The Annals of Statistics
1988, Vol. 16, No. 4, 1428–1449

STRONG UNIFORM CONSISTENCY RATES FOR ESTIMATORS OF CONDITIONAL FUNCTIONALS[1]

BY W. HÄRDLE, P. JANSSEN AND R. SERFLING

Universität Bonn, Limburgs Universitair Centrum and The Johns Hopkins University

Strong uniform consistency rates are established for kernel type estimators of functionals of the conditional distribution function, under general conditions. The present treatment unifies a number of specific problems previously studied separately in the literature. Some of these applications we treat in detail, including regression curve estimation, density estimation, estimation of conditional df's, L-smoothing and M-smoothing. Various previous results in the literature are extended and/or sharpened.

1. Introduction, basic formulation and applications. Let (X, Y) be a bivariate random vector with joint df $F(x, y)$, joint density $f(x, y)$, conditional df $F(y|x)$ for Y given X, conditional density $f(y|x)$ for Y given X and marginal density $f_0(x)$ for X, x and $y \in \mathbb{R}$. Let $\{\beta_t,\ t \in I\}$ be a family of real-valued measurable functions on $\mathbb{R}$ for which it is desired to estimate

$$r_t(x) = E\{\beta_t(Y)|X = x\} = \int \beta_t(y)\, dF(y|x), \tag{1.1}$$

with a good almost sure (a.s.) convergence rate holding uniformly for $t \in I$ and $x \in J$, where I is a possibly infinite, or possibly degenerate, interval in $\mathbb{R}$ and J is a possibly infinite interval in $\mathbb{R}$. In general, we may think of this type of problem as one of nonparametric estimation of linear functionals of the conditional df $F(y|x)$. As will be seen from the examples, such a problem may arise in nonparametric regression and related contexts, either as a given target problem or as a technical problem to which a given target problem becomes reduced.

Expressing $r_t(x)$ in the form $r_t(x) = d_t(x)/f_0(x)$, with

$$d_t(x) = \int \beta_t(y) f(x, y)\, dy, \tag{1.2}$$

we shall consider estimators of the form

$$r_{tn}(x) = d_{tn}(x)/f_n(x), \tag{1.3a}$$

with

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K_{0n}\left(\frac{x - X_i}{h_n}\right) \tag{1.3b}$$

Received July 1986; revised January 1988.

[1]Research supported by the U.S. Department of Navy under Office of Naval Research Contract N00014-79-C-0801 and by NATO under Research Grant 0034/87. Reproduction in whole or in part is permitted for any purpose of the United States Government. Research of the third author also supported under a U.S. Senior Scientist Award by the Alexander von Humboldt-Stiftung.

AMS 1980 *subject classifications*. Primary 62G05; secondary 60F15.

Key words and phrases. Strong uniform consistency rates, nonparametric kernel estimators, density estimation, L-smoother, M-smoother, regression function estimation.

and

$$d_{tn}(x) = (nh_n)^{-1} \sum_{i=1}^{n} \beta_t(Y_i) K_n\left(\frac{x - X_i}{h_n}\right), \tag{1.3c}$$

where $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent observations on F, $\{K_{0n}\}$ and $\{K_n\}$ are sequences of *kernel* functions $K: \mathbb{R} \to \mathbb{R}$ and $\{h_n\}$ is a sequence of positive constants (*bandwidths*) tending to 0 as $n \to \infty$. We recognize $f_n(\cdot)$ to be in the form of the familiar Rosenblatt–Parzen type of density estimator for $f_0(\cdot)$, except that we consider a sequence $\{K_{0n}\}$ instead of a fixed kernel K_0.

The kernels under consideration may be smooth or discrete, although we shall give some emphasis to the discrete case. Since smooth kernels become discretized in computations with data, this case has considerable relevance to estimators actually computed in practice. We consider sequences instead of fixed kernels K_0 and K in order to include the case that K_{0n} and K_n are step-function kernels providing increasingly close approximation to given smooth kernels. The sequences $\{K_{0n}\}$ and $\{K_n\}$ may be selected to coincide, but this is not necessary, and we avoid such an assumption in order to provide greater flexibility in applications.

Under suitable restrictions, we shall establish the uniform a.s. rate

$$\sup_{t \in I} \sup_{x \in J} |r_{tn}(x) - r_t(x)| = O\left(\max\left\{\left(\frac{\log n}{nh_n}\right)^{1/2}, h_n^{\alpha}\right\}\right) \quad \text{a.s., } n \to \infty, \tag{1.4}$$

where α is the order of uniform local Lipschitz (ulL) conditions imposed on $f_0(\cdot)$ and $\{d_t(\cdot),\ t \in I\}$. [Under stronger smoothness conditions, the component h_n^{α} in (1.4) can be improved.] By allowing a *family* of $\beta(\cdot)$ functions instead of a single one, we obtain a very useful type of extension of previous results in the literature, and our results also yield certain improvements in previously considered special cases.

Section 2 provides general theory, in which the most fundamental results are Theorems 2.1 and 2.2. These yield, in particular, a new result on density estimation (Corollary 2.1), our main result on (1.4) (Theorem 2.3) and a corollary giving conditions under which (1.4) provides the rate $O((n/\log n)^{-\alpha/(2\alpha+1)})$. Optimality of this rate, in the case of nonparametric regression function estimation, is shown in Stone (1982).

The crucial role of (1.4) in establishing uniform strong consistency rates for a variety of estimators involving conditional functionals may be seen from the following examples, which will be treated technically in Section 3 by systematically applying the theory of Section 2.

EXAMPLE 1. *Nonparametric regression function estimation.* This corresponds to (1.1) with the single $\beta(\cdot)$ function $\beta(y) = y$, in which case $r(x) = E(Y|X = x)$, the classical regression function, and $r_n(\cdot)$ represents the classical Nadaraya–Watson estimator [Nadaraya (1964) and Watson (1964)]. For general background, see Collomb (1981) and Mack and Silverman (1982), with whose results we make comparison in Section 3.

EXAMPLE 2. *Nonparametric scale curve estimation.* A nonparametric approach to the problem of heteroscedasticity in linear models involves estimation of the conditional variances

$$v(x) = E(Y^2|X = x) - [E(Y|X = x)]^2$$

[see Carroll (1982)]. Here the first component is given by (1.1) with the single function $\beta(y) = y^2$, and the second component is handled by Example 1. (Higher-order conditional moments may be treated similarly.)

