删失非线性模型的中位数回归

王海康, 刘网定, 周秀轻

(南京师范大学数学与计算机科学学院, 江苏 南京 210097)

[摘要] 研究了带有不完全数据的非线性模型的中位数回归问题. 将完全数据线性回归模型的 L_1 方法推广到随机右删失非线性中位数回归模型中, 提出了一种估计非线性中位数回归模型参数的半参数方法, 并得到了估计量的强相合性和渐近正态性.

[关键词] 非线性模型, 中位数回归, 删失

[中图分类号] O212.3 [文献标识码] A [文章编号] 1001-4616(2007)02-0027-06

Median Regression for Censoring Nonlinear Model

Wang Haikang Liu Wangding Zhou Xiuqing

(School of Mathematics and Computer Science, Nanjing Normal University, Nanjing 210097, China)

Abstract Median regression for nonlinear model with incomplete observations is studied. A semiparametric procedure is proposed to estimate the parameter in nonlinear median regression model with randomly right censored data. The strong consistency and asymptotic normality of the estimator are also obtained.

Key words nonlinear model, median regression, censoring

0 引言

删失数据回归模型是生存分析中一个很有价值且常用的模型. 带有右删失数据的参数回归模型为

$$T_i = f(\beta, \boldsymbol{X}_i) + e_i, \qquad i = 1, 2, \cdots, n, \tag{1}$$

其中 β 是 $\mathbf{R}^p$ 中未知的参数向量, $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ 是协变量, $e_1, \cdots, e_n$ 是独立同分布的误差项. 由于删失的存在, 观测值为 $Y_i = \min(T_i, C_i)$ 和 $\delta_i = I(T_i \leqslant C_i)$ 而非 T_i, 其中 $I[A]$ 为集合 A 的指示函数, $C_1, \cdots, C_n$ 为独立同分布的随机删失变量, 且与 $T_1, \cdots, T_n$ 独立. 在删失线性回归模型 (即在模型 (1) 中 $f(\beta, \boldsymbol{X}_i) = \beta'\boldsymbol{X}_i$) 下, 文献 [1] ~ [5] 等给出了参数 β 的众多最小二乘估计 (简称 L_2 估计). 这些估计方法大部分都需要调整删失数据, 然后对调整后的数据使用普通最小二乘.

曾有文献对删失数据已知时回归模型 (1) 的 L_1 估计做过研究. 文献 [6] ~ [8] 研究了观测数据为 $Y_i = \min(T_i, 0)$, 即删失点固定且为 0时的删失线性回归模型参数的 L_1 估计, 并证明了该模型参数的 L_1 估计的渐近正态性. 文献 [9] 在误差项 e_i 的条件中位数为 0的条件下, 针对删失线性回归模型参数的统计推断, 提出了一个估计方程的方法, 并证明了该统计量的强相合性和渐近正态性. 文献 [10] 研究了在随机删失变量 C_i 均已知的条件下删失回归模型 (1) 的 L_1 估计, 并证明了参数 L_1 估计量的渐近正态性.

本文考虑回归模型 (1), 并假设删失变量 $C_1, \cdots, C_n$ 有相同的连续生存函数 G, β_0 是参数 β 的真值. 我们将文献 [9] 中提出的估计方程的方法应用于回归问题 (1), 并研究了该模型参数估计量的渐近性质.

1 估计方程

为得到参数 β 的估计, 首先回顾文献 [9] 中的方法. 该文考虑在无删失且误差项 e_i 的条件中位数为 0

收稿日期: 2006-05-16. 修回日期: 2006-07-11.
基金项目: 国家自然科学基金 (10626028)资助项目.
作者简介: 王海康 (1981—), 硕士研究生, 主要从事生存分析的学习与研究. E-mail: whk_lk@chinaren.com
通讯联系人: 周秀轻 (1975—), 女, 博士, 讲师, 主要从事生存分析和回归分析的教学与研究. E-mail: zhouxiuqing@njnu.edu.cn

