The Annals of Statistics
1992, Vol. 20, No. 1, 77–97

NONPARAMETRIC FUNCTION ESTIMATION INVOLVING TIME SERIES

By Young K. Truong and Charles J. Stone[1]

University of North Carolina, Chapel Hill, and University of California, Berkeley

Consider a stationary time series $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \dots$, with $\mathbf{X}_t$ being $\mathbb{R}^d$-valued and Y_t real-valued. The conditional mean function is given by $\theta(\mathbf{X}_0) = E(Y_0 | \mathbf{X}_0)$. Under appropriate regularity conditions, 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 $n^{-1/(2+d)}$ both pointwise and in L_2 norms restricted to a compact; and it can also be chosen to achieve the optimal rate of convergence $(n^{-1}\log(n))^{1/(2+d)}$ in L_∞ norm restricted to a compact. Similar results hold for local median estimators of the conditional median function, which is given by $\theta(\mathbf{X}_0) = \operatorname{med}(Y_0 | \mathbf{X}_0)$.

1. Statement of results. Let $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \dots$, denote a (strictly) stationary time series with $\mathbf{X}_t$ being $\mathbb{R}^d$-valued and Y_t being real-valued. Let $\theta(\cdot)$ denote either the conditional mean (regression function) on $\mathbb{R}^d$, which is given by $\theta(\mathbf{X}_0) = E(Y_0 | \mathbf{X}_0)$, or the conditional median function on $\mathbb{R}^d$, which is given by $\theta(\mathbf{X}_0) = \operatorname{med}(Y_0 | \mathbf{X}_0)$. Here $E(Y_0 | \mathbf{X}_0)$ and $\operatorname{med}(Y_0 | \mathbf{X}_0)$ denote the mean and median, respectively, of the conditional distribution of Y_0 given $\mathbf{X}_0$.

Example 1 (Univariate time-series). Let X_t, $t = 0, \pm 1, \pm 2, \dots$, be a real-valued stationary time series, let d be a positive integer and let m be an integer. Set

$$\mathbf{X}_t = (X_{t+1}, \dots, X_{t+d}) \quad \text{and} \quad Y_t = X_{t+d+m}.$$

Then $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \dots$, is a stationary time series,

$$E(Y_0 | X_0) = E(X_{d+m} | X_1, \dots, X_d)$$

and

$$\operatorname{med}(Y_0 | X_0) = \operatorname{med}(X_{d+m} | X_1, \dots, X_d).$$

In the context of forecasting m units of time into the future, m is a positive integer.

Example 2 (Bivariate time-series). Let (X_t, Z_t), $t = 0, \pm 1, \dots$, be an $\mathbb{R}^2$-valued stationary time series, and let d be a positive integer and m a

Received September 1988; revised June 1990.
[1]Research partly supported by NSF Grant DMS-86-00409.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20.
Key words and phrases. Stationarity, conditional mean function, local average, conditional median function, local median, rate of convergence.

nonnegative integer. Set

$$\mathbf{X}_t = (X_{t+1}, \dots, X_{t+d}) \quad \text{and} \quad Y_t = Z_{t+d+m}.$$

Then $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \dots$, is a stationary time series,

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

and

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

EXAMPLE 3 (Bivariate time-series). Let (X_t, Z_t), $t = 0, \pm 1, \dots$, be an $\mathbb{R}^2$-valued stationary time series, and let d, k and m be positive integers such that $k \le d$. Set

$$\mathbf{X}_t = (X_{t+1}, \dots, X_{t+k}, Z_{t+k+1}, \dots, Z_{t+d}) \quad \text{and} \quad Y_t = Z_{t+d+m}.$$

Then $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \dots$, is a stationary time series

$$E(Y_0|\mathbf{X}_0) = E(Z_{d+m}|X_1, \dots, X_k, Z_{k+1}, \dots, Z_d)$$

and

$$\operatorname{med}(Y_0|\mathbf{X}_0) = \operatorname{med}(Z_{d+m}|X_1, \dots, X_k, Z_{k+1}, \dots, Z_d).$$

In this paper, we use local averages to estimate the conditional mean function and local medians to estimate the conditional median function. These estimators will be shown to possess optimal rates of convergence under various conditions, which will now be listed.

Let U be a nonempty open subset of the origin of $\mathbb{R}^d$. The following smoothness condition is imposed on the conditional mean function or the conditional median function.

CONDITION 1. There is a positive constant M_0 such that

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

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

[Denote the conditional distribution function of Y_0 given $\mathbf{X}_0 = \mathbf{x}$ by $G(y|\mathbf{x})$ and its density by $g(y|\mathbf{x})$. Set $\theta(\mathbf{x}) = \operatorname{med}(Y_0|\mathbf{X}_0 = \mathbf{x})$ and let c_1, c_2 and c_3 be positive constants. Suppose $g(y|\mathbf{x}) > c_1$ and $|G(y|\mathbf{x}) - G(y|\mathbf{x}')| \le c_2\|x - x'\|$ for $|y - \theta(\mathbf{x})| < c_3$ and $\mathbf{x}, \mathbf{x}' \in U$. Then Condition 1 holds for the conditional median function $\theta(\cdot)$.]

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

CONDITION 3. For $j \ge 1$, the conditional distribution of $\mathbf{X}_j$ given $\mathbf{X}_0 = \mathbf{x}$ has a density $f_j(\cdot|\mathbf{x})$; there is a positive constant M_2 such that

$$M_2^{-1} \le f_j(\mathbf{x}'|\mathbf{x}) \le M_2 \quad \text{for } \mathbf{x}, \mathbf{x}' \in U \text{ and } j \ge 1.$$

Each conclusion of Theorems 1–3 below requires (i), (ii) or (iii) of the following condition.

CONDITION 4. (i) There is a positive constant $\nu > 2$ such that

$$\sup_{\mathbf{x} \in U} E\big(|Y_0|^{\nu} | \mathbf{X}_0 = \mathbf{x}\big) < \infty.$$

(ii) There is a positive constant M_3 such that

$$P\big(|Y_0| \le M_3 | \mathbf{X}_0 = \mathbf{x}\big) = 1, \qquad \mathbf{x} \in U.$$

(iii) The conditional distribution of Y_0 given $\mathbf{X}_0 = \mathbf{x}$ is absolutely continuous and its density $g(y|\mathbf{x})$ is bounded away from zero and infinity over a neighborhood of the median; that is, there are positive constants ε_0 and M_4 such that

$$M_4^{-1} \le g(y|\mathbf{x}) \le M_4, \qquad y \in \big(\theta(\mathbf{x}) - \varepsilon_0, \theta(\mathbf{x}) + \varepsilon_0\big) \text{ and } \mathbf{x} \in U.$$

Let $\mathscr{F}_t$ and $\mathscr{F}^t$ denote the σ-fields generated, respectively, by $(\mathbf{X}_i, Y_i)$, $-\infty < i \le t$, and $(\mathbf{X}_i, Y_i)$, $t \le i < \infty$. Given a positive integer k, set

$$\alpha(k) = \sup\{|P(A \cap B) - P(A)P(B)|: A \in \mathscr{F}_t, \text{And } B \in \mathscr{F}^{t+k}\}.$$

The stationary sequence is said to be α-mixing or strongly mixing if $\alpha(k) \to 0$ as $k \to \infty$. Each conclusion of Theorems 1–3 requires (i), (ii) or (iii) of the following condition. [Note that (i), (ii) and (iii) are increasingly strong forms of α-mixing.]

CONDITION 5. (i) $\sum_{i \ge N} \alpha(i) = O(N^{-1})$ as $N \to \infty$.
(ii) $\sum_{i \ge N} \alpha^{1-(2/\nu)}(i) = O(N^{-1})$ as $N \to \infty$ ($\nu > 2$).
(iii) $\alpha(N) = O(\rho^N)$ as $N \to \infty$ for some ρ with $0 < \rho < 1$.

Given positive numbers a_n and b_n, $n \ge 1$, let $a_n \sim b_n$ mean that a_n/b_n is bounded away from zero and infinity. Given random variables V_n, $n \ge 1$, let $V_n = O_P(b_n)$ mean that the random variables $b_n^{-1}V_n$, $n \ge 1$, are bounded in probability; that is, that

$$\lim_{c \to \infty} \limsup_n P(|V_n| > cb_n) = 0.$$

Let δ_n, $n \ge 1$, be positive numbers that tend to zero as $n \to \infty$. For $\mathbf{x} \in \mathbb{R}^d$ and $n \ge 1$, set

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

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

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

the local median estimator of the conditional median function is given by

$$\hat{\theta}(\mathbf{x}) = \operatorname{med}(Y_i : \mathbf{x} \in I_n(\mathbf{x})), \qquad \mathbf{x} \in \mathbb{R}^d.$$

Set $r = 1/(2+d)$. The local (pointwise) rate of convergence of $\hat{\theta}_n(\cdot)$ is given in the following result.

