Lec5: 点估计(二)

张伟平

September 29, 2009

§1 极大似然估计

一、引言及定义

极大似然法是在参数分布族场合下常用的参数估计方法. 设有一参数分布族$\mathscr{F}=\{F_\theta,\ \theta\in\Theta\}$,此处$\Theta$为参数空间. 令$\mathbf{X}=(X_1,\cdots,X_n)$为从$\mathscr{F}$中抽取的简单随机样本, $f(\mathbf{x},\theta)=f(x_1,\ldots,x_n,\theta)$为样本$\mathbf{X}$的概率函数. 即当总体分布为连续型时,$f(\mathbf{x},\theta)$表示样本$\mathbf{X}$的密度函数; 当总体分布为离散形时,$f(\mathbf{x},\theta)$为样本$\mathbf{X}$的概率分布, 即$f(\mathbf{x},\theta)=P_\theta(X_1=x_1,\cdots,X_n=x_n)$. 首先引入似然函数的概念.

定义 1. *设$f(\mathbf{x},\theta)=f(x_1,\cdots,x_n,\theta)$为样本$\mathbf{X}=(X_1,\cdots,X_n)$的概率函数. 当$\mathbf{x}$固定时, 把$f(\mathbf{x},\theta)$看成$\theta$的函数, 称为似然函数(Likelihood function),记为*

$$L(\theta,\mathbf{x})=f(\mathbf{x},\theta),\qquad \theta\in\Theta,\quad \mathbf{x}\in\mathscr{X} \tag{1.1}$$

此处Θ为参数空间, $\mathscr{X}$为样本空间. 称$\log L(\theta,\mathbf{x})$为对数似然函数, 记为$l(\theta,\mathbf{x})$.

注1. *似然函数和概率函数是同一表达式(1.1), 但表示两种不同含意. 当把θ固定, 将其看成定义在样本空间$\mathscr{X}$上的函数时, 称为概率函数; 当把$\mathbf{x}$固定, 将其看成定义在参数空间Θ上的函数时, 称为似然函数. 这是两个不同的概念.*

为解释极大似然原理, 考虑下列一个简单的实例.

例1. *设罐子里有许多黑球和红球. 假定已知它们的比例是$1:3$,但不知道是黑球多还是红球多. 也就是说抽出一个黑球的概率或者是$1/4$或者是$3/4$. 如果有放回地从罐子中抽n个球, 要根据抽样数据, 说明抽到黑球的概率是$1/4$, 还是$3/4$.*

解 将此问题用统计模型来表述. 令X_i表示第i次抽球的结果,即

$$X_i=\begin{cases}1 & \text{抽出为黑球}\\ 0 & \text{否则.}\end{cases}$$

记每次抽样中抽到黑球的概率为θ,此处θ 只取可能的两个值$\theta_1=1/4,\ \theta_2=3/4$之一. 记$X=\sum_{i=1}^n X_i$,则$X\sim b(n,\theta)$.亦即样本分布族$\mathscr{F}=\{F_{\theta_1},F_{\theta_2}\}$, 其中$F_{\theta_1}$为$b(n,\theta_1)$, F_{θ_2}为$b(n,\theta_2)$.要根据抽样结果对θ作出估计, 即θ取值为$1/4$还是$3/4$? 或曰样本来自总体F_{θ_1}还是F_{θ_2}?

显然, 当样本X给定时, 似然函数为

$$L(\theta, x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x=0,1,2,\cdots,n.$$

为简单计, 取$n=3$.当$x=0,1,2,3$时似然函数取值如下表:

x	0	1	2	3
$L(\theta_1, x)$	27/64	27/64	9/64	1/64
$L(\theta_2, x)$	1/64	9/64	27/64	27/64

可见

当$x=0,1$时, $\quad L(\theta_1,x)>L(\theta_2,x)$,

当$x=2,3$时, $\quad L(\theta_2,x)>L(\theta_1,x)$.

因此我们得出结论: 当样本观察值$x=\sum_{i=1}^3 x_i$取值为$0,1$ 时认为样本来自总体F_{θ_1},即取参数θ的估计值为$\hat{\theta}_1=1/4$;当$x=2,3$时认为样本来自总体F_{θ_2}, 即取θ的估计值为$\hat{\theta}_2=3/4$.

将此例模型化如下: 若样本$\mathbf{X}=(X_1,\cdots,X_n)$为从总体$\mathscr{F}=\{F_\theta, \theta\in\Theta\}$中抽取的简单随机样本, 其中$\Theta=\{\theta_1,\theta_2\}$. 此即等价地说分布族$\mathscr{F}$中只有两个总体$F_{\theta_1}, F_{\theta_2}$. 一旦获得了样本$\mathbf{x}$,如何用极大似然方法求出真参数$\theta$的估计值呢? 上例表明若

$$L(\theta_1,\mathbf{x})>L(\theta_2,\mathbf{x}),$$

则我们倾向于认为样本$\mathbf{X}$ 来自总体F_{θ_1} (即真参数θ为θ_1)的理由比认为样本$\mathbf{X}$来自总体F_{θ_2} (即真参数θ为θ_2)的理由更充分些. 或者说, 真参数θ为θ_1的“似然性”更大些. 这样, 我们自然把“似然性”最大(即看起来最像)的那个值作为真参数θ的估计值. 这正是“极大似然”一词的由来.

