系统科学与数学
J. Sys. Sci. & Math. Scis.
5 (3) (1985), 192—200

多参数情形下 MLE 的大偏差

陈 家 骅
(中国科学院系统科学研究所,北京)

在大样本点估计理论中,我们一般只考虑相合估计,对于任何一个相合估计 T_n 以及任何给定的 $\varepsilon>0$, 其尾概率

$$\alpha(T_n, \theta, \varepsilon)=P\{\|T_n-\theta\|>\varepsilon\},$$

随 n 趋无穷而趋于零. 于是有人[1,2,3] 提出了将此概率趋于零的速度作为考查此相合估计的渐近性质的一个准则. 一般地,对于一个相合估计而言,在一定的条件下, $\alpha(T_n, \theta, \varepsilon)$ 有如下的形式:

$$\alpha(T_n, \theta, \varepsilon)=e^{-n\beta(T,\theta,\varepsilon)}(C_{0n}+n^{-1}C_{1n}+\cdots).$$

其中 $\beta(T, \theta, \varepsilon)$ 及 C_{in} 是只依赖于分布族,估计 T_n 以及 ε 的常数. 我们称 $\beta(T, \theta, \varepsilon)$ 为其指数速度.

近来很多人对相合估计的收敛速度作了大量研究,这里就不一一列举了,其中 James, C. F_u.[5]. 研究了在单参数情形下, $\beta(\cdot)$的 Bahadur 界以及 MLE 估计的指数速度, 指出在一定的条件下,如果将这两函数作关于 ε 在原点的 Taylor 展开, 那么它们展开项系数直到三次项都是相同的,而四次项在简单指数族情形下也是相同的, 这就表明了 MLE 估计在这时是 Fisher, Rao, Efron 意义下的二阶渐近有效估计.

本文继续 James, C. Fu. 的工作,指出在一定的条件下,可以把 James, C. Fu. 的工作推广到多参数情形. 即在多参数情形下, MLE 估计的指数速度与 Bahadur 界关于 ε 在 0 点的展开有直到 ε^3 的相同的系数, 而 ε^4 的系数在简单指数族下相等,于是,这时的 MLE 也是 Fisher, Rao, Efron 意义下的二阶有效估计.

下面我们分两步进行,先给出 Bahadur 界,而后求出 MLE 的指数速度. 最后将它们进行比较,并举几个例子.

设 $X_1,\cdots,X_n$, 为一列 i. i. d 随机变量序列, 具有分布密度 $f(x, \theta)$, 其中 $\theta\in\Theta$, Θ 为 R^q 中的一开凸集. θ 的估计序列为 $T_n(X_1,\cdots,X_n)$, 简记为 T_n, 则若 T_n 为 θ 的相合估计,就有

$$\lim_{n\to\infty}\frac{1}{n}\log P\{\|T_n-\theta\|\geqslant\varepsilon\}\geqslant -B(\theta, \varepsilon),$$

$B(\theta, \varepsilon)$ 为 Bahadur 阶,由下式给出

$$B(\theta, \varepsilon)=\inf_{\theta'\in\Theta}\{K(\theta', \theta),\ \|\theta'-\theta\|>\varepsilon\},$$

其中 $K(\theta', \theta)$ 为 Kullback 信息,由

1984 年 9 月 1 日收到.

$$K(\theta',\theta)=E_{\theta'}\log\frac{f(X,\theta')}{f(X,\theta)}$$

给出.

由于此结果与 [5] 中的不尽相同,所以这里给一简单证明.

引理 1. 设 $\theta_1\neq\theta_2$, $\theta_1,\theta_2\in\boldsymbol{\Theta}$. 以 $\alpha_n(\beta)$ 记样本量为 n 时所有功效不小于 $1-\beta$ 的关于 $\theta_1\longleftrightarrow\theta_2$ 的检验的显著性水平的下界,即

$$\alpha_n(\beta)=\inf_{\phi\in\mathscr{T}}\{E_{\theta_1}\phi(X^{(n)})\},$$
$$\mathscr{T}=\{\phi:\ E_{\theta_2}\phi(X^{(n)})\geqslant 1-\beta\},$$
$$(X^{(n)}=(X_1,\cdots,X_n)\text{以后同此}),$$

则有

$$\lim_{n\to\infty}\frac{1}{n}\log\alpha_n(\beta)=-K(\theta_2,\theta_1).$$

证明见[1].