THEOREM 1. *Suppose that* $\delta_n \sim n^{-r}$ *and that Conditions* 1–3 *hold. Suppose also that Conditions* 4(i) *and* 5(ii) *hold for estimation of the conditional mean and that Conditions* 4(iii) *and* 5(i) *hold for estimation of the conditional median. Then*

$$|\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| = O_P(n^{-r}), \qquad \mathbf{x} \in U.$$

Let C be a fixed compact subset of U having a nonempty interior. Given a real-valued function h on C, set

$$\|h\|_q = \left\{\int_C |h(\mathbf{x})|^q \, d\mathbf{x}\right\}^{1/q}, \qquad 1 \le q < \infty \quad \text{and} \quad \|h\|_\infty = \sup_{\mathbf{x}\in C} |h(\mathbf{x})|.$$

The L_q rate of convergence is given in the following result.

THEOREM 2. *Suppose that* $\delta_n \sim n^{-r}$ *and that Conditions* 1–3 *and* 5(iii) *hold. Suppose also that Condition* 4(i) *holds and* $q = 2$ *for estimation of the conditional mean and that Condition* 4(iii) *holds for estimation of the conditional median. Then*

$$\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_q = O_P(n^{-r}), \qquad 1 \le q < \infty.$$

The L_∞ rate of convergence is given in the following result.

THEOREM 3. *Suppose that* $\delta_n \sim (n^{-1}\log n)^r$ *and that Conditions* 1–3 *and* 5(iii) *hold. Suppose also that Condition* 4(ii) *holds for estimation of the conditional mean and that Condition* 4(iii) *holds for estimation of the conditional median. Then there is a positive constant c such that*

$$\lim_n P\left(\|\hat{\theta}_n(\cdot) - \theta(\cdot)\|_\infty \ge c\left[n^{-1}\log(n)\right]^r\right) = 0.$$

The proofs of Theorems 1–3 for estimation of the conditional mean will be given in Section 2 and the proofs for estimation of the conditional median will be given in Section 3.

Under the iid assumption, asymptotic results for the conditional mean function estimation were established by Stone (1977, 1980, 1982). Some of these results have been extended by Bierens (1983), Collomb (1984), Doukhan and Ghindes (1980), Robinson (1983) and Yakowitz (1985, 1987) to time series under various mixing conditions. In particular, Collomb (1984) and Bierens (1983) considered the uniform consistency for kernel estimators based on local averages under the ϕ-mixing condition, which is considerably stronger than the α-mixing condition adopted in this paper. Also, the approach taken by Collomb (1984) is only valid for bounded time series. Doukhan and Ghindes (1980) and Yakowitz (1985, 1987) obtained similar (pointwise) results in the context of density estimation and prediction for Markov sequences satisfying the G_2 condition, which is basically equivalent to the ϕ-mixing condition. Robinson (1983) established pointwise consistency and a central limit theorem under the α-mixing condition. In this paper, we address the problem on rates of convergence of local means under the (weaker) α-mixing condition. Note that the boundedness condition [Condition 4(ii)] is not required by Theorem 1 or 2. An interesting open problem is to verify the L_∞ rate of convergence in Theorem 3 when Condition 4(ii) is replaced by a weaker condition such as the following:

$$\sup_{\mathbf{x}\in U} E\big(\exp(t|Y_0|)\big) < \infty \quad \text{for some } t > 0.$$

In the problem of conditional median function estimation for iid observations, a consistency result was obtained in Stone (1977). Rates of convergence were considered by Härdle and Luckhaus (1984) and Truong (1989). In particular, the former considered the L_∞ rate of convergence for a class of robust nonparametric estimators, while the latter considered the problem of L_q, $1 \le q \le \infty$, rates of convergence for the local medians. In this paper, the above results are generalized to the estimation based on local medians involving dependent observations. Robust estimation was addressed by Collomb and Härdle (1986) on uniform consistency under ϕ-mixing and by Boente and Fraiman (1989, 1990) under α-mixing conditions. The class of estimators considered there did not include local medians. Robinson (1984) established a central limit theorem for the local M-estimators under the α-mixing condition.

REMARK 1. Since a sequence of independent random variables is also a stationary sequence, the rates of convergence established in Theorems 1–3 are in fact optimal; see Stone (1980, 1982).

REMARK 2. With a simple modification of Condition 4(iii), Theorems 1–3 are easily extended to yield rates of convergence for conditional quantile estimators.

2. Estimation of the conditional mean. Throughout this section, $\theta(\cdot)$ is the conditional mean function and $\hat{\theta}_n(\cdot)$ is the local average estimator of this function.

The proofs start with some Hölder-type inequalities for stationary sequences satisfying the α-mixing condition. Let $u(\cdot,\cdot)$ and $v(\cdot,\cdot)$ be real-valued, measurable functions on $\mathbb{R}^{d+1}$. Set $U = u(\mathbf{X}_i, Y_i)$, $V = v(\mathbf{X}_j, Y_j)$ and $\alpha = \alpha(|i-j|)$. Proofs of the following two results can be found on pages 277–278 of Hall and Heyde (1980).

LEMMA 1. *Suppose that* $|u(\cdot,\cdot)| < B_1$ *and* $|v(\cdot,\cdot)| < B_2$. *Then*

$$|E(UV) - E(U)E(V)| \le 4B_1B_2\alpha.$$

LEMMA 2. *Suppose that* $E|U|^p < \infty$, $E|V|^q < \infty$, *where* $p, q > 1$ *and* $p^{-1} + q^{-1} < 1$. *Then*

$$|E(UV) - E(U)E(V)| \le 8\|U\|_p\|V\|_q\alpha^{1-p^{-1}-q^{-1}}.$$

Given $\mathbf{x} \in C$, set $K_i = K_i(\mathbf{x}) = 1_{\{\|\mathbf{X}_i - \mathbf{x}\| \le \delta_n\}}$, $i = 1, \dots, n$. The next lemma is easily established.

LEMMA 3. *Suppose that Conditions* 2 *and* 3 *hold. Then there is a positive constant* c_1 *such that*

$$E(K_iK_{i+j}) \le \begin{cases} c_1\delta_n^{2d}, & \text{for } j > 0,\\ c_1\delta_n^{d}, & \text{for } j = 0.\end{cases}$$

LEMMA 4. *Suppose that Conditions* 2, 3 *and* 5(i) *hold. Then*

$$\operatorname{var}\Big(\sum_i K_i\Big) = O(n\delta_n^d).$$

PROOF. By Lemma 1, $|\operatorname{cov}(K_iK_{i+j})| \le 4\alpha(j)$. Thus by Condition 5(i) and Lemma 3,

$$\begin{aligned}\operatorname{var}\Big(\sum_i K_i\Big) &= n\operatorname{var}(K_1) + 2\sum_i\sum_j \operatorname{cov}(K_i, K_{i+j})\\ &= O\bigg(n\delta_n^d + n\sum_1^n \min\big(\alpha(j), \delta_n^{2d}\big)\bigg) = O(n\delta_n^d),\end{aligned}$$

as desired. □

The following result follows from Chebyshev's inequality and Lemma 4.