更一般, 若样本$\mathbf{X}$的分布族$\mathscr{F}=\{F_\theta,\theta\in\Theta\}$, 参数空间$\Theta$为$R_1$的有限子集或无限子集. 当样本$\mathbf{x}$给定时, 若$\hat{\theta}^*$使似然函数$L(\hat{\theta}^*,\mathbf{x})$ 为似然函数的集合$\{L(\theta,\mathbf{x})$, 一切 $\theta\in\Theta\}$中最大者, 即参数θ的真值为$\hat{\theta}^*$的“似然性”比θ取Θ 中其它值的“似然性”更大, 则取“似然性”最大的$\hat{\theta}^*$作为θ的估计值, 这一方法得到参数θ的估计,称为“极大似然估计”. 将这一直观想法用数学语言来描述, 得到如下定义:

定义 2. *设$\mathbf{X}=(X_1,\cdots,X_n)$是从参数分布族$\mathscr{F}=\{F_\theta,\theta\in\Theta\}$中抽取的简单随机样本, $L(\theta,\mathbf{x})$是似然函数, 若存在统计量$\hat{\theta}^*(\mathbf{X})=\hat{\theta}^*(X_1,\cdots,X_n)$, 满足条件*

$$L\big(\hat{\theta}^*(\mathbf{x}),\mathbf{x}\big)=\sup_{\theta\in\Theta}L(\theta,\mathbf{x}), \quad x\in\mathscr{X}, \tag{1.2}$$

或等价地使得

$$l\big(\hat{\theta}^*(\mathbf{x}),\mathbf{x}\big)=\sup_{\theta\in\Theta}l(\theta,\mathbf{x}), \quad x\in\mathscr{X}, \tag{1.3}$$

则称$\hat{\theta}^*(\mathbf{X})$为$\theta$ 的极大似然估计 (Maximum Likelihood Estimation,简记为MLE). 若待估函数是$g(\theta)$,则定义$g\big(\hat{\theta}^*(\mathbf{x})\big)$为$g(\theta)$的MLE.

极大似然估计是R.A. Fisher在1912年的一项工作中提出来的. 在正态分布这个特殊情况下, 这方法可追溯到Gauss在19世纪初关于最小二乘法的工作. Fisher后来在1922年工作中, 尤其1925年发表的《Theory of Statistical Estimation》一文中对这一估计作了许多研究. 因此这个方法应归功于R.A. Fisher.

二、极大似然估计的求法及例

获得参数θ的极大似然估计有下列两种方法:

1. 用微积分中求极值的方法

若似然函数$L(\theta,\mathbf{x})$是θ的连续可微函数, 则我们可用微积分中求极值点的方法去求θ的MLE. 找使$L(\theta,\mathbf{x})$达到最大时θ的值. 由于$L(\theta,\mathbf{x})$和$\log L(\theta,\mathbf{x}) = l(\theta,\mathbf{x})$具有相同的极值点, 我们可用$l(\theta,\mathbf{x})$来代替$L(\theta,\mathbf{x})$. 因为$L(\theta,\mathbf{x})$一般是若干个函数的乘积, $l(\theta,\mathbf{x})$ 为若干个函数之和而较易于处理.

设$\theta=(\theta_1,\cdots,\theta_k)$为参数向量(特别当$k=1$, θ为参数). 若$l(\theta,\mathbf{x})$的极大值在参数空间Θ的内点处(而非边界点)达到, 则此点必为似然方程组

$$\frac{\partial l(\theta,\mathbf{x})}{\partial\theta_i}=0,\quad i=1,2,\cdots,k$$

的解.

因此求极大似然估计首先求似然方程的解$\hat{\theta}$.但此解是否一定是θ 的MLE 呢? $\hat{\theta}$满足似然方程, 只是MLE的必要条件, 而非充分条件. 一般只有满足下列条件: (i) 似然函数的极大值在参数空间Θ内部达到, (ii) 似然方程只有唯一解, 则似然方程之解$\hat{\theta}$必为θ的MLE.

因此我们求出似然方程(或方程组)的解后, 要验证它为θ的MLE, 有时并非易事. 但对样本分布族是指数族的场合, 有非常满意的结果, 叙述如下:

设$\mathbf{X}=(X_1,\cdots,X_n)$为从某总体中抽取的简单随机样本, X_1的分布为指数族, 即

$$f(x,\theta)=C(\theta)\exp\Big\{\sum_{i=1}^{k}\theta_iT_i(x)\Big\}h(x),\quad \theta\in\Theta$$

此处Θ为自然参数空间, Θ_0为Θ之内点集, 这时$\mathbf{X}$的联合密度为

$$L(\theta,\mathbf{x})=C^n(\theta)\exp\Big\{\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)\Big\}h(\mathbf{x}),$$

其中$h(\mathbf{x})=\prod\limits_{i=1}^{n}h(x_i)$.对上式取对数得

$$l(\theta,\mathbf{x})=\log L(\theta,\mathbf{x})=n\log C(\theta)+\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)$$

定理 1. *若对任何样本$\mathbf{X}=(X_1,\cdots,X_n)$,方程组*

$$\frac{n}{C(\theta)}\frac{\partial C(\theta)}{\partial\theta_i}=-\sum_{j=1}^{n}T_i(X_j),\quad i=1,2,\cdots,k$$

在Θ_0内有解, 则解必唯一, 且为θ的MLE.

