固定设计下半参数回归模型中的随机加权逼近

薛留根, 田 萍
(北京工业大学 应用数理学院, 北京 100022)

摘 要: 为了模拟参数估计的误差分布,研究了固定设计下半参数回归模型中参数估计的误差分布的随机加权逼近问题,构造了参数估计的随机加权统计量,在适当的条件下,证明了用随机加权统计量的分布去逼近原估计量的误差分布的精度可达到 $o(n^{-1/2})$.

关键词: 半参数回归模型; 随机加权逼近; 参数估计
中图分类号: O 212.7　　**文献标识码**: A　　**文章编号**: 0254-0037(2004)02-0247-05

考虑半参数回归模型

$$Y_i = x_i\beta + g(t_i) + e_i,\quad i = 1,\cdots,n \tag{1}$$

其中:$\{(x_i,t_i),1\leqslant i\leqslant n\}$是取值于 $R^1\times[0,1]$ 上的已知设计点列;β 是一维未知参数;g 是定义在闭区间 [0,1] 上的未知函数;$\{e_i,1\leqslant i\leqslant n\}$ 是相互独立的随机误差序列,且 $Ee_i=0, Ee_i^2=\sigma_i^2>0$.

关于对模型(1)中估计的大样本性质已有很多研究,得到了令人满意的结果. 文献[1-5]分别在不同的条件下研究了模型(1)中估计的渐近正态性,强相合性及其收敛速度. 文献[6]给出了β的估计量的分布的渐近展开,由此可得正态逼近的逼近精度. 文献[7,8]研究了模型(1)中随机加权估计和小波估计的强逼近速度. 本文在较一般的情况下研究模型(1)中参数β的估计$\hat\beta_n$之误差的分布的随机加权逼近问题. 用随机加权法构造了$\hat\beta_n-E\hat\beta_n$的随机加权估计 H_n,在适当的条件下证明了用$(H_n-E^*H_n)/(\mathrm{Var}^*(H_n))^{1/2}$的分布去逼近$(\hat\beta_n-E\hat\beta_n)/(\mathrm{Var}(\hat\beta_n))^{1/2}$或$(\hat\beta_n-\beta)/(\mathrm{Var}(\hat\beta_n))^{1/2}$的分布的精度可达到 $o(n^{-1/2})$,其中 E^*, Var^*分别表示在给定 $Y_1,\cdots,Y_n$下的条件期望和条件方差. 该逼近速度与线性模型中参数β的估计的随机加权逼近速度相当(参见文献[9]). 同时指出,本文结果不要求误差序列$e_1,\cdots,e_n$同分布.

1 主要结果

设$\{W_{ni}(t)=W_{ni}(t;t_1,\cdots,t_n),1\leqslant i\leqslant n\}$是一列定义于闭区间[0,1]上的非负权函数. 记 $\tilde Y_i = Y_i - \sum_{j=1}^n W_{nj}(t_i)Y_j$, $\tilde x_i = x_i - \sum_{j=1}^n W_{nj}(t_i)x_j$, $1\leqslant i\leqslant n$, $\tilde S_n^2=\sum_{i=1}^n\tilde x_i^2$,则利用通常的非参数权函数法再结合最小二乘法给出$\beta$的估计量为

$$\hat\beta_n = \tilde S_n^{-2}\sum_{i=1}^n \tilde x_i\tilde Y_i \tag{2}$$

记 $Q_n=(\hat\beta_n-E\hat\beta_n)/(\mathrm{Var}\hat\beta_n)^{1/2}$, $T_n=(\hat\beta_n-\beta)/(\mathrm{Var}\hat\beta_n)^{1/2}$. 又记 $u_{ni}=\tilde S_n^{-2}(\tilde x_i-\sum_{j=1}^n W_{ni}(t_j)\tilde x_j)$, $h_{ni}=u_{ni}/(\sum_{j=1}^n u_{nj}^2)^{1/2}$,则有 $Q_n=\sum_{i=1}^n h_{ni}e_i/(\sum_{j=1}^n h_{nj}^2\sigma_j^2)^{1/2}$. 为了模拟 Q_n 和 T_n 的分布,采用随机加权统计量