LEMMA 5. *Suppose that Conditions* 2, 3 *and* 5(i) *hold. If* $\delta_n \sim n^{-r}$, *then there is a positive constant* c_2 *such that*

$$\lim_n P\Big(\sum_i K_i \le c_2 n\delta_n^d\Big) = 0.$$

LEMMA 6. *Suppose that Conditions* 2, 3, 4(i) *and* 5(ii) *hold. Then*

$$\operatorname{var}\Big(\sum_i K_i[Y_i - \theta(\mathbf{X}_i)]\Big) = O(n\delta_n^d).$$

PROOF. Set $W_i = Y_i - \theta(\mathbf{X}_i)$. Applying Hölder's inequality twice,

$$E\left(K_i|W_i|K_{i+j}|W_{i+j}|\right)$$
$$(2.1)\qquad = E\left[\left(K_i|W_i|^\nu\right)^{1/\nu}\left(K_{i+j}|W_{i+j}|^\nu\right)^{1/\nu}\left(K_iK_{i+j}\right)^{1-(2/\nu)}K_i^{1/\nu}K_{i+j}^{1/\nu}\right]$$
$$\leq \left\{E\left[K_i|W_i|^\nu\right]\right\}^{2/\nu}\left\{E\left[K_iK_{i+j}\right]\right\}^{1-(2/\nu)}.$$

By Lemma 2,

$$(2.2)\qquad \left|E\left(K_iW_iK_{i+j}W_{i+j}\right)\right| \leq 8\left\{E\left(K_i|W_i|^\nu\right)\right\}^{2/\nu}\left\{\alpha(j)\right\}^{1-(2/\nu)}.$$

According to Condition 2,

$$E\left(K_i|W_i|^s\right) = E\left(K_iE\left(|W_i|^s|\mathbf{X}_i\right)\right)$$
$$(2.3)\qquad \leq M_1 \sup_{\|\mathbf{y}\|\leq\delta_n} Q(\mathbf{y})\int K_i(\mathbf{z})\,d\mathbf{z} = O\left(\delta_n^d\right) \quad \text{for } 1 \leq s \leq \nu,$$

where $Q(\mathbf{y}) = E(|W_i|^s|\mathbf{X}_i = \mathbf{y})$ is bounded in $\mathbf{y} \in U$ by Condition 4. By (2.1)–(2.3), Lemma 3 and Condition 5(ii) [note that $E(W_i|\mathbf{X}_i) = 0$],

$$\operatorname{var}\left(\sum_i K_iW_i\right) = n\operatorname{var}(K_1W_1) + 2\sum_i\sum_j \operatorname{cov}(K_iW_i, K_{i+j}W_{i+j})$$
$$= O\left(n\delta_n^d + n\left(\delta_n^d\right)^{2/\nu}\sum_1^n \min\left\{\alpha^{1-(2/\nu)}(j), \left(\delta_n^{2d}\right)^{1-(2/\nu)}\right\}\right)$$
$$= O\left(n\delta_n^d\right),$$

which completes the proof of Lemma 6. □

LEMMA 7. *Suppose that Conditions* 2–4(i), 5(ii) *and* 5(iii) *hold. If* $\delta_n \sim n^{-r}$ *or* $\delta_n \sim (n^{-1}\log n)^r$, *then there is a positive constant* c_3 *such that*

$$\lim_n P\left(N_n(\mathbf{x}) \geq c_3 n\delta_n^d \textit{ for } \mathbf{x} \in C\right) = 1.$$

Under the assumption of independence, there are several known results than can be used to prove the above lemma: Vapnik and Cervonenkis inequality [see Theorem 12.2 of Breiman, Friedman, Olsen and Stone (1984)]; Bernstein's inequality [see Theorem 3 of Hoeffding (1963)]; Markov's inequality applied to sufficient high order moments; and Lemma 1 of Stone (1982). Collomb (1984) obtained a Bernstein-type inequality for dependent random variables satisfying the ϕ-mixing condition, which is stronger than α-mixing and is too restrictive for many applications. In particular, this ϕ-mixing condition is equivalent to m-dependence for stationary Gaussian time series. In what follows, we will present a proof of Lemma 7 under the α-mixing condition based on an inequality establised by Philipp (1982). (We thank Magda Peligrad for pointing out this result to us.)

LEMMA 8. *Let $\{\xi_j,\ j \geq 1\}$ be a strictly stationary sequence of real-valued random variables, centered at expectations and uniformly bounded by 1. Suppose that $\{\xi_j,\ j \geq 1\}$ is α-mixing and that $\sigma^2 = E\xi_1^2 + 2\Sigma_{j\geq 2}E\xi_1\xi_j < \infty$. Let c_4, c_5 and γ denote positive constants such that $0 < \gamma < 1/2$. Then for any $R > 0$,*

$$P\left(\left|\sum_{j\leq n}\xi_j\right| > Rn^{1/2}\right)$$

$$\leq \begin{cases} O\big(\exp(-c_4R^2/\sigma^2) + n\alpha([n^\gamma])(\sigma^{-4} + R^{-2})\big), & \text{if } R \leq \sigma^2\sqrt{n}/n^\gamma; \\ O\big(\exp(-c_5n\sigma^2/n^{2\gamma}) + n\alpha([n^\gamma])(\sigma^{-4} + R^{-2})\big), & \text{if } R > \sigma^2\sqrt{n}/n^\gamma. \end{cases}$$

PROOF. See Theorem 4 and Proposition 5.1 of Philipp (1982). □

PROOF OF LEMMA 7. We assume $C = [-1/2, 1/2]^d$. Write C as the disjoint union of M_n^d cubes $C_{n\alpha}$ with length of each side $\sim \delta_n$, where $M_n \sim \delta_n^{-1}$ and $\alpha = 1, \dots, M_n^d$. Set $K_{i\alpha} = 1_{\{\mathbf{X}_i \in C_{n\alpha}\}}$, $\mu = \mu_\alpha = E(K_{i\alpha}) \sim \delta_n^d$ and $N_{n\alpha} = \#\{i\colon 1 \leq i \leq n;\ \mathbf{X}_i \in C_{n\alpha}\} = \Sigma_i K_{i\alpha}$. Suppose that $\delta_n \sim n^{-r}$ or $(n^{-1}\log n)^r$. Then

$$\lim_n P\big(N_{n\alpha} \geq \tfrac{1}{2}M_1^{-1}n\delta_n^d \text{ for } \alpha = 1, \dots, M_n^d\big) = 1.$$

Indeed, set $V_i = V_{i\alpha} = K_{i\alpha} - \mu$ and $\sigma^2 = EV_1^2 + 2\Sigma_{j\geq 2}EV_1V_j$. Then, by Condition 5(ii) and the argument given in the proof of Lemma 4, $\Sigma_{j\geq 2}EV_1V_j = o(\delta_n^d)$. Thus $\sigma^2 \sim \delta_n^d$. According to the second inequality of Lemma 8 with $R = \sqrt{n}\,\mu/2$ and Condition 5(iii), there is a positive constant a_1 such that

$$P\big(N_{n\alpha} \leq \tfrac{1}{2}n\mu\big) = P\Big(\sum_i V_i \leq -\tfrac{1}{2}n\mu\Big)$$

$$\leq O\Big(\exp\big(-a_1n\delta_n^d/n^{2\gamma}\big) + np^{[n^\gamma]}\Big(\big(\delta_n^{2d}\big)^{-1} + 4\big(n\mu^2\big)^{-1}\Big)\Big).$$

The conclusion of the lemma follows easily from this result. □

PROOF OF THEOREM 1. According to Condition 1,

$$|\theta(\mathbf{X}_i) - \theta(\mathbf{x})| \leq M_0\delta_n \quad \text{for } i \in I_n(\mathbf{x}).$$

Set $I_n = I_n(\mathbf{x})$ and $N_n = N_n(\mathbf{x})$. Then

$$\left|N_n^{-1}\sum_{I_n}[\theta(\mathbf{X}_i) - \theta(\mathbf{x})]\right| = O_P(\delta_n). \tag{2.4}$$

On the other hand, by Lemma 5,

$$P\left(N_n^{-1}\left|\sum_{I_n}[Y_i-\theta(\mathbf{X}_i)]\right|\geq c\delta_n\right)$$
$$\leq P\left(N_n^{-1}\left|\sum_{I_n}[Y_i-\theta(\mathbf{X}_i)]\right|\geq c\delta_n;\, N_n>c_2n\delta_n^d\right)+P\left(N_n\leq c_2n\delta_n^d\right)$$
$$\leq P\left(\left|\sum_{I_n}[Y_i-\theta(\mathbf{X}_i)]\right|\geq c_2cn\delta_n^{d+1}\right)+o(1).$$

Since $n\delta_n^{d+1}\sim\delta_n^{-1}$ and $n\delta_n^d\sim\delta_n^{-2}$, it follows from Lemma 6 and Chebyshev's inequality that

$$\left|N_n^{-1}\sum_{I_n}[Y_i-\theta(\mathbf{X}_i)]\right|=O(\delta_n). \tag{2.5}$$

The conclusion of Theorem 1 follows from (2.4) and (2.5) □

Proof of Theorem 2. According to Condition 1,

$$|\theta(\mathbf{X}_i)-\theta(\mathbf{x})|\leq M_0\|\mathbf{X}_i-\mathbf{x}\|\leq M_0\delta_n\quad\text{for } i\in I_n(\mathbf{x})\text{ and }\mathbf{x}\in C.$$

Thus there is a positive constant c_6 such that

$$\lim_n P\left(\left|N_n(\mathbf{x})^{-1}\sum_{I_n(\mathbf{x})}[\theta(\mathbf{X}_i)-\theta(\mathbf{x})]\right|\geq c_6\delta_n\text{ for some }\mathbf{x}\in C\right)=0. \tag{2.6}$$

