The Annals of Statistics
1993, Vol. 21, No. 2, 1040–1057

NONPARAMETRIC FUNCTION ESTIMATION FOR TIME SERIES BY LOCAL AVERAGE ESTIMATORS

By Lanh Tat Tran

Indiana University

Let $(\mathbf{X}_t, Y_t)$ be a stationary time series with $\mathbf{X}_t$ being R^d-valued and Y_t real valued, and where Y_t is not necessarily bounded. Let $E(Y_0|\mathbf{X}_0)$ be the conditional mean function. Under appropriate regularity conditions, local average estimators of this function can be chosen to achieve the optimal rate of convergence $(n^{-1}\log n)^{1/(d+2)}$ in L_∞ norm restricted to a compact. The result answers a question raised by Truong and Stone.

1. Introduction. Let $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \dots$ be a stationary time series with $\mathbf{X}_t$ being R^d-valued and Y_t being real valued. Let $\theta(\cdot)$ denote the conditional mean function on R^d, which is given by $\theta(\mathbf{X}_0) = E(Y_0|\mathbf{X}_0)$. Here $E(Y_0|\mathbf{X}_0)$ denotes the mean of the conditional distribution of Y_0 given $\mathbf{X}_0$. We will denote $(\mathbf{X}_t, Y_t)$ by $\mathbf{L}_t$. Let $\mathscr{F}(\mathbf{L}_t)$ be the σ-field generated by $\mathbf{L}_t$. For σ-fields $\mathscr{F}$ and $\mathscr{G}$, define

$$\alpha(\mathscr{F}, \mathscr{G}) = \sup\{|P(A \cap B) - P(A)P(B)|: A \in \mathscr{F}, B \in \mathscr{G}\}.$$

Definition 1.1. Let g be a given nonnegative function defined on $N \times N$, where N is the set of natural numbers. The process $\mathbf{L}_t$ is said to satisfy the strong mixing property in the locally transitive sense (SMLT) with respect to the function g if for all positive integers m, p,

$$\alpha\big(\mathscr{F}(\mathbf{L}_0), \mathscr{F}(\mathbf{L}_p)\big) \le \chi(p) \tag{1.1}$$

and

$$\alpha(m, p) \equiv \sup\{|P(A \cap B) - P(A)P(B)|\} \le g(mp, p)\chi(p), \tag{1.2}$$

for some constant $C > 0$ and some function $\chi(p)\downarrow 0$ as $p \to \infty$. The supremum in (1.2) is taken over all sets A, B with

$$A \in \mathscr{F}(\mathbf{L}_t: 1 \le t \le mp), \qquad B \in \mathscr{F}(\mathbf{L}_t: (m+1)p + 1 \le t \le (m+2)p),$$

where $\mathscr{F}(\mathbf{L}_t: 1 \le t \le mp)$, $\mathscr{F}(\mathbf{L}_t: (m+1)p + 1 \le t \le (m+2)p)$ are, respectively, the σ-fields generated by $\mathbf{L}_1, \dots, \mathbf{L}_{mp}$ and $\mathbf{L}_{(m+1)p+1}, \dots, \mathbf{L}_{(m+2)p}$.

The SMLT condition is weaker than the strong mixing condition defined as follows:

Definition 1.2. Let $\mathscr{F}_{-\infty}^0$ and $\mathscr{F}_n^\infty$ denote, respectively, the σ-fields generated by $\mathbf{L}_t$, $t \le 0$ and by $\mathbf{L}_t$, $t \ge n$. Then $\mathbf{L}_t$ is strong mixing if

$$\alpha(n) = \sup\{|P(A \cap B) - P(A)P(B)|: A \in \mathscr{F}_{-\infty}^0, B \in \mathscr{F}_n^\infty\}\downarrow 0 \quad \text{as } n \to \infty.$$

Received May 1989; revised September 1992.
AMS 1991 *subject classifications.* Primary 62G07; secondary 62G05, 62G20.
Key words and phrases. Nonparametric estimation, strong mixing, local mean.

For more information on strong mixing processes, see Rosenblatt (1956) and Ibragimov (1962).

Note that if $\mathbf{L}_t$ is strong mixing, then $\mathbf{L}_t$ satisifes the SMLT property with respect to the function $g \equiv 1$. To achieve generality of the SMLT condition the function $g(mp, p)$ is included to account for the cardinalities of the two groups of r.v.'s involved in Definition 1.1. We will assume throughout that for some constant $C_1 > 0$,

$$g(mp, p) \le C_1(mp + p)^a \tag{1.3}$$

or

$$g(mp, p) \le C_1(mp)^b p^b, \tag{1.4}$$

where a and b are nonnegative constants. Instead of treating each case separately, for brevity, we will simply assume throughout that

$$g(mp, p) \le C_1(mp + p)^a (mp)^b p^b. \tag{1.5}$$

Clearly, both (1.3) and (1.4) can be, respectively, obtained from (1.5) by setting $b = 0$ and $a = 0$ in (1.5). Let x be a real number. Occasionally, $[x]$ will be used to denote the integer part of x. Given positive numbers a_n, b_n, let $a_n \sim b_n$ mean that a_n/b_n is bounded away from zero and infinity. Let δ_n, $n \ge 1$, be positive numbers that tend to zero as $n \to \infty$. For $\mathbf{x} \in R^d$, define

$$I_n(\mathbf{x}) = \{i: 1 \le i \le n \text{ and } \|\mathbf{X}_i - \mathbf{x}\| \le \delta_n\},$$

and let $N_n(\mathbf{x}) = \#I_n(\mathbf{x})$ denote the number of points in $I_n(\mathbf{x})$. The local average estimator of the conditional mean function is given by

$$\hat{\theta}_n(\mathbf{x}) = (N_n(\mathbf{x}))^{-1} \sum_{I_n(\mathbf{x})} Y_i, \qquad \mathbf{x} \in R^d.$$

Under appropriate regularity conditions, Truong and Stone (1992) have shown that a local average estimator of this function based on a finite realization $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ can be chosen to achieve the optimal rate of convergence $n^{-1/(2+d)}$ in L_2 norm restricted to a compact; and it can be chosen to achieve the optimal rate of convergence $(n^{-1} \log n)^{1/(2+d)}$ in L_∞ norm restricted to a compact. Let $\mathbf{U} \subset R^d$ be a bounded nonempty set containing the origin of R^d, and let $\mathbf{C}$ be a fixed compact subset of $\mathbf{U}$. Given a real-valued function h on $\mathbf{C}$, set $\|h\|_\infty = \sup_{\mathbf{x} \in \mathbf{C}} |h(\mathbf{x})|$. Let

$$r = 1/(d + 2) \tag{1.6}$$

and let $\hat{\theta}_n$ be local average estimators with $\delta_n \sim (n^{-1} \log n)^r$. Under certain regularity conditions, and under the condition

$$P[|Y_0| \le M | \mathbf{X}_0 = \mathbf{x}] = 1, \qquad \mathbf{x} \in \mathbf{U}, \tag{1.7}$$

for some $M > 0$, Truong and Stone (1992) showed that

$$\lim_{n \to \infty} P\left[\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_\infty \ge C_2(n^{-1} \log n)^r\right] = 0, \tag{1.8}$$

for some constant $C_2 > 0$. They further raised the question whether (1.8) continues to hold if (1.7) is replaced by weaker conditions. Under conditions weaker than (1.7), we show that indeed (1.8) continues to hold. These conditions are given in Assumption 4. Details can be found in Section 2.