由于 T_n 是 θ 的相合估计,所以对任何 $\theta'\in\boldsymbol{\Theta}$, $\|\theta'-\theta\|>\varepsilon$, 作关于 $\theta\longleftrightarrow\theta'$ 的检验函数

$$\phi(T_n)=\begin{cases}0, & \|T_n-\theta\|<\varepsilon,\\ 1, & \|T_n-\theta\|\geqslant\varepsilon.\end{cases}$$

因由 $E_{\theta'}\phi(T_n)\to 1$, 可知存在 $M>0$, 当 $n>M$ 时, $E_{\theta'}\phi(T_n)>1-\beta$, 所以有 $E_\theta\phi(T_n)\geqslant\alpha_n(\beta)$. 于是

$$\varliminf_{n\to\infty}\frac{1}{n}\log P\{\|T_n-\theta\|\geqslant\varepsilon\}\geqslant\lim_{n\to\infty}\frac{1}{n}\log\alpha_n(\beta)=-K(\theta',\theta).$$

由于此不等式对任何 $\|\theta'-\theta\|>\varepsilon$ 成立,所以

$$\begin{aligned}\varliminf_{n\to\infty}\frac{1}{n}\log P\{\|T_n-\theta\|\geqslant\varepsilon\}&\geqslant\sup_{\theta'\in\Theta}\{-K(\theta',\theta),\ \|\theta'-\theta\|>\varepsilon\}\\&=-\inf_{\theta'\in\Theta}\{K(\theta',\theta),\ \|\theta'-\theta\|>\varepsilon\}.\end{aligned}$$

下面我们约定一些记号. 令

$$l(x,\theta)=\log f(x,\theta),$$
$$l(x^{(n)},\theta)=\sum_{i=1}^{n}l(x_i,\theta),$$
$$l^{(1)}(x,\theta)=\left(\frac{\partial l}{\partial\theta_1},\cdots,\frac{\partial l}{\partial\theta_q}\right)^{\tau},$$
$$l^{(2)}(x,\theta)=\left(\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}(x,\theta)\right)_{q\times q}.$$

以 ν 记 R^q 中的单位向量,

$$l_\nu^{(i)}(x,\theta+s\nu)=\frac{\partial^i l}{\partial t^i}(x,\theta+t\nu)|_{t=s},\ i=1,2,3,4.$$
$$I^\nu=E_\theta(l_\nu^{(1)}(X,\theta))^2,$$
$$f_\nu^{(i)}=\frac{\partial^i f}{\partial t^i}(x,\theta+t\nu)|_{t=0}.$$

$$\mu^{\nu}_{i,j,k} = E\left(\frac{f^{(1)}_{\nu}}{f}\right)^{i}\cdot\left(\frac{f^{(2)}_{\nu}}{f}\right)^{j}\cdot\left(\frac{f^{(3)}_{\nu}}{f}\right)^{k}.$$

以上记号尽管复杂，但只要把任何带 ν 的记号都将一维时的导数改为多维时对 ν 的偏导即可.

为了得到本文结果，假设如下条件成立：

1. $l(x,\theta)=\log f(x,\theta)$ 对任何 x 都是 θ 的上凸函数，即 $l^{(2)}=\left(\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right)_{q\times q}$ 为一负定矩阵.

2. 对任何 $\theta\in\boldsymbol{\Theta}$，有一个 θ 的 δ 邻域 $N(\theta,\delta)$ 及可测函数 $A_i(x,\theta)$，$i=0,1,2,3,4$。

$$E_\theta e^{tA_0(X,\theta)}<\infty.$$
$$E_\theta e^{tA_1(X,\theta)}<\infty,\quad \text{对某 } t>0 \text{ 成立}.$$
$$E_\theta A_i(X,\theta)<\infty,\quad \text{在 } i=2,3,4 \text{ 成立}.$$

当 $\theta+s\nu\in N(\theta,\delta)$ 时，

$$|l^{(1)}_{\nu}(x,\theta+s\nu)|<A_0(x,\theta).$$

$\theta',\theta''\in N(\theta,\delta)$ 时，

$$\|l^{(i)}(x,\theta')-l^{(i)}(x,\theta'')\|\leqslant A_i(x,\theta)\|\theta'-\theta''\|,\quad i=1,2,3,4.$$

3. 对任何 $\theta\in\boldsymbol{\Theta}$，存在常数 $u=u(\theta)>0$，使概率 $P_\theta\{l^{(1)}_{\nu}(X,\theta+\varepsilon\nu)>0\}>0$ 对一切 $0<\varepsilon<u$ 成立.