的条件下,线性回归模型(即模型(1)$f(\beta, \boldsymbol{X}_i) = \beta' \boldsymbol{X}_i$)参数的 L_1估计. 在这种情况下,参数 β的 L_1估计是极小化问题

$$\min_{\beta} \sum_i | T_i - \beta' \boldsymbol{X}_i |$$

的解. 该问题等价于

$$\min_{\beta} \sum_i \{I(T_i \geqslant \beta' \boldsymbol{X}_i) - \frac{1}{2}\}(T_i - \beta' \boldsymbol{X}_i),$$

其中记号"$\sum_i$"代表"$\sum_{i=1}^{n}$",本文中将一直使用此简便记号. 注意到上述问题的解是方程

$$\boldsymbol{U}_n(\beta) = \sum_i \{I(T_i \geqslant \beta' \boldsymbol{X}_i) - \frac{1}{2}\} \boldsymbol{X}_i \approx \boldsymbol{0} \tag{2}$$

的根. 这里使用"≈"是因为 $\boldsymbol{U}_n(\beta)$是 β的不连续函数. 一般而言,当误差 e_i的分布连续时,(2)式的根可以使 $\boldsymbol{U}_n$为 $O(1)$. 由于给定 $\boldsymbol{X}_i$时 $\boldsymbol{U}_n(\beta_0)$的条件期望为 0,所以 $\boldsymbol{U}_n(\beta)$是 β的一个合理的估计函数. 其次对于无删失非线性回归模型(1),参数 β的 L_1估计是极小化问题

$$\min_{\beta} \sum_i | T_i - f(\beta, \boldsymbol{X}_i) |$$

的解. 它等价于

$$\min_{\beta} \sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\}(T_i - f(\beta, \boldsymbol{X}_i)).$$

我们将上式改写为

$$\min_{\beta} \sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\}(T_i - f(\beta_0, \boldsymbol{X}_i) - \nabla f(\beta_0, \boldsymbol{X}_i)'(\beta - \beta_0) - d_i(\beta, \beta_0, \boldsymbol{X}_i)),$$

其中

$$d_i(\beta, \beta_0, \boldsymbol{X}_i) = f(\beta, \boldsymbol{X}_i) - f(\beta_0, \boldsymbol{X}_i) - \nabla f(\beta_0, \boldsymbol{X}_i)'(\beta - \beta_0).$$

当 β和 β_0充分接近且 n充分大时,在一定条件下,$\sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\} d_i(\beta, \beta_0, \boldsymbol{X}_i)$对上述问题解的影响与$\sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\}(T_i - f(\beta_0, \boldsymbol{X}_i) - \nabla f(\beta_0, \boldsymbol{X}_i)'(\beta - \beta_0))$相比就会充分小. 因此参数 β的估计可以重新定义为极小化问题

$$\min_{\beta} \sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\}(T_i - f(\beta_0, \boldsymbol{X}_i) - \nabla f(\beta_0, \boldsymbol{X}_i)'(\beta - \beta_0))$$

的解. 注意到上述问题的解在一定条件下是方程

$$\tilde{U}_n(\beta) = \sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\} \nabla f(\beta_0, \boldsymbol{X}_i) \approx \boldsymbol{0}$$

的根. 由于上式中 β_0未知,我们重新定义参数 β的估计为方程

$$\tilde{U}_n(\beta) = \sum_i \{I(T_i \geqslant f(\beta, \boldsymbol{X}_i)) - \frac{1}{2}\} \nabla f(\beta, \boldsymbol{X}_i) \approx \boldsymbol{0} \tag{3}$$

的根. 对于删失非线性回归模型(1),观测值为 Y_i而非 T_i. 在下一节条件 A1下,因为

$$E\left[I(Y_i \geqslant f(\beta_0, \boldsymbol{X}_i)) \mid \boldsymbol{X}_i \right] = \frac{1}{2} G(f(\beta_0, \boldsymbol{X}_i)),$$

所以参数 β的一个自然的估计方程为

$$\boldsymbol{S}_n(\beta) = \sum_i \left[\frac{I(Y_i \geqslant f(\beta, \boldsymbol{X}_i))}{\hat{G}(f(\beta, \boldsymbol{X}_i))} - \frac{1}{2} \right] \nabla f(\beta, \boldsymbol{X}_i) \approx \boldsymbol{0}, \tag{4}$$

其中 $\hat{G}$是 G的乘极限估计. 方程(4)的根 $\hat{\beta}_n$定义为 $\| \boldsymbol{S}_n(\beta) \|$的极小值点,$\| \cdot \|$为 $\mathbf{R}^p$中欧氏范数.

