缺失数据下条件分位数的估计及其渐近性质

Estimates and Their Asymptotic Properties of Conditional Quantile Under Missing Data

郑李玲[1,2]

ZHENG Li-ling[1,2]

(1. 广西师范大学数学科学学院，广西桂林　541004；2. 钦州学院数学与计算机科学系，广西钦州　535000)

(1. College of Mathematical Science, Guangxi Normal University, Guilin, Guangxi, 541004, China; 2. Department of Mathematics and Computer Science, Qinzhou University, Qinzhou, Guangxi, 535000, China)

摘要：在响应变量满足随机缺失机制下，利用完全记录单元方法分别给出不含附加信息和含附加信息时条件分位数的估计，并在给定的正则条件下证明估计的渐近正态性.

关键词：条件分位数　缺失数据　渐近正态性　经验似然

中图法分类号：O212.7　**文献标识码**：A　**文章编号**：1005-9164(2009)03-0264-04

Abstract: In response to the variables in the random missing mechanism, we employ the C-C method to propose estimators of conditional quantiles in the absence and presence of some auxiliary information. It is shown that proposed estimators are asymptotic normally distributed in certain regularity conditions.

Key words: conditional quantile, missing data, asymptotic normality, empirical likelihood

设 $(X_i,Y_i),1\leqslant i\leqslant n$，为取自于总体 $(X,Y)((X,Y)\in R^d\times R)$ 的iid. 样本，记 $F(x)=P(X\leqslant x)$，$F_x(y)=P(Y\leqslant y|X=x)$，$x\in R^d$，$Y\in R$. 设 $0<q<1$，定义 $\rho(q|x_0)=\inf\{y|F_{x_0}(y)\geqslant q\}$ 为给定 $X=x_0$ 下 Y 的条件 q 分位数. 条件分位数的估计及其渐近性质是统计界十分关注的一个问题，文献[1～4]在完全样本下对条件分位数的估计及渐近性质进行研究，利用附加信息给出了条件分位数一类新的估计，并在一定的正则条件下证明新估计比未利用附加信息的估计更渐近有效.

缺失数据现象在现实领域中普遍存在，比如市场调研、民意调查、生存分析、可靠性寿命试验、医药研究等领域经常会有缺失数据情况发生，因此，对缺失数据的统计性质进行研究具有非常重要的实际意义. 处理缺失数据的方法有很多，可归结为4大类[5]. 本文利用文献[5]中基于完全记录单元的方法(即Complete-Case，简称C·C方法)分别研究不含附加信息和含附加信息时条件分位数的估计及其渐近性质.

1　两种条件分位数的估计及相关引理

文中设有不完全 iid. 样本 $\{X_i,Y_i,\delta_i\}_{i=1}^n$，其中 $\{X_i\}_{i=1}^n$ 被完全观测，$\{Y_i\}_{i=1}^n$ 有缺失，δ_i 为指示 Y_i 缺失的变量，即

$$\delta_i=\begin{cases}0, & \text{若 } Y_i \text{ 缺失},\\ 1, & \text{若 } Y_i \text{ 不缺失},\end{cases}$$

且满足随机缺失(MAR)机制，即

$$P\{\delta_i=1|x_i,y_i\}=P\{\delta_i=1|x_i\}=P(x_i),1\leqslant i\leqslant n. \tag{1}$$

1.1　条件分位数估计

1.1.1　不含附加信息时条件分位数的估计

取 Borel 可测函数 $K(u)(u\in R^d)$ 和正数列 $h=h_n\to 0$，定义

$$W_{ni}(x)=K(\frac{x-X_i}{h})/\sum_{j=1}^n\delta_jK(\frac{x-X_j}{h}). \tag{2}$$

定义 $F_{x_0}(y)$ 的估计为

收稿日期：2009-03-05

作者简介：郑李玲(1978-)女，硕士，助教，主要从事概率论与数理统计研究。

$$F^{(1)}(y) \triangleq F_{x_0,n,1}(y) = \sum_{i=1}^{n} \delta_i W_{ni}(x_0) I(Y_i \leqslant y), \tag{3}$$

