第七章

参数估计

统计推断的基本问题

参数估计问题

点估计

区间估计

假设检验问题

什么是参数估计？

参数是刻画总体某方面概率特性的数量.

当此数量未知时，从总体抽出一个样本，用某种方法对这个未知参数进行估计就是参数估计.

例如，$X \sim N(\mu, \sigma^2)$,

若μ, σ^2未知，通过构造样本的函数，给出它们的估计值或取值范围就是参数估计的内容.

点估计 **区间估计**

第一节 点估计

引入：

医院就诊人数

一个地区的男性成年人的身高

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 $E(X)$.

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

故 $E(X) = \lambda$ 的估计为 1.22 .

一、点估计问题的一般提法

设总体 X 的分布函数 $F(x;\theta)$ 的形式为已知，θ 是待估参数．$X_1, X_2, \Lambda\ , X_n$ 是 X 的一个样本，$x_1, x_2, \Lambda\ , x_n$ 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 $\hat{\theta}(X_1, X_2, \Lambda\ , X_n)$，用它的观察值 $\hat{\theta}(x_1, x_2, \Lambda\ , x_n)$ 来估计未知参数 θ.

$$\left.\begin{array}{l} \hat{\theta}(X_1, X_2, \Lambda\ , X_n)\text{称为}\theta\text{的估计量.} \\ \hat{\theta}(x_1, x_2, \Lambda\ , x_n)\text{称为}\theta\text{的估计值.} \end{array}\right\} \begin{array}{l}\text{通称估计，} \\ \text{简记为}\hat{\theta}. \end{array}$$

例2 在某纺织厂细纱机上的断头次数 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, 现检查了150只纱锭在某一时间段内断头的次数, 数据如下, 试估计参数 λ.

断头次数 k	0	1	2	3	4	5	6	
断头 k 次的纱锭数 n_k	45	60	32	9	2	1	1	150

解 先确定一个统计量 $\overline{X}$, 再计算出 $\overline{X}$ 的观察值 $\overline{x}$, 把 $\overline{x}$ 作为参数 λ 的估计值.

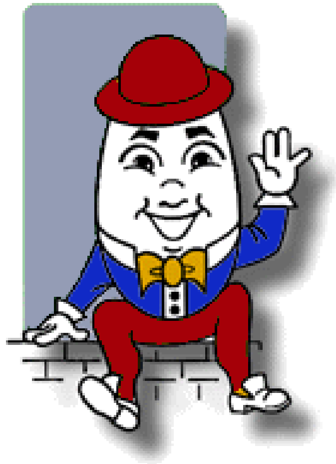
$\overline{x} = 1.133$. λ 的估计值为 1.133.

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 $f(x;\theta_1,\theta_2,\Lambda\ ,\theta_k)$, 或 X 为离散型随机变量,其分布律为 $P\{X=x\}=p(x;\theta_1,\theta_2,\Lambda\ ,\theta_k)$,其中 $\theta_1,\theta_2,\Lambda\ ,\theta_k$ 为待估参数,

若 $X_1,X_2,\Lambda\ ,X_n$ 为来自 X 的样本,

假设总体 X 的前 k 阶矩存在,

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

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

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

其中 R_X 是 x 可能取值的范围，$l = 1,2,\Lambda\ ,k$

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

总体矩 $\mu_l\ (l = 1, 2,\Lambda\ , k)$,

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.

点估计的思想方法

设总体X 的分布函数的形式已知, 但含有一个或多个未知参数：$\theta_1, \theta_2, \ldots, \theta_k$

设 $X_1, X_2, \ldots, X_n$为总体的一个样本

构造 k 个统计量：

$$\left.\begin{array}{l} \theta_1(X_1, X_2, \Lambda, X_n) \\ \theta_2(X_1, X_2, \Lambda, X_n) \\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda \\ \theta_k(X_1, X_2, \Lambda, X_n) \end{array}\right\} \text{随机变量}$$

当测得样本值$(x_1, x_2,\ldots, x_n)$时,代入上述统计量，即可得到 k 个数：

$$\left.\begin{array}{l}\hat{\theta}_1(x_1, x_2, \Lambda\ , x_n) \\ \hat{\theta}_2(x_1, x_2, \Lambda\ , x_n) \\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda\ \Lambda \\ \hat{\theta}_k(x_1, x_2, \Lambda\ , x_n)\end{array}\right\}\text{数 值}$$

称数 $\theta_1 \quad , \hat{\theta}_k$为未知参数 $\theta_1, \quad ,\theta_k$ 的估计值

对应统计量 为未知参数 $\theta_1, \quad ,\theta_k$ 的估计量

问题 $\left\{\begin{array}{l}\text{如何构造统计量？} \\ \text{如何评价估计量的好坏？}\end{array}\right.$

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.