Set $Z_n(\mathbf{x})=\sum_{i\in I_n(\mathbf{x})}[Y_i-\theta(\mathbf{X}_i)]$. By Lemma 6,

$$E\left[Z_n^2(\mathbf{x})\right]=O\left(n\delta_n^d\right)\quad\text{uniformly over }\mathbf{x}\in C.$$

Consequently,

$$E\left[\int_C|Z_n(\mathbf{x})|^2\,d\mathbf{x}\right]=\int_C E\left[|Z_n(\mathbf{x})|^2\right]d\mathbf{x}=O\left(n\delta_n^d\right). \tag{2.7}$$

By Lemma 7,

$$\lim_n P(\Omega_n)=1, \tag{2.8}$$

where $\Omega_n=\{N_n(\mathbf{x})\geq c_3n\delta_n^d\text{ for }\mathbf{x}\in C\}$. By (2.7) and (2.8),

$$\begin{aligned}
&P\left(\left\{\int_C\left|N_n(\mathbf{x})^{-1}\sum_{I_n(\mathbf{x})}[Y_i-\theta(\mathbf{X}_i)]\right|^2d\mathbf{x}\right\}^{1/2}\geq c\left(n^{-1}\delta_n^{-d}\right)^{1/2}\right)\\
&\quad\leq P(\Omega_n^c)+P\left(\int_C|Z_n(\mathbf{x})|^2\,d\mathbf{x}\geq c^2c_3^2n\delta_n^d\right)\\
&\quad=P(\Omega_n^c)+\frac{O(1)n\delta_n^d}{c^2n\delta_n^d}=o(1)\quad\text{as }n,c\to\infty.
\end{aligned} \tag{2.9}$$

It follows from (2.6) and (2.9) that

$$\lim_{c\to\infty}\lim_n P\Big(\big\|\hat\theta_n-\theta\big\|_2\ge c\Big(\delta_n+\big(n^{-1}\delta_n^{-d}\big)^{1/2}\Big)\Big)=0.$$

The conclusion of Theorem 2 now follows by choosing δ_n so that $\delta_n = (n^{-1}\delta_n^{-d})^{1/2}$, or equivalently, $\delta_n = n^{-r}$. □

Proof of Theorem 3. We can assume $C=[-1/2,1/2]^d\subset U$. Let s be a positive constant such that $0<s<1$ and set $L_n=[\delta_n^{-(2+s)}\log n]$. Let W_n be the collection of $(2L_n+1)^d$ points in C each of whose coordinates is of the form $j/(2L_n)$ for some integer j such that $|j|\le L_n$. Then C can be written as the union of $(2L_n)^d$ subcubes, each having length (of each side) $2\lambda_n=(2L_n)^{-1}$ and all of its vertices in W_n. For each $\mathbf{x}\in C$, there is a subcube $Q_{\mathbf{w}}$ with center $\mathbf{w}$ such that $\mathbf{x}\in Q_{\mathbf{w}}$. Let C_n denote the collection of centers of these subcubes. Then

$$P\Big(\sup_{\mathbf{x}\in C}\big|\hat\theta_n(\mathbf{x})-\theta(\mathbf{x})\big|\ge c\big(n^{-1}\log n\big)^r\Big)$$
$$=P\Big(\max_{\mathbf{w}\in C}\sup_{\mathbf{x}\in Q_{\mathbf{w}}}\big|\hat\theta_n(\mathbf{x})-\theta(\mathbf{x})\big|\ge c\big(n^{-1}\log n\big)^r\Big).$$

It follows from $\lambda_n\sim\delta_n^{2+s}/\log n=o(\delta_n)$ and Condition 1 that (for n sufficiently large)

$$|\theta(\mathbf{x})-\theta(\mathbf{w})|\le M_0\|\mathbf{x}-\mathbf{w}\|\le M_0\delta_n\quad\text{for }\mathbf{x}\in Q_{\mathbf{w}},\ \mathbf{w}\in C_n.$$

Therefore, to prove the theorem, it is sufficient to show that there is a positive constant c such that

$$\lim_n P\Big(\max_{\mathbf{w}\in C_n}\sup_{\mathbf{x}\in Q_{\mathbf{w}}}\big|\hat\theta_n(\mathbf{x})-\theta(\mathbf{w})\big|\ge c\big(n^{-1}\log n\big)^r\Big)=0. \tag{2.10}$$

Set $\bar I_n=\bar I_n(\mathbf{w})=\{i\colon 1\le i\le n \text{ and } \|\mathbf{X}_i-\mathbf{w}\|\le\delta_n+\lambda_n\sqrt{d}\}$, $\bar N_n=\bar N_n(\mathbf{w})=\#\bar I_n(\mathbf{w})$ and $\bar\theta_n(\mathbf{w})=\operatorname{ave}\{Y_i\colon i\in\bar I_n(\mathbf{w})\}$, $\mathbf{w}\in C_n$. Then (2.10) follows from

$$\lim_n P\Big(\max_{\mathbf{w}\in C_n}\sup_{\mathbf{x}\in Q_{\mathbf{w}}}\big|\hat\theta_n(\mathbf{x})-\bar\theta_n(\mathbf{w})\big|\ge c\big(n^{-1}\log n\big)^r/2\Big)=0 \tag{2.11}$$

and

$$\lim_n P\Big(\max_{\mathbf{w}\in C_n}\big|\bar\theta_n(\mathbf{w})-\theta(\mathbf{w})\big|\ge c\big(n^{-1}\log n\big)^r/2\Big)=0. \tag{2.12}$$

To verify (2.11) and (2.12), set $\underline{N}_n=\underline{N}_n(\mathbf{w})=\#\{i\colon\|\mathbf{X}_i-\mathbf{w}\|\le\delta_n-\lambda_n\sqrt{d}\}$. By Conditions 2–5 and Lemma 8 there are positive constants c_7 and c_8 such that

$$\lim_n P(\Psi_n)=1, \tag{2.13}$$

where $\Psi_n=\{\bar N_n(\mathbf{w})-\underline{N}_n(\mathbf{w})\le c_7\delta_n^{-1+s}$ and $\bar N_n(\mathbf{w})\ge c_8 n\delta_n^d$ for all $\mathbf{w}\in C_n\}$.

Indeed, note that $\bar{N}_n - \underline{N}_n = \#\{i: \delta_n - \lambda_n\sqrt{d} \le \|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d}\}$ is a sum of n Bernoulli random variables with probability of success $\pi_n = P(\delta_n - \lambda_n\sqrt{d} \le \|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d})$. By Condition 2,

$$\pi_n \sim (\delta_n + \lambda_n\sqrt{d})^d - (\delta_n - \lambda_n\sqrt{d})^d \sim \delta_n^{d-1}\lambda_n \quad \text{for } n \text{ sufficiently large}.$$

It follows from $n\delta_n^{d+2} \sim \log n$ and $\lambda_n \sim \delta_n^{2+s}/\log n$ that $n\pi_n \sim \delta_n^{-1+s} \to \infty$ as $n \to \infty$. Thus by Condition 5(ii) and the second inequality of Lemma 8 (with $\sigma^2 \sim \pi_n$, $R^2 \sim n\pi_n^2$), there is a positive constant c_9 such that

$$\begin{aligned} &P(\bar{N}_n(\mathbf{w}) - \underline{N}_n(\mathbf{w}) \ge 2n\pi_n \text{ for some } \mathbf{w} \in C_n) \\ &\quad = [2L_n]^d O\left(\exp\left(-c_9\frac{n\pi_n}{n^{2\gamma}}\right) + n\alpha([n^\gamma])\left(\frac{1}{\pi_n^2} + \frac{1}{n\pi_n^2}\right)\right) \\ &\quad = o(1) \quad \text{as } n \to \infty \end{aligned}$$

for $\gamma < (1-s)r/2$. Similarly,

$$\lim_n P(\bar{N}_n(\mathbf{w}) \le \tfrac{1}{2}np_n(\mathbf{w}) \text{ for some } \mathbf{w} \in C_n) = 0,$$

where $p_n(\mathbf{w}) = P(\|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d}) \sim \delta_n^d$. Thus (2.13) is proven.

