第11卷 第1期
1991年2月

数学研究与评论
JOURNAL OF MATHEMATICAL
RESEARCH AND EXPOSITION

Vol.11 No.1
Feb. 1991

随机设计的截断数据回归分析*

郑 祖 康

（复旦大学统计运筹系，上海）

摘 要

本文讨论了随机设计的截断数据回归分析．参数 β 的估计 $\hat{\beta}_n$ 来自 Class K 的某些子集，证明了 $\sqrt{n}(\hat{\beta}_n-\beta)$ 收敛到正态分布．

线性模型 $y_i=\alpha+\beta x_i+\varepsilon_i$（$i=1,2,\cdots,n$），其中 ε_i 为零均值独立随机变量．当 y_i 被另一列随机变量 t_i 截断时，我们能观察到的是：

$$z_i=\min(y_i,t_i), \qquad \delta_i=I_{(y_i\le t_i)}=\begin{cases}1 & \text{当 } y_i\le t_i\\ 0 & \text{当 } y_i>t_i\end{cases}.$$

如何利用这些观察量来估计 α 和 β 呢？这就是所谓截断数据的回归分析问题．1976年由Miller[1]提出，1981年Koul，Susarla和Van Ryzin[2]找到了收敛解．他们用 $\delta_i z_i/[1-G(z_i)]$ 来代替 y_i，再用最小二乘法求出 $\hat{\alpha}_n$，$\hat{\beta}_n$，其中 t_i 是独立同分布的随机变量，具有分布函数 G．当 G 为未知时，他们建议用 G 的估计 $\hat{G}_n$ 来代替．1984年，T.L.Lai 和本人提出了更广泛的 Class K 估计．本文讨论 x_i 为随机设计，G 为未知的情况．

我们总假定 ε_i 独立同分布 $E\varepsilon_i=0$，$\mathrm{Var}\varepsilon_i<\infty$．$t_i$ 独立同分布具有连续分布函数 G．x_i 独立同分布，且 $\{x_i\}$、$\{\varepsilon_i\}$、$\{t_i\}$ 三叙列都独立．此时 y_i 也为独立同分布，我们假定它也有连续分布函数 F．

记 $H(t)$ 为 z_i 的分布函数，$1-H(t)=(1-F(t))(1-G(t))$．$\tau_F=\inf\{t:F(t)=1\}$，$\tau_G=\inf\{t:G(t)=1\}$，$\tau_H=\inf\{t:H(t)=1\}$，$T=T(n)=\max\limits_{1\le i\le n}z_i$．对随机过程 $Q(t)$，记 $Q^T(t)=Q(T\wedge t)$．本文还假定 $\tau_F<\tau_G$（i.e. $G(\tau_F)<1$），且 x_i 是有界随机变量，从而 $E|x_i|<\infty$．$E|x_i|^2<\infty$．

先回顾一下Class K 的估计方法[8]：令

$$y_i^*=\varphi_1(z_i)\delta_i+\varphi_2(z_i)(1-\delta_i) \tag{1}$$

代替 y_i，然后用最小二乘法求出 $\hat{\alpha}_n$，$\hat{\beta}_n$，其中 φ_1，φ_2 在 $(-\infty,\tau_G)$ 上连续且满足

（i） $[1-G(y_i)]\varphi_1(y_i)+\int_{-\infty}^{y_i}\varphi_2(t)\,dG(t)=y_i$；

（ii） φ_1，φ_2 与 y_i 的分布函数无关．

* 1989年3月9日收到．国家自然科学基金资助．

我们称具有上述性质的函数对(φ_1,φ_2)属于Class$K$，它满足

$$E(\varphi_1(z_i)\delta_i+\varphi_2(z_i)(1-\delta_i))=Ey_i\ .$$

注意到φ_1,φ_2可能是G的函数，当G未知时，用$\hat{G}_n$代替G，这里$\hat{G}_n$是Kaplan-Maier[3]估计．此时把t_i看作为y_i所截断，$\delta'_i=1-\delta_i$．为了表示φ_1,φ_2对G或$\hat{G}_n$的依赖关系，我们写成$\varphi_1(t,G(t))$，$\varphi_2(t,G(t))$或$\varphi_1(t,\hat{G}_n(t))$，$\varphi_2(t,\hat{G}_n(t))$．下面只对$\beta$的估计$\hat{\beta}_n$进行讨论，$\alpha$的估计$\hat{\alpha}_n$可完全类似进行．定义：

$$\hat{\beta}_n=\frac{1}{\sum_{j=1}^{n}(x_j-\bar{x}_n)^2}\sum_{i=1}^{n}(x_i-\bar{x}_n)[\varphi_1(z_i,\hat{G}_n(z_i))\delta_i+\varphi_2(z_i,\hat{G}_n(z_i))(1-\delta_i)] \qquad (2)$$

其中 $\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$ ．

为了使β_n有较好的性质，我们在Class$K$中选出一些理想的函数对，现给出如下定义：$K_R(s)$是所有$(\varphi_1,\varphi_2)\in K$中满足下列条件的函数对：当$s$满足$G(s)<1$时

（i）存在常数C^*使得 $\max\limits_{\substack{j=1,2\\ t\leq s}}|\varphi_j(t,G)|\leq C^*$成立．

