The Annals of Statistics
1989, Vol. 17, No. 2, 606–617

ASYMPTOTIC PROPERTIES OF KERNEL ESTIMATORS BASED ON LOCAL MEDIANS

BY YOUNG K. TRUONG

University of North Carolina at Chapel Hill

The desire to make nonparametric regression robust leads to the problem of conditional median function estimation. Under appropriate regularity conditions, a sequence of local median estimators can be chosen to achieve the optimal rate of convergence $n^{-1/(2+d)}$ both pointwise and in the L^q $(1 \le q < \infty)$ norm restricted to a compact. It can also be chosen to achieve the optimal rate of convergence $(n^{-1} \log n)^{1/(2+d)}$ in the L^∞ norm restricted to a compact. These results also constitute an answer to an open question of Stone.

1. Introduction. Let $(\mathbf{X}, Y)$ be a pair of random variables which are, respectively, d and one dimensional, and let $\theta(\cdot)$ denote the conditional median of the response Y on the measurement variable $\mathbf{X}$, so that $\operatorname{med}(Y|\mathbf{X}) = \theta(\mathbf{X})$. Consider the problem of estimating the function $\theta(\cdot)$ based on a training sample from the distribution of $(\mathbf{X}, Y)$. Under appropriate regularity conditions, asymptotic properties (rates of convergence) of nonparametric estimators constructed by kernel methods based on local medians are established and these results address two issues: robustification in nonparametric regression estimation and an answer to a question of Stone (1982).

Nearest neighbor, kernel and recursive partition methods of nonparametric regression, as usually defined, are based on local averages. Recently, there has been an interest in adopting local medians (or some other robust methods such as M-estimates) as nonparametric regression estimates, especially when outliers may be present. The latter approach indeed offers an interesting alternative for modeling functional relationships between the responses and the measurement variables. For instance, it would be appropriate to estimate the regression function based on local means when the response variable has light-tailed and symmetric (conditional) distribution. But if the distribution (such as the distribution of housing values, annual incomes) is heavy-tailed or asymmetric, then local medians should be considered since they are highly resistant against outliers and the results based on them are easier to interpret. In a seminal work on nonparametric regression, Stone (1977) obtained a consistency theorem for nearest neighbor estimators based on local medians, and in the course of the discussion, Brillinger addressed the importance of conditional M-estimates. The usefulness of these procedures in exploratory data analysis was discussed by Härdle and Gasser (1984) and Härdle and Tsybakov (1988). Numerical examples

Received November 1987; revised September 1988.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20.
Key words and phrases. Kernel estimator, nonparametric regression, conditional median function, local median, rate of convergence.

on robust nonparametric regression for recursive partitioning based on local medians are given in Breiman, Friedman, Olshen and Stone (1984).

In addition to the above desirable robustness property, the current approach can also be given theoretical justification in terms of rates of convergence. In fact, it is shown in this paper that local medians and local averages have the same optimal asymptotic properties in regression-type problems.

The asymptotic results presented here also constitute an answer to an open question of Stone (1982). In the context of estimating the regression function $E(Y|\mathbf{X})$, Stone (1977) obtained a consistency theorem for a large class of nonparametric regression estimators and used this to establish the consistency of nearest neighbor estimators based on local averages. Since then, consistency has been established for kernel estimators by Devroye and Wagner (1980a, b) and Spiegelman and Sacks (1980), and for partition estimators by Gordon and Olshen (1980) and Breiman, Friedman, Olshen and Stone (1984). If the pth derivative of the regression function is bounded, then Stone (1980, 1982) showed that $\{n^{-p/(2p+d)}\}$ is the optimal rate of convergence in both pointwise and L^q $(1 \le q < \infty)$ norms, while $\{(n^{-1}\log n)^{p/(2p+d)}\}$ is the optimal rate of convergence in L^∞ norm. Under some regularity conditions, a sequence of kernel estimators based on local polynomials can be chosen to achieve the optimal rates of convergence. The desire to robustify nonparametric regression leads to Question 4 of Stone [(1982), page 1044]: Is $\{n^{-p/(2p+d)}\}$ still an achievable rate of convergence in estimating the conditional median?

Rates of convergence of conditional M-estimators have been considered by several authors. Under some regularity conditions and the boundedness assumption of the first derivative (i.e., $p = 1$), Härdle and Luckhaus (1984) presented the L^∞ rate of convergence for a class of robust nonparametric estimators including an estimator of the conditional median. Härdle (1984) and Härdle and Tsybakov (1988) considered the pointwise results for a class of local M-estimates that did not cover the local medians. The later paper also considered the pointwise joint estimation of the regression and scale functions. However, the problem of L^q $(1 \le q < \infty)$ rates of convergence was still unsolved.

