Lec6: 点估计(三)

张伟平

2011 年 3 月 24 日

1 一致最小方差无偏估计

一、引言及定义

设有一参数分布族$\mathscr{F} = \{F_\theta, \theta \in \Theta\}$,其中$\Theta$为参数空间. 设$g(\theta)$是定义在$\Theta$上的函数, $\mathbf{X} = (X_1, \cdots, X_n)$为自总体$F_\theta$ 中抽取的简单样本, $\hat{g}(\mathbf{X}) = \hat{g}(X_1, \cdots, X_n)$为$g(\theta)$ 的一个估计量, 如何评价$\hat{g}(X)$的优劣? 一般用$\hat{g}(\mathbf{X}) - g(\theta)$ 作为其偏差, 为消除$\hat{g}(\mathbf{X}) - g(\theta)$取值出现"$+, -$"可能抵消的影响, 一般用$(\hat{g}(\mathbf{X}) - g(\theta))^2$来代替. 由于这个量是随机的, 将其平均, 即计算其均值, 以得到一个整体性的指标$E_\theta(\hat{g}(\mathbf{X}) - g(\theta))^2$, 这就是估计量$\hat{g}(\mathbf{X})$的*均方误差*.

定义 1. *设$\hat{g}(\mathbf{X})$为$g(\theta)$的估计, 则称$E_\theta(\hat{g}(\mathbf{X}) - g(\theta))^2$ 为$\hat{g}(\mathbf{X})$的均方误差 (Mean Square Error,简记MSE)*

设$\hat{g}_1(\mathbf{X})$和$\hat{g}_2(\mathbf{X})$为$g(\theta)$的两个不同的估计, 若

$$E_\theta\big(\hat{g}_1(\mathbf{X}) - g(\theta)\big)^2 \le E_\theta\big(\hat{g}_2(\mathbf{X}) - g(\theta)\big)^2, \quad \text{对一切 } \theta \in \Theta,$$

则称在MSE准则下$\hat{g}_1(\mathbf{X})$优于$\hat{g}_2(\mathbf{X})$.

若存在$\hat{g}^*(\mathbf{X})$,使得对$g(\theta)$的任一估计量$\hat{g}(\mathbf{X})$,都有

$$E_\theta\big(\hat{g}^*(\mathbf{X}) - g(\theta)\big)^2 \le E_\theta\big(\hat{g}(\mathbf{X}) - g(\theta)\big)^2, \quad \text{对一切 } \theta \in \Theta,$$

则称$\hat{g}^*(X)$为$g(\theta)$的一致最小均方误差估计.

可惜的是一致最小均方误差估计常不存在. 解决这个问题的办法之一, 是把最优性准则放宽一些, 使适合这种最优性准则的估计一般能存在. 从直观上想, 在一个大的估计量的类中找一致最优的估计不存在, 把估计量的类缩小, 就有可能存在一致最优的估计量. 因此我们把估计类缩小为无偏估计类来考虑. 在无偏估计类中, 估计量的*均方误差*就变为其方差. 即当$\hat{g}(\mathbf{X})$为$g(\theta)$的无偏估计时, $MSE(\hat{g}(\mathbf{X})) = D_\theta(\hat{g}(\mathbf{X}))$, 此处$D_\theta(\hat{g}(\mathbf{X}))$表示$\hat{g}(\mathbf{X})$的方差.

存在这样的情形, 对参数$g(\theta)$它的无偏估计不存在. 请看下例:

例1. *设样本$X \sim$ 二项分布 $b(n, p)$, n已知而p未知. 令$g(p) = 1/p$,则参数$g(p)$的无偏估计不存在.*

证 采用反证法: 若不然, $g(p)$有无偏估计$\hat{g}(X)$.由于X只取$0, 1, \cdots, n$这些值, 令$\hat{g}(X)$的取值用$\hat{g}(i) = a_i$表示, $i = 0, 1, \cdots, n$.由$\hat{g}(X)$的无偏性,应有

$$E_p(\hat{g}(X)) = \sum_{i=0}^{n} a_i \binom{n}{i} p^i (1-p)^{n-i} = 1/p, \quad 0 < p < 1.$$

于是有

$$\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1}(1-p)^{n-i} - 1 = 0, \quad 0 < p < 1.$$

但上式左端是p的$n+1$次多项式, 它最多在$(0,1)$区间有$n+1$个实根, 可无偏性要求对$(0,1)$中的任一实数p上式都成立. 这个矛盾说明$g(p) = 1/p$无的偏估计不存在.

今后我们把不存在无偏估计的参数除外. 参数的无偏估计若存在, 则此参数为可估参数; 若参数函数的无偏估计存在, 则称此函数为可估函数 (Estimable function). 因此可估函数的无偏估计类是非空的.

假如可估函数的无偏估计类中的无偏估计不止一个, 怎样比较它们的优劣? 故引入下列的定义.

定义 2. *设$\mathscr{F} = \{F_\theta, \theta \in \Theta\}$是一个参数分布族, 其中$\Theta$为参数空间,$g(\theta)$为定义在$\Theta$上的可估函数. 设$\hat{g}^*(\mathbf{X}) = \hat{g}^*(X_1, \cdots, X_n)$为$g(\theta)$的一个无偏估计, 若对$g(\theta)$的任一无偏估计$\hat{g}(\mathbf{X}) = \hat{g}(X_1, \cdots, X_n)$,都有*

$$D_\theta(\hat{g}^*(\mathbf{X})) \leq D_\theta(\hat{g}(\mathbf{X})), \quad \text{对一切}\ \theta \in \Theta,$$

则称$\hat{g}^*(\mathbf{X})$是$g(\theta)$的一致最小方差无偏估计 (Uniformly Minimum Variance Unbiased Estimation, 简记为UMVUE).

对给定参数分布族, 如何寻找可估参数的UMVUE呢? 本节以下将介绍两种方法: 零无偏估计法和充分完全统计量法, 下一节的Cramer-Rao不等式法也是寻找UMVUE的一种方法.

在前面我们曾介绍过Rao-Blackwell定理, 这一定理提供了一个改进无偏估计的方法, 它在本节以下寻找UMVUE中,起到简化问题的作用. 重新表述此定理如下

定理 1 (Rao-Blackwell)**.** *设$T = T(\mathbf{X})$是一个充分统计量, 而$\hat{g}(\mathbf{X})$是$g(\theta)$ 的一个无偏估计, 则*

$$h(T) = E(\hat{g}(\mathbf{X})|T)$$

是$g(\theta)$的无偏估计, 并且

$$D_\theta(h(T)) \leq D_\theta(\hat{g}(\mathbf{X})), \quad \text{一切}\ \theta \in \Theta, \tag{1.1}$$

其中等号当且仅当$P_\theta\big(\hat{g}(\mathbf{X}) = h(T)\big) = 1$,即$\hat{g}(\mathbf{X}) = h(T)$, a.e. P_θ成立.

这个引理提供了一个改进无偏估计的方法, 即一个无偏估计$\hat{g}(\mathbf{X})$ 对充分统计量$T(\mathbf{X})$的条件期望$E\{\hat{g}(\mathbf{X})|T\}$ 将能导出一个新的无偏估计, 且它的方差不会超过原估计量$\hat{g}(\mathbf{X})$的方差. 若原估计$\hat{g}(\mathbf{X})$不是$T(\mathbf{X})$的函数, 则新的无偏估计$E(\hat{g}(\mathbf{X})|T)$一定比原估计$\hat{g}(\mathbf{X})$具有更小的方差. 这个定理还表明一致最小方差无偏估计一定是充分统计量的函数, 否则可以通过充分统计量构造出一个具有更小方差的无偏估计来.

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. 显然,$X_1$是$p$的一个无偏估计, $T(\mathbf{X}) = \sum_{i=1}^n X_i$是$p$ 的充分统计量, 试利用$T = T(\mathbf{X})$构造一个具有比X_1方差更小的无偏估计.*

解 由引理3.4.1可知,容易构造p的一个无偏估计如下：

$$h(T) = E(X_1|T = t) = 1 \cdot P(X_1 = 1|T = t) + 0 \cdot P(X_1 = 0|T = t)$$

$$
\begin{aligned}
&= \frac{P(X_1=1,T=t)}{P(T=t)} = \frac{P(X_1=1,X_2+\cdots+X_n=t-1)}{P(T=t)} \\
&= \frac{p\cdot\binom{n-1}{t-1}p^{t-1}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \frac{t}{n} = \bar{x}.
\end{aligned}
$$

显然样本均值$h(T)=\bar{X}$的方差为$p(1-p)/n$, 而X_1的方差为$p(1-p)$,当$n\geq 2$时$\bar{X}$的方差更小.

二、零无偏估计法

本段介绍一个一般性的定理, 用以判断某一估计量是否为UMVUE.