（ii）记$\mathscr{B}_s$为$(-\infty,s]$上所有有界Borel函数所构成的Banach空间(sup-norm)存在连续函数φ_j^*：$(-\infty,s]\times\mathscr{B}_s\to R\ (j=1,2)$使得

（A）对固定的$t\leq s$，$\varphi_j^*(t,\cdot)$是$\mathscr{B}_s$上的有界线性泛函，且

$$\max_{-\infty<t\leq s}\|\varphi_j^*(t,\cdot)\|<\infty .$$

（B）存在$L=L(s)$以及$\eta>0$使得

$$\max_{\substack{j=1,2\\ -\infty<t\leq s}}|\varphi_j(t,\tilde{G})-\varphi_j(t,G)-\varphi_j^*(t,\tilde{G}-G)|\leq L(\max_{t\leq s}|\tilde{G}(t)-G(t)|)^2$$

对一切满足$\max\limits_{t\leq s}|\tilde{G}(t)-G(t)|<\eta$的分布函数$\tilde{G}$成立．

由于本文假定$G(\tau_F)<1$，我们总可以找到$L=L(\tau_F)$使(B)成立．例如

$$\frac{\delta_i z_i}{1-G(z_i)}\in K_R(\tau_F),\quad \int_{-\infty}^{z_i}\frac{ds}{1-G(s)}\in K_R(\tau_F)\ \text{等}.$$

引理1[5] 在足够大的概率空间$(\Omega,\mathscr{F},P)$中，可定义高斯过程$B_n^0(t)$．$B_n^1(t)\ (n=1,2,\cdots)$，使得

$$P(\sup_{t\leq T_n}|\sqrt{n}(\hat{G}_n(t)-G(t))-W_n(t)|>r(n))\leq Qn^{-(1+\delta)}, \qquad (3)$$

其中

$$W_n(t)=(1-G(t))[\int_{-\infty}^{t}B_n^0(s)(1-H(s))^{-2}dG^1(s)+B_n^1(t)(1-H(t))^{-1}-$$

$$-\int_{-\infty}^{t}B_n^1(s)(1-H(s))^{-2}dH(s)],\quad G^1(s)=\int_{-\infty}^{s}(1-F(u))dG(u);$$ Q,δ都是正常数；

$r(n)=O(\max(n^{-\frac{1}{3}}b_n^2(\log n)^{\frac{3}{2}},\ n^{-\frac{1}{2}}b_n^4(\log n),\ n^{-\frac{3}{2}}b_n^6(\log n)^2))$，$b_n=(1-H(T_n))^{-1}$，$T_n<\tau_H$满足 $1-H(T_n)\geq(2(1+\delta)\frac{\log n}{n})^{\frac{1}{2}}$．

高斯过程 $B_n^0(t)$，$B_n^1(t)$ 满足（对任意t,s）

$$EB_n^0(t)=EB_n^1(t)=0,$$

$$EB_n^0(t)B_n^0(s)=\min(H(t),H(s))-H(t)H(s),$$
$$EB_n^1(t)B_n^1(s)=\min(G^1(t),G^1(s))-G^1(t)G^1(s),$$
$$EB_n^1(t)B_n^0(s)=\min(G^1(t),G^1(s))-G^1(t)H(s).$$

我们总假定工作在这个足够大的空间上．取 $b_n^{-1}=n^{-\frac{1}{8}+c}$ $(c=\frac{1}{60})$ 当 n 充分大时 $b_n=n^{\frac{1}{8}-c}\le(\frac{n}{2(1+\delta)\log n})^{\frac{1}{2}}$ 满足引理要求，且 $r(n)=O(\max(n^{-\frac{1}{3}}n^{\frac{1}{4}-2c}(\log n)^{\frac{3}{2}},\ n^{-\frac{1}{2}}n^{\frac{1}{2}-4c}(\log n),\ n^{-\frac{3}{2}}n^{\frac{3}{4}-6c}(\log n)^2))=O(n^{-\frac{1}{15}}\log n)$．注意到 b_n 取得太大时会有 $r(n)\to\infty\ (n\to\infty)$．另一方面，$b_n$ 受 T_n 制约，当 b_n 变小时，$1-H(T_n)$ 将变大，T_n 将变小，这意味着收敛区间在缩小．为了求得 $\hat{\beta}_n$ 的渐近行为，我们作如下分解：

$$\begin{aligned}\sqrt{n}(\hat{\beta}_n-\beta)=&\frac{\sqrt{n}}{\sum(x_j-\bar{x}_n)^2}\sum_{i=1}^n(x_i-\bar{x}_n)\{[\varphi_1(z_i,G(z_i))\delta_i+\varphi_2(z_i,G(z_i))(1-\delta_i)]-\\&-E_{x_i}[\varphi_1(z_i,G(z_i))\delta_i+\varphi_2(z_i,G(z_i))(1-\delta_i)]\}\\&+\frac{\sqrt{n}}{\sum(x_j-\bar{x}_n)^2}\sum_{i=1}^n(x_i-\bar{x}_n)\{\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))\delta_i+\varphi_2^*(z_i,\hat{G}_n(z_i)-\\&-G(z_i))(1-\delta_i)\}I_{z_i\le T_n}\\&+\frac{\sqrt{n}}{\sum(x_j-\bar{x}_n)^2}\sum_{i=1}^n(x_i-\bar{x}_n)\{\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))\delta_i+\varphi_2^*(z_i,\hat{G}_n(z_i)\\&-G(z_i))(1-\delta_i)\}I_{z_i>T_n}\\&+\frac{\sqrt{n}}{\sum(x_j-\bar{x}_n)^2}\sum_{i=1}^n(x_i-\bar{x}_n)\{[\varphi_1(z_i,\hat{G}_n(z_i))-\varphi_1(z_i,G(z_i))\\&-\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))]\delta_i+[\varphi_2(z_i,\hat{G}_n(z_i))-\varphi_2(z_i,G(z_i))-\\&-\varphi_2^*(z_i,\hat{G}_n(z_i)-G(z_i))](1-\delta_i)\}\triangleq\xi_{1n}+\xi_{2n}+\xi_{3n}+\xi_{4n}.\end{aligned}\tag{4}$$