Asymptotic results for the conditional mean function also in the i.i.d. case have been established by Stone (1977, 1980, 1982). Estimation of the regression function or conditional mean for time series under various settings has been considered by Bierens (1983), Collomb (1984, 1985), Collomb and Härdle (1986), Robinson (1983, 1986), Yakowitz (1985, 1987) and Roussas (1990).

The difficulty arising from the unboundedness of Y is handled by a truncation argument similar to those used earlier by Mack and Silverman (1982). Relevant probability inequalities involving SMLT r.v.'s are obtained by using approximations of SMLT r.v.'s by independent ones. In the rest of the paper, for definiteness, we assume $\delta_n = (n^{-1} \log n)^r$, where r is as defined in (1.6); however, our results would also hold if $\delta_n \sim (n^{-1} \log n)^r$ as can be seen from the relevant proofs. Our paper is organized as follows: In Section 2, the assumptions and main results are stated. Section 3 presents preliminary lemmas and proof of the main result (Theorem 2.1). Lemma 3.4 shows the approximation of SMLT's r.v.'s by independent ones. Its proof is given in a separate appendix. Local average estimators are investigated in Section 3. Throughout the paper, we assume that $\mathbf{L}_t$ is SMLT with $\chi(n) = O(n^{-\rho})$ for some $\rho > 0$.

The SMLT condition is weaker than many other dependence conditions, for example, the absolute regularity condition or the ϕ-mixing condition. A large class of time series models satisfy the strong mixing condition and hence are SMLT. Indeed, autoregressive moving average and bilinear time series models are strong mixing under weak conditions [see Gorodetskii (1977) and Pham (1986)]. The assumptions of the present paper are also rather weak. Thus, the results here apply to many natural situations. An example is presented below to illustrate the theory.

EXAMPLE 1.1. Let X_t, $t = 0, \pm 1, \pm 2 \dots$ be a real valued stationary time series. Let d, m be positive integers. Let $\mathbf{X}_t = (X_{t+1}, \dots, X_{t+d})$ and $Y_t = X_{t+d+m}$. Then $\mathbf{L}_t = (\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \pm 2, \dots$ is a stationary time series, and

$$E(Y_0 | \mathbf{X}_0) = E(X_{d+m} | X_1, \dots, X_d).$$

Suppose X_t is SMLT, then $\mathbf{L}_t$ is also SMLT. Theorem 2.1 shows that if (2.1) and some additional regularity conditions (see Assumptions 1–4) are met, then

$$\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_\infty = O(\delta_n) \quad \text{a.s.}$$

If in addition, X_t is strong mixing, then a and b can be zero. Then (2.1) reduces to

$$\rho > (9vd + 2d^2 + 7v + 7d + 6)/[2(v - d - 2)], \tag{1.9}$$

where v is as defined in Assumption 4. Suppose X_t is an autoregressive process or a bilinear time series model, then X_t is strong mixing with geometric rates under weak conditions [see Gorodetskii (1977), Pham and Tran (1985), Pham (1986) and Athreya and Pantula (1986)]. In this case ρ can be chosen arbitrarily large and (1.9) is always satisfied. If Assumption 4 holds for all $v > 0$, then (1.9) is satisfied if $\rho > (9d + 7)/2$.

2. Assumptions and main results. We will employ the following assumptions.

ASSUMPTION 1. There is a positive constant M_0 such that

$$|\theta(\mathbf{x}) - \theta(\mathbf{x}')| \le M_0\|\mathbf{x} - \mathbf{x}'\| \quad \text{for } \mathbf{x}, \mathbf{x}' \in \mathbf{U},$$

where $\|\mathbf{x}\| = (x_1^2 + \cdots + x_d^2)^{1/2}$ for $\mathbf{x} = (x_1, \dots, x_d) \in R^d$.

ASSUMPTION 2. The distribution of $\mathbf{X}_0$ is absolutely continuous and its density f is bounded from zero and infinity on $\mathbf{U}$. That is, there exists a positive constant M_1 such that $M_1^{-1} \le f(\mathbf{x}) \le M_1$ for $\mathbf{x} \in \mathbf{U}$.

ASSUMPTION 3. Let $f_{\mathbf{X}_i, \mathbf{X}_j}$ denote the joint density of $\mathbf{X}_i$ and $\mathbf{X}_j$. Assume that there exists a positive constant M_2 such that $M_2^{-1} < f_{\mathbf{X}_i, \mathbf{X}_j}(\mathbf{x}, \mathbf{x}') \le M_2$ for $\mathbf{x}, \mathbf{x}' \in \mathbf{U}$ and all i, j with $j > i$.

ASSUMPTION 4. (i) There exists a constant $v > d + 2$ such that $E|Y|^v < \infty$. (ii) Suppose $\sup_{\mathbf{x}} \int |y|^v f(\mathbf{x}, y)\, dy < \infty$

REMARK 2.1. Note that Assumptions 3 and 4(ii) imply that

$$\sup_{\mathbf{x} \in \mathbf{U}} E\big(|Y_1|^v | \mathbf{X}_1 = \mathbf{x}\big) \le M_1 \sup_{\mathbf{x} \in \mathbf{U}} \int |y|^v f(\mathbf{x}, y)\, dy < \infty.$$

THEOREM 2.1. *Suppose that* $(\mathbf{X}_t, Y_t)$ *satisfies the SMLT condition and Assumptions* 1–4 *hold. Then*:

(i) $\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_\infty = O(\delta_n)$ *a.s. if for some* $v > d + 2$,

$$\rho > (2avd + 2bvd + 9vd + 4av + 6bv - 2bd + 2d^2 - 4b + 7v + 7d + 6)/[2(v - d - 2)]; \tag{2.1}$$

and

(ii) *there exists a positive constant* C_3 *such that*

$$\lim_{n \to \infty} P\big[\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_\infty > C_3\delta_n\big] = 0,$$

if for some $v > d + 2$,

$$\rho > (2avd + 2bvd + 7vd + 4av + 6bv - 2bd + 2d^2 - 4b + 3v + 7d + 6)/[2(v - d - 2)]. \tag{2.2}$$

REMARK 2.2. Our theorem applies to many general situations. First, the SMLT condition is quite weak and is satisfied by many important time series. Second, our approach removes the restriction that the time series is bounded, an assumption that appears in the work of Truong and Stone (1992). One notes that many standard time series models involve unbounded observations.

3. Preliminaries and proof of the main result. Let $u(\cdot,\cdot)$ and $v(\cdot,\cdot)$ be real valued, measurable funcitons on R^{d+1}. Set $U = u(\mathbf{X}_i, Y_i)$, $V = v(\mathbf{X}_j, Y_j)$ for $i \neq j$.

LEMMA 3.1. *Suppose that* $|u(\cdot,\cdot)| < C_4$ *and* $|v(\cdot,\cdot)| < C_5$, *where* C_4 *and* C_5 *are positive constants. Then* $|E(UV) - E(U)E(V)| \le C_6\chi(|j-i|)$ *for some constant* $C_6 > 0$.