其中 $I(\cdot)$ 为示性函数. $\rho(q|x_0)$ 的估计定义为

$$\rho^{(1)}(q|x_0) \triangleq \rho_{x_0,n,1}(q|x_0) = \inf\{y|F^{(1)}(y) \geqslant q\}. \tag{4}$$

1.1.2 含附加信息时条件分位数的估计

设存在 $r(r \geqslant 1)$ 个已知函数 $g_1(y),\cdots,g_r(y)(g_i(y) \in R, i=1,\cdots,r)$,使得

$$E(g(Y)|X = x_0) = 0, \tag{5}$$

其中

$$g(y) = (g_1(y),\cdots,g_r(y))^\tau. \tag{6}$$

若令

$$M(x) = E(g(Y)|X = x), \tag{7}$$

则由(5)式知 $M(x_0) = 0$. 引入经验似然函数 $L = \prod_{i=1}^{n} p_i$,其中 $p_1,\cdots,p_n$ 满足

$$p_i \geqslant 0, i = 1,\cdots,n, \sum_{i=1}^{n} p_i = 1, \tag{8}$$

$$\sum_{i=1}^{n} p_i \delta_i K(\frac{x_0 - X_i}{h}) g(Y_i) = 0. \tag{9}$$

用Lagrange乘子法[6]求出在约束条件(8)和(9)下的最大值 $L_{\max} = \prod_{i=1}^{n} \hat{p}_i$,其中

$$\hat{p}_i = \frac{1}{n} \frac{1}{1 + \eta^\tau \delta_i K(\frac{x_0 - X_i}{h}) g(Y_i)}. \tag{10}$$

η 由下式确定,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i K(\frac{x_0 - X_i}{h}) g(Y_i)}{1 + \eta^\tau \delta_i K(\frac{x_0 - X_i}{h}) g(Y_i)}. \tag{11}$$

与文献[7]类似可证,当 n 充分大时,$\sum_{i=1}^{n} \delta_i K^2(\frac{x_0 - X_i}{h}) g(Y_i) g^\tau(Y_i) > 0$(正定),且 0 在 $\delta_1 K(\frac{x_0 - X_1}{h}) g(Y_1),\cdots,\delta_n K(\frac{x_0 - X_n}{h}) g(Y_n)$ 的凸包内. 以下均假定 n 充分大,则由文献[6]的说明知,η 由(11)式唯一确定.

由上述分析,我们定义 $F_{x_0}(y)$ 的估计为

$$F^{(2)}(y|x_0) \triangleq F_{x_0,n,2}(y|x_0) = \sum_{i=1}^{n} \hat{p}_i \delta_i I(Y_i \leqslant y) K(\frac{x_0 - X_i}{h}) / \sum_{j=1}^{n} \hat{p}_j \delta_j K(\frac{x_0 - X_j}{h}), \tag{12}$$

从而在含附加信息(5)时,$\rho(q|x_0)$ 的估计定义为

$$\rho^{(2)}(q|x_0) \triangleq \rho_{x_0,n,2}(q|x_0) = \inf\{y|F^{(2)}(y|x_0) \geqslant q)\}. \tag{13}$$

令 $V(x) = E\{g(Y)g^\tau(Y)|X = x\}$,$f(x) = F'(x)$,$f_x(y) = \partial F_x(y)/\partial y$,$\mu_x(y) = E\{g(y)I(Y > y)|X = x_0\}$,$\mu_0 = \mu_{x_0}(\rho(q|x_0))$.

给出正则条件:

(i) $F_x(y)$ 关于变量 x 的所有 $r(r \geqslant 2)$ 阶混合偏导在 $(x_0,\rho(q|x_0))$ 的某领域内存在且连续,$f_{x_0}(y)$ 和 $\mu_x(y)$ 在 $(x_0,\rho(q|x_0))$ 的某领域内连续,且 $f_{x_0}(\rho(q|x_0)) > 0$.

(ii) $M(\cdot)$,$f(\cdot)$ 及 $p(\cdot)$ 的所有 $r(r \geqslant 2)$ 阶混合偏导在点 x_0 的某邻域内存在且连续,$V(\cdot)$ 在点 x_0 的某邻域内连续,且 $M(x_0) = 0$,$p(x_0) > 0$,$f(x_0) > 0$,$V(x_0) > 0$.