证明思路 由于自然参数空间为一凸集, 若方程组有两个不同的解, θ_0和θ_1, 则构造函数

$$H(t)=l(t\theta_0+(1-t)\theta_1), 0\le t\le 1.$$

由于θ_0和θ_1都是方程$\partial l(\theta)/\partial\theta = 0$的解, 因此$H'(0) = H'(1) = 0$, 所以存在$0 < t_0 < 1$, 使得$H''(t_0) = 0$. 记

$$\tilde{\theta} = t_0\theta_0 + (1 - t_0)\theta_1$$

则

$$H''(t_0) = (\theta_0 - \theta_1)' Q(t_0)(\theta_0 - \theta_1) = 0$$

其中$Q(\theta) = \frac{\partial^2 logC(\theta)}{\partial\theta\partial\theta^T} = -Var(T_1(X), \cdots, T_k(X)) < 0$, 对任意的$\Theta$内点$\theta$, 上式与此特性矛盾. 因此$\theta_0 = \theta_1$.

另一方面, 若θ_0为一个解, 而用θ_1表示Θ中任意一点, 则前面构造的函数$H(t)$满足$H'(1) = 0, H''(t) < 0$. 因此$H'(t) > 0, 0 \le t < 1$. 所以$H(1) > H(0)$, 即$l(\theta_0) > l(\theta_1)$, θ_0为l的极大点.

因此若样本分布为指数族, 只要似然方程解属于自然参数空间的内点集(这点很容易验证), 则其解必为θ的MLE. 常见的分布族, 如二项分布族、Poisson分布族、正态分布族、Gamma分布族等都是指数族, 定理3.3.1的条件皆成立. 因此似然方程的解, 就是有关参数的MLE.

2. 从定义出发

若似然函数$L(\theta, \mathbf{x})$不是θ的可微函数,特别当似然函数$L(\theta, \mathbf{x})$不是θ的连续函数时, 我们不能采用上述方法, 必须直接从定义3.3.2出发去求参数θ的极大似然估计.

下面我们分别用上述两种方法考察一些例子.

例2. *设$\mathbf{X} = (X_1, \cdots, X_n)$是从两点分布族$\{b(1, p) : 0 < p < 1\}$中抽取的简单样本, 求$p$和$g(p) = p(1 - p)$的MLE.*

解 似然函数为

$$L(p, \mathbf{x}) = p^{\sum\limits_{i=1}^{n} x_i}(1 - p)^{n - \sum\limits_{i=1}^{n} x_i},$$

故有

$$l(p, \mathbf{x}) = \log L(p, \mathbf{x}) = \left(\sum_{i=1}^{n} x_i\right)\log p + \left(n - \sum_{i=1}^{n} X_i\right)\log(1 - p)$$

对数似然方程为

$$\frac{\partial l(p, \mathbf{x})}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1 - p}\left(n - \sum_{i=1}^{n} x_i\right) = 0,$$

解得

$$\hat{p}^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

由于两点分布族为指数族, 故$\hat{p}^* = \bar{X}$为p的MLE. 它与例3.2.2中给得出的p的矩估计量相同. 按定义可知$g(p) = p(1 - p)$的MLE为

$$\hat{g}^*(\mathbf{X}) = \hat{p}^*(1 - \hat{p}^*) = \bar{X}(1 - \bar{X}).$$

例3. *设$\mathbf{X} = (X_1, \cdots, X_n)$是从Poisson分布族$\{P(\lambda) : \lambda > 0\}$中抽取的简单样本, 求$\lambda$和$g(\lambda) = e^{-\lambda}$的MLE.*

解 似然函数,即样本$\mathbf{X}=(X_1,\cdots,X_n)$的分布为

$$L(\lambda,\mathbf{x})=P(X_1=x_1,\cdots,X_n=x_n)=\frac{\lambda^{\sum_{i=1}^{n}x_i}e^{-n\lambda}}{x_1!\cdots x_n!},\quad \lambda>0.$$

故对数似然函数为

$$l(\lambda,\mathbf{x})=\Big(\sum_{i=1}^{n}x_i\Big)\log\lambda-n\lambda-\sum_{i=1}^{n}\log x_i!\ .$$

由对数似然方程

$$\frac{\partial l(\lambda,\mathbf{x})}{\partial\lambda}=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0,$$

解得

$$\hat{\lambda}^*=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\ .$$

由于Poisson分布族是指数族, 故$\hat{\lambda}^*=\bar{X}$为λ的MLE, 它与λ的矩估计量相同.