$$H_n=\sum_{i=1}^n u_{ni}\hat e_i\xi_i$$

其中$\xi_1,\cdots,\xi_n$ 为随机变量,其共同分布为$\Gamma(2,4)$,即ξ_1的密度函数为$f(x)=[2^4/\Gamma(4)]x^3e^{-2x}I(x>$

收稿日期: 2003-10-16.
基金项目: 北京市自然科学基金资助项目(1042002).
作者简介: 薛留根(1957-),男,河南鄢陵县人,教授,博士生导师.

$0)$，$\{\xi_i, 1 \leqslant i \leqslant n\}$与$\{Y_i, 1 \leqslant i \leqslant n\}$相互独立．$\hat{e}_i$为残差$e_i$的估计．即$\hat{e}_i = \tilde{Y}_i - \tilde{x}_i \hat{\beta}_n, i = 1, \cdots, n$.

首先列出如下条件

A1: (i)存在$\sigma_0 > 0$，使$Ee_i^2 = \sigma_i^2 \geqslant \sigma_0^2, i = 1, 2, \cdots$

(ii)e_i的特征函数$v_i(t)$一致满足 Cramer 条件，即$\forall \varepsilon > 0$，存在$0 < \eta < 1$，使

$$\sup_{|t| \geqslant \varepsilon} |v_i(t)| < 1 - \eta, i = 1, 2, \cdots$$

A2: 存在$[0,1]$上某函数$h(t)$，对$1 \leqslant i \leqslant n$，有$x_i = h(t_i) + v_i$，其中

$$\frac{1}{n} \sum_{i=1}^{n} v_i \to 0, \quad \liminf_{n \to \infty} n^{-1} \sum_{i=1}^{n} v_i^2 \geqslant A > 0$$

$$\frac{1}{n} \sum_{i=1}^{n} |v_i|^3 = O(1), \quad \max_{1 \leqslant i \leqslant n} |v_i| = O(n^{1/6} (\log n)^{-\delta/6}), \quad \delta > 1$$

A3: (i)$h(t)$在$[0,1]$上连续；

(ii)$g(t)$在$[0,1]$上满足一阶 Lipschitz 条件．

A4: (i) $\max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^{n} W_{nj}(t_i) - 1 \right| = o(1)$；

(ii) $\max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} |W_{nj}(t_i)| I(|t_j - t_i| > a_n) = O(b_n)$；

(iii) $\max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} |W_{nj}(t_i)| = O(1)$； $\max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} |W_{ni}(t_j)| = O(1)$；

(iv) $\max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^{n} W_{nj}(t_i) v_j \right| = o(1)$； $\max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^{n} W_{ni}(t_j) v_j \right| = O(1)$；

(v) $\max_{1 \leqslant i, j \leqslant n} |W_{nj}(t_i)| = O(d_n)$.

这里a_n, b_n, d_n皆为收敛于 0 的正的常数列．

注 条件 A1 是对误差e_i的基本假设．条件 A2 和 A3 已被广泛采用[4,6]．条件 A4 给出的权函数是存在的[4,6]．

得到了如下结果：

定理 1 设 A1, A2, A3(i), A4(i)～(iv)成立，若$\sup_i E|e_i|^{3+\rho} < \infty, \rho > 0$，则

$$\sqrt{n} \sup_y \left| P(Q_n \leqslant y) - \Phi(y) + \frac{1}{6} \varphi(y)(y^2 - 1) \frac{\sum_{i=1}^{n} h_{ni}^3 E e_i^3}{(\sum_{j=1}^{n} h_{nj}^2 \sigma_j^2)^{3/2}} \right| \to 0$$

