第六章 参数估计

参数估计的目的： 用样本观察值估计总体的某些数字特征. 如：数学期望 $E(X)$，方差 $D(X)$ 等等.

内容： 1. 估计法

- 点估计
  - 矩估计法
  - 最大似然估计法
- 区间估计

下页 返回

第六章 参数估计

内容： 2. 估计量的评选标准

- 无偏性
- 有效性
- 一致性(或相合性)

第六章 参数估计

第一节 参数的点估计

- 一、点估计问题的提法
- 二、估计量的求法
- 三、内容小结

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**.

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. $X_1, X_2, \cdots, X_n$ 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 θ.

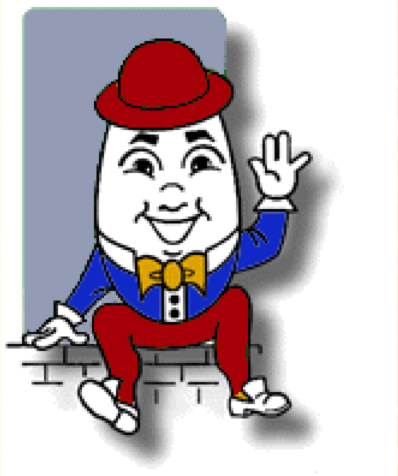
$$\left.\begin{array}{l}\hat{\theta}(X_1, X_2, \cdots, X_n)\text{称为}\theta\text{的估计量.}\\ \hat{\theta}(x_1, x_2, \cdots, x_n)\text{称为}\theta\text{的估计值.}\end{array}\right\}\begin{array}{l}\text{通称估计,}\\ \text{简记为}\hat{\theta}.\end{array}$$

二、估计量的求法

如何求得参数θ的估计量便是问题的关键所在.

常用构造估计量的方法: (两种)

1. 矩估计法
2. 最(极)大似然估计法.



1. 矩估计法

它是基于一种简单的“替换”思想建立起来的一种估计方法 .

是英国统计学家*K*.皮尔逊最早提出的 .

基本思想：用样本矩估计总体矩 .

理论依据： 大数定律

或格列汶科定理

目录 上页 下页 返回 结束

记总体k阶原点矩为 $\alpha_k = E(X^k)$

样本k阶原点矩为 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

记总体k阶中心矩为 $\mu_k = E[X - E(X)]^k$

样本k阶中心矩为 $B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k$

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.

设总体 X 的分布函数为 $F(x;\underbrace{\theta_1,\theta_2,\cdots,\theta_m}_{})$

m个待估参数 (未知)

$\alpha_k = E(X^k)$存在$(k=1,2,\cdots,m)$,

$(X_1,X_2,\cdots,X_n)$ 为来自总体X的简单随机样本.

矩估计法的具体步骤:

1° 求出 $\alpha_k = E(X^k) = \alpha(\theta_1,\theta_2,\cdots,\theta_m)$, $\quad k=1,2,\cdots,m$;

2° 要求$\alpha_k = A_k$, $\quad k=1,2,\cdots,m$

这是一个包含 m个未知参数$\theta_1,\theta_2,\cdots,\theta_m$的方程组.

3° 解出其中$\theta_1,\theta_2,\cdots,\theta_m$, 用$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m$表示.

4° 用方程组的解$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m$ 分别作为 $\theta_1,\theta_2,\cdots,\theta_m$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.

注 方程组

$$\alpha_k = A_k, \quad k = 1,2,\cdots,m$$

方程中的个数= 待估参数的个数.

例2　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　$\because \quad \alpha_1 = E(X) = \dfrac{\theta}{2}$,

根据矩估计法，令　$\alpha_1 = A_1$,

即　$$\frac{\theta}{2} = \alpha_1 = A_1 = \overline{X},$$

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

例3 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的矩估计量.

解 1° $\alpha_1 = E(X) = \dfrac{a+b}{2},$

$$\alpha_2 = E(X^2) = D(X)+[E(X)]^2$$

$$= \frac{(a-b)^2}{12} + \frac{(a+b)^2}{4},$$

2° 要求: $\begin{cases} \alpha_1 = A_1 = \overline{X} \\ \alpha_2 = A_2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \end{cases}$

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

即 $$\begin{cases} \dfrac{a+b}{2} = \overline{X}, \\ \dfrac{(a-b)^2}{12} + \dfrac{(a+b)^2}{4} = A_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2, \end{cases}$$

亦即 $$\begin{cases} a+b = 2\overline{X} \\ b-a = \sqrt{12(A_2 - \overline{X}^2)} = 2\sqrt{3}\,S_n \end{cases}$$

解方程组得到a, b的矩估计量分别为

$$\hat{a} = \overline{X} - \sqrt{3}\,S_n$$

$$\hat{b} = \overline{X} + \sqrt{3}S_n$$

小结：矩法的优点：简单易行，并不需要事先知道总体是什么分布．

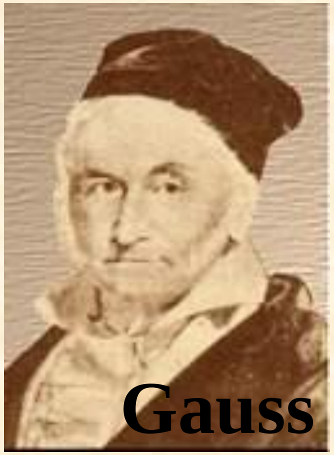
缺点：当总体类型已知时，没有充分利用分布提供的信息．一般场合下，矩估计量不唯一．