Both the pointwise (local) and the L^q $(1 \le q \le \infty)$ (global) rates of convergence for kernel estimators based on local medians will be described in Section 2. For this class of nonparametric estimators, the results presented there settle an open question of Stone (1982) when the first derivative of the conditional median is bounded. Proofs of these results are given in Section 3 and include a different and more intuitive proof [compared to Härdle and Luckhaus (1984)] of the uniform rate of convergence.

2. Statement of results. Results on the local and global rates of convergence of nonparametric estimators of the conditional median, based on a random sample from the distribution of $(\mathbf{X}, Y)$, will be treated in this section.

Let U be a nonempty bounded open neighborhood of the origin of $\mathbf{R}^d$. Given $\mathbf{x} = (x_1, \dots, x_d) \in \mathbf{R}^d$, set $\|\mathbf{x}\| = (x_1^2 + \cdots + x_d^2)^{1/2}$. The kernel estimator $\hat{\theta}_n(\cdot)$

of $\theta(\cdot)$ will now be described. Given $n \geq 1$, let $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ denote a random sample of size n from the distribution of $(\mathbf{X}, Y)$. Let δ_n, $n \geq 1$, be positive numbers that tend to zero as n tends to infinity. For $\mathbf{x} \in U$, set $I_n(\mathbf{x}) = \{i\colon\ 1 \leq i \leq n \text{ and } \|\mathbf{X}_i - \mathbf{x}\| \leq \delta_n\}$, $N_n(\mathbf{x}) = \# I_n(\mathbf{x})$ and $\hat{\theta}_n(\mathbf{x}) = \operatorname{med}\{Y_i\colon i \in I_n(\mathbf{x})\}$ [use the average of the two middle order statistics if $N_n(\mathbf{x})$ is even].

The rates of convergence of the estimators treated here depend on the following smoothness condition on $\theta(\cdot)$.

Condition 1. There is a positive constant M_0 such that

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

(If U is convex, the above condition is implied by an appropriate boundedness condition on the restriction to U of the first derivative of θ.)

A condition on the distribution of the measurement variable is required to guarantee the achievability of the desired rate of convergence.

Condition 2. The distribution of $\mathbf{X} = (X_1, \dots, X_d)$ 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 $M_1^{-1} < f(\mathbf{x}) < M_1$ for $\mathbf{x} \in U$.

A condition on the conditional distribution of Y given $\mathbf{X}$ is required to guarantee the uniqueness of the conditional median (uniqueness will ensure consistency) and also the achievability of the desired rate of convergence. If the conditional density is not bounded away from zero around the median, the desired rate of convergence will not be achievable. (The same condition is required in order to obtain the usual asymptotic result about the sample median in the univariate case.)

Condition 3. The conditional distribution of Y given $\mathbf{X} = \mathbf{x}$ is absolutely continuous and its density $h(y|\mathbf{x}, \theta)$ is bounded away from zero and infinity over a neighborhood of the median, that is, there is a positive constant ε_0 such that $M_1^{-1} \leq h(y|\mathbf{x}, \theta) \leq M_1$ for $\mathbf{x} \in U$ and $y \in (\theta(\mathbf{x}) - \varepsilon_0, \theta(\mathbf{x}) + \varepsilon_0)$.

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

$$\lim_{c \to \infty} \limsup_{n} P(|V_n| > c b_n) = 0.$$

Set $r = (2 + d)^{-1}$.

Theorem 1. *Suppose that Conditions 1–3 hold and that $\delta_n \sim n^{-r}$. 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 nonempty interior and let $g(\cdot)$ be a real-valued function on U. Set

$$\|g\|_q = \left\{ \int_C |g(\mathbf{x})|^q \, d\mathbf{x} \right\}^{1/q}, \qquad 1 \le q < \infty,$$

$$\|g\|_\infty = \sup_{\mathbf{x} \in C} |g(\mathbf{x})|.$$

THEOREM 2. *Suppose that Conditions* 1–3 *hold and that* $\delta_n \sim (n^{-1} \log n)^r$. *Then there exists a* $c > 0$ *such that*

$$\lim_n P\Big(\|\hat{\theta}_n - \theta\|_\infty \ge c\big(n^{-1} \log n\big)^r \Big) = 0.$$

THEOREM 3. *Suppose that Conditions* 1–3 *hold and that* $\delta_n \sim n^{-r}$. *Then there exists a* $c > 0$ *such that*

$$\lim_n P\Big(\|\hat{\theta}_n - \theta\|_q \ge c n^{-r} \Big) = 0, \qquad 1 \le q < \infty.$$

Proofs of these theorems are given in Section 3.