其中$\Phi(y)$，$\varphi(y)$分别为标准正态分布函数及密度函数．

定理 2 设条件 A1(i), A2, A3, A4 成立，若$\sup_i E e_i^6 < \infty$，则当$a_n = b_n = O(n^{-1/5})$，$d_n = O(n^{-1/5})$时，对几乎所有的样本序列，有

$$\sqrt{n} \sup_y \left| P^*(R_n \leqslant y) - \Phi(y) + \frac{1}{6} \varphi(y)(y^2 - 1) \frac{\sum_{i=1}^{n} h_{ni}^3 \hat{e}_i^3}{(\sum_{j=1}^{n} h_{nj}^2 \hat{e}_j^2)^{3/2}} \right| \to 0, \quad \text{a.s.}$$

其中$R_n = (H_n - E^* H_n) / (\mathrm{Var}^*(H_n))^{1/2} = \sum_{i=1}^{n} h_{ni} \hat{e}_i (\xi_i - 2) / (\sum_{j=1}^{n} h_{nj}^2 \hat{e}_j^2)^{1/2}$. P^*, E^*和Var^*分别表示在给定$Y_1, \cdots, Y_n$下的条件概率、条件期望和条件方差．

定理 3 在定理 2 的条件下，若还有 A1(ii)成立，则对几乎所有的样本序列，有

$$\sqrt{n} \sup_y |P(Q_n \leqslant y) - P^*(R_n \leqslant y)| \to 0, \quad \text{a.s.}$$

定理 4 在定理 3 的条件下，若还有

$$\sum_{i=1}^{n} \tilde{x}_i \left(g(t_i) - \sum_{j=1}^{n} W_{nj}(t_i) g(t_j) \right) = O(n(a_n^2 + b_n^2))$$

则当$a_n = b_n = o(n^{-1/2})$，$d_n = O(n^{-1/2} \log n)$时，对几乎所有的样本序列，有

$$\sqrt{n}\sup_{y}|P(T_n \leqslant y)-P^{*}(R_n \leqslant y)| \to 0,\ \text{a.s.}$$

2 定理的证明

为了证明本文第 1 节的 4 个定理，首先给出 2 个引理．引理的证明省略．

引理 1　设条件 A2，A3(i)，A4(i)～(iv)成立，则

$$\max_{1 \leqslant i \leqslant n}|h_{ni}| = O[n^{-1/3}(\log n)^{-\delta/6}],\ \delta > 1 \tag{3}$$

$$\sqrt{n}\sum_{i=1}^{n}|h_{ni}|^{3} = O(1) \tag{4}$$

引理 2　设定理 2 的条件成立，则

$$\sum_{i=1}^{n}h_{ni}^{2}(\hat{e}_i^{2}-Ee_i^{2}) \to 0,\ \text{a.s.} \tag{5}$$

$$\sqrt{n}\sum_{i=1}^{n}h_{ni}^{3}(\hat{e}_i^{3}-Ee_i^{3}) \to 0,\ \text{a.s.} \tag{6}$$

定理 1 的证明　利用文献[10]中关于独立和的分布的渐近展开定理 1，得

$$|P(Q_n \leqslant y)-\Phi(y)+\frac{1}{6}\varphi(y)(y^{2}-1)B_n^{-3}\sum_{i=1}^{n}h_{ni}^{3}Ee_i^{3}| \leqslant$$

$$C[(1+|y|)^{-3}B_n^{-3}\sum_{i=1}^{n}E|U_{ni}(y)|^{3}+(1+|y|)^{-4}B_n^{-4}\sum_{i=1}^{n}E|Z_{ni}(y)|^{4}]+$$

$$C\{(1+|y|)^{-4}n^{6}[\sup_{|t|\geqslant\delta_n}\frac{1}{n}\sum_{i=1}^{n}|v_i(th_{ni})|+\frac{1}{2n}]^{n}\} =: J_{n1}+J_{n2}+J_{n3} \tag{7}$$