(iii) $K(\cdot)$ 有界且有紧支撑,同时对于满足 $1 \leqslant s_1 + \cdots + s_d \leqslant r - 1$ 的非负整数有 $\int_{R^d} u_1^{s_1} \cdots u_d^{s_d} K(u_1,\cdots,u_d) \mathrm{d}_{u_1},\cdots,\mathrm{d}_{ud} = 0$.

(iv) 存在 $s \geqslant 3$,使得 $E\|g(Y)\|^s < \infty$.

(v) $h \to 0, nh^d \to \infty, nh^{d+2r} \to 0, n^{s-2}h^{sd} \to \infty$.

(vi) $q(1-q) \neq 0, q(1-q) - \mu_0^\tau V^{-1}(x_0)\mu_0 \neq 0$.

1.2 相关引理

为了方便,记 $\omega_i = \delta_i K(\frac{x_0 - X_i}{h}) g(Y_i)$,$\bar{\omega} = (nh^d)^{-1} \sum_{i=1}^{n} \omega_i$.

引理 1 设条件(i) ~ (v) 满足,则

$$\sqrt{nh^d}\, \bar{\omega} \xrightarrow{d} N(0, p(x_0) f(x_0) V(x_0) \cdot \int_{R^d} K^2(u) \mathrm{d}u). \tag{14}$$

证明 因为

$$E(\bar{\omega}) = h^{-d} E(\delta K(\frac{x_0 - X}{h}) g(Y)) = h^{-d} E\{E(\delta K(\frac{x_0 - X}{h}) g(Y))|X = u\} = h^{-d} \int_{R^d} K(\frac{x_0 - u}{h}) p(u) M(u) f(u) \mathrm{d}u = \int_{R^d} K(u) p(x_0 - u) M(x_0 - u) f(x_0 - u) \mathrm{d}u.$$

由 $M(x_0) = 0$,Taylor 展开和条件(ii) 知

$$E(\bar{\omega}) = O(h^r). \tag{15}$$

对任意 $l \in R^r, l \neq 0$,记 $s_n^2 = \mathrm{Var}(l^\tau \bar{\omega})$,$\Gamma_n = s_n^{-3}(nh^d)^{-3} \sum_{i=1}^{n} E|l^\tau \omega_i - E(l^\tau \omega_i)|^3$. 结合条件(v) 知

$$s_n^2 = E(l^\tau \bar{\omega})^2 - (E(l^\tau \bar{\omega}))^2 = (nh^d)^{-1}(l^\tau \cdot p(x_0) f(x_0) V(x_0) \int_{R^d} K^2(u) \mathrm{d}u \cdot l + o(1)) + O(h^{2r}) = (nh^d)^{-1}(l^\tau \cdot p(x_0) f(x_0) V(x_0) \int_{R^d} K^2(u) \mathrm{d}u \cdot l + o(1)). \tag{16}$$

又由条件(v) 知

$$\Gamma_n \leqslant C s_n^{-3}(nh^d)^{-2} \to 0(C\text{ 为常数}). \tag{17}$$

由中心极限定理知

$$l^\tau(\bar{\omega} - E\bar{\omega})/s_n \xrightarrow{d} N(0,1). \tag{18}$$

于是由 Crammer-Wold 定理知引理 1 成立.

引理 2[4] 设条件(i) ~ (v) 满足,则

$$\eta = (p(x_0)f(x_0)V(x_0)\int_{R^d}K^2(u)\mathrm{d}u)^{-1}\bar{\omega} + o_p((nh^d)^{-1/2}). \tag{19}$$

其中 η 由(11)式确定.