又由定义可知$g(\lambda)=e^{-\lambda}$的MLE为

$$\hat{g}^*(\mathbf{X})=e^{-\bar{X}}.$$

例4. *设$\mathbf{X}=(X_1,\cdots,X_n)$是从正态分布族$\{N(a,\sigma^2):\ \sigma^2>0,\ -\infty<\mu<\infty\}$中抽取的简单样本, 求$a$, σ^2和$g(\theta)=a/\sigma^2$的MLE, 此处$\theta=(a,\sigma^2)$.*

解 样本$\mathbf{X}=(X_1,\cdots,X_n)$的分布为

$$f(\mathbf{x},\theta)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right\}.$$

对数似然函数为

$$l(\theta,\mathbf{x})=\log f(\mathbf{x},\theta)=-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-a)^2.$$

由对数似然方程组

$$\begin{aligned}\frac{\partial l(\theta,\mathbf{x})}{\partial a}&=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-a)=0,\\ \frac{\partial l(\theta,\mathbf{x})}{\partial\sigma^2}&=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i-a)^2=0,\end{aligned}$$

解得

$$\hat{a}^*=\bar{X},\qquad \hat{\sigma}^2_*=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=S_n^2\ .$$

由于正态分布族为指数族, 故$\hat{a}^*=\bar{X}$和$\hat{\sigma}^2_*=S_n^2$ 分别是a和σ^2的MLE,它们也分别是a和σ^2的矩估计量. 前者是a的无偏估计, 后者不是σ^2的无偏估计. 可见极大似然估计不一定具有无偏性.

又由定义可知$g(\theta)=a/\sigma^2$的MLE为

$$\hat{g}^*(\mathbf{X})=\bar{X}/S_n^2.$$

例5. 设元件的寿命服从下列指数分布$EP(\lambda)$,其密度函数为

$$f(x,\lambda)=\begin{cases}\lambda e^{-\lambda x} & \text{当 } x>0\\ 0 & \text{当 } x\le 0.\end{cases}$$

设 $X_1,\cdots,X_n$分别表示接受试验的n个元件寿命. 由于时间的限制, 试验实际上只进行到有r $(r\le n)$个元件失效时就停止了, 以$X_{(1)}\le X_{(2)}\le\cdots\le X_{(r)}$记这$r$个元件的寿命. 即, 我们只观察到了样本 $X_1,\cdots,X_n$前r个次序统计量$X_{(1)},X_{(2)},\cdots,X_{(r)}$. 基于这前$r$个次序统计量, 求$\lambda$和$g(\lambda)=1/\lambda$的$MLE$.

解 为叙述方便, 记$t_i=x_{(i)}$, $i=1,2,\cdots,r$,则有$t_1\le t_2\le\cdots\le t_r$. 为确定似然函数, 我们要知道上述观察结果的概率. 一个产品在$[t_i,t_i+\Delta t_i]$内失效的概率为$f(t_i)\Delta t_i$, $i=1,2,\cdots,r$.其余$n-r$个产品寿命超过t_r的概率为$\left[e^{-\lambda t_r}\right]^{n-r}$,所以上述结果出现的概率近似为

$$k(\lambda e^{-\lambda t_1}\Delta t_1)(\lambda e^{-\lambda t_2}\Delta t_2)\cdots(\lambda e^{-\lambda t_r}\Delta t_r)[e^{-\lambda t_r}]^{n-r},$$

其中k为某一常数. 忽略一个常数因子对求MLE无影响. 故取似然函数为

$$L\left(\lambda,x_{(1)},\cdots,x_{(r)}\right)=k\lambda^r\exp\left\{-\lambda\left(\sum_{i=1}^r x_{(i)}+(n-r)x_{(r)}\right)\right\},$$

其对数似然函数为

$$l\left(\lambda,x_{(1)},\cdots,x_{(r)}\right)=r\log\lambda-\lambda\left(\sum_{i=1}^r x_{(i)}+(n-r)x_{(r)}\right).$$

对λ求导, 得似然方程为

$$\frac{\partial l}{\partial\lambda}=\frac{r}{\lambda}-\left(\sum_{i=1}^r x_{(i)}+(n-r)x_{(r)}\right)=0,$$

解得

$$\hat\lambda^*=r\Big/\left(\sum_{i=1}^r X_{(i)}+(n-r)X_{(r)}\right).$$

由于似然函数$L\left(\lambda,x_{(1)},\cdots,x_{(r)}\right)$是指数族的形式, 故$\hat\lambda^*$为$\lambda$的MLE. 由定义可知$g(\lambda)=1/\lambda$的MLE为

$$\hat g^*(X_{(1)},\cdots,X_{(r)})=\left(\sum_{i=1}^r X_{(i)}+(n-r)X_{(r)}\right)\Big/r.$$

本例中所述产品寿命试验进行到第r个产品失效时就终止, 这种试验叫 定数截尾试验. 另一种方式是, 先定下一个时间$T>0$, 当试验进行T时试验就终止, 这种试验叫做定时截尾试验.

例6. 设$\mathbf{X}=(X_1,\cdots,X_n)$是从均匀分布族$\{U(0,\theta):\ \theta>0\}$中抽取的简单样本. *(1)* 求$\theta$的$MLE$ $\hat\theta^*$; *(2)* 说明$\hat\theta^*$是否为θ的无偏估计. 若不然, 作适当修正获得θ的无偏估计$\hat\theta_1^*$; *(3)* 试将$\hat\theta_1^*$ 与θ的矩估计进行比较, 看哪一个有效? *(4)*证明θ的极大似然估计$\hat\theta^*=X_{(n)}$是θ的弱相合估计.