PROOF. For any r.v.'s U, V with $|U| \le C_4$ and $|V| \le C_5$, we have

$$|\operatorname{cov}(U,V)| \le C_7 C_4 C_5 \alpha(\mathscr{F}(U), \mathscr{F}(V)) \tag{3.1}$$

for some positive constant C_7. For more information on this inequality, see Deo (1973), Hall and Heyde (1980) or Lemma 1 in Nakhapetyan (1987). By stationarity, (3.1) and (1.1),

$$\begin{aligned}|E(UV) - E(U)E(V)| &= \left|\operatorname{cov}\left(u(\mathbf{X}_0, Y_0), v(\mathbf{X}_{j-i}, Y_{j-i})\right)\right| \\ &\le C_7 C_4 C_5 \alpha\left(\mathscr{F}(\mathbf{L}_0), \mathscr{F}(\mathbf{L}_{|j-i|})\right) \le C_6\chi(|j-i|)\end{aligned}$$

for some positive constant C_6. We can choose $C_6 = CC_7C_4C_5$, where C is the constant in Definition 1.1. $\square$

LEMMA 3.2. *Suppose that* $E|U|^p < \infty$ *and* $E|V|^p < \infty$ *where* $p, q > 1$ *and* $p^{-1} + q^{-1} < 1$. *Then there exists a positive constant* C_8 *such that*

$$|E(UV) - E(U)E(V)| \le C_8 \|U\|_p \|V\|_q \chi(|j-i|)^{1-p^{-1}-q^{-1}}.$$

PROOF. By Lemma 1 in Nakhapetyan (1987), there exists a positive constant C_9 such that

$$|E(UV) - E(U)E(V)| \le C_9 \|U\|_p \|V\|_q \left(\alpha(\mathscr{F}(U), \mathscr{F}(V))^{1-p^{-1}-q^{-1}}\right).$$

Lemma 3.2 then follows from stationarity and (1.1). $\square$

Lemmas 3.1 and 3.2 are standard inequalities which will often be used. The rest of the lemmas in this section provide the main tools for the proof of Theorem 2.1. The idea is to decompose $\mathbf{C}$ into small subcubes. Each subcube has length depending on δ_n chosen as in (3.2) below. For each $\mathbf{x} \in \mathbf{C}$ there is a subcube $\mathbf{Q}_{\mathbf{w}}$ with center $\mathbf{w}$ such that $\mathbf{x} \in \mathbf{Q}_{\mathbf{w}}$. Let $\mathbf{C}_n$ denote the collection of centers of these subcubes. Using Assumption 1, it is clear that to prove Theorem 3.1, it is sufficient to prove (3.48). The general line of argument for proving (3.48) is similar to that of Truong and Stone (1992).

Without loss of generality we assume that $\mathbf{U}$ contains $\mathbf{C} = [-1/2, 1/2]^d$. Let s be a positive constant such that $(d + 2)/v < s < 1$. Let

$$l_n = \left[\delta_n^{-(2+s)} \log n\right]. \tag{3.2}$$

Let W_n be the collection of $(2l_n + 1)^d$ points in $\mathbf{C}$ each of whose coordinates is of the form $j/(2l_n)$ for some integer j such that $|j| \le l_n$. Observe that $\mathbf{C}$ can be written as the union of $(2l_n)^d$ subcubes, each having length $2\lambda_n = (2l_n)^{-1}$ and all its vertices in W_n.

Let $\mathbf{x} \in R^d$ and $\hat{r} > 0$. Denote the sphere with center at $\mathbf{x}$ and radius $\hat{r}$ as $S(\mathbf{x}, \hat{r})$.

LEMMA 3.3. *Assume that $\rho > 2$ and that Assumptions 2 and 3 hold. Let $\bar{r}_n = \delta_n + \lambda_n d^{1/2}$ and $\underline{r}_n = \delta_n - \lambda_n d^{1/2}$. Let $\mathbf{w} \in \mathbf{C}_n$ and let $S(\mathbf{w}, \bar{r}_n) - S(\mathbf{w}, \underline{r}_n)$ denote the elements of R^d in $S(\mathbf{w}, \bar{r}_n)$ but outside $S(\mathbf{w}, \underline{r}_n)$. Let $A \subset R^d$. Define $I(A, \mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $I(A, \mathbf{x}) = 0$ otherwise. Let*

$$\begin{aligned} \psi_i &= I(S(\mathbf{w}, \bar{r}_n) - S(\mathbf{w}, \underline{r}_n), \mathbf{X}_i) \quad \textit{and} \\ \pi_n &= P[\mathbf{X}_i \in S(\mathbf{w}, \bar{r}_n) - S(\mathbf{w}, \underline{r}_n)]. \end{aligned} \tag{3.3}$$

Then there exists a positive constant C_{10} such that

$$\sum_{j=1}^{n} \sum_{i=1}^{n} \left|\operatorname{cov}\{\psi_i, \psi_j\}\right| \le C_{10} n \pi_n.$$

PROOF. Let $D = \{\mathbf{x} \in R^d \colon \mathbf{x} \in S(\mathbf{w}, \bar{r}_n) - S(\mathbf{w}, \underline{r}_n)\}$. Assumptions 2 and 3 imply that $|f_{\mathbf{x}_i, \mathbf{x}_{i+j}}(\mathbf{x}, \mathbf{y}) - f(\mathbf{x}) f(\mathbf{y})|$ is bounded above for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{C}$ by $M_2 + M_1^2$. By Assumption 2,

$$\int\int_{D \times D} d\mathbf{x}\, d\mathbf{y} < M_1^2 \pi_n^2.$$

Thus there exists a positive constant C_{11} such that

$$\begin{aligned} \left|\operatorname{cov}\{\psi_i, \psi_{i+j}\}\right| &\le \int\int_{D \times D} |f_{\mathbf{x}_i, \mathbf{x}_{i+j}}(\mathbf{x}, \mathbf{y}) - f(\mathbf{x}) f(\mathbf{y})|\, d\mathbf{x}\, d\mathbf{y} \\ &\le \int\int_{D \times D} (M_2 + M_1^2)\, d\mathbf{x}\, d\mathbf{y} \le C_{11} \pi_n^2, \end{aligned} \tag{3.4}$$

where C_{11} can be any number greater than or equal to $(M_2 + M_1^2) M_1^2$.

By Lemma 3.1, for $j > 0$,

$$\left|\operatorname{cov}\{\psi_i, \psi_{i+j}\}\right| \le C_6 \chi(j). \tag{3.5}$$

Since $\rho > 2$, by (3.4) and (3.5),

$$\sum_{j=1}^{n} \sum_{i=1}^{n} \left|\operatorname{cov}\{\psi_i, \psi_j\}\right| \le n \operatorname{var} \psi_1 + C_{12} n \sum_{j=1}^{n} \min\{\chi(j), \pi_n^2\} \le C_{13} n \pi_n,$$

for some positive constants C_{12} and C_{13}. The last summation is obtained by summing over all j's between 1 and $[1/\pi_n]$ and then over all remaining j's. $\square$

The following result will be needed to approximate SMLT r.v.'s by independent ones.

LEMMA 3.4. *Suppose $n = 2pq$ for some positive integer q. Suppose V_j, $1 \le j \le q$ is a sequence of r.v.'s with V_j being measurable with respect to the σ-field*

$$\mathscr{F}(\mathbf{L}_t: (2j-1)p + 1 \le t \le 2jp).$$

Let ξ and γ be positive numbers such that $\xi \le \|V_j\|_\gamma < \infty$ for all $1 \le j \le q$. Then there exists a constant $C_{14} > 0$ and a sequence of independent r.v.'s W_j, $1 \le j \le q$ such that W_j has the same distribution as V_j and satisfies

$$P\big[|V_j - W_j| > \xi\big] \le C_{14}\big(\|V_j\|_\gamma/\xi\big)^\tau \gamma_n, \tag{3.6}$$

where

$$\gamma_n = \big\{n^a(np)^b\chi(p)\big\}^{2r} \quad \text{and} \quad \tau = \gamma/(2\gamma+1). \tag{3.7}$$

The proof of Lemma 3.4 is given in the Appendix.