其中 E_{x_i} 表示对 x_i 的条件期望．

我们需要以下几个引理

引理 2 若 $A_1(n)$, $A_2(n)$ 满足条件 $1-H(A_1(n))=n^{-a}$, $1-H(A_2(n))=n^{-2a+\varepsilon}$，其中 $0<a<\frac{1}{2}$，$0<\varepsilon<a$．令

$$Q(t)=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_iI_{z_i\le t}$$，那么对任意 $\eta>0$

$$\lim_{n\to\infty}\sup P(\sup_{A_1(n)<t\le A_2(n)\wedge T}|Q(t)-Q(A_1(n))|>\eta)=0\tag{5}$$

证明 $Z(t)\triangleq\frac{\sqrt{n}(\hat{G}_n(t)-G(t))}{1-G(t)}$ 是 $(-\infty,T\wedge A_2(n))$ 上的鞅（cf[4]）．

$$Q(t)-Q(A_1(n))=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_iI_{A_1(n)<z_i\le t}$$

$$\begin{aligned}&P(\sup_{A_1(n)<t\le T\wedge A_2(n)}|Q(t)-Q(A_1(n))|>\eta)\\\le&P(\sup_{A_1(n)<t\le T\wedge A_2(n)}|\frac{C_0}{n}Z(t)^{\#}(A_1(n),A_2(n))|<\eta)\end{aligned}$$

其中 $^{\#}(a,b)$ 表示满足条件 $a<|z_i\le b$ 的 z_i 的个数，C_0 为有限常数．记 M^* 为一个大于 1 的常

数，那么上式不大于

$$P(\sup_{A_1(n)<t\leq T\wedge A_2(n)}\left|\frac{C_0}{n}Z(t)^{\#}(A_1(n),A_2(n))\right|>\eta^{\#}(A_1(n),A_2(n)))$$

$$\leq nM^*[H(A_2(n)-H(A_1(n))]) + P(^{\#}(A_1(n),A_2(n))>nM^*[H(A_2(n))-H(A_1(n))])$$

$$\leq P(\sup_{A_1(n)<t\leq T\wedge A_2(n)}\left|\int_{-\infty}^{t}(1-H(A_1(n)))\mathrm{d}Z(s)\right|>\frac{\eta}{M^*C_0})$$

$$+P(^{\#}(A_1(n),A_2(n))>nM^*[H(A_2(n))-H(A_1(n))]).$$

当 n 充分大时，由中心极限定理第二项趋于零，第一项利用Lenglart不等式[6]对任意 $\eta^*>0$ 不大于

$$\frac{\eta^*}{(\eta/M^*C_0)^2}+P[\int_{-\infty}^{T\wedge A_2(n)}(1-H(A_1(n)))^2(\frac{1-\hat{G}_n(s)}{1-G(s)})^2\frac{n}{Y(s)}\mathrm{d}\Lambda_G(s)>\eta^*].$$

其中 $\Lambda_G(s)=\int_{-\infty}^{s}\frac{\mathrm{d}G(r)}{1-G(r)}$，$Y(s)$ 是大于等于 s 的 z_i 的个数．又由[4]的引理6，引理7对任意 $\beta\in(0,1)$，有

$$P(1-\hat{G}_n(t)\leq\beta^{-1}(1-G(t)),\quad\forall t\leq T)\geq 1-\beta,$$

$$P(Y(t)/n\geq\beta(1-H(t)),\quad\forall t\leq T)\geq 1-e(\frac{1}{\beta})e^{-\frac{1}{\beta}}.$$

再令

$\eta^*=\int_{-\infty}^{A_2(n)}\beta^{-3}\frac{(1-H(A_1(n)))^2}{1-H(s)}\mathrm{d}\Lambda_G(s)$，当 n 充分大时，我们有

$$P(\sup_{A_1(n)<t\leq T\wedge A_2(n)}|Q(t)-Q(A_1(n))|>\eta)$$

$$\leq\frac{\eta^*}{(\eta/(M^*C_0))^2}+\beta+e(\frac{1}{\beta})e^{-\frac{1}{\beta}}+P(\int_{-\infty}^{A_2(n)}\frac{\beta^{-3}(1-H(A_1(n)))^2}{1-H(s)}\mathrm{d}\Lambda_G(s)>\eta^*)$$