解 (1) 样本$\mathbf{X}=(X_1,\cdots,X_n)$的分布为

$$f(\mathbf{x},\theta)=\begin{cases}\frac{1}{\theta^n} & \text{当 } 0<x_1,\cdots,x_n<\theta\\ 0 & \text{其 他.}\end{cases}\tag{1.4}$$

因为均匀分布$U(0,\theta)$的支撑集依赖于θ,似然函数$L(\theta,\mathbf{x})=f(\mathbf{x},\theta)$作为$\theta$的函数不是连续函数,因此不能用对然函数求微商的办法去求θ的MLE. 只能从MLE的定义出发来讨论.

为使$L(\theta,\mathbf{x})$达到极大, 由(1.4)式可见, 应使分母上的θ 尽可能地小, 但θ又不能太小, 太小了使似然函数变为0了. 这个界限就取在

$$\hat{\theta}^*=\max(X_1,\cdots,X_n)=X_{(n)}.$$

因此$\hat{\theta}^*=X_{(n)}$ 就是θ的MLE.

(2) 为求$E\left(X_{(n)}\right)$,就要算出$T=X_{(n)}$的密度函数, 易求T 的密度函数

$$g(t,\theta)=\begin{cases}\frac{nt^{n-1}}{\theta^n} & \text{当 } 0\le t\le\theta\\ 0 & \text{其 它},\end{cases} \tag{1.5}$$

故有

$$E(\hat{\theta}^*)=E(T)=\int_0^\theta\frac{nt^n}{\theta^n}dt=\frac{n}{n+1}\theta,$$

故$\hat{\theta}^*=X_{(n)}$不是θ的无偏估计. 显见

$$\hat{\theta}_1^*=\frac{n+1}{n}X_{(n)}$$

为θ的无偏估计.

(3) θ的矩估计$\hat{\theta}_1=2\bar{X}$是θ的无偏估计. 由于

$$D(\hat{\theta}_1^*)=\theta^2\big/[n(n+2)],\qquad D(\hat{\theta}_1)=\theta^2\big/3n,$$

所以在$n\ge 2$时$\hat{\theta}_1^*$比$\hat{\theta}_1$有效. 在$n=1$时$\hat{\theta}_1^*=\hat{\theta}_1$,即这两个估计是相同的.

(4) 已知T的密度函数由(1.5)给出, 故有

$$\begin{aligned}P\left(|\hat{\theta}^*-\theta|\ge\varepsilon\right)&=1-P(|\hat{\theta}^*-\theta|<\varepsilon)=1-P(\theta-\varepsilon<T<\theta+\varepsilon)\\&=1-\int_{\theta-\varepsilon}^{\theta}\frac{nt^{n-1}}{\theta^n}dt=1-\frac{1}{\theta^n}\left[\theta^n-(\theta-\varepsilon)^n\right]=\left(1-\frac{\varepsilon}{\theta}\right)^n.\end{aligned}$$

因此有

$$\lim_{n\to\infty}P(|\hat{\theta}^*-\theta|\ge\varepsilon)=\lim_{n\to\infty}\left(1-\varepsilon/\theta\right)^n=0,$$

故知$\hat{\theta}^*=X_{(n)}$为θ的弱相合估计.

例7. 设$\mathbf{X}=(X_1,\cdots,X_n)$是从均匀分布族$\{U(\theta,\theta+1):\ -\infty<\theta<+\infty\}$中抽取的简单样本,求$\theta$的*MLE*.

解 给定样本$\mathbf{x}$时, θ的似然函数为

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

这时, 似然函数只取1和0两个值, 只要$\theta\le x_{(1)}$和$\theta+1\ge x_{(n)}$都可使L达到极大. 故θ的MLE不止一个, 如

$$\hat{\theta}_1^*(\mathbf{X})=X_{(1)},\quad \hat{\theta}_2^*(\mathbf{X})=X_{(n)}-1$$

都是θ的MLE.事实上对任给的$0<\lambda<1$,

$$\hat{\theta}^*(\mathbf{X})=\lambda\hat{\theta}_1^*(\mathbf{X})+(1-\lambda)\hat{\theta}_2^*(\mathbf{X})=\lambda X_{(1)}+(1-\lambda)\left(X_{(n)}-1\right)$$

都是θ的MLE, 故知θ的MLE有无穷多个.