定理 1. 条件 2 成立时，我们有

$$\begin{aligned}B(\theta,\varepsilon)&=\inf_{\theta'\in\Theta}\{K(\theta',\theta),\ \|\theta'-\theta\|>\varepsilon\}\\&\leqslant\inf\{K(\theta',\theta),\ \|\theta'-\theta\|=\varepsilon\}\\&=\inf_{\nu}\Big\{\frac{\varepsilon^2}{2}I^{\nu}(\theta)+\frac{\varepsilon^3}{3!}(3\mu^{\nu}_{110}-\mu^{\nu}_{300})\\&\quad+\frac{\varepsilon^4}{4!}(2\mu^{\nu}_{400}-6\mu^{\nu}_{210}+4\mu^{\nu}_{101}+3\mu^{\nu}_{020})+o(\varepsilon^4),\\&\quad\text{任何单位向量 }\nu\in R^q\Big\}.\end{aligned}$$

证. 由条件 2，易见 $K(\theta',\theta)$ 关于 θ' 连续，故有

$$\begin{aligned}&\inf_{\theta'\in\Theta}\{K(\theta',\theta),\ \|\theta'-\theta\|>\varepsilon\}\\&\quad=\inf\{K(\theta',\theta),\ \|\theta'-\theta\|\geqslant\varepsilon\}\\&\quad\leqslant\inf\{K(\theta',\theta),\ \|\theta'-\theta\|=\varepsilon\}.\end{aligned}$$

以下在每一方向 ν 作 Taylor 展开再由控制收敛定理便可得证.

引理 2. 在条件 2 成立时，下列等式成立.

i) $El^{(1)}_{\nu}(\varepsilon)=\varepsilon El^{(2)}_{\nu}+\frac{1}{2!}\varepsilon^2El^{(3)}_{\nu}+\frac{1}{3!}\varepsilon^3El^{(4)}_{\nu}+o(\varepsilon^3)$,

ii) $E[l^{(1)}_{\nu}(\varepsilon)]^2=E(l^{(1)}_{\nu})^2+2\varepsilon El^{(1)}_{\nu}l^{(2)}_{\nu}+\varepsilon^2[E(l^{(2)}_{\nu})^2+El^{(1)}_{\nu}l^{(3)}_{\nu}]+o(\varepsilon^2)$,

iii) $E[l^{(1)}_{\nu}(\varepsilon)]^3=E(l^{(1)}_{\nu})^3+3\varepsilon E(l^{(1)}_{\nu})^2l^{(2)}_{\nu}+3\varepsilon^2[El^{(1)}_{\nu}(l^{(2)}_{\nu})^2$

$$+\frac{1}{2}E(l_\nu^{(1)})^2 l_\nu^{(3)}]+o(\varepsilon^2),$$

iv) $E[l_\nu^{(1)}(\varepsilon)]^4=E(l_\nu^{(1)})^4+o(1).$

这里 $l_\nu^{(i)}(\varepsilon)$ 表示 $l_\nu^{(i)}(X,\theta+\varepsilon\nu)$，$l_\nu^{(i)}=l_\nu^{(i)}(X,\theta)$，$E[\cdot]=E_\theta[\cdot]$. 此证明如[5]，只要利用 Taylor 展开即可.

记

$$\phi(t,\theta,\varepsilon,\nu)=E_\theta[\exp\{tl_\nu^{(1)}(X,\theta+\varepsilon\nu)\}],$$ 那么有

引理 3. 设条件 2，3 成立，则存在一个唯一的单值函数 $T_\theta(\varepsilon,\nu)$ 定义于 $0<\varepsilon<u$，使得

$$\phi_t'(T_\theta(\varepsilon,\nu),\theta,\varepsilon,\nu)=0,$$

$$\inf_{t>0}\phi(t,\theta,\varepsilon,\nu)=\phi(T_\theta(\varepsilon,\nu),\theta,\varepsilon,\nu).$$

而且

$$T_\theta(\varepsilon,\nu)=\varepsilon+A\varepsilon^2+B\varepsilon^3+o(\varepsilon^3),$$