REMARK 1. Theorem 2 holds with a fixed constant c instead of taking a limit when $c \to \infty$ as in Theorem 1. This is because the L^∞ rate is slower than the pointwise rate. See the inequality before (3.11) in the proof of Theorem 2.

REMARK 2. Theorem 3 also holds with a fixed constant c because of the following heuristic argument. Consider the weighted sum $\sum_{i=1}^m w_i |\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q$ with $w_i \sim \delta_n^d$ and $\sum_{i=1}^m w_i = \operatorname{vol}(C)$ (the volume of C). For simplicity, suppose $\operatorname{vol}(C) = 1$ and $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ is nonrandom. Also, suppose $\{|\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q\}$ is a sequence of independent random variables and there is a positive constant c_0 such that $E(|\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q) \le c_0 \delta_n^q$ and $\operatorname{Var}(|\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q) \le c_0 \delta_n^{2q}$ for $i = 1, 2, \dots, m$ [see (3.20)]. Let c denote a positive constant so that $c^q > c_0$. Then by Chebyshev's inequality,

$$\begin{aligned} P\Big(\sum_i w_i |\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q \ge (c\delta_n)^q \Big) &\le \frac{\sum_i w_i^2 \operatorname{Var}\big(|\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q\big)}{(c^q - c_0)^2 \delta_n^{2q}} \\ &\le \frac{c_0 \max_i w_i}{(c^q - c_0)^2} = O\big(\delta_n^d\big) \to 0. \end{aligned}$$

Note that $\delta_n \sim n^{-r}$ and $\|\hat{\theta}_n - \theta\|_q^q = \int_C |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q \, d\mathbf{x}$ may be interpreted as the weighted sum $\sum_i w_i |\hat{\theta}_n(\mathbf{x}_i) - \theta(\mathbf{x}_i)|^q$. We conclude that Theorem 3 holds with a fixed constant c. This heuristic argument is justified in the proof of Theorem 3; see especially (3.20)–(3.22).

REMARK 3. With a simple modification of Condition 3, Theorems 1–3 are easily extended to yield rates of convergence for nonparametric estimators of other conditional quantiles.

REMARK 4. The proof of Theorem 2 is simpler and more intuitive than the corresponding proof given by Härdle and Luckhaus (1984) (only the calculation of binomial probabilities is required).

3. Proofs.

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

$$\lim_{c\to\infty} \limsup_{n} P\big(\hat{\theta}_n(\mathbf{x}) > \theta(\mathbf{x}) + cn^{-r}\big) = 0, \qquad \mathbf{x} \in U. \tag{3.1}$$

Given $\mathbf{x} \in U$, set $N_n = N_n(\mathbf{x})$ and $I_n = I_n(\mathbf{x})$. The proof of (3.1) depends on the following lemma, whose proof uses the Bernstein and Hoeffding inequalities. Let ε_n denote a sequence of positive numbers tending to zero as $n \to \infty$.

LEMMA 1. *Suppose that Conditions* 1–3 *hold and that c is a positive constant greater than M_0. Then there are positive constants c_1 and c_2 such that*

$$\begin{aligned} &P\Big(N_n^{-1}\sum_{I_n} 1_{\{Y_i \geq \theta(\mathbf{x}) + c\delta_n\}} \geq \tfrac{1}{2} - \varepsilon_n\delta_n\Big) \\ &\quad \leq \exp\big(-(c - M_0)^2 c_1 n\delta_n^{d+2}\big) + \exp\big(-c_2 n\delta_n^d\big), \qquad \mathbf{x} \in U. \end{aligned}$$

PROOF. According to Condition 1, $\theta(\mathbf{X}_i) \leq \theta(\mathbf{x}) + M_0\delta_n$ for $i \in I_n$. Thus

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

Hence by Condition 3, there is a positive constant η such that if $c > M_0$ and n is sufficiently large, then

$$\tfrac{1}{2} - \varepsilon_n\delta_n - P\big(Y_i \geq \theta(\mathbf{x}) + c\delta_n | \mathbf{X}_i\big) \geq (c - M_0)\eta\delta_n, \qquad i \in I_n. \tag{3.2}$$