$$=\beta^{-3}(\frac{\eta}{M^*C_0})^{-2}\int_{-\infty}^{A_2(n)}\frac{(1-H(A_1(n)))^2}{1-H(s)}\cdot\frac{\mathrm{d}G(s)}{1-G(s)}+\beta+e(\frac{1}{\beta})e^{-\frac{1}{\beta}}$$

$$\leq\beta^{-3}(\frac{\eta}{M^*C_0})^{-2}\int_{-\infty}^{A_2(n)}\frac{(1-H(A_1(n)))^2}{1-H(A_2(n))}[-\mathrm{d}\log(1-G(s))]+\beta+e(\frac{1}{\beta})e^{-\frac{1}{\beta}}$$

$$=\beta^{-3}(\frac{\eta}{M^*C_0})^{-2}n^{-\varepsilon}[-\log(1-G(A_2(n)))]+\beta+e(\frac{1}{\beta})e^{-\frac{1}{\beta}}$$

$$\leq\beta^{-3}(\frac{\eta}{M^*C_0})^{-2}n^{-\varepsilon}[-\log(1-H(A_2(n)))]+\beta+e(\frac{1}{\beta})e^{-\frac{1}{\beta}}$$

$$=\beta^{-3}(\frac{\eta}{M^*C_0})^{-2}n^{-\varepsilon}(2a-\varepsilon)\log n+\beta+e(\frac{1}{\beta})e^{-\frac{1}{\beta}}\to 0\quad(n\to\infty,\ \beta\to 0)$$ ■

引理 3 若 $A_1(n)$ 满足 $1-H(A_1(n))=n^{-\frac{1}{8}+c}$ $(c=\frac{1}{60})$，令

$$D_n=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_iI_{z_i>A_1(n)}\tag{6}$$

那么 $D_n\xrightarrow{p}0$.

证明 $D_n=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_iI_{A_1(n)<z_i\leq A_2(n)}$

$$+\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_iI_{A_2(n)<z_i\leq A_3(n)}$$

$$+\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_i I_{A_3(n)<z_i\le A_4(n)}$$

$$+\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_n)\varphi_1^*(z_i,\sqrt{n}(\hat{G}_n(z_i)-G(z_i)))\delta_i I_{A_4(n)<z_i}$$

$$\triangleq D_n^1+D_n^2+D_n^3+D_n^4,$$

其中 $1-H(A_2(n))=n^{-\frac{1}{4}+3c}$，$1-H(A_3(n))=n^{-\frac{1}{2}+7c}$，$1-H(A_4(n))=n^{-1+15c}$. 对任意 $\gamma>0$，有

$$P(|D_n|>\gamma)\le P(|D_n^1|>\frac{\gamma}{4})+P(|D_n^2|>\frac{\gamma}{4})+P(|D_n^3|>\frac{\gamma}{4})+P(|D_n^4|>\frac{\gamma}{4}).$$

由引理2，a 取 $\frac{1}{8}-\varepsilon$，c 取 $\frac{1}{60}$，$\varepsilon=c$ 得

$$P(|D_n^1|>\frac{\gamma}{4})=P(|Q(A_2(n))-Q(A_1(n))|>\frac{\gamma}{4})$$

$$\le P(\sup_{A_1(n)<t\le T\wedge A_2(n)}|Q(t)-Q(A_1(n))|>\frac{\gamma}{4})\to 0.$$

同理，$P(|D_n^2|>\frac{\gamma}{4})\to 0$ $(a=\frac{1}{4}-3c,\ \varepsilon=c)$，$P(|D_n^3|>\frac{\gamma}{4})\to 0$ $(a=\frac{1}{2}-7c,\ \varepsilon=c)$，而 $|D_n^4|\le\frac{c'}{\sqrt{n}}\Sigma I_{z_i>A_4(n)}\sim\frac{c'}{\sqrt{n}}n\,n^{-1+15c}\to 0$. 此处 c' 是有限常数. ■

显然本引理对 φ_2^* 的部分也成立. 以下记 z_i 的经验分布函数为 $H_n(t)=\frac{1}{n}\Sigma I_{z_i\le t}$，用 $N^+(t)$ 表示 $z_1,z_2,\cdots,z_n$ 中大于 t 的个数.

引理 4 [6]

$$\log(1-\hat{G}_n(t))-\log(1-G(t))=(R_{n1}(t)-ER_{n1}(t))+R_{n2}(t)+R_{n3}(t).$$

其中
$$R_{n1}(t)=-\frac{1}{n}\sum_{j=1}^{n}[I_{z_j\le t}\delta_j'(1-H(z_j))^{-1}],$$

$$ER_{n1}(t)=-EI_{z_j\le t}\delta_j'(1-H(z_j))^{-1}=\log(1-G(t)),$$

$$R_{n2}(t)=-\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\sum_{l=2}^{\infty}\frac{1}{l}(1+N^+(z_j))^{-l},$$

$$R_{n3}(t)=-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\{n(1-N^+(z_j))^{-1}-(1-H(z_j))^{-1}\}.$$

引理 5 当 n 充分大时，对一切 $t\le T_n$ 都有

$$1-H_n(t)\ge\widetilde{C}(1-H(t))$$

其中 $\widetilde{C}$ 是一适当正常数.

