2000,20(3):370—377

Mathematica Acta Scientia
数学物理学报

一种Sieve极大似然估计的渐近性质*

宋立新　　　　薛宏旗
（吉林大学数学系　长春130021）　（中国科技大学研究生院　北京100039）

摘要　该文针对部分线性模型，在响应变量的观测值为I型区间删失数据的情形下，讨论Sieve极大似然估计的渐近性质．用三角级数来构造Sieve空间，在一定条件下证明了该估计具有强相合性；得到了该估计的弱收敛速度，并且非参数部分的估计达到了最优收敛速度；还算出了参数部分的信息界．

关键词　部分线性模型，区间删失，Sieve，强相合，信息界．

MR(1991)主题分类　62F

1　引言

部分线性模型是一种半参数回归模型，由Engle R F等[1]在研究用电量与天气、收入以及季节等变量之间的关系时首先提出，其后，在理论和应用两方面都得到了发展，特别是我国学者在理论方面做了不少工作．

关于此模型，大部分文献都是研究完全数据情形，文献[2]较全面地介绍了这方面的研究成果．但在许多实际问题，如生存分析、可靠性寿命试验等方面的研究中，观测数据往往是不完全的，例如右删失数据、区间删失数据和分组数据等．对于右删失情形的部分线性模型，文献[3]用核估计代替非参数部分研究了估计的渐近性质．区间删失分为I型和II型两类，大量发生在工程、医学和经济学等领域．这类数据所提供的信息比右删失更少，更难以研究，也是当前国外研究较多的数据类型．关于区间删失数据类型，[4]等文研究了线性模型、Cox模型和比例差异模型(proportional odds model)以及其它一些模型下估计的性质，但还没有文献讨论部分线性模型．

本文在响应变量的观测值为I型区间删失数据的情形下，讨论部分线性模型Sieve极大似然估计(简写为MLE)的渐近性质．我们用三角级数来构造Sieve空间(定义见正文)，此种构造方法文献[5][6]使用较多，文献[4]中是用分段线性函数，还有利用B-样条以及小波的在第2节中介绍模型和似然函数，第3节介绍Sieve空间和Sieve MLE，第4节证明了该估计的强相合性，第5节求该估计的弱收敛速度，第6节求出了参数部分的信息界，但参数部分的Sieve MLE的渐近正态性没有解决．II型区间删失情形与此类似，本文不予讨论．

在本文中，$\|\cdot\|$ 为欧氏模，$\|\cdot\|_\infty$为一致模，$\|\cdot\|_{P_2}$为L_{P_2}模，即对于向量$a=(a_1,\cdots,a_m)^T$和函数f，$\|a\|=(a_1^2+\cdots+a_m^2)^{\frac{1}{2}}$，$\|f\|_\infty=\sup_t|f(t)|$，$\|f\|_{P_2}=(\int f^2dP)^{\frac{1}{2}}$．

* 国家自然科学基金资助项目
1999-03-24日收到

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$I_{[A]}$ 表示集合 A 的示性函数.

2 模型和似然函数

考虑以下的部分线性模型

$$Y = \beta^T X + g(T) + \varepsilon, \tag{1}$$

其中:未知参数 $\beta=(\beta_1,\cdots,\beta_d)^T\in A$, A 是空间 R^d 中有界闭集,g 是未知的光滑函数;自变量(X,T)可观测,其联合密度 $\varphi(x,t)$不依赖于(β,g),而且 $T\in[0,1]$,X 有界,即存在一个常数 M, 使得 $P\{\|X\|\leqslant M\}=1$;响应变量 Y 不能直接观测,只能得到它的 I 型区间删失观测值,即有一个有界随机变量 Z,使 $\delta=I_{[Y\leqslant Z]}$和 Z 是可观测的(注意:I 型区间删失与右删失有区别,右删失是当 $Y\leqslant Z$ 时,可以观测到 Y 的确切值,而区间删失则不能);在本文中,Z 与 (X,T) 独立,其密度 $h(z)$不依赖于(β,g);随机误差 ε 与(X,T,Z)独立,其分布函数 F 完全已知,密度 $f>0,E\varepsilon=0$.