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

$$E(Z_i | \mathbf{X}_1, \dots, \mathbf{X}_n) = 0, \qquad i = 1, \dots, n,$$

and, by (3.2),

$$N_n^{-1}\sum_{I_n}\big[\tfrac{1}{2} - \varepsilon_n\delta_n - P\big(Y_i \geq \theta(\mathbf{x}) + c\delta_n | \mathbf{X}_i\big)\big] \geq (c - M_0)\eta\delta_n, \qquad c > M_0.$$

Set $P^X(\cdot) = P(\cdot | \mathbf{X}_1, \dots, \mathbf{X}_n)$. Then

$$\begin{aligned} &P\Big(N_n^{-1}\sum_{I_n} 1_{\{Y_i \geq \theta(\mathbf{x}) + c\delta_n\}} \geq \tfrac{1}{2} - \varepsilon_n\delta_n\Big) \\ &\quad \leq E\Big[P^X\Big(N_n^{-1}\sum_{I_n} Z_i \geq N_n^{-1}\sum_{I_n}\big[\tfrac{1}{2} - \varepsilon_n\delta_n - P\big(Y_i \geq \theta(\mathbf{x}) + c\delta_n | \mathbf{X}_i\big)\big]\Big)\Big] \\ &\quad \leq E\Big[P^X\Big(N_n^{-1}\sum_{I_n} Z_i \geq (c - M_0)\eta\delta_n\Big)\Big]. \end{aligned} \tag{3.3}$$

By Hoeffding's inequality [see Theorem 2 of Hoeffding (1963)]

$$P^X\Big(N_n^{-1}\sum_{I_n} Z_i \geq (c - M_0)\eta\delta_n\Big) \leq \exp\big(-2N_n\big[(c - M_0)\eta\delta_n\big]^2\big). \tag{3.4}$$

Set $p_n = p_n(\mathbf{x}) = P(\|\mathbf{X}_1 - \mathbf{x}\| \le \delta_n)$. According to Bernstein's inequality [see Theorem 3 of Hoeffding (1963)]

$$P\left(N_n < \tfrac{1}{2}np_n\right) \le \exp\left(-\frac{n\left(\frac{1}{2}p_n\right)^2}{2p_n + p_n}\right). \tag{3.5}$$

By (3.3)–(3.5),

$$\begin{aligned} &P\left(N_n^{-1}\sum_{I_n} 1_{\{Y_i \ge \theta(\mathbf{x}) + c\delta_n\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\right) \\ &\quad \le E\left[\exp\left(-2[(c - M_0)\eta]^2 N_n\delta_n^2\right)1_{\{N_n \ge (1/2)np_n\}}\right] + P\left(N_n < \tfrac{1}{2}np_n\right) \\ &\quad \le \exp\left(-[(c - M_0)\eta]^2 np_n\delta_n^2\right) + \exp\left(-\tfrac{1}{12}np_n\right). \end{aligned} \tag{3.6}$$

By Condition 2, $p_n \sim \delta_n^d$. The conclusion of Lemma 1 follows from (3.6). □

The proof of (3.1) will now be given. Note that the event $\{\hat{\theta}_n(\mathbf{x}) \ge \theta(\mathbf{x}) + c\delta_n\}$ is contained in the event $\{N_n^{-1}\Sigma_{I_n}1_{\{Y_i \ge \theta(\mathbf{x}) + c\delta_n\}} \ge \frac{1}{2}\}$. It follows from Lemma 1 and $\delta_n \sim n^{-r}$ or, equivalently, $n\delta_n^{d+2} \sim 1$, that

$$\begin{aligned} P\left(\hat{\theta}_n(\mathbf{x}) \ge \theta(\mathbf{x}) + c\delta_n\right) &\le P\left(N_n^{-1}\sum_{I_n} 1_{\{Y_i \ge \theta(\mathbf{x}) + c\delta_n\}} \ge \tfrac{1}{2}\right) \\ &\le \exp\left(-(c - M_0)^2 c_1 n\delta_n^{d+2}\right) + \exp\left(-c_2 n\delta_n^d\right) = o(1) \end{aligned}$$

as $n, c \to \infty$. This completes the proof of Theorem 1. □

PROOF OF THEOREM 2. Without loss of generality it can be assumed that $C = [-\frac{1}{2}, \frac{1}{2}]^d$. Choose $s > 1$ and let $\{L_n\}$ denote a sequence of positive integers such that $L_n \sim n^s$. 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