EXAMPLE 3. *The conditional df.* The conditional df itself, i.e., the function $F(t|x)$, $t \in \mathbb{R}$, is given by (1.1) with $\beta_t(y) = I(y \leq t)$, $y \in \mathbb{R}$, $t \in I = \mathbb{R}$. The corresponding estimator $F_n(t|x)$ given by (1.3a) has been treated by Collomb (1980). He proved consistency results, without rates, which are uniform in x and pointwise in t. A Glivenko–Cantelli type theorem for $F_n(t|x)$, uniform in t and pointwise in x, is given in Stute (1986). Besides the intrinsic interest of the additional information provided by (1.4) in this case, such a result also plays a fundamental role in obtaining uniform strong consistency rates in other problems, as in Example 5.

EXAMPLE 4. *The marginal density* f_0. With the single trivial function $\beta(y) \equiv 1$, $d_n(\cdot)$ given by (1.3c) becomes a density estimator for $f_0(\cdot)$. A key theoretical tool (Theorem 2.2) in Section 2 concerns the behavior of $\sup_t \sup_x |d_{tn}(x) - d_t(x)|$ and, for this choice of $\{\beta_t,\ t \in I\}$, yields new results on density estimation [see Corollary 2.1 and Remark 2.3(i)].

EXAMPLE 5. *L-smoothing.* Denote by $F^{-1}(v|x) = \inf\{y\colon\ F(y|x) \geq v\}$, $0 < v < 1$, the conditional quantile function associated with $F(\cdot|x)$ and consider estimation of a conditional L-functional

$$l(x) = \int_0^1 J(v) F^{-1}(v|x)\, dv.$$

For $J(v) \equiv 1$, $l(x)$ reduces to the *regression function* $r(x)$ considered in Example 1. The same occurs in the case $J(v) = I\{p \leq v \leq 1 - p\}/(1 - 2p)$, where $0 < p < 1/2$, with $f(y|x)$ symmetric about $r(x)$. Letting $F_n(t|x)$ denote the estimator of $F(t|x)$ considered in Example 3, we consider for $l(x)$ the estimator $l_n(x)$ produced by substituting $F_n^{-1}(v|x)$ for $F^{-1}(v|x)$. In our treatment in Section 3, we obtain uniform strong consistency rates for *trimmed* L-smoothers by reduction of the problem to an application of results obtained for Example 3.

EXAMPLE 6. *M-smoothing.* For any given real function $\psi(\cdot)$, a corresponding M-functional $T_\psi(\cdot)$ may be defined on df's G by letting $T_\psi(G)$ denote the solution t_0 of the equation

$$\int \psi(y - t_0)\, dG(y) = 0.$$

[The case $\psi(x) = x$ yields $T_\psi(G) = \int y\, dG(y)$, the mean functional.] In the case that $G(\cdot)$ is symmetric about θ, any antisymmetric ψ yields $T_\psi(G) = \theta$. Thus, for a class of such $\psi(\cdot)$, a class of competing estimators of θ is given by $T_\psi(\hat{G})$, with $\hat{G}$ estimating G.

Adapting this to regression curve estimation, we let $r(x)$ be as in Example 1 and assume that, for each $x \in J$, the conditional density $f(y|x)$ is symmetric about $r(x)$. Then, for antisymmetric ψ, $r(x)$ is the solution of the preceding equation with $G(\cdot)$ replaced by $F(\cdot|x)$ and an estimator $r_{\psi n}(x)$ is given by solving this equation with $G(\cdot)$ replaced by $F_n(\cdot|x)$ defined as in Example 3. For suitable choice of $\psi(\cdot)$, the function $r_{\psi n}(x)$ for estimation of $r(x)$, $x \in J$, is more resistant to the presence of outliers than is the estimator $r_n(x)$ treated previously. We call $r_{\psi n}(x)$, $x \in J$, the *M-smoother* corresponding to ψ. Pointwise consistency of M-smoothers has been treated by Stone (1977), Tsybakov (1983) and Härdle (1984). Uniform *weak* consistency rates have been established by Härdle and Luckhaus (1984), by reduction, with $\beta_t(y) \doteq \psi(y - t)$, $y \in \mathbb{R}$, to the analogous problem for the estimators $r_{tn}(x)$ of $r_t(x)$, for t in a small neighborhood of $r(x)$. Following this approach, we establish a.s. uniform consistency rates for M-smoothers in Section 3.

Our method in Section 2 will be to handle $r_{tn} - r_t$ via the decomposition

$$r_{tn} - r_t = R_{tn} + S_{tn}, \tag{1.5}$$

with $R_{tn} = (d_{tn} - d_t)/f_n$ and $S_{tn} = d_t(f_0 - f_n)/(f_0 f_n)$. As noted in Example 4, results for $f_n - f_0$ may be obtained by specialization of results for $d_{tn} - d_t$. Thus our treatment of $r_{tn} - r_t$ via (1.5) will flow from study of $d_{tn} - d_t$. For this we shall provide key foundational results in Theorems 2.1 and 2.2, from which our target results, Corollary 2.1, Theorem 2.3 and Corollary 2.2 will be derived. We shall deal with the stochastic component $d_{tn} - Ed_{tn}$ of $d_{tn} - d_t$ by analyzing, in effect, the modulus of continuity of a certain randomly weighted empirical df. The related bias component $Ed_{tn} - d_t$ will be handled by imposing mild local Lipschitz conditions. Without such conditions (for example, assuming only uniform continuity), the *rate* of convergence of the bias to 0 cannot be precisely characterized and thus a rate cannot properly be asserted in (1.4).