It follows from the boundedness of Y_i and the first inequality of Lemma 8 (with $\gamma < r$, $\sigma^2 \sim \delta_n^d$ and $R^2 = c^2c_8^2n\delta_n^{2d+2}$) that there is a positive constant c_{10} such that

$$\begin{aligned} &P\left(\left|\sum_{\bar{I}_n(\mathbf{w})}[Y_i - \theta(\mathbf{X}_i)]\right| \ge cc_8n\delta_n^{d+1}\right) \\ &\quad = O(1)\exp(-c_{10}c^2n\delta_n^{d+2}) + O(1)\left[n\rho^{[n^\gamma]}\left(\frac{1}{\delta_n^{2d}} + \frac{1}{\delta_n^d\log n}\right)\right]. \end{aligned}$$

Note that there is a positive constant κ such that $\#C_n \le n^\kappa$. According to (2.13),

$$\begin{aligned} &P\left(\max_{\mathbf{w}\in C_n}\left|\bar{N}_n(\mathbf{w})^{-1}\sum_{\bar{I}_n(\mathbf{w})}[Y_i - \theta(\mathbf{X}_i)]\right| \ge c\delta_n\right) \\ &\quad \le P(\Psi_n^c) + P\left(\max_{\mathbf{w}\in C_n}\left|\sum_{\bar{I}_n(\mathbf{w})}[Y_i - \theta(\mathbf{X}_i)]\right| \ge cc_8n\delta_n^{d+1}\right) \\ &\quad = o(1) + O(1)n^\kappa\exp(-c^2c_{10}n\delta_n^{d+2}) + 2n^{\kappa+2}O(\rho^{n^\gamma}). \end{aligned}$$

Since $n\delta_n^{d+2} \sim \log n$, we conclude that for c sufficiently large,

$$\begin{aligned} &P\left(\max_{\mathbf{w}\in C_n}\left|\bar{N}_n(\mathbf{w})^{-1}\sum_{\bar{I}_n(\mathbf{w})}[Y_i - \theta(\mathbf{X}_i)]\right| \ge c\delta_n\right) \\ &\quad \le O(1)n^\kappa\exp(-c^2\log n) + o(1). \end{aligned}$$

Consequently, for $c^2 > \kappa$,

$$\lim_n P\left(\max_{\mathbf{w} \in C_n} \left| \overline{N}_n(\mathbf{w})^{-1} \sum_{\bar{I}_n(\mathbf{w})} [Y_i - \theta(\mathbf{X}_i)] \right| \geq c\delta_n \right) = 0. \tag{2.14}$$

Observe that (2.12) follows from (2.6) and (2.14).

Given $\mathbf{x} \in C$, set $N_n = N_n(\mathbf{x})$ and $I_n = I_n(\mathbf{x})$ and choose $\mathbf{w}$ such that $\mathbf{x} \in Q_{\mathbf{w}}$. Then $\underline{N}_n \leq N_n \leq \overline{N}_n$ and

$$\frac{\Sigma_{\bar{I}_n} Y_i}{\overline{N}_n} - \frac{\Sigma_{I_n} Y_i}{N_n} = \frac{N_n \Sigma_{\bar{I}_n \setminus I_n} Y_i - (\overline{N}_n - N_n) \Sigma_{I_n} Y_i}{\overline{N}_n N_n}.$$

Thus

$$\left| \frac{\Sigma_{\bar{I}_n} Y_i}{\overline{N}_n} - \frac{\Sigma_{I_n} Y_i}{N_n} \right| \leq \frac{(\overline{N}_n - \underline{N}_n)}{\overline{N}_n} \max_{\bar{I}_n \setminus I_n} |Y_i| + \frac{(\overline{N}_n - \underline{N}_n)}{\overline{N}_n} \max_{\bar{I}_n} |Y_i|$$

and hence

$$\left| \frac{\Sigma_{\bar{I}_n} Y_i}{\overline{N}_n} - \frac{\Sigma_{I_n} Y_i}{N_n} \right| \leq 2 \frac{(\overline{N}_n - \underline{N}_n)}{\overline{N}_n} \max_{\bar{I}_n} |Y_i|.$$

Consequently, (2.11) follows from (2.13) and the boundedness of $\{Y_i\}$. □

3. Estimation of the conditional median. Throughout this section, $\theta(\cdot)$ is the conditional median function and $\hat{\theta}_n(\cdot)$ is the local median estimator of this function.

PROOF OF THEOREM 1. By symmetry, it suffices to show that

$$\lim_{c \to \infty} \limsup_n \left(\hat{\theta}_n(\mathbf{0}) > \theta(\mathbf{0}) + cn^{-r} \right) = 0. \tag{3.1}$$

Set $I_n = I_n(\mathbf{0})$. It follows from Condition 1 that $\theta(\mathbf{X}_i) \leq \theta(\mathbf{0}) + M_0 \delta_n$ for $i \in I_n$. Thus

$$\tfrac{1}{2} - P(Y_i \geq \theta(\mathbf{0}) + c\delta_n | \mathbf{X}_i) \geq P(0 \leq Y_i - \theta(\mathbf{X}_i) \leq (c - M_0)\delta_n | \mathbf{X}_i), \qquad i \in I_n.$$

Hence by Condition 4(iii), there is a positive constant c_0 such that if $c > M_0$, then

$$\tfrac{1}{2} - P(Y_i \geq \theta(\mathbf{0}) + c\delta_n | \mathbf{X}_i) \geq (c - M_0) c_0 \delta_n, \qquad n \gg 1 \text{ and } i \in I_n. \tag{3.2}$$

Set

$$Z_i = 1_{\{Y_i \geq \theta(\mathbf{0}) + c\delta_n\}} - P(Y_i \geq \theta(\mathbf{0}) + c\delta_n | \mathbf{X}_i).$$

Then

$$E\left[\sum_{I_n} Z_i \right] = 0$$

and, by an argument analogous to that given in the proof of Lemma 6 (see also Lemma 4),

$$\operatorname{var}\Big(\sum_{I_n^*} Z_i\Big) = O(n\delta_n^d).$$

Let $c > M_0$. Then by (3.2),

$$\tfrac{1}{2} - N_n^{-1} \sum_{I_n} P(Y_i \ge \theta(\mathbf{0}) + c\delta_n | \mathbf{X}_i) \ge (c - M_0)c_0\delta_n, \qquad n \gg 1.$$

It now follows from (3.2) and Lemma 5 that, for some $c_1 > 0$ and $n \gg 1$,

$$\begin{aligned}
P\big(\hat{\theta}_n(\mathbf{0}) \ge \theta(\mathbf{0}) + c\delta_n\big) &\le P\Big(N_n^{-1} \sum_{I_n} 1_{\{Y_i \ge \theta(\mathbf{0}) + c\delta_n\}} \ge \tfrac{1}{2}\Big)\\
&= P\Big(N_n^{-1} \sum_{I_n} Z_i \ge \tfrac{1}{2} - N_n^{-1} \sum_{I_n} P(Y_i \ge \theta(\mathbf{0}) + c\delta_n | \mathbf{X}_i)\Big)\\
&\le P\Big(N_n^{-1} \sum_{I_n} Z_i \ge (c - M_0)c_0\delta_n\Big)\\
&\le P\Big(N_n^{-1} \sum_{I_n} Z_i \ge (c - M_0)c_0\delta_n ; N_n \ge c_1 n\delta_n^d\Big)\\
&\quad + P(N_n < c_1 n \delta_n^d)\\
&\le P\Big(\sum_{I_n} Z_i \ge (c - M_0)c_0 c_1 n \delta_n^{d+1}\Big) + o(1).
\end{aligned}$$

Since $n\delta_n^{d+2} \sim 1$, (3.1) now follows from Chebyshev's inequality. This completes the proof of Theorem 1. □

The proof of Theorem 2 depends on Theorem 3, which will be considered next.