其主要原因在于建立矩法方程时，选取那些总体矩用相应样本矩代替带有一定的随意性．

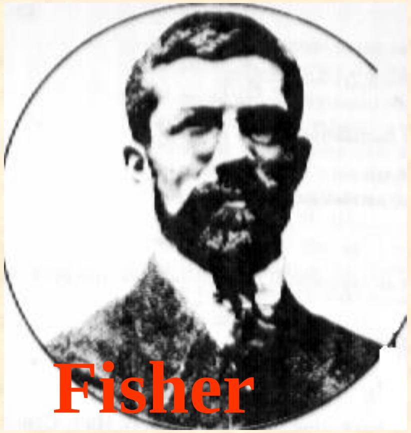
2. 最大似然估计法

是在总体类型已知条件下使用的一种参数估计方法 .

它首先是由德国数学家高斯在1821年提出的 ,



然而，这个方法常归功于英国统计学家费歇 .



费歇在1922年重新发现了这一方法，并首先研究了这种方法的一些性质 .

(1) 最大似然法的基本思想

先看一个简单例子：

某位同学与一位猎人一起外出打猎 .

一只野兔从前方窜过 .

只听一声枪响，野兔应声倒下 .

如果要你推测，

是谁打中的呢 ?

你会如何想呢?

你就会想，只发一枪便打中,猎人命中的概率一般大于这位同学命中的概率. 看来这一枪是猎人射中的 .

这个例子所作的推断已经体现了最大似然法的基本思想 .

下面我们再看一个例子,进一步体会最大似然法的基本思想 .

例6 设 $X \sim B(1, p)$, p未知. 设想我们事先知道 p 只有两种可能:

$$p=0.7 \quad 或 \quad p=0.3$$

如今重复试验3次,得结果: 0 , 0, 0

问: 应如何估计p?

由概率论的知识, 3次试验中出现“1”的次数

$$Y \sim B(3, p)$$

$$P\{Y = k\} = C_3^k p^k (1-p)^{n-k} \quad (k=0, 1, 2, 3)$$

$$P\{Y=k\}=C_3^k p^k(1-p)^{n-k} \quad (k=0, 1, 2, 3)$$

Y		0	1	2	3
$P\{Y=k\}$	$p=0.7$时	0.027	0.189	0.441	0.343
	$p=0.3$时	0.343	0.441	0.189	0.027

依题设，“重复试验3次, 得结果： 0， 0， 0”

即事件$\{Y=0\}$发生

$\because \quad P\{Y=0; p=0.3\}>P\{Y=0; p=0.7\}$

$\therefore$ 选 $\hat{p}=0.3$ 作为p的估计.

一般地，(1) 如果有$p_1, p_2, \ldots, p_m$可供选择, 又如何合理地选 p 呢?

若重复进行试验n次,结果“1”出现k次 $(0 \le k \le n)$, 我们计算一切可能的

$$P\{Y=k; p_i\}= Q_i, \quad i=1, 2, \ldots, m$$

从中选取使Q_i最大的p_i作为p的估计.

比方说, 当 $p = p_{i_0}$ 时Q_i最大,

$$P\{Y = k; p_{i_0}\} \ge P\{Y = k; p_i\} \quad i=1, 2, \ldots, m$$

则估计参数p为 $\hat{p} = p_{i_0}$.

(2) 如果只知道$0<p<1$，并且实测记录是 $Y=k\ (0\le k\le n)$，又应如何估计p呢?

注意到:

$$P\{Y=k;p\}=C_n^k p^k(1-p)^{n-k}=f(p)$$

是 p 的函数, 可用求导的方法找到使 $f(p)$达到极大值的p .

但因 $f(p)$ 与 $\ln f(p)$ 达到极大值的自变量相同,故问题可转化为求$\ln f(p)$的极大值点 .

$$\ln f(p) = \ln C_n^k + k\ln p + (n-k)\ln(1-p)$$

将 $\ln f(p)$ 对p求导并令其为 0,

$$\frac{\mathrm{d}\ln f(p)}{\mathrm{d}p} = \frac{k}{p} - \frac{n-k}{1-p} = 0$$

便得 $p(n-k) = k(1-p)$

参数 p的估计值

从中解得 $\hat{p} = \dfrac{k}{n}$

这时, 对一切 $0 < p < 1$, 均有

$$P\{Y = k; \hat{p}\} \geq P\{Y = k; p\}$$

目录 上页 下页 返回 结束

综上所述：

设某试验的可能结果为：

$$A_1,\ A_2,\ \cdots;\ A_i\ ,\ \cdots$$

若在一次试验中，某结果 A_i 出现，则应选择参数使A_i出现的概率最大.