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

It follows from $\lambda_n \sim n^{-s} = o((n^{-1}\log n)^r)$ 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

$$\lim_n P\left(\max_{\mathbf{w} \in C_n}\sup_{\mathbf{x} \in Q_{\mathbf{w}}}\left|\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})\right| \ge c\left(n^{-1}\log n\right)^r\right) = 0. \tag{3.7}$$

To prove (3.7), let $\mathbf{x} \in Q_{\mathbf{w}}$ and $N_n' = N_n'(\mathbf{w}) := \#\{i\colon \|\mathbf{X}_i - \mathbf{w}\| \le \delta_n - \lambda_n\sqrt{d}\}$. It follows from $N_n = N_n(\mathbf{x}) := \#\{i\colon \|\mathbf{X}_i - \mathbf{x}\| \le \delta_n\} \ge N_n'$ for $\mathbf{x} \in Q_{\mathbf{w}}$ that

$$\{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \ge c\delta_n\} \subseteq \Big\{N_n^{-1} \sum_{I_n} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}\Big\}$$

$$\subseteq \Big\{\sum_{I_n^*} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}N_n'\Big\},$$

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

$$\bigcup_{Q_{\mathbf{w}}} \{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \ge c\delta_n\} \subseteq \Big\{\sum_{I_n^*} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}N_n'\Big\}. \tag{3.8}$$

Set $N_n^* = N_n^*(\mathbf{w}) = \#I_n^*(\mathbf{w})$. Then $N_n^* - N_n' = \#\{i\colon \delta_n - \lambda_n\sqrt{d} \le \|\mathbf{X}_i - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d}\}$ is a binomial random variable with parameters n and π_n, where (by Condition 2)

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

By Condition 2 and Theorem 12.2 of Breiman, Friedman, Olshen and Stone [(1984), page 320], there are positive constants c_3 and c_4 such that

$$\lim_n P(\Psi_n) = 1, \tag{3.9}$$

where $\Psi_n = \{N_n^*(\mathbf{w}) - N_n'(\mathbf{w}) \le c_3 \text{ and } N_n^*(\mathbf{w}) \ge c_4 n\delta_n^d \text{ for all } \mathbf{w} \in C_n\}$.

According to (3.8), (3.9) and $N_n^{*-1} \le (c_4 n\delta_n^d)^{-1}$ on Ψ_n, there is a sequence of positive constants ε_n ($\sim \delta_n/\log n$) such that

$$P\Big(\max_{\mathbf{w}\in C_n} \sup_{\mathbf{x}\in Q_{\mathbf{w}}} [\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w})] \ge c\delta_n\Big)$$

$$\le P\Big(\bigcup_{C_n}\bigcup_{Q_{\mathbf{w}}} \{\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{w}) \ge c\delta_n\}\Big)$$

$$\le P\Big(\bigcup_{C_n}\Big\{\sum_{I_n^*} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}N_n'\Big\}\Big) \tag{3.10}$$

$$\le P\Big(\bigcup_{C_n}\Big\{\sum_{I_n^*} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2}N_n^* - \tfrac{1}{2}c_3\Big\} \cap \Psi_n\Big) + P(\Psi_n^c)$$

$$\le P\Big(\bigcup_{C_n}\Big\{N_n^{*-1} \sum_{I_n^*} 1_{\{Y_i \ge \theta(\mathbf{w}) + c\delta_n\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\Big\}\Big) + P(\Psi_n^c).$$

According to Condition 2, $P(\|\mathbf{X}_1 - \mathbf{w}\| \le \delta_n + \lambda_n\sqrt{d}) \sim \delta_n^d$ for $\mathbf{w} \in C_n$. Thus by Lemma 1, there are positive constants c_5 and c_6 such that

$$\begin{aligned} P&\left(\bigcup_{C_n}\left\{N_n^{*-1}\sum_{I_n^*}1_{\{Y_i \ge \theta(\mathbf{w})+c\delta_n\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\right\}\right) \\ &\le (2L_n)^d \max_{C_n} P\left(N_n^{*-1}\sum_{I_n^*}1_{\{Y_i \ge \theta(\mathbf{w})+c\delta_n\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\right) \\ &\le (2L_n)^d \exp\left(-c^2c_5 n\delta_n^{d+2}\right) + (2L_n)^d \exp\left(-c_6 n\delta_n^d\right). \end{aligned}$$

Note that $(2L_n)^d \sim n^{sd}$ and δ_n is chosen so that $\delta_n \sim (n^{-1}\log n)^r$ or, equivalently $n\delta_n^{d+2} \sim \log n$. Consequently, for c sufficiently large

$$P\left(\bigcup_{C_n}\left\{N_n^{*-1}\sum_{I_n^*}1_{\{Y_i \ge \theta(\mathbf{w})+c\delta_n\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\right\}\right) = o(1). \tag{3.11}$$