2　假设条件和主要结论

首先假设参数 β的真值 β_0在有界凸区域 D的内部,其次给出下列条件:

A1. $e_1, \cdots, e_n$独立同分布. 给定 $\boldsymbol{X}_i$时 e_i的条件生存函数记为 H,它连续、未知但中位数为 0. $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$

是独立同分布的协变量.

A2 $C_1, \cdots, C_n$ 为独立同分布且具有连续生存函数 G 的非负随机删失变量,并与 $T_1, \cdots, T_n$ 独立. 函数 g 和 h,分别为 $1-G$ 和 $1-H$ 的密度函数,均一致有界.

A3 $f(\beta, \boldsymbol{X}_i)$ 在 $\beta=\beta_0$ 处对于 β 连续可微,且 $f(\beta_0, \boldsymbol{X}_i)\geqslant 0$, $i=1,2,\cdots,n$. 存在 β_0 的邻域 B_0 使得对一切 $\beta\in B_0$, $|f(\beta, \boldsymbol{X}_i)-f(\beta_0, \boldsymbol{X}_i)-\nabla f(\beta_0, \boldsymbol{X}_i)'(\beta-\beta_0)|\leqslant r_i(\beta_0)\|\beta-\beta_0\|$. 且

$$\overline{\lim_{n\to\infty}}\sum_i r_i^2(\beta_0)<\infty.$$

此外,对一切 $\beta\in D$ 和 $\boldsymbol{X}$,梯度向量 $\nabla f(\beta, \boldsymbol{X})$ 有界.

A4 对 $\beta\in D$ 和 $\boldsymbol{X}$,设存在常量 t_0 使得 $P(Y\geqslant t_0)\geqslant\delta>0$ 和 $f(\beta, \boldsymbol{X})\leqslant t_0$ 几乎处处成立, δ 是常量.

A5 假定矩阵 $E[\nabla f(\beta_0, \boldsymbol{X})\nabla f(\beta_0, \boldsymbol{X})'h(0|\boldsymbol{X})]$ 正定,这里 $h(\cdot|\boldsymbol{X})$ 是给定 $\boldsymbol{X}$ 时 H 的条件密度函数.

注 条件 A2, A3 中分别要求删失变量 C_i 以及 $f(\beta_0, \boldsymbol{X}_i)$ 非负, $i=1,2,\cdots,n$,仅仅是为了叙述上的方便,这两个假设可以很容易的拓宽为 $C_i\geqslant -k$, $f(\beta_0, \boldsymbol{X}_i)\geqslant -k$, $i=1,2,\cdots,n$,其中 k 为任意的有限正常数.

主要结论为如下 3 个定理:

定理 1 在条件 A1 ~ A5 下, $\hat\beta_n\xrightarrow{a.s}\beta_0$,其中"$\xrightarrow{a.s}$"意为几乎处处收敛.

定理 2 在条件 A1 ~ A5 下,对任意的 $\varepsilon>0$,

$$\boldsymbol{S}_n(\beta)=\boldsymbol{S}_n(\beta_0)+n\boldsymbol{A}(\beta-\beta_0)+o_p(\max(n^{\frac{1}{2}+\varepsilon}, n\|\beta-\beta_0\|))$$

成立,其中

$$\boldsymbol{A}=E[\nabla f(\beta_0, \boldsymbol{X})\nabla f(\beta_0, \boldsymbol{X})'h(0|\boldsymbol{X})].$$

定理 3 令

$$p(t)=\lim_{n\to\infty}\frac{1}{n}\sum_j I(Y_j\geqslant t),\quad \boldsymbol{q}(t)=\lim_{n\to\infty}\frac{1}{n}\sum_j I(f(\beta_0, \boldsymbol{X}_j)\geqslant t)\nabla f(\beta_0, \boldsymbol{X}_j).$$