2. Some general results on strong uniform consistency rates. Our target results will follow from a basic theorem we establish on convergence of quantities of the form

$$\sup_{t \in I} \sup_{x \in J} |D_{tn}(x) - D_t(x)|, \tag{2.1}$$

with

$$D_t(x) = \int_{\mathbb{R}} \gamma_t(y) f(x, y)\, dy, \qquad x \in \mathbb{R}, \tag{2.2a}$$

$$D_{tn}(x) = c_n^{-1}[G_{tn}(x + c_n') - G_{tn}(x - c_n'')], \qquad x \in \mathbb{R}, \tag{2.2b}$$

$\{c_n'\}$ and $\{c_n''\}$ nonnegative sequences tending to 0, $c_n = c_n' + c_n''$ and

$$G_{tn}(x) = n^{-1} \sum_{i=1}^{n} \gamma_t(Y_i) I\{X_i \le x\}, \qquad x \in \mathbb{R}. \tag{2.2c}$$

Here (X, Y), (X_i, Y_i), $1 \le i \le n$, $f(x, y)$, $f(y|x)$, $f_0(x)$, I and J will be as in Section 1, but the functions $\{\beta_t,\ t \in I\}$ there are replaced for the present by a family $\{\gamma_t,\ t \in I]$ satisfying some specialized assumptions, and for the moment we do not concern ourselves with kernels $\{K_n\}$. The "randomly weighted" empirical df G_{tn} has mean function

$$G_t(x) = EG_{tn}(x) = \int_{-\infty}^{x} D_t(z)\,dz, \tag{2.3}$$

and we readily see by the classical SLLN that for each fixed pair t and x, $D_{tn}(x) \to D_t(x)$ a.s., $n \to \infty$. Our purpose here is to strengthen this by giving a rate for this convergence uniformly in $t \in I$ and $x \in J$.

The following assumptions come into play. We define a function g on $\mathbb{R}$ to be *uniformly locally Lipschitz of order* α (ulL-α), where $0 < \alpha \le 1$, if for some $\delta > 0$ and $M < \infty$, $\sup_{x \in \mathbb{R}} |g(x + z) - g(x)| \le M|z|^\alpha$, for $|z| \le \delta$.

ASSUMPTIONS.

$$\sup_{t\in I}\sup_{x\in J}\int_{\mathbb{R}}\gamma_t^2(y)f(y|x)\,dy = M_0 < \infty. \tag{A.1}$$

$$\sup_{x\in J} f_0(x) = M_1 < \infty. \tag{A.2}$$

$$0 \le \gamma_t(y) \le \gamma_{t'}(y), \qquad t < t' \in I,\ y \in \mathbb{R}. \tag{A.3}$$

(A.4) For some α, $0 < \alpha \le 1$, $D_t(\cdot)$ is ulL-α on J, uniformly for $t \in I$; i.e., for some $\delta_\alpha > 0$ and $M^{(\alpha)} < \infty$,

$$\sup_{t\in I}\ \sup_{\substack{x,x'\in J\\ |x-x'|\le\delta_\alpha}} |D_t(x) - D_t(x')| \le M^{(\alpha)}|x - x'|^\alpha.$$

(A.5) $E\gamma_t(Y)$ is a continuous function of t in I.

(A.6) The limit functions $\gamma_{t_*} = \lim_{t\to t_*}\gamma_t$ and $\gamma_{t^*} = \lim_{t\to t^*}\gamma_t$ exist and are finite a.s. (w.r.t. the df of Y), where $t_* = \inf I\ (\ge -\infty)$ and $t^* = \sup I\ (\le +\infty)$.

(A.7) $(E|\gamma_{t^*}(Y)|^\lambda)^{1/\lambda} = M_\lambda < \infty$ for some λ, $2 < \lambda \le \infty$ [in the case $\lambda = \infty$, M_∞ denotes $\sup_{y\in\mathbb{R}}|\gamma_{t^*}(y)|$].

REMARKS 2.1. Consider the assumptions:

$$\inf_{x\in J} f_0(x) = m_1 > 0; \tag{B.1}$$

$$\sup_{t\in I}\sup_{x\in J}\int_{\mathbb{R}}\gamma_t^2(y)f(x, y)\,dy = M_0^* < \infty; \tag{B.2}$$

$$\sup_{t\in I}\sup_{x\in J}|D_t(x)| = M_2 < \infty. \tag{B.3}$$

By simple arguments, we obtain:

(i) Under (A.1) and (A.2), (B.2) holds with $M_0^* \le M_0M_1$.
(ii) Under (B.1) and (B.2), (A.1) holds with $M_0 \le M_0^*/m_1$.
(iii) Under (A.2) and (B.2), (B.3) holds with $M_2 \le (M_1M_0^*)^{1/2}$.
(iv) If $\sup_{t\in I}\sup_{y\in\mathbb{R}}|\gamma_t(y)| < \infty$, then (A.1) holds with $J = \mathbb{R}$.

For the case $\gamma_t(y) = y$, (B.2) is a type of assumption used by Mack and Silverman (1982), who also assumed (B.1) and (A.2). Statements (i) and (ii) indicate that we are enabled to have (A.1), (A.2) and (B.2) while bypassing (B.1), thus providing our result on the quantity (2.1) with a broader scope of potential application. [However, in dealing with r_{tn} and establishing (1.4), we will need (B.1).] Assumption (A.1) may be interpreted as requiring the conditional second moments $E[\gamma_t^2(Y)|X = x]$ to be uniformly bounded for $t \in I$, $x \in J$. Statements (i) and (iii) will be used in the proof of Theorem 2.3.

THEOREM 2.1. *Assume* (A.1)–(A.7). *Let* $\{c_n\}$ *satisfy* (i) $0 \le c_n \to 0$, (ii) $\Delta_n = nc_n/\log n \to \infty$ *and* (iii) $1 \le c_n^{-1} \le (n/\log n)^{1-2/\lambda}$, *for* λ *as in* (A.7). *Then, with* α *as in* (A.4),

$$\text{(2.4a)} \qquad \sup_{t\in I}\sup_{x\in J}|D_{tn}(x) - D_t(x)| = O\big(\max\{\Delta_n^{-1/2}, c_n^\alpha\}\big) \quad a.s., \ n \to \infty.$$

Further, in the case $\lambda = \infty$ *in* (A.7), *there exists a number* n_0 *and for each real* $\kappa > 0$ *there exists a constant* A_κ, *not depending on the sequence* $\{c_n\}$, *such that*

$$\text{(2.4b)} \qquad P\Big\{\sup_{t\in I}\sup_{x\in J}|D_{tn}(x) - D_t(x)| > A_\kappa\Delta_n^{-1/2} + M^{(\alpha)}c_n^\alpha\Big\} < n^{-\kappa},$$

all $\kappa > 0$ *and* $n \ge n_0$, *with* $M^{(\alpha)}$ *as in* (A.4).

REMARKS 2.2. (i) By the Borel–Cantelli lemma, if (2.4b) holds, then so does (2.4a).