记 $\theta=(\beta,g)^T$,并称之为模型参数,习惯称 β 为参数部分,g 为非参数部分,$\theta_0=(\beta_0,g_0)^T$ 为参数真值.令

$$B=\{g\in C^m[0,1]:\|g^{(j)}\|\leqslant L_j,j=0,\cdots,m,$$
$$|g^{(m)}(t_1)-g^{(m)}(t_2)|\leqslant L_{m+1}|t_1-t_2|^\gamma\},$$

其中 $m+\gamma$ 称为光滑度,$m\geqslant1$, $L_j,j=0,\cdots,m+1$ 都已知.设参数空间为 $\Theta=\{\theta:\theta\in A*B\}$.

记 $W=(X,T,Z,\delta)^T$,w 为 W 的观测值,W 的密度函数为

$$[F(Z-\beta^TX-g(T))]^\delta[1-F(Z-\beta^TX-g(T))]^{1-\delta}\varphi(X,T)h(Z),$$

控制测度为由$\{0,1\}$上的计数测度与 $\mathscr{R}^3$ 上的 Lebesgue 测度组成的乘积测度.记相应的似然函数为 $Q(\theta,w)$,对数似然函数为 $l(\theta,w)=\log Q(\theta,w)$.

用 P_θ表示模型参数为 θ 时 W 的概率分布,E_0 表示相应于 P_{θ_0}求期望.

可识性:利用函数 $\log x$ 的凸性以及 Jensen 不等式,可知 θ_0 是 $E_0l(\theta,W)$的一个最大值点.若假设 $\log f(x)$严凸,则 θ_0 是 $E_0l(\theta,W)$的唯一最大值点.事实上,根据[7],知 $\log F(x)$也严凸,$\log(1-F(x))$亦然,于是结论成立.

3 Sieve 空间和 Sieve MLE

由于参数空间 Θ 是无穷维的,我们用一列有穷维空间来逼近它,并限制在这些有穷维空间上估计 θ,从而得到 θ 的估计序列,这就是所谓 Sieve 方法.具体地说,设 ρ 是 Θ 上的距离,$L_n=L_n(\theta,\tilde{X}_n)$是观测值 $\tilde{X}_n$ 和 θ 的函数,用来反映参数为 θ 时,模型与观测值的拟合程度,称之为经验准则函数;又 $\Theta_1,\Theta_2,\cdots,\Theta_n,\cdots$是一列与 Θ 相交的集合,大多数情形下,它们是 Θ 的子集,可用来逼近 Θ,即对任意 $\theta\in\Theta$,都存在 $\pi_n\theta\in\Theta_n$, 使当 $n\to\infty$时,$\rho(\pi_n\theta,\theta)\to0$,文献[8]称$\{\Theta_n\}$为 Sieve 空间,称满足如下条件的 $\hat{\theta}_n$ 为 Sieve MLE

$$L_n(\hat{\theta}_n,\tilde{X}_n)\geqslant\sup_{\theta\in\Theta_n}L_n(\theta,\tilde{X}_n)-\eta_n, \tag{2}$$

其中 $\eta_n\to0,n\to\infty$.

现在回到第 2 节模型.设 $W_i=(X_i,T_i,Z_i,\delta_i)^T(i=1,2,\cdots,n)$是 i.i.d. 样本,记 $\tilde{W}_n=(W_1,\cdots,W_n)^T$,$P_n$ 是相应的经验概率测度,我们来构造 θ_0 的 Sieve MLE.

对 $\theta_i=(\beta_i,g_i)^T\in\Theta,i=1,2$,定义距离 ρ

$$\rho(\theta_1,\theta_2)=(E_0((\beta_1-\beta_2)^TX+g_1(T)-g_2(T))^2)^{\frac{1}{2}}. \tag{3}$$

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

设 $$B_n=\{g\in B: g(t)=\alpha_0+\sum_{j=1}^{r_n}(\alpha_j\cos 2\pi jt+\beta_j\sin 2\pi jt),$$

$$\alpha_0^2+\sum_{j=1}^{r_n}j^{2p}(\alpha_j^2+\beta_j^2)\leqslant C_1, |\alpha_0|+\sum_{j=1}^{r_n}j^p(|\alpha_j|+|\beta_j|)\leqslant C_2\},$$