这里

$$A=-El_\nu^{(1)}l_\nu^{(2)}/(2I^\nu),$$

$$B=-\frac{1}{I^\nu}\Big[\frac{1}{3!}El_\nu^{(4)}-\frac{1}{I^\nu}(El_\nu^{(1)}l_\nu^{(2)})^2+E(l_\nu^{(2)})^2+El_\nu^{(1)}l_\nu^{(3)}$$
$$+\frac{3}{2}E(l_\nu^{(1)})^2l_\nu^{(2)}+\frac{1}{2I^\nu}El_\nu^{(1)}l_\nu^{(2)}E(l_\nu^{(1)})^3+\frac{1}{3!}E(l_\nu^{(1)})^4\Big].$$

证. 在条件 2，3 成立时，由矩母函数性质，ϕ 关于 t 在 ϕ 有穷的区间内严凸，且极小值在 $t>0$ 处取得. 因此，$T_\theta(\varepsilon,\nu)$ 存在唯一. 而 $T_\theta(\varepsilon,\nu)$ 的具体表达式，通过 Taylor 展开，利用引理 2 的结果及控制收敛性便得.

类似于[5]，通过对随机变量 $l_\nu^{(1)}(X,\theta+\varepsilon\nu)$ 的累积数的计算，就有

$$\log\phi(T_\theta(\varepsilon,\nu),\theta,\varepsilon,\nu)$$
$$=-\Big\{\frac{\varepsilon^2}{2}I^\nu+\frac{\varepsilon^3}{3!}(3\mu_{110}^\nu-\mu_{300}^\nu)+\frac{\varepsilon^4}{4!}[(2\mu_{400}^\nu-6\mu_{211}^\nu$$
$$+4\mu_{101}^\nu+3\mu_{020}^\nu)-3(I^\nu)^2\gamma_\nu^2(\theta)]+o(\varepsilon^4)\Big\}.$$

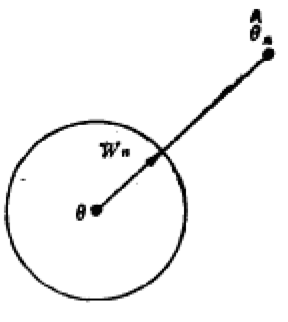
这里 $\gamma_\nu^2(\theta)$ 表示分布族 $\{f(x,\theta+s\nu),s\in(-\varepsilon,\varepsilon)\}$ 的统计曲率，即

$$\gamma_\nu^2(\theta)=\frac{1}{(I^\nu)^2}(\mu_{020}^\nu-2\mu_{210}^\nu+\mu_{400}^\nu)-1-\frac{1}{(I^\nu)^3}(\mu_{110}^\nu-\mu_{300}^\nu)^2.$$

有了这些准备后，我们来求多参数下极大似然估计的 $\beta(\hat\theta_n,\theta,\varepsilon)$.

假设条件 1，2，3 成立，如右图，若 $\|\hat\theta_n-\theta\|\geqslant\varepsilon$，令

$W_n=\dfrac{\hat\theta_n-\theta}{\|\hat\theta_n-\theta\|}$，则由于 $l(x,\theta)$ 为凸函数，因此，$l(x^{(n)},\theta)$ 也是凸函数，由可导性便知 $l^{(1)}(x^{(n)},\hat\theta_n)=0$，所以



$$\frac{\partial l}{\partial s}(x^{(n)},\theta+sW_n)|_{s=\|\hat\theta_n-\theta\|}=W_n^\tau l^{(1)}(x^{(n)},\hat\theta_n)=0.$$

同时

$$\frac{\partial^2 l}{\partial s^2}(x^{(n)},\theta+sW_n)|_{s=\|\hat\theta_n-\theta\|}=W_n^\tau l^{(2)}W_n<0\ \text{(负定性)},$$

从而知 $\frac{\partial l}{\partial s}(x^{(n)}, \theta + sW_n)$ 关于 s 单调减，因其在 $s = \|\hat{\theta}_n - \theta\| \geqslant \varepsilon$ 处值为 0. 所以

$$\frac{\partial l}{\partial s}(x^{(n)}, \theta + sW_n)|_{s=\varepsilon} \geqslant 0.$$

这就有

$$\{\|\hat{\theta}_n - \theta\| > \varepsilon\} \subset \left\{\frac{\partial l}{\partial s}(X^{(n)}, \theta + sW_n)\Big|_{s=\varepsilon} \geqslant 0 \text{ 对某 } W_n \text{ 成立}\right\}$$
$$\subset \{\sup_\nu l_\nu^{(1)}(X^{(n)}, \theta + \varepsilon\nu) \geqslant 0\}.$$

于是得到

引理 4. $P\{\|\hat{\theta}_n - \theta\| \geqslant \varepsilon\} \leqslant P\{\sup_\nu l_\nu^{(1)}(X^{(n)}, \theta + \varepsilon\nu) \geqslant 0\}$.