例8. 设k个事件$A_1, A_2, \cdots, A_k$构成完备事件群, 事件A_i发生的概率为p_i, $i=1,\cdots,k$且$\sum\limits_{i=1}^{k} p_i = 1$.将试验独立重复$n$次, 以$X_i$记$A_i$发生的次数, $i=1,2,\cdots,k$,则$\mathbf{X}=(X_1,\cdots,X_k)$ 服从多项分布$M(n,p_1,\cdots,p_k)$.求$p_1,\cdots,p_k$的*MLE.*

解 记$\mathbf{p}=(p_1,\cdots,p_k)$.给定样本$\mathbf{x}$时, $\mathbf{p}$的似然函数为

$$L(\mathbf{p},\mathbf{x}) = \frac{n!}{x_1!\cdots x_k!} p_1^{x_1}\cdots p_k^{x_k} = \frac{n!}{x_1!\cdots x_k!} p_1^{x_1}\cdots p_{k-1}^{x_{k-1}} \left(1-\sum_{i=1}^{k-1} p_i\right)^{x_k}.$$

对数似然函数为

$$l(\mathbf{p},\mathbf{x}) = \log n! - \sum_{i=1}^{k} \log x_i! + \sum_{i=1}^{k-1} x_i \log p_i + x_k \log\left(1-\sum_{i=1}^{k-1} p_i\right). \tag{1.6}$$

对p_i求偏导数, 得似然方程组

$$\frac{\partial l(\mathbf{p},\mathbf{x})}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_k}{p_k} = 0, \quad i=1,2,\cdots,k-1.$$

若令

$$\frac{x_1}{p_1} = \frac{x_2}{p_2} = \cdots = \frac{x_k}{p_k} = \lambda,$$

则有

$$x_i = \lambda p_i, \qquad i=1,2,\cdots,k. \tag{1.7}$$

将这k个等式两边分别相加得

$$n = \sum_{i=1}^{k} x_i = \lambda \sum_{i=1}^{k} p_i = \lambda.$$

因此由(1.7)可知p_i的MLE如下:

$$\hat{p}_i^* = X_i/n, \quad i=1,2,\cdots,k.$$

例9. 设$X_1,\cdots,X_n$为从如下分布中抽取的简单样本，求θ的*MLE.*

$$f(x) = \frac{1}{x!(2-x)!}[\theta^x(1-\theta)^{2-x} + \theta^{2-x}(1-\theta)^x], \quad x=0,1,2; \quad \theta\in(0,\frac{1}{2})$$

解: 由题设知$f(x)$为离散型，其分布律为

X	0	1	2
P	$\frac{1}{2}[(1-\theta)^2+\theta^2]$	$2\theta(1-\theta)$	$\frac{1}{2}[(1-\theta)^2+\theta^2]$

若直接从此分布出发，则不能得到θ的MLE的显式表达。为此，我们重新参数化，记$\eta=2\theta(1-\theta)$. 则由题设知$0<\eta<1/2$。则

X	0	1	2
P	$\frac{1}{2}(1-\eta)$	η	$\frac{1}{2}(1-\eta)$

再记$n_i = \#\{X_1,\cdots,X_n$中等于i的个数$\}$, $\quad i=0,1,2$, 则得到似然函数为

$$L(\eta) = (\frac{1}{2}(1-\eta))^{n_0}\eta^{n_1}(\frac{1}{2}(1-\eta))^{n_2} = (\frac{1}{2}(1-\eta))^{n-n_1}\eta^{n_1}$$

求解并注意η的下界即得到η的MLE为

$$\hat{\eta} = \max\{\frac{n_1}{n}, \frac{1}{2}\}$$

再由$\theta = \frac{1-\sqrt{1-2\eta}}{2}$得到$\theta$的MLE为

$$\hat{\theta} = \frac{1-\sqrt{1-2\hat{\eta}}}{2}$$

□

例10. 设从总体

X	0	1	2	3
P	$\theta/2$	θ	$3\theta/2$	$1-3\theta$

抽取的一个简单样本$X_1, \cdots, X_{10}$的观察值为$(0,3,1,1,0,2,0,0,3,0)$，

(1) 求θ的矩估计量$\hat{\theta}_M$和极大似然估计量$\hat{\theta}_L$，并求出估计值。

(2) 上述估计量是否为无偏的？若不是，请作修正.

(3) 比较修正后的两个估计量，指出那个更有效.

在有些问题中，$p_1, \cdots, p_k$都是另一些参数$\theta_1, \cdots, \theta_r$ $(r \leq k)$的函数，这时去掉(1.6)中与$\theta_1, \cdots, \theta_r$无关的部分后, 得对数似然函数为

$$l(\theta_1, \cdots, \theta_r; \mathbf{x}) = \sum_{i=1}^{k} x_i \log p_i(\theta_1, \cdots, \theta_r)$$

对θ_i求偏导数得似然方程组

$$\frac{\partial l}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, r. \tag{1.8}$$

若似然方程之解$\hat{\theta}_1^*, \cdots, \hat{\theta}_r^*$ 为$\theta_1, \cdots, \theta_r$的MLE, 则$p_1, \cdots, p_k$的MLE分别是$\hat{p}_1(\hat{\theta}_1^*, \cdots, \hat{\theta}_r^*) \cdots,$ $\hat{p}_k(\hat{\theta}_1^*, \cdots, \hat{\theta}_r^*)$. 但有时似然方程组(1.8)无显式解, 就只能用数值方法.

三、极大似然估计的性质*

前面例子中已指出极大似然估计不一定是无偏的. 至于相合性问题, 远没有矩估计那么简单. 极大似然估计的相合性问题, 引起许多统计学者的兴趣, 直到现在都不能说已彻底解决了. 1946年Cramer在一些条件下, 证明了似然方程有一根是参数θ的弱相合估计. 由于似然方程的根不一定是极大似然估计, 这个结果还是没有解决似然估计的相合性问题. 直到1949年, Wald才首次证明了极大似然估计的强相合性, 但所要求的条件很复杂. 此后有一些学者继续进行研究, 希望在较少的条件下证明相合性. 这些结果, 不便在此一一细述.

存在反例, 说明极大似然估计可以不相合, 有兴趣的读者可查看参考文献[2] P_{81}的反例.

下面来介绍极大似然估计的几条性质:

1. 极大似然估计与充分统计量

设$\mathbf{X} = (X_1, \cdots, X_n)$为自总体$\{f(x,\theta), \theta \in \Theta\}$ 中抽出的简单随机样本, $T = T(X_1, \cdots, X_n)$是参数θ的充分统计量, 如果θ的极大似然估计存在, 则它必为T的函数.

由因子分解定理可知样本$\mathbf{X}$的概率函数, 亦即似然函数可表为

$$L(\theta, \mathbf{x}) = \prod_{i=1}^{n} f(x_i, \theta) = g(T(\mathbf{x}), \theta) h(\mathbf{x}).$$

因此使$\sup\limits_{\theta} L(\theta,x)$达到上确界之点$\hat{\theta}^*$, 即为使$\sup\limits_{\theta} g(T(\mathbf{x}),\theta)$达到上确界之点, 它必为$T(\mathbf{x})$的函数.

此性质说明θ的极大似然估计$\hat{\theta}^*=\hat{\theta}^*(X_1,\cdots,X_n)$ 可表为$T(\mathbf{X})$的函数即$\hat{\theta}^*=\hat{\theta}^*(T(X_1,\cdots,X_n))$. 如例3.3.2—例3.3.8中的极大似然估计皆为充分统计量的函数.

2. 极大似然估计与有效估计

设$\mathbf{X}=(X_1,\cdots,X_n)$为自分布族$\{f(x,\theta),\theta\in\Theta\}$ 抽取的简单随机样本, 若$g(\theta)$的有效估计 $\hat{g}(\mathbf{X})=\hat{g}(X_1,\cdots,X_n)$存在, 则$g(\theta)$的MLE $\hat{g}^*(\mathbf{X})=\hat{g}^*(X_1,\cdots,X_n)$ 必与$\hat{g}(\mathbf{X})$重合, 即$\hat{g}^*(\mathbf{X})=\hat{g}(\mathbf{X})$.

这一性质的证明放在§3.5,即介绍了有效估计的概念之后, 其证明将由例3.5.11给出.

3. 相合渐近正态性

我们只考虑参数θ为一维的情形. 设$\mathscr{F}=\{f_\theta(x),\theta\in\Theta\}$ 为一概率函数族, $\Theta=(a,b)$为R_1上开区间. 设$f(x,\theta)$满足下列条件:

(1) 对一切$\theta\in\Theta$,偏导数

$$\frac{\partial\log f_\theta(x)}{\partial\theta},\quad \frac{\partial^2\log f_\theta(x)}{\partial\theta^2},\quad \frac{\partial^3\log f_\theta(x)}{\partial\theta^3}$$

存在.

(2) 存在定义于实轴上的函数$F_1(x)$、$F_2(x)$和$H(x)$, 使对一切$\theta\in\Theta$和$x\in R_1$有

$$\Big|\frac{\partial f_\theta(x)}{\partial\theta}\Big|<F_1(x),\quad \Big|\frac{\partial^2 f_\theta(x)}{\partial\theta^2}\Big|<F_2(x),\quad \Big|\frac{\partial^3 f_\theta(x)}{\partial\theta^3}\Big|\le H(x),$$

其中

$$\int_{-\infty}^{\infty}F_i(x)dx<\infty,\ i=1,2;\quad \int_{-\infty}^{\infty}H(x)f_\theta(x)dx<M,\quad \theta\in\Theta,$$

这里M与θ无关.

(3) 对一切$\theta\in\Theta$有

$$0<I(\theta)=E\Big[\Big(\frac{\partial\log f_\theta(x)}{\partial\theta}\Big)^2\Big]=\int_{-\infty}^{\infty}\Big(\frac{\partial\,\mathrm{log}f_\theta(x)}{\partial\theta}\Big)^2 f_\theta(x)dx<\infty.$$

关于极大似然估计的相合渐近正态(CAN)性, 有下列结果:

定理 2. *设$\mathbf{X}=(X_1,\cdots,X_n)$为自满足上述条件(1)-(3)的总体中抽取的简单随机样本, 且设对数似然方程*

$$\sum_{i=1}^{n}\frac{\partial\log f_\theta(x_i)}{\partial\theta}=0$$

有唯一根$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$, θ_0为真值, 则$\hat{\theta}$为θ_0的CAN估计. 即

$$\sqrt{n}(\hat{\theta}-\theta_0)\ \xrightarrow{\mathscr{L}}\ N\Big(0,\frac{1}{I(\theta_0)}\Big),$$

且$\hat{\theta}$为θ_0的弱相合估计.

证明思路 我们先不管条件, 记$L_n(\theta)=\frac{1}{n}\sum_{i=1}^n \log f_\theta(x_i), \theta\in\Theta$, 由大数律知

$$L_n(\theta)=\frac{1}{n}\sum_{i=1}^n \log f_\theta(x_i) \longrightarrow E_{\theta_0} log f_\theta(X)=L(\theta),\ n\to\infty.$$

其中$L(\theta)=E_{\theta_0}\log f_\theta(X)=\int \log f_\theta(x) f_{\theta_0}(x)dx$. 显然$L(\theta)$与样本无关, 且易知$\theta_0$为其最大值点. 这是由于

$$L(\theta)\le L(\theta_0), \forall\theta\in\Theta.$$

且等号成立当且仅当$P(f_\theta(X)=f_{\theta_0}(X))=1$. 事实上, 由于对任意的$t>0$都有$\log t\le t-1$, 因此

$$\begin{aligned}L(\theta)-L(\theta_0) &= E_{\theta_0}\log f_\theta(X)-E_{\theta_0}\log f_{\theta_0}(X)=E_{\theta_0} log\frac{f_\theta(X)}{f_{\theta_0}(X)}\\ &\le E_{\theta_0}\left[\frac{f_\theta(X)}{f_{\theta_0}(X)}-1\right]=0.\end{aligned}$$

因此, 由于$\hat\theta$为$L_n(\theta)$的最大值点, 而$L_n(\theta)$收敛到$L(\theta)$, θ_0 为$L(\theta)$的最大值点, 因此$\hat\theta$收敛到θ_0, 即$\hat\theta$是θ_0的相合估计.

对渐近正态性, 记$l(\theta)=\frac{\partial L_n(\theta)}{\partial\theta}$, 由Taylor展开式

$$0=l(\hat\theta)=l(\theta_0)+(\hat\theta-\theta_0)l'(\tilde\theta)$$

其中$\tilde\theta\in[\hat\theta,\theta_0]$. 所以

$$\sqrt{n}(\hat\theta-\theta_0)=\frac{-\sqrt{n}l(\theta_0)}{l'(\tilde\theta)}.$$

注意到由中心极限定理有

$$-\sqrt{n}l(\theta_0)=-\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\partial\log f_\theta(X_i)}{\partial\theta}\Big|_{\theta=\theta_0}\ \xrightarrow{\mathscr{L}}\ N(0,\frac{1}{I(\theta_0)})$$

其中$I(\theta)=E[\frac{\partial\log f_\theta(X)}{\partial\theta}]^2$.

另外, 由大数律知对所有的$\theta\in\Theta$有

$$l'(\theta)=\frac{1}{n}\sum_{i=1}^n\frac{\partial^2\log f_\theta(X_i)}{\partial\theta^2}\longrightarrow E\frac{\partial^2\log f_\theta(X_i)}{\partial\theta^2}=-I(\theta)$$

右边等式是因为

$$\frac{\partial^2\log f_\theta(X_i)}{\partial\theta^2}=\frac{f''}{f}-\left[\frac{f'}{f}\right]^2$$

以及$Ef''=0$.

因此,结合$\hat\theta\to\theta_0$ 以及$\tilde\theta\in[\hat\theta,\theta_0]$ 知道

$$l'(\tilde\theta)\longrightarrow -I(\theta_0).$$

从而由Slutsky定理得证.

例11. 设$X_1,\cdots,X_n$是取自正态总体$N(\mu,\sigma^2)$的样本. 证明例*3.3.4*给出的a和σ^2的MLE分别具有渐近正态性.

证 (1) 显然正态分布满足定理3.3.2的条件(1)-(3). 因此由例3.3.4已知μ 的MLE是$\hat{\mu}=\bar{X}$.由于

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

故有

$$\log f(x,\mu)=-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma^2-\frac{(x-\mu)^2}{2\sigma^2}.$$

所以

$$I_1(\theta)=E\Big[\Big(\frac{\partial\log f(x,\mu)}{\partial\mu}\Big)^2\Big]=E\Big[\Big(\frac{x-\mu}{\sigma^2}\Big)^2\Big]=\frac{1}{\sigma^2},$$

从而由定理3.3.2可知

$$\sqrt{n}(\hat{\mu}-\mu)\xrightarrow{\mathscr{L}}N(0,\sigma^2),$$

即$\hat{\mu}$近似服从$N(\mu,\sigma^2/n)$.

(2) 由例3.3.4可知σ^2的MLE是$\hat{\sigma}^2=S_n^2$. 由于

$$I_2(\theta)=E\Big[\Big(\frac{\partial\log f(x,\sigma^2)}{\partial\sigma^2}\Big)^2\Big]=E\Big[-\frac{1}{2\sigma^2}+\frac{(x-\mu)^2}{2\sigma^4}\Big]^2=\frac{1}{2\sigma^4},$$

故由定理3.3.2可知

$$\sqrt{n}(\hat{\sigma}^2-\sigma^2)\xrightarrow{\mathscr{L}}N(0,2\sigma^4),$$

即$\hat{\sigma}^2$近似服从正态分布$N(\sigma^2,2\sigma^4/n)$.