其中 r_n 待定,记 $r_n=n^{2\tau},\tau>0$. 当 $m=1$ 时, $p=1$;当 $m\geqslant 2$ 时, $p=2$. 常数 C_1,C_2 定义如下:对于 $g\in B_n,k\geqslant 1$ 时,易算得

$$\|g^{(k)}\|_\infty\leqslant\sum_{j=1}^{r_n}(2\pi j)^k(|\alpha_j|+|\beta_j|),$$

$$\|g^{(k)}\|_2^2\leqslant\frac{1}{2}\sum_{j=1}^{r_n}(2\pi j)^{2k}(\alpha_j^2+\beta_j^2),$$

又因为 $g\in B$,于是可定义

$$C_1=\max_{0\leqslant j\leqslant p}2L_j^2/(2\pi)^{2j}, C_2=\max_{0\leqslant j\leqslant p}L_j/(2\pi)^j.$$

根据文[9]定理 8 和定理 9(P_{62}),对任意 $g\in B$,存在一个 $\pi_n g\in B_n$,使得

$$\|\pi_n g-g\|_\infty\leqslant O(\frac{1}{r_n^{m+\gamma}}).$$

记 $\Theta_n=A*B_n$,于是对任意 $\theta=(\beta,g)^T\in\Theta$,存在一个 $\pi_n\theta=(\beta,\pi_n g)^T\in\Theta_n$,使得

$$\rho(\pi_n\theta,\theta)\leqslant\|\pi_n g-g\|_\infty\leqslant O(\frac{1}{r_n^{m+\gamma}}),\tag{4}$$

当 $n\to+\infty$, $r_n\to+\infty$,有 $\rho(\pi_n\theta,\theta)\to 0$. 于是 Θ_n 可作为 Θ 的 Sieve 空间. 在(2)中取经验准则函数为 $L_n(\theta,\widetilde{W}_n)=P_n l(\theta,w)=\frac{1}{n}\sum_{i=1}^{n}l(\theta,W_i)$, $\eta_n=0$,记 Sieve MLE 为 $\hat\theta_n=(\hat\beta_n,\hat g_n)$. 因为 Θ_n 为有界闭集, L_n 连续, $\hat\theta_n$ 一定存在.

4 强相合

对于列向量 a,记 $a^{\otimes 2}=aa^T$. 记 $\xi(\theta,W)=Z-\beta'X-g(T)$, $D(\xi)=\frac{\partial}{\partial\xi}l(\theta,W)=\delta\frac{f(\xi)}{F(\xi)}-(1-\delta)\frac{f(\xi)}{1-F(\xi)}$.

定理 1 在前 3 节的条件和记号下,有 $\rho(\hat\theta_n,\theta_0)\to 0$, a. s. P_{θ_0}. 若下面条件成立

$$E_0(X-E_0(X|T))^{\otimes 2}>0,\tag{5}$$

则进一步有

$$\|\hat\beta_n-\beta_0\|\to 0, \|\hat g_n-g_0\|_2\to 0, \text{ a. s. } P_{\theta_0}.$$

证明定理 1 需以下几个引理.

引理 1 定义两个函数集合 $\mathscr{F}=\{f(g(\cdot))\}$, $\mathscr{G}=\{g(\cdot)\}$,如果对于任意 $f(g_1)$, $f(g_2)\in\mathscr{F}$,存在常数 C,使得

$$|f(g_1)-f(g_2)|\leqslant C|g_1-g_2|,$$

则对于任意概率测度 P,集合 $\mathscr{F}$ 关于测度 P 的 ϵ-覆盖数 $N(\epsilon,\mathscr{F},\|\cdot\|_P)$(含义见[10]定义 II. 23, II. 32)满足

$$N(\epsilon,\mathscr{F},\|\cdot\|_{P_2})\leqslant N(\frac{\epsilon}{C},\mathscr{G},\|\cdot\|_{P_2}).$$