对于任意两个 R^q 中的单位向量 ν_1, ν_2. 在 $\varepsilon \leqslant \delta$ 时

$$\begin{aligned}
&|l_{\nu_1}^{(1)}(x, \theta + \varepsilon\nu_1) - l_{\nu_2}^{(1)}(x, \theta + \varepsilon\nu_2)| \\
&\quad = |(\nu_1 - \nu_2)^\tau l_{\nu_1}^{(1)}(x, \theta + \varepsilon\nu_1) + \nu_2^\tau[l_{\nu_1}^{(1)}(x, \theta + \varepsilon\nu_1) - l_{\nu_2}^{(1)}(x, \theta + \varepsilon\nu_2]| \\
&\quad \leqslant \|\nu_1 - \nu_2\|A_0(x, \theta) + \varepsilon\|\nu_1 - \nu_2\|A_1(x, \theta) \\
&\quad \triangleq \|\nu_1 - \nu_2\|A(x, \theta),
\end{aligned}$$

其中 $A(x, \theta) = A_0(x, \theta) + \varepsilon A_1(x, \theta)$.

基于上述事实，对事先给定的 $\sigma > 0$，利用 R^q 中单位球面的紧致性，可取到有限个单位向量 $\nu_1, \cdots, \nu_k$，使得任何单位向量 ν，都有某 ν_i，[$0 < i \leqslant k$ 存在，使 $\|\nu - \nu_i\| < \sigma$. 这样又得

引理 5. $\{\sup_\nu l_\nu^{(1)}(X^{(n)}, \theta + \varepsilon\nu) \geqslant 0\}$

$$\subset \bigcup_{j=1}^{k}\left\{l_{\nu_j}^{(1)}(X^{(n)}, \theta + \varepsilon\nu_j) + \sigma\sum_{i=1}^{n} A(X_i, \theta) \geqslant 0\right\},$$

在 $\varepsilon < \delta$ 时成立.

证. 因$\{\nu, \nu \in R^q$ 为单位向量$\}$是一个紧致集，所以对任一 $x^{(n)}$，$\sup_\nu l_\nu^{(1)}(x^{(n)}, \theta + \varepsilon\nu) \geqslant 0$ 时，必有一 ν_0，使 $l_{\nu_0}^{(1)}(x^{(n)}, \theta + \varepsilon\nu_0) \geqslant 0$.

由 $\nu_1, \cdots, \nu_k$ 的取法，存在一 ν_{i_0}，使 $\|\nu_{i_0} - \nu_0\| < \sigma$，因此，

$$\begin{aligned}
&|l_{\nu_0}^{(1)}(x^{(n)}, \theta + \varepsilon\nu_0) - l_{\nu_{i_0}}^{(1)}(x^{(n)}, \theta + \varepsilon\nu_{i_0})| \\
&\leqslant \sum_{i=1}^{n}|l_{\nu_0}^{(1)}(x_i, \theta + \varepsilon\nu_0) - l_{\nu_{i_0}}^{(1)}(x_i, \theta + \varepsilon\nu_{i_0})| \\
&\leqslant \sigma\sum_{i=1}^{n} A(x_i, \theta),
\end{aligned}$$

于是，

$$l_{\nu_{i_0}}^{(1)}(x^{(n)}, \theta + \varepsilon\nu_{i_0}) + \sum_{i=1}^{n}\sigma A(x_i, \theta) \geqslant 0.$$

引理由此便得证.