Let A_n be an event. We denote the event that A_n occurs infinitely often by $[A_n \text{ i.o.}]$. Let $\mathbf{w} \in \mathbf{C}_n$ and let $\bar{I}_n = \bar{I}_n(\mathbf{w}) = \{i\colon 1 \le i \le n \text{ and } \mathbf{X}_i \in S(\mathbf{w}, \bar{r}_n)\}$. Denote

$$\overline{N}_n = \overline{N}_n(\mathbf{w}) = \#\bar{I}_n(\mathbf{w}) \quad \text{and} \quad \underline{N}_n = \underline{N}_n(\mathbf{w}) = \#\{i: \mathbf{X}_i \in S(\mathbf{w}, \underline{r}_n)\}; \tag{3.8}$$

$$\Delta_n = \big\{\overline{N}_n(\mathbf{w}) - \underline{N}_n(\mathbf{w}) \ge 2n\pi_n \text{ for some } \mathbf{w} \in \mathbf{C}_n\big\}. \tag{3.9}$$

LEMMA 3.5. *Suppose that Assumptions* 2 *and* 3 *hold. Then*:

(i) $P[\Delta_n\ i.o.] = 0$ *if* (2.1) *holds*

and

(ii) $\lim_{n\to\infty} P[\Delta_n] = 0$ *if* (2.2) *holds.*

PROOF. (i) We will employ an approximation of weakly dependent r.v.'s by independent r.v.'s as done in Tran (1989, 1990). Let

$$\lambda = \mu \log n (n\pi_n)^{-1} \quad \text{and} \quad p = \big[n\pi_n/(2\mu \log n)\big], \tag{3.10}$$

where μ is a large number to be specified later. Without loss of generality, assume $n = 2pq$ for some positive integer q. Note that $\overline{N}_n(\mathbf{w}) - \underline{N}_n(\mathbf{w}) = \sum_{i=1}^n \psi_i$, where ψ_i is defined in (3.3). The random variables $\psi_i - E\psi_i$ can be grouped successively into $2q$ blocks of size p. Write $\sum_{i=1}^n(\psi_i - E\psi_i) = S_{1n} +$

S_{2n}, where

$$S_{1n} = \sum_{j=1}^{q} V(n, 2j), \qquad S_{2n} = \sum_{j=1}^{q} V(n, 2j-1),$$

and

$$V(n, j) = \sum_{i=(j-1)p+1}^{jp} (\psi_i - E\psi_i) \quad \text{for} \quad j \geq 1.$$

Observe that

$$P\big[\overline{N}_n(\mathbf{w}) - \underline{N}_n(\mathbf{w}) \geq 2n\pi_n\big] \leq P[S_{1n} > n\pi_n/2] + P[S_{2n} > n\pi_n/2],$$

where $\overline{N}_n(\mathbf{w})$ and $\underline{N}_n(\mathbf{w})$ are as defined in (3.8).

We will refer to $V(n, 2j)$ simply as V_j for simplicity. We have

$$V_j = \sum_{i=(2j-1)p+1}^{2jp} (\psi_i - E\psi_i).$$

By Lemma 3.4, there exists a sequence of independent r.v.'s W_j, $1 \leq j \leq q$ such that W_j has the same distribution as V_j and satisfies (3.6). Now,

$$\begin{aligned} P[S_{1n} > n\pi_n/2] &\leq P\left[\sum_{j=1}^{q} W_j > n\pi_n/4\right] \\ &\quad + P\left[\sum_{j=1}^{q} (V_j - W_j) > n\pi_n/4\right]. \end{aligned} \tag{3.11}$$

Clearly, $\lambda|W_j| \leq \lambda p \leq 1/2$ a.s. and

$$\exp(\lambda W_j) \leq 1 + \lambda W_j + W_j^2 \lambda^2. \tag{3.12}$$

A simple computation using $n\pi_n \sim \delta_n^{-1+s}$ shows

$$\lambda^2 n\pi_n \sim \mu^2 (\log n)^2 (n^{-1} \log n)^{(1-s)/(d+2)} \to 0, \tag{3.13}$$

since $(d+2)/v < s < 1$. By Lemma 3.3,

$$\sum_{j=1}^{q} EW_j^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\operatorname{cov}\{\psi_i, \psi_j\}\right| \leq C_{10} n\pi_n. \tag{3.14}$$

Let l be an arbitrary large positive number. Using the independence of the W_j's, Markov's inequality and (3.12)–(3.14), we have for sufficiently large μ

$$\begin{aligned} P\left[\sum_{j=1}^{q} W_j > n\pi_n/4\right] &\leq \exp\left((-\lambda n\pi_n/4) + \lambda^2 \sum_{j=1}^{q} EW_j^2\right) \\ &\leq \exp(-(\mu \log n/4) + C_{10}\lambda^2 n\pi_n) \leq n^{-l}. \end{aligned} \tag{3.15}$$

Next,

$$(3.16)\quad P\left[\sum_{j=1}^{q}(V_j - W_j) > n\pi_n/4\right] \le q_1 \max_{1\le j\le q} P\big[(V_j - W_j) > n\pi_n/(4q)\big].$$

If $\|V_j\|_\gamma \ge n\pi_n/(4q)$, then for some positive constant C_{15},

$$(3.17)\quad P\left[\sum_{j=1}^{q}(V_j - W_j) > n\pi_n/4\right] \le C_{15} q \pi_n^{-\tau}\gamma_n$$

by (3.10), (3.6), (3.7) and (3.16) and since $\|V_j\|_\gamma^\tau \le p^\tau$ for all $1 \le j \le q$.

If $\|V_j\|_\gamma \le n\pi_n/(4q)$, then

$$(3.18)\quad P\left[\sum_{j=1}^{q}(V_j - W_j) > n\pi_n/4\right] \le q \max_{1\le j\le q} P\big[V_j - W_j > \|V_j\|_\gamma\big],$$

which is again bounded by the last term of (3.17) for large n since $\pi_n^{-\tau} \to \infty$ as $n \to \infty$. Finally by (3.11), (3.15), (3.17), (3.18) and a simple computation involving the last term of (3.17), there exists a positive constant C_{16} such that

$$(3.19)\quad P[S_{1n} > n\pi_n/2] \le n^{-l} + C_{16}(\log n)\pi_n^{-1-\tau}\gamma_n.$$

Similarly, $P[S_{2n} > n\pi_n/2]$ is bounded by the right-hand side of (3.19). Hence for some positive constant C_{17},

$$(3.20)\quad P[\Delta_n] \le C_{17} l_n^d n^{-l} + C_{17} l_n^d (\log n)\pi_n^{-1-\tau}\gamma_n,$$

where Δ_n is as defined in (3.9).

By (3.2), $l_n^d \le C_{18}(\delta_n^{-(2+s)}\log n)^d$ for some positive constant C_{18}. Thus $l_n^d n^{-l} \le n^{-2}$ for sufficiently large l and n. Using $n\pi_n \sim \delta_n^{-1+s} \to \infty$ and (3.7), after some computation, it is seen that for some positive constant C_{19}

$$(3.21)\quad l_n^d(\log n)\pi_n^{-1-\tau}\gamma_n \le C_{19} n^\alpha (\log n)^\beta,$$

with

$$(3.22)\quad \begin{aligned}\alpha = {}& [1 + \tau + 2\tau a + 2\tau b] \\ & -[(-1+s)(-1-\tau+2\tau b - 2\tau\rho) - (2+s)d]/(d+2),\end{aligned}$$

and where β is a constant. By (3.20)–(3.22) and Borel–Cantelli lemma, it follows that (i) holds if $\alpha < -1$, that is,

$$(3.23)\quad \begin{aligned}&[1 + \tau + 2\tau a + 2\tau b + 1](d+2) + (1-s)(-1-\tau+2\tau b) \\ &\quad +(2+s)d < (1-s)2\tau\rho.\end{aligned}$$

Since both quantities on the left-hand side and right-hand side of (3.23) are functions jointly continuous in τ and s, clearly (3.23) is satisfied for some $0 < \tau < 1/2$ and some $s > (d+2)/v$ if it is satisfied for $\tau = 1/2$ and $s = (d+2)/v$. Replace τ and s in (3.23) by these values and simplify, to obtain (2.1).