证 利用覆盖数定义. 设存在 s 个函数 $g_1,\cdots,g_s$,对于任意 $g(x)\in\mathscr{G}$,有 $\min_i(P(g_i-$

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$g)^2)^{\frac{1}{2}}\leqslant\varepsilon$. 则 $f(g_1),\cdots,f(g_s)$ 可覆盖 $\mathscr{F}$,因为

$$\min_i(P(f(g_i)-f(g))^2)^{\frac{1}{2}}\leqslant C\min_i(P(g_i-g)^2)^{\frac{1}{2}}\leqslant C\varepsilon.$$

引理 2 记 $\mathscr{F}_n=\{l(\theta,.)-l(\pi_n\theta_0,.):\theta\in\Theta_n\}$,则该集合的熵 $H(\varepsilon,\mathscr{F}_n,\|\cdot\|_\infty)=\log N(\varepsilon,\mathscr{F}_n,\|\cdot\|_\infty)$ 满足

$$H(\varepsilon,\mathscr{F}_n,\|\cdot\|_\infty)\leqslant K\varepsilon^{-\frac{1}{p}},$$

K 为常数.

证 因为对于任意 $\theta_1,\theta_2\in\Theta_n$,由 Taylor 展式有

$$\begin{aligned}&|l(\theta_1,W)-l(\theta_2,W)|\\ &=|D(\xi')|*|((\beta_1-\beta_2)^TX+g_1(T)-g_2(T)|\\ &\leqslant C*(\|\beta_1-\beta_2\|+\|g_1-g_2\|_\infty),\end{aligned}$$

其中 ξ' 是一个 $\xi(\theta_0,W)$ 与 $\xi(\theta,W)$ 之间的数,有界,C 是一个常数.于是利用引理 1,得

$$N(\varepsilon,\mathscr{F}_n,\|\cdot\|_\infty)\leqslant N(\frac{\varepsilon}{2},A,\|\cdot\|)*N(\frac{\varepsilon}{2},B_n,\|\cdot\|_\infty).$$

设集合 A 的直径为 M_1,利用[11]引理 4.1,有

$$N(\frac{\varepsilon}{2},A,\|\cdot\|)\leqslant(\frac{3M_1}{\frac{\varepsilon}{2}})^d=(\frac{6M_1}{\varepsilon})^d.$$

下面算 $N(\frac{\varepsilon}{2},B_n,\|\cdot\|_\infty)$.利用[12]定理 2.7.1(记号也见该文),

$$\log N(\varepsilon,C_1^\alpha(\mathscr{X}),\|\cdot\|_\infty)\leqslant K\lambda(\mathscr{X}')(\frac{1}{\varepsilon})^{\frac{d}{\alpha}},$$

现在 $\mathscr{X}=[0,1]$,d 为其维数,即 $d=1$. $\mathscr{X}'=\{x:\|x-\mathscr{X}\|<1\}$,$\lambda(\mathscr{X}')$ 是集合 $\mathscr{X}'$ 的 Lebesgue 测度,$\alpha=p$. $C_C^\alpha(\mathscr{X})\overset{\text{def}}{=}\{f:\mathscr{X}\to R,f\text{ 连续},\|f\|_\alpha\leqslant C\}$,计算

$$\begin{aligned}\|g\|_\alpha&=\max_{k\leqslant p-1}\sup_t|g^{(k)}(t)|+\max_{k=p-1}\sup_{x,y}\frac{|g^{(k)}(x)-g^{(k)}(y)|}{|x-y|}\\ &\leqslant\max_{k\leqslant p-1}\sup_t|g^{(k)}(t)|+\max_{k=p-1}\sup_\xi|g^{(k+1)}(\xi)|\\ &\leqslant\max_{k\leqslant p}\sup_t|g^{(k)}(t)|\\ &\leqslant\max_{k\leqslant p}L_k.\end{aligned}$$

令 $C=\max_{k\leqslant p}L_k$,可见 $\{g:g\in B_n\}\subset C_C^\alpha(\mathscr{X})$,于是

$$H(\frac{\varepsilon}{2},B_n,\|\cdot\|_\infty)\leqslant H(\frac{\varepsilon}{2},C_C^\alpha(\mathscr{X}),\|\cdot\|_\infty)\leqslant K(\frac{2C}{\varepsilon})^{\frac{1}{p}}.$$