矩估计原则

- 用样本均值估计总体均值$E(X)$，$\hat{E}(X)=\bar{X}$
- 用样本方差估计总体方差$\mathrm{Var}(X)$，$\hat{\mathrm{V}}\mathrm{ar}(X)=S_n^2$
- 用样本的 p 分位数估计总体的 p 分位数，
- 用样本中位数估计总体中位数。

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

$\hat{\theta}_{矩} = 2\overline{x}$ 为所求 θ 的估计值.

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

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

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

2. 最大似然估计法

思想方法：一次试验就出现的
事件有较大的概率

例如: 有两外形相同的箱子,各装100个球

一箱	99个白球	1 个红球
一箱	1 个白球	99个红球

现从两箱中任取一箱, 并从箱中任取一球, 结果所取得的球是白球.

问：所取的球来自哪一箱？ 答：第一箱.

(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, Λ, X_n是来自总体 X 的样本,

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

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

则样本 X_1, X_2, Λ, X_n 取到观察值 x_1, x_2, Λ, x_n 的概率,

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

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

$L(\theta)$称为样本似然函数.

最大似然估计法

得到样本值 x_1, x_2, Λ, 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 $L(x_1, x_2, \Lambda, x_n; \hat{\theta}) = \max\limits_{\theta \in \Theta} L(x_1, x_2, \Lambda, x_n; \theta).$

(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, Λ, x_n 有关,记为

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

$\boxed{\hat{\theta}(X_1, X_2, \Lambda, X_n)}$ 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 $f(x;\theta), \theta$ 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.

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

$L(\theta)$称为样本的似然函数.

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

$\boxed{\hat{\theta}(x_1,x_2,\cdots,x_n)}$ 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.

费舍尔

求最大似然估计量的步骤:

(一) 写出似然函数

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

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

(二) 取对数

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

(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, Λ, 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, Λ, x_n为相应于样本X_1, X_2, Λ, 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 $\dfrac{\mathrm{d}}{\mathrm{d}p}\ln L(p)=\dfrac{\sum_{i=1}^{n}x_i}{p}-\dfrac{n-\sum_{i=1}^{n}x_i}{1-p}=0,$

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

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

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，$X_1,X_2,\Lambda\ ,X_n$ 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

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

所以 λ 的似然函数为

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

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

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

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

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.

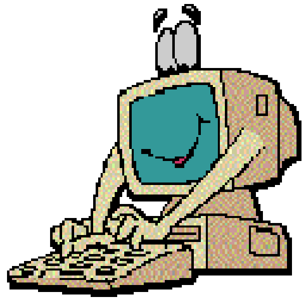
例10 设总体 X 在 $[a,b]$ 上服从均匀分布,其中 a, b 未知, x_1, x_2, Λ, x_n 是来自总体 X 的一个样本值, 求 a,b 的最大似然估计量.

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

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

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$a \le x_1, x_2, \Lambda, x_n \le b$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

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

典型例题

例1.

设总体X的概率密度为：$f(x)=\begin{cases}\dfrac{6x(\theta-x)}{\theta^3} & 0<x<\theta \\ 0 & \text{其它}\end{cases}$.

$X_1, X_2, \cdots, X_n$是取自总体X的样本。

求：θ的矩估计量。

典型例题

例2. 设总体X具有分布律

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中，$\theta(0<\theta<1)$是未知参数，已知取得样本值$x_1=1, x_2=2, x_3=1$，试求θ的矩估计值和最大似然估计值。

典型例题

例3. 设总体X的密度函数为：

$$f(x,a)=\begin{cases}(a+1)x^{a} & 0<x<1\\ 0 & \text{其它}\end{cases}$$

其中，$a>-1$,是未知参数，$X=(X_1,X_2,\cdots,X_n)$是其样本，试求a的矩估计量和最大似然估计量。

现得样本值0.1，0.2，0.8，0.9，0.7，0.7。

求参数a的估计值。

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

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

费舍尔资料

Ronald Aylmer Fisher

Born: 17 Feb. 1890 in London, England

Died: 29 Jul. 1962 in Adelaide, Australia

第三节　估计量的评选标准

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, 矩估计、极大似然估计、**Bayes**估计等. 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 $X_1, X_2, \Lambda\ , X_n$ 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 $\hat{\theta} = \theta(X_1, X_2, \Lambda\ , X_n)$ 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 $\mu_k = E(X^k)\ (k \ge 1)$ 存在，又设 X_1, X_2, Λ, 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.