证明 令 $\eta = \lambda\theta$,其中 $\lambda \geqslant 0, \theta \in R^r, \|\theta\| = 1, Z_n = \max\limits_{1\leqslant i\leqslant n}|\theta^\tau\omega_i|$. 由(11)式知

$$0 = \| \frac{1}{nh^d}\sum_{i=1}^n \frac{\omega_i}{1+\lambda\theta^\tau\omega_i} \| \geqslant \frac{1}{nh^d}|\theta^\tau(\sum_{i=1}^n\omega_i - \lambda\sum_{i=1}^n\frac{\omega_i\theta^\tau\omega_i}{1+\lambda\theta^\tau\omega_i})| \geqslant \frac{\lambda}{nh^d}\theta^\tau\sum_{i=1}^n\frac{\omega_i\theta^\tau\omega_i}{1+\lambda\theta^\tau\omega_i}\theta - \frac{1}{nh^d}|\theta^\tau\sum_{i=1}^n\omega_i| \geqslant \frac{\lambda\theta^\tau Q_n\theta}{1+\lambda Z_n} - \frac{1}{nh^d}|\theta^\tau\sum_{i=1}^n\omega_i|,$$

其中 $Q_n = \frac{1}{nh^d}\sum_{i=1}^n\omega_i\omega_i^\tau$. 用 t_1 记 $p(x_0)f(x_0)V(x_0)\cdot\int_{R^d}K^2(u)\mathrm{d}u$ 的最小特征根,则由大数定理知 $\theta^\tau Q_n\theta \geqslant t_1 + o_p(1)$. 结合引理 1 知

$$\frac{\lambda}{1+\lambda Z_n} = O_p((nh^d)^{-1/2}). \tag{20}$$

又 $Z_n = o(n^{1/s})$, a.s., 故由 $n^{s-2}h^{sd} \to \infty$ 知 $\lambda = O_p((nh^d)^{-1/2})$,从而

$$\eta = O_p((nh^d)^{-1/2}). \tag{21}$$

令 $\gamma_i = \eta^\tau\omega_i$,则 $\max\limits_{1\leqslant i\leqslant n}|\gamma_i| = O_p((nh^d)^{-1/2})o_p(n^{1/s}) = o_p(1)$. 由(11)式知 $0 = \frac{1}{nh^d}\sum_{i=1}^n\omega_i(1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i}) = \bar{\omega} - Q_n\eta + \frac{1}{nh^d}\sum_{i=1}^n\frac{\omega_i\gamma_i^2}{1+\gamma_i}$, 而 $|\frac{1}{nh^d}\sum_{i=1}^n\frac{\omega_i\gamma_i^2}{1+\gamma_i}| = \frac{1}{nh^d}\sum_{i=1}^n\|\omega_i\|^3\|\eta\|^2\|1+\gamma_i\|^{-1} = o_p((nh^d)^{-1/2})$,故 $\eta = Q_n^{-1}\bar{\omega} + o_p((nh^d)^{-1/2})$. 由大数定理可以知道引理 2 成立.

引理 3[8] 设 $U, U_{ni}, 1\leqslant i\leqslant 3$,均为随机变量,且 $U_{n1}\xrightarrow{d}U, U_{n2} = a + o_p(1), U_{n3} = b + o_p(1), a, b$ 均为常数,则 $U_{n1}U_{n2} + U_{n3} \xrightarrow{d} aU + b$.

2 主要结果

定理 1 设条件(i) ~ (vi) 满足,则 $\sqrt{nh^d}(\rho^{(1)}(q|x_0) - \rho(q|x_0)) \xrightarrow{d} N(0,\sigma_1^2)$,其中 $\sigma_1^2 = \dfrac{q(1-q)\int_{R^d}K^2(u)\mathrm{d}u}{p(x_0)f(x_0)f_{x_0}^2(\rho(q|x_0))}$.

证明 对任意 $t \in R$,令 $\varepsilon_n = \sigma_1 t(nh^d)^{-1/2}$,用 $\Phi(\cdot)$ 记标准正态随机变量的分布函数,并且设 $H_n = P\{(nh^d)^{1/2}(\rho^{(1)}(q|x_0) - \rho(q|x_0))/\sigma_1 \leqslant t\}$. 要证明定理 1,只需证明

$$H_n \longrightarrow \Phi(t). \tag{22}$$

易知

$$H_n = P\{(\rho^{(1)}(q|x_0) \leqslant \rho(q|x_0) + \varepsilon_n\} = P\{F^{(1)}(\rho(q|x_0)+\varepsilon_n) \geqslant q\} = P\{p(x_0)f(x_0)\cdot[(nh^d)^{-1}\sum_{i=1}^n\delta_i(q - I(Y_i \leqslant \rho(q|x_0) + \varepsilon_n))K(\frac{x_0-X_i}{h})]/[(nh^d)^{-1}\sum_{j=1}^n\delta_jK(\frac{x_0-X_j}{h})] \leqslant 0\}. \tag{23}$$

