DOI:10.13657/j.cnki.gxkxyxb.2013.04.014

广西科学院学报 2013,29(4):269~274,277
Journal of Guangxi Academy of Sciences Vol. 29,No. 4 November 2013

缺失数据情形条件分位数的经验似然置信区间*

Empirical Likelihood Confidence Intervals for Conditional Quantiles with Missing Data

郑李玲

ZHENG Li-ling

(钦州学院数学与计算机科学学院,广西钦州 535000)

(School of Mathematics and Computer Science,Guangxi Qinzhou University,Qinzhou,Guangxi,535000,China)

摘要:在响应变量满足随机缺失机制下,采用经验似然方法和 C-C 方法,分别构造不含附加信息和含附加信息时条件分位数的经验似然比统计量的渐近分布,以及条件分位数的经验似然置信区间,并证明一类检验问题的渐近功效随着信息量的增加而非下降.

关键词:缺失数据 条件分位数 经验似然 附加信息

中图法分类号:O212.7 **文献标识码**:A **文章编号**:1005-9164(2013)04-0269-06

Abstract:When the response variable is missing at random, empirical likelihood and the C-C method is combined to obtain asymptotic distributions of empirical likelihood ratio statistics in the absence and presence of some auxiliary information. Then empirical likelihood ratio confidence intervals for conditional quantiles are constructed. It is shown that asymptotic efficacy of some test problems increases along with raise of the information quantity.

Key words:missing data, conditional quantile, empirical likelihood, auxiliary information

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,$F(x)=P(X\leqslant x)$,$F_x(y)=P(Y\leqslant y\mid X=x)$,$x\in R^d$,$Y\in R$.若 $0<q<1$,对给定 $X=x_0$ 情况下,Y 的条件 q 分位数 θ_0 唯一确定,则记 $\theta_0=F^{-1}(q\mid x_0)$.$\theta_0$ 的估计及其渐近性质是统计界十分关注的一个问题[1~4],文献[5,6]在完全样本下用经验似然方法构造条件分位数 θ_0 的置信区间,目前尚未发现缺失数据情形条件分位数 θ_0 的经验置信区间的相关结论报道.经验似然是一种重要的统计推断方法,由文献[7,8]首先提出,并用来处理非参数统计问题.有些学者将经验似然与其它一些统计学方法(如 Bootstrap 法)作比较,证明经验似然方法有很多突出的优点,如:用经验似然方法构造置信区间.该方法除了有域保持性、变换不变性及置信域的形状由数据自行决定等诸多优点外,还有 Bartlett 纠偏性及无需构造枢轴统计量等优点.因此,经验似然方法已广泛应用到统计模型中[9~12].

缺失数据现象在实际领域中普遍存在,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,因此,缺失数据情形的统计分析具有重要的实际意义.学者们已提出很多处理缺失数据的方法,如:将缺失数据的记录删除(即 Complete-Case,简称 C-C 方法)、加权法、借补法等[13].本文把经验似然方法和处理缺失数据的 C-C 方法相结合讨论条件分位数 θ_0 的似然置信区间的构造.

1 条件分位数经验似然比统计量的构造

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

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

收稿日期:2013-04-27

修回日期:2013-06-02

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

*国家自然科学基金项目(10971038)资助。

且满足随机缺失(MAR)机制,即

$$P\{\delta_i=1\mid x_i,y_i\}=P\{\delta_i=1\mid x_i\}=p(x_i),1\leqslant i\leqslant n. \quad (1)$$

为了方便使用,引入记号

$$s_r=\{i:\delta_i=1,i=1,\cdots,n\}.$$

1.1 不含附加信息时的经验似然比统计量

类似文献[14]中的方法,考虑经验似然函数

$$L_1^{(1)}=\prod_{i=1}^n p_i, \quad (2)$$

其中 p_i 满足 $p_i\geqslant 0,i=1,\cdots,n,\sum_{i=1}^n p_i=1$. 易知(2)式的最大值为 n^{-n}.

取Borel可测函数 $K_1(u),K_2(v)(u\in R^d,v\in R)$ 和正数列 $h=h_n\to 0$,令

$$G(t)=\int_{-\infty}^{t/h}K_2(v)\mathrm{d}v.$$

考虑关于 θ_0 的经验似然函数 $L_1^{(2)}=\prod_{i=1}^n p_i$,其中 p_i 满足(3)式和(4)式:

$$p_i\geqslant 0,i=1,\cdots,n,\sum_{i=1}^n p_i=1, \quad (3)$$

$$\sum_{i=1}^n p_i\delta_i K_1(\frac{x_0-X_i}{h})(G(\theta_0-Y_i)-q)=0. \quad (4)$$

