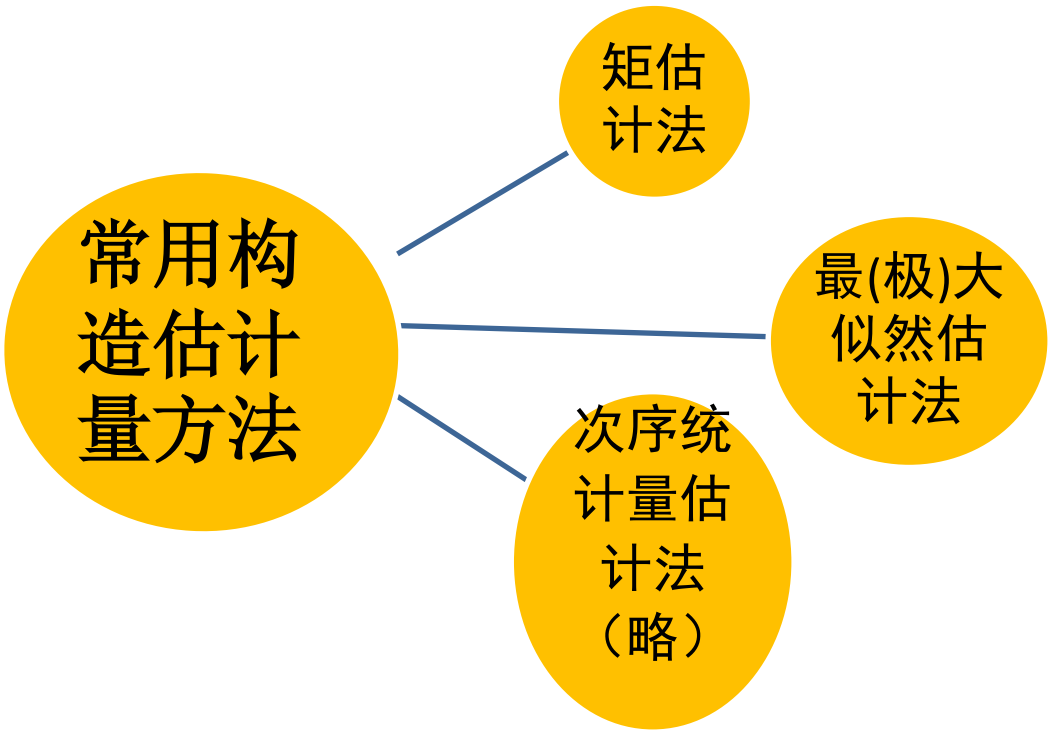
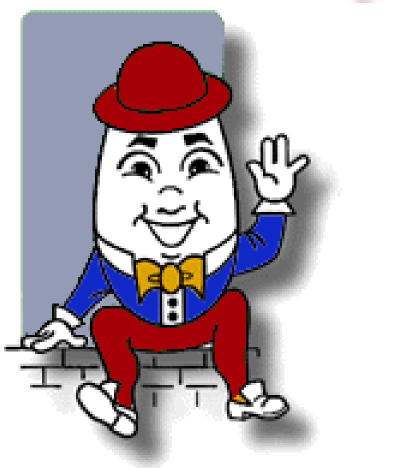
2.2 点估计量的求法

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, 因此如何求得参数θ的估计量便是问题的关键所在.





一、矩估计法

1. 矩估计法

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

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 $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$

用样本矩来估计总体矩，这种估计法称为矩估计法.

设总体 X 的分布函数为 $F(x;\underbrace{\theta_1,\theta_2,\cdots,\theta_m}_{m\text{个待估参数(未知)}})$

$\alpha_k=E(X^k)$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 若$\hat{\theta}_k$是θ_k的矩估计，$g(\theta)$为连续函数，则也称$g(\hat{\theta}_k)$是$g(\theta_k)$的矩估计.

即用样本矩的函数来估计总体矩的函数。

例1 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的矩估计量.

证 因为 $\alpha_1 = E(X) = \dfrac{\theta}{2}$,

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

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

例2 设总体 X 在$[\theta_1,\theta_2]$上服从均匀分布,其中θ_1,θ_2未知,$(X_1, X_2,\cdots, X_n)$是来自总体X的样本,求θ_1,θ_2的矩估计量.