证明 T_n 满足 $1-H(T_n)=n^{-\frac{1}{8}+c}$ $(c=\frac{1}{60})$，仿[6]引理1的证法即得. ■

在下面一些引理证明中，我们用 C_i 表示不同的有限常数.

引理 6 (i) $\sup\limits_{t\le T_n}R_{n2}(t)=O(n^{-\frac{47}{60}})$.

(ii) $\sup\limits_{t\le T_n}\left|R_{n3}(t)+\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{H_n(z_j)-H(z_j)}{[1-H(z_j)]^2}\right|=O(n^{-\frac{81}{120}}(\log\log n)).$

证明 (i) 由[6]及引理5，$|R_{n2}(t)|\le C_1\frac{1}{n(1-H(t))^2}$ 立即得到

$$\sup_{t\le T_n}|R_{n2}(t)|\le C_1 n^{-1+\frac{1}{4}-2c}=O(n^{-\frac{47}{60}}).$$

(ii) $$R_{n3}(t)=-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'[n(N^+(z_j))^{-1}-(1-H(z_j))^{-1}]$$

$$+\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'[n(N^+(z_j))^{-1}-n(1+N^+(z_j))^{-1}]$$

$$=-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{H_n(z_j)-H(z_j)}{[1-H(z_j)]^2}-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{[H_n(z_j)-H(z_j)]^2}{[1-H_n(z_j)][1-H(z_j)]^2}$$

$$+\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'[n(N^+(z_j))^{-1}-n(1+N^+(z_j))^{-1}].$$

因为

$$\left|\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'[n(N^+(z_j))^{-1}-n(1+N^+(z_j))^{-1}]\right|$$

$$\le\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\left(\frac{n}{(N^+(z_j))^2}\right)\le\frac{1}{n}\frac{1}{[(N^+(t))/n]^2}\le C_2\frac{1}{n(1-H(t))^2}$$

以及

$$\left|-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{[H_n(z_j)-H(z_j)]^2}{[1-H_n(z_j)][1-H(z_j)]^2}\right|\le C_3\sup_{u\le t}\left[\frac{[H_n(u)-H(u)]^2}{[1-H(u)]^3}\right].$$

当 n 充分大时，得

$$\sup_{t\le T_n}\left|R_{n3}(t)+\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{H_n(z_j)-H(z_j)}{(1-H(z_j))^2}\right|$$

$$\le C_2\sup_{t\le T_n}\frac{1}{n(1-H(t))^2}+C_3\sup_{t\le T_n}\frac{\sup_{u\le t}[H_n(u)-H(u)]^2}{[1-H(t)]^3}$$

$$\le C_4 n^{-\frac{3}{4}-2c}+C_3 n^{\frac{3}{8}-3c}(\sup_{t\le T_n}|H_n(u)-H(u)|)^2$$

$$\le C_4 n^{-\frac{47}{60}}+C_5 n^{\frac{3}{8}-\frac{3}{60}-1}(\log\log n)$$

$$=o(n^{-\frac{81}{120}}\log\log n).$$

这里我们用到了经验分布函数的重对数律． ■

引理 7 $\xi_{40}=o_p(1)$.

证明 因为$\sum(x_i-\overline{x}_n)^2/n\to M$（常数）a.s．我们只须考虑

$$\left|\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_n)\sqrt{n}[\varphi_1(z_i,\hat{G}_n(z_i))-\varphi_1(z_i,G(z_i))-\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))]\delta_i\right|$$

$$\le\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{x}_n|\sqrt{n}|\varphi_1(z_i,\hat{G}_n(z_i))-\varphi_1(z_i,G(z_i))-\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))|I_{z_i\le T_n}$$

$$+\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{x}_n|\sqrt{n}|\varphi_1(z_i,\hat{G}_n(z_i))-\varphi_1(z_i,G(z_i))-\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))|I_{z_i>T_n}$$

$$\le\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{x}_n|\sqrt{n}L\cdot\sup_{t\le T_n}|\hat{G}_n(t)-G(t)|^2+o_p(1).\qquad(7)$$

其中用到了引理2,3的证明（$A_1(n)=T_n$）．另一方面，注意到

$$
\begin{aligned}
\sqrt{n}\sup_{t\le T_n}|\hat{G}_n(t)-G(t)| &= \sup_{t\le T_n}\sqrt{n}|\hat{G}_n(t)-G(t)|-\sup_{t\le T_n}|W_n(t)|+\sup_{t\le T_n}|W_n(t)| \\
&\le \sup_{t\le T_n}(|\sqrt{n}(\hat{G}_n(t)-G(t))|-|W_n(t)|)+\sup_{t\le T_n}|W_n(t)| \\
&\le \sup_{t\le T_n}|\sqrt{n}(\hat{G}_n(t)-G(t))-W_n(t)|+\sup_{t\le T_n}|W_n(t)| \\
&= \sup_{t\le T_n}|W_n(t)|+C_6(n^{-\frac{1}{15}}\log n). \qquad (8)
\end{aligned}
$$