用Lagrange乘子法求出 $L_1^{(2)}$ 在约束条件(3)和(4)下的最大值 $\max L_1^{(2)}=\prod_{i=1}^n\overline{p}_i$,其中

$$\overline{p}_i=\frac{1}{n}\frac{1}{1+\lambda(\theta_0)\delta_i K_1(\frac{x_0-X_i}{h})(G(\theta_0-Y_i)-q)},$$

$i=1,\cdots,n$,

且 $\lambda(\theta_0)$ 由(5)式唯一确定

$$\frac{1}{n}\sum_{i=1}^n[\delta_i K_1(\frac{x_0-X_i}{h})(G(\theta_0-Y_i)-q)]/[1+\lambda(\theta_0)\delta_i K_1(\frac{x_0-X_i}{h})(G(\theta_0-Y_i)-q)]. \quad (5)$$

于是在不含附加信息时关于 θ_0 的经验似然比统计量为

$$l_1(\theta_0)=\prod_{i=1}^n\frac{n^{-n}}{\overline{p}_i}=\prod_{i=1}^n[1+\lambda(\theta_0)\delta_i K_1(\frac{x_0-X_i}{h})(G(\theta_0-Y_i)-q)]^{-1}.$$

1.2 含附加信息时的经验似然比统计量

设附加信息:

$$E(g_{(r)}(Y)\mid X=x_0)=0. \quad (6)$$

其中 $g_{(r)}(y)=(g_1(y),\cdots,g_r(y))^\tau(g_i(y)\in R,i=1,\cdots,r)$ 为 $r(r\geqslant 1)$ 个已知函数.

考虑经验似然函数

$$L_2^{(1)}=\prod_{i=1}^n p_i,$$

其中 p_i 满足约束条件(7)和(8):

$$p_i\geqslant 0,i=1,\cdots,n,\sum_{i=1}^n p_i=1, \quad (7)$$

$$\sum_{i=1}^n p_i\delta_i K_1(\frac{x_0-X_i}{h})g_{(r)}(Y_i)=0. \quad (8)$$

用Lagrange乘子法求出 $L_2^{(1)}$ 在约束条件(7)和(8)下的最大值 $\max L_2^{(1)}=\prod_{i=1}^n\hat{p}_i$,其中

$$\hat{p}_i=\frac{1}{n}\frac{1}{1+\eta_1^\tau\delta_i K_1(\frac{x_0-X_i}{h})g_{(r)}(Y_i)},$$

$i=1,\cdots,n$,

且 η_1 由下式唯一确定

$$\frac{1}{n}\sum_{i=1}^n\frac{\delta_i K_1(\frac{x_0-X_i}{h})g_{(r)}(Y_i)}{1+\eta_1^\tau\delta_i K_1(\frac{x_0-X_i}{h})g_{(r)}(Y_i)}.$$

再考虑经验似然函数

$$L_2^{(2)}=\prod_{i=1}^n p_i,$$

其中 p_i 满足约束条件(7)、(8)和(5)式. 又用Lagrange乘子法求出 $L_2^{(2)}$ 在约束条件(7)、(8)和(5)式下的最大值 $\max L_2^{(2)}=\prod_{i=1}^n\widetilde{p}_i$,其中

$$\widetilde{p}_i=\frac{1}{n}\frac{1}{1+\eta_2^\tau(\theta_0)\delta_i K_1(\frac{x_0-X_i}{h})(g_{(r)}^\tau(Y_i),G(\theta_0-Y_i)-q)^\tau},$$

$i=1,\cdots,n$,

且 $\eta_2(\theta_0)$ 由下式唯一确定

$$\frac{1}{n}\sum_{i=1}^n\frac{\delta_i K_1(\frac{x_0-X_i}{h})(g_{(r)}^\tau(Y_i),G(\theta_0-Y_i)-q)^\tau}{1+\eta_2^\tau(\theta_0)\delta_i K_1(\frac{x_0-X_i}{h})(g_{(r)}^\tau(Y_i),G(\theta_0-Y_i)-q)^\tau}.$$

因此,在含附加信息(6)时,θ_0 的经验似然比统计量为

$$l_2(\theta_0)=\prod_{i=1}^n\frac{\widetilde{p}_i}{\hat{p}_i}=\prod_{i=1}^n\frac{1+\eta_1^\tau\delta_i K_1(\frac{x_0-X_i}{h})g_{(r)}(Y_i)}{1+\eta_2^\tau(\theta_0)\delta_i K_1(\frac{x_0-X_i}{h})(g_{(r)}^\tau(Y_i),G(\theta_0-Y_i)-q)^\tau}.$$

2 条件分位数经验似然置信区间的构造

令

$$f(x)=F'(x), f(y\mid x)=\partial F(y\mid x)/\partial y,$$
$$M(x)=E(g_{(r)}(Y)\mid X=x), V(x)=$$
$$E\{g_{(r)}(Y)g_{(r)}^{\tau}(Y)\mid X=x\}.$$

给出正则条件:

(i)$F_1(x,y)\triangleq F(y\mid x)$ 的所有 $r_0(r_0\geqslant 2)$ 阶混合偏导数在 (x_0,θ_0) 的某领域内存在,且在 (x_0,θ_0) 连续.