则

$$H(\varepsilon,\mathscr{F}_n,\|\cdot\|_\infty)\leqslant d\log\frac{6M_1}{\varepsilon}+K(\frac{2C}{\varepsilon})^{\frac{1}{p}}\leqslant K'\varepsilon^{-\frac{1}{p}}.$$

引理 3 当 $E_0(X-E_0(X|T))^{\otimes 2}>0$ 时,设其最小特征根为 λ_1,我们可将 $\rho(\theta,\theta_0)$ 中参数部分 β 与非参数部分 g 分开,

$$\|\beta-\beta_0\|\leqslant\frac{1}{\sqrt{\lambda_1}}\rho(\theta,\theta_0),\ \|g-g_0\|_2\leqslant(1+\frac{M}{\sqrt{\lambda_1}})\rho(\theta,\theta_0).$$

证

$$\rho(\theta,\theta_0)^2=E_0[(\beta-\beta_0)^TX+g_n(T)-g_0(T)]^2$$

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$= E_0[(\beta-\beta_0)^T(X-E_0(X|T))+(\beta-\beta_0)^TE_0(X|T)+g_n(T)-g_0(T)]^2,$$

我们可以证明交叉项

$$E_0[(\beta-\beta_0)^T(X-E_0(X|T))]*[(\beta-\beta_0)^TE_0(X|T)+g_n(T)-g_0(T)]=0.$$

事实上，若令 $J(T)=(\beta-\beta_0)^TE_0(X|T)+g_n(T)-g_0(T)$，有

$$\begin{aligned}&E_0(X-E_0(X|T))*J(T)\\&=E_0XJ(T)-E_0E_0(X|T)J(T)\\&=E_0XJ(T)-E_0E_0(XJ(T)|T)\\&=E_0XJ(T)-E_0XJ(T)=0.\end{aligned}$$

那么

$$E_0[(\beta-\beta_0)^T(X-E_0(X|T))]^2\leqslant\rho(\theta,\theta_0)^2,$$
$$(\beta-\beta_0)^TE_0(X-E_0(X|T))^{\otimes 2}(\beta-\beta_0)\leqslant\rho(\theta,\theta_0)^2,$$

于是有

$$\lambda_1(\beta-\beta_0)^T(\beta-\beta_0)\leqslant\rho(\theta,\theta_0)^2,$$
$$\|\beta-\beta_0\|\leqslant\frac{1}{\sqrt{\lambda_1}}\rho(\theta,\theta_0),$$

另外，

$$\begin{aligned}&[E_0(g_n(T)-g_0(T))^2]^{\frac12}\\&\leqslant[E_0((\beta-\beta_0)^TX)^2]^{\frac12}+[E_0((\beta-\beta_0)^TX+g_n(T)-g_0(T))^2]^{\frac12}\\&\leqslant M\|\beta-\beta_0\|+\rho(\theta,\theta_0)\\&=(1+\frac{M}{\sqrt{\lambda_1}})\rho(\theta,\theta_0).\end{aligned}$$

▌

定理 1 证明 当 ε 固定时，由引理 2 知

$$\log N(\varepsilon,\mathscr{F}_n,P_n)\leqslant\log N(\varepsilon,\mathscr{F}_n,\|\cdot\|_\infty)\leqslant K\varepsilon^{-\frac1p}=o_p(n),$$

于是由[10]定理 II. 24，有

$$\sup_{\mathscr{F}_n}|P_nl-Pl|\to 0.\quad \text{a.s.}\ P_{\theta_0},\tag{6}$$

于是

$$\begin{aligned}&P_nl(\pi_n\theta_0,W)-Pl(\pi_n\theta_0,W)+Pl(\pi_n\theta_0,W)-Pl(\theta_0,W)\\&\leqslant P_nl(\hat\theta,W)-Pl(\theta_0,W)\leqslant P_nl(\hat\theta,W)-Pl(\hat\theta,W),\end{aligned}$$

由 (6) 得 $P_nl(\pi_n\theta_0,W)-Pl(\pi_n\theta_0,W)\to0$，$P_nl(\hat\theta,W)-Pl(\hat\theta,W)\to0$ 再根据 (4)，有 $Pl(\pi_n\theta_0,W)-Pl(\theta_0,W)\to0$，因此 $|P_nl(\hat\theta,W)-Pl(\theta_0,W)|\to0$. a.s..