(ii) Clearly $\lim_{n\to\infty} P[\Delta_n] = 0$ if $\alpha < 0$ for $\tau = 1/2$ and $s = (d+2)/v$. Solving for ρ yields (2.2). □

LEMMA 3.6. *Suppose Assumptions* 2 *and* 3 *hold. Let* $\mathbf{w} \in \mathbf{C}_n$ *and* $\varphi_i = I(S(\mathbf{w}, \bar{r}_n), \mathbf{X}_i)$. *Then* $\operatorname{var}(\sum_{j=1}^n \varphi_i) \le C_{20} n \delta_n^d$ *for some positive constant* C_{20}.

Lemma 3.6 can be obtained by the same line of argument in Lemma 3.3 and is omitted. See also Lemma 4 of Truong and Stone (1992) for a similar result.

Let

$$\Gamma_n = \left[\bar{N}_n(\mathbf{w}) \le (1/2) n p_n \text{ for some } \mathbf{w} \in \mathbf{C}_n\right] \tag{3.24}$$

with

$$p_n \equiv P[\mathbf{X}_i \in S(\mathbf{w}, \bar{r}_n)]. \tag{3.25}$$

LEMMA 3.7. *Let* Γ_n *be as defined in* (3.24) *above. Suppose Assumptions* 2 *and* 3 *hold. Then*:

(i) $P[\Gamma_n \text{ i.o.}] = 0$ *if for some* $v > d + 2$,

$$\rho > [2avd + 2bvd + 9vd + 4av + 8bv + 2d^2 + 4d + 4v]/(4v), \tag{3.26}$$

and

(ii) $\lim_{n \to \infty} P[\Gamma_n] = 0$ *if for some* $v > d + 2$,

$$\rho > [2avd + 2bvd + 7vd + 4av + 8bv + 2d^2 + 4d]/(4v). \tag{3.27}$$

PROOF. (i) Since $\lambda_n = o(\delta_n)$, using Assumption 2 and (3.25), it is easy to see that $p_n \sim \delta_n^d$. Let $\mu > 0$ and $p = [np_n(2\mu \log n)^{-1}]$. Choose $\lambda = \mu \log n (np_n)^{-1}$. Then

$$P\left[\bar{N}_n(\mathbf{w}) \le (1/2) n p_n\right] = P\left[\sum_{i=1}^{n} (\varphi_i - E\varphi_i) \le -(1/2) n p_n\right]. \tag{3.28}$$

Using Lemma 3.6 and (3.28), and following the proof of Lemma 3.5, for some positive constant C_{21},

$$P\left[\bar{N}_n(\mathbf{w}) \le (1/2) n p_n\right] \le C_{21} l_n^d n^{-l} + C_{21} l_n^d \log n p_n^{-1-\tau} \gamma_n, \tag{3.29}$$

where $0 < \tau < 1/2$ and γ_n is given in (3.7). After a simple computation, the last term of (3.29) is bounded by $C_{22} n^{\alpha} (\log n)^{\beta}$ for some positive constant C_{22}, where

$$\begin{aligned} \alpha = &-[d(-1 - \tau + 2\tau b - 2\tau\rho - 2 - s)/(d + 2)] \\ &+ 2\tau a + 2\tau b + b - 2\tau\rho, \end{aligned} \tag{3.30}$$

and β is a constant. By the Borel–Cantelli lemma, (i) holds if $\alpha < -1$ for $\tau = 1/2$ and $s = (d + 2)/v$. We obtain (3.26) after simplification.

(ii) Part (ii) follows since (3.27) implies that α of (3.30) is negative for $\tau = 1/2$ and $s = (d + 2)/v$. □

Let ε be a positive number. Denote

$$h(n, \varepsilon) = n \log n (\log \log n)^{1+\varepsilon}. \tag{3.31}$$

LEMMA 3.8. *Let*

$$B_n = [h(n,\varepsilon)]^{1/v}. \tag{3.32}$$

Then, under Assumption 4(i),

$$P[|Y_n| > B_n \ i.o.] = 0.$$

PROOF. By Markov inequality and Assumption 4(i), for some positive constant $C_3 > 0$,

$$P[|Y_n| > B_n] \le B_n^{-v} E|Y_n|^v \le C_{23}[h(n,\varepsilon)]^{-1}.$$

The proof follows from the Borel–Cantelli lemma and by noting that $\sum_{n=1}^{\infty}[h(n,\varepsilon)]^{-1} < \infty$. □

For $1 \le i \le n$, define $I(|Y_i|B_n) = 1$ if $|Y_i| \le B_n$ and $I(|Y_i| \le B_n) = 0$ otherwise, where B_n is as defined in Lemma 3.8. Denote

$$K_i = I(S(\mathbf{w},\bar{r}_n),\mathbf{X}_i) \tag{3.33}$$

and

$$Z_i = Y_i I(|Y_i| \le B_n) - \theta(\mathbf{X}_i). \tag{3.34}$$

Let

$$\eta_i = K_i Z_i. \tag{3.35}$$

LEMMA 3.9. *If Assumption* 4(ii) *holds, then* $|E\eta_i| \le C_{24}\delta_n^d B_n^{1-v}$ *for some positive constant* C_{24}.

PROOF. Since $\lambda_n = o(\delta_n)$, we have for some positive constant C_{25},

$$\int_{S(\mathbf{w},\bar{r}_n)} d\mathbf{x} \le C_{25}\bar{r}_n^d \sim \delta_n^d. \tag{3.36}$$

Clearly $E[\theta(\mathbf{X}_i)K_i] = E(Y_i K_i)$. Using (3.36), there exists a positive constant C_{26} such that

$$\begin{aligned} |E\eta_i| &\le \int_{S(\mathbf{w},\bar{r}_n)} d\mathbf{x} \sup_{\mathbf{x}} \int_{\{|y|>B_n\}} |y| f(\mathbf{x},y)\,dy \\ &\le C_{26}\delta_n^d B_n^{1-v} \sup_{\mathbf{x}} \int_{\{|y|>B_n\}} |y|^v f(\mathbf{x},y)\,dy. \end{aligned} \tag{3.37}$$

The lemma follows from (3.37) by Assumption 4(ii) and by choosing C_{24} sufficiently large. □

LEMMA 3.10. *If Assumption* 4(ii) *holds, then* $\sum_{i=1}^n |E\eta_i| = o(n\delta_n^{d+1})$.

PROOF. A simple computation shows that

$$n\delta_n^d = n\delta_n^{d+1}(n^{-1}\log n)^{-1/(d+2)}. \tag{3.38}$$

Using the value of B_n in Lemma 3.8, (3.38) and Lemma 3.9,

$$\sum_{i=1}^{n} |E\eta_i| \le C_{24} n \delta_n^d B_n^{1-v} = o(n\delta_n^{d+1}),$$

since $v > (d+2)/(d+1)$. □

LEMMA 3.11. *Let K_i be as defined in* (3.33). *If Assumptions* 2 *and* 3 *are satisfied, then for some positive constants C_{27} and C_{28},*

$$E[K_i K_{i+j}] \le C_{27}\delta_n^{2d} \quad \text{for } j > 0 \quad \text{and} \quad E[K_i K_{i+j}] \le C_{28}\delta_n^d \quad \text{for } j = 0.$$

Lemma 3.11 follows easily by Assumptions 2 and 3.