在条件 A1 ~ A5 下,

$$\hat\beta_n-\beta_0\xrightarrow{d}N(0, n^{-1}(\boldsymbol{A}^{-1})\Gamma(\boldsymbol{A}^{-1})'),$$

其中

$$\Gamma=\lim_{n\to\infty}\frac{1}{n}\sum_i\left[\frac{I(Y_i\geqslant f(\beta_0, \boldsymbol{X}_i))}{G(f(\beta_0, \boldsymbol{X}_i))}-\frac{1}{2}\right]^2\nabla f(\beta_0, \boldsymbol{X}_i)\nabla f(\beta_0, \boldsymbol{X}_i)'+\frac{1}{4}\int\frac{\boldsymbol{q}(t)\boldsymbol{q}(t)'}{p(t)}\mathrm{d}\Lambda_G(t),$$

"$\xrightarrow{d}$"意为依分布收敛, Λ_G 是删失变量 C 的累积失效率函数.

3 定理的证明

首先,我们证明 $\hat\beta_n$ 的强相合性. 令

$$\tilde{\boldsymbol{S}}_n(\beta)=\sum_i\left\{H(f(\beta, \boldsymbol{X}_i)-f(\beta_0, \boldsymbol{X}_i))-\frac{1}{2}\right\}\nabla f(\beta, \boldsymbol{X}_i).$$

因为(见[11])对一切 $\varepsilon>0$,

$$\sup_{t\leqslant t_0}|\hat G(t)-G(t)|=o(n^{-\frac{1}{2}+\varepsilon})\quad a.s,\tag{5}$$

所以对一切 $\beta\in D$ 以概率 1 有下式成立:

$$\boldsymbol{S}_n(\beta)-\tilde{\boldsymbol{S}}_n(\beta)=\sum_i G(f(\beta, \boldsymbol{X}_i))^{-1}\left\{I(Y_i\geqslant f(\beta, \boldsymbol{X}_i))-P(Y_i\geqslant f(\beta, \boldsymbol{X}_i)|\boldsymbol{X}_i)\right\}\cdot\nabla f(\beta, \boldsymbol{X}_i)+o(n^{\frac{1}{2}+\varepsilon}).$$

依据重对数定律及条件 A3,有

$$\sup_{\beta\in D}\left\|\sum_i G(f(\beta, \boldsymbol{X}_i))^{-1}\left\{I(Y_i\geqslant f(\beta, \boldsymbol{X}_i))-P(Y_i\geqslant f(\beta, \boldsymbol{X}_i)|\boldsymbol{X}_i)\right\}\nabla f(\beta, \boldsymbol{X}_i)\right\|=o(n^{\frac{1}{2}+\varepsilon})\quad a.s.$$

因此

$$\sup_{\beta\in D}\|n^{-1}\boldsymbol{S}_n(\beta)-n^{-1}\tilde{\boldsymbol{S}}_n(\beta)\|=o(n^{\frac{1}{2}+\varepsilon})\quad a.s.\tag{6}$$

令

$$\boldsymbol{A}_n(\beta)=\frac{1}{n}\frac{\partial}{\partial\beta}\widetilde{\boldsymbol{S}}_n(\beta)=\frac{1}{n}\sum_i h(f(\beta,\boldsymbol{X}_i)-f(\beta_0,\boldsymbol{X}_i))\nabla f(\beta,\boldsymbol{X}_i)\nabla f(\beta,\boldsymbol{X}_i)'$$
$$+\frac{1}{n}\sum_i\left\{H(f(\beta,\boldsymbol{X}_i)-f(\beta_0,\boldsymbol{X}_i))-\frac{1}{2}\right\}\frac{\partial}{\partial\beta}\nabla f(\beta,\boldsymbol{X}_i).$$

