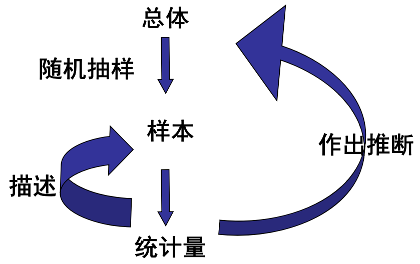
§7 参数估计



研究统计量的性质和评价一个统计推断的优良，完全取决于其抽样分布的性质.

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的概率特性的数量.

当这个数量是未知的时候，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.

在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数.

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

参数估计的类型

点估计(point Estimation) —— 估计未知参数的值

区间估计(interval Estimation)—— 估计未知参数的取值范围，使得这个范围包含未知参数真值的概率为给定的值.

§7.1　点估计

一、点估计问题的一般提法

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：

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

并建立k个方程。

当测得样本值$(x_1, x_2,\dots, x_n)$时,代入上述方程组,即可得到 k 个数:

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

称数$\hat{\theta}_1\cdots,\hat{\theta}_k$为未知参数$\theta_1,\cdots,\theta_k$的**估计值**
对应统计量为未知参数$\theta_1,\cdots,\theta_k$ 的**估计量**
} 通称估计，简记为 $\hat{\theta}$.

例1 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.

（由大数定律, $\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\mu|<\varepsilon\}=1$ ）

$\overline{x}=1.133$. 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,如何构造统计量、如何评价估计量的好坏是关键问题.

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


K.皮尔逊
英国统计学家

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

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

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

其中R_X是x可能取值的范围, $l=1,2,\cdots,k$

若总体X的 数学期望 $E(X)=\mu$ 存在

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}E(X)=\mu$$
$$\Downarrow$$
$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}E(X^k)=\mu_k\ (k=1,2,\cdots)$$
$$g(A_1,A_2,\cdots,A_k)\xrightarrow{P}g(\mu_1,\mu_2,\cdots,\mu_k)$$

其中 g 为连续函数.

这表明，当样本容量很大时，在统计上，可以用样本矩去估计总体矩.这一事实导出矩估计法。

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.

理论依据: 大数定律

矩估计法的具体做法:

设总体的分布函数中含有k个未知参数 $\theta_1,\theta_2,\cdots,\theta_k$,那么它的前$k$阶矩 $\mu_1,\mu_2,\cdots,\mu_k$，一般都是这 k 个参数的函数,记为:

$$\mu_i = \mu_i(\theta_1,\theta_2,\cdots,\theta_k) \qquad i=1,2,\dots,k$$

从这 k 个方程中解出

$$\theta_j = \theta_j(\mu_1,\mu_2,\cdots,\mu_k) \qquad j=1,2,\dots,k$$

那么用诸 μ_i 的估计量 A_i 分别代替上式中的诸 μ_i ，

即可得诸 θ_j 的矩估计量：

$$\hat{\theta}_j = \theta_j(A_1,A_2,\cdots,A_k) \qquad j=1,2,\dots,k$$

矩估计量的观察值称为矩估计值 .

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

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

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

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

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

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

例3 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 $\hat{\mu} = A_1 = \overline{X}$,

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

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

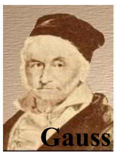
用样本二阶中心矩 $B_2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 作为总体 X 的方差的矩估计.

❑ 最大似然估计法(maximum likelihood)

2. 最大似然估计法

它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舍儿 .



费舍儿在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 .

思想方法：一次试验就出现的事件有较大的概率

例如: 有两个外形相同的箱子,都装有100个球

一号箱　99个白球，　1个红球

一号箱　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：极有可能是第一箱.

基本思想:

如：甲.乙两人比较射击技术,分别射击目标一次,甲中而乙未中, 则可以认为:甲射击技术优于乙射击技术.

实际问题(医生看病、公安人员破案、技术人员进行质量检验等)尽管千差万别,但他们具有一个共同的规律,即在获得了观察资料之后,给参数选取一个数值, 使得前面的观察结果出现的可能性最大.

最大似然估计就是通过样本值$x_1,\cdots,x_n$来求得总体的分布参数,使得 $X_1,\cdots,X_n$ 取值为$x_1,\cdots,x_n$的概率最大.

似然函数的定义

设总体X的概率密度为$f(x;\theta)$(或分布律 $P\{X=k\}=p(x;\theta)$), 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 $X_1,X_2,\cdots,X_n$ 的联合密度为 $\prod_{i=1}^{n} f(x_i;\theta)$

(或分布律为 $\prod_{i=1}^{n} p(x_i;\theta)$).

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.

离散型：

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

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

连续型：

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$dx_1,dx_2,\cdots,dx_n$的n维立方体)内的概率近似地为 $\prod_{i=1}^{n} f(x_i;\theta)dx_i$，

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

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

(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

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

$\hat{\theta}(X_1,X_2,\cdots,X_n)$ 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

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

(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.

$$L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$$

若 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$ 关于 $\theta_1,\ldots,\theta_k$ 可微,则

$$\boxed{\frac{\partial}{\partial\theta_i}\ln L=0,\quad i=1,2,\cdots,k.}$$ 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4 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 $\dfrac{\mathrm{d}}{\mathrm{d}p}\ln L(p)=\dfrac{\sum_{i=1}^{n}x_i}{p}-\dfrac{n-\sum_{i=1}^{n}x_i}{1-p}=0,$

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

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

这一估计量与矩估计量是相同的.

例5 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 $\dfrac{1}{\sigma^2}\left[\sum_{i=1}^{n}x_i-n\mu\right]=0$ 解得 $\hat{\mu}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},$

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

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

例6 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

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

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

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

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

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

如 在正态总体$N(\mu,\sigma^2)$中, σ^2的极大似然估计值为

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

$\sigma=\sqrt{\sigma^2}$ 是σ^2的单值函数, 且具有单值反函数,

故σ的极大似然估计值为

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

$\lg\sigma$的极大似然估计值为

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

矩估计就不具有这个性质.

例如 设 X 的密度函数为

$$f(x)=\begin{cases}\sqrt{\dfrac{2}{\pi}}\dfrac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} & x>0\\ 0 & x\le 0\end{cases}\qquad E(X)=\sqrt{\frac{2}{\pi}}\sigma,\ D(X)=\left(1-\frac{2}{\pi}\right)\sigma^2$$

$X_1,X_2,\ldots,X_n$为总体的样本

由矩法，令 $E(X)=\sqrt{\dfrac{2}{\pi}}\sigma=\bar X$

$$E(X^2)=D(X)+E^2(X)=\sigma^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$$

得σ与σ^2的矩法估计量为 $\hat\sigma=\sqrt{\dfrac{\pi}{2}}\bar X$

不具有不变性—— $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2\neq(\hat\sigma)^2$

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

37