证

$$\alpha_1 = E(X) = \frac{\theta_1+\theta_2}{2},$$

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

$$= \frac{(\theta_2-\theta_1)^2}{12} + \frac{(\theta_1+\theta_2)^2}{4},$$

令 $$\frac{\theta_1+\theta_2}{2} = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\frac{(\theta_2-\theta_1)^2}{12} + \frac{(\theta_1+\theta_2)^2}{4} = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

即$\begin{cases} \theta_1+\theta_2=2A_1, \\ \theta_2-\theta_1=\sqrt{12(A_2-A_1^2)}, \end{cases}$

解方程组得到 θ_2,θ_1 的矩估计量分别为

$$\hat{\theta}_1=A_1-\sqrt{3(A_2-A_1^2)}=\overline{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

$$\hat{\theta}_2=A_1+\sqrt{3(A_2-A_1^2)}=\overline{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

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

$$\alpha_1 = E(X) = \mu,$$

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

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$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$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.

例4 设总体X的分布密度为

$$p(x;\theta)=\frac{1}{2\theta}\mathrm{e}^{-\frac{|x|}{\theta}}\quad (x\in R,\ \theta>0)$$

$(X_1,X_2,\cdots,X_n)$为来自总体X的样本. 求参数θ的矩估计量.

分析: $p(x;\theta)$中只含有一个未知参数θ, 一般地,

只需要求: $E(X)=\alpha_1=A_1=\overline{X}\ \Rightarrow\theta$的矩估计量.

然而 $E(X)=\int_{-\infty}^{+\infty}xp(x;\theta)\mathrm{d}x$

然而 $E(X)=\int_{-\infty}^{+\infty} x p(x;\theta)\,\mathrm{d}x$

$$=\int_{-\infty}^{+\infty} x\cdot\frac{1}{2\theta}\mathrm{e}^{-\frac{|x|}{\theta}}\,\mathrm{d}x=0$$

不含有θ，故不能由此得到θ的矩估计量.

解(方法1) 要求：$E(X^2)=\alpha_2=A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$

$\because\quad E(X^2)=\int_{-\infty}^{+\infty} x^2 p(x;\theta)\,\mathrm{d}x$

$$=\int_{0}^{+\infty} x^2\cdot\frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}\,\mathrm{d}x=2\theta^2.$$

$$\therefore\quad \hat{\theta}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}X_i^2}$$

— θ的矩估计量

(方法2) 要求：

$$E(|X|)=\frac{1}{n}\sum_{i=1}^{n}|X_i|$$

$$\because\quad E(|X|)=\int_{-\infty}^{+\infty}|x|p(x;\theta)\mathrm{d}x=\int_{-\infty}^{+\infty}|x|\cdot\frac{1}{2\theta}\mathrm{e}^{-\frac{|x|}{\theta}}\mathrm{d}x$$

$$=\int_{0}^{+\infty}x\cdot\frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}\mathrm{d}x=\left(-x\mathrm{e}^{-\frac{x}{\theta}}\Big|_{0}^{+\infty}+\int_{0}^{+\infty}\mathrm{e}^{-\frac{x}{\theta}}\mathrm{d}x\right)=\theta$$

$\therefore$ θ的矩估计量：$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}|X_i|$

注　此例表明：同一参数的矩估计量可不唯一.

例5(p43例2.9)已知水文站最高水位X服从$\Gamma(\alpha,\beta)$,

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x}, & x>0,\\ 0, & x\le 0,\end{cases}$$

其中未知参数$\alpha>0,\beta>0$，试求α，β的矩估计

解 由Γ分布的性质1可知，

$$E(X)=\frac{\alpha}{\beta},\quad E(X^2)=\frac{(\alpha+1)\alpha}{\beta^2},$$

建立方程

$$\begin{cases} \dfrac{\alpha}{\beta} = \bar{X}, \\ \dfrac{(\alpha+1)\alpha}{\beta^2} = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 \end{cases}$$

求解方程可得

$$\hat{\alpha} = \frac{\bar{X}^2}{S_n^2}, \qquad \hat{\beta} = \frac{\bar{X}}{S_n^2}$$

小结： 矩法的**性质**：矩估计一定是相合估计。

矩法的**优点**：简单易行, 并不需要事先知道总体是什么分布 .