LEMMA 3.12. *Suppose Assumptions* 2, 3 *and* 4 *hold. Assume in addition that $\rho > 2v/(v-2)$. Then* $\mathrm{var}[\sum_{i=1}^n \eta_i] = O(n\delta_n^d)$.

PROOF. Using Lemma 3.11, Lemma 3.12 can be obtained by a slight variation of Lemma 6 of Truong and Stone (1992). Employing Hölder's inequality, Lemmas 3.3, 3.11, Assumptions 2, 3 and 4 and Remark 2.1,

$$\mathrm{cov}\{\eta_i, \eta_j\} \le C_{29}(\delta_n^d)^{2/v}\{\chi(|j-i|)\}^{1-(2/v)} \tag{3.39}$$

for some positive constant C_{29}.

By Hölder's inequality and Lemma 3.11, for some positive constant C_{30}, we have

$$\mathrm{cov}\{\eta_i, \eta_j\} \le C_{30}(\delta_n^d)^{2/v}\{\delta_n^{2d}\}^{1-(2/v)}. \tag{3.40}$$

Let $K = [(\delta_n^d)^{-1+(2/v)}]$. By (3.39) and (3.40), for some positive constants C_{31} and C_{32},

$$\begin{aligned}\mathrm{var}\left[\sum_{i=1}^n \eta_i\right] &\le C_{31} n\delta_n^d \left(1 + \sum_{j=1}^{K} (\delta_n^d)^{1-(2/v)} + \sum_{j=K+1}^{n} (\delta_n^d)^{(2/v)-1}\{\chi(j)\}^{1-(2/v)}\right)\\ &\le C_{32} n \delta_n^d,\end{aligned}$$

since $v > 2$ and $-\rho[1-(2/v)] + 2 < 0$ by assumption. □

Let

$$\Lambda_n = \left[\max_{\mathbf{w}\in\mathbf{C}_n}\left|\sum_{i=1}^n (\eta_i - E\eta_i)\right| \ge \mu n \delta_n^{d+1}\right].$$

LEMMA 3.13. *Suppose that Assumptions* 2, 3 *and* 4 *are satisfied and that $\rho > 2v/(v-2)$. Then there exists a number $\mu > 0$ such that*:

(i) $P[\Lambda_n$ *i.o.*$] = 0$ *if* (2.1) *holds*,

and

(ii) $\lim_{n\to\infty} P[\Lambda_n] = 0$ *if* (2.2) *holds*.

PROOF. (i) Note that $|\theta(\mathbf{X}_i)K_i| \le \sup_{\mathbf{x}\in\mathbf{C}}|\theta(\mathbf{x})|$, which is bounded above by a positive constant since $S(\mathbf{w}, \bar{r}_n) \subset \mathbf{C}$. Let $\varepsilon > 0$ and let B_n be as in (3.32). Let η_i be as defined in (3.35). Then there exists a positive constant C_{33} such that $|\eta_i| \le C_{33}\{h(n,\varepsilon)\}^{1/v}$, where $h(n,\varepsilon)$ is as defined in (3.31). Assume $n = 2pq$ as in Lemma 3.5. Set the random variables $\eta_i - E\eta_i$ successively into $2q$ blocks of size p. Define $S_{1n}, S_{2n}, V(n,j), W_j$ as in Lemma 3.5 except with ψ_i replaced by η_i. Let

$$\lambda = \delta_n \quad \text{and} \quad p = \left[\delta_n^{-1}\{h(n,\hat{\varepsilon})\}^{-1/v}\right], \tag{3.41}$$

where $\hat{\varepsilon}$ is a positive number greater than ε. Then $p \to \infty$ since $v > d+2$. Now, $|V(n,j)| \le C_{33}p\{h(n,\varepsilon)\}^{1/v}$ and

$$\lambda|V(n,j)| \le C_{33}(\log\log n)^{(\varepsilon-\hat{\varepsilon})(1/v)}, \tag{3.42}$$

which tends to zero as $n \to \infty$.

Using (3.42), Lemma 3.12, Markov inequality and arguing as in Lemma 3.5, we obtain that for some positive constant C_{34},

$$P\left[\left|\sum_{j=1}^{q} W_j\right| > (\mu/4)n\delta_n^{d+1}\right] \le \exp((-\mu/4) + C_{34})\log n \le n^{-l}, \tag{3.43}$$

by choosing μ sufficiently large.

We next find an upper bound for $P[|\sum_{j=1}^{q}(V_j - W_j)| > \mu n\delta_n^{d+1}/4]$. If $\mu n\delta_n^{d+1}/4q \le \|V_j\|_\gamma$, we have by using Lemma 3.4 that for some positive constant C_{35}

$$P\left[\left|\sum_{j=1}^{q}(V_j - W_j)\right| > \mu n\delta_n^{d+1}/4\right] \le C_{35}q\nu_n\gamma_n, \tag{3.44}$$

where γ_n is defined in (3.7), $\nu_n = (q/(n\delta_n^{d+1}))^\tau[p\{h(n,\varepsilon)\}^{1/v}]^\tau$ and $0 < \tau < 1/2$. If $\mu n\delta_n^{d+1}/4q \le \|V_j\|_\gamma$, then following (3.18) and (3.6),

$$P\left[\left|\sum_{j=1}^{q}(V_j - W_j)\right| > \mu n\delta_n^{d+1}/4\right] \le C_{14}q\gamma_n, \tag{3.45}$$

which is again bounded by the last term of (3.44) for sufficiently large n since ν_n tends to infinity as $n \to \infty$. Recall that C_{14} is the constant in (3.6).

Finally, using (3.43)–(3.45), for some positive constant C_{36},

$$P[\Lambda_n] \le C_{36}l_n^d n^{-l} + C_{36}l_n^d q\nu_n\gamma_n. \tag{3.46}$$

Using the value p in (3.41), the last term of (3.46) is bounded by $C_{37}n^\alpha(\log n)^\beta$ for some positive constant C_{37}, where β is a constant and

$$\begin{aligned}\alpha &= 1 + 2\tau(a+b) + \frac{\tau(d+1) + (2+s)d}{d+2}\\ &\quad + \frac{\tau}{v} + \frac{(v-d-2)(2\tau(b-\rho)-1)}{vd+2v}.\end{aligned} \tag{3.47}$$

Thus $P[\Lambda_n \text{ i.o.}] = 0$ if $\alpha < -1$ for $\tau = 1/2$ and $s = (d+2)/v$. Solving for ρ, we obtain again (2.1).