定理 2. 设$\hat{g}(\mathbf{X})$是$g(\theta)$的一个无偏估计, $D_\theta(\hat{g}(\mathbf{X}))<\infty$,对任何$\theta\in\Theta$. 且对任何满足条件"$E_\theta l(\mathbf{X})=0$,对一切 $\theta\in\Theta$" 的统计量$l(\mathbf{X})$,必有

$$
Cov_\theta\big(\hat{g}(\mathbf{X}),l(\mathbf{X})\big)=E_\theta\big[\hat{g}(\mathbf{X})\cdot l(\mathbf{X})\big]=0,\quad \text{一切 } \theta\in\Theta, \tag{1.2}
$$

则$\hat{g}(\mathbf{X})$是$g(\theta)$的*UMVUE.*

从形式上看, 条件"$E_\theta l(\mathbf{X})=0$,对一切 $\theta\in\Theta$"可解释为"$l(\mathbf{X})$ 是零的无偏估计", 由此得到求UMVUE的方法之一的名称"零无偏估计法".

此定理还可进一步加强为: 设$\hat{g}(\mathbf{X})$为$g(\theta)$的一个无偏估计, $D_\theta(\hat{g}(\mathbf{X}))<\infty$,一切 $\theta\in\Theta$, 则$\hat{g}(\mathbf{X})$是$g(\theta)$的UMVUE的充分必要条件是: 对任何满足条件"$E_\theta l(\mathbf{X})=0$,对一切 $\theta\in\Theta$"的统计量$l(\mathbf{X})$, 必有(1.2)成立.

证 设$\hat{g}_1(\mathbf{X})$为$g(\theta)$的任一无偏估计. 记$l(\mathbf{X})=\hat{g}_1(\mathbf{X})-\hat{g}(\mathbf{X})$ 为零的无偏估计, 由于(1.2)式成立, 因而

$$
\begin{aligned}
&D_\theta\big(\hat{g}_1(\mathbf{X})\big)=D_\theta\big[\hat{g}(\mathbf{X})+l(\mathbf{X})\big] \\
&=D_\theta\big(\hat{g}(\mathbf{X})\big)+D_\theta\big(l(\mathbf{X})\big)+2Cov_\theta\big(\hat{g}(\mathbf{X}),l(\mathbf{X})\big) \\
&=D_\theta\big(\hat{g}(\mathbf{X})\big)+D_\theta\big(l(\mathbf{X})\big)\geq D_\theta\big(\hat{g}(\mathbf{X})\big).
\end{aligned}
$$

另一方面, 如果$\delta=\hat{g}(X)$为$g(\theta)$的UMVUE, 则对任意的零无偏估计量$l(x)$, 令

$$\delta'=\delta+\lambda l(X)$$

则δ'仍为$g(\theta)$的无偏估计, 且

$$Var(\delta')=Var(\delta+\lambda l(X))=\lambda^2Var(l(X))+2\lambda Cov(\delta,l(X))+Var(\delta)\geq Var(\delta).$$

对所有的λ成立. 等价于

$$\lambda^2Var(l(X))+2\lambda Cov(\delta,l(X))\geq 0,\ \forall\ \lambda.$$

左边有两个根$\lambda=0$和$\lambda=-Cov(\delta,l(X))/Var(l(X))$, 推出$Cov(\delta,l(X))=0$.

这就证明了所要的结果.

从定理的内容看, 它是一个验证某个特定的估计量$\hat{g}(\mathbf{X})$为UMVUE的方法. 至于这个特定的估计$\hat{g}(\mathbf{X})$从何而来, 定理3.4.1不能提供任何帮助, 它不是UMVUE的构造性定理. $\hat{g}(\mathbf{X})$可以从直观的想法提出, 如由矩估计或极大似然估计等方法获得, 然后利用此定理验证它是否

为$g(\theta)$的UMVUE. 条件(1.2)的验证也不容易, 因为零无偏估计很多. 下面的几个例子说明, 本章前面一些例子中提到的几个常用估计, 都可以用此法验证其为UMVUE.

例　求上例中$g(p)=p$的UMVUE.

解　由上例已知$T=\sum_{i=1}^n X_i$为充分统计量, $\hat{g}(\mathbf{X})=T/n$,故只要验证它满足定理的条件. 显然$\hat{g}(\mathbf{X})$是p的无偏估计. 且$D_\theta(\hat{g}(\mathbf{X}))=p(1-p)/n<\infty$,对一切$0<p<1$.现设$l=l(T)$为任一零无偏估计, 并记$a_i=l(i)$, $i=0,1,2,\cdots,n$, 则因$T\sim b(n,p)$,故有

$$E_p l(T)=\sum_{i=0}^n a_i\binom{n}{i}p^i(1-p)^{n-i}=0,\quad 0<p<1.$$

约去因子$(1-p)^n$,并记$\varphi=p/(1-p)$ (φ 取值$(0,\infty)$), 将上式改写为

$$\sum_{i=0}^n a_i\binom{n}{i}\varphi^i=0,\quad \text{一切}\ 0<\varphi<\infty.$$

上式左边是φ的多项式, 要使其为0, 必有$a_i\binom{n}{i}=0$,即$a_i=0$, $i=1,2,\cdots,n$.故$l(T)$在其定义域中处处为0, 因而有$l(T)\equiv 0$.从而有

$$Cov_p\big(\hat{g},l(T)\big)=E\big(\hat{g}\cdot l(T)\big)=0.$$

即定理条件成立, 故$\hat{g}(\mathbf{X})=\bar{X}=T/n$为$p$的UMVUE.

例　设$\mathbf{X}=(X_1,\cdots,X_n)$为从指数分布$EP(\lambda)$ 中抽取的简单样本, 求总体均值$g(\lambda)=1/\lambda$的UMVUE.

解　由于在指数分布族中$T=\sum_{i=1}^n X_i$为λ的充分统计量, 则$T\sim\Gamma(n,\lambda)$,其密度函数为

$$\phi(t,\lambda)=\begin{cases}\frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t} & \text{当}\ t>0\\ 0 & \text{当}\ t\le 0,\end{cases}$$

其中$\lambda>0$.取$\hat{g}(\mathbf{X})=T/n$, 显然$E(\hat{g}(\mathbf{X}))=1/\lambda$,即$\hat{g}(\mathbf{X})$ 为$g(\lambda)=1/\lambda$的无偏估计, 且$D_\lambda(\hat{g}(\mathbf{X}))=1/(n\lambda^2)<\infty$.现设$l=l(T)$为任一零无偏估计, 故有

$$El(T)=\int_0^\infty l(t)\frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}dt=0,$$

即$\int_0^\infty l(t)t^{n-1}e^{-\lambda t}dt=0$.两边对$\lambda$求导得

$$\int_0^\infty l(t)t^n e^{-\lambda t}dt=0,$$

此式等价于$E_\lambda\big[T/n\cdot l(T)\big]=E_\lambda(\hat{g}\cdot l(T))=Cov_\lambda(\hat{g}\cdot l)=0$,即定理条件成立. 因此$\hat{g}(\mathbf{X})=T/n$为$g(\lambda)=1/\lambda$的UMVUE.

例　设$\mathbf{X}=(X_1,\cdots,X_n)$为从均匀分布$U(0,\theta)$中抽取的简单样本, 求$\theta$的UMVUE.

解　由$T=T(\mathbf{X})=X_{(n)}$是参数θ的充分统计量, 又知$\hat{g}(\mathbf{X})=\frac{n+1}{n}T$是$\theta$的无偏估计, 且$D_\theta(\hat{g}(T))=\frac{1}{n(n+2)}\theta^2<\infty$. 现设$l(T)$为任一零无偏估计, T的密度函数如(**??**)所示, 因此有

$$E_\theta l(T)=\int_0^\theta l(t)\cdot\frac{nt^{n-1}}{\theta^n}dt=0,\quad \text{一切}\ \theta>0,$$

于是有

$$\int_0^\theta l(t)t^{n-1}dt=0, \quad 一切\ \theta>0.$$

将上式两边对θ 求导得$l(\theta)\theta^{n-1}=0$,一切 $\theta>0$.故有$l(\theta)\equiv 0$ 对一切$\theta>0$.可见$Cov(\hat{g},l(T))=E(\hat{g}\cdot l(T))=0$,即定理条件成立. 因此$\hat{g}(\mathbf{X})=\frac{n+1}{n}X_{(n)}$为$g(\theta)=\theta$的UMVUE.

例 设$\mathbf{X}=(X_1,\cdots,X_n)$为从正态分布$N(a,\sigma^2)$中抽取的简单随机样本, 求$a$和$\sigma^2$的UMVUE.