分别考虑(23)式的分子和分母,根据(1)式,条件(ii),(v)和 Bernstein 不等式,得分母为

$$H_{n1} \triangleq (nh^d)^{-1}\sum_{j=1}^n\delta_jK(\frac{x_0-X_j}{h}) = p(x_0)f(x_0) + o_p(1). \tag{24}$$

(23)式的分子中去掉 $p(x_0)f(x_0)$ 的部分,记

$$H_{n2} \triangleq (nh^d)^{-1}\sum_{i=1}^n\delta_i(q - I(Y_i \leqslant \rho(q|x_0) + \varepsilon_n))K(\frac{x_0-X_i}{h}),$$

$$\sqrt{nh^d}EH_{n2} = \sqrt{nh^d}(nh^d)^{-1}\sum_{i=1}^nE\{\delta_i[q - I(Y_i\leqslant \rho(q|x_0)+\varepsilon_n)]K(\frac{x_0-X_i}{h})\} = \sqrt{nh^d}\{\int_{R^d}[q - F_{x_0-hu}(\rho(q|x_0)+\varepsilon_n)]K(u)p(x_0-hu)f(x_0-hu)\mathrm{d}u + O(h^r)\} = \sqrt{nh^d}\{\int_{R^d}[F_{x_0}(\rho(q|x_0)) - F_{x_0}(\rho(q|x_0)+\varepsilon_n)]K(u)p(x_0)f(x_0)\mathrm{d}u + O(h^r)\} = -\sigma_1 t f_{x_0}(\rho(q|x_0))p(x_0)f(x_0) + o(1), \tag{25}$$

$$nh^d\mathrm{Var}H_{n2} = (nh^d)^{-1}\sum_{i=1}^n\mathrm{Var}\{\delta_i[q - I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n)]K(\frac{x_0-X_i}{h})\} = (nh^d)^{-1}\cdot\sum_{i=1}^nE\{K^2(\frac{x_0-X_i}{h})[\delta_i(q - I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))]^2\} + o(1) = p(x_0)f(x_0)[q(1-q)]\int_{R^d}K^2(u)\mathrm{d}u + o(1), \tag{26}$$

结合条件(vi)知

$$-\frac{EH_{n2}}{\sqrt{\mathrm{Var}H_{n2}}} \longrightarrow t. \tag{27}$$

由中心极限定理知

$$\frac{H_{n2} - EH_{n2}}{\sqrt{\mathrm{Var}H_{n2}}} \xrightarrow{d} N(0,1). \tag{28}$$

由(22)～(28)式及引理3知定理1成立.

定理2 设条件(i)～(vi)满足,则 $\sqrt{nh^d}(\rho^{(2)}(q|x_0)-\rho(q|x_0))\xrightarrow{d}N(0,\sigma_2^2)$,其中 $\sigma_2^2=\dfrac{[q(1-q)-\mu_0^{\tau}V^{-1}(x_0)\mu_0]\int_{R^d}K^2(u)\mathrm{d}u}{p(x_0)f(x_0)f_{x_0}^2(\rho(q|x_0))}$.

证明 对任意 $v\in R$,令 $\varepsilon_n=\sigma_2 v(nh^d)^{-1/2}$,用 $\Phi(\cdot)$ 记标准正态随机变量的分布函数,并设 $T_n=P\{(nh^d)^{1/2}(\rho^{(2)}(q|x_0)-\rho(q|x_0))/\sigma_2\leqslant v\}$. 要证明定理2,只需证明

$$T_n\longrightarrow\Phi(v). \tag{29}$$

易知

$$T_n=P\{(\rho^{(2)}(q|x_0)\leqslant\rho(q|x_0)+\varepsilon_n\}=P\{F^{(2)}((\rho(q|x_0)+\varepsilon_n)|x_0)\geqslant q\}=P\{p(x_0)f(x_0)[h^{-d}\sum_{i=1}^n\hat{p}_i\delta_i(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))K(\frac{x_0-X_i}{h})]/[h^{-d}\sum_{j=1}^n\hat{p}_j\delta_jK(\frac{x_0-X_j}{h})]\leqslant 0\}. \tag{30}$$