则 $\boldsymbol{A}_n(\beta_0)\to\boldsymbol{A}$ 以概率 1成立, 其中 $\boldsymbol{A}$ 是正定矩阵. 由于 $\widetilde{\boldsymbol{S}}_n(\beta_0)=0$, 故对充分大的 n, 如果 $\beta\neq\beta_0$ 且 $\beta\in B_0$, 则存在常量 $\triangle_n>0$, 使得

$$\|n^{-1}\widetilde{\boldsymbol{S}}_n(\beta)\|\geqslant\triangle_n>0 \tag{7}$$

由 $\hat\beta_n$ 的定义及等式(6)可得

$$\|n^{-1}\boldsymbol{S}_n(\hat\beta_n)\|\leqslant\|n^{-1}\boldsymbol{S}_n(\beta_0)\|=o(n^{-\frac{1}{2}+\varepsilon})\quad a.s.$$

所以

$$\|n^{-1}\widetilde{\boldsymbol{S}}_n(\hat\beta_n)\|\leqslant\|n^{-1}\widetilde{\boldsymbol{S}}_n(\hat\beta_n)-n^{-1}\boldsymbol{S}_n(\hat\beta_n)\|+\|n^{-1}\boldsymbol{S}_n(\hat\beta_n)\|=o(n^{-\frac{1}{2}+\varepsilon})\quad a.s. \tag{8}$$

依据(7)和(8), 知 $\hat\beta_n\xrightarrow{a.s.}\beta$, 定理 1证毕.

其次, 为证明定理 2, 需要下面的引理:

引理 1　设 μ 为连续可微函数, 则对任意 $\varepsilon>0$,

$$\sup_{t\leqslant t_0}|\mu(\hat G(t))-\mu(G(t))-\mu(\hat G(s))+\mu(G(s))|=o_p(n^{-\frac{1}{2}+\varepsilon}).$$

证明　由等式(5)知引理的结论显然.

定理 **2**的证明　令

$$\triangle_1=\boldsymbol{S}_n(\beta)-\sum_i\left[G(f(\beta,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))-\frac{1}{2}\right]\nabla f(\beta,\boldsymbol{X}_i)$$
$$-\sum_i\left[\hat G(f(\beta_0,\boldsymbol{X}_i))^{-1}-G(f(\beta_0,\boldsymbol{X}_i))^{-1}\right]I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))\nabla f(\beta,\boldsymbol{X}_i),$$

由引理 1及条件 A3知

$$\|\triangle_1\|=o_p(n^{\frac{1}{2}+\varepsilon})$$

成立. 因此

$$\begin{aligned}\boldsymbol{S}_n(\beta)=&\sum_i\left[G(f(\beta,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))-\frac{1}{2}\right]\nabla f(\beta,\boldsymbol{X}_i)\\&+\sum_i\left[\hat G(f(\beta_0,\boldsymbol{X}_i))^{-1}-G(f(\beta_0,\boldsymbol{X}_i))^{-1}\right]I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))\nabla f(\beta,\boldsymbol{X}_i)+o_p(n^{\frac{1}{2}+\varepsilon})\\&=\boldsymbol{S}_n(\beta_0)+\sum_i\left[G(f(\beta,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))-\frac{1}{2}\right]\nabla f(\beta,\boldsymbol{X}_i)\\&-\sum_i\left[G(f(\beta_0,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta_0,\boldsymbol{X}_i))-\frac{1}{2}\right]\nabla f(\beta_0,\boldsymbol{X}_i)\\&+\sum_i G(f(\beta_0,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta_0,\boldsymbol{X}_i))\nabla f(\beta_0,\boldsymbol{X}_i)\\&-\sum_i G(f(\beta_0,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))\nabla f(\beta,\boldsymbol{X}_i)\\&+\sum_i \hat G(f(\beta_0,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta,\boldsymbol{X}_i))\nabla f(\beta,\boldsymbol{X}_i)\\&-\sum_i \hat G(f(\beta_0,\boldsymbol{X}_i))^{-1}I(Y_i\geqslant f(\beta_0,\boldsymbol{X}_i))\nabla f(\beta_0,\boldsymbol{X}_i)+o_p(n^{\frac{1}{2}+\varepsilon})\\&=:\boldsymbol{S}_n(\beta_0)+\boldsymbol{I}_1-\boldsymbol{I}_2+\boldsymbol{I}_3-\boldsymbol{I}_4+\boldsymbol{I}_5-\boldsymbol{I}_6+o_p(n^{\frac{1}{2}+\varepsilon}).\end{aligned}$$