(ii) From (2.7), Lemma 2.1 and the proof of Lemma 2.2, it will be seen that the constant A_κ in (2.4b) may be taken as $6A + 4$, where A is chosen (sufficiently large) to satisfy

$$\frac{(A - M_1)^2}{2M_0M_1 + \frac{2}{3}(A - M_1)M_\infty} \ge \kappa + 2,$$

with M_0, M_1, M_∞ as in (A.1), (A.2) and (A.7).

(iii) Note that for $2 < \lambda < \infty$ condition (iii) implies condition (ii). For $\lambda = \infty$ the right inequality of condition (iii) follows from condition (ii).

We prove the theorem by decomposing $D_{tn} - D_t$ into a stochastic component $A_{tn} = D_{tn} - ED_{tn}$ and a deterministic (bias) component $B_{tn} = ED_{tn} - D_t$, each to be treated separately. For the bias part, we readily obtain, using (2.3),

LEMMA 2.1. *Under* (A.4) *and for* $c_n \to 0$,

$$\text{(2.5)} \qquad \sup_{t\in I}\sup_{x\in J}|B_{tn}(x)| \le M^{(\alpha)}c_n^\alpha, \quad \textit{for all large } n.$$

For the stochastic component, it is easily checked that $|A_{tn}(x)| \le 2c_n^{-1}V_{tn}(x, c_n)$, where

$$\text{(2.6)} \qquad V_{tn}(x, \delta) = \sup_{|z|\le\delta}|G_{tn}(x + z) - G_{tn}(x) - [G_t(x + z) - G_t(x)]|.$$

Putting

$$V_n = \sup_{t \in I} \sup_{x \in J} V_{tn}(x, c_n),$$

we have

$$\sup_{t \in I} \sup_{x \in J} |A_{tn}(x)| \le 2c_n^{-1} V_n. \tag{2.7}$$

Consequently, Theorem 2.1 follows from Lemma 2.1 with the following central result.

LEMMA 2.2. *Under the conditions of Theorem* 2.1, *excepting* (A.4),

$$V_n = O\left(\Delta_n^{-1/2} c_n\right) \quad a.s., \ n \to \infty. \tag{2.8a}$$

Further, in the case $\lambda = \infty$ *in* (A.7), *there exists a number* n_0 *and for each real* $\kappa > 0$ *there exists a constant* B_κ, *not depending on the sequence* $\{c_n\}$, *such that*

$$P\left\{V_n > B_\kappa \Delta_n^{-1/2} c_n\right\} < n^{-\kappa}, \quad \text{all } \kappa > 0 \text{ and } n \ge n_0. \tag{2.8b}$$

The proof of this lemma is given in Section 4.

From Theorem 2.1, we now can establish analogous results for the estimators d_{tn} and r_{tn} introduced in Section 1, in the case of $\{\beta_t, t \in I\}$.

We shall assume that the family $\{\beta_t, t \in I\}$ has a representation

$$\beta_t(y) = \sum_{i=1}^{i_0} q_i \gamma_{ti}(y), \qquad y \in \mathbb{R}, t \in I, \tag{2.9}$$

with fixed and finite i_0, $q_1, \dots, q_{i_0}$ and with the families $\{\gamma_{ti}, t \in I\}$, $1 \le i \le i_0$, satisfying assumptions (A.1) and (A.3)–(A.7), with common α in (A.4) and common λ in (A.7).

We first consider kernel sequences of *step-function* form,

$$K_n(u) = \sum_{j=1}^{j_n} a_{nj} I\{-b''_{nj} \le u < b'_{nj}\}, \qquad u \in \mathbb{R}, \tag{2.10a}$$

with $\{j_n\}, \{a_{nj}\}, \{b'_{nj}\}, \{b''_{nj}\}$ nonnegative constants such that, with $b_{nj} = b'_{nj} + b''_{nj}$,

$$\sum_{j=1}^{j_n} a_{nj} b_{nj} = 1 \quad \left[\text{i.e., } \int K_n(u)\, du = 1\right], \tag{2.10b}$$

$$\sup_n \sum_{j=1}^{j_n} a_{nj} b_{nj}^{1/2} < \infty \tag{2.10c}$$

and

$$\sup_n \sum_{j=1}^{j_n} a_{nj} b_{nj}^2 < \infty. \tag{2.10d}$$

THEOREM 2.2. *Let* $d_{tn}(\cdot)$ *be defined by* (1.3c), *with* $\{\beta_t, t \in I\}$ *having representation* (2.9), $\{K_n\}$ *having form* (2.10) *and* $\{h_n\}$ *satisfying* (i) $h_n B_n \to 0$,

(ii) $nh_n b_n/\log n \to \infty$ *and* (iii) $B_n \le h_n^{-1} \le b_n(n/\log n)^{1-2/\lambda}$, *where* $b_n = \min_{j \le j_n} b_{nj}$, $B_n = \max_{j \le j_n} b_{nj}$ *and* λ *is given in* (A.7). *Assume also* (A.2) *and either*

$$\text{(2.11a)} \qquad \lambda < \infty; \qquad j_n \equiv j_0 < \infty$$

or

$$\text{(2.11b)} \qquad \lambda = \infty; \qquad j_n = O(n^s), \quad \textit{some } s > 0.$$

Then

$$\text{(2.12)} \quad \sup_{t \in I} \sup_{x \in J} |d_{tn}(x) - d_t(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^\alpha\}\big) \quad a.s., n \to \infty,$$

with $\Delta_n = nh_n/\log n$ *and* α *as in* (A.4).