Hence, by (3.9)–(3.11) there is a positive constant c such that

$$\lim_n P\left(\max_{\mathbf{w}\in C_n}\sup_{\mathbf{x}\in Q_\mathbf{w}}\left[\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})\right] \ge c\left(n^{-1}\log n\right)^r\right) = 0. \tag{3.12}$$

Similarly,

$$\lim_n P\left(\max_{\mathbf{w}\in C_n}\sup_{\mathbf{x}\in Q_\mathbf{w}}\left[\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})\right] \le -c\left(n^{-1}\log n\right)^r\right) = 0. \tag{3.13}$$

It follows from (3.12) and (3.13) that (3.7) is valid. This completes the proof of Theorem 2. □

PROOF OF THEOREM 3. By Condition 3 and the argument given in the proof of Lemma 1, there are positive constants c_7 and c_8 such that

$$\begin{aligned} P&\left(\bigcup_{C_n}\left\{N_n^{*-1}\sum_{I_n^*}1_{\{Y_i \ge \theta(\mathbf{w})+c\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\right\}\right) \\ &\le (2L_n)^d \exp\left(-c_7[\varepsilon_0 \wedge c]^2 n\delta_n^d\right) + (2L_n)^d \exp\left(-c_8 n\delta_n^d\right). \end{aligned}$$

Since $(2L_n)^d \sim n^{sd}$ and δ_n is chosen so that $n\delta_n^d \sim \delta_n^{-2} \sim n^{2r}$, we conclude that for $c > 0$,

$$P\left(\bigcup_{C_n}\left\{N_n^{*-1}\sum_{I_n^*}1_{\{Y_i \ge \theta(\mathbf{w})+c\}} \ge \tfrac{1}{2} - \varepsilon_n\delta_n\right\}\right) = o(1). \tag{3.14}$$

It follows from (3.8)–(3.10), (3.14) [with $c\delta_n$ replaced by c in (3.8) and (3.10)] and the boundedness of $\theta(\cdot)$ on C that there is a positive constant $T \ge 1$ such that

$$\lim_n P(\Pi_n) = 1, \tag{3.15}$$

where $\Pi_n := \{\|\hat{\theta}_n\|_\infty \le T\}$.

For $i = 1, \dots, n$, 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 $\bar{\theta}_n(\mathbf{x}) = \operatorname{med}\{Y_i' \colon i \in I_n(\mathbf{x})\}$. Note that $\bar{\theta}_n(\mathbf{x}) = \hat{\theta}_n(\mathbf{x})$ for $\mathbf{x} \in C$ except on Π_n^c. Thus by (3.15), in order to prove the theorem, it is sufficient to show

$$\lim_n P\left(\|\bar{\theta}_n - \theta\|_q \ge cn^{-r}\right) = 0. \tag{3.16}$$

To verify (3.16), we may assume that C is contained in the interior of the cube $C_0 = [-\frac{1}{2}, \frac{1}{2}]^d \subset U$ and $\|\theta(\cdot)\|_\infty \le T$ on C. Set $E^X(\cdot) = E(\cdot\,|\mathbf{X}_1, \dots, \mathbf{X}_n)$. Then

$$E^X\left(\|\bar{\theta}_n - \theta\|_q^q\right) = \int_C E^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\right) d\mathbf{x}$$

and

$$\begin{aligned} E^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\right) &= \int_0^\infty qt^{q-1}P^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right) dt \\ &= \int_0^{2M_0\delta_n} qt^{q-1}P^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right) dt \\ &\quad + \int_{2M_0\delta_n}^{2T} qt^{q-1}P^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right) dt \\ &\le (2M_0\delta_n)^q + \int_{2M_0\delta_n}^{2T} qt^{q-1}P^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right) dt. \end{aligned} \tag{3.17}$$

Recall that $N_n(\mathbf{x}) = \sharp I_n(\mathbf{x})$. By Condition 2 and Theorem 12.2 of Breiman, Friedman, Olshen and Stone [(1984), page 320], there is a positive constant c_9 such that

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

where $\Omega_n := \{N_n(\mathbf{x}) \ge c_9 n\delta_n^d \text{ for all } \mathbf{x} \in C\}$. By Condition 3, there is a positive constant c_{10} such that

$$\begin{aligned} &\int_{2M_0\delta_n}^{2T} qt^{q-1}P^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\right) dt \\ &\qquad \le c_{10}T^q\left[N_n(\mathbf{x})\right]^{-q/2}, \qquad \mathbf{x} \in C. \end{aligned} \tag{3.19}$$