由条件 3，$P_\theta\{l_\nu^{(1)}(X, \theta + \varepsilon\nu) > 0\} > 0$，因而

$$P_\theta\{l_\nu^{(1)}(X, \theta + \varepsilon\nu) + \sigma A(X, \theta) > 0\} > 0.$$

又从 $l_\nu^{(1)}(x, \theta + s\nu)$ 关于 s 的严格单调性，得

$$E_\theta l_\nu^{(1)}(X, \theta + \varepsilon\nu) < E_\theta l_\nu^{(1)}(X, \theta) = 0.$$

又因 $E_\theta A(X,\theta)<\infty$，故可取 σ 充分小，使得

$$E_\theta[l_{\nu_i}^{(1)}(X,\theta+\varepsilon\nu_i)+\sigma A(X,\theta)]<0,\quad i=1,2,\cdots,k.$$

以 $\rho(\theta,\varepsilon,\nu,\sigma)$ 记 $E_\theta\exp\{t[l_\nu^{(1)}(X,\theta+\varepsilon\nu)+\sigma A(X,\theta)]\}$ 在 $t\geqslant 0$ 时的最小值.（注意：因 $l_\nu^{(1)}(X,\theta+\varepsilon\nu)$, $A(X,\theta)$ 的矩母函数在某 $t>0$ 处有限，所以其和的母函数也在某 $t>0$ 处有限，这只要利用 Cauchy 不等式即可证）. 有了上述这些关于 $l_\nu^{(1)}(X,\theta+\varepsilon\nu)+\sigma A(X,\theta)$ 的性质，利用 [1] 的第七节定理 3.1，得下述引理

引理 6. $$\lim_{n\to\infty}\frac{1}{n}\log P_\theta\left\{\{l_{\nu_i}^{(1)}(X^{(n)},\theta+\varepsilon\nu_i)+\sigma\sum_{i=1}^{n}A(X_i,\theta)\geqslant 0\right\}$$
$$=\log\rho(\theta,\varepsilon,\nu_i,\sigma).$$

综合上述各引理，我们便得

$$(*)\qquad \lim_{n\to\infty}\frac{1}{n}\log P\{\|\hat\theta_n-\theta\|\geqslant\varepsilon\}\leqslant\max_{1\leqslant i\leqslant K}\log\rho(\theta,\varepsilon,\nu_i,\sigma)$$
$$\leqslant\sup_\nu\{\log\rho(\theta,\varepsilon,\nu,\sigma)\}.$$

最后证明 $\rho(\theta,\varepsilon,\nu,\sigma)$ 关于 σ 连续，而且在零点对于 ν **是均匀的.**

引理 7. $\rho(\theta,\varepsilon,\nu,\sigma)$ 对 σ 在零点关于 ν 均匀连续.

证. 令

$$\varphi_0(t,\sigma,\nu)=E_\theta\{\exp[t(l_\nu^{(1)}(X,\theta+\varepsilon\nu)+\sigma A(X,\theta)\}.$$

且设 $t(\sigma)$ 为使 $\varphi_0(t(\sigma),\sigma,\nu)=\rho(\theta,\varepsilon,\nu,\sigma)$ 者，于是

$$\int(l_\nu^{(1)}+\sigma A)e^{t(\sigma)(l_\nu^{(1)}+\sigma A)}f(x,\theta)d\mu=0.$$

利用隐函数法则对 σ 求导得

$$t'(\sigma)\int(l_\nu^{(1)}+\sigma A)^2e^{t(\sigma)(l_\nu^{(1)}+\sigma A)}f(x,\theta)d\mu$$
$$=-\int Ae^{t(\sigma)(l_\nu^{(1)}+\sigma A)}f(x,\theta)d\mu-t(\sigma)\int A(l_\nu^{(1)}+\sigma A)e^{t(\sigma)(l_\nu^{(1)}+\sigma A)}f(x,\theta)d\mu,$$

上式两边都是 ν 的连续函数. $t'(\sigma)$ 的系数恒大于 0. 由 $\{\nu:\ \nu\in R^q,\ \|\nu\|=1\}$ 的紧性，便知 $t'(\sigma)$ 有一与 ν 无关的界. 因此，$t(\sigma)$ 对 σ 关于 ν 均匀连续.

易见，$\varphi_0(t,\sigma,\nu)$ 对于 t 关于 σ, ν 均匀连续（σ 充分小），而

$$|\varphi_0(t(\sigma),\sigma,\nu)-\varphi_0(t(0),0,\nu)|$$
$$\leqslant|\varphi_0(t(\sigma),\sigma,\nu)-\varphi_0(t(0),\sigma,\nu)|$$
$$+|\varphi_0(t(0),\sigma,\nu)-\varphi_0(t(0),0,\nu)|,$$

由于 $t(\sigma)$ 对 σ 关于 ν 均匀连续，便得 $\varphi_0(t(\sigma),\sigma,\nu)$ 的均匀连续性，即 $\rho(\theta,\varepsilon,\nu,\sigma)$ 的均匀连续性.