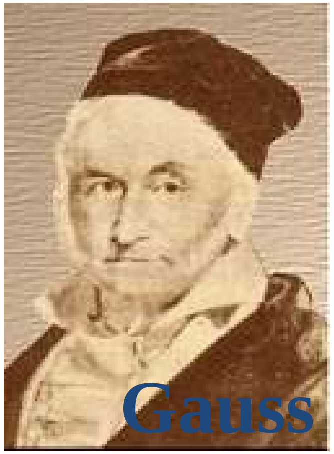
缺点：当总体类型已知时，没有充分利用分布提供的信息. 有时候，矩估计量不唯一 .

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

二、最大似然估计法

(Maximum-likelihood Extimation)

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．



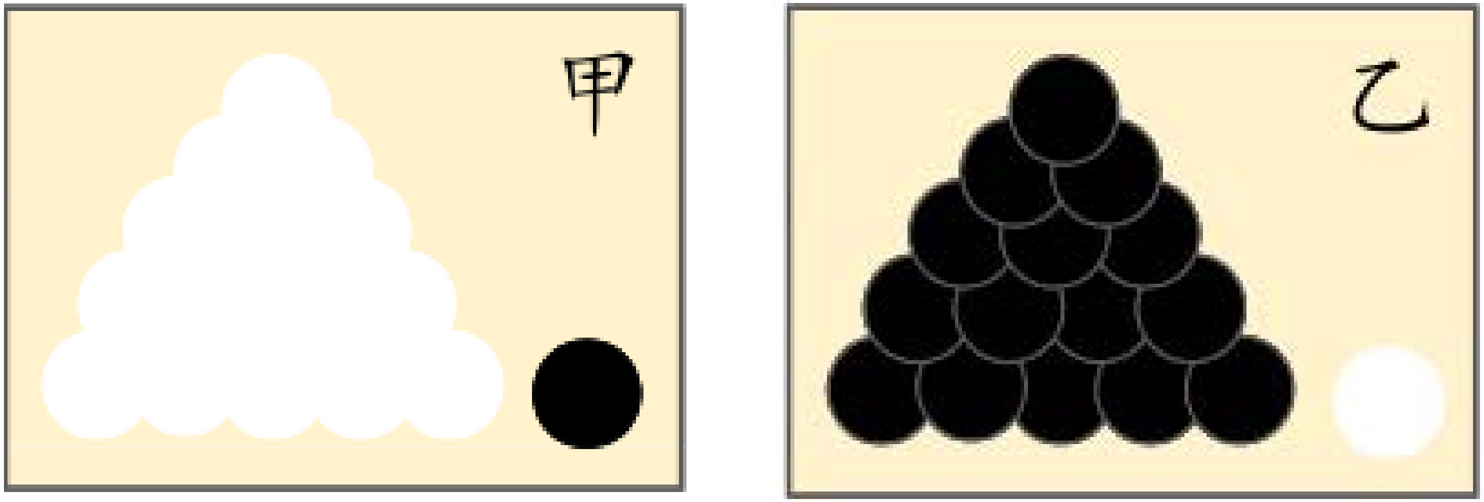
它首先是由德国数学家高斯在1821年提出的，

然而，这个方法常归功于英国统计学家**Fisher**．

Fisher在1921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．

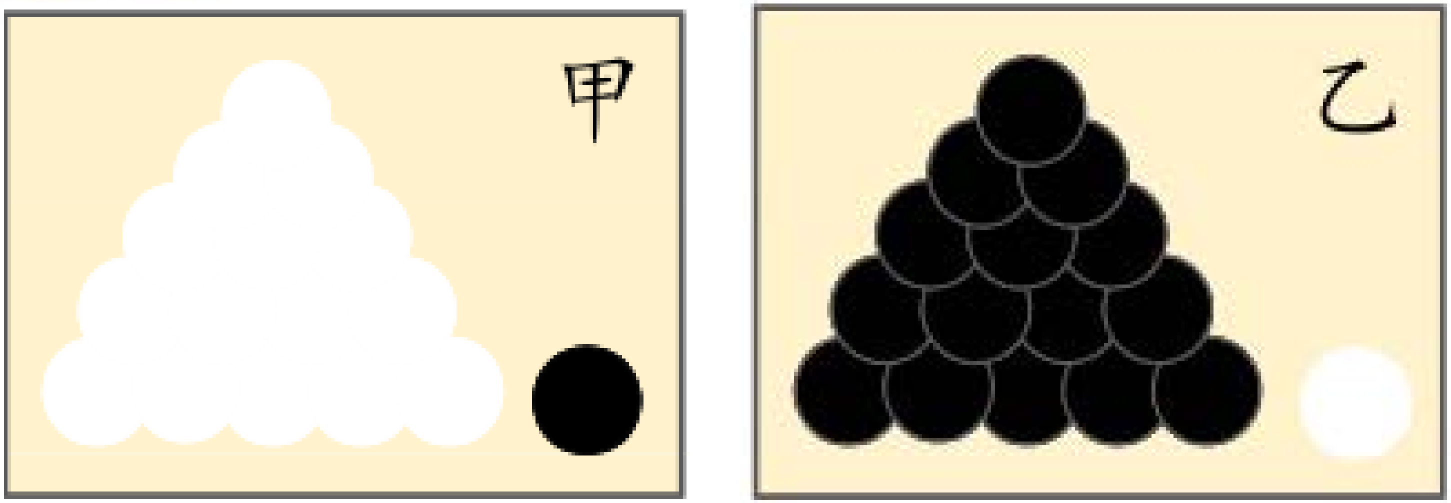