PROOF. With the assumed forms for $\{\beta_t, t \in I\}$ and $\{K_n\}$, the estimator d_{tn} has a decomposition into terms of the type treated in Theorem 2.1, and accordingly we obtain

$$\text{(2.13)} \qquad \sup_{t \in I} \sup_{x \in J} |d_{tn}(x) - d_t(x)| \le \sum_{i=1}^{i_0} |q_i| S_{ni},$$

where

$$\text{(2.14)} \qquad S_{ni} = \sum_{j=1}^{j_n} a_{nj} b_{nj} \sup_{t \in I} \sup_{x \in J} |D_{tn}^{(i,j)}(x) - D_t^{(i)}(x)|,$$

with

$$D_t^{(i)}(x) = \int_{\mathbb{R}} \gamma_{ti}(y) f(x, y)\, dy,$$

$$D_{tn}^{(i,j)}(x) = c_{nj}^{-1}\big[G_{tn}^{(i)}(x + c'_{nj}) - G_{tn}^{(i)}(x - c''_{nj})\big],$$

$$G_{tn}^{(i)}(x) = n^{-1} \sum_{k=1}^{n} \gamma_{ti}(Y_k) I\{X_k \le x\}$$

and

$$c'_{nj} = h_n b'_{nj}, \qquad c''_{nj} = h_n b''_{nj}, \qquad c_{nj} = h_n b_{nj}.$$

Fix i and j and put $\Delta_{nj} = \Delta_n b_{nj}$. Then conditions (i), (ii) and (iii) assumed in the present theorem yield their counterparts in Theorem 2.1 with $\{c_n\}$ replaced by $\{c_{nj}\}_{n \ge 1}$ and Δ_n replaced by Δ_{nj}, $n \ge 1$, and Theorem 2.1 thus yields

$$\text{(2.15)} \quad \sup_{t \in I} \sup_{x \in J} |D_{tn}^{(i,j)}(x) - D_t^{(i)}(x)| = O\big(\max\{\Delta_{nj}^{-1/2}, c_{nj}^\alpha\}\big) \quad \text{a.s., } n \to \infty.$$

Now suppose that (2.11a) holds. Then (2.15) yields

$$\text{(2.16)} \qquad S_{ni} = O\left(\sum_{j=1}^{j_0} a_{nj} b_{nj} \max\{\Delta_{nj}^{-1/2}, c_{nj}^\alpha\}\right) \quad \text{a.s., } n \to \infty.$$

It is easily seen, using (2.10c) and (2.10d) that the right-hand side of (2.16) is $O(\Delta_n^{-1/2}, h_n^\alpha)$, and thus (2.12) follows, via (2.13).

Alternatively, assume (2.11b), choose real $\kappa > 0$ and put

$$\varepsilon_n = A_\kappa \Delta_n^{-1/2} \sum_{j=1}^{j_n} a_{nj} b_{nj}^{1/2} + M^{(\alpha)} h_n^\alpha \sum_{j=1}^{j_n} a_{nj} b_{nj}^{1+\alpha},$$

with A_κ, $M^{(\alpha)}$ as in Theorem 2.1. Then

$$P\{S_{ni} > \varepsilon_n\} \le \sum_{j=1}^{j_n} P\Big\{ \sup_{t \in I} \sup_{x \in J} |D_{tn}^{(i,j)}(x) - D_t^{(i)}(x)| > A_\kappa \Delta_{nj}^{-1/2} + M^{(\alpha)} c_{nj}^\alpha \Big\}$$

and by (2.4b) of Theorem 2.1 we obtain

$$P\{S_{ni} > \varepsilon_n\} \le j_n n^{-\kappa}, \quad \text{all } n \ge n_0.$$

Choosing $\kappa > s + 1$, with s as in (2.11b), and applying the Borel–Cantelli lemma, we obtain

$$S_{ni} = O(\varepsilon_n) \quad \text{a.s., } n \to \infty. \tag{2.17}$$

Again using (2.10c) and (2.10d), we have $\varepsilon_n = O(\max\{\Delta_n^{-1/2}, h_n^\alpha\})$, $n \to \infty$, and thus (2.12) follows, via (2.13). □

As discussed in Example 4 of Section 1, Theorem 2.2 yields a result on density estimation with discrete kernels, as follows.

COROLLARY 2.1. *Let $f_n(\cdot)$ be defined by* (1.3b) *with $\{K_{0n}\}$ having form* (2.10) *with $j_n = O(n^s)$, some $s > 0$, and with $\{h_n\}$ satisfying* (i), (ii) *and* (iii) *of Theorem* 2.2 *with $\lambda = \infty$. Assume* (A.2) *and that f_0 is ulL-α on J for some α, $0 < \alpha \le 1$. Then*

$$\sup_{x \in J} |f_n(x) - f_0(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^\alpha\}\big) \quad a.s., \ n \to \infty, \tag{2.18}$$

with $\Delta_n = nh_n/\log n$.

PROOF. With the family $\{\beta_t,\ t \in I\}$ reduced to the single function $\beta(y) \equiv 1$, d_{tn} given by (1.3c) reduces to f_n in form, and, under the present assumptions, the conditions of Theorem 2.2 are satisfied with the option (2.11b). Thus (2.12) holds, which is the same as (2.18). □

This corollary not only extends the results of Serfling (1982) to a wider scope of kernels and thus is of independent interest, but also serves as a tool in developing our result for r_{tn}, as follows.

THEOREM 2.3. (i) *Discrete kernels. Let $r_{tn}(\cdot)$ be defined by* (1.3) *with $\{\beta_t,\ t \in I\}$ having representation* (2.9), *$\{K_n\}$ having form* (2.10) *and $\{K_{0n}\}$ having form* (2.10) *with constants $\{\tilde{j}_n\}$, $\{\tilde{a}_{nj}\}$, $\{\tilde{b}'_{nj}\}$, $\{\tilde{b}''_{nj}\}$ and with $\tilde{j}_n = O(n^{\tilde{s}})$ for some*

$\tilde{s} > 0$. *Let* $\{h_n\}$ *satisfy*

(a) $h_n \max\{B_n, \tilde{B}_n\} \to 0$,
(b) $(\log n)^{-1} n h_n \min\{b_n, \tilde{b}_n\} \to \infty$,
(c) $B_n \le h_n^{-1} \le b_n (n/\log n)^{1-2/\lambda}$ *and*
(d) $\tilde{B}_n \le h_n^{-1} \le \tilde{b}_n (n/\log n)$,

with $b_n = \min_{j \le j_n} b_{nj}$, $B_n = \max_{j \le j_n} b_{nj}$, $\tilde{b}_n = \min_{j \le \tilde{j}_n} \tilde{b}_{nj}$, $\tilde{B}_n = \max_{j \le \tilde{j}_n} \tilde{b}_{nj}$ *and* λ *as in* (A.7). *Assume* (A.2), (B.1), f_0 *ulL*-α_0 *on* J *for some* α_0, $0 < \alpha_0 \le 1$, *and either* (2.11a) *or* (2.11b). *Then*

$$\sup_{t \in I} \sup_{x \in J} |r_{tn}(x) - r_t(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^{\tilde{\alpha}}\}\big) \quad a.s.,\ n \to \infty, \tag{2.19}$$

with $\Delta_n = n h_n / \log n$ *and* $\tilde{\alpha} = \min\{\alpha, \alpha_0\}$, *for* α *as in* (A.4).