(ii)$M(\cdot)$, $f(\cdot)$ 及 $p(\cdot)$ 的所有 $r_0(r_0\geqslant 2)$ 阶混合偏导数在点 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,$p(x_0)>0$, $f(x_0)>0$, $f(\theta_0\mid x_0)>0$ 且 $V(x_0)>0$.

(iii)$K_i(\cdot)(i=1,2)$ 有界且有紧支撑,$\int_{R^d}K_1(u)\mathrm{d}u=1$, $K_2''(\cdot)$ 存在且有界,$\int_{R^d}u_1^{m_1}\cdots u_d^{m_d}K(u_1,\cdots,u_d)\mathrm{d}u_1\cdots\mathrm{d}u_d=0(m_1,\cdots,m_d$ 为非负整数且 $1\leqslant m_1+\cdots+m_d\leqslant r_2-1)$, $\int_R v^{r_2}K_2(v)\mathrm{d}v\neq 0$ 且

$$\int_R v^m K_2(v)\mathrm{d}v=\begin{cases}1, m=0,\\ 0, 1\leqslant m\leqslant r_2-1.\end{cases}$$

(iv)$h\to 0$, $nh^{d+2r_0}\to 0$, $nh^{d+2}\to\infty$.

(v) 存在 $s\geqslant 3$,使得 $E\|g(Y)\|^s<\infty$($\|\cdot\|$ 表示欧氏范数).

(vi)$h\to 0$, $nh^{d+2r_0}\to 0$ 且 $n^{s-2}h^{sd}\to\infty$.

(vii) $\begin{pmatrix}V(x_0) & qg_{(r)}(\theta_0)\\ qg_{(r)}^{\tau}(\theta_0) & q(1-q)\end{pmatrix}$ 正定.

定理1 设条件(i)、(vi) 满足,则对任意常数 δ, $c_0(c_0>0)$,有

$$\lim_{n\to\infty}P\{-2\log l_1(\theta_0)<c_0\}=P\{\chi_1^2<c_0\}, \tag{9}$$
$$\lim_{n\to\infty}P\{-2\log l_1(\theta_n)<c_0\}=P\{\chi_1^2(\delta^2)<c_0\}, \tag{10}$$

其中

$$\theta_n=\theta_0+(nh^d)^{-1/2}\cdot\frac{(q(1-q)p(x_0)f(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u)^{1/2}}{p(x_0)f(x_0)f(\theta_0\mid x_0)}\delta. \tag{11}$$

$\chi_1^2(\delta^2)$ 表示自由度为 1,非中心参数为 δ^2 的非中心 χ^2-分布.

定理2 设条件(i)~(iii)及(v)~(vii)满足,则对任意常数 δ 和 $c_0(c_0>0)$,有

$$\lim_{n\to\infty}P\{-2\log l_2(\theta_0)<c_0\}=P\{\chi_1^2<c_0\}, \tag{12}$$
$$\lim_{n\to\infty}P\{-2\log l_2(\theta_n)<c_0\}=P\{\chi_1^2(\rho^2)<c_0\},$$

其中 θ_n 与(11)式相同,且

$$\rho^2=\frac{q(1-q)}{q(1-q)-q^2g_{(r)}^{\tau}(\theta_0)V^{-1}(x_0)g_{(r)}(\theta_0)}\delta^2.$$

若取定 α, $0<\alpha<1$,又设 c_α 满足 $P\{\chi_1{}^2>c_\alpha\}=\alpha$,那么在不含附加信息和含附加信息两种情况下,由定理1和定理2可确定,θ_0 的渐近置信系数为 $1-\alpha$ 的渐近置信域分别为 $\{\theta\mid -2\log l_1(\theta)\leqslant c_\alpha\}$ 和 $\{\theta\mid -2\log l_2(\theta)\leqslant c_\alpha\}$.

根据文献[15]237页中的说明,再类似文献[16]和文献[17]来考虑检验问题:$H_0:\theta=\theta_0$, $H_1:\theta=\theta_n$. 由定理1和定理2确定的检验规则的检验水平为 α 的渐近否定域,分别为 $\{(X_i,Y_i), i\in s_r\mid -2\log l_1(\theta_0)>c_\alpha\}$ 和 $\{(X_i,Y_i), i\in s_r\mid -2\log l_2(\theta_0)>c_\alpha\}$. 这两种检验规则在 $\theta=\theta_n$ 时的渐近功效分别为

$$\lim_{n\to\infty}P_{\theta_n}\{-2\log l_1(\theta_0)>c_\alpha\}=$$
$$\lim_{n\to\infty}P_{\theta_n}\{-2\log l_1(\theta_n-\Delta_n\delta)>c_\alpha\}=P\{\chi_1^2(\delta^2)>c_\alpha\},$$
$$\lim_{n\to\infty}P_{\theta_n}\{-2\log l_2(\theta_0)>c_\alpha\}=$$
$$\lim_{n\to\infty}P_{\theta_n}\{-2\log l_2(\theta_n-\Delta_n\delta)>c_\alpha\}=P\{\chi_1^2(\rho^2)>c_\alpha\},$$

