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

引言

统计推断问题分为两大类:

(1)参数估计 $\begin{cases} \text{点估计} \\ \\ \text{区间估计} \end{cases}$

(2) 假设检验问题

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 $E(X)$.

$$\bar{x}=\frac{\sum_{k=0}^{6}kn_k}{\sum_{k=0}^{6}n_k}=\frac{1}{250}(0\times75+1\times90+2\times54+3\times22+4\times6+5\times2+6\times1)=1.22.$$

故 $E(X)=\lambda$ 的估计为 1.22.

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.

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

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

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

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 $\mu_l \ (l = 1, 2, \cdots, k)$,

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.

所以可以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.

矩估计法的一般步骤：

1. 求出总体的矩：$\mu_l = E(X^l),(l=1,2...k)$

2.列矩估计方程组：

$$\mu_l = A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l,(l=1,2...k)$$

3.解方程组得：$\hat{\theta}_l = \theta_l(X_1, X_2....X_n),(l=1,2...k)$

即：θ_l 的矩估计量是

$$\hat{\theta}_l = \theta_l(X_1, X_2....X_n),(l=1,2...k)$$

4. $\hat{\theta}_l(l=1,2...k)$ 的观察值$\hat{\theta}_l = \theta_l(x_1, x_2....x_n)$称为 θ_l 的矩估计值。

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 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

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

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

例4 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 $\mu_1 = E(X) = \mu,$

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

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

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

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

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,

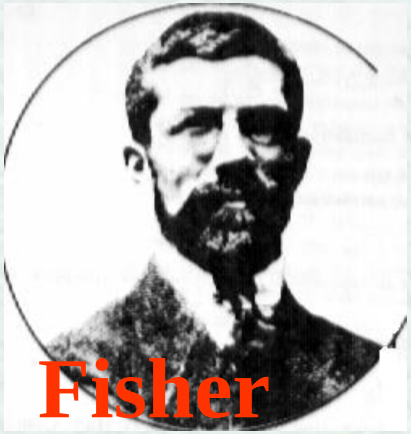
用样本二阶中心矩 $B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$作为总体

X的方差的矩估计.

二、 最大似然估计法

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舍尔.



费舍尔在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.

最大似然估计法的基本思想：

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. 故一般猜测这一枪是猎人命中的 .

这个例子所作的推断已经体现了最大似然估计法的基本思想。

也就是在已经得到试验结果的情况下，寻找这个结果出现可能性最大的那个值 θ 作为未知参数的估计。

似然函数的定义

(1) 设总体 X 属离散型

设分布律 $P\{X=x\}=p(x;\theta)$, 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

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

$L(\theta)$称为样本的似然函数.

(2) 设总体 X 属连续型

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,

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. 定义样本的似然函数为

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

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.

求最大似然估计量的步骤:

(一) 写出似然函数

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

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

(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$\hat{\theta}$.

若方程无解，则只能按最大似然估计的基本思想来求出最大值点。

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 θ_i $(i = 1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

例5　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

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

所以 λ 的似然函数为

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

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

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

解得λ的最大似然估计值 $\lambda = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = \overline{x},$

λ的最大似然估计量为 $\hat{\lambda} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i = \overline{X}.$

这一估计量与矩估计量是相同的.

例6 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 $f(x;\mu,\sigma^2)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,

X 的似然函数为

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

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

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

令

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

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

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.

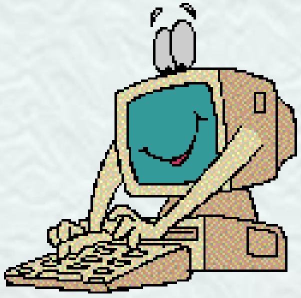
例7 设总体 X 在 $[a,b]$ 上服从均匀分布,其中 a, b 未知, $x_1, x_2, \cdots, x_n$ 是来自总体 X 的一个样本值,求 a,b 的最大似然估计量.

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（最大似然估计的不变性）

设 θ 的函数 $u=u(\theta),\theta\in\Theta$ 具有单值反函数 $\theta=\theta(u),u\in$. 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.

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6中, σ^2 的最大似然估计值为 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$,

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

最小二乘法与最大似然估计

https://www.zhihu.com/question/20447622

https://blog.csdn.net/zouxy09/article/details/8537620

最大似然估计到期望最大算法

最大似然估计：观察到的数据默认是来自同一个分布中的。
如果观察到的数据不是来自同一个分布怎么办？

EM的意思是“Expectation Maximization”，在我们上面这个问题里，
我们是先随便猜一下男生（身高）的正态分布的参数：
如均值和方差是多少（当然了，刚开始肯定没那么准），
然后计算出每个人更可能属于第一个还是第二个正态分布中的
这个是属于Expectation一步。

有了每个人的归属，也就是说将这200个人分为男生和女生两部分，
我们就可以根据之前的最大似然估计，通过大概分为男生的n个人来
重新估计男生分布的参数，再重新估计女生的分布参数。
这个是Maximization。（对应最大似然估计）