PROOF OF THEOREM 3. We can assume that $C = [-1/2, 1/2]^d \subset U$. Set $L_n = [n^{2r}]$. Let W_n be the collection of $(2L_n + 1)^d$ points in C each of whose coordinates is of the form $j/(2L_n)$ for some integer j such that $|j| \le L_n$. Then C can be written as the union of $(2L_n)^d$ subcubes, each having length $2\lambda_n = (2L_n)^{-1}$ and all of its vertices in W_n. For each $\mathbf{x} \in C$, there is a subcube $Q_{\mathbf{w}}$ with center $\mathbf{w}$ such that $\mathbf{x} \in Q_{\mathbf{w}}$. Let C_n denote the collection of the centers of these subcubes. Then

$$\begin{aligned}
&P\Big(\sup_{\mathbf{x} \in C} \big|\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})\big| \ge c\big(n^{-1}\log n\big)^r\Big)\\
&\quad = P\Big(\max_{\mathbf{w} \in C_n} \sup_{\mathbf{x} \in Q_{\mathbf{w}}} \big|\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})\big| \ge c\big(n^{-1}\log n\big)^r\Big).
\end{aligned}$$

It follows from $\lambda_n \sim n^{-2r}$ and Condition 1 that (for n sufficiently large)

$$|\theta(\mathbf{x}) - \theta(\mathbf{w})| \le M_0\|\mathbf{x} - \mathbf{w}\| \le M_0\delta_n \quad \text{for } \mathbf{x} \in Q_{\mathbf{w}}, \mathbf{w} \in C_n.$$

Therefore, to prove the theorem, it is sufficient to show that there is a positive constant c such that

$$\lim_n P\Big(\max_{\mathbf{w}\in C_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} \big|\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})\big| \ge c(n^{-1}\log n)^r\Big) = 0. \tag{3.3}$$

Given $\mathbf{x} \in Q_{\mathbf{w}}$, set $\underline{N}_n = \underline{N}_n(\mathbf{w}) = \#\{i\colon \|\mathbf{X}_i - \mathbf{w}\| \le \delta_n - \lambda_n\sqrt{d}\}$. It follows from $N_n = N_n(\mathbf{x}) = \#\{i\colon \|\mathbf{X}_i - \mathbf{x}\| \le \delta_n\} \ge \underline{N}_n$ for $\mathbf{x} \in Q_{\mathbf{w}}$ that

$$\begin{aligned}\big\{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \ge c\delta_n\big\} &\subseteq \Big\{N_n^{-1}\sum_{I_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}\Big\} \\ &\subseteq \Big\{\sum_{\bar{I}_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}\underline{N}_n\Big\},\end{aligned}$$

where $\bar{I}_n = \bar{I}_n(\mathbf{w}) = \{i\colon 1 \le i \le n \text{ and } \|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d}\}$. Thus

$$\bigcup_{Q_{\mathbf{w}}} \big\{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \ge c\delta_n\big\} \subseteq \Big\{\sum_{\bar{I}_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}\underline{N}_n\Big\}. \tag{3.4}$$

Set $\bar{N}_n = \bar{N}_n(\mathbf{w}) = \#\bar{I}_n(\mathbf{w})$. By Conditions 2, 3 and 5(iii) and Lemma 8, there are positive constants c_2 and c_3 such that

$$\lim_n P(\Psi_n) = 1, \tag{3.5}$$

where $\Psi_n = \{\bar{N}_n(\mathbf{w}) - \underline{N}_n(\mathbf{w}) \le c_2 n\delta_n^{d-1}\lambda_n \text{ and } \bar{N}_n(\mathbf{w}) \ge c_3 n\delta_n^d \text{ for all } \mathbf{w} \in C_n\}$.

Note that $n\delta_n^{d-1}\lambda_n\bar{N}_n^{-1} = O(\lambda_n/\delta_n) = o(\delta_n)$ on Ψ_n. It follows from (3.4) that there is a positive constant c_4 such that

$$\begin{aligned}&P\Big(\max_{\mathbf{w}\in C_n}\sup_{\mathbf{x}\in Q_{\mathbf{w}}}\big[\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})\big] \ge c\delta_n\Big) \\ &\quad \le P\Big(\bigcup_{C_n}\bigcup_{Q_{\mathbf{w}}}\big\{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \ge c\delta_n\big\}\Big) \\ &\quad \le P\Big(\bigcup_{C_n}\Big\{\sum_{\bar{I}_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}\underline{N}_n\Big\}\Big) \\ &\quad \le P\Big(\bigcup_{C_n}\Big\{\sum_{\bar{I}_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}\bar{N}_n - \tfrac{1}{2}c_2 n\delta_n^{d-1}\lambda_n\Big\} \cap \Psi_n\Big) + P(\Psi_n^c) \\ &\quad \le P\Big(\bigcup_{C_n}\Big\{\bar{N}_n^{-1}\sum_{\bar{I}_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2} - c_4\delta_n\Big\}\Big) + P(\Psi_n^c).\end{aligned} \tag{3.6}$$

According to Condition 1, $\theta(\mathbf{X}_i) \le \theta(\mathbf{w}) + M_0(\delta_n + \lambda_n\sqrt{d})$ whenever $\|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d}$. Thus

$$\begin{aligned}&\tfrac{1}{2} - P\big(Y_i \ge \theta(\mathbf{w}) + c\delta_n|\mathbf{X}_i\big)\\ &\qquad \ge P\big(0 \le Y_i - \theta(\mathbf{X}_i) \le (c - M_0)\delta_n - M_0\lambda_n\sqrt{d}\,|\mathbf{X}_i\big), \qquad i \in \bar{I}_n.\end{aligned}$$

By Condition 4(iii), there is a positive constant c_5 such that for $c \ge 2M_0$,

$$\tfrac{1}{2} - P\big(Y_i \ge \theta(\mathbf{w}) + c\delta_n|\mathbf{X}_i\big) \ge cc_5\delta_n, \qquad n \gg 1 \text{ and } i \in \bar{I}_n. \tag{3.7}$$

Thus, (3.7) implies

$$\tfrac{1}{2} - \bar{N}_n^{-1} \sum_{\bar{I}_n} P\big(Y_i \ge \theta(\mathbf{w}) + c\delta_n|\mathbf{X}_i\big) \ge cc_5\delta_n, \qquad n \gg 1. \tag{3.8}$$

Set $Z_i = 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} - P(Y_i \ge \theta(\mathbf{w}) + c\delta_n|\mathbf{X}_i)$. It now follows from (3.8) that there are positive constants c_6 and κ such that for $cc_5 > 2c_4$ and $n \gg 1$,

$$P\left(\bigcup_{C_n}\left\{\bar{N}_n^{-1}\sum_{\bar{I}_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2} - c_4\delta_n\right\}\right)$$

$$\begin{aligned}&= P\left(\bigcup_{C_n}\left\{\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \ge \tfrac{1}{2} - \bar{N}_n^{-1}\sum_{\bar{I}_n} P\big(Y_i \ge \theta(\mathbf{w}) + c\delta_n|\mathbf{X}_i\big) - c_4\delta_n\right\}\right)\\ &\le n^{\kappa}\max_{C_n} P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \ge cc_5\delta_n - c_4\delta_n\right)\\ &\le n^{\kappa}\max_{C_n} P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \ge cc_6\delta_n\right).\end{aligned} \tag{3.9}$$

Set $p_n = p_n(\mathbf{w}) = P(\|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d})$ (which, by stationarity, does not depend on i). Then $p_n \sim \delta_n^d$. Note that $\sum_{\bar{I}_n} Z_i = \sum_i K_i Z_i$ and $E(K_i Z_i) = 0$. By Lemma 6, $\operatorname{var}(\sum_i K_i Z_i) = O(n\delta_n^d)$. It follows from $\alpha(n) = O(\rho^n)$ and a double application of Lemma 8 (with $\gamma < r$, $\sigma^2 \sim \delta_n^d$, $R^2 = M_1^{-1}n\delta_n^{2d}$ and $R^2 = M_1^{-1}c^2c_6^2 n\delta_n^{2d+2}$, respectively) that there are positive constants c_7 and c_8 such that

$$\begin{aligned}P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \ge cc_6\delta_n\right) &\le P\big(\bar{N}_n < \tfrac{1}{2}np_n\big) + P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \ge cc_6\delta_n;\, \bar{N}_n \ge \tfrac{1}{2}np_n\right)\\ &\le \exp\big(-c_7 n\delta_n^d/n^{2\gamma}\big) + \exp\big(-c^2c_8 n\delta_n^{d+2}\big)\\ &\quad + O(1)\left[n\rho^{[n^{\gamma}]}\left(\frac{1}{\delta_n^{2d}} + \frac{1}{\delta_n^d\log n}\right)\right] \quad \text{for } \mathbf{w} \in C_n.\end{aligned}$$

Now it follows from $n\delta_n^{d+2} \sim \log(n)$ that there is a positive constant c such that

$$n^{\kappa}\max_{C_n} P\left(\bar{N}_n^{-1}\sum_{\bar{I}_n} Z_i \ge cc_6\delta_n\right) \to 0 \quad \text{as } n \to \infty. \tag{3.10}$$

Hence by (3.5), (3.6), (3.9) and (3.10),

$$\lim_n P\Big(\max_{C_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} \big[\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})\big] \ge c\delta_n\Big) = 0 \quad \text{for } c > 0. \tag{3.11}$$

Similarly,

$$\lim_n P\Big(\max_{C_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} \big[\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})\big] \le -c\delta_n\Big) = 0 \quad \text{for } c > 0. \tag{3.12}$$

It follows from (3.11) and (3.12) that (3.3) is valid. This completes the proof of Theorem 3. □

The proof of Theorem 2 depends on the following result on bounds for the moments of sum of weakly dependent random variàbles. Let $\{\nu_n\}$ be a sequence of positive numbers such that $\nu_n \sim n^{-\gamma}$ for some $\gamma \in (0,1)$.