其中

$$\Delta_n=(nh^d)^{-1/2}\cdot\frac{(q(1-q)p(x_0)f(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u)^{1/2}}{p(x_0)f(x_0)f(\theta_0\mid x_0)}.$$

由 $\rho^2\leqslant\delta^2$ 及非卡方分布的性质知,$P\{\chi_1{}^2(\rho^2)>c_\alpha\}\geqslant P\{\chi_1{}^2(\delta^2)>c_\alpha\}$,故若 $g_{(r)}(\theta_0)\neq 0$,则 $P\{\chi_1{}^2(\rho^2)>c_\alpha\}>P\{\chi_1{}^2(\delta^2)>c_\alpha\}$. 此时,在同一检验水平下,由定理2确定的检验规则的渐近功效高于定理1确定的检验规则的渐近功效. 由 $g_{(r)}(\theta_0)$ 和 $V^{-1}(x_0)$ 的结构知,对上述检验问题,由定理2所确定的检验规则的渐近功效随信息量的增加而非降. 另外,由本文的结论还可以推出文献[6]的主要结果.

3 数值模拟

取不同的响应概率,在有限样本下模拟第2节中给出的基于(9)式和(12)式得到的经验似然置信区间. 为此,设 $(X,Y)\sim N(0,0,1,1,0.5)$,取核函数为 $K(u)=15/16(1-u^2)^2I(|u|\leqslant 1)$,窗宽 $h=n^{-1/4}$, $d=1$, $x_0=0.5$, $q=0.6$,且都满足定理的条件. 假设附加信息为 $E(Y\mid X=0.5)=0.025$,即 $g(y)=y-0.025$.

在 MAR 和 MCAR 缺失机制下考虑两种缺失情形:

情形1 (MAR):$p_1(x)=P(\delta=1\mid X=x)=0.8+0.2(|x-1|)$,当 $|x-1|\leqslant 1$ 时;其它情况下,$p_1(x)=0.9$.

情形2 (MCAR):对于所有 x, $p_2(x)=P(\delta=1\mid X=x)=0.6$.

对上述两种情形，产生1000个不完全样本$\{X_i,Y_i,\delta_i,i=1,\cdots,n\}$，其中$n=60,120$和200. 取$1-\alpha=0.95$，在不完全样本下分别计算出$\theta_0$的经验似然置信区间的覆盖概率($CP$)和平均区间长度($AL$)(表1).

表1 θ_0在各种缺失机制和样本容量下的经验似然置信区间的覆盖概率(CP)和平均区间长度(AL)

$p(x)$	n	CP		AL	
		$\theta_0^{(\mathrm{I})}$	$\theta_0^{(\mathrm{II})}$	$\theta_0^{(\mathrm{I})}$	$\theta_0^{(\mathrm{II})}$
$p_1(x)$	60	0.843	0.912	0.8322	0.6201
	120	0.851	0.934	0.6315	0.5382
	200	0.889	0.942	0.5232	0.3879
$p_2(x)$	60	0.832	0.895	1.0165	0.7815
	120	0.846	0.903	0.9023	0.6101
	200	0.875	0.917	0.8731	0.5023

表1结果表明，经验似然置信区间的覆盖概率(CP)随着样本容量n的增加而逐步接近名义覆盖水平0.95，且平均区间长度随着样本容量的增加逐渐变小.

4 定理的证明

引理1 设条件(i)～(v)满足，则当$\theta=\theta_0+O_p((nh^d)^{-1/2})$时，

$$\lambda(\theta)=O_p((nh^d)^{-1/2}+h^{r_0}),\tag{13}$$

$$\lambda(\theta)=[(nh^d)^{-1}\sum_{i=1}^n\omega_{1i}^2(\theta)]^{-1}(nh^d)^{-1}\sum_{i=1}^n\omega_{1i}(\theta)+O_p(((nh^d)^{-1/2}+h^{r_0})^2),\tag{14}$$

其中

$$\omega_{1i}(\theta)=\delta_iK_1(\frac{x_0-X_i}{h})(G(\theta-Y_i)-q),i=1,\cdots,n.\tag{15}$$

证明 由(5)式知

$$\sum_{i=1}^n\omega_{1i}(\theta)-\lambda(\theta)\sum_{i=1}^n\frac{\omega_{1i}^2(\theta)}{1+\lambda(\theta)\omega_{1i}(\theta)}=0,\tag{16}$$