[The proof of (3.19) will be given at the end of this section.] It follows from (3.17)–(3.19) that there is a positive constant c_{11} such that

$$E^X\left(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q\right) \le c_{11}\left[N_n(\mathbf{x})^{-q/2} + \delta_n^q\right], \qquad \mathbf{x} \in C.$$

Note that $\delta_n \sim n^{-r}$. Thus there is a positive constant c_{12} such that

$$E^X\left(\|\bar{\theta}_n - \theta\|_q^q\right) \le c_{12}(n^{-r})^q \quad \text{on } \Omega_n. \tag{3.20}$$

Set $\operatorname{Var}^X(\cdot) = \operatorname{Var}(\cdot|\mathbf{X}_1, \dots, \mathbf{X}_n)$ and $\operatorname{Cov}^X(\cdot, \cdot) = \operatorname{Cov}((\cdot, \cdot)|\mathbf{X}_1, \dots, \mathbf{X}_n)$. Then

$$\begin{aligned}
&\operatorname{Var}^X\left(\|\bar{\theta}_n - \theta\|_q^q\right) \\
&\quad = \operatorname{Var}^X\left(\int_C |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q \, d\mathbf{x}\right) \\
&\quad = \int\int_D \operatorname{Cov}^X\left\{|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^q, |\bar{\theta}_n(\mathbf{x}') - \theta(\mathbf{x}')|^q\right\} d\mathbf{x}\, d\mathbf{x}' \\
&\quad \le \int\int_D \left\{E^X(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})|^{2q}) E^X(|\bar{\theta}_n(\mathbf{x}') - \theta(\mathbf{x}')|^{2q})\right\}^{1/2} d\mathbf{x}\, d\mathbf{x}',
\end{aligned}$$

where $D = \{\mathbf{x}, \mathbf{x}' \in C: \|\mathbf{x} - \mathbf{x}'\| \le 2\delta_n\}$. It now follows from (3.20) that there is a sequence of positive numbers κ_n tending to 0 such that

$$\operatorname{Var}^X\left(\|\bar{\theta}_n - \theta\|_q^q\right) \le \kappa_n (n^{-r})^{2q} \quad \text{on } \Omega_n. \tag{3.21}$$

By (3.20), (3.21) and Chebyshev's inequality, we conclude that for c sufficiently large,

$$\begin{aligned}
P\left(\|\bar{\theta}_n - \theta\|_q^q \ge (cn^{-r})^q | \mathbf{X}_1, \dots, \mathbf{X}_n\right) &\le \frac{\operatorname{Var}^X\left(\|\bar{\theta}_n - \theta\|_q^q\right)}{(c^q - c_{12})^2 (n^{-r})^{2q}} \\
&= O(\kappa_n) \quad \text{on } \Omega_n.
\end{aligned} \tag{3.22}$$

It follows from (3.18) and (3.22) that there is a positive constant c such that (3.16) holds, as desired.

Proof of (3.19). Given $\mathbf{x} \in C$, it follows from Condition 1 that $\theta(\mathbf{X}_i) \le \theta(\mathbf{x}) + M_0\delta_n$ for $i \in I_n$. Thus, by Condition 3 and $T \ge 1$ (for n sufficiently large),

$$\begin{aligned}
\tfrac{1}{2} - P(Y_i > \theta(\mathbf{x}) + t|\mathbf{X}_i) &\ge P(0 \le Y_i - \theta(\mathbf{X}_i) \le t - M_0\delta_n|\mathbf{X}_i) \\
&\ge M_1^{-1}(t - M_0\delta_n) \\
&\ge M_1^{-1}(2T)^{-1}(t - M_0\delta_n), \qquad 2M_0\delta_n \le t \le \varepsilon_0,\ i \in I_n,
\end{aligned}$$

and

$$\tfrac{1}{2} - P(Y_i > \theta(\mathbf{x}) + t|\mathbf{X}_i) \ge M_1^{-1}(\varepsilon_0 - M_0\delta_n)(2T)^{-1}(t - M_0\delta_n),$$
$$\varepsilon_0 \le t \le 2T,\ i \in I_n.$$

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

$$\begin{aligned}
&\tfrac{1}{2} - P(Y_i > \theta(\mathbf{x}) + t|\mathbf{X}_i) \ge c_{13}T^{-1}(t - M_0\delta_n), \\
&\qquad\qquad 2M_0\delta_n \le t \le 2T,\ i \in I_n.
\end{aligned} \tag{3.23}$$