分别考察(30)式的分子和分母,由(10)式,(11)式,(19)式,条件(ii),(v)和Bernstain不等式,可知分母变为

$$T_{n1}\triangleq h^{-d}\sum_{j=1}^n\hat{p}_j\delta_jK(\frac{x_0-X_j}{h})=(nh^d)^{-1}\sum_{j=1}^n\delta_jK(\frac{x_0-X_j}{h})-(nh^d)^{-1}\eta^{\tau}\cdot\sum_{j=1}^n\frac{\omega_jK(\frac{x_0-X_j}{h})}{1+\eta^{\tau}\omega_j}=p(x_0)f(x_0)+o_p(1). \tag{31}$$

(30)式分子中去掉 $p(x_0)f(x_0)$ 的部分,记 $T_{n2}\triangleq h^{-d}\sum_{i=1}^n\hat{p}_i\delta_i(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))K(\frac{x_0-X_i}{h})$, $T_{n3}\triangleq T_{n1}^{(1)}-\mu_0^{\tau}V^{-1}(x_0)\overline{\omega}=(nh^d)^{-1}\sum_{i=1}^n\delta_i[(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))-\mu_0^{\tau}V^{-1}(x_0)g(Y_i)]K(\frac{x_0-X_i}{h})$. ($T_{n1}^{(1)}$ 见(33)式)

再证明

$$T_{n2}=T_{n3}+o_p((nh^d)^{-1/2}). \tag{32}$$

我们记

$$T_{n1}^{(1)}\triangleq(nh^d)^{-1}\sum_{i=1}^n\delta_i(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))K(\frac{x_0-X_i}{h}), \tag{33}$$

$I_{n2}\triangleq(nh^d)^{-1}\sum_{i=1}^n\delta_i(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))g(Y_i)K^2(\frac{x_0-X_i}{h})$, $I_{n3}\triangleq(nh^d)^{-1}\sum_{i=1}^n\dfrac{\eta^{\tau}\delta_iK(\frac{x_0-X_i}{h})g(Y_i)}{1+\eta^{\tau}\delta_iK(\frac{x_0-X_i}{h})g(Y_i)}(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))g(Y_i)K^2(\frac{x_0-X_i}{h})$, $T_{n2}^{(2)}\triangleq I_{n2}^{\tau}\cdot\eta$, $T_{n3}^{(3)}\triangleq I_{n3}^{\tau}\cdot\eta$.

由(10)式和(11)式知

$$T_{n2}=T_{n1}^{(1)}-T_{n2}^{(2)}+T_{n3}^{(3)}. \tag{34}$$

由条件(v)知 $I_{n3}=o_p(1)$,因此结合(21)式,得

$$T_{n3}^{(3)}=I_{n3}^{\tau}\cdot\eta=o_p((nh^d)^{-1/2}). \tag{35}$$

由 $\mu_x(y)$ 在 $(x_0,\rho(q|x_0))$ 的邻域内连续且 $M(x_0)=0$,可知

$$I_{n2}\triangleq(nh^d)^{-1}\sum_{i=1}^n\delta_i(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))g(Y_i)K^2(\frac{x_0-X_i}{h})=(nh^d)^{-1}\sum_{i=1}^n\delta_i(I(Y_i\geqslant\rho(q|x_0)+\varepsilon_n)-(1-q))g(Y_i)K^2(\frac{x_0-X_i}{h})=p(x_0)f(x_0)\mu_0\int_{R^d}K^2(u)\mathrm{d}u+o_p(1).$$