(ii) Part (ii) follows since (2.2) ensures that α of (3.47) is negative for $\tau = 1/2$ and $s = (d+2)/v$. $\square$

PROOF OF THEOREM 2.1. (i) Using Assumption 1 and following the proof of Theorem 3 of Truong and Stone (1992), to complete the theorem it is sufficient to show that

$$\max_{\mathbf{w}\in\mathbf{C}_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})| = O(\delta_n) \quad \text{a.s.} \tag{3.48}$$

Set $\bar{\theta}_n(\mathbf{w}) = \text{ave}\{Y_i\colon i \in \bar{I}_n(\mathbf{w})\}$, $\mathbf{w} \in \mathbf{C}_n$. Then (3.48) follows from

$$\max_{\mathbf{w}\in\mathbf{C}_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} |\hat{\theta}_n(\mathbf{x}) - \bar{\theta}_n(\mathbf{w})| = O(\delta_n) \quad \text{a.s.,} \tag{3.49}$$

and

$$\max_{\mathbf{w}\in\mathbf{C}_n} |\bar{\theta}_n(\mathbf{w}) - \theta(\mathbf{w})| = O(\delta_n) \quad \text{a.s.} \tag{3.50}$$

We now verify (3.49) and (3.50). Let C_{38} and C_{39} be positive constants. Define

$$\begin{aligned} E_n &= \left\{\bar{N}_n(\mathbf{w}) - N_n(\mathbf{w}) \le C_{38}\delta_n^{-1+s} \text{ for all } \mathbf{w} \in \mathbf{C}_n\right\},\\ H_n &= \left\{\bar{N}_n(\mathbf{w}) \ge C_{39} n \delta_n^d \text{ for all } \mathbf{w} \in \mathbf{C}_n\right\},\\ G_n &= \left\{|Y_i| \le B_n,\ 1 \le i \le n\right\} \quad \text{and} \quad \Psi(n) = E_n \cap H_n \cap G_n. \end{aligned} \tag{3.51}$$

A simple computation shows that (2.1) implies (3.26). Recall from the proof of Lemma 3.5 that $\delta_n^{-1+s} \sim n\pi_n$. By Lemmas 3.6 and 3.8, there exist constants C_{38} and C_{39} such that $\lim_{n\to\infty} I(E_n, \omega) = 1$ a.s. and $\lim_{n\to\infty} I(H_n, \omega) = 1$ a.s. Since B_n is increasing in n, by Lemma 3.8, $\lim_{n\to\infty} I(G_n, \omega) = 1$ a.s. Therefore $\lim_{n\to\infty} I(\Psi(n), \omega) = 1$ a.s. for some constants C_{38} and C_{39}.

We now proceed to prove (3.50). Let H_n^c be the complement of H_n. Let Z_i be as defined in (3.34). It is easy to see that for any positive constant C_{40},

$$\begin{aligned} &\left[\max_{\mathbf{w}\in\mathbf{C}_n} \left|\bar{N}_n(\mathbf{w})^{-1} \sum_{i\in\bar{I}_n(\mathbf{w})} Z_i\right| \ge C_{40}\delta_n \text{ i.o.}\right]\\ &\quad\subseteq [H_n^c \text{ i.o.}] \cup \left[\max_{\mathbf{w}\in\mathbf{C}_n} \left|\sum_{i\in\bar{I}_n(\mathbf{w})} Z_i\right| \ge C_{40}C_{39} n \delta_n^{d+1} \text{ i.o.}\right], \end{aligned} \tag{3.52}$$

where C_{39} is the constant in the definition of H_n in (3.51). By Lemma 3.7,

$$P[H_n^c \text{ i.o.}] = 0.$$

Let η_i be as defined in (3.35). Then

$$(3.53)\quad \begin{aligned} &P\left[\max_{\mathbf{w}\in\mathbf{C}_n}\left|\sum_{i\in\bar{I}_n(\mathbf{w})} Z_i\right| \geq C_{40}C_{39}n\delta_n^{d+1} \text{ i.o.}\right] \\ &\leq P\left[\max_{\mathbf{w}\in\mathbf{C}_n}\left|\sum_{i=1}^{n}(\eta_i - E\eta_i)\right| \geq C_{40}C_{39}n\delta_n^{d+1} - \sum_{i=1}^{n}|E\eta_i| \text{ i.o.}\right], \end{aligned}$$

which is equal to zero for sufficiently large C_{40} by Lemmas 3.10 and 3.13. Thus, for some constant $C_{40} > 0$,

$$(3.54)\qquad P\left[\max_{\mathbf{w}\in\mathbf{C}_n}\left|\bar{N}_n(\mathbf{w})^{-1}\sum_{i\in\bar{I}_n(\mathbf{w})} Z_i\right| > C_{40}\delta_n \text{ i.o.}\right] = 0.$$

Clearly,

$$(3.55)\qquad \left|\sum_{i\in\bar{I}_n(\mathbf{w})}[Y_i - \theta(\mathbf{X}_i)]\right| = \left|\sum_{i\in\bar{I}_n(\mathbf{w})} Z_i + [Y_i I(|Y_i| > B_n)]\right|.$$

Note that since B_n is increasing in n,

$$(3.56)\qquad P[Y_i I(|Y_i| > B_n) \neq 0 \text{ for some } i \leq n \text{ i.o.}] = 0,$$

by Lemma 3.8. Employing (3.54)–(3.56),

$$(3.57)\quad P\left[\max_{\mathbf{w}\in\mathbf{C}_n}\left|\bar{N}_n(\mathbf{w})^{-1}\sum_{i\in\bar{I}_n(\mathbf{w})}[Y_i - \theta(\mathbf{X}_i)]\right| > C_{40}\delta_n \text{ i.o.}\right] = 0 \quad \text{a.s.}$$

By Assumption 1, and since $\delta_n = o(\lambda_n)$,

$$|\theta(\mathbf{X}_i) - \theta(\mathbf{w})| \leq M_0\|\mathbf{X}_i - \mathbf{w}\| \leq C_{41}\delta_n \quad \text{for } i \in \bar{I}_n(\mathbf{w}),$$

for some positive constant C_{41}. Let

$$(3.58)\quad D_n = \left[\left(\bar{N}_n(\mathbf{w})\right)^{-1}\sum_{i\in\bar{I}_n(\mathbf{w})}[\theta(\mathbf{X}_i) - \theta(\mathbf{w})]| \geq C_{40}\delta_n \text{ for some } \mathbf{w}\right].$$

Then for C_{40} in (3.58) sufficiently large,

$$(3.59)\qquad \lim_{n\to\infty} I(D_n, \omega) = 0 \quad \text{a.s.}$$

The proof of (3.50) now follows from (3.57) and (3.59).

Given $\mathbf{x} \in \mathbf{C}$, let $N_n = N_n(\mathbf{x})$ and $I_n = I_n(\mathbf{x})$ and choose $\mathbf{w}$ such that $\mathbf{x} \in Q_{\mathbf{w}}$. Then $\underline{N}_n \leq N_n \leq \bar{N}_n$ and following the proof of Theorem 3 of Truong and Stone (1992), for $\omega \in \Psi_n$,

$$(3.60)\quad \begin{aligned} \left|\left(\bar{N}_n\right)^{-1}\sum_{i\in\bar{I}_n} Y_i\right| &\leq 2\left(\bar{N}_n - \underline{N}_n\right)\left(\bar{N}_n\right)^{-1}\max_{i\in\bar{I}_n}|Y_i| \\ &\leq 2C_{38}\delta_n^{-1+s}\left(C_{39}n\delta_n^d\right)^{-1}B_n = o(\delta_n) \end{aligned}$$

because $s > (d+2)/v$. Since $\lim_{n\to\infty} I(\Psi(n), \omega) = 1$ a.s., by (3.60),

$$\max_{\mathbf{w}\in\mathbf{C}_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} \left| (\bar{N}_n)^{-1} \sum_{i\in\bar{I}_n} Y_i - (N_n)^{-1} \sum_{i\in I_n} Y_i \right| = o(\delta_n) \quad \text{a.s.} \tag{3.61}$$

Finally (3.49) follows from (3.61).