LEMMA 9. *Let $V_{n1},\dots,V_{nn}$ be uniformly bounded random variables such that V_{ni} has mean zero and is a function of $\mathbf{X}_i$. Suppose that $E|V_{ni}| \le \nu_n$ and $E|V_{ni}V_{nj}| \le \nu_n^2$ for $1 \le i < j \le n$. Suppose $\alpha(N) = O(\rho^N)$, $N = 1,2,\dots$ and let k be a positive integer. Then*

$$E\Big[\Big(\sum_i V_{ni}\Big)^k\Big] = O\big((n\nu_n)^{k/2}\big) \quad \text{as } n \to \infty.$$

PROOF. In the following discussion, write V_i for V_{ni}. We may assume that $|V_i| \le 1$. Observe that

$$E\Big[\Big(\sum_i V_i\Big)^k\Big] \le k!\sum{}'\sum{}''\big|E\big(V_{i_1}^{\tau_1}\cdots V_{i_1+\cdots+i_t}^{\tau_t}\big)\big|, \tag{3.13}$$

where the indices in the first sum Σ' on the right side of (3.13) are on values of $t, \tau_1,\dots,\tau_t$ constrained by $\tau_1,\dots,\tau_t > 0$ and $\tau_1 + \cdots + \tau_t = k$ and the indices in the second sum Σ'' are on values of $i_1,\dots,i_t$ constrained by $i_1,\dots,i_t > 0$ and $i_1 + \cdots + i_t < n$. Let N be a positive integer less than than n. Partition the second sum in (3.13) into a finite number of sums such that the indices in each of these sums are constrained by: certain of the indices are larger than N and all others are less than equal to N. More precisely, let $\psi_t = (\phi_1,\dots,\phi_t)$ be a t-tuple of 0's and 1's and let $\Sigma_{\psi_t}|E(V_{i_1}^{\tau_1}\cdots V_{i_1+\cdots+i_t}^{\tau_t})|$ mean that (a) if $\phi_l = 1$, then the index i_l in the sum ranges over $N+1,\dots,n$; (b) if $\phi_l = 0$, then the index i_l in the sum ranges over $1,\dots,N$. Thus

$$\sum{}''\big|E\big(V_{i_1}^{\tau_1}\cdots V_{i_1+\cdots+i_t}^{\tau_t}\big)\big| = \sum_{\text{all }\psi_t}\sum_{\psi_t}\big|E\big(V_{i_1}^{\tau_1}\cdots V_{i_1+\cdots+i_t}^{\tau_t}\big)\big|. \tag{3.14}$$

Let ψ_t be fixed. By induction on m, where $m = \tau_1 + \cdots + \tau_t$,

$$\sum_{\psi_t}\big|E\big(V_{i_1}^{\tau_1}\cdots V_{i_1+\cdots+i_t}^{\tau_t}\big)\big| = O\big((n\nu_n)^{m/2}\big). \tag{3.15}$$

Indeed, (3.15) is valid for $m = 1,2$. $[\Sigma_{i,j}|E(V_iV_j)| = O(n\Sigma_i\min(\alpha(i),\nu_n^2)) =$

$O(n\nu_n)$.] Suppose $m > 2$ and assume that (3.15) holds for $\tau_1, \ldots, \tau_t$ with $\tau_1 + \cdots + \tau_t \le m - 1$. Set $N = [m\gamma^{-1}(\gamma + 1)\log \nu_n/(2\log\rho)]$. Suppose that $\phi_j = 0$ for $2 \le j \le t$. Then, since $m > 2$ and $|V_i| \le 1$ for $i = 1, \ldots, n$,

$$\begin{aligned}\sum_{\psi_t} \left|E\left(V_{i_1}^{\tau_1} \cdots V_{i_1+\cdots+i_t}^{\tau_t}\right)\right| &\le N^{t-1} n\nu_n \\ &= O\left((\log n)^t\right) n\nu_n \\ &= o\left((n\nu_n)^{(m/2)-1}\right) n\nu_n = o\left((n\nu_n)^{m/2}\right).\end{aligned}$$

Suppose instead $\phi_j = 1$ for some j such that $2 \le j \le t$. Set $b = \min\{j\colon 2 \le j \le t, \phi_j = 1\}$. Since the V_i's are bounded by 1, it follows from Lemma 1 that

$$\begin{aligned}&\left|E\left(V_{i_1}^{\tau_1} \cdots V_{i_1+\cdots+i_{b-1}}^{\tau_{b-1}} V_{i_1+\cdots+i_b}^{\tau_b} \cdots V_{i_1+\cdots+i_t}^{\tau_t}\right)\right| \\ &\qquad \le \left|E\left(V_{i_1}^{\tau_1} \cdots V_{i_1+\cdots+i_{b-1}}^{\tau_{b-1}}\right)\right| \left|E\left(V_{i_1+\cdots+i_b}^{\tau_b} \cdots V_{i_1+\cdots+i_t}^{\tau_t}\right)\right| + 4\alpha(i_b).\end{aligned}$$

Consequently, by the inductive hypothesis,

$$\begin{aligned}&\sum_{\psi_t} \left|E\left(V_{i_1}^{\tau_1} \cdots V_{i_1+\cdots+i_{b-1}}^{\tau_{b-1}} V_{i_1+\cdots+i_b}^{\tau_b} \cdots V_{i_1+\cdots+i_t}^{\tau_t}\right)\right| \\ &\quad \le \sum_{\psi_t} \left|E\left(V_{i_1}^{\tau_1} \cdots V_{i_1+\cdots+i_{b-1}}^{\tau_{b-1}}\right)\right| \left|E\left(V_{i_1+\cdots+i_b}^{\tau_b} \cdots V_{i_1+\cdots+i_t}^{\tau_t}\right)\right| + 4\sum_{\psi_t} \alpha(i_b) \\ &\quad = O\left((n\nu_n)^{(\tau_1+\cdots+\tau_{b-1})/2}\right) O\left((n\nu_n)^{(\tau_b+\cdots+\tau_t)/2}\right) + 4n^{t-1} \sum_{i>N} \alpha(i) \\ &\quad = O\left((n\nu_n)^{m/2}\right),\end{aligned}$$

for it follows from $N = [m\gamma^{-1}(\gamma + 1)\log \nu_n/(2\log\rho)]$ and $\sum_{i>N}\alpha(i) \sim \rho^N$ that (with $t \le m$)

$$n^t \sum_{i>N} \alpha(i) \le n^m \sum_{i>N} \alpha(i) \sim n^m \nu_n^{m(\gamma+1)/2\gamma} \sim (n\nu_n)^{m/2}.$$

This completes the proof of (3.15). The conclusion of the lemma follows from (3.13)–(3.15). □

Proof of Theorem 2. By Condition 1, $\theta(\cdot)$ is bounded on C (compact). Thus it follows from Theorem 3 that there is a positive constant $T > 1$ such that $\|\theta(\cdot)\| \le T$ and

$$\lim_n P(\Phi_n) = 1, \tag{3.16}$$

where $\Phi_n := \{\|\hat{\theta}_n(\cdot)\|_\infty \le T\}$. For $i = 1, \ldots, n$, set

$$Y_i' = \begin{cases} -T, & \text{if } Y_i < -T; \\ Y_i, & \text{if } |Y_i| \le T; \\ T, & \text{if } Y_i > T. \end{cases}$$

Set $\bar{\theta}_n(\mathbf{x}) = \operatorname{med}\{Y_i'\colon i \in I_n(\mathbf{x})\}$. Then $\bar{\theta}_n(\mathbf{x}) = \hat{\theta}_n(\mathbf{x})$ for $\mathbf{x} \in C$ except on Φ_n^c.

Together with (3.16), it is sufficient to prove the theorem by showing

$$\lim_{c\to\infty}\lim_n P\left(\|\bar{\theta}_n - \theta\|_q \ge cn^{-r}\right) = 0. \tag{3.17}$$

To verify (3.17), we may assume that $C = [-1/2, 1/2]^d \subset U$. According to $\alpha(n) = O(\rho^n)$ and Lemma 8 (see also the argument given in Lemma 7), there is a positive constant c_9 such that

$$\lim_n P(\Omega_n) = 1, \tag{3.18}$$

where $\Omega_n := \{N_n(\mathbf{x}) \ge c_9 n\delta_n^d \text{ for } \mathbf{x} \in C\}$.