因此由(21)式和(19)式可知

$$T_{n2}^{(2)}=(p(x_0)f(x_0)\mu_0\int_{R^d}K^2(u)\mathrm{d}u+o_p(1))\cdot\eta=p(x_0)f(x_0)\mu_0\int_{R^d}K^2(u)\mathrm{d}u[(p(x_0)f(x_0)V(x_0)\cdot\int_{R^d}K^2(u)\mathrm{d}u)^{-1}\overline{\omega}+o_p((nh^d)^{-1/2})]+o_p((nh^d)^{-1/2})=\mu_0^{\tau}V^{-1}(x_0)\overline{\omega}+o_p((nh^d)^{-1/2}). \tag{36}$$

由(34)～(36)式知(32)式成立. 又因为 $M(x_0)=0$,由Taylor展开式知

$$\sqrt{nh^d}ET_{n3}=\sqrt{nh^d}(nh^d)^{-1}\sum_{i=1}^nE\{\delta_i[(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))-\mu_0^{\tau}V^{-1}(x_0)g(Y_i)]\cdot K(\frac{x_0-X_i}{h})\}=\sqrt{nh^d}\{\int_{R^d}[q-F_{x_0-hu}(\rho(q|x_0)+\varepsilon_n)]K(u)p(x_0-hu)f(x_0-hu)\mathrm{d}u+O(h^r)\}=\sqrt{nh^d}\{\int_{R^d}[F_{x_0}(\rho(q|x_0))-F_{x_0}(\rho(q|x_0)+\varepsilon_n)]K(u)p(x_0)f(x_0)\mathrm{d}u+O(h^r)\}=-\sigma_2vf_{x_0}(\rho(q|x_0))p(x_0)f(x_0)+o(1).$$

$$nh^d\mathrm{Var}T_{n3}=(nh^d)^{-1}\sum_{i=1}^n\mathrm{Var}\{\delta_i[(q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))-\mu_0^{\tau}V^{-1}(x_0)g(Y_i)]K(\frac{x_0-X_i}{h})\}=(nh^d)^{-1}\sum_{i=1}^nE\{K^2(\frac{x_0-X_i}{h})[\delta_i((q-I(Y_i\leqslant\rho(q|x_0)+\varepsilon_n))-\mu_0^{\tau}V^{-1}(x_0)g(Y_i))]^2\}+o(1)=p(x_0)f(x_0)[q(1-q)-\mu_0^{\tau}V^{-1}(x_0)\mu_0]\int_{R^d}K^2(u)\mathrm{d}u+o(1).$$

结合条件(vi)知 $-ET_{n3}/\sqrt{\mathrm{Var}T_{n3}}\longrightarrow v$. 由中心极限定理知 $(T_{n3}-ET_{n3})/\sqrt{\mathrm{Var}T_{n3}}\xrightarrow{d}N(0,1)$.

(下转第273页 Continue on page 273)

$\leqslant \sup_{0\leqslant t\leqslant 1}|T_{R_{k,t}}-t|+C\sum_{s>k}v_{ns}=O_p(\frac{k}{n})$，$|J_{n61}|\leqslant o_p(1)\ \frac{1}{n}\sum_{i=1}^{n}|\sum_{s=1}^{n}W_{ns}(T_i)\delta_s\{g(T_s)-g(T_i)\}|\leqslant o_p(1)O_p(\frac{k}{n})=o_p(1)$. 又由于$\{e_i,1\leqslant i\leqslant n\}$ iid. 且与$\{(X_i,T_i),1\leqslant i\leqslant n\}$相互独立，$Ee_i=0$，故

$$E(\sum_{s=1}^{n}W_{ns}(T_i)\delta_s e_s)^2=$$

$$\sum_{s=1}^{n}E(W_{ns}(T_i)^2\delta_s^2 e_s^2)\leqslant\sum_{s=1}^{n}v_{ns}^2\sigma_0^2=O(1/k),$$

因此

$$|J_{n62}|\leqslant|\frac{1}{n}\sum_{i=1}^{n}o_p(\frac{1}{\sqrt{k}})|\leqslant o_p(\frac{1}{\sqrt{k}}),$$

所以$J_{n6}\xrightarrow{P}0$.