式(7)中 $v_j(t)$是 e_j 的特征函数，并且

$$B_n^{2}=\sum_{i=1}^{n}h_{ni}^{2}\sigma_i^{2},\quad \delta_n=\frac{1}{12}B_n^{2}(\sum_{i=1}^{n}E|Y_{ni}|^{3})^{-1}$$

$$Y_{ni}=h_{ni}e_i I(|h_{ni}e_i| \leqslant B_n)$$

$$U_{ni}(y)=h_{ni}e_i I[|h_{ni}e_i| > B_n(1+|y|)]$$

$$Z_{ni}(y)=h_{ni}e_i I[|h_{ni}e_i| \leqslant B_n(1+|y|)]$$

下面估计式(7)右端各项．由条件 A1(i)及$\sum_{j=1}^{n}h_{nj}^{2}=1$，知

$$B_n^{2}=\sum_{i=1}^{n}h_{ni}^{2}\sigma_i^{2} \geqslant \min_{1\leqslant i\leqslant n}\sigma_j^{2} \geqslant \sigma_0 > 0 \tag{8}$$

因此，由引理 1，得

$$\sqrt{n}J_{n1} \leqslant C\sqrt{n}\sum_{i=1}^{n}E[|\frac{h_{ni}e_i}{B_n}|^{3}I(|\frac{h_{ni}e_i}{B_n}| \geqslant 1)] \leqslant$$

$$C\sup_{i}E|e_i|^{3+\rho}(\max_{1\leqslant i\leqslant n}|h_{ni}|)^{\rho}\sqrt{n}\sum_{i=1}^{n}|h_{ni}|^{3} \leqslant \sqrt{C}n^{-\rho/3} \to 0 \tag{9}$$

$$\sqrt{n}J_{n2} \leqslant C\sqrt{n}\sum_{i=1}^{n}E|h_{ni}e_i/B_n|^{3+\rho} \leqslant Cn^{-\rho/3} \to 0 \tag{10}$$

现估计 J_{n3}．由引理 1 的(4)知存在 $M>0$，使$\sqrt{n}\sum_{i=1}^{n}|h_{ni}|^{3} \leqslant M$．记 $M_1=\sigma_0^{2}/(12M\sup_i E|e_i|^{3})$，则

$$\delta_n=B_n^{2}[12\sum_{i=1}^{n}E|h_{ni}e_i|^{3}I(|h_{ni}e_i| \leqslant B_n)]^{-1} \geqslant \sigma_0^{2}/(12\sup_{i}E|e_i|^{3}\sum_{i=1}^{n}|h_{ni}|^{3}) \geqslant M_1\sqrt{n} \tag{11}$$

适当选取$\varepsilon_1 > 0$,使得$\varepsilon = M_1^{-1}\varepsilon_1 \in (0,1)$. 记 $N_n = \#\{j, \sqrt{n}\,|h_{nj}| \geqslant \varepsilon\}$. 其中记号$\#(A)$表示$A$中元素个数. 由引理1的(3),知存在$K > 0$,使$n^{1/3}\max\limits_{1\leqslant i\leqslant n}|h_{ni}| \leqslant \sqrt{K}$. 于是,由$\sum\limits_{i=1}^{n} h_{ni}^2 = 1$,可得

$$n = \sum_{i=1}^{n}(\sqrt{n}\,|h_{ni}|)^2 \leqslant N_n K n^{1/3} + (n - N_n)\varepsilon$$

因此

$$N_n \geqslant [(1-\varepsilon)n]/(Kn^{1/3} - \varepsilon) \geqslant K^{-1}(1-\varepsilon)n^{2/3} \tag{12}$$

由式(11)、(12),利用条件 A1(ii)及不等式 $1 + x \leqslant e^x$,得

$$\sqrt{n}\,J_{n3} \leqslant n^7\Big[\sup_{|t|\geqslant\delta_n}\frac{1}{n}\sum_{i=1}^{n}|v_i(h_{ni}t)| + \frac{1}{2n}\Big]^n \leqslant n^7\Big[\frac{1}{n}\sum_{i=1}^{n}\sup_{|t|\geqslant M_1\sqrt{n}|h_{ni}|}|v_i(t)| + \frac{1}{2n}\Big]^n \leqslant$$

$$n^7\Big[\frac{N_n}{n}(1-\eta) + \big(1 - \frac{N_n}{n}\big) + \frac{1}{2n}\Big]^n \leqslant n^7\Big[1 - K^{-1}\eta(1-\varepsilon)n^{-1/3} + \frac{1}{2n}\Big]^n \leqslant \sqrt{e}\,n^7 e^{-cn^{2/3}} \to 0 \tag{13}$$

将式(9)、(10)、(13)代入到式(7)即证得定理1.

定理2的证明 利用文献[10]中定理1关于$k = 3$的结果以及引理2的式(5)、(8),依照定理1的证明不难证得定理2. 证明省略.

定理3的证明 由定理1和定理2,只需证明

$$\sup_y|\varphi(y)(y^2-1)\Big[\frac{\sum_{i=1}^{n}h_{ni}^3\hat{e}_i^3}{(\sum_{i=1}^{n}h_{ni}^2\hat{e}_i^2)^{3/2}} - \frac{\sum_{i=1}^{n}h_{ni}^3 e_i^3}{(\sum_{i=1}^{n}h_{ni}^2\sigma_i^2)^{3/2}}\Big]| = o\Big(\frac{1}{\sqrt{n}}\Big),\ \text{a.s.} \tag{14}$$

注意到 $\sup\limits_y|\varphi(y)(y^2-1)| < \infty$,由引理2和式(8)可直接推出式(14).

定理4的证明 记

$$A_n^2 = \mathrm{Var}(\hat{\beta}_n) = \sum_{i=1}^{n}u_{ni}^2\sigma_i^2,\quad r_n = A_n^{-1}\tilde{S}_n^{-2}\sum_{i=1}^{n}\tilde{x}_i\Big[g(t_i) - \sum_{i=1}^{n}W_{nj}(t_i)g(t_j)\Big]$$

由于 $P(T_n \leqslant y) = P(Q_n \leqslant y - r_n)$. 因此,由定理1得

$$\sqrt{n}\,\sup_y|P(T_n \leqslant y) - \Phi(y - r_n) + \frac{1}{6}\varphi(y - r_n)[(y - r_n)^2 - 1]\frac{\sum_{i=1}^{n}h_{ni}^3 Ee_i^3}{(\sum_{j=1}^{n}h_{nj}^2\sigma_j^2)^{3/2}}| \to 0$$

由条件 A1(i)和 A4,并注意到 $\tilde{S}_n^{-2}\sum_{i=1}^{n}u_{ni}^2 \geqslant 1/2$ 和 $n\tilde{S}_n^{-2} = O(1)$,容易得到

$$|r_n| \leqslant C\tilde{S}_n^{-1}\Big|\sum_{i=1}^{n}\tilde{x}_i\Big[g(t_i) - \sum_{j=1}^{n}W_{nj}(t_i)g(t_j)\Big]\Big| = o\Big(\frac{1}{\sqrt{n}}\Big)$$

利用微分中值定理不难证得

$$\sup_y|\Phi(y - r_n) - \Phi(y)| \leqslant \sup_y|y|\varphi(y)|r_n| = o\Big(\frac{1}{\sqrt{n}}\Big)$$

$$\sup_y|\varphi(y - r_n)[(y - r_n)^2 - 1] - \varphi(y)(y^2 - 1)| \leqslant \sup_y[|3y - y^3|\varphi(y)]|r_n| = o\Big(\frac{1}{\sqrt{n}}\Big)$$

综上所述,可得

$$\sqrt{n}\,\sup_y|P(T_n \leqslant y) - \Phi(y) + \frac{1}{6}\varphi(y)(y^2 - 1)\sum_{i=1}^{n}h_{ni}^3 Ee_i^3/\Big(\sum_{j=1}^{n}h_{nj}^2\sigma_j^2\Big)^{3/2}| \to 0 \tag{15}$$

故由式(14)、(15)及定理2即证得定理4.

参考文献:

[1] ENGLE R, GRANGER C, RICE J, et al. Nonparametric estimates of the relation between weather and electricity sales[J]. J Amer Statist Assoc, 1986, 81: 310-320.

[2] CHEN H. Convergence rates for parametric components in a partly linear model[J]. Ann Statist, 1988, 16(1): 136-146.

[3] SPECKMAN P. Kernel smoothing in partial linear models[J]. J Roy Statist Soc, 1988, 50B: 413-436.

[4] 高集体，陈希孺，赵林城．部分线性模型中估计的渐近正态性 [J]．数学学报，1994，37(2)：256-268.
GAO Ji-ti, CHEN Xi-ru, ZHAO Lin-cheng. Asymptotic normality in partially linear model[J]. Acta Mathematica Sinica, 1994, 37(2): 256-268. (in Chinese)

[5] 陈明华．固定设计下半参数回归模型估计的收敛速度 [J]．应用概率统计，1998，14(5)：149-158.
CHEN Ming-hua. Convergence rates of the estimations in semiparametric regression model under fixed design[J]. J Appl Prob Statist, 1998, 14(5): 149-158. (in Chinese)

[6] 石坚．部分线性模型中的 Edgeworth 展开 [J]．数学学报，1998，41(4)：683-686.
SHI Jian. Edgeworth expansion in partially linear model[J]. Acta Mathematica Sinica, 1998, 41(4): 683-686. (in Chinese)

[7] 薛留根．半参数回归模型中随机加权 *M* 估计的强逼近 [J]．应用数学学报，2002，25(4)：591-603.
XUE Liu-gen. Strong approximation of random weighting *M*-estimates in semiparametric regression mode[J]. Acta Mathematicae Applicatae Sinica, 2002, 25(4): 591-603. (in Chinese)

[8] 薛留根．半参数回归模型中小波估计的随机加权逼近速度 [J]．应用数学学报，2003，26(1)：11-25.
XUE Liu-gen. Rates of random weighting approximation of wavelet estimates in semiparametric regression model[J]. Acta Mathematicae Applicatae Sinica, 2003, 26(1): 11-25. (in Chinese)

[9] 郑忠国，涂冬生．线性模型中的随机加权逼近 [J]．中国科学(A 辑)，1988，18(6)：561-575.
ZHENG Zhong-guo, TU Dong-sheng. Reandom weighting asymptotic in linear models[J]. Science in China(Series A), 1988, 18(6): 561-575. (in Chinese)

[10] 白志东，赵林城．独立随机变量之和的分布函数的渐近展开 [J]．中国科学(A 辑)，1985，15(8)：677-679.
BAI Zhi-dong, ZHAO Lin-cheng. Edgeworth expansion of the distribution function of sum of independent random variable[J]. Science in China(Series A), 1985, 15(8): 677-679. (in Chinese)

Random Weighted Approximation in Semiparametric Regression Model Under Fixed Design

XUE Liu-gen, TIAN Ping

(College of Applied Sciences, Beijing University of Technology, Beijing 100022, China)

Abstract: To simulate the error distribution of parametric estimation, the random weighted approximation of the error distribution of parametric estimation in the semiparametric regression model under fixed design is studied and the random weighted statistic is structured by means of random weighted method. Under proper conditions the distribution of the random weighted statistic is used to approximate the error distribution of the preestimation, the accuracy of which can reach $o(n^{-1/2})$.

Key words: semiparametric regression model; random weighted approximation; parametric estimation