且易证

$$(nh^d)^{-1}\sum_{i=1}^n\omega_{1i}(\theta)=(nh^d)^{-1}\sum_{i=1}^n\omega_{1i}(\theta_0)+O_p((nh^d)^{-1/2}),\tag{17}$$

故

$$(nh^d)^{-1}\sum_{i=1}^n\omega_{1i}(\theta_0)+O_p((nh^d)^{-1/2})=\lambda(\theta)(nh^d)^{-1}\sum_{i=1}^n\{\delta_iK_1{}^2(\frac{x_0-X_i}{h})[(G(\theta_0-Y_i)-q)^2+o_p(1)]\}\div[1+\lambda(\theta)\omega_{1i}(\theta)].\tag{18}$$

由(1)式，条件(i)～(vi)和Taylor展式知

$$(nh^d)^{-1}\sum_{i=1}^nE\omega_{1i}(\theta_0)=\int_{R^d}K_1(u)(\int(F_1(x_0-hu,\theta_0-hv)-F_1(x_0,\theta_0))K_2(v)\mathrm{d}v)p(x_0-hu)f(x_0-hu)\mathrm{d}u=O(h^{r_0}).\tag{19}$$

又由(15)式和(19)式知

$$(nh^d)^{-2}\sum_{i=1}^n\mathrm{Var}\omega_{1i}(\theta_0)=(nh^d)^{-1}(q(1-q)p(x_0)f(x_0)\int_{R^d}K_1{}^2(u)\mathrm{d}u+o(1)),$$

再由中心极限定理得

$$\sqrt{nh^d}\sum_{i=1}^n[\omega_{1i}(\theta_0)-E\omega_{1i}(\theta_0)]\xrightarrow{d}N(0,q(1-q)p(x_0)f(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u).\tag{20}$$

类似可证

$$(nh^d)^{-1}\sum_{i=1}^n\omega_{1i}^2(\theta_0)=q(1-q)p(x_0)f(x_0)\int_{R^d}K_1{}^2(u)\mathrm{d}u+o_p(1).\tag{21}$$

令$Z_{n1}=\max\limits_{1\leqslant i\leqslant n}|\omega_{1i}(\theta)|$，则由(17)～(20)式知

$$\frac{|\lambda(\theta)|}{1+|\lambda(\theta)|Z_{n1}}(q(1-q)p(x_0)f(x_0)\int_{R^d}K_1{}^2(u)\mathrm{d}u+o_p(1))=O_p((nh^d)^{-1/2}+h^{r_0}),$$

由此知(13)式成立. 最后由(5)式，(13)式和(16)式知

$$\sum_{i=1}^n\omega_{1i}(\theta)=\lambda(\theta)\sum_{i=1}^n\omega_{1i}^2(\theta)+O_p(nh^d((nh^d)^{-1/2}+h^{r_0})^2),$$

即(14)式成立，引理1证毕.

引理2 设条件(i)～(vi)满足，则

$$(nh^d)^{-1}\sum_{i=1}^n\omega_{2i}\omega_{2i}^{\tau}=\Sigma_{(1)}+o_p(1),\tag{22}$$

$$\eta_1=\Sigma_{(1)}^{-1}(nh^d)^{-1}\sum_{i=1}^n\omega_{2i}+o_p((nh^d)^{-1/2}),\tag{23}$$

$$\sqrt{nh^d}(nh^d)^{-1}\sum_{i=1}^n\omega_{2i}\xrightarrow{d}N(0,\Sigma_{(1)}).\tag{24}$$

其中

$$\omega_{2i}(\theta)=\delta_iK_1(\frac{x_0-X_i}{h})g_{(r)}^{\tau}(Y_i),\Sigma_{(1)}=p(x_0)f(x_0)V(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u.$$

证明 由于(22)式的证明类似(21)式. 在此只证(23)式和(24)式. 记$\overline{\omega}=(nh^d)^{-1}\sum_{i=1}^n\omega_{2i}$. 先证

(24)式.因为

$$E(\overline{\omega})=h^{-d}\int_{R^d}K(\frac{x_0-u}{h})p(u)M(u)f(u)\mathrm{d}u$$

$$=\int_{R^d}K(u)p(x_0-u)M(x_0-u)f(x_0-u)\mathrm{d}u,$$

由 $M(x_0)=0$,Taylor 展式和条件(iii)知

$$E(\overline{\omega})=O(h^{r_0}).\tag{25}$$

又对任意 $l\in R^r,l\neq 0$,记

$$s_n^2=\mathrm{Var}(l^\tau\overline{\omega}),\Gamma_n=s_n^{-3}(nh^d)^{-3}\sum_{i=1}^n E\mid l^\tau\omega_{2i}-E(l^\tau\omega_{2i})\mid^3.$$