1 最大似然法的基本思想

引例1： 有两个外形完全相同的箱子，



甲箱中有99只白球，1只黑球；乙箱中有99只黑球，1只白球；一次试验取出一球，结果取出的是黑球。

问：黑球从哪个箱子中取出？

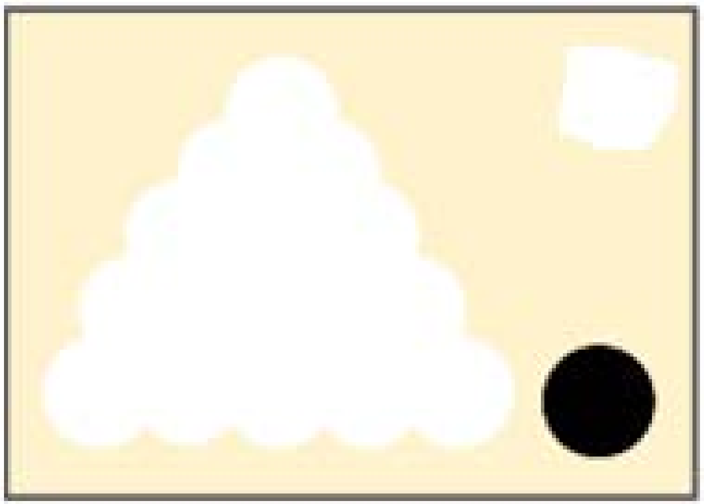


我们的第一印象就是：此黑球最像是从乙箱中取出的。

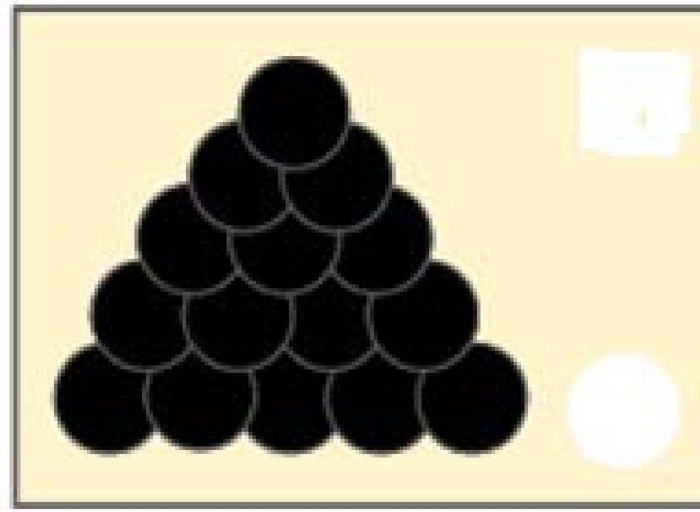
这个推断符合人们的经验事实。

“最像” 就是“最大似然” 之意。

引例2： 有一个箱子，



or



箱中要么有99只白球，1只黑球；要么有99只黑球，1只白球；从箱中任取出一球，是黑球。

问：估算箱子中的黑球数目是99只还是1只？

上问题可以转化为

$X \sim B(1, p)$, p表示取得黑球$\{X=1\}$的概率，

通过试验结果来估算 $p = 0.01 \; or \; 0.99.$

显然 p最可能取0.99.