(ii) *Smooth kernels. Let* $r_{tn}(\cdot)$ *be defined by* (1.3) *with* $\{\beta_t,\ t \in I\}$ *having representation* (2.9) *and with* $K_n(\cdot) \equiv K(\cdot)$, $K_{0n}(\cdot) \equiv K_0(\cdot)$, *where* K *is symmetric, has bounded support and bounded first two derivatives and* K_0 *satisfies similar conditions. Assume* (A.2), (B.1) *and* f_0 *ulL*-α_0 *on* J *for some* α_0, $0 < \alpha_0 \le 1$. *Assume that* $\{h_n\}$ *satisfies*

(a) $h_n \to 0$,
(b) $n h_n / \log n \to \infty$ *and*
(c) $B \le h_n^{-1} \le b(n/\log n)^{1-2\lambda}$,

for some constants b *and* B *and for* λ *as in* (A.7). *Then* (2.19) *holds.*

Proof. (i) It is immediate that, under the assumptions of the present theorem, the conditions of Theorem 2.2 and Corollary 2.1 are satisfied, and we have (2.12) as well as

$$\sup_{x \in J} |f_n(x) - f_0(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^{\alpha_0}\}\big) \quad \text{a.s.},\ n \to \infty. \tag{2.20}$$

We now apply relation (1.5). By (B.1), (2.12) and (2.20), we have

$$\sup_{t \in I} \sup_{x \in J} |R_{tn}(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^{\alpha}\}\big) \quad \text{a.s.},\ n \to \infty. \tag{2.21}$$

Using (B.1) again, as well as (B.3) [see statements (i) and (iii) of Remarks 2.1] and (2.20), we have

$$\sup_{t \in I} \sup_{x \in J} |S_{tn}(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^{\alpha_0}\}\big) \quad \text{a.s.},\ n \to \infty. \tag{2.22}$$

Combining (2.21) and (2.22), we have (2.19).

(ii) Let K be symmetric with bounded support, say $\subset [-1, 1]$ and let us introduce an associated sequence $\{K_n\}$ of discrete kernels, defined by

$$K_n(u) = \sum_{i=1}^{j_n} I\{(i-1)\delta_n < u \le i\delta_n\} K(i\delta_n), \qquad u > 0,$$

where $j_n = [\delta_n^{-1}] + 1$, with $0 < \delta_n \to 0$ and $K_n(u) = K_n(-u)$, for $u < 0$, and

$K_n(0) = K(\delta_n)$. For this kernel the regularity conditions (2.10c) and (2.10d) reduce to

$$\sup_n \sum_{i=1}^{j_n} [K(i\delta_n) - K((i+1)\delta_n)](i\delta_n)^{1/2} < \infty$$

and

$$\sup_n \sum_{i=1}^{j_n} [K(i\delta_n) - K((i+1)\delta_n)](i\delta_n)^2 < \infty.$$

For K'' bounded, these reduce to

$$\sup_n \delta_n^{3/2} \sum_{i=1}^{j_n} i^{1/2}|K'(i\delta_n)| < \infty$$

and

$$\sup_n \delta_n^3 \sum_{i=1}^{j_n} i^2|K'(i\delta_n)| < \infty,$$

which in turn are satisfied if we have $\int x^2|K'(x)|\,dx < \infty$, which indeed follows from our restrictions on $K(\cdot)$. Similar considerations apply in connection with $K_0(\cdot)$.

Now note that for $\gamma_t(\cdot)$ and $K(\cdot)$ bounded, we have

$$\sup_{t\in I} \sup_{x\in J} |d_{tn}(x; K) - d_{tn}(x; K_n)| = O\left(\frac{\delta_n}{h_n}\right),$$

where $d_{tn}(x; L)$ denotes (1.3c) based on the kernel $L(\cdot)$. Therefore, we can take $\delta_n = O(h_n^2)$ so that $j_n = O(h_n^{-2})$, which is $O(n^s)$, for some $s > 0$ under our condition on the bandwidth. Thus we may now apply part (i) to obtain (2.19) again in the present case. □

Useful corollaries of Theorem 2.3 are obtained by choosing h_n to make the rates $\Delta_n^{-1/2}$ and $h_n^{\tilde{\alpha}}$ agree. In particular, for the case of discrete kernels we have

COROLLARY 2.2. *Let $r_{tn}(\cdot)$ be defined by* (1.3) *with $\{\beta_t, t \in I\}$ having representation* (2.9), *$\{K_n\}$ having form* (2.10) *and $\{K_{0n}\}$ having form* (2.10) *with constants $\{\tilde{j}_n\}, \{\tilde{a}_{nj}\}, \{\tilde{b}'_{nj}\}, \{\tilde{b}''_{nj}\}$ and with $\tilde{j}_n = O(n^{\tilde{s}})$, for some $\tilde{s} > 0$. Assume $\alpha = 1$, in* (A.4) *and $\lambda > 3$ in* (A.7). *Assume* (A.2), (B.1), *f_0 ulL-1 on J and either* (2.11a) *or* (2.11b). *With the notation of Theorem* 2.3, *assume*

(i) $$\max\{B_n, \tilde{B}_n\} = o\left((n/\log n)^{1/3}\right),$$

(ii) $$(n/\log n)^{-2/3} = o\left(\min\{b_n, \tilde{b}_n\}\right) \le c_0\tilde{b}_n,$$

(iii) $$(n/\log n)^{-2(1/3-1/\lambda)} \le c_0 b_n,$$

for some constant $c_0 > 0$. Let $h_n \sim c_0(n/\log n)^{-1/3}$ in (1.3). *Then*

$$\sup_{t\in I}\sup_{x\in J}|r_{tn}(x) - r_t(x)| = O\left((n/\log n)^{-1/3}\right) \quad a.s., n \to \infty. \tag{2.23}$$

Proof. It is readily seen that this choice of h_n and (i), (ii) and (iii) in the preceding discussion yield (a)–(d) of Theorem 2.3(i). Also, the other assumptions of Theorem 2.3(i) are obviously fulfilled by the present assumptions. Thus (2.19) holds and yields (2.23). □

Remarks 2.3. (i) An analogue of (2.23) for the density estimator f_n may be easily derived.