Write $P_{\Omega_n}(\cdot) = P(\cdot\,;\Omega_n) = P(\cdot \cap \Omega_n)$ and $E_{\Omega_n}(W) = E(W1_{\Omega_n})$, where W is a real-valued random variable. By (3.18), there is a sequence of positive numbers $\varepsilon_n \to 0$ such that

$$\begin{aligned}
&P\left(\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\,d\mathbf{x} \ge (cn^{-r})^q\right)\\
&\quad\le P\left(\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\,d\mathbf{x} \ge (cn^{-r})^q;\Omega_n\right) + \varepsilon_n\\
&\quad\le \frac{E_{\Omega_n}\left[\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\,d\mathbf{x}\right]}{(cn^{-r})^q} + \varepsilon_n.
\end{aligned} \tag{3.19}$$

By Condition 1, $|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|$ is bounded by $2T$ for $\mathbf{x} \in C$. Thus there is a positive constant c_{10} such that

$$\begin{aligned}
E_{\Omega_n}\left[|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\right] &= \int_0^{2T} qt^{q-1}P_{\Omega_n}\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right)dt\\
&= \int_0^{2M_0\delta_n} qt^{q-1}P_{\Omega_n}\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right)dt\\
&\quad + \int_{2M_0\delta_n}^{2T} qt^{q-1}P_{\Omega_n}\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right)dt\\
&\le c_{10}\delta_n^q + \int_{2M_0\delta_n}^{2T} qt^{q-1}P_{\Omega_n}\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right)dt.
\end{aligned} \tag{3.20}$$

By Conditions 1–3, 4(iii) and 5(iii), there is a positive number c_{11} such that

$$\int_{2M_0\delta_n}^{2T} qt^{q-1}P_{\Omega_n}\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right)dt \le c_{11}\delta_n^q \quad \text{for } \mathbf{x} \in C. \tag{3.21}$$

[The proof of (3.21) will be given shortly.] It follows from (3.20) and (3.21) that there is a positive constant c_{12} such that

$$E_{\Omega_n}\left[|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\right] \le c_{12}\delta_n^q \quad \text{for } \mathbf{x} \in C.$$

Thus there is a positive constant c_{13} such that

$$E_{\Omega_n}\left[\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\,d\mathbf{x}\right] = \int_C E_{\Omega_n}\left[|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\right]d\mathbf{x} \le c_{13}\delta_n^q. \tag{3.22}$$

The conclusion of Theorem 3 follows from (3.19) and (3.22).

Finally, (3.21) will be proven. Let $\mathbf{x} \in C$ be fixed. By Condition 4(iii), there is a positive constant c_{14} such that

$$\tfrac{1}{2} - N_n^{-1} \sum_{I_n} P\big(Y_i \geq \theta(\mathbf{x}) + t|\mathbf{X}_i\big) \geq c_{14}t, \qquad 2M_0\delta_n \leq t \leq 2T.$$

Set

$$Z_i = 1_{\{Y_i \geq \theta(\mathbf{x})+t\}} - P\big(Y_i \geq \theta(\mathbf{x}) + t|\mathbf{X}_i\big).$$

Then (since $\{Y_i' > \theta(\mathbf{x}) + t\} \subset \{Y_i > \theta(\mathbf{x}) + t\}$)

$$\begin{aligned} &P_{\Omega_n}\big(\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) > t\big) \\ &\quad \leq P_{\Omega_n}\Big(N_n^{-1} \sum_{I_n} 1_{\{Y_i' > \theta(\mathbf{x})+t\}} \geq \tfrac{1}{2}\Big) \\ &\quad \leq P_{\Omega_n}\Big(N_n^{-1} \sum_{I_n} 1_{\{Y_i > \theta(\mathbf{x})+t\}} \geq \tfrac{1}{2}\Big) \\ &\quad \leq P_{\Omega_n}\Big(N_n^{-1} \sum_{I_n} Z_i \geq \tfrac{1}{2} - N_n^{-1} \sum_{I_n} P\big(Y_i \geq \theta(\mathbf{x}) + t|\mathbf{X}_i\big)\Big) \\ &\quad \leq P\Big(\sum_{I_n} Z_i \geq c_9 c_{14} t n \delta_n^d\Big). \end{aligned} \tag{3.23}$$

Set $K_i = K_i(\mathbf{x}) = 1_{\{\|\mathbf{X}_i - \mathbf{x}\| \leq \delta_n\}}$. Note that $\sum_{I_n} Z_i = \sum_i K_i Z_i$, $E(K_i Z_i) = 0$, $E|K_i Z_i| = O(\delta_n^d)$ and $E|K_i Z_i K_j Z_j| = O(\delta_n^{2d})$. Since Z_i is bounded, it follows from Lemma 9 that

$$E\left(\left|\sum_{I_n} Z_i\right|^{2k}\right) = E\left(\left|\sum_i K_i Z_i\right|^{2k}\right) = O\big(n\delta_n^d\big)^k \quad \text{for } k = 1, 2, 3, \ldots.$$

Consequently, by Markov's inequality,

$$\begin{aligned} P\Big(\sum_{I_n} Z_i \geq c_9 c_{14} t n \delta_n^d\Big) &\leq \frac{E\left|\sum_{I_n} Z_i\right|^{2k}}{\big(c_9 c_{14} t n \delta_n^d\big)^{2k}} \\ &= \frac{O\big(n\delta_n^d\big)^k}{\big(c_9 c_{14} t n \delta_n^d\big)^{2k}}, \qquad 2M_0\delta_n \leq t \leq 2T. \end{aligned} \tag{3.24}$$

By (3.23) and (3.24), there is a positive constant c_{15} such that (note that $n\delta_n^d \sim \delta_n^{-2}$)

$$P_{\Omega_n}\big(\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) > t\big) \leq c_{15} t^{-2k} \delta_n^{2k}, \qquad 2M_0\delta_n \leq t \leq 2T. \tag{3.25}$$

Similarly,

$$P_{\Omega_n}\big(\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x}) < -t\big) \leq c_{15} t^{-2k} \delta_n^{2k}, \qquad 2M_0\delta_n \leq t \leq 2T. \tag{3.26}$$

Note that c_{14} and c_{15} do not depend on $\mathbf{x}$. It now follows from (3.25) and (3.26) by choosing $k > q/2$ that

$$\int_{2M_0\delta_n}^{2T} t^{q-1} P_{\Omega_n}\big(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\big)\, dt \le 2\delta_n^{2k} c_{15} \int_{2M_0\delta_n}^{2T} t^{q-2k-1}\, dt = O(\delta_n^q). \quad \square$$

REMARK. Why is it necessary to use Lemma 9 to establish the above inequality, instead of using Lemma 8? The main reason is: For simplicity, suppose $t = 2M_0\delta_n$. Then the exponential inequality (from lemma 8) contains the term $\exp(-c^2 n\delta_n^{d+2}) = O(1)$, because $n\delta_n^{d+2} \sim 1$. [See the inequality before (3.10).] Consequently, that would not yield the desired result. However, the exponential inequality is useful for establishing the L_∞ convergent rates in that $\exp(-c^2 n\delta_n^{d+2}) \sim \exp(-c^2 \log(n))$ as δ_n is now chosen so that $n\delta_n^{d+2} \sim \log(n)$.

Acknowledgments. We wish to thank the editors and referees for helpful comments that improved the readability of the paper.

REFERENCES

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

BOENTE, G. and FRAIMAN, R. (1989). Robust nonparametric regression estimation for dependent variables. *Ann. Statist.* **17** 1242–1256.

BOENTE, G. and FRAIMAN, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.* **18** 891–906.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, Calif.

COLLOMB, B. G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. *Z. Wahrsch. Verw. Gebiete* **66** 441–460.

COLLOMB, B. 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.

DOUKHAN, P. and GHINDES, M. (1980). Estimations dans le processus $X_{n+1} = f(X_n) + \varepsilon_n$. *C. R. Acad. Sci. Paris Ser. A-B* **291** 61–64.

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

HÄRDLE, W. and LUCKHAUS, S. (1984). Uniform consistency of a class of regression function estimators. *Ann. Statist.* **12** 612–623.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

PHILIPP, W. (1982). Invariance principles for sums of mixing random elements and the multivariate empirical process. In *Limit Theorems in Probability and Statistics. Colloq. Math. Soc. János Bolyai* **36** 843–873.

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

ROBINSON, P. M. (1984). Robust nonparametric regression. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 247–255. Springer, New York.

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

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.

TRUONG, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** 606–617.

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

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

SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-7400

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720