进一步，有

$$\begin{aligned}&|Pl(\hat\theta,W)-Pl(\theta_0,W)|\\&\leqslant|P_nl(\hat\theta,W)-Pl(\hat\theta,W)|+|P_nl(\hat\theta,W)-Pl(\theta_0,W)|\\&\to0.\ \text{a.s.}.\end{aligned}$$

由于 K-L 信息量大于 Hellinger 距离的平方([13]，P_{346})，即

$$\begin{aligned}&|Pl(\hat\theta,W)-Pl(\theta_0,W)|=E_0(l(\theta_0,W)-l(\theta,W))\\&\geqslant\|Q(\theta_0,W)^{\frac12}-Q(\theta,W)^{\frac12}\|_2^2\\&=\|\frac{q(\xi')}{2Q^{\frac12}(\xi')}[(\hat\beta_n-\beta_0)^TX+\hat g_n(T)-g_0(T)]\|_2^2,\end{aligned}\tag{7}$$

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

其中 $q(\xi)=f(\xi)^{\delta}(-f(\xi))^{1-\delta}$，$\varphi$ 与 h 已在第一个等式中消掉，所以此时 $Q(.,.)$ 中不含有它们 ξ' 是 $\xi(\theta_0,W)$ 与 $\xi(\theta,W)$ 之间的一个量，有界，因此 $\frac{q(\xi')}{2Q^{\frac{1}{2}}(\xi')}$ 也有界，于是

$$\rho(\hat{\theta},\theta_0)\to 0,\text{a.s.}.$$

利用引理 3 将参数部分与非参数部分分开，可得定理 1. ▌

5 弱收敛速度

定理 2 在第 1 节～第 3 节的条件和记号下，有 $\rho(\hat{\theta}_n,\theta_0)=O_p(\max(n^{-\frac{p}{1+2p}},n^{-2\tau(m+\gamma)}))$. 若选择 $\tau=\frac{p}{2(1+2p)(m+\gamma)}$，可得到最优弱收敛速度 $\rho(\hat{\theta}_n,\theta_0)=O_p(n^{-\frac{p}{1+2p}})$. 若条件(5)成立，则进一步有

$$\|\hat{\beta}_n-\beta_0\|=O_p(n^{-\frac{p}{1+2p}}),\ \|\hat{g}_n(T)-g_0(T)\|_2=O_p(n^{-\frac{p}{1+2p}}),$$

这与文献[14]提出的非参数最优收敛速度一致.

定理 2 证明 验证文[6] 定理 1 条件.

C1 同(7)一样，可得

$$\begin{aligned}&\inf_{\rho(\theta,\theta_0)\geqslant\varepsilon,\theta\in\Theta_n}E(l(\theta_0,W)-l(\theta,W))\\ \geqslant&\inf_{\rho(\theta,\theta_0)\geqslant\varepsilon,\theta\in\Theta_n}\|(\beta-\beta_0)^TX+g(T)-g_0(T)\|_2^2\\ \geqslant&C\varepsilon^2,\end{aligned}$$

可见 C1 成立，$\alpha=1$.

C2 利用中值定理，有

$$\begin{aligned}&\mathrm{Var}(l(\theta_0,W)-l(\theta,W))\leqslant E_0(l(\theta_0,W)-l(\theta,W))^2\\ &=E_0D(\xi')^2[(\beta-\beta_0)^TX+g(T)-g_0(T))]^2,\end{aligned}$$

其中 ξ' 位于 $\xi(\theta,W)$ 与 $\xi(\theta_0,W)$ 两者之间，

$$\begin{aligned}&\sup_{\rho(\theta,\theta_0)\leqslant\varepsilon,\theta\in\Theta_n}\mathrm{Var}(l(\theta_0,W)-l(\theta,W))\\ \leqslant&C\sup_{\rho(\theta,\theta_0)\leqslant\varepsilon,\theta\in\Theta_n}\|(\beta-\beta_0)'X+g(T)-g_0(T)\|_2^2\\ \leqslant&C\varepsilon^2,\end{aligned}$$

C2 成立，$\beta=1$.