有了此引理后，在(＊)式两边同时令 $\sigma\to 0$，并在右边把极限取进去，得

定理. $$\lim_{n\to\infty}\frac{1}{n}\log P\{\|\hat\theta_n-\theta\|>\varepsilon\}$$
$$\leqslant\sup_\nu\{\log\rho(\theta,\varepsilon,\nu,0)\}$$
$$=-\inf\{\log\phi(T_\theta(\varepsilon,\nu),\varepsilon,\nu)\}$$
$$=-\inf_\nu\left\{\frac{\varepsilon^2}{2!}(I^\nu(\theta))^2+\frac{\varepsilon^3}{3!}(3\mu_{110}^\nu-\mu_{300}^\nu)+\frac{\varepsilon^4}{4}[(2\mu_{400}^\nu-6\mu_{210}^\nu\right.$$

$$+ 4\mu_{101}^{\nu} + 3\mu_{020}^{\nu}) - 3(I^{\nu}(\theta))^2\gamma_{\nu}^2(\theta)] + o(\varepsilon^4)\Big\},$$

以 $\beta(\hat{\theta},\theta,\varepsilon)$ 记 $-\lim_{n\to\infty}\frac{1}{n}\log P\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}$，便有

$$\lim_{\varepsilon\to 0}\varepsilon^{-4}[B(\theta,\varepsilon)-\beta(\hat{\theta},\theta,\varepsilon)]\leqslant \sup_{\nu}\left\{\frac{1}{8}(I^{\nu}(\theta))^2\gamma_{\nu}^2(\theta)\right\}.$$

这样我们便证明了在条件 1，2，3 下，MLE 是一阶有效的，并且其指数速度与 Bahadur 界至多在 ε^4 的量级上出现差异，而且这一偏差在指数族情形下消失。因而这时的 MLE 是在 Fisher, Rao, Efron 意义下的二阶渐近有效估计。

下面我们举两个满足定理条件的分布族。

例 1. 在指数族 $e^{\theta^\tau x-\psi(\theta)}d\mu$ 之下，$\theta\in\boldsymbol{\Theta}$ 为 R^q 中的开凸集，且不退化。则

1. $f(x,\theta)=e^{\theta^\tau x-\psi(\theta)}$ 显然有共同支撑。

$$l(x,\theta)=\theta^\tau x-\psi(\theta),$$

$$l^{(2)}(\theta)=-\left(\frac{\partial^2\psi}{\partial\theta_i\partial\theta_j}\right)=-E(X-EX)(X-EX)^\tau<0.$$

2. 对任何 $\theta_0\in\boldsymbol{\Theta}$，取 $\delta>0$，使 $\{\|\theta-\theta_0\|<\delta\}\subset\boldsymbol{\Theta}$，则

$$\begin{aligned}|l_{\nu}^{(1)}(x,\theta+s\nu)| &= |\nu^\tau x-\nu^\tau\psi^{(1)}(\theta_0+s\nu)|\\ &\leqslant \|x\|+\|\psi^{(1)}(\theta_0+s\nu)\|\leqslant A_0(\theta,x),\text{（连续性）}\end{aligned}$$

$$\|l^{(1)}(x,\theta')-l^{(1)}(x,\theta'')\|\leqslant\|\psi^{(1)}(\theta')-\psi^{(1)}(\theta'')\|\leqslant A_1(\theta_0),$$

当 $\theta',\theta''\in N(\theta_0,\delta)$ 时成立。

而 $\|l^{(i)}(\theta')-l^{(i)}(\theta'')\|$ 在 $i\geqslant 2$ 时都与 x 无关，由连续性，定理条件中的 A_i 显然存在。

3. 因分布族不退化，$El_{\nu}^{(1)}(X,\theta)=0$，所以有

$$P\{l_{\nu}^{(1)}(X,\theta)>0\}>0,$$

再由连续性，必存在 $u(\theta)>0$，当 $0<\varepsilon<u$ 时

$$P\{l_{\nu}^{(1)}(X,\theta+\varepsilon\nu)>0\}>0.$$

由此可知，我们的定理对简单指数族是适用的。

例 2. Berkeson问题。

以 x 表示某种农药的剂量，$\pi_x=\dfrac{1}{(1+e^{\alpha+\beta x})}$ 表示在这一剂量下一个虫子被杀死的概率，样本为

剂量	$x_1,\cdots,x_k$,
杀死数	$l_1,\cdots,l_k$,
总数	$n_1,\cdots,n_k$.