(ii) It would be of interest, in the case $\lambda < \infty$, to relax the restriction (2.11a) on $\{j_n\}$ in Theorems 2.2 and 2.3(i) to a condition of form $j_n = O(n^s)$, for some $s > 0$. However, this would require a strengthened version of Lemma 2.2 with (2.8b) extended to the case $\lambda < \infty$. The present proof of Lemma 2.2, given in Section 4, does not appear to yield such a strengthening, due to the complication presented by the truncation step involving the random variable W_n in (4.9). A possible approach could be to control the rate at which the r.h.s. of (4.11) converges to 0.

(iii) From the proofs of Theorem 2.2, Corollary 2.1, Theorem 2.3(i) and Corollary 2.2, it is clear that the restrictions on $\{K_n\}$ and $\{K_{0n}\}$ may be dropped or relaxed, at the expense of introducing further factors (involving b_n, $\tilde{b}_n$, B_n, $\tilde{B}_n$, etc.) into the rates expressed in the relations (2.12), (2.18), (2.19) and (2.22).

(iv) In the case of single step-function kernels $K(\cdot)$ and $K_0(\cdot)$, with finitely many jumps in place of the sequences $\{K_n\}$ and $\{K_{0n}\}$, the restrictions on $\{h_n\}$ in Theorem 2.2, Corollary 2.1 and Theorem 2.3(i) reduce to those given by (a), (b) and (c) in Theorem 2.3(ii), with $b = \min\{b_1, \dots, b_{j_0}, \tilde{b}_1, \dots, \tilde{b}_{\tilde{j}_0}\}$, $B = \max\{b_1, \dots, b_{j_0}, \tilde{b}_1, \dots, \tilde{b}_{\tilde{j}_0}\}$ and λ as in (A.7)

(v) From the proof of Theorem 2.2 it is easily seen that in the case $\lambda = \infty$ we may express (2.12) in the form

$$\sup_{t \in I} \sup_{x \in J} |d_{tn}(x) - d_t(x)| \le A\Delta_n^{-1/2} + A'h_n^\alpha, \quad \text{all large } n, \text{a.s.},$$

with A and A' constants not depending on n.

(vi) We may also consider smooth kernels with *unbounded* support, by restricting attention to a finite interval of increasing length. For example, in the case of the standard normal density, we restrict to $[-t_n, t_n]$ with $t_n = n^\alpha$ for some $\alpha > 0$, take $\delta_n = n^{-\beta} \ge \exp(-n^{2\alpha}/2)$ for some $\beta > 0$ and finally note that $j_n = O(t_n \delta_n^{-1}) = O(n^{\alpha+\beta})$.

3. Strong consistency rates in selected applications. Using Theorem 2.3(i) and Corollary 2.2, we develop strong consistency rates for the applications discussed in Section 1, except for density estimation (Example 4), which has been treated in Corollary 2.1.

For convenience and simplicity, we confine our attention to the case that the kernels in (1.3b) and (1.3c) are step-functions not depending on n and having finitely many jumps. Thus [see Remark 2.3(iv)] throughout this section we shall assume the following standard conditions and notation with respect to the bandwidth sequence $\{h_n\}$ and kernels K and K_0 in (1.3b) and (1.3c):

(3.1a) $$h_n \to 0,$$

(3.1b) $$\Delta_n = nh_n/(\log n) \to \infty,$$

(3.1c) $$B \le h_n^{-1} \le b(n/\log n)^{1-2/\lambda},$$

with b, B defined as in Remark 2.3(iv) and λ a constant to be specified in each particular application.

All of the results to be given have extensions to general kernel sequences of form (2.39), at the expense of complicating the formulation. We also could develop some analogous results for smooth kernels, but we omit this in the interest of brevity.

3.1. *Nonparametric regression function estimation.* As in Example 1, we take $\beta_t(y) \equiv \beta(y) = y$, in which case the representation (2.38) becomes $\beta(y) = \gamma_1(y) - \gamma_2(y)$, with $\gamma_1(y) = \max\{0, y\}$ and $\gamma_2(y) = -\min\{0, y\}$. Then the assumptions (A.1)–(A.7), (B.1) and f_0 ulL may be reduced to:

(3.2a) $$\sup_{x \in J} \int_{\mathbf{R}} y^2 f(y|x)\,dy = M_0 < \infty;$$

(3.2b) $$E|Y|^\lambda < \infty, \quad \text{with } 2 < \lambda \le \infty;$$

(3.2c) $$0 < m_1 \le f_0(x) \le M_1 < \infty, \qquad x \in J;$$

(3.2d) the functions $f_0(x)$ and $g_0(x) = \int y f(x, y)\,dy$ are ulL-α on J, with $0 < \alpha \le 1$.

Thus Theorem 2.3(i) and Corollary 2.2 yield the following result.

THEOREM 3.1. *Assume* (3.2) *and let* $\{h_n\}$ *satisfy* (3.1) *with* λ *as in* (3.2b). *Then*

(3.3) $$\sup_{x \in J} |r_n(x) - r(x)| = O\left(\max\{\Delta_n^{-1/2}, h_n^\alpha\}\right) \quad a.s., n \to \infty.$$

In the case $\alpha = 1$ *and* $\lambda \ge 3$ *and for* $h_n \sim C_0(n/\log n)^{-1/3}$, *we have* $O((n/\log n)^{-1/3})$ *in* (3.3).

Let us compare, for example, with Theorem B of Mack and Silverman (1982). There the kernels $K(\cdot)$ and $K_0(\cdot)$ are taken to be equal, symmetric and subject to a set of smoothness conditions. Our (3.2b) and (3.2c) are also assumed, but in place of (3.2a) is the stronger requirement (see Remark 2.1) $\sup_{x \in J} \int_{\mathbf{R}} |y|^\lambda f(x, y)\,dy < \infty$, with λ as in (3.2b). Also, the functions f_0 and g_0 in our (3.2d) are assumed to have bounded second derivatives [thus implying (3.2d) with $\alpha = 1$]. As bandwidth assumptions, our (3.1a) and an equivalent of (3.1b) are assumed, and a slightly stronger version of (3.1c), namely that $n^\eta h_n \to \infty$ for some $\eta < 1 - 2/\lambda$, is assumed. Also, $\Sigma_n h_n^s < \infty$ for some $s > 0$, and $(*)$ $h_n = O(\tilde{\Delta}_n^{-1/2})$, where $\tilde{\Delta}_n = nh_n/\log(1/h_n)$, are assumed. (Note that $\tilde{\Delta}_n \sim \Delta_n$ for all typical choices of $\{h_n\}$.) Their theorem asserts for the quantity in (3.3) the a.s.

rate $O(\tilde{\Delta}_n^{-1/2})$, which is compatible with our rate under their additional assumption ($*$). In summary, our theorem considers step-function kernels instead of smooth kernels, requires weaker moment assumptions, weaker regularity assumptions and weaker bandwidth restrictions and provides a rate in (3.3) more sensitive to the regularity assumptions.

In particular, given (3.2d) with $\alpha = 1$ (implied by Mack and Silverman's conditions), the optimal rate in (3.3) is $n^{-1/3}$ (ignoring log factors). This, in view of ($*$), is also the optimal rate attainable in Mack and Silverman's Theorem B. For such a rate, our theorem requires $\lambda \geq 3$ in (3.2b), whereas their theorem requires $\lambda > 3$ and regularity stronger than (3.2d) with $\alpha = 1$.

3.2. *Nonparametric scale curve estimation.* This may be handled very much like the previous application (see discussion of Example 2 in Section 1), and we shall leave the details implicit.

3.3. *The conditional distribution function.* With the family $\{\beta_t,\ t \in I\}$ given by $\beta_t(y) = I\{y \leq t\}$, $y \in \mathbb{R}$, $t \in I = \mathbb{R}$, the quantity $r_t(x)$ defined by (1.1) becomes the conditional df $F(t|x)$. Let us take $K(\cdot)$ and $K_0(\cdot)$ in (1.3b) and (1.3c) to be equal, in which case the quantity $r_{tn}(x)$ in (1.3a) becomes a df (in the variable t), which we shall denote by $F_n(t|x)$, $t \in \mathbb{R}$, for each x. For the present choice of $\beta_t(\cdot)$, representation (2.38) holds trivially and assumptions (A.1)–(A.7), (B.1) and f_0 ulL may be reduced to:

(3.4a) $\qquad F_y$, the marginal df of Y, is continuous;

(3.4b) $\qquad 0 < m_1 \leq f_0(x) \leq M_1 < \infty, \qquad x \in J;$

(3.4c) $\qquad f_0(\cdot)$ is ulL-α on J, and the functions $F(t|\cdot)$, $t \in \mathbb{R}$, are ulL-α on J, uniformly in $t \in \mathbb{R}$, with $0 < \alpha \leq 1$.

Thus Theorem 2.3(i) and Corollary 2.2 yield

THEOREM 3.2. *Assume* (3.4) *and let* $\{h_n\}$ *satisfy* (3.1) *with* $\lambda = \infty$. *Then*

$$\sup_{x \in J} \sup_{t \in \mathbb{R}} |F_n(t|x) - F(t|x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^\alpha\}\big) \quad a.s.,\ n \to \infty. \tag{3.5}$$

3.4. *L-smoothing.* For L-smoothers with trimmed weight functions, uniform strong consistency rates may be obtained by reduction to an application of rates established for conditional df estimators, for example as given in the preceding section. As discussed in Example 5 of Section 1, we consider a conditional L-functional of the form

$$l(x) = \int_0^1 J_0(v) F^{-1}(v|x)\,dv, \qquad x \in J. \tag{3.6}$$

Let $\hat{l}_n(x)$ be a corresponding estimator defined by replacing $F^{-1}(v|x)$ in (3.6) by $\hat{F}_n^{-1}(v|x)$, where $\hat{F}_n(\cdot|x)$ is a df for each x and is uniformly strongly consistent for estimation of $F(\cdot|x)$, in the sense that

$$Z_n = \sup_{x \in J} \sup_{t \in \mathbb{R}} |\hat{F}_n(t|x) - F(t|x)| \to 0 \quad \text{a.s.},\ n \to \infty. \tag{3.7}$$

Assume that $J_0(\cdot)$ satisfies

(3.8) $J_0(\cdot)$ is bounded on $[v_0, v_1]$ and vanishes elsewhere, with $0 < v_0 < v_1 < 1$.

We shall utilize the following elementary reduction lemma. For any function $g(\cdot)$, let $\|g\|_\infty$ denote $\sup|g(\cdot)|$.

LEMMA 3.1. *Let $J_0(\cdot)$ satisfy* (3.8) *and let F and G be arbitrary df's. Then*

$$\left|\int_0^1 J_0(G^{-1} - F^{-1})\right| \le \|J_0\|_\infty \left[F^{-1}(v_1 + \delta) - F^{-1}(v_0 - \delta)\right]\delta, \tag{3.9}$$

where $\delta = \|G - F\|_\infty$ and v_0, v_1 are as in (3.8).

PROOF. Put $H_0(u) = \int_0^u J_0(v)\,dv$, $0 < u < 1$, and $y_0 = F^{-1}(v_0 - \delta)$, $y_1 = F^{-1}(v_1 + \delta)$. Then, using integration by parts, (3.8) and the inequalities $\max\{F(y), G(y)\} < v_0$ for $y < y_0$, $\min\{F(y), G(y)\} > v_1$ for $y > y_1$, we have

$$\int_0^1 J_0(v)\left[G^{-1}(v) - F^{-1}(v)\right] dv = -\int_{-\infty}^{\infty}\left[H_0(G(y)) - H_0(F(y))\right] dy$$
$$= -\int_{y_0}^{y_1}\left[H_0(G(y)) - H_0(F(y))\right] dy.$$

Now, using $|H_0(u) - H_0(u')| \le \|J_0\|_\infty |u - u'|$, we obtain (3.9). □

We now give a general uniform strong convergence result for estimators $\hat{l}_n(x)$ formulated as before. We shall suppose that the given family of conditional df's, $\{F(\cdot|x),\ x \in J\}$, satisfies

$$a_0 < F^{-1}(v_0 - \varepsilon_0|x) < F^{-1}(v_1 + \varepsilon_0|x) < a_1, \quad \text{all } x \in J, \tag{3.10}$$

with $-\infty < a_0 < a_1 < \infty$, $\varepsilon_0 < \min\{v_0, 1 - v_1\}$ and v_0, v_1 as in (3.8).

THEOREM 3.3. *Let $l(\cdot)$ be defined by* (3.6), *with $J_0(\cdot)$ satisfying* (3.8) *and $\{F(\cdