

固定设计下混合序列回归函数的小波估计*

柴根象, 刘元金

(同济大学应用数学系, 上海 200092)

摘 要: 考虑非参数回归模型 $Y_i = r(t_i) + \epsilon_i$, 其中 t_i 为固定设计点列, (ϵ_i) 为 ρ 混合相依随机变量. 本文用小波方法建立 $r(t)$ 的估计 $\hat{r}(t)$, 并讨论了它的渐近无偏性, 弱收敛, 强收敛, 及其收敛速度.

关键词: 混合序列; 回归函数; 小波估计; 弱收敛; 强收敛; 收敛速度.

分类号: AMS(1991) 62G05, 62E20/CLC O212.7

文献标识码: A

文章编号: 1000-341X(2001)04-0555-06

1 引 言

考虑非参数回归模型 $Y_i = r(t_i) + \epsilon_i, i = 1, \dots, n$. t_i 是固定设计点列, 不失一般性设有 $0 \leq t_1 \leq \dots \leq t_n \leq 1$. 当 (ϵ_i) 是 i. i. d 随机变量时, 回归函数 $r(\cdot)$ 的一些大样本性质已被研究的比较透彻, 当 ϵ_i 是混合相依随机变量时, 柴根象^[1] 曾用核估计方法讨论了回归函数的一些大样本性质并得到了一系列满意的结果.

近几年来小波方法在统计中的应用日益广泛而且显示出很多优点, 为此, Antoniadias^[2] 等用小波方法讨论了当 ϵ_i 是 i. i. d 随机变量时回归函数 $r(\cdot)$ 的一些大样本性质. 本文用小波方法, 沿用 Antoniadias 的估计 $\hat{r}(t) = \sum_{i=1}^n Y_i \int_{A_i} E_m(t, s) ds, A_i = [s_{i-1}, s_i)$, 讨论当 (ϵ_i) 是混合相依序列时回归函数 $r(\cdot)$ 的一些大样本性质, 而且得到的结果与 ϵ_i 是 i. i. d 随机变量的情形很接近. 本文在第二节给出一些引理, 第三节给出主要结果及其证明.

2 引理及假定

在本节给出一些基本定义、假设及引理.

定义 设 (x_i) 为 r. v. 序列, 记 $F_n^m = \sigma(x_i, n \leq i < m)$, 定义混合系数

$$\rho(n) = \sup_m \sup_{(x \in L_2(F_n^m), y \in L_2(F_{n+m}^m))} \frac{|E(x - Ex)E(y - Ey)|}{\sqrt{\text{Var}(x)\text{Var}(y)}}.$$

* 收稿日期: 1998-07-13

基金项目: 国家自然科学基金资助项目(19571061)

作者简介: 柴根象(1939-), 男, 浙江宁海人, 教授, 博导.

如果 $n \rightarrow \infty$ 时 $\rho(n) \rightarrow 0$, 则称序列 $\{x_i\}$ 为 ρ 混合的.

如果 $\sup_i P(|x_i| \geq t) \leq P(|x| \geq t), \forall t > 0$, 称 $\{x_i, i \geq 1\}$ 为随机变量 x 所界.

定义 小波再生核 $E_m(t, s) = 2^m \sum_{k \in \mathbb{Z}} \delta(2^m t - k) \delta(2^m s - k)$ 并假设

(i) 刻度函数 $\delta(\cdot)$ 为 q 阶正则且具有紧支撑满足 Lipschitz 条件;

(ii) $\max_i |s_i - s_{i-1}| = O(n^{-1})$;

(iii) 某个函数 $l(\cdot)$ 满足 Lipschitz 条件且使 $l(n) = \max_i |s_i - s_{i-1} - \frac{l(s_i)}{n}| = o(n^{-1})$;

(iv) 回归函数 $r(\cdot)$ 满足 γ 阶 Lipschitz 条件且 $r(\cdot) \in H^v, v > \frac{1}{2}$.

下面给出主要结果要用到的一些引理.

引理 1^[2] 若刻度函数 q 阶正则且具有紧支撑, 则 $\sup_{x, m} \int_0^1 |E_m(x, y)| dy < \infty$.

引理 2^[2] $r(\cdot)$ 满足 γ 阶 Lipschitz 条件, $r \in H^v, v > 1/2$, 则

$$\int_0^1 E_m(t, s) r(s) ds = r(t) + O(\eta_m),$$

$$\text{其中 } \eta_m = \begin{cases} (\frac{1}{2^m})^{v-1/2}, & \frac{1}{2} < v < \frac{3}{2}, \\ \frac{\sqrt{m}}{2^m}, & v = \frac{3}{2}, \\ \frac{1}{2^m}, & v > \frac{3}{2}. \end{cases}$$

引理 3^[3] 若 $Ex_i = 0, \sup_i E|x_i|^p < \infty, p > 1, \sum_{i=1}^{\infty} \rho(i) < \infty$ 且存在 $\theta > \frac{1}{p}, k > 0$ 使

$$\sum_{i=1}^n |a_{ni}| \leq k, |a_{ni}| \leq kn^{-\theta}, \text{ 则 } \sum_{i=1}^n a_{ni} x_i \xrightarrow{a.s.} 0.$$

引理 4 $|\int_{\Lambda_i} E_m(t, s) ds| = O(\frac{2^m}{n})$.

易得, 证明从略.

引理 5 $\sum_{i=1}^n (\int_{\Lambda_i} E_m(t, s) ds)^2 = O(\frac{2^m}{n})$.

易得, 证明从略.

引理 6 (Bernstein 不等式) 设 $\{x_i\}$ 是 ρ 混合序列且 $Ex_i = 0, \{a_{ni}, 1 \leq i \leq n\}$ 为常数数

列. $|x_i| \leq d_i$ a. s. $S_n = \sum_{i=1}^n a_{ni} x_i$, 则对任意 $\varepsilon > 0$, 有

$$P(|S_n| > \varepsilon) \leq C_1 \exp(-t\varepsilon + 2C_2 t^2 \Delta + 2(l+1) + 2\ln(\rho(k))),$$

其中 C_1 是与 n 无关的常数, $C_2 = 2(1+2\delta)(1+8\sum_{i=1}^n \rho(i)), \Delta = \sum_{i=1}^n a_{ni}^2 d_i^2, t > 0, l, k$ 满足如下

关系式, $2lk \leq n \leq 2(l+1)k, tk \max_{1 \leq i \leq n} |a_{ni} d_i| \leq \delta \leq \frac{1}{2}$.

证明 采用 Bosq 的方法, 对每一 $n > 1$, 存在整数 $l = l(n), k = k(n)$, 使得 $2lk \leq n \leq 2(l+1)k$. 令 $\xi_i = a_{ni} x_i$ 构造序列

$$Y_1 = \xi_1 + \xi_2 + \dots + \xi_k, Y_2 = \xi_{2k+1} + \dots + \xi_{3k}, Y_i = \xi_{2(i-1)k+1} + \dots + \xi_{(2i-1)k}, Y_l$$

$$= \xi_{2(l-1)k+1} + \dots + \xi_{(2l-1)k},$$

$$Z_1 = \xi_{k+1} + \dots + \xi_{2k}, Z_2 = \xi_{3k+1} + \dots + \xi_{4k}, Z_i = \xi_{(2i-1)k+1} + \dots + \xi_{2ik}, Z_l$$

$$= \xi_{(2l-1)k+1} + \dots + \xi_{2lk},$$

$$Z_{l+1} = \begin{cases} 0, & \text{当 } 2lk = n \text{ 时,} \\ \xi_{2lk+1} + \dots + \xi_n, & \text{当 } 2lk < n \text{ 时,} \end{cases}$$

则 $S_n = \sum_{i=1}^l Y_i + \sum_{i=1}^{l+1} Z_i$. 因为 $0 \leq n - 2lk \leq 2k$, 所以 $|Z_{l+1}| \leq 2lk \max_{1 \leq i \leq n} |a_n d_i| \leq 2\delta \leq 1$,

$$P(S_n > \epsilon) \leq e^{-\alpha} E(e^{\epsilon S_n}) \leq e^{-\alpha} E(e^{i \sum_{i=1}^l Y_i} e^{i \sum_{i=1}^{l+1} Z_i} e^{\epsilon Z_{l+1}}) \leq e^{1-\alpha} \sqrt{E(e^{2i \sum_{i=1}^l Y_i}) E(e^{2i \sum_{i=1}^{l+1} Z_i})} \quad (1)$$

记 $t_{i1} = 2(i-1)k + 1, t_{i2} = (2i-1)k, \Delta(i) = \sum_{j=t_{i1}}^{t_{i2}} a_n^2 d_j^2$ 易知 $EY_i^2 \leq (1 + 4 \sum_{i=1}^k \rho(i)) \Delta(i)$,

因为 $EY_i = 0$ 且 $|2tY_i| \leq 2\delta \leq 1$, 所以

$$Ee^{2tY_i} = \sum_{n=1}^{\infty} \frac{E(2tY_i)^n}{n!} \leq 1 + 2t^2(1 + 2\delta)EY_i^2 \leq \exp(C_2 t^2 \Delta(i)).$$

由定义知 $|EXY - EXEY| \leq \rho(k) \sqrt{\text{Var}(X)\text{Var}(Y)}$,

$$E(e^{2i \sum_{j=1}^l Y_j}) \leq E(e^{2i \sum_{j=1}^l Y_j}) E(e^{2tY_l}) + \rho(k) \sqrt{Ee^{2i \sum_{j=1}^{l-1} Y_j} Ee^{4tY_l}},$$

因为 $2|tY_l| \leq 2\delta \leq 1$, 所以 $e^{4tY_l} \leq e^2$. $E(e^{2i \sum_{j=1}^l Y_j}) \leq E(e^{2i \sum_{j=1}^{l-1} Y_j}) Ee^{2tY_l} + \rho(k)e^2$, 反复利用递推法可得

$$E(e^{2i \sum_{j=1}^l Y_j}) \leq [(E(e^{2i \sum_{j=1}^{l-1} Y_j}) Ee^{2tY_{l-1}} + \rho(k)e^{l-2}) Ee^{2tY_{l-1}} + \rho(k)e^{l-1}] Ee^{2tY_l} + \rho(k)e^l$$

$$\leq \rho(k)[e^l + e^{l-1} \exp\{C_2 t^2 \Delta(l)\} + e^{l-2} \exp\{C_2 t^2 \sum_{i=1}^l \Delta(i)\} + \dots + \exp\{C_2 t^2 \sum_{i=1}^l \Delta(i)\}]$$

$$\leq \rho(k)e^{l+1} \exp\{C_2 t^2 \Delta\}. \quad (2)$$

同理可得

$$E(e^{2i \sum_{j=1}^l Z_j}) \leq \rho(k)e^{l+1} \exp\{C_2 t^2 \Delta\}. \quad (3)$$

由(1)–(3)式可得 $P(|S_n| > \epsilon) \leq C_1 \exp(-t\epsilon + 2C_2 t^2 \Delta + 2(l+1) + 2\ln(\rho(k)))$. \square

3 本文主要结果及证明

定理 1 若 ϵ_i 是 ρ 混合序列且 $E\epsilon_i = 0$, 则 $E\hat{r}(t) - r(t) = O(\eta_m) + O(n^{-\gamma})$.

证明

$$E\hat{r}(t) = E\left(\sum_{i=1}^n Y_i \int_{A_i} E_m(t, s) ds\right) = E\left(\sum_{i=1}^n (r(t_i) + \epsilon_i) \int_{A_i} E_m(t, s) ds\right),$$

$$\begin{aligned}
&= \sum_{i=1}^n r(t_i) \int_{A_i} E_m(t, s) ds \\
E(\hat{r}(t)) &= \sum_{i=1}^n (r(t_i) - r(\xi_i)) \int_{A_i} E_m(t, s) ds + \sum_{i=1}^n r(\xi_i) \int_{A_i} E_m(t, s) ds \\
&\leq O(n^{-\gamma}) \sum_{i=1}^n \int_{A_i} |E_m(t, s)| ds + \sum_{i=1}^n r(\xi_i) \int_{A_i} E_m(t, s) ds \\
&= O(n^{-\gamma}) \int_0^1 |E_m(t, s)| ds + \int_0^1 E_m(t, s) r(s) ds.
\end{aligned}$$

由引理 1 和引理 4 可得 $E\hat{r}(t) - r(t) = O(\eta_m) + O(n^{-\gamma})$.

定理 2 $\text{Var}(\hat{r}(t)) = O((1 + 4 \sum_{i=1}^n \rho(n)) \frac{2^m}{n})$.

证明

$$\begin{aligned}
\text{Var}(\hat{r}(t)) &= E(\hat{r}(t) - E\hat{r}(t))^2 = E(\sum_{i=1}^n \epsilon_i \int_{A_i} E_m(t, s) ds)^2 \\
&= \sum_{i=1}^n \epsilon_i^2 (\int_{A_i} E_m(t, s) ds)^2 + 2 \sum_{1 \leq i < j \leq n} E(\epsilon_i \epsilon_j \int_{A_i} E_m(t, s) ds \int_{A_j} E_m(t, s) ds) \\
&\leq \sigma^2 \sum_{i=1}^n (\int_{A_i} E_m(t, s) ds)^2 + 4 \sum_{i < j} \rho(j-i) E(\epsilon_i^2 (\int_{A_i} E_m(t, s) ds)^2)^{\frac{1}{2}} E(\epsilon_j^2 (\int_{A_j} E_m(t, s) ds)^2)^{\frac{1}{2}} \\
&= \sigma^2 \sum_{i=1}^n (\int_{A_i} E_m(t, s) ds)^2 + 2\sigma^2 \sum_{k=1}^{n-1} \rho(k) \sum_{i=1}^{n-k} [(\int_{A_i} E_m(t, s) ds)^2 + (\int_{A_{i+k}} E_m(t, s) ds)^2] \\
&\leq \sigma^2 \sum_{i=1}^n (\int_{A_i} E_m(t, s) ds)^2 + 2\sigma^2 \sum_{k=1}^{n-1} \rho(k) 2 \sum_{i=1}^n [(\int_{A_i} E_m(t, s) ds)^2] \\
&= (1 + 4 \sum_{i=1}^n \rho(i)) \sum_{i=1}^n [(\int_{A_i} E_m(t, s) ds)^2] \sigma^2,
\end{aligned}$$

由引理 5 知 $\sum_{i=1}^n [(\int_{A_i} E_m(t, s) ds)^2] = O(\frac{2^m}{n})$, 结论立得.

定理 3 若 $\{\epsilon_i\}$ 是 ρ 混合序列且 $E\epsilon_i = 0, E\epsilon_i^2 = \sigma^2, \sum_{i=1}^{\infty} \rho(n) < \infty, \frac{2^m}{n} \rightarrow 0$, 则

$$E(\hat{r}(t) - r(t))^2 \rightarrow 0$$

证明 $E(\hat{r}(t) - r(t))^2 = \text{Var}(\hat{r}(t)) + (E\hat{r}(t) - r(t))^2$. 由定理 1 及定理 2 立得结论.

系 在定理 3 的条件下 $\hat{r} \xrightarrow{p} r(t)$.

定理 4 若 $\{\epsilon_i\}$ 是 ρ 混合序列, $\sum_{i=1}^{\infty} \rho(n) < \infty, E\epsilon_i = 0, E\epsilon_i^2 = \sigma^2, \frac{2^m}{n^{1-\theta}} \rightarrow 0, \frac{1}{2} < \theta < 1$, 则

$$\hat{r}(t) \xrightarrow{a.s.} r(t).$$

证明 $\hat{r}(t) - r(t) = \hat{r}(t) - E\hat{r}(t) + E\hat{r}(t) - r(t)$.

由定理 1 知, 只需证 $\hat{r}(t) - E\hat{r}(t) \rightarrow 0$, 即 $\sum_{i=1}^n \epsilon_i \int_{A_i} E_m(t, s) ds \rightarrow 0, a.s.$ 由引理 3 知只需验证

存在 $\theta > \frac{1}{2}, k > 0$ 使得 1). $\sum_{i=1}^n |\int_{A_i} E_m(t, s) ds| < k; 2). |\int_{A_i} E_m(t, s) ds| < kn^{-\theta}$. 由引理 1 和

引理 4 知, 易知 1). 和 2). 满足.

定理 5 若 $\sum_{i=1}^{\infty} \rho(i) < \infty, C_n = \max(n^{-\gamma}, \eta_m)$, 且存在 $d = d(n) \in N$ 满足 $dC_n \rightarrow \infty, \frac{2^m d^3}{nC_n} \rightarrow 0, \frac{d^2 C_n}{n} \rightarrow 0$, 则 $\hat{r}(t) - r(t) = O_p(C_n)$.

证明 由定理 1 知 $E(\hat{r}(t)) - r(t) = O(C_n)$, 下面证明 $\hat{r}(t) - E(\hat{r}(t)) = o_p(C_n)$ 即

$$\sum_{i=1}^n \int_{A_i} E_m(t, s) d\epsilon_i = o_p(C_n),$$

记 $a_{ni} = \int_{A_i} E_m(t, s) ds, S_n = \sum_{i=1}^n a_{ni} \epsilon_i$, 则

$$\begin{aligned} P(|S_n| > \eta C_n) &\leq P(|\sum_{i=1}^n a_{ni} \epsilon_i I(|\epsilon_i| \leq d)| > \eta C_n) + P(|\sum_{i=1}^n a_{ni} \epsilon_i I(|\epsilon_i| > d)| > \eta C_n) \\ &= J_{n1} + J_{n2}. \end{aligned}$$

下面特取引理 5 中 $d=t=k$, 易知引理 5 条件满足

$$\begin{aligned} J_{n1} &\leq C_1 \exp\{-tC_n \eta + C_2 t^2 \Delta + 2(l+1) + 2\ln(\rho(k))\} \\ &= C_1 \exp\{-dC_n \eta + C_2 d^2 \Delta + 2(l+1) + 2\ln(\rho(d))\} \\ &\leq C_1 \exp\{-dC_n \eta + C_2 d^2 \Delta + 2(l+1)\}, \quad \text{当 } n \text{ 充分大时,} \end{aligned} \quad (4)$$

$$\Delta = O\left(\frac{2^m d^2}{n}\right), \quad \frac{d^2 \Delta}{dC_n} = \frac{d^4 2^m}{dC_n} = \frac{2^m d^3}{nC_n} \rightarrow 0, \quad (5)$$

$$\frac{l+1}{dC_n} \leq \frac{n}{d_1 C_n} = \frac{n}{d^2 C_n} \rightarrow 0. \quad (6)$$

由 (4), (5), (6) 得 $\frac{d^2 \Delta + (l+1)}{dC_n} \rightarrow 0$, 故 d 充分大时 $J_{n1} \leq C_1 \exp(-\frac{d}{2} C_n \eta) \rightarrow 0$. 对于 J_{n2} 由 Markov 不等式得

$$J_{n2} \leq \frac{E(|\sum_{i=1}^n \epsilon_i I(|\epsilon_i| > d)| \int_{A_i} E_m(t, s) ds|)}{C_n \eta} \leq \frac{E \epsilon^2 \int_0^1 |E_m(t, s)| ds}{C_n d \eta}.$$

由引理 1 知 $J_{n2} \rightarrow 0$, 故 $P(|S_n| > \eta C_n) \rightarrow 0$, 依概率有 $S_n = o(C_n)$. \square

定理 6 若 $\sum_{i=1}^{\infty} \rho(i) < \infty, C_n = \max(n^{-\gamma}, \eta_m), \forall \epsilon > 0, E \epsilon_i^{2+\epsilon} < \infty$ 且存在 $d = d(n) \in N, p > 0$ 满足 $\frac{\ln(n)C_n}{2^m n^p} \rightarrow 0, \frac{2^{4m}}{n^{1-3p} C_n^7} \rightarrow 0, \frac{n^{1-2p} C_n^3}{2^{2m}} \rightarrow 0$, 则

$$\hat{r}(t) - r(t) = O(C_n), \quad \text{a. s.}$$

证明 由定理 1 知 $E(\hat{r}(t)) - r(t) = O(C_n)$. 下面证明 $\hat{r}(t) - E(\hat{r}(t)) = o(C_n)$, a. s. 即

$\sum_{i=1}^n \int_{A_i} E_m(t, s) d\epsilon_i = o(C_n)$, 记 $a_{ni} = \int_{A_i} E_m(t, s) ds, S_n = \sum_{i=1}^n a_{ni} \epsilon_i$, 则

$$\begin{aligned} P(|S_n| > \eta C_n) &\leq P(|\sum_{i=1}^n a_{ni} \epsilon_i I(|\epsilon_i| \leq d)| > \eta C_n) + P(|\sum_{i=1}^n a_{ni} \epsilon_i I(|\epsilon_i| > d)| > \eta C_n) \\ &= P(J_{n1}) + P(J_{n2}). \end{aligned}$$

下面特取引理 5 中 $d=t=k$, 易知引理 5 条件满足.

对于 $P(J_{n2})$ 由 Markov 不等式并类似定理 2 的证明可得

$$P(J_{n_2}) \leq \frac{E(|\sum_{i=1}^n \epsilon_i I(|\epsilon_i| > d) \int_{A_i} E_m(t,s) ds|)^2}{C_n^2 \eta^2} \leq \frac{CE(|\sum_{i=1}^n E \epsilon_i^{2+p} I(\int_{A_i} E_m(t,s) ds|)^2}{C_n^2 d \eta^2}.$$

由定理 2 知 $P(J_{n_2}) \leq \frac{2^m}{nC_n^2 d \eta^2}$, 由 B-C 引理知要求 $J_{n_2} \xrightarrow{a.s.} 0$ 只需

$$\frac{2^m}{nC_n^2 d} = O(\frac{1}{n^{1+p}}), \quad p > 0, \quad (7)$$

$$P(J_{n_1}) \leq C_1 \exp\{-tC_n \eta + C_2 t^2 \Delta + 2(l+1) + 2\ln(\rho(k))\} \\ = C_1 \exp\{-dC_n \eta + C_2 d^2 \Delta + 2(l+1) + 2\ln(\rho(d))\}.$$

由 B-C 引理知要求 $J_{n_1} \xrightarrow{a.s.} 0$ 只需 $\frac{\ln(n)}{dC_n} \rightarrow 0, \frac{2^m d^3}{nC_n} \rightarrow 0, \frac{n}{d^2 C_n} \rightarrow 0$. 把 (7) 式代入此三式可知条件成立. \square

参考文献:

- [1] 柴根象. 相依样本分布函数, 回归函数的非参数估计的强相合性 [J]. 系统科学与数学, 1998, 8: 281-288.
CHAI Gen-xiang. Based on Dependent sample, the strong consistant of Estimations for distribution and regresion function [J]. J. Sys. Sci. Math. Sci., 1998, 8: 281-288. (in Chinese)
- [2] ANTONIADIS A, GREGOIRE G, MCKEAGUE I W. Wavelet methods for curve estimation [J]. American Statistical Association Journal of the American Statistical Association, 1994, 89: 1340-1352.
- [3] 杨善朝. 混合序列加权求和的强收敛性 [J]. 系统科学与数学, 1995, 15: 254-265.
YANG Shan-chao. Strong convergence of weighted sumation for mixed sequence [J]. Sys. Sci. Math. Sci., 1995, 15: 254-265. (in Chinese)
- [4] GILBERT G. Walter Wavelet and other orthogonal System with Application CRC Press, 1994.
- [5] 陆传荣, 林正炎. 混合相依变量的极限理论 [M]. 北京: 科学出版社, 1997.
LU Chuan-rong, LIN Zheng-yan. Limit Theory for Mixed Dependent Variables [M]. Science Pres, Beijing, 1997. (in Chinese)

Wavelet Estimation of Regression Function Based on Mixing Sequence Under Fixed Design

CHAI Gen-xiang, LIU Yuan-jin

(Dept. of Appl. Math., Tongji University, Shanghai 200092, China)

Abstract; Consider nonparametric regression modal $Y_i = r(t_i) + \epsilon_i$, t_i is fixed design point and $\{\epsilon_i\}$ is ρ depending r. v. we get the estimation of $r(t)$ with wavelet method and discuss its weak convergence, strong convergence and the rate of convergence.

Key words; mixed series; wavelet estimate; regression function; week convergence; strong convergence; rate of convergence.