又因为$\tau_H<\tau_G$. 由引理1 $W_n(t)$的表达式

$$
|W_n(t)|\le(1-G(t))\{|\int_{-\infty}^{t}B_n^0(s)(1-H(s))^{-2}\mathrm{d}G^1(s)|+|B_n^1(t)(1-H(t))^{-1}| + |\int_{-\infty}^{t}B_n^1(s)(1-H(s))^{-2}\mathrm{d}H(s)|\}
$$

$$
\begin{aligned}
\sup_{t\le T_n}|W_n(t)| &\le C_7\{\sup_{t\le T_n}\int_{-\infty}^{t}|B_n^0(s)|\frac{1-F(s)}{(1-H(s))^2}\mathrm{d}G(s)+\sup_{t\le T_n}\left|\frac{B_n^1(t)}{1-H(t)}\right| \\
&\quad +\sup_{t\le T_n}\int_{-\infty}^{t}|B_n^1(s)|\frac{\mathrm{d}H(s)}{(1-H(s))^2}\} \\
&\le C_8\{\sup_{t\le T_n}|B_n^0(t)|\int_{-\infty}^{T_n}\frac{\mathrm{d}G(s)}{1-H(s)}+\sup_{t\le T_n}|B_n^1(t)|\frac{1}{1-H(T_n)} \\
&\quad +\sup_{t\le T_n}|B_n^1(t)|\int_{-\infty}^{T_n}\frac{\mathrm{d}H(s)}{(1-H(s))^2}\} \\
&\le C_9 n^{\frac{1}{8}-c}\{\sup_{t\le T_n}|B_n^0(t)|+\sup_{t\le T_n}|B_n^1(t)|\}.
\end{aligned}
$$

注意B_n^0; B_n^1的协方差结构，以及

$$
\{B_n^1(t),\quad -\infty<t<\infty\}\stackrel{\mathscr{D}}{=}\{B(G^1(u)),\quad -\infty<u<\infty\}
$$
$$
\{B_n^0(t),\quad -\infty<t<\infty\}\stackrel{\mathscr{D}}{=}\{B(H(u)).\quad -\infty<u<\infty\}
$$

(cf[5]) 其中 $B(\cdot)$ 是Brownian桥，有

$$
P(\sup_{-\infty<t<\infty}|B_n^i(t)|>u)\le 2\exp(-2u^2),\quad u>0,\ i=0,1.
$$

取 $u=\sqrt{\log n}$ ，由 Borel-Cantelli 引理，

$$
\sup_t|B_n^i(t)|=O(\sqrt{\log n}),\qquad \text{a.s.},
$$

所以

$$
\sup_{t\le T_n}|W_n(t)|\le C_{10}n^{\frac{1}{8}-c}(\log n)^{\frac{1}{2}}=C_{10}n^{\frac{13}{120}}(\log n)^{\frac{1}{2}}.
$$

由(8)式得

$$
n\sup_{t\le T_n}|\hat{G}_n(t)-G(t)|^2\le 2[\sup_{t\le T_n}|W_n(t)|]^2+O(n^{-\frac{2}{15}}(\log n)^2).
$$

从而

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_n)\sqrt{n}[\varphi_1(z_i,\hat{G}_n(z_i))-\varphi_1(z_i,G(z_i))-\varphi_1^*(z_i,\hat{G}_n(z_i)-G(z_i))]\delta_i\right| \\
&\le C_{11}\frac{1}{\sqrt{n}}[n\sup_{t\le T_n}|\hat{G}_n(t)-G(t)|^2]+o_p(1) \\
&\le C_{12}n^{-\frac{1}{2}+\frac{13}{120}}(\log n)^{\frac{1}{2}}+O(n^{-\frac{1}{2}-\frac{2}{15}}(\log n)^2)+o_p(1) \\
&= o_p(1).
\end{aligned}
$$

对φ_2的部分也同样处理得$\xi_{4n}=o_p(1)$. ■

引理 8

$$\left|\frac{\sqrt{n}}{\sum(x_j-\overline{x}_n)^2}\sum_{i=1}^n(x_i-\overline{x}_n)I_{z_i\le T_n}\{\varphi_1^*[z_i,(1-G(z_i))(\log(1-\hat{G}_n(z_i))-\log(1-G(z_i)))]\delta_i+\varphi_2^*[z_i,(1-G(z_i))(\log(1-\hat{G}_n(z_i))-\log(1-G(z_i)))](1-\delta_i)\}+\xi_{n2}\right|=o_p(1).$$

证明

$$-[\hat{G}_n(z_i)-G(z_i)]=[1-\hat{G}_n(z_i)]-[1-G(z_i)]=e^{\log(1-\hat{G}_n(z_i))}-e^{\log(1-G(z_i))}$$

$$=(1-G(z_i))[\log(1-\hat{G}_n(z_i))-\log(1-G(z_i))]+\frac{1}{2}(1-G^*(z_i))[\log(1-\hat{G}_n(z_i))-\log(1-G(z_i))]^2.$$

其中$1-G^*(z_i)$介于$1-\hat{G}_n(z_i)$与$1-G(z_i)$之间，由于φ_j^*的线性，我们只须证明

$$\left|\frac{\sqrt{n}}{\sum(x_j-\overline{x}_n)^2}\sum_{i=1}^n(x_i-\overline{x}_n)I_{z_i\le T_n}\{\varphi_1^*[z_i,(1-G^*(z_i))(\log(1-\hat{G}_n(z_i))-\log(1-G(z_i)))^2]\delta_i+\varphi_2^*[z_i,(1-G^*(z_i))(\log(1-\hat{G}_n(z_i))-\log(1-G(z_i)))^2](1-\delta_i)\}\right|\xrightarrow[p]{}0$$