(ii) By Assumption 1, it is sufficient to show that for some constant $C_{42} > 0$,

$$\lim_{n\to\infty} P\left[\max_{\mathbf{w}\in\mathbf{C}_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} |\hat{\theta}_n(x) - \bar{\theta}_n(\mathbf{w})| \geq C_{42}\delta_n \right] = 0 \tag{3.62}$$

and

$$\lim_{n\to\infty} P\left[\max_{\mathbf{w}\in\mathbf{C}_n} |\bar{\theta}_n(\mathbf{w}) - \theta(\mathbf{w})| \geq C_{42}\delta_n \right] = 0. \tag{3.63}$$

A simple computation shows that (2.2) is a stronger condition than (3.27). Thus the conditions of Lemmas 3.5(ii), 3.7(ii) and 3.13(ii) are met. By Lemma 3.8, $\lim_{n\to\infty} I(G_n, \omega) = 1$ a.s. Therefore $\lim_{n\to\infty} P[\Psi(n)] = 1$ for some C_{38} and C_{39}. The proof of (3.62) and (3.63) can now be obtained by a slight variation of the proof of Part (i). □

APPENDIX

We will need the following lemma of Bradley (1983):

LEMMA A.1. *Suppose X and Y are random variables taking their values on $\mathscr{S}$ and R, respectively, where $\mathscr{S}$ is a Borel space, and let U be a uniform-*[0, 1] *r.v. independent of (X, Y); furthermore, suppose ξ and γ are positive numbers such that $\xi \leq \|Y\|_\gamma < \infty$. Then there exists a real-valued r.v. $Y^* = f(X, Y, U)$, where f is a measurable function defined on $\mathscr{S} \times R \times [0,1]$, such that:*

(i) *Y^* is independent of X,*
(ii) *the probability distributions of Y and Y^* are identical, and*

$$P[|Y^* - Y| \geq \xi] \leq 18(\|Y\|_\gamma/\xi)^{\gamma/(2\gamma+1)}\{\alpha(\mathscr{B}(X), \mathscr{B}(Y))\}^{2\gamma/(2\gamma+1)},$$

where $\mathscr{B}(X)$ and $\mathscr{B}(Y)$ are the σ-fields induced by X and Y, respectively, and $\|Y\|_\gamma = (E|Y|^\gamma)^{1/\gamma}$.

PROOF OF LEMMA 3.4. By enlarging the probability space if necessary, introduce a sequence $U_1, \ldots, U_q$ of independent uniform [0, 1] r.v.'s, this sequence of independent r.v.'s being independent of $V_1, \ldots, V_q$. Define $W_1 = V_1$. By Lemma A.1, for each $j \geq 1$, there exists a r.v. W_j which is a measurable function of $V_1, \ldots, V_j, U_j$ such that W_j is independnet of $V_1, \ldots, V_{j-1}$, has the same distribution as V_j and satisfies

$$\begin{aligned} P\left[|V_j - W_j| > \xi\right] &\le 18\left(\|V_j\|_\gamma/\xi\right)^\tau\left\{\alpha\left(\mathscr{F}(V_1,\dots,V_{j-1}),\mathscr{F}(V_j)\right)\right\}^{2\tau} \\ &\le 18\left(\|V_j\|_\gamma/\xi\right)^\tau\left\{\alpha\left(\mathscr{F}(\mathbf{L}_1,\dots,\mathbf{L}_{2(j-1)p}),\right.\right. \\ &\qquad\qquad \left.\left.\mathscr{F}(\mathbf{L}_{(2j-1)p+1},\dots,\mathbf{L}_{2jp})\right)\right\}^{2\tau} \\ &\le 18\left(\|V_j\|_\gamma/\xi\right)^\tau\left\{\alpha(2(j-1),p)\right\}^{2\tau}. \end{aligned}$$

Since $2(j-1)p \le n$, by (1.2) and (1.5),

$$\alpha(2(j-1),p) \le C_1(n+p)^a n^b p^b.$$

Relation (3.6) is then obtained by choosing C_{14} sufficiently large.

It remains to show that $W_1,\dots,W_q$ are independent. We will follow the argument of (3.10) in Izenman and Tran (1990). To prove this it is sufficient to show that W_j and $(W_1,\dots,W_{j-1})$ are independent for $j>1$. Note that $(V_1,\dots,V_j)$, $U_1,\dots,U_j$ are independent. Thus $(V_1,\dots,V_j,U_j)$, $U_1,\dots,U_{j-1}$ are independent. Since W_j is a measurable function of $V_1,\dots,V_j,U_j$, it follows that $(W_j,V_1,\dots,V_{j-1})$, $U_1,\dots,U_{j-1}$ are independent. Now W_j is independent of $V_1,\dots,V_{j-1}$. Hence $W_j,(V_1,\dots,V_{j-1}),U_1,\dots,U_{j-1}$ are independent. Finally W_j and $(W_1,\dots,W_{j-1})$ are independent since $(W_1,\dots,W_{j-1})$ is measurable with respect to the σ-field generated by $V_1,\dots,V_{j-1},U_1,\dots,U_{j-1}$. $\square$

Acknowledgment. I would like to thank the referees for many valuable comments.

REFERENCES

ATHREYA, K. B. and PANTULA, S. G. (1986). Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23** 880–892.

BIERENS, H. J. (1983). Uniform consistency of kernel estimatiors of a regression under general conditions. *J. Amer. Statist. Assoc.* **78** 699–707.

BRADLEY, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.* **30** 69–81.

COLLOMB, G. (1984). Propriétes de convergence presque complete du prédicteur a noyau. *Z. Wahrsch. Verw. Gebiete* **66** 441–460.

COLLOMB, G. (1985). Nonparametric time series analysis and prediction: Uniform almost sure convergence of the window and $k - NN$ autoregression estimates. *Math. Operationsforsch. Statist. Ser. Statist.* **16** 297–307.

COLLOMB, G. and HÄRDLE, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic Process. Appl.* **23** 77–89.

DEO, C. M. (1973). A note on empirical processes of strong mixing sequences. *Ann. Probab.* **1** 870–875.

GORODETSKII, V. V. (1977). On the strong mixing properties for linear sequences. *Theory Probab. Appl.* **22** 411–413.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic, New York.

HART, J. D. and VIEU, P.(1988). Nonparametric regression under dependence: A class of asymptotically optimal data-driven bandwidths. Technical Report, Dept. Statistics, Texas A & M Univ., College Station, TX.

IBRAGIMOV, I. A. (1962). Some limit theorems for stationary sequences. *Theory Probab. Appl.* **4** 347–382.

IZENMAN, A. J. and TRAN, L. T. (1990). Estimation of the survival function and hazard rate under weak dependence, *J. Statist. Plann. Inference* **24** 233–247.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415.

NAKHAPETYAN, B. S. (1987). An approach to proving limit theorems for dependent random variables. *Theory Probab. Appl.* **32** 535–539.

PHAM, D. T. (1986). The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Process. Appl.* **23** 291–300.

PHAM, D. T. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

ROBINSON, P. M. (1986). On the consistency and finite sample properties of nonparametric kernel time series regression, autoregression and density estimators. *Ann. Inst. Statist. Math.* **38** 539–549.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43–47.

ROUSSAS, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Process. Appl.* **36** 107–116.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

TRAN, L. T. (1989). Recursive density estimation under dependence. *IEEE Trans. Inform. Theory* **35** 1103–1108.

TRAN, L. T. (1990). Kernel density estimation under dependence. *Statist. Probab. Lett.* **10** 193–201.

TRUONG, Y. K. and STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.

YAKOWITZ, S. (1985). Nonparametric density estimation, prediction and regression for Markov sequences. *J. Amer. Statist. Assoc.* **80** 215–221.

YAKOWITZ, S. (1987). Nearest-neighbor methods for time series analysis. *J. Time Ser. Anal.* **8** 235–247.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
SWAIN HALL EAST
BLOOMINGTON, INDIANA 47405