由引理 2 知 C3 成立，$r_0=0$，$r=\frac{1}{p}$. 另外，

$$\rho(\pi_n\theta_0,\theta_0)=O_p(n^{-2\tau(m+\gamma)}),$$

$$\begin{aligned}K(\pi_n\theta_0,\theta_0)&\overset{\text{def}}{=}E_0(l(\theta_0,W)-l(\pi_n\theta_0,W))\\ &=-E_0l^{(1)}(\theta_0,W)(\pi_n\theta_0-\theta_0)-E_0l^{(2)}(\xi',W)(\pi_n\theta_0-\theta_0)^2\\ &=0-E_0l^{(2)}(\xi',W)(\pi_n\theta_0-\theta_0)^2\\ &\leqslant CE_0(\pi_n\theta_0-\theta_0)^2=O_p(n^{-4\tau(m+\gamma)}),\end{aligned}$$

于是利用[6]定理 1，得

$$\rho(\hat{\theta}_n,\theta_0)=O_p(\max(n^{-\frac{p}{1+2p}},n^{-2\tau(m+\gamma)})).$$

再利用引理 3 可将速度中参数部分 β 与非参数部分 g 分开. ▌

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

6 信息界

记 $C_1=E_0D(\xi(\theta_0,W))^2$，$J(w)=C_1^{-1}D(\xi(\theta_0,w))^2Q(\theta_0,w)$，显然 J 是一个密度函数，E_JX 表示 X 的密度为 J 时的期望.

定理 3(有效得分函数与信息界) 若 θ_0 是 Θ 的内点，条件(5)满足，以及第 1 节～第 3 节的记号和条件下，按照[15]中第三章的定义和方法算得 β 的有效得分函数为

$$l_\beta^* = -D(\xi(\theta_0,W))(X-E_J(X|T)),$$

信息阵为

$$I(\beta_0)=E_0(l_\beta^* l_\beta^{*T})=C_1E_J(X-E_J(X|T))^{\otimes 2},$$

其逆称为信息界.

定理 3 简要证明 利用[15]Corollary 3.4.1.

首先，我们计算 β 和 g 的得分函数，因为 φ 与 h 中不含模型参数，可以不予考虑. 对于 β 的得分函数就是对数似然关于 β 的偏导数，即

$$\dot{l}_\beta=-D(\xi(\theta_0,W))X.$$

下面假设 $B_0=\{g_\eta(.):g_\eta(t)|_{\eta=0}=g_0(t),|\eta|<1\}$，它是集合 B 的一个正则参数子族(含义见[15]定义 2.1.2). 设

$$\frac{\partial}{\partial\eta}g_\eta(t)|_{\eta=0}=a(t),$$

那么，$a(t)\in L_2(P_0)\overset{\text{def}}{=\!=\!=}\{f:\int f^2\mathrm{d}P_0<\infty\}$. g 的得分算子为

$$\dot{l}_g(a)=\frac{\partial}{\partial\eta}l(\beta_0,g_\eta)|_{\eta=0}=-D(\xi(\theta_0,W)a.$$

按照[15]第三章的方法，要算 β 的信息阵，需计算其有效得分函数 l_β^*. 记 $\dot{P}_2$ 为 $\dot{l}_g(a)$ 在 $L_2(P)$ 上的闭线性张空间，l_β^* 可解释为 $\dot{l}_\beta$ 在 $\dot{P}_2$ 的正交余集上的投影，取 $a_*\in\dot{P}_2$，$\dot{l}_g(a_*)=\Pi_0(\dot{l}_\beta|\dot{P}_2)$($\Pi_0$ 表示投影)，则

$$l_\beta^*=\dot{l}_\beta-\dot{l}_g(a_*).$$

显然，a_* 应是使

$$E_0(\dot{l}_\beta-\dot{l}_g(a))^2,$$

达到最小的 a. 记 $\alpha(a)=E_0D(\xi(\theta_0,W))^2(X-a(T))^2$，则

$$\alpha(a)=\int(X-a(T))^2D(\xi(\theta_0,W))^2Q(\theta_0,W)\mathrm{d}\delta\mathrm{d}X\mathrm{d}T\mathrm{d}Z,$$