结合条件(vi)知

$$s_n^2=E(l^\tau\overline{\omega})^2-(E(l^\tau\overline{\omega}))^2=(nh^d)^{-1}(l^\tau\cdot p(x_0)f(x_0)V(x_0)\int_{R^d}K^2(u)\mathrm{d}u\cdot l+o(1))+O(h^{2r_0})=(nh^d)^{-1}(l^\tau\cdot p(x_0)f(x_0)V(x_0)\int_{R^d}K^2(u)\mathrm{d}u\cdot l+o(1)).\tag{26}$$

又由条件(vi)知

$$\Gamma_n\leqslant Cs_n^{-3}(nh^d)^{-2}\to 0.\tag{27}$$

最后由中心极限定理知

$$l^\tau(\overline{\omega}-E\overline{\omega})/s_n\xrightarrow{d}N(0,1).\tag{28}$$

于是由 Crammer -Wold 定理和条件(vi)知,(24)式成立.

再证(23)式.令 $\eta_1=\lambda\theta$,其中 $\lambda\geqslant 0,\theta\in R^r$,$\|\theta\|=1$,并令 $Z_{n2}=\max\limits_{1\leqslant i\leqslant n}\mid\theta^\tau\omega_{2i}\mid$.由(6)式知

$$0=\|\frac{1}{nh^d}\sum_{i=1}^n\frac{\omega_{2i}}{1+\lambda\theta^\tau\omega_{2i}}\|\geqslant\frac{1}{nh^d}\mid\theta^\tau(\sum_{i=1}^n\omega_{2i}-\lambda\sum_{i=1}^n\frac{\omega_{2i}\theta^\tau\omega_{2i}}{1+\lambda\theta^\tau\omega_{2i}})\mid\geqslant\frac{\lambda}{nh^d}\theta^\tau\sum_{i=1}^n\frac{\omega_{2i}\theta^\tau\omega_{2i}}{1+\lambda\theta^\tau\omega_{2i}}\theta-\frac{1}{nh^d}\mid\theta^\tau\sum_{i=1}^n\omega_{2i}\mid\geqslant\frac{\lambda\theta^\tau Q_n\theta}{1+\lambda Z_{n2}}-\frac{1}{nh^d}\mid\theta^\tau\sum_{i=1}^n\omega_{2i}\mid,$$

其中

$$Q_n=\frac{1}{nh^d}\sum_{i=1}^n\omega_{2i}\omega_{2i}^\tau.$$

用 t 记 $p(x_0)f(x_0)V(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u$ 的最小特征根,则由大数定理知 $\theta^\tau Q_n\theta\geqslant t+o_p(1)$. 结合(4)式知

$$\frac{\lambda}{1+\lambda Z_{n2}}=O_p((nh^d)^{-1/2}).\tag{29}$$

又因为 $Z_{n2}=o(n^{1/s})$,a. s.,故由 $n^{s-2}h^{sd}\to\infty$ 知,$\lambda=O_p((nh^d)^{-1/2})$,从而

$$\eta_1=O_p((nh^d)^{-1/2}).\tag{30}$$

再令 $\gamma_i=\eta_1^\tau\omega_{2i}$,则

$$\max_{1\leqslant i\leqslant n}\mid\gamma_i\mid=O_p((nh^d)^{-1/2})o_p(n^{1/s})=o_p(1).$$

由(6)式知

$$0=\frac{1}{nh^d}\sum_{i=1}^n\omega_{2i}(1-\gamma_i+\frac{\gamma_i}{1+\gamma_i})=\overline{\omega}-Q_n\eta_1+\frac{1}{nh^d}\sum_{i=1}^n\frac{\omega_{2i}\gamma_i^2}{1+\gamma_i},$$

而又由于

$$\mid\frac{1}{nh^d}\sum_{i=1}^n\frac{\omega_{2i}r_i^2}{1+r_i}\mid=\frac{1}{nh^d}\sum_{i=1}^n\|\omega_{2i}\|^3\|\eta_1\|^2\|1+r_i\|^{-1}=o_p((nh^d)^{-1/2}),$$

故

$$\eta_1=Q_n^{-1}\overline{\omega}+o_p((nh^d)^{-1/2}).$$

再由大数定理知(23)式成立,引理 2 证毕.

