}^2$ 是否为 θ^2 的无偏估计量，并说明理由。

解:

(1) 矩估计量:

先求数学期望

$$
\begin{aligned}
E(X) &= \int_0^\theta x \cdot \frac{1}{2\theta}\,dx + \int_\theta^1 x \cdot \frac{1}{2(1-\theta)}\,dx \\
&= \frac{1}{2\theta} \cdot \frac{\theta^2}{2} + \frac{1}{2(1-\theta)} \cdot \frac{1-\theta^2}{2} \\
&= \frac{\theta}{4} + \frac{1-\theta^2}{4(1-\theta)} = \frac{1+2\theta}{4}.
\end{aligned}
$$

令 $E(X) = \bar{X}$，得

$$
\frac{1+2\theta}{4} = \bar{X} \quad \Rightarrow \quad \hat{\theta} = \frac{4\bar{X}-1}{2} = 2\bar{X} - \frac{1}{2}.
$$

(2) 判断 $4\bar{X}^2$ 是否为 θ^2 的无偏估计量:

计算

$$
\begin{aligned}
E(X^2) &= \int_0^\theta x^2 \cdot \frac{1}{2\theta}\,dx + \int_\theta^1 x^2 \cdot \frac{1}{2(1-\theta)}\,dx \\
&= \frac{1}{2\theta} \cdot \frac{\theta^3}{3} + \frac{1}{2(1-\theta)} \cdot \frac{1-\theta^3}{3} \\
&= \frac{2\theta^2+\theta+1}{6}.
\end{aligned}
$$

因此

$$
\begin{aligned}
D(X) &= E(X^2) - \big(E(X)\big)^2 \\
&= \frac{2\theta^2+\theta+1}{6} - \left(\frac{1+2\theta}{4}\right)^2 = \frac{4\theta^2-4\theta+5}{48}.
\end{aligned}
$$

又

$$
\begin{aligned}
E(\bar{X}^2) &= D(\bar{X}) + \big(E(\bar{X})\big)^2 \\
&= \frac{D(X)}{n} + \left(\frac{1+2\theta}{4}\right)^2 \\
&= \frac{4\theta^2-4\theta+5}{48n} + \frac{(1+2\theta)^2}{16}.
\end{aligned}
$$

从而

$$
\begin{aligned}
E\big(4\bar{X}^2\big) &= 4E(\bar{X}^2) \\
&= \frac{4\theta^2-4\theta+5}{12n} + \frac{(1+2\theta)^2}{4} \neq \theta^2.
\end{aligned}
$$

因此，$4\bar{X}^2$ 不是 θ^2 的无偏估计量。