依据条件 A3及等式(5)可得

$$\|\boldsymbol{I}_3-\boldsymbol{I}_6\|=o_p(n^{\frac{1}{2}+\varepsilon}),\quad\|\boldsymbol{I}_5-\boldsymbol{I}_4\|=o_p(n^{\frac{1}{2}+\varepsilon}). \tag{9}$$

令

$$\triangle_2=\boldsymbol{I}_1-\boldsymbol{I}_2-\sum_i\left[G(f(\beta,\boldsymbol{X}_i))^{-1}P(Y_i\geqslant f(\beta,\boldsymbol{X}_i)\mid\boldsymbol{X}_i)-\frac{1}{2}\right]\nabla f(\beta,\boldsymbol{X}_i),$$

依据重对数定律及条件 A3知

$$\| \Delta_2 \| = o_p(n^{\frac{1}{2}+\varepsilon}).$$

故

$$\begin{aligned} \boldsymbol{I}_1 - \boldsymbol{I}_2 &= \sum_i \left[G(f(\beta, \boldsymbol{X}_i))^{-1} P(Y_i \geqslant f(\beta, \boldsymbol{X}_i) \mid \boldsymbol{X}_i) - \frac{1}{2} \right] \nabla f(\beta, \boldsymbol{X}_i) + o_p(n^{\frac{1}{2}+\varepsilon}) \\ &= \widetilde{\boldsymbol{S}}_n(\beta) + o_p(n^{\frac{1}{2}+\varepsilon}). \end{aligned} \tag{10}$$

由(9) ~ (10)知

$$\boldsymbol{S}_n(\beta) = \boldsymbol{S}_n(\beta_0) + \widetilde{\boldsymbol{S}}_n(\beta) + o_p(n^{\frac{1}{2}+\varepsilon}). \tag{11}$$

从而, 在(11)中对 $\widetilde{\boldsymbol{S}}_n(\beta)$ 在 β_0 处用泰勒展式即得 $\boldsymbol{S}_n(\beta)$ 的局部线性性质. 定理 2证毕.

最后为证明定理 3, 即 $\hat{\beta}_n - \beta_0$ 的渐近正态性, 我们首先证明 $n^{-\frac{1}{2}}\boldsymbol{S}_n(\beta_0)$ 渐近于正态分布 $N(0, \Gamma)$, 然后由定理 2知, 定理 3的结论显然. 为证明 $n^{-\frac{1}{2}}\boldsymbol{S}_n(\beta_0)$ 的渐近正态性, 我们需要下面的两个引理.

引理 **2** 定义 M 为

$$M(t) = N(t) - \int_0^t Y(s)\,\mathrm{d}\Lambda_G(s).$$

则 M 是 $[0, \infty)$ 上的平方可积鞅, 其可料二次变差过程《M, M》由

$$《M, M》(t) = \int_0^t Y(s)\,\mathrm{d}\Lambda_G(s)$$

给定, 其中

$$N(t) = \sum_i I(Y_i \leqslant t, \delta_i = 0), \quad Y(t) = \sum_i I(Y_i \geqslant t).$$

证明 见文献[12].

引理 **3** 对一切 $t > 0$, 有下式成立

$$\frac{\hat{G}(t)}{G(t)} = 1 - \int_0^t \frac{\hat{G}(s)}{G(s)Y(s)}\,\mathrm{d}M(s).$$

证明 见文献[12].

定理 **3**的证明 方法是用一列独立的零均值随机变量和逼近 $n^{-\frac{1}{2}}\boldsymbol{S}_n(\beta_0)$.