似然函数是二项概率函数的积

$$\prod_{i=1}^{k}\binom{n_i}{l_i}(\pi_{x_i})^{l_i}(1-\pi_{x_i})^{n_i-l_i},$$

未知参数是 α,β.

$$l(Y,\theta)=C(Y)-\sum_{i=1}^{k}n_i\log(1+e^{\alpha+\beta x_i})+\sum_{i=1}^{k}(n_i-l_i)(\alpha+\beta x_i),$$

其中　$Y=(l_1,\cdots,l_k)^\tau,\ \theta=(\alpha,\beta)^\tau.$

它的一阶偏导为

$$\begin{pmatrix}\sum_{i=1}^{k}(n_i-l_i)-\sum[n_i/(1+e^{-(\alpha+\beta x_i)})]\\ \sum_{i=1}^{k}(n_i-l_i)x_i-\sum[n_ix_i/(1+e^{-(\alpha+\beta x_i)})]\end{pmatrix},$$

二阶偏导为

$$-\begin{pmatrix}\sum_{i=1}^{k}\frac{n_ie^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2} & \sum_{i=1}^{k}\frac{n_ix_ie^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2}\\ * & \sum_{i=1}^{k}\frac{n_ix_ie^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2}\end{pmatrix}.$$

因

$$\left[\sum_{i=1}^{k}\frac{n_ix_ie^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2}\right]^2=\left[\sum_{i=1}^{k}\frac{\sqrt{n_i}\,x_ie^{-\frac{1}{2}(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})}\cdot\frac{\sqrt{n_i}\,e^{-\frac{1}{2}(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})}\right]^2$$
$$\leqslant\left[\sum_{i=1}^{k}\frac{n_ie^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2}\right]\cdot\left[\sum_{i=1}^{k}\frac{n_ix_i^2e^{-(\alpha+\beta x_i)}}{(1+e^{-(\alpha+\beta x_i)})^2}\right],$$

等号成立当且仅当 x_i 全相等.

显然,在 x_i 全相同时,考虑 α,β 的估计是无意义的. 因此我们设其不全相同.

这时 $\left(\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right)_{2\times2}<0.$

由样本的有界性以及 l 的二阶导数存在连续，便知定理的其它二个条件显然满足. 因此本问题的 MLE 是二阶有效. 事实上,这里的分布是指数族.

例 3. 记号如上,但把 x_i 也看作随机变量,它有边际分布密度 $\varphi(x)$，且

$$\int e^{tx^2}\varphi(x)dx<\infty$$

在某 $t>0$ 处成立. 于是,似然函数就成了

$$\prod_{i=1}^{k}\binom{n_i}{l_i}\varphi(x_i)(\pi_{x_i})^{l_i}(1-\pi_{x_i})^{n_i-l_i}.$$

这就不再是指数分布族了. 但完全和上例相同,可验证它满足定理的三个条件,只要注意到 x 的分布连续,因而 x_i 全相同的概率为 0 即可.

最后,谨向不断给我有益指导的成平导师表示感谢.

参 考 文 献

[1] Bahadu, R. R., Some limit theorems in statistics, Society for Industrial and Applied Mathematics, Philadephia, 1971.

[2] Basu, D., On the concept of asymptotic efficiency. *Sankhya*, 17 (1956), 193—196.

[3] Fu, J. C., On a theorem of Bahadur on the rate of convergence of point estimators, *Ann. Statist*, 1(1973), 741—749.

[4] Fu, J. C., The rate of convergence of consistent point estimators, *Ann. Statist*, 3(1975), 234—240

[5] Fu, J. C., Large sample point estimations: A large deviation theory approach, *Ann. Statist*, 3(1982), 762—771.

[6] 成平，极大似然估计的渐近有效性，数学学报，**23**: 6(1980)，883—900.

THE LARGE DEVIATION OF MLE IN MULTIPARAMETER CASE

CHEN JIA-HUA

(Institute of Systems Science, Academia Sinica)

ABSTRACT

In this paper the exponential rates of decrease and bounds on tail probabilities for consistent estimators in multiparameter case are studied using the large deviation method. The asymptotic expansions of Bahadur bounds and the upper bounds of exponetial rates in the case of maximum likelihood estimator are obtained. Based on those results we obtain a result parallel to the result of J. C. Fu, which deals with the one-parameter case.