由定理 3 前的记号，有

$$\alpha(a)=C_1E_J(X-a(T))^2.$$

显然，$a_*(T)=E_J(X|T)\in L_2(P_0)$，则

$$l_\beta^*=-D(\xi(\theta_0,W))(X-E_J(X|T)),$$

β 的信息阵为

$$\begin{aligned}I(\beta_0)&=E_0(l_\beta^* l_\beta^{*T})\\&=E_0D(\xi(\theta_0,W))^2(X-E_J(X|T))^{\otimes 2}\\&=C_1E_J(X-E_J(X|T))^{\otimes 2}.\end{aligned}$$

虽然本文中参数部分的 Sieve MLE. $\hat{\beta}_n$ 的渐近正态性没有解决，但有了信息界之后，可以研究各种估计的方差与信息界差别有多大，以及什么样的估计是渐近有效的，这都是一些

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

有意义的问题.

作者在写作过程中，得到了中科院系统所李国英研究员大力帮助，在此表示衷心感谢！

参考文献

1 Engle R, Granger C, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. JASA, 1986, **81**(394): 310—320

2 柴根象，洪圣岩编著. 半参数回归模型. 合肥：安徽教育出版社，1995

3 王启华，郑忠国. 随机删失半参数回归模型中估计的渐近性质. 中国科学(A 辑)，1997，**27**(7)：583—594

4 Huang Jian, Rossini A J. Sieve estimation for the proportional-odds failure-time regression model with interval censoring. JASA 1997, **92**(439): 960—967

5 Shen Xiaotong. On methods of sieves and penalization. Ann Stat, 1997, **25**(6): 2555—2597

6 Shen Xiaotong, Wong Winghung. Convergence rate of sieve estimates. Ann Stat, 1994, **22**(2): 580—615

7 Burridge J. A note on maximum likelihood estimation for regression models using grouped data. JRSS B, 1981, **43**(1): 41—45

8 Grenander U. Abstract inference. New York: Wiley, 1981

9 Lorentz G G. Approximation of functions. New York: Holt, Rinehart and Winston, 1966

10 Pollard D. Convergence of stochastic processes. New York: Springer-Verlag, 1984

11 Pollard D. Empirical processes theory and applications. California: Institute of Mathematical Statistics, Hayward, 1990

12 Vaart A W, Wellner J A. Weak convergence and empirical processes. New York: Springer-Verlag, 1996

13 Wong Winghung, Shen Xiaotong. Probability inequalities for likelihood ratios and convergence rates of sieve MLE. Ann Stat, 1995, **23**(2): 339—362

14 Stone C J. Optimal global rates of convergence for nonparametric regression. Ann Stat, 1982, **10**(4): 1040—1053

15 Bickel P J. Klaassen C A J, Ritov Y, Wellner J A. Efficient and adaptive estimation for semiparametric models. The Johns Hopkins University Press, 1993

Asymptotic Properties of a Kind of Sieve MLE

Song Lixin

(*Department of Mathematics, Jilin University, Chang chun* 130023)

Xue Hongqi

(*Graduate School at Beijing, University of Science and Technology of China, Beijing* 100039)

Abstract For a partly linear model, when observations of the respond variable are case one interval censored, asymptotic properties of sieve MLEs are discussed. Triangle series are used to struct sieve spaces, under mild conditions, sieve MLEs are shown to be strong consistent, sieve MLE of the nonparametric part has an optimal convergence rate, and the information bound of sieve MLE of the parametric part is obtained.

Key words Partly linear model, Interval censoring, Sieve MLE, Strong consistent, Information bound.

MR(1991) Subject Classification 62F

© 1994-2006 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net