令

$$\Delta_3 = \boldsymbol{S}_n(\beta_0) - \sum_i \left[\frac{I(Y_i \geqslant f(\beta_0, \boldsymbol{X}_i))}{G(f(\beta_0, \boldsymbol{X}_i))} - \frac{1}{2} \right] \nabla f(\beta_0, \boldsymbol{X}_i).$$

则

$$\Delta_3 = -n \int_0^\infty \frac{\hat{G}(t) - G(t)}{\hat{G}(t)G(t)}\,\mathrm{d}Q(t),$$

其中 $Q(t) = n^{-1}\sum_i I(f(\beta_0, \boldsymbol{X}_i) \leqslant \min(t, Y_i)) \nabla f(\beta_0, \boldsymbol{X}_i)$.

由引理 3得

$$\begin{aligned} \Delta_3 &= n \int_0^\infty \int_0^t \frac{\hat{G}(s)}{\hat{G}(t)G(s)Y(s)}\,\mathrm{d}M(s)\,\mathrm{d}Q(t) \\ &= \frac{1}{2}\sum_i \int_0^\infty \frac{\boldsymbol{q}(s)}{p(s)} \left\{ \mathrm{d}I(Y_i \leqslant s, \delta_i = 0) - I(Y_i \geqslant s)\,\mathrm{d}\Lambda_G(s) \right\} + o_p(n^{\frac{1}{2}}), \end{aligned}$$

其中

$$p(s) = \lim_{n\to\infty} \frac{1}{n}\sum_i I(Y_j \geqslant s), \quad \boldsymbol{q}(s) = \lim_{n\to\infty} \frac{1}{n}\sum_i I(f(\beta_0, \boldsymbol{X}_j) \geqslant s) \nabla f(\beta_0, \boldsymbol{X}_j).$$

因此, $n^{-\frac{1}{2}}\boldsymbol{S}_n(\beta_0)$ 渐近等价于 $n^{-\frac{1}{2}}\sum_i \tau_i =: \boldsymbol{I}_7 + \boldsymbol{I}_8$, 其中

$$\begin{aligned} \tau_i &= \left[\frac{I(Y_i \geqslant f(\beta_0, \boldsymbol{X}_i))}{G(f(\beta_0, \boldsymbol{X}_i))} - \frac{1}{2} \right] \nabla f(\beta_0, \boldsymbol{X}_i) \\ &\quad + \frac{1}{2}\int_0^\infty \frac{\boldsymbol{q}(t)}{p(t)} \left\{ \mathrm{d}I(Y_i \leqslant t, \delta_i = 0) - I(Y_i \geqslant t)\,\mathrm{d}\Lambda_G(t) \right\}. \end{aligned}$$

在给定 $\boldsymbol{X}_i$ 的条件下, $E\boldsymbol{I}_7 = \boldsymbol{0}$,

$$\mathrm{Var}\boldsymbol{I}_7 = \mathrm{Var}\left\{\left[\frac{I(Y_i \geqslant f(\beta_0, \boldsymbol{X}_i))}{(f(\beta_0, \boldsymbol{X}_i))} - \frac{1}{2}\right] \cdot \nabla f(\beta_0, \boldsymbol{X}_i)\right\}$$

$$= \lim_{n\to\infty} \frac{1}{n} \cdot \sum_i \left[\frac{I(Y_i \geqslant f(\beta_0, \boldsymbol{X}_i))}{G(f(\beta_0, \boldsymbol{X}_i))} - \frac{1}{2}\right]^2 \cdot \nabla f(\beta_0, \boldsymbol{X}_i) \cdot \nabla f(\beta_0, \boldsymbol{X}_i)'.$$

由引理 2及定理 4[13] 知

$$EI_8 = \frac{1}{2} \cdot n^{-\frac{1}{2}} \cdot E\left\{\int_0^\infty \frac{\boldsymbol{q}(t)}{p(t)} \mathrm{d}N(t) - \int_0^\infty \frac{\boldsymbol{q}(t)}{p(t)} \mathrm{d}\int_0^t Y(s)\,\mathrm{d}\Lambda_G(s)\right\} = \boldsymbol{0},$$