$p = 0.99$ 使 $P\{X=1\} = p^{1}(1-p)^{1-1} = p = 0.99$

在p的两个取值0.01和0.99中达到最大.

一次试验中概率值最大的那个事件最有可能发生。
哪个参数值使得发生事件的概率值最大，参数
就最可能取哪个值。

引例3 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?

解 问题可以转化为

总体$X\sim B(1,p)$, p表示$\{X=1\}$的概率,

通过试验结果来估算 $p=0.7 \text{ or } 0.3.$

这里重复试验3次, 得结果： 0， 0， 0，

相当于从总体中取得一个样本容量为3的样本。

样本的联合分布律

$$P\{X_1 = x_1, X_2 = x_2, X_3 = x_3\} = \prod_{i=1}^{3} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{x_1+x_2+x_3}(1-p)^{3-(x_1+x_2+x_3)}$$

$$P\{X_1 = 0, X_2 = 0, X_3 = 0\} = p^0(1-p)^{3-0} = (1-p)^3$$

$= L(p)$ 应该最大

显然 $L(0.3) > L(0.7)$, 所以 $\hat{p} = 0.3.$

参数的估计值跟样本值有关。

2 似然函数

（1）设总体是离散型

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

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)设总体是连续型

似然函数的定义

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

$X_1, X_2, \cdots, X_n$ 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 $dx_1, dx_2, \cdots, dx_n$ 的 n 维立方体)内的概率近似地为 $\prod_{i=1}^{n} f(x_i;\theta)dx_i = \prod_{i=1}^{n} f(x_i;\theta) \underline{\prod_{i=1}^{n} dx_i}$，

固定

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

$L(\theta)$ 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

3. 求最大似然估计的步骤

(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. 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6(p46例2. 12) 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

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

所以 λ 的似然函数为

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

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

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

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

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

这一估计量与矩估计量是相同的.

例7(p47例2. 13)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

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

X 的似然函数为

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

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

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

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

它们与相应的矩估计量相同.

例7-1 设总体 $X \sim N(\mu,\sigma^2)$, μ为已知参数，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

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

X 的似然函数为

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

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

令 $\frac{d}{d\sigma^2}\ln L(\sigma^2) = 0,$

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

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

故当μ已知时，σ^2的最大似然估计为

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

？矩估计和最大似然估计结果一样吗

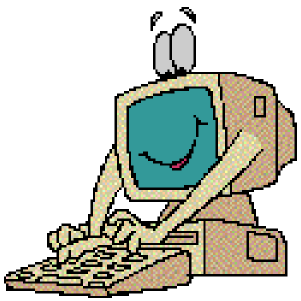
例8(p48例2.15) 设总体 X 在 $[\theta_1,\theta_2]$ 上服从均匀分布,其中 θ_1,θ_2 未知, $x_1,x_2,\cdots,x_n$ 是来自总体 X 的一个样本值,求 θ_1,θ_2 的最大似然估计量.

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

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

X 的概率密度为

$$f(x;\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1}, & \theta_1\le x\le\theta_2\\ 0, & \text{其它}\end{cases}$$



因为$\theta_1 \le x_1, x_2, \cdots, x_n \le \theta_2$等价于$\theta_1 \le x_{(1)}, x_{(n)} \le \theta_2$,

作为θ_1, θ_2的函数的似然函数为

$$L(\theta_1, \theta_2) = \begin{cases} \dfrac{1}{(\theta_2 - \theta_1)^n}, & \theta_1 \le x_{(1)},\ x_{(n)} \le \theta_2 \\ 0, & \text{其它} \end{cases}$$

于是对于满足条件$\theta_1 \le x_{(1)}, \ \theta_2 \ge x_{(n)}$的任意$\theta_1, \theta_2$有

$$L(\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} \le \frac{1}{(x_{(n)} - x_{(1)})^n},$$