引理 3 设条件(i)～(iv)及(v)～(vii)满足,则

$$(nh^d)^{-1}\sum_{i=1}^n\omega_{3i}(\theta)\omega_{3i}^\tau(\theta)=\Sigma_{(2)}+o_p(1),\tag{31}$$

$$\eta_2(\theta)=\Sigma_{(2)}^{-1}(nh^d)^{-1}\sum_{i=1}^n\omega_{3i}(\theta)+o_p((nh^d)^{-1/2}),\tag{32}$$

$$\sqrt{nh^d}(nh^d)^{-1}\sum_{i=1}^n\omega_{3i}(\theta)\xrightarrow{d}N(\mu,\Sigma_{(2)}).\tag{33}$$

其中

$$\mu=(0^\tau,c_1p(x_0)f(x_0)f(\theta_0\mid x_0))^\tau,$$

$$c_1=\frac{(q(1-q)p(x_0)f(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u)^{1/2}}{p(x_0)f(x_0)f(\theta_0\mid x_0)}\delta,$$

$$\omega_{3i}(\theta)=\delta_iK_1(\frac{x_0-X_i}{h})(g_{(r)}^\tau(Y_i),G(\theta_0-Y_i)-q)^\tau,$$

$$\Sigma_{(2)}=\begin{pmatrix}V(x_0) & qg_{(r)}(\theta_0)\\ qg_{(r)}^\tau(\theta_0) & q(1-q)\end{pmatrix}p(x_0)f(x_0)\int_{R^d}K_1{}^2(u)\mathrm{d}u.$$

证明 由条件(vii)知 $\Sigma_{(2)}$ 正定.由(21)式和(22)式类似可知(31)式成立.由(14)式和(23)式可知(32)式成立.由(20)式和(24)式类似可知(33)式成立.

引理 4 设 $A_{p\times p}$ 对称,$X\sim N(\mu_0,\Sigma_{p\times p})$,$\Sigma>0$,$\Sigma A$ 为幂等阵,且 $\mathrm{Rank}(A)=k$,则 $X^\tau AX\sim\chi_k^2(\mu_0^\tau A\mu_0)$.

引理 4 证明参考文献[18]的 Corollary2. 11. 2.

定理 1 的证明 因为(9)式为(10)式的特殊情

况,故只需证明(10)式.设 c_1 与引理3中的相同,当 $\theta=\theta_0+(nh^d)^{-1/2}c_1$ 时,由 Taylor 展式知

$$(nh^d)^{-1}\sum_{i=1}^{n}\omega_{1i}(\theta)=(nh^d)^{-1}\sum_{i=1}^{n}\omega_{1i}(\theta_0)+c_1(nh^d)^{-1/2}(nh^{d+1})^{-1}\sum_{i=1}^{n}\delta_iK_1(\frac{x_0-X_i}{h})K_2(\frac{\theta_0-Y_i}{h})+O_p((nh^{d+1})^{-1}). \quad (34)$$

易证

$$(nh^{d+1})^{-1}\sum_{i=1}^{n}\delta_iK_1(\frac{x_0-X_i}{h})K_2(\frac{\theta_0-Y_i}{h})=p(x_0)f(x_0)f(\theta_0\mid x_0)+o_p(1). \quad (35)$$

由(19)式,(20)式,(34)式,(35)式和定理1的条件知

$$\sqrt{nh^d}(nh^d)^{-1}\sum_{i=1}^{n}\omega_{1i}(\theta)\xrightarrow{d}N(c_1p(x_0)f(x_0)f(\theta_0\mid x_0),q(1-q)p(x_0)f(x_0)\int_{R^d}K_1^2(u)\mathrm{d}u). \quad (36)$$

再由 Taylor 展式知

$$-2\log l_1(\theta)=2\lambda(\theta)\sum_{i=1}^{n}\omega_{1i}(\theta)-\lambda^2(\theta)\sum_{i=1}^{n}\omega_{1i}{}^2(\theta)+O_p(nh^d((nh^d)^{-1/2}+h^{r_0})^3)=\lambda(\theta)\sum_{i=1}^{n}\omega_{1i}(\theta)+O_p(nh^d((nh^d)^{-1/2}+h^{r_0})^3)=[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{1i}^2(\theta)]^{-1}\times nh^d[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{1i}(\theta)]^2+O_p(nh^d((nh^d)^{-1/2}+h^{r_0})^3). \quad (37)$$

最后由(21)式,(36)式和(37)式知,(10)式成立,定理1证毕.

定理2的证明 由引理2和引理3并类似(37)式的证明知

$$-2\log l_2(\theta)=nh^d[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)]^\tau\times[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)\omega_{3i}^\tau(\theta)]^{-1}\times[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)]-nh^d[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{2i}]^\tau[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{2i}\omega_{2i}{}^\tau]^{-1}\times[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{2i}]+o_p(1)=nh^d[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)]^\tau\Sigma_{(2)}^{-1}[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)]-nh^d[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{2i}]^\tau\Sigma_{(1)}^{-1}[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{2i}]+o_p(1)=nh^d[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)]^\tau\times B[(nh^d)^{-1}\sum_{i=1}^{n}\omega_{3i}(\theta)]+o_p(1), \quad (38)$$

其中

$$B=(\Sigma_{(2)}^{-1}-\begin{pmatrix}\Sigma_{(1)}^{-1} & 0\\ 0 & 0\end{pmatrix}).$$

注意到 $B\Sigma_{(2)}$ 为幂等阵且秩为1,又

$$\mu^\tau B\mu=c_1^2p^2(x_0)f^2(x_0)f^2(\theta_0\mid x_0)c_2^{-1}(p(x_0)f(x_0)\int_R^d K_1^2(u)\mathrm{d}u)^{-1},$$

其中

$$c_2=q(1-q)-q^2g_{(r)}^\tau(\theta_0)V^{-1}(x_0)g_{(r)}(\theta_0),$$

故

$$\mu^\tau B\mu=\frac{q(1-q)}{q(1-q)-q^2g_{(r)}^\tau(\theta_0)V^{-1}(x_0)g_{(r)}(\theta_0)}\delta^2.$$

结合(38)式,引理3和引理4知定理2成立.