$$\mathrm{Cov}(\boldsymbol{I}_7, \boldsymbol{I}_8) = \frac{1}{n} \cdot E(\boldsymbol{I}_7 \cdot \boldsymbol{I}'_8) = \frac{1}{n} \cdot E\left(\frac{1}{2}\int_0^\infty \frac{\boldsymbol{I}_7 \boldsymbol{q}(t)'}{p(t)} \mathrm{d}M(t)\right) = \boldsymbol{0},$$

$$\mathrm{Var}\boldsymbol{I}_8 = \frac{1}{4n} \cdot \mathrm{Var}\left\{\int_0^\infty \frac{\boldsymbol{q}(t)}{p(t)} \mathrm{d}M(t)\right\} = \frac{1}{4n} \cdot E\left\{\int_0^\infty \frac{\boldsymbol{q}(t)}{p(t)} \mathrm{d}M(t) \cdot \int_0^\infty \frac{\boldsymbol{q}(t)'}{p(t)} \mathrm{d}M(t)\right\}$$

$$= \frac{1}{4n} \cdot E\left\langle \int_0^\infty \frac{\boldsymbol{q}(t)}{p(t)} \mathrm{d}M(t)\right\rangle = \frac{1}{4} \cdot \int_0^\infty \frac{\boldsymbol{q}(t)\boldsymbol{q}(t)'}{p(t)} \mathrm{d}\Lambda_G(t).$$

从而, 依据多元中心极限定理, $n^{-\frac{1}{2}}\boldsymbol{S}_n(\beta_0)$ 渐近于正态分布 $N(0, \Gamma)$, 其中

$$\Gamma = \lim_{n\to\infty} \frac{1}{n} \cdot \sum_i \left[\frac{I(Y_i \geqslant f(\beta_0, \boldsymbol{X}_i))}{G(f(\beta_0, \boldsymbol{X}_i))} - \frac{1}{2}\right]^2 \cdot \nabla f(\beta_0, \boldsymbol{X}_i) \cdot \nabla f(\beta_0, \boldsymbol{X}_i)'$$

$$+ \frac{1}{4} \cdot \int_0^\infty \frac{\boldsymbol{q}(t)\boldsymbol{q}(t)'}{p(t)} \mathrm{d}\Lambda_G(t).$$

定理 3证毕.

[参考文献]

[1] Buckley J, James I. Linear regression with censored data[J]. Biometrika, 1979, 66: 429-436.

[2] Chatterjee S, McLeish D L. Fitting linear regression models to censored data by least squares and maximum likelihood methods[J]. Comm Statist-Theor Meth, 1986, 15: 3227-3243.

[3] Koul H, Susarla V, Van Ryzin J. Regression analysis with randomly right-censored data[J]. Ann Stat, 1981, 9: 1276-1288.

[4] Leurgans S. Linear models, random censoring and synthetic data[J]. Biometrika, 1987, 74: 301-309.

[5] Miller R G. Least squares regression with censored data[J]. Biometrika, 1976, 63: 449-464.

[6] Powell J L. Least absolute deviations estimation for the censored regression model[J]. J Econometrics, 1984, 25: 303-325.

[7] Powell J L. Censored regression quantiles[J]. J Econometrics, 1986, 32: 143-155.

[8] Rao C R, Zhao L C. Asymptotic normality of LAD estimator in censored regression models[J]. Math Methods of Statist, 1993, 2(3): 228-239.

[9] Ying Z, Jung S H, Wei L J. Survival analysis with median regression models[J]. Journal of the Ammerican Statistical Association, 1995, 90(429): 178-184.

[10] Zhou Xiuqing, Wang Jinde. LAD estimation for nonlinear regression models with randomly censored data[J]. Science in China Ser A, 2005, 35(4): 387-403.

[11] Csörgö S, Horvath L. The rate of strong uniform consistency for the product-limit estimator[J]. ZW, 1983, 62: 411-426.

[12] Gill R D. Large sample behavior of the product-limit estimator on the whole line[J]. The Annals of Statisitcs, 1983, 11: 49-58.

[13] Tze Leunglai, Zheng Zukang. Survival Analysis[M]. Hangzhou: Zhejiang Publishing House of Science and Technology, 1993: 152-160.

[责任编辑: 陆炳新]