即似然函数 $L(\theta_1,\theta_2)$ 在 $\theta_1=x_{(1)}$，$\theta_2=x_{(n)}$ 时

取到最大值 $(x_{(n)}-x_{(1)})^{-n}$，

θ_1,θ_2 的最大似然估计值

$$\hat{\theta}_1=x_{(1)}=\min_{1\le i\le n} x_i,\quad \hat{\theta}_2=x_{(n)}=\max_{1\le i\le n} x_i,$$

θ_1,θ_2 的最大似然估计量

$$\hat{\theta}_1=\min_{1\le i\le n} X_i,\quad \hat{\theta}_2=\max_{1\le i\le n} X_i\ .$$

矩估计

$$\hat{\theta}_1=A_1-\sqrt{3(A_2-A_1^2)}=\overline{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

$$\hat{\theta}_2=A_1+\sqrt{3(A_2-A_1^2)}=\overline{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2},$$

最大似然估计与矩估计不同！

例9 设总体 X 在$[\theta,\theta+1]$上服从均匀分布,其中θ未知, $x_1, x_2, \cdots, x_n$是来自总体 X的一个样本值,求θ的最大似然估计量.

解 X的概率密度为

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

$$L(\theta)=\begin{cases}1, & \theta \le x_{(1)},\ x_{(n)} \le \theta+1, \\ 0, & \text{其它}\end{cases}$$

似然函数在不为零的区域上是常数。

那么，只要θ不超过$x_{(1)}$, $\theta+1$不小于$x_{(n)}$，都可使L达到最大。

也即$x_{(n)}-1 \le \theta \le x_{(1)}$时,都可使L达到最大。

故θ的最大似然估计是区间$[X_{(n)}-1, X_{(1)}]$中任一值。

此例说明最大似然估计有时不唯一。

特殊情形总结

当似然函数是参数的严格单调函数时，利用似然函数的定义直观得到参数的最大似然估计。

（1）似然函数是参数的严格单调增函数时，参数最大似然估计值取其取值范围内最大值点。

（2）似然函数是参数的严格单调减函数时，参数最大似然估计值取其取值范围内最小值点。

（3）似然函数是参数的常函数时，参数最大似然估计值取其取值范围内任一点。

例10(p47例2.14) 设总体X服从柯西分布，其分布密度为

$$f(x,\theta)=\frac{1}{\pi[1+(x-\theta)^2]},\quad -\infty<x<+\infty,$$

解 由分布可知，其似然函数为

$$L(\theta)=\prod_{i=1}^{n}\frac{1}{\pi[1+(x_i-\theta)^2]}$$

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

此方程只能求解其数值解，可以以样本中位数为初始值进行迭代。又因为此分布均值不存在，不能用矩估计.

4. 最大似然估计的性质

定理2.4 设$\hat{\theta}$是θ的最大似然估计，如果函数$g(\theta)$是$\theta\in\Theta$的连续函数，则$g(\hat{\theta})$是$g(\theta)$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.

例10(p48例2. 16)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$g(\mu,\sigma^2)=P\{X>2\}$的最大似然估计.

解 $$g(\mu,\sigma^2)=P\{X>2\}=1-P\{X\le 2\}$$

$$=1-P\{\frac{X-\mu}{\sigma}\le\frac{2-\mu}{\sigma}\}=1-\Phi(\frac{2-\mu}{\sigma})$$

而μ和σ^2的最大似然估计为$\bar{X}, S_n^2$,同时$\Phi(\mu,\sigma^2)$为连续函数，因而$g(\mu,\sigma^2)$的最大似然估计为

$$g(\hat{\mu},\hat{\sigma}^2)=1-\Phi(\frac{2-\bar{X}}{S_n})$$

定理2.5 设$T=T(X_1,X_2,\cdots,X_n)$为θ的任一充分统计量，则θ的最大似然估计$\hat{\theta}$一定可以表示成T的函数.

证 由因子分解定理可知

$$L(\theta)=h(x_1,x_2,\cdots,x_n)g(T(x_1,x_2,\cdots,x_n),\theta)$$

其中h与θ无关，因此，最大化$L(\theta)$等于最大化$g(T,\theta)$,因此，最大似然估计$\hat{\theta}$（若存在）一定是T的函数.

注 该定理说明最大似然估计是充分统计量的函数。

通常，最大似然估计是渐近正态估计，也是相合估计。

Thank You!