解 由$T=(T_1,T_2)$为$\theta=(a,\sigma^2)$ 的充分统计量.其中$T_1=\bar{X}$, $T_2=\sum_{i=1}^n(X_i-\bar{X})^2$, 又$T_1$和$T_2$独立, 且$T_1\sim N\left(a,\sigma^2/n\right)$, $T_2/\sigma^2\sim\chi^2_{n-1}$. 因此$(T_1,T_2)$的联合密度为

$$\begin{aligned}f(t_1,t_2)=&\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}e^{-\frac{n(t_1-a)^2}{2\sigma^2}}\cdot\left(2^{\frac{n-1}{2}}\Gamma\Big(\frac{n-1}{2}\Big)\sigma^{n-1}\right)^{-1}t_2^{\frac{n-3}{2}}e^{-\frac{t_2}{2\sigma^2}},\\ &当\ -\infty<t_1<+\infty,\ t_2>0;\quad 其它处为\ 0.\end{aligned} \tag{1.3}$$

先考虑a的UMVUE, 令$\hat{g}_1(\mathbf{X})=T_1$,显然$\hat{g}_1(\mathbf{X})$ 为$g_1(\theta)=a$的无偏估计, 且$D_\theta(\hat{g}_1(\mathbf{X}))=\sigma^2/n<\infty$.现设$l(T)=l(T_1,T_2)$为任一零无偏估计, 则有

$$E_\theta(l(T))=\int_0^\infty\int_{-\infty}^\infty l(t_1,t_2)f_\theta(t_1,t_2)dt_1dt_2=0,$$

此处$-\infty<a<\infty$, $\sigma>0$.将上式两边对a求导数, 得

$$\int_0^\infty\int_{-\infty}^\infty l(t_1,t_2)(t_1-a)\cdot t_2^{\frac{n-3}{2}}\exp\Big\{-\frac{1}{2\sigma^2}\left[n(t_1-a)^2+t_2\right]\Big\}dt_1dt_2=0,$$

按(1.3),此即

$$Cov_\theta\big(\hat{g}_1,l(T)\big)=E_\theta\big(\hat{g}_1\cdot l(T_1,T_2)\big)=0,\ -\infty<a<+\infty,\ \sigma>0,$$

故定理3.4.1的条件满足. 所以$\hat{g}_1(\mathbf{X})=T_1$为$g_1(\theta)=a$的UMVUE.

同理可验证$T_2=S^2$为$g_2(\theta)=\sigma^2$的UMVUE, 这一验证留为练习.

三、充分完全统计量法

下列定理给出的求UMVUE的方法, 即充分完全统计量法是由E.L. Lehmann和H. Scheffe给出的, 完全统计量的概念也是由他们在1950年提出的.

定理 (Lehmann-Scheffe定理) 设$T(\mathbf{X})$为一个充分完全统计量, 若$\hat{g}(T(\mathbf{X}))$为$g(\theta)$ 的一个无偏估计, 则$\hat{g}(T(\mathbf{X}))$是$g(\theta)$ 的唯一的UMVUE (唯一性是在这样的意义下: 若$\hat{g}$和$\hat{g}_1$都是$g(\theta)$的UMVUE, 则$P_\theta(\hat{g}\neq\hat{g}_1)=0$,对一切$\theta\in\Theta$).

证 先证唯一性. 设$\hat{g}_1(T(\mathbf{X}))$为$g(\theta)$的任一无偏估计, 令$\delta(T(\mathbf{X}))=\hat{g}(T(\mathbf{X}))-\hat{g}_1(T(\mathbf{X}))$, 则$E_\theta\delta(T(\mathbf{X}))=E_\theta\hat{g}(T(\mathbf{X}))-E_\theta\hat{g}_1(T(\mathbf{X}))=0$,对一切$\theta\in\Theta$. 由$T(\mathbf{X})$为完全统计量, 可知$\delta(T(\mathbf{X}))=0$, $a.e.\ P_\theta$.即$\hat{g}(T(\mathbf{X}))=\hat{g}_1(T(\mathbf{X}))$, $a.e.\ P_\theta$, 故唯一性成立.

设$\varphi(\mathbf{X})$为$g(\theta)$的任一无偏估计. 令$h(T(\mathbf{X}))=E(\varphi(\mathbf{X})|T)$,由$T(\mathbf{X})$为充分统计量, 故知$h(T(\mathbf{X}))$与$\theta$无关, 因此是统计量, 可知

$$\begin{aligned}E_\theta(h(T(\mathbf{X})))&=g(\theta), &&一切\ \theta\in\Theta,\\ D_\theta(h(T(\mathbf{X})))&\le D_\theta(\varphi(\mathbf{X})), &&一切\ \theta\in\Theta.\end{aligned}$$

由唯一性得$\hat{g}(T(\mathbf{X}))=h(T(\mathbf{X}))$, $a.e.\ P_\theta$故有

$$D_\theta(\hat{g}(T(\mathbf{X}))\le D_\theta(\varphi(\mathbf{X})),\quad \text{一切}\ \theta\in\Theta,$$

所以$\hat{g}(T(\mathbf{X}))$ 为$g(\theta)$的UMVUE, 且唯一.

推论 设样本$\mathbf{X}=(X_1,\cdots,X_n)$的分布为指数族

$$f(\mathbf{x},\theta)=C(\theta)\exp\Big\{\sum_{j=1}^k\theta_jT_j(\mathbf{x})\Big\}h(\mathbf{x}),\quad \theta=(\theta_1,\cdots,\theta_k)\in\Theta.$$

令$T(\mathbf{X})=(T_1(\mathbf{X}),\cdots,T_k(\mathbf{X}))$, 若自然参数空间$\Theta$作为$R_k$的子集有内点, 且$h(T(\mathbf{X}))$为$g(\theta)$的无偏估计, 则$h(T(\mathbf{X}))$为$g(\theta)$的唯一的UMVUE.

证 在推论的条件下, 由指数族的性质可知$T(\mathbf{X})$为充分完全统计量. 故由*Lehmann-Scheffe*定理 (简记为L-S定理), 得知$h(T(\mathbf{X}))$为$g(\theta)$的唯一的UMVUE.

例 证明两点分布的p的无偏估计$\hat{g}(\mathbf{X})=T/n=\bar{X}$ 为p的UMVUE.

证 由因子分解定理可知$T(\mathbf{X})=\sum\limits_{i=1}^n X_i$为两点分布$b(1,p)$中参数$p$ 的充分统计量, 由例2.8.1知$T(\mathbf{X})$也是完全统计量, 故$\hat{g}(\mathbf{X})=T/n=\bar{X}$是充分完全统计量$T(\mathbf{X})$的函数, 且$E_p\,\hat{g}(\mathbf{X})=p$,对$0<p<1$. 因此由L-S定理可知$\hat{g}(\mathbf{X})$为$p$的唯一的UMVUE.

例 在上例中, 已知$T=\sum\limits_{i=1}^n X_i$服从二项分布$b(n,p)$, 且$T(\mathbf{X})$为充分完全统计量, 求$g(p)=p(1-p)$的UMVUE.

解 设$\delta(T)$为$g(p)=p(1-p)$的一个无偏估计, 要导出$\delta(T)$的表达式. 按无偏估计的定义及$T\sim b(n,p)$,可得

$$\sum_{t=0}^n\binom{n}{t}\delta(t)p^t(1-p)^{n-t}=p(1-p),\quad \text{一切}\ 0<p<1.$$

令$\rho=p/(1-p)$, 故有$p=\rho/(1+\rho)$, $1-p=1/(1+\rho)$,将它们代入上式得

$$\sum_{t=0}^n\binom{n}{t}\delta(t)\rho^t=\rho(1+\rho)^{n-2},\quad 0<\rho<\infty.$$

将$\rho(1+\rho)^{n-2}$ 展开得$\rho(1+\rho)=\sum\limits_{l=0}^{n-2}\binom{n-2}{l}\rho^{l+1}=\sum\limits_{t=1}^{n-1}\binom{n-2}{t-1}\rho^t$,将其代入上式右边得

$$\sum_{t=0}^n\binom{n}{t}\delta(t)\rho^t=\sum_{t=1}^{n-1}\binom{n-2}{t-1}\rho^t,\quad 0<\rho<\infty.$$

上式两边为ρ的多项式, 比较其系数得

$$\begin{aligned}&\delta(t)=0,\quad \text{当}\ t=0,n;\\ &\delta(t)=\binom{n-2}{t-1}\Big/\binom{n}{t}=\frac{t(n-t)}{n(n-1)},\quad \text{当}\ t=1,2,\cdots,n-1.\end{aligned}$$

综合上述两式得

$$\delta(T)=\frac{T(n-T)}{n(n-1)},\quad t=0,1,\cdots,n$$

为$g(p) = p(1-p)$的无偏估计, 它又是充分完全统计量$T = \sum_{i=1}^n X_i$的函数, 由L-S定理可知$\delta(T)$为$g(p)$的UMVUE.