再考虑J_{n2}. 记$\beta=(\beta_1,\cdots,\beta_p)'$，$\hat{\beta}_r=(\hat{\beta}_1,\cdots,\hat{\beta}_p)'$，可以得到

$$J_{n2}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\cdot\frac{\sum_{s=1}^{n}\sum_{l=1}^{p}W_{ns}(T_i)\delta_sX_{sl}(\hat{\beta}_{rl}-\beta_l)}{\sum_{s=1}^{n}W_{ns}(T_i)\delta_s}.$$

注意到$\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})=0$，利用大数定律和引理6可得

$$|\frac{1}{n}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\frac{\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}}{\sum_{s=1}^{n}W_{ns}(T_i)\delta_s}|$$

$$\leqslant|\frac{1}{n}\sum_{i=1}^{n}o_p(n^{-1/4})\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\sum_{s=1}^{n}W_{ns}(T_i)\cdot\delta_sX_{sl}|+|p^{-1}\frac{1}{n}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\cdot\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}\triangleq J_{21}+J_{22},$$

$$J_{21}=o_p(n^{-1/4})\frac{1}{n}\sum_{i=1}^{n}|\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}|\leqslant 2o_p(n^{-1/4})\cdot\frac{1}{n}\sum_{i=1}^{n}|X_i|\max_{1\leqslant i\leqslant n}|\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}-E(\delta X_{1l})|=o_p(1),$$

$$J_{22}=p^{-1}|\frac{1}{n}\sum_{i=1}^{n}\delta_i(X_i-\frac{\sum_{i=1}^{n}\delta_iX_i}{\sum_{i=1}^{n}\delta_i})\cdot(\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}-E(\delta_1X_{1l}))|\leqslant 2p^{-2}\cdot\frac{1}{n}\sum_{i=1}^{n}|X_i|\max_{1\leqslant i\leqslant n}|\sum_{s=1}^{n}W_{ns}(T_i)\delta_sX_{sl}-E(\delta_1X_{1l})|\xrightarrow{P}0,$$

最后考虑J_{n3}. 用类似于证明定理2的方法可以证明$J_{n3}\xrightarrow{L}N(0,\sigma_0{}^2\Sigma^{-1})$.

综上所述，定理4成立.

注　文献[1]中定理1～4是本文结果($\delta_i=1$)的特例.

参考文献：

[1]　薛留根，韩建国. 半参数回归模型中二阶段估计的渐近性质[J]. 高校应用数学学报：A辑，2001，16(1)：87-94.

[2]　洪圣岩. 一类半参数回归模型的估计理论[J]. 中国科学：A辑，1991，21(12)：1258-1272.

[3]　赵林城，白志东. 非参数回归函数最近邻估计的强相合性[J]. 中国科学：A辑，1984，14(5)：387-393.

[4]　洪圣岩. 最近邻回归估计的渐近正态性[J]. 应用概率统计，1991，7(2)：187-191.

（责任编辑：尹　闯）

（上接第267页 Continue from page 267）

注　当$\delta_i\equiv 1$时，由本文的结果可以推出文献[2]和文献[4]的主要结果. 易知$\sigma_2^2\leqslant\sigma_1^2$，即附加信息下的估计比未含附加信息的估计更渐近有效.

致谢：

感谢秦永松教授给予的指导和帮助.

参考文献：

[1]　郑忠国. 条件中位数的最近邻估计和它的Bootstrap统计量的渐进性质[J]. 中国科学：A辑，1984，12：1074-1089.

[2]　Liu Z J，Tu D S. Kernel method on conditona median estimation [J]. Chinese Science Bulletin，1987，5：642-643.

[3]　Xiang X J. A kernel estimation of a conditional quantile [J]. Multivarate Anal，1996，59：206-216.

[4]　秦永松，苏淳. 附加信息时条件分位数的估计及其渐进性质应[J]. 应用数学学报，2000，23(1)：55-62.

[5]　范承华. 缺失数据半参数回归分析[D]. 北京：北京工业大学，2007.

[6]　Qin J，Lawless J. Empirical likelihood and general estimating equation[J]. Ann Statist，1994，22：300-325.

[7]　Owen A B. Empirical likelihood confidence regions[J]. Ann Statist，1990，18：90-120.

[8]　Chow Y S，Teicher H. Probbability theory[M]. New York：Spring-Verlag，1988.

（责任编辑：尹　闯）