参考文献:

[1] 郑忠国.条件中位数的最近邻估计和它的 Bootstrap 统计量的渐近性质[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]. J Multivarate Anal,1996,59:206-216.

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

[5] 秦永松.条件分位数和条件密度的经验似然置信区间[J].数学年刊,1999,21A(3):333-342.

[6] 秦永松,苏淳.条件分位数的经验似然置信区间[J].数学年刊,2000,21A(2):231-240.

[7] Owen A B. Empirical likelihood ratio confidence intervals for a single functional[J]. Biomertrika,1988,75:237-249.

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

[9] Kolaczyk E D. Empirical likelihood for linear models [J]. Statistica Sinica,1994,4:199-218.

[10] Wang Q H,Jing B Y. Empirical likelihood for partly linear modles with fixed design[J]. Statist Probab Lett,1999,41(4):425-433.

(下转第277页)

a_i 的确定可以由专家意见得到，$0 < a_i < 1$，$\sum_{i=1}^{4} a_i \leqslant 1$.

在实践中，由于自身因素和非自身因素的不同，则 a_i 的赋值也不相同. 例如对 a_i 确定时，如果是采用直接借鉴他人的科研成果的方式，则 a_i 的值相对可能低一点，如果采用自主创新的方式，则 a_i 的值相对可以高一点. 分析还可以看出，由公式(2)确定的大学生数学应用能力培养指数的值域为[0,100]，由公式(3)确定的大学生数学应用能力培养指数的值域也为[0,100]. 指数越大，说明数学应用能力培养越好.

3　模型应用效果

将本文建立的模型在我校进行实践. 首先给老师、学生各发放100份问卷调查，然后对问卷结果统计分析得出各要素的指标值 z_{ij}，取 $r_{ij}=\frac{1}{4}$，由公式(1)计算得 N_i，取 $r_i=\frac{1}{4}$，再把 r_i 和 N_i 代入公式(2)，得出大学生数学应用能力培养的评价指数 N，最后将评价指数 N 与我校实际情况进行比较. 评价结果符合我校实际情况.

参考文献：

[1]　刘坚. 基于多元统计法的综合素质评价模型研究[J]. 曲阜师范大学学报:自然科学版，2005(3):25-28.
[2]　曲国禹，刘学铭. 对建立企业技术创新能力评价指标体系的探讨[J]. 辽宁工学院学报:社会科学版，1999(1):81-84.
[3]　林国耀. 大学生学习能力的量表编制与现状测查研究[D]. 福建:福建师范大学，2006.
[4]　魏宗舒. 概率论与数理统计教程[M]. 北京:高等教育出版社，2008.
[5]　曹弋编. MATLAB 实用教程[M]. 北京:电子工业出版社，2007.
[6]　陈光亭，裘哲勇. 数学建模[M]. 北京:高等教育出版社，2010.
[7]　姜启源，谢金星，叶俊. 数学模型[M]. 北京:高等教育出版社，2003.

(责任编辑:尹　闯)

(上接第274页)

[11]　Zhong Bob，Rao J N K. Empiricallikelihood inference under stratified random sampling usingauxiliary populaion information[J]. Biometrika，2000，87:929-938.
[12]　Kitamura Y. Asymptotic optimality of empiricallikelihood for testing moment restrictions[J]. Econometrica，2001，69:1661-1672.
[13]　范承华. 缺失数据半参数回归分析[D]. 北京:北京工业大学，2007.
[14]　Chen S X，Hall P. Smoothed empirical likelihood confidence intervals[J]. Ann Statist，1993，21:1166-1181.
[15]　Sen P K，Singer J M. Large sample methods in statistics-an introduction with application[M]. New York:Chapman Hall，1993.
[16]　Chen S X. Comparing empirical likelihood and Bootstrap hypothesis test[J]. Multivariate Analysis，1994，51:277-293.
[17]　Zhang B. Empirical likelihood confidence intervals in the presence of auxiliary informatioin[J]. Statist Proba Lett，1997，32:87-97.
[18]　Srivastava M，Khatri C G. An introduction to multivariate statistics[M]. New York:North-Holland，1979.

(责任编辑:尹　闯)