对$j=1,2$，我们有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n|x_i-\overline{x}_n|\cdot|\varphi_j^*[z_i,(1-G^*(z_i))(\log(1-\hat{G}_n(z_i))-\log(1-G(z_i)))^2]|I_{z_i\le T_n}$$

$$\le C_{13}\sqrt{n}\sup_{t\le T_n}|\log(1-\hat{G}_n(t))-\log(1-G(t))|^2$$

$$=C_{13}\sqrt{n}\sup_{t\le T_n}|(R_{n1}(t)-ER_{n1}(t))+R_{n2}(t)+R_{n3}(t)|^2$$

$$\le C_{13}\sqrt{n}\cdot 3\{\sup_{t\le T_n}|R_{n1}(t)-ER_{n1}(t)|^2+\sup_{t\le T_n}|R_{n2}(t)|^2+\sup_{t\le T_n}|R_{n3}(t)|^2\}$$

现在分别考虑上式右端括号内的三项.

(i) $R_{n1}(t)-ER_{n1}(t)$: 当t固定时为零均值独立同分布随机变量之和对n的比值. 由重对数律，其阶为$\left(\frac{\log\log n}{n}\right)^{\frac{1}{2}}$，且此同分布随机变量的方差不大于$(1-H(t))^{-2}$ 所以

$$\sup_{t\le T_n}|R_{n1}(t)-ER_{n1}(t)|^2\le C_{14}[(\log\log n)^{\frac{1}{2}}n^{-\frac{1}{2}}(n^{-\frac{1}{8}+c})^{-1}]^2$$

$$=C_{14}n^{-\frac{3}{4}-2c}\log\log n=C_{14}n^{-\frac{47}{60}}\log\log n.$$

(ii) $\sup_{t\le T_n}|R_{n2}(t)|^2=C_{15}n^{-\frac{47}{30}}$ (引理 6).

(iii) 由引理 6 的证明可知

$$|R_{n3}(t)|\le\left|\frac{1}{n}\sum_{j=1}^n I_{z_j\le t}\delta_j'\frac{H_n(z_j)-H(z_j)}{[1-H(z_j)]^2}\right|+\left|\frac{1}{n}\sum_{j=1}^n I_{z_j\le t}\delta_j'\frac{[H_n(z_j)-H(z_j)]^2}{[1-H_n(z_j)][1-H(z_j)]^2}\right|$$

$$+C_{16}\frac{1}{n(1-H(t))^2}$$

$$\sup_{t\le T_n}|\dot{R}_{n3}(t)|^2\le 3\sup_{t\le T_n}\left|\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{H_n(z_j)-H(z_j)}{[1-H(z_j)]^2}\right|^2$$

$$+C_{17}\sup_{t\le T_n}\left|\frac{\sup_{u\le t}[H_n(u)-H(u)]^2}{(1-H(t))^3}\right|^2+C_{18}n^{-\frac{47}{30}}$$

$$\le C_{19}\left[\frac{\sup_{t\le T_n}|H_n(t)-H(t)|}{[1-H(T_n)]^2}\right]^2+C_{17}n^{-\frac{5}{4}-6c}(\log\log n)^2+C_{18}n^{-\frac{47}{30}}$$

$$=C_{20}n^{-\frac{1}{2}-4c}\log\log n+C_{17}n^{-\frac{5}{4}-6c}(\log\log n)^2+C_{18}n^{-\frac{47}{30}}$$

$$=O(n^{-\frac{1}{2}-4c}\log\log n).$$

从而

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|x_i-\overline{x}_n|\cdot|\varphi_j^*[\dot{z}_j,(1-G^*(z_j))(\log(1-\hat{G}_n(z_i))-\log(1-G(z_i)))^2]|I_{z_i\le T_n}$$

$$=O(n^{-4c}\log\log n)=o(1)$$

又 $\sum(x_j-\overline{x}_n)^2/n\to M$（常数）. 即得引理. ■

引理9 记

$$\widetilde{R}_{n3}=-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le t}\delta_j'\frac{H_n(z_j)-H(z_j)}{[1-H(z_j)]^2},$$

$$\widetilde{\xi}_{2n}=\frac{\sqrt{n}}{\sum(x_j-\overline{x}_n)^2}\sum_{i=1}^{n}(x_i-\overline{x}_n)I_{z_i\le T_n}\{\varphi_1^*[z_i,(1-G(z_i))(R_{n1}(z_i)$$

$$-ER_{n1}(z_i)+\widetilde{R}_{n3}(\dot{t})]\delta_i+\varphi_2^*[z_i,(1-G(z_i))(R_{n1}(z_i)$$

$$-ER_{n1}(z_i+\widetilde{R}_{n3}(t))](1-\delta_i)\},$$

则 $\xi_{2n}=\widetilde{\xi}_{2n}+o_p(1)$.

证明 由引理 8 的证明即得. ■

引理10 $\xi_{3n}=o_p(1)$.