Set $Z_i = 1_{\{Y_i > \theta(\mathbf{x}) + t\}} - P(Y_i > \theta(\mathbf{x}) + t | \mathbf{X}_i)$ and $N_n = N_n(\mathbf{x})$. Then it follows from $\{Y_i' > \theta(\mathbf{x}) + t\} \subset \{Y_i > \theta(\mathbf{x}) + t\}$ and (3.23) that

$$\begin{aligned} P^X\big(\bar{\theta}_n(\mathbf{x}) > \theta(\mathbf{x}) + t\big) &\le P^X\Big(N_n^{-1} \sum_{I_n} 1_{\{Y_i' > \theta(\mathbf{x}) + t\}} \ge \tfrac{1}{2}\Big) \\ &\le P^X\Big(N_n^{-1} \sum_{I_n} 1_{\{Y_i > \theta(\mathbf{x}) + t\}} \ge \tfrac{1}{2}\Big) \\ &\le P^X\Big(N_n^{-1} \sum_{I_n} Z_i \ge N_n^{-1} \sum_{I_n} \big[\tfrac{1}{2} - P(Y_i > \theta(\mathbf{x}) + t | \mathbf{X}_i)\big]\Big) \\ &\le P^X\Big(N_n^{-1} \sum_{I_n} Z_i \ge c_{13} T^{-1}(t - M_0\delta_n)\Big), \\ &\qquad\qquad 2M_0\delta_n \le t \le 2T. \end{aligned}$$

By Hoeffding's inequality (see the proof of Lemma 1), there is a positive constant c_{14} such that

$$P^X\big(\bar{\theta}_n(\mathbf{x}) > \theta(\mathbf{x}) + t\big) \le \exp\big[-c_{14} N_n T^{-2}(t - M_0\delta_n)^2\big], \quad 2M_0\delta_n \le t \le 2T \tag{3.24}$$

and, similarly,

$$P^X\big(\bar{\theta}_n(\mathbf{x}) < \theta(\mathbf{x}) - t\big) \le \exp\big[-c_{14} N_n T^{-2}(t - M_0\delta_n)^2\big], \quad 2M_0\delta_n \le t \le 2T. \tag{3.25}$$

It follows from (3.24) and (3.25) that there is a positive constant c_{15} such that

$$\begin{aligned} &\int_{2M_0\delta_n}^{2T} t^{q-1} P^X\big(|\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| > t\big)\, dt \\ &\quad\le 2\int_{2M_0\delta_n}^{2T} t^{q-1} \exp\big[-c_{14} N_n T^{-2}(t - M_0\delta_n)^2\big]\, dt \\ &\quad\le 2\int_{M_0\delta_n}^{2T} (s + M_0\delta_n)^{q-1} \exp\big(-c_{14} N_n T^{-2} s^2\big)\, ds \\ &\quad\le 2^q \int_0^\infty s^{q-1} \exp\big(-c_{14} N_n T^{-2} s^2\big)\, ds \\ &\quad= c_{15} T^q N_n^{-q/2}, \qquad \mathbf{x} \in C, \end{aligned}$$

as desired. □

Acknowledgments. This work is part of the author's Ph.D. thesis at University of California, Berkeley, written under the supervision of Professor C. Stone, whose generous guidance and suggestions are gratefully acknowledged and appreciated. Thanks are also due Professors P. Bickel and L. Le Cam for their support, and the referee and Associate Editor for helpful comments that improved the readability of the paper. Finally, I would like to thank my wife, Psyche Lee, for her patience and inspiration.

REFERENCES

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

DEVROYE, L. P. and WAGNER, T. J. (1980a). On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. *Z. Wahrsch. verw. Gebiete* **51** 15–25.

DEVROYE, L. P. and WAGNER, T. J. (1980b). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239.

GORDON, L. and OLSHEN, R. A. (1980). Consistent nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.* **10** 611–627.

HÄRDLE, W. (1984). Robust regression function estimation. *J. Multivariate Anal.* **14** 169–180.

HÄRDLE, W. and GASSER, T. (1984). Robust non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46** 42–51.

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

HÄRDLE, W. and TSYBAKOV, A. B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* **16** 120–135.

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

SPIEGELMAN, C. and SACKS, J. (1980). Consistent window estimation in nonparametric regression. *Ann. Statist.* **8** 240–246.

STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *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.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514