以上这种选择一个参数使得实验结果具有最大概率的思想就是最大似然法的基本思想 .

(2) 似然函数

定义6.1 设总体X的分布密度(或分布律)为 $p(x;\theta)$，其中$\theta=(\theta_1,\theta_2,\cdots,\theta_m)$为未知参数. 又设$(x_1,x_2,\cdots,x_n)$ 为自总体X的样本$(X_1,X_2,\ldots,X_n)$ 的一个观察值，则称样本的联合分布

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

为似然函数.

(3) 最大似然估计量(值)

定义 若对于任意给定的样本值$(x_1, x_2, \cdots, x_n)$,

存在 $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m)$，使

$$L(\hat{\theta}) = \max L(\theta)$$

则称 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$分别为$\theta_1, \theta_2, \cdots, \theta_m$的

最大似然估计值 (MLE).

maximum likelihood estimate

注 1°对于给定的样本值$(x_1, x_2, \cdots, x_n)$，在最大似然估计值处，$L(\theta)$ 取得最大值;

2° $\because$ $L(\theta)$与样本值$(x_1, x_2, \cdots, x_n)$有关，

$$\therefore \quad \hat{\theta}_k = \hat{\theta}_k(x_1, x_2, \cdots, x_n) \quad (k = 1,2,\cdots,m)$$

将 $\hat{\theta}_k$中的 $x_1, x_2, \cdots, x_n$换成 $X_1, X_2, \cdots, X_n$，则称 $\hat{\theta}_k(X_1, X_2, \cdots, X_n)$为$\theta_k$的最大似然估计量.

$$(k = 1,2,\cdots,m)$$

(4) 求最大似然估计(MLE)的步骤:

$$\boxed{\theta = (\theta_1, \theta_2, \cdots, \theta_m)}$$

1° 写出似然函数

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

2° 取对数 $\ln L(\theta) = \sum_{i=1}^{n} \ln p(x_i; \theta)$

3° 解似然方程(组)

$$\begin{cases} \dfrac{\partial \ln L}{\partial \theta_1} = 0 \\ \dfrac{\partial \ln L}{\partial \theta_2} = 0 \\ \quad\vdots \\ \dfrac{\partial \ln L}{\partial \theta_m} = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial \ln L}{\partial \theta_1} = 0 \\ \dfrac{\partial \ln L}{\partial \theta_2} = 0 \\ \quad\vdots \\ \dfrac{\partial \ln L}{\partial \theta_m} = 0 \end{cases}$$

所求得的解 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$ 为最大似然估计值.

4° 将 $\hat{\theta}_k$ 中的 $x_1, x_2, \cdots, x_n$ 替换成 $X_1, X_2, \cdots, X_n$,

便得到 θ_k 的最大似然估计量 $\hat{\theta}_k(X_1, X_2, \cdots, X_n)$.

注 1°上述求最大似然估计的方法，要求$\ln L$可微，若不满足此条件，则须从定义出发最大似然估计.

2°似然估计方程组与最大似然估计之间没有必然的关系.

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

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

其中 $x_1, x_2, \cdots, x_n$ 在集合$\{0,1\}$中取值.

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

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 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

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

X 的似然函数为

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

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

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

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 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 $x_{(1)} = \min(x_1, x_2, \cdots, x_n)$,

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

1° X 的概率密度为

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

2° 似然函数

$\because\ a \le x_1, x_2, \cdots, x_n \le b$

等价于 $a \le x_{(1)}, x_{(n)} \le b$

$$L(a,b)=\prod_{i=1}^{n} p(x_i;a,b)$$

$$=\begin{cases} \dfrac{1}{(b-a)^n}, & a \le x_i \le b \ \ (i=1,2,\cdots n) \\ 0, & 其它 \end{cases}$$

$$=\begin{cases} \dfrac{1}{(b-a)^n}, & a \le x_{(1)},\ x_{(n)} \le b \\ 0, & 其它 \end{cases}$$

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

分析 $\because$
$$\frac{\partial \ln L}{\partial b} = -\frac{n}{b-a} \ne 0$$

$$\frac{\partial \ln L}{\partial a} = \frac{n}{b-a} \ne 0$$

$\therefore$ 不可用微分法求$\hat{a}, \hat{b}$.

3° 从定义出发求$\hat{a},\hat{b}$.

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

于是对于满足条件 $a\le x_{(1)},\ b\ge x_{(n)}$的任意$a,b$有

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

即似然函数$L(a,b)$在$a=x_{(1)},\ b=x_{(n)}$时， 取得最大值$[x_{(n)}-x_{(1)}]^{-n}$,

$\therefore$ a, b的最大似然估计值

$$\hat{a} = x_{(1)} = \min_{1 \le i \le n} x_i,$$

$$\hat{b} = x_{(n)} = \max_{1 \le i \le n} x_i,$$

4° a, b的最大似然估计量

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

三、内容小结

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

费希尔资料

Ronald Aylmer Fisher

Born: 17 Feb 1890 in London, England
Died: 29 July 1962 in Adelaide, Australia