证明 取 $A_1(n)=T_n$，本引理为引理 3 的直接推论. ■

至此为止，我们已证明了 $\sqrt{n}(\hat{\beta}_n-\beta)=\xi_{1n}+\widetilde{\xi}_{2n}+o_p(1)$. 令

$$U_n=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_n)\{[y_i^*-E_{x_i}y_i^*]+I_{z_i\le T_n}(\varphi_1^*[z_i,(1-G(z_i))(-\frac{1}{n}\sum I_{z_j\le z_i}\delta_j'(1-H(z_j))^{-1}$$

$$+E_iI_{z_j\le z_i}\delta_j'(1-H(z_j))^{-1}-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le z_i}\delta_j'\frac{H_n(z_j)-H(z_j)}{(1-H(z_j))^2})]\delta_i+\varphi_2^*[z_i,(1-G(z_i))$$

$$\cdot(-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le z_i}\delta_j'(1-H(z_j))^{-1}+E_iI_{z_j\le z_i}\delta_j'(1-H(z_j))^{-1}$$

$$-\frac{1}{n}\sum_{j=1}^{n}I_{z_j\le z_i}\delta_j'\frac{H_n(z_j)-H(z_j)}{(1-H(z_j))^2})]\cdot(1-\delta_i))\}$$

$$=\frac{\sum(x_j-\overline{x}_n)^2}{n}\cdot\frac{\xi_{1n}+\widetilde{\xi}_{2n}}{\sqrt{n}}$$

显然 U_n 是一个 U- 统计量，可以算出它的对称核为：

$$h=\frac{1}{3}\sum_{\substack{\text{下标为}i,j,\\ k\,3\text{种交换}}}^{*}(x_i-\overline{x}_n)\{(y_i^*-E_{x_i}y_i^*)+I_{z_i\le T_n}\varphi_1^*[z_i,(1-G(z_i))(E_iI_{z_0\le z_i}\delta_0'$$

$$\cdot(1-H(z_0))^{-1})]\delta_i + I_{z_i\le T_n}\varphi_2^*[z_i,\ (1-G(z_i))(E_i I_{z_0\le z_i}\delta_0'(1-H(z_0))^{-1})](1-\delta_i)\}$$

$$+\frac{1}{6}\sideset{}{^*}\sum_{\substack{\text{下标为}\\(i,j),(i,k),(j,k)\\(k,i),(j,i),(k,j)\\6\text{种交换}}}(x_i-\overline{x_n})I_{z_i\le T_n}\{\varphi_1^*[z_i,\ (1-G(z_i))I_{z_j\le z_i}\delta_j'(2H(z_j)-1)(1-H(z_j))^{-2}]\delta_i$$

$$+\varphi_2^*[z_i,\ (1-G(z_i))I_{z_j\le z_i}\delta_j'(2H(z_j)-1)(1-H(z_j))^{-2}](1-\delta_i)\}$$

$$+\frac{1}{6}\sideset{}{^*}\sum_{\substack{\text{下标为}\\(i,j,k),(i,k,j),(j,i,k)\\(j,k,i),(k,i,j),(k,j,i)\\6\text{种交换}}}(x_i-\overline{x_n})I_{z_i\le T_n}\{\varphi_1^*[z_i,\ (1-G(z_i))I_{z_j\le z_i}\delta_j'(1-H(z_j))^{-2}I_{z_k\le z_j}]\delta_i$$

$$+\varphi_2^*[z_i,\ (1-G(z_i))I_{z_j\le z_i}\delta_j'(1-H(z_j))^{-2}I_{z_k\le z_j}](1-\delta_i)\}.$$

其中(z_0,δ_0)与(z_i,δ_i)独立同分布. ■

由$\sum(x_j-\overline{x_n})^2/n\to M$（常数）以及U-统计量的渐近正态性：$\sqrt{n}\,U_n\xrightarrow{\mathcal{L}}N(0,\sigma^{*2})$立即得到：

定理 $\sqrt{n}(\hat{\beta}_n-\beta)\xrightarrow{\mathcal{L}}N(0,\sigma^{*2}/M^2)$. 其中$\sigma^{*2}$由U-统计量$U_n$的对称核$h$所决定.

参 考 文 献

[1] Miller, R. G., Least Squares Regression with Censored Data, Biometrika 63, 1976, 449—464.

[2] Koul, H., Susarla, V. and Van Ryzin, J., Ann. Stat. 9, 1981, 1276—1288.

[3] Kaplan, E. L. and Meier, P., JASA 53, 1958, 457—481.

[4] Gill, R., Ann. Stat. 11, 1983, 49—58.

[5] Burke, M. D., Csörgö, S., Horváth, L., Strong Approximations of some Biometric Estimates under Random Censorship, Z. Wahrscheinlichkeitstheorie verw, Gebiete 56, 1981, 87—112.

[6] Zheng Zukang, Acta Mathematica Sinia, New Series, 1986, Vol. 2, No. 2, 144—151.

[7] Lenglart, E., Relation de domination entre deux processus, Ann. Inst. Henri Poincaré 13, 1977, 171—179.

[8] Zheng Zukang, Acta Mathematicae Applicatae Sinica Vol. 3, No. 3.

Random Design Regression Analysis with Censored Data

Zheng Zukang

(Fudan University)

Abstract

In this paper, the random design regression analysis with censored data is discussed. The estimators $\hat{\beta}_n$ of the parameter β belong to certain subset of Class K. Under some conditions the limit distribution of $\sqrt{n}(\hat{\beta}_n-\beta)$ is normal distribution.