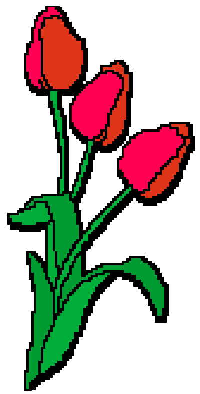
证 因为 X_1, X_2, Λ, X_n 与 X 同分布，

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

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

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（略）设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，试求 σ^2 的无偏估计量.

解 由定理二知 $$\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1),$$

$$E\left(\frac{\sqrt{n-1}S}{\sigma}\right) = \int_0^{+\infty} \sqrt{x}\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \mathrm{e}^{-\frac{x}{2}} x^{\frac{n-1}{2}-1} \mathrm{d}x$$

$$= \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \int_0^{+\infty} \mathrm{e}^{-\frac{x}{2}} x^{\frac{n}{2}-1} \mathrm{d}x = \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

$$E(S)=\sqrt{\frac{2}{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\sigma,$$

故 S 不是 σ 的无偏估计量，

$$\sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}S \text{ 是 } \sigma \text{ 的无偏估计量.}$$

例4 设总体 X 在 $[0,\theta]$ 上服从均匀分布，参数 $\theta>0$，X_1,X_2,Λ,X_n 是来自总体 X 的样本，试证明 $2\overline{X}$ 和 $\dfrac{n}{n+1}\max(X_1,X_2,\Lambda,X_n)$ 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 $X_h=\max(X_1,X_2,\Lambda,X_n)$ 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 $\frac{n}{n+1}\max(X_1,X_2,\Lambda\ ,X_n)$ 也是 θ 的无偏估计量.

例5 设总体 X 服从参数为 θ 的指数分布, 概率密度

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

其中参数 $\theta>0$, 又设 $X_1,X_2,\Lambda\ ,X_n$ 是来自总体 X 的样本, 试证 $\overline{X}$ 和 $nZ=n[\min(X_1,X_2,\Lambda\ ,X_n)]$ 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 $Z=\min(X_1,X_2,\Lambda\ ,X_n)$ 服从参数为 $\dfrac{\theta}{n}$ 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 $E(Z)=\dfrac{\theta}{n},\quad E(nZ)=\theta,$

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.

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6 (续例5)

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7 (续例4) 在例4中已证明 $\hat{\theta}_1 = 2\overline{X}$

和 $\hat{\theta}_2 = \dfrac{n+1}{n}\max\{X_1, X_2, \Lambda\ , X_n\}$ 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（一致性）

若$\hat{\theta} = \hat{\theta}(X_1, X_2, \Lambda, X_n)$为参数$\theta$的估计量，若对于任意$\theta \in \Theta$，当$n \to \infty$时，$\hat{\theta}(X_1, X_2, \Lambda, X_n)$依概率收敛于$\theta$，则称$\hat{\theta}$为$\theta$的相合估计量.

例如 由第六章第二节知，样本$k(k \geq 1)$阶矩是总体X的k阶矩$\mu_k = E(X^k)$的相合估计量，进而若待估参数$\theta = g(\mu_1, \mu_2, \Lambda, \mu_n)$，其中$g$为连续函数，则$\theta$的矩估计量$\hat{\theta} = g(\hat{\mu}_1, \hat{\mu}_2, \Lambda, \hat{\mu}_n) = g(A_1, A_2, \Lambda, A_n)$是$\theta$的相合估计量.

例8 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,

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

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,

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

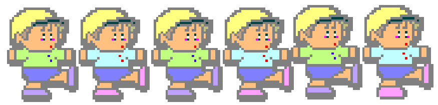
$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\text{依概率收敛于}E(X),$$

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.