例 设$\mathbf{X} = (X_1, \cdots, X_n)$为从Poisson分布$P(\lambda)$ 中抽取的简单随机样本, 求 (1) $g_1(\lambda) = \lambda$; (2) $g_2(\lambda) = \lambda^r$, $r > 0$ 为自然数; (3) $g_3(\lambda) = P_\lambda(X_1 = x)$的UMVUE.

解 由§2.7和§2.8可知$T(\mathbf{X}) = \sum_{i=1}^n X_i$ 为关于Poisson分布的充分完全统计量.

(1) 令$\hat{g}_1(T) = T(\mathbf{X})/n$, $E(\hat{g}_1(T)) = E(\bar{X}) = \lambda$, 故$\hat{g}_1(T)$是分完全统计量$T$的函数，且是$\lambda$的无偏估计,故由L-S定理可知$\hat{g}_1(T)$是$\lambda$的UMVUE.

(2) 由于$T(\mathbf{X}) = \sum_{i=1}^n X_i \sim P(n\lambda)$,令$\delta(T)$为$g_2(\lambda) = \lambda^r$的无偏估计, 故有$E_\lambda \delta(T) = g_2(\lambda)$,即

$$\sum_{t=0}^{\infty} \delta(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!} = \lambda^r.$$

此式等价于

$$\sum_{t=0}^{\infty} \delta(t) \frac{n^t \lambda^t}{t!} = \lambda^r e^{n\lambda}.$$

将上式右边作展开得

$$\lambda^r e^{n\lambda} = \sum_{l=0}^{\infty} \frac{n^l \lambda^{l+r}}{l!} = \sum_{t=r}^{\infty} \frac{n^{t-r} \lambda^t}{(t-r)!}.$$

将其代入上式右边得

$$\sum_{t=0}^{\infty} \delta(t) \frac{n^t \lambda^t}{t!} = \sum_{t=r}^{\infty} \frac{n^{t-r} \lambda^t}{(t-r)!}.$$

上述等式两边是λ的幂级数, 比较其系数得

$$\begin{aligned} &\delta(t) = 0, \quad \text{当 } t = 0, 1, \cdots, r-1, \\ &\delta(t) = \frac{t!\, n^{t-r}}{(t-r)! n^t} = \frac{t(t-1)\cdots(t-r+1)}{n^r}, \quad \text{当 } t = r, r+1, \cdots \end{aligned}$$

综合上述两式得

$$\delta(T) = \frac{T(T-1)\cdots(T-r+1)}{n^r}, \quad T = 0, 1, 2, \cdots$$

为$g_2(\lambda) = \lambda^r$的无偏估计, $\delta(T)$是充分完全统计量T的函数, 故由L-S定理可知$\delta(T)$为$g_2(\lambda)$的UMVUE.

(3) 由$P_\lambda(X_1 = x) = e^{-\lambda}\lambda^x/x!$,可见它是参数λ的函数,故可用$g_3(\lambda)$表示. 令$\varphi(X_1) = I_{[X_1 = x]}$, 则$E_\lambda[\varphi(X_1)] = P_\lambda(X_1 = x)$.因此$\varphi(X_1)$ 为$g_3(\lambda)$的无偏估计, 注意到$T = \sum_{i=1}^n X_i \sim P(n\lambda)$和$\sum_{i=2}^n X_i \sim P((n-1)\lambda)$,故有

$$\begin{aligned} \delta_1(T) = \delta_1(T(\mathbf{X})) &= E(\varphi(X_1)|T = t) = \frac{P_\lambda(X_1 = x, T = t)}{P_\lambda(T = t)} \\ &= \frac{P_\lambda(X_1 = x) P_\lambda(X_2 + \cdots + X_n = t - x)}{P_\lambda(X_1 + \cdots + X_n = t)} = \frac{(n-1)^{t-x} t!}{n^t (t-x)! x!} \\ &= \binom{t}{x} \frac{(n-1)^{t-x}}{n^t} = \binom{t}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{t-x}, \quad t \geq x \end{aligned}$$

$\delta_1(T)$为$g_3(\lambda)$的无偏估计, 它又是充分完全统计量$T(\mathbf{X})$的函数, 所以

$$\delta_1(T(\mathbf{X})) = \binom{T}{x} \left(\frac{1}{n}\right) \left(1 - \frac{1}{n}\right)^{T-x}$$

为$g_3(\lambda)$的UMVUE.

例　设$\mathbf{X}=(X_1,\cdots,X_n)$为从指数分布$EP(\lambda)$中抽取的简单随机样本, 求$g(\lambda)=\lambda$的UMVUE.

解　由$T(\mathbf{X})=\sum\limits_{i=1}^{n}X_i$的充分完全统计量, 以及$T(\mathbf{X})\sim\Gamma(n,\lambda)$,即参数为$n$和$\lambda$的Gamma分布, 故有

$$E\Big(\frac{1}{T}\Big)=\int_0^{\infty}\frac{1}{t}\cdot\frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}dt=\frac{\lambda}{n-1}.$$

因此$\hat{g}(T(\mathbf{X}))=(n-1)/T(\mathbf{X})$ 为λ的无偏估计, 由L-S定理可知它是λ的UMVUE.

例　设$\mathbf{X}=(X_1,\cdots,X_n)$为从正态分布$N(a,\sigma^2)$中抽取的简单随机样本, 记$\theta=(a,\sigma^2)$. (1)求$a$和$\sigma^2$的UMVUE, (2)求$g(\theta)=\sigma^r$的UMVUE.

解　由$T(\mathbf{X})=(T_1(\mathbf{X}),T_2(\mathbf{X}))$, 其中$T_1(\mathbf{X})=\bar{X}$, $T_2(\mathbf{X})=\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$,为充分完全统计量.

(1)由于$\hat{g}_1(\mathbf{X})=\bar{X}=T_1$和$\hat{g}_2(\mathbf{X})=T_2/(n-1)$分别为$a$和$\sigma^2$的无偏估计, 它们又是充分完全统计量, 故由L-S定理可知它们分别是a和σ^2的UMVUE.

(2)由于$T_2/\sigma^2\sim\chi^2_{n-1}$,故$\sigma^r$的无偏估计与$T_2$的幂函数有关. 先计算下式:

$$\begin{aligned}E\left(\frac{T_2}{\sigma^2}\right)^{r/2}&=\frac{1}{\sigma^r}E\left(T_2^{\frac{r}{2}}\right)=\int_0^{\infty}t^{\frac{r}{2}}\cdot\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}t^{\frac{n-1}{2}-1}e^{-\frac{t}{2}}dt\\&=\frac{2^{r/2}\Gamma\left(\frac{n+r-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\triangleq\frac{1}{K_{n-1,r}}.\end{aligned}$$

由上式可知

$$E(K_{n-1,r}\cdot T_2^{\frac{r}{2}})=\sigma^r.$$

因此估计量

$$\hat{g}_3(T(\mathbf{X}))=K_{n-1,r}T_2^{r/2}=\frac{\Gamma\left(\frac{n-1}{2}\right)}{2^{r/2}\Gamma\left(\frac{n+r-1}{2}\right)}T^{r/2}$$

是σ^r的无偏估计, 又是充分完全统计量$T=(T_1,T_2)$的函数, 故由L-S定理可知它是$g(\theta)=\theta^r$的UMVUE.

例　用Lehmann-Scheffe定理再考虑均匀分布$U(0,\theta)$, $\theta>0$, 中的参数θ的UMVUE.

解　由$T(\mathbf{X})=\max(X_1,\cdots,X_n)=X_{(n)}$ 为充分完全统计量, 和$\hat{g}(T(\mathbf{X}))=\frac{n+1}{n}T(\mathbf{X})$为$\theta$的无偏估计, 故由L-S定理立得$\hat{g}(T(\mathbf{X}))$ 为θ的UMVUE.可见此处证明要简单的多.

例　考虑均匀分布$U(0,\theta)$, $\theta>1$, 中的参数θ的UMVUE.

解　由因子分解定理知道$T(\mathbf{X})=\max(X_1,\cdots,X_n)=X_{(n)}$为充分统计量，但它并不为完全统计量。实际上对任意满足$El(T)=0$,即

$$\int_0^1 l(t)t^{n-1}dt+\int_1^{\theta}l(t)t^{n-1}dt=0\qquad(*)$$

因此可取

$$l_0(t)=\begin{cases}(n+1)t-n, & 0<t<1\\ 0,1<t<\theta\end{cases}$$

因而存在非恒为零的函数$l(T) = l_0(T)$满足$El(T) = 0$.所以T不是完全统计量, 从而不能利用Lehmann-Scheffe定理验证基于T的无偏估计为UMVUE, 但是我们可以使用零无偏方法. 注意到欲使$Eg(T)l(T) = 0$, 即

$$\int_0^1 g(t)l(t)t^{n-1}dt + \int_1^\theta g(t)l(t)t^{n-1}dt = 0$$

结合$(*)$式从而可取

$$g(t) = \begin{cases} c, & 0 < t < 1 \\ bt, 1 < t < \theta \end{cases}$$

由无偏性有

$$Eg(T) = \int_0^1 cf_T(t)dt + \int_1^\theta btf_T(t)dt = \frac{c}{\theta^n} + \frac{b}{\theta^n}\frac{n}{n+1}(\theta^{n+1} - 1) = \theta$$

从而可取

$$c = 1, \qquad b = \frac{n+1}{n}$$

因此

$$g(T) = \begin{cases} 1, & 0 < T < 1 \\ \frac{n+1}{n}T, 1 < T < \theta \end{cases}$$

为参数θ的UMVUE.

注 可以证明统计量

$$\tilde{T} = \begin{cases} 1, & 0 < T < 1 \\ T, 1 < T < \theta \end{cases}$$

为完全统计量.

2 Cramer-Rao不等式

一、引言

Cramer-Rao不等式(简称C-R不等式)是判别一个无偏估计量是否为UMVUE的方法之一. 这一方法的思想如下: 设$\mathscr{U}_g$是$g(\theta)$的一切无偏估计构成的类. $\mathscr{U}_g$中的估计量的方差有一个下界, 这个下界称为C-R下界. 因此, 如果$g(\theta)$的一个无偏估计$\hat{g}$的方差达到这个下界, 则$\hat{g}$ 就是$g(\theta)$的一个UMVUE, 当然样本分布族和$\hat{g}$要满足一定的正则条件. 这个不等式是由C.R. Rao和H. Cramer在1945和1946年分别证明的. 以后一些统计学者将条件作了一些改进和精确化, 但结果的基本形式并无重大变化.

这一方法的缺陷是: 由于C-R不等式确定的下界常比真下界为小. 在一些场合, 虽然$g(\theta)$的UMVUE $\hat{g}$存在, 但其方差大于C-R下界. 在这一情况下, 用C-R不等式就无法判定$g(\theta)$的UMVUE存在. 因此这一方法的适用范围不广. C-R不等式除了用于判别$g(\theta)$的UMVUE之外, 它在数理统计理论上还有其它的用处, 如估计的效率和有效估计的概念以及Fisher信息量都与之有关.

C-R不等式成立需要样本分布族满足一些正则条件, 适合这些条件的分布族称为C-R正则分布族, 下面给出其定义.

定义 若单参数概率函数族$\mathscr{F} = \{f(x,\theta), \theta \in \Theta\}$满足下列条件:

(i)参数空间Θ是直线上的某个开区间;

(ii)导数$\frac{\partial f(x,\theta)}{\partial \theta}$对一切$\theta \in \Theta$存在;

(iii)概率函数的支撑集$\{x : f(x,\theta) > 0\}$与θ无关;

(iv)概率函数$f(x,\theta)$的积分与微分运算可交换, 即

$$\frac{\partial}{\partial \theta}\int f(x,\theta)dx = \int \frac{\partial}{\partial \theta} f(x,\theta)dx.$$

若$f(x,\theta)$为离散随机变量的概率分布, 上述条件改为无穷级数和微分运算可交换;

(v)下列数学期望存在, 且

$$0 < I(\theta) = E_\theta\left[\frac{\partial \log f(X,\theta)}{\partial \theta}\right]^2 < \infty,$$

则称该分布族为*C-R正则分布族*. 其中(i)-(v)称为*C-R正则条件*. $I(\theta)$称为该分布的*Fisher信息量*(或称为Fisher信息函数).

二、单参数C-R不等式

1. C-R不等式及例

定理 设$\mathscr{F} = \{f(x,\theta), \theta \in \Theta\}$是C-R正则分布族, $g(\theta)$ 是定义于参数空间Θ上的可微函数. 设$\mathbf{X} = (X_1, \cdots, X_n)$ 是由总体$f(x,\theta) \in \mathscr{F}$中抽取的简单随机样本, $\hat{g}(\mathbf{X})$ 是$g(\theta)$的任一无偏估计, 且满足下列条件:

(vi) 积分

$$\int \cdots \int \hat{g}(\mathbf{x}) f(\mathbf{x},\theta) d\mathbf{x}$$

可在积分号下对θ求导数, 此处$d\mathbf{x} = dx_1 \cdots dx_n$, 则有

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{(g'(\theta))^2}{nI(\theta)}, \quad 一切\ \theta \in \Theta. \tag{2.1}$$

特别当$g(\theta) = \theta$时(2.1)即为

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{1}{nI(\theta)}, \quad 一切\ \theta \in \Theta. \tag{2.2}$$

当$f(x,\theta)$为离散$r.v.\ X$的概率分布时, (2.1)变为

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{[g'(\theta)]^2}{n\sum\limits_i \left\{\left[\frac{\partial \log f(x_i,\theta)}{\partial \theta}\right]^2 f(x_i,\theta)\right\}}, \quad 一切\ \theta \in \Theta. \tag{2.3}$$

证 由于$X_1, \cdots, X_n$为i.i.d.样本, 故有$f(\mathbf{x},\theta) = \prod\limits_{i=1}^n f(x_i,\theta)$. 记

$$S(\mathbf{x},\theta) = \frac{\partial \log f(\mathbf{x},\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(x_i,\theta)}{\partial \theta},$$

因此由正则条件(iii)和(iv)可知

$$\begin{aligned}E\{S(\mathbf{X},\theta)\} &= \sum_{i=1}^{n} E\Big\{\frac{\partial \log f(X_i,\theta)}{\partial\theta}\Big\} = \sum_{i=1}^{n}\int \frac{1}{f(x_i,\theta)}\frac{\partial f(x_i,\theta)}{\partial\theta}\cdot f(x_i,\theta)dx_i \\ &= \sum_{i=1}^{n}\int \frac{\partial f(x_i,\theta)}{\partial\theta}dx_i = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\int f(x_i,\theta)dx = 0.\end{aligned}$$

由$\hat{g}(\mathbf{x})$为$g(\theta)$的无偏估计和正则条件(v)和(vi)可知

$$\begin{aligned}&Cov(\hat{g}(\mathbf{X}),S(\mathbf{X},\theta)) = E\{\hat{g}(\mathbf{X})\cdot S(\mathbf{X},\theta)\}\\ &= \int\cdots\int \hat{g}(\mathbf{x})\frac{\partial f(\mathbf{x},\theta)}{\partial\theta}d\mathbf{x}\\ &= \frac{\partial}{\partial\theta}\int\cdots\int \hat{g}(\mathbf{x})f(\mathbf{x},\theta)d\mathbf{x} = \frac{\partial g(\theta)}{\partial\theta} = g'(\theta),\\ &D_\theta(S(\mathbf{X},\theta)) = \sum_{i=1}^{n} D_\theta\Big\{\frac{\partial \log f(X_i,\theta)}{\partial\theta}\Big\}\\ &= \sum_{i=1}^{n} E\Big\{\frac{\partial \log f(X_i,\theta)}{\partial\theta}\Big\}^2 = nI(\theta). \end{aligned} \tag{2.4}$$

由Cauchy-Schwartz不等式, 得

$$D_\theta\{\hat{g}(\mathbf{X})\}\cdot D_\theta\{S(\mathbf{X},\theta)\} \geq [Cov(\hat{g}(\mathbf{X}),S(\mathbf{X},\theta))]^2 = [g'(\theta)]^2$$

将(2.4)代入上式得

$$D_\theta[\hat{g}(\mathbf{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)},$$

定理得证.

不等式(2.1)称为**Cramer-Rao不等式**, 简称*C-R*不等式. 因此*C-R*不等式可视为验证某一无偏估计是否为UMVUE的方法. 用*C-R*不等式寻找$g(\theta)$的UMVUE时, 首先要验证样本分布族是否满足正则条件(i)-(v)和(vi), 然后再计算Fisher信息量 $I(\theta)$和无偏估计$\hat{g}(\mathbf{X})$的方差$D_\theta(\hat{g}(\mathbf{X}))$,看其是否达到C-R下界. 验证正则条件(i)-(v)和条件(vi)十分麻烦. 但幸运的是对指数族上述正则条件(i)-(v)皆成立.

但要注意一点的是: 若$D_\theta(\hat{g}(\mathbf{X}))$达不到C-R下界, 并不能得出结论说$g(\theta)$的UMVUE就不存在, 而只能说用此法无法判别. 存在这样的例子, $\hat{g}(\mathbf{X})$是$g(\theta)$的UMVUE, 但其方差大于C-R下界. 后面将给出这一例子.

例 设$\mathbf{X} = (X_1,\cdots,X_n)$为从两点分布$b(1,p)$中抽取的简单样本, 证明样本均值$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$为$p$的UMVUE.

证 设随机变量$X \sim b(1,p)$,则其概率分布为$f(x,p) = p^x(1-p)^x,\ x = 0,1,\ 0<p<1$. 由于两点分布族是指数族, C-R正则条件成立. Fisher信息函数为

$$I(p) = E\bigg[\frac{\partial \log f(X,p)}{\partial p}\bigg]^2 = E\left[\frac{X-p}{p(1-p)}\right]^2 = \frac{Var_p(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)},$$

因此C-R下界为$1/[nI(p)] = p(1-p)/n$.

而$\bar{X}$为p的无偏估计, 其方差$D_p(\bar{X}) = p(1-p)/n$达到C-R下界. 故$\bar{X}$为p的UMVUE.

例 设$\mathbf{X}=(X_1,\cdots,X_n)$为从Poisson分布$P(\lambda)$中抽取的简单样本, 用C-R不等式验证样本均值$\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$为$\lambda$的UMVUE.

证 设随机变量$X\sim P(\lambda)$,则其概率分布为$f(x,\lambda)=e^{-\lambda}\lambda^x/x!$, $x=0,1,2,\cdots$, $\lambda>0$. 由于Poisson分布族为指数族, 故C-R正则条件成立. Fisher信息函数为

$$I(\lambda)=E\left[\frac{\partial\log f(X,\lambda)}{\partial\lambda}\right]^2=E\left[\frac{X-\lambda}{\lambda}\right]^2=\frac{1}{\lambda^2}\,D_\lambda(X)=\frac{1}{\lambda},$$

C-R下界为$1/[nI(\lambda)]=\lambda/n$.

$\hat{g}(\mathbf{X})=\bar{X}$为$g(\lambda)=\lambda$的无偏估计, 且方差$D_\lambda(\bar{X})=\lambda/n$达到C-R下界, 故$\bar{X}$为$\lambda$的UMVUE.

例 设$\mathbf{X}=(X_1,\cdots,X_n)$为从指数分布$EP(\lambda)$中抽取的简单样本, 用C-R不等式验证$\hat{g}(\mathbf{X})=\bar{X}$为$g(\lambda)=1/\lambda$的UMVUE.

证 指数分布$EP(\lambda)$的密度函数为$f(x,\lambda)=\lambda e^{-\lambda x}I_{[x>0]}$, $\lambda>0$. 指数分布族是指数族, 故C-R正则条件成立. Fisher信息函数为

$$I(\lambda)=E\left[\frac{\partial\log f(X,\lambda)}{\partial\lambda}\right]^2=E\left(\frac{1}{\lambda}-X\right)^2=D_\lambda(X)=\frac{1}{\lambda^2},$$

故C-R下界为$(g'(\lambda))^2/[nI(\lambda)]=1/(n\lambda^2)$.

而$\hat{g}(\mathbf{X})=\bar{X}$为$g(\lambda)=1/\lambda$的无偏估计, 其方差$D_\lambda(\bar{X})=1/(n\lambda^2)$达到C-R下界, 故$\bar{X}$为$g(\lambda)=1/\lambda$的UMVUE.

2. C-R不等式等号成立的条件*

指数族不但能使定理的条件(i)-(vi)成立, 它还是能使C-R不等式等号成立的唯一分布族. 下列的定理指出除了指数族外, $g(\theta)$的无偏估计的方差不能处处达到C-R下界.

定理 若C-R正则条件(i)-(v)成立, $\hat{g}(\mathbf{X})$为$g(\theta)$的无偏估计. $g(\theta)$在Θ上不恒等于常数, $D_\theta(\hat{g}(\mathbf{X}))<\infty$对一切$\theta\in\Theta$, 则C-R不等式中等号对一切$\theta\in\Theta$成立的充要条件是

$$f(\mathbf{x},\theta)=C(\theta)\exp\{Q(\theta)T(\mathbf{X})\}h(\mathbf{x}), \tag{2.5}$$

其中$C(\theta)$和$Q(\theta)$为θ 的连续可微函数, 且$Q'(\theta)\neq 0$在Θ上处处成立.

注3.5.1 定理的严格叙述要将C-R不等式对一切$\theta\in\Theta$成立的充要条件(2.5)改为如下形式

$$P_\theta\left(f(\mathbf{x},\theta)=C(\theta)\exp\{Q(\theta)T(\mathbf{X})\}h(\mathbf{x})\right)=1, \tag{2.6}$$

其余不变. 即要求(2.5)式关于测度P_θ概率为1成立. 这对定理的证明并无实质影响.

证 定理证明中导出(2.1)的关键步骤是用到Cauchy-Schwartz不等式: $Cov^2(\xi,\eta)\le D(\xi)\cdot D(\eta)$. 这一不等式中等号成立的充要条件是$\xi$和$\eta$有线性关系, 即存在常数$a$, b 使得$\xi=a\eta+b$. 在C-R不等式定理中$\xi=S(\mathbf{x},\theta)$, $\eta=\hat{g}(\mathbf{X})$, 因此(2.1)式等号成立的充要条件为

$$S(\mathbf{x},\theta)=a(\theta)\hat{g}(\mathbf{X})+b(\theta), \tag{2.7}$$

这里$a(\theta)\neq 0$, $b(\theta)$皆与x无关. 故问题转化为证明: “(2.5)对一切$\theta\in\Theta$成立的充要条件为(2.7)式成立”. 下面来证明之.

若(2.7)式成立, 两边对θ求积分得

$$\int_{\theta_0}^{\theta} \frac{\partial \log f(\mathbf{x}, u)}{\partial u} du = \int_{\theta_0}^{\theta} a(u) du \cdot \hat{g}(\mathbf{x}) + \int_{\theta_0}^{\theta} b(u) du.$$

即

$$\log f(\mathbf{x}, \theta) = Q(\theta)\hat{g}(\mathbf{x}) + R(\theta) + l(\mathbf{x}), \tag{2.8}$$

此处

$$Q(\theta) = \int_{\theta_0}^{\theta} a(u) du, \quad R(\theta) = \int_{\theta_0}^{\theta} b(u) du, \quad l(\mathbf{x}) = \log f(\mathbf{x}, \theta_0).$$

记$C(\theta) = e^{R(\theta)}$, $h(\mathbf{x}) = e^{l(\mathbf{x})}$, 由(2.8)得

$$f(\mathbf{x}, \theta) = C(\theta) \exp\{Q(\theta)\hat{g}(\mathbf{x})\} h(\mathbf{x}), \quad 一切\ \theta \in \Theta,$$

即(2.5)式成立.

反之, 若(2.5)式成立, 将其两边取对数得

$$\log f(\mathbf{x}, \theta) = \log C(\theta) + Q(\theta)\hat{g}(\mathbf{x}) + \log h(\mathbf{x}),$$

两边对θ求导得

$$S(\mathbf{x}, \theta) = \frac{C'(\theta)}{C(\theta)} + Q'(\theta)\hat{g}(\mathbf{x}). \tag{2.9}$$

由$E_\theta S(\mathbf{x}, \theta) = 0$可得$C'(\theta)/C(\theta) + Q'(\theta)g(\theta) = 0$, 即$C'(\theta)/C(\theta) = -Q'(\theta)g(\theta)$,将此式代入到(2.9)可得

$$S(\mathbf{x}, \theta) = Q'(\theta)(\hat{g}(\mathbf{x}) - g(\theta)) = a(\theta)\hat{g}(\mathbf{x}) + b(\theta),$$

此处$a(\theta) = Q'(\theta) \neq 0$, $b(\theta) = Q'(\theta)g(\theta)$.因此(2.7)成立, 定理证毕.

下面的定理指出, 即使在指数族的情形, $g(\theta)$的无偏估计的方差能处处达到C-R下界的情形也不多.

定理 设样本$\mathbf{X}$的分布族为如下的单参数指数族

$$f(\mathbf{x}, \theta) = C(\theta) \exp\{Q(\theta)T(\mathbf{x})\} h(\mathbf{x}),$$

则当且仅当

$$g(\theta) = E_\theta(aT(\mathbf{x}) + b)$$

时, 才有其方差处处达到C-R下界的无偏估计$\hat{g}(\mathbf{X}) = aT(\mathbf{X}) + b$,其中$a$, b为与θ无关的两个常数.

证 见参考文献[1] P_{117} 定理2.2.3.

例 设$\mathbf{X} = (X_1, \cdots, X_n)$为自Poisson分布$P(\lambda)$中抽取的简单随机样本,证明只有$g(\lambda)$是$\lambda$的线性函数时,才存在$g(\lambda)$的无偏估计,其方差能处处达到C-R下界.

证 将样本$\mathbf{X}$的概率分布

$$f(\mathbf{x}, \lambda) = \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \Big/ \prod_{i=1}^{n} x_i!$$

表成指数族的形式

$$f(\mathbf{x}, \lambda) = e^{-n\lambda} \exp\left\{\log \lambda \cdot \sum_{i=1}^{n} x_i\right\} \Big/ \prod_{i=1}^{n} x_i!$$

$$= C(\lambda)\exp\{Q(\lambda)T(\mathbf{x})\}h(\mathbf{x}).$$

此处$C(\lambda) = e^{-n\lambda}$, $Q(\lambda) = \log\lambda$, $T(\mathbf{x}) = \sum_{i=1}^{n} X_i$, $h(\mathbf{x}) = 1/\prod_{i=1}^{n} x_i!$.

由定理可知当且仅当

$$g(\lambda) = E(aT(\mathbf{x}) + b) = a \cdot n\lambda + b = c\lambda + b, \quad c,\ b\text{为常数}$$

时, 无偏估计$\hat{g}(\mathbf{X}) = aT(\mathbf{X}) + b$的方差处处达到C-R下界, 且是$g(\lambda)$的UMVUE. 其它形式的参数都不存在其UMVUE.

特别取$g(\lambda) = \lambda$,即$a = 1/n$, $b = 0$, 则$\hat{g}(\mathbf{X}) = T/n = \bar{X}$的方差处处达到C-R下界, 因此它是$\lambda$的UMVUE. 这与例3.5.2给出的结果相同.

3. Fisher信息函数

将C-R不等式中的

$$I(\theta) = E_\theta\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right]^2$$

称为*Fisher*信息函数 (或称为*Fisher*信息量 . 为解释$I(\theta)$,不妨令$g(\theta) = \theta$并假定C-R不等式下界$1/(nI(\theta))$可达到. 这时$nI(\theta)$越大, $g(\theta)$的无偏估计$\hat{g}(\mathbf{X})$的方差越小, 表明$g(\theta) = \theta$ 可以估计得越精. $nI(\theta)$与n和$I(\theta)$成正比, n是样本容量, 这表明若以估计量的方差的倒数作为估计量精度的指标, 则精度与n成正比. 比例因子, 即$I(\theta)$,反映总体分布的一种性质. 就是说, 总体分布的$I(\theta)$越大, 意味着总体的参数越容易估计, 或者说, 该总体模型本身提供的信息量越多. 故有理由把$I(\theta)$视为衡量总体模型所含信息多少的量—信息量. $I(\theta)$也可以解释成单个样本提供的信息量. 由于$X_1,\cdots,X_n$是i.i.d.的, 它们的地位是平等的, 故每个样品提供同样多的信息$I(\theta)$, 即整个样本$(X_1,\cdots,X_n)$所含信息量为$nI(\theta)$.

*Fisher*信息量 $I(\theta)$的重要意义还在于, 在点估计大样本理论的研究中, 它起相当作用. 如在寻求θ的极大似然估计$\hat{\theta}^*$ 的渐近分布, $\hat{\theta}^*$的渐近正态分布的方差就可以用*Fisher*信息量表示, 即

$$\left\{nE_\theta\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)\right\}^{-1} = \frac{1}{nI(\theta)}.$$

所以在定理3.3.2的条件下, 参数θ的极大似然估计$\hat{\theta}^*$ 的渐近分布可表示为$N\big(\theta, 1/[nI(\theta)]\big)$. 这表明其渐近方差与样本的*Fisher*信息量成反比. 因此, 当$I(\theta)$越大时, 渐近方差越小, 用θ的极大似然估计$\hat{\theta}^*$来估计θ就越精. 由此看来, Fisher将$I(\theta)$称为信息量, 确有一定的根据.

三、多参数C-R不等式简介*

以上讨论的都是参数θ为一维的情形, 对θ为高维的情形也可以建立类似的结果. 为此先引进一些记号. 设$A = (a_{ij})$和$B = (b_{ij})$是同阶的非负定方阵, 若$A - B$是非负定的, 则记为$A \geq B$. 这时必有$a_{ii} \geq b_{ii}$,对一切i.

现设$\theta = (\theta_1,\cdots,\theta_k)$,总体概率函数记为$f(x,\theta)$, $\mathbf{X} = (X_1,\cdots,X_n)$为从总体中抽取的简单随机样本. 设$\hat{\theta} = \hat{\theta}(\mathbf{X}) = (\hat{\theta}_1,\cdots,\hat{\theta}_k)$是$\theta$的一个无偏估计. 以$Cov_\theta(\hat{\theta})$ 记θ的协方差阵, 它是一个k阶非负定方阵, 其(i,j) 元为$E_\theta\big[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\big]$, 则在类似于$\theta$为一维的正则条件下, 可以证明

$$Cov_\theta(\hat{\theta}) \geq (nI(\theta))^{-1}, \tag{2.10}$$

这里$I(\theta)=(I_{ij}(\theta))$是一个$k$阶正定方阵, 且

$$I_{ij}(\theta)=E_\theta\left[\left(\frac{\partial\log f(x,\theta)}{\partial\theta_i}\right)\left(\frac{\partial\log f(x,\theta)}{\partial\theta_j}\right)\right],\quad i,j=1,2,\cdots,k, \tag{2.11}$$

则(2.10)就是多维的*C-R*不等式. 若记$(I(\theta))^{-1}=I^*(\theta)=(I^*_{ij}(\theta))$,故由(2.10)可得

$$D_\theta(\hat\theta_i)\geq I^*_{ii}(\theta)/n,\ \ i=1,2,\cdots,k. \tag{2.12}$$

这给出了θ的每个分量θ_i的无偏估计$\hat\theta_i$的方差的下限.

例 设$\mathbf{X}=(X_1,\cdots,X_n)$为从正态总体$N(a,\sigma^2)$中抽取的简单样本, 记$\theta=(a,\sigma^2)$, 其中$\theta_1=a$, $\theta_2=\sigma^2$.求θ的C-R下界,并将其与θ_1和θ_2 的无偏估计$\bar X$和S^2的方差进行比较。

解 正态随机变量的密度函数为

$$f(x,\theta)=(2\pi\theta_2)^{-\frac12}\exp\Big\{-\frac{1}{2\theta_2}(x-\theta_1)^2\Big\},$$

可知

$$\frac{\partial\log f(x,\theta)}{\partial\theta_1}=\frac{x-\theta_1}{\theta_2},\quad \frac{\partial\log f(x,\theta)}{\partial\theta_2}=\frac{-\theta_2+(x-\theta_1)^2}{2\theta_2^2}.$$

由此算出

$$I_{11}(\theta)=\frac{1}{\sigma^2},\quad I_{22}(\theta)=\frac{1}{2\sigma^4},\quad I_{12}(\theta)=I_{21}(\theta)=0,$$

故有

$$nI(\theta)=\begin{pmatrix} n/\sigma^2 & 0\\ 0 & n/2\sigma^4\end{pmatrix},\qquad (nI(\theta))^{-1}=\begin{pmatrix}\sigma^2/n & 0\\ 0 & 2\sigma^4/n\end{pmatrix},$$

若记

$$\hat\theta_1=\bar X=\frac1n\sum_{i=1}^n X_i,\qquad \hat\theta_2=S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar X)^2,$$

则由多维C-R不等式可知

$$Cov_\theta\begin{pmatrix}\hat\theta_1\\ \hat\theta_2\end{pmatrix}\geq\begin{pmatrix}\sigma^2/n & 0\\ 0 & 2\sigma^4/n\end{pmatrix}.$$

而$D_\theta(\hat\theta_1)=\sigma^2/n$达到C-R下界, 故它是$\theta_1=a$的UMVUE. 利用$(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$及$D_\theta[(n-1)S^2/\sigma^2]=2(n-1)$可得

$$D_\theta(\hat\theta_2)=\frac{2\sigma^4}{n-1}>\frac{2\sigma^4}{n}.$$

因此, $S^2=\hat\theta_2$的方差大于C-R下界. 在例3.4.11中我们已证明了$\bar X$和S^2 分别是a和σ^2的UMVUE.

由本例看出, 即使在正态总体方差这样简单的场合, 方差σ^2的UMVUE S^2 也达不到C-R下界. 因此在一些问题中C-R下界常比真下界为小. 这表明以*C-R*不等式作为寻找UMVUE的方法是不够理想的. 多年来有一些学者一直在研究改进这个不等式的问题.

四、有效估计和估计的效率

定义3.5.2 设$\hat g(\mathbf{X})$为$g(\theta)$的无偏估计, 比值

$$e_{\hat g}(\theta)=\frac{[g'(\theta)]^2/nI(\theta)}{D_\theta[\hat g(\mathbf{X})]}$$

称为无偏估计$\hat{g}(\mathbf{X})$ 的效率(Efficiency). 显然$0 < e_{\hat{g}}(\theta) \le 1$, 当$e_{\hat{g}}(\theta) = 1$时, 称$\hat{g}(\mathbf{X})$是$g(\theta)$ 的有效估计 (Effective estimation). 若$\lim\limits_{n\to\infty} e_{\hat{g}}(\theta) = 1$, 则称$\hat{g}(\mathbf{X})$是$g(\theta)$的渐近有效估计 (Asymptotically effective estimation).

这一概念有其不足之处: 有效估计是无偏估计类中最好的估计, 人们当然希望使用它. 可惜, 有效估计是不多的, 但渐近有效估计却不少. 从有效估计的定义看, 有效估计一定是UMVUE, 但很多UMVUE不是有效估计. 这是因为C-R下界偏小, 在很多场合UMVUE的方差达不到C-R下界. 另外C-R不等式成立的前提要求样本分布族满足C-R正则条件. 当这些条件不成立时, C-R不等式可以不对, 这时依据它所提供的C-R下界去定义估计的效率就不合理了.

定义3.5.3 设$\hat{g}$为$g(\theta)$的一个CAN估计, 其渐近方差为$\sigma^2(\theta)$, 则称

$$ae_{\hat{g}}(\theta) = \frac{1}{nI(\theta)}\Big/\sigma^2(\theta) = \big(nI(\theta)\sigma^2(\theta)\big)^{-1}$$

为$\hat{g}$的渐近效率 (Asymptotic efficiency).

据此, 极大似然估计有渐近效率1. 至于矩估计, 我们在§3.2 已证明它们在很一般条件下为CAN估计, 不过除了常见的几个例子(在其中矩估计与MLE重合)之外, 矩估计渐近效率一般都低于1, 通常人们说矩估计不如极大似然估计, 大概就是指这一点.

例3 设$\mathbf{X} = (X_1, \cdots, X_n)$为从$N(a, \sigma^2)$中抽取的简单随机样本, (i) 当$a$未知时, 证明样本方差$S^2$不是$\sigma^2$的有效估计，但是渐近有效估计. (ii) 当$a$已知时, 求$\sigma^2$的有效估计.

解 (i) 当a未知时, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ 之方差为$2\sigma^4/(n-1)$达不到C-R下界$2\sigma^4/n$, 故它不是σ^2的有效估计. 估计的效率为$e_{_{S^2}}(\sigma^2) = (n-1)/n < 1$,但是

$$\lim_{n\to\infty} e_{_{S^2}}(\sigma^2) = \lim_{n\to\infty} \frac{n-1}{n} = 1,$$

因此S^2是σ^2的渐近有效估计.

(ii) 由于a已知, 令$S_a^2 = \frac{1}{n}\sum_{i=1}^n (X_i - a)^2$, 由于$nS_a^2/\sigma^2 \sim \chi_n^2$, 故$D\left(nS_a^2/\sigma^2\right) = 2n$, 因此有$D(S_a^2) = 2\sigma^4/n$,它达到C-R下界, 故$S_a^2$为$\sigma^2$的有效估计. 当$a$已知时, 利用Lehmann-Scheffe定理也容易证明S_a^2为σ^2的UMVUE.

例3 设$\mathbf{X} = (X_1, \cdots, X_n)$是从下列含有位置参数的指数分布族中抽取的简单样本,

$$f(x, a) = e^{-(x-a)} I_{[a,\infty)}(x), \ \ -\infty < a < +\infty,$$

求a的UMVUE.

解 上述分布族不是C-R正则族, 即C-R正则条件不成立. 因为在$a = x$时$\frac{\partial f(x,a)}{\partial a}$不存在, 且密度函数的支撑集$\{x : f(x, a) > 0\} = \{x : x \ge a\}$与未知参数$a$有关, 因此它不满足C-R正则条件.因此我们不能用C-R不等式来求a的UMVUE.

显然,样本的最小次序统计量$X_{(1)}$是a的充分完全统计量. $X_{(1)}$的密度函数为

$$f(y, a) = ne^{-n(y-a)} I_{[a,\infty)}(y).$$

于是

$$E(X_{(1)}) = n\int_a^\infty ye^{-n(y-a)}dy = a + \frac{1}{n},$$

故由L-S定理可知$\hat{a}(\mathbf{X}) = X_{(1)} - \frac{1}{n}$是$a$ 的UMVUE; 但不可以讨论$\hat{a}(\mathbf{X})$是否为有效估计, 因为C-R正则条件不成立, 去讨论C-R下界就失去了意义.

下面是一个确定渐近效率的例子:

例　设$\mathbf{X}=(X_1,\cdots,X_n)$为从正态分布$N(a,\sigma^2)$中抽取的简单随机样本. 记$\theta=(a,\sigma^2)$,设$\hat{g}(\mathbf{X})$为样本中位数, 它是$a$的一个估计. 求它的渐近效率。

解　由分位数估计的渐近分布, 并注意到此处$f(\xi_{1/2})=f(a)=1\big/\sqrt{2\pi}\sigma$,故有

$$2\sqrt{n}\cdot\frac{1}{\sqrt{2\pi}\sigma}(\hat{g}(\mathbf{X})-a)\xrightarrow{\mathscr{L}}N(0,1),$$

或改写为

$$\sqrt{n}(\hat{g}(\mathbf{X})-a)\xrightarrow{\mathscr{L}}N\Big(0,\frac{\pi}{2}\sigma^2\Big),$$

即$\hat{g}(\mathbf{X})$的渐近方差为$\pi\sigma^2\big/(2n)$.由例3.5.4可知$I(\theta)=1/\sigma^2$, 故按定义3.5.3可知$\hat{g}(\mathbf{X})$的渐近效率是

$$ae_{\hat{g}}(\theta)=\frac{1}{nI(\theta)}\Big/\sigma^2(\theta)=\frac{\sigma^2}{n}\Big/\frac{\pi}{2n}\sigma^2=\frac{2}{\pi}.$$

例　利用有效估计的概念和定理可以证明下列极大似然估计的一条重要性质, 叙述如下:

设$\mathbf{X}=(X_1,\cdots,X_n)$为自分布族$\{f(x,\theta),\theta\in\Theta\}$中抽取的简单随机样本, Θ为R_1上的开区间,记为A. 若$g(\theta)$的有效估计$\hat{g}(\mathbf{X})=\hat{g}(X_1,\cdots,X_n)$存在, 则$g(\theta)$之MLE $\hat{g}^*(\mathbf{X})=\hat{g}^*(X_1,\cdots,X_n)$ 必与$\hat{g}(\mathbf{X})$重合, 即$\hat{g}^*(\mathbf{X})=\hat{g}(\mathbf{X})$.

证　若$\hat{g}(\mathbf{X})$为$g(\theta)$的有效估计, 即其方差达到C-R下界, 由定理3.5.2可知这一事实成立的充要条件为(2.7)式成立, 即$S(\mathbf{x},\theta)=a(\theta)\hat{g}(\mathbf{x})+b(\theta)$.现将此式改写一下, 因为$E_\theta\{S(X,\theta)\}=a(\theta)g(\theta)+b(\theta)=0$,即$b(\theta)=-a(\theta)g(\theta)$,将其代入上式得

$$S(\mathbf{x},\theta)=a(\theta)\hat{g}(\mathbf{x})-a(\theta)g(\theta)=a(\theta)\left(\hat{g}(\mathbf{x})-g(\theta)\right),\tag{2.13}$$

其中$a(\theta)\neq0$,对一切$\theta\in\Theta$,且在Θ上连续, 故$a(\theta)$在Θ或全大于0, 或全小于0. 无妨令$a(\theta)>0$,对一切$\theta\in\Theta$,则由(2.13)可知, 当$\hat{g}(\mathbf{x})>g(\theta)$时, $S(\mathbf{x},\theta)=\partial\log f(\mathbf{x},\theta)\big/\partial\theta>0$, 表示对数似然函数在$(-\infty,\hat{g}(\mathbf{x}))\cap A$内严格单调上升; 当$\hat{g}(\mathbf{x})<g(\theta)$时, $\partial\log f(\mathbf{x},\theta)\big/\partial\theta<0$表示对数似然函数在$(\hat{g}(\mathbf{x}),\infty)\cap A$上严格单调下降, 所以$\hat{g}(\mathbf{x})$是似然方程的唯一解, 且是极大化的, 故$\hat{g}(\mathbf{X})$即为$g(\theta)$的MLE, 因此$\hat{g}(\mathbf{X})=\hat{g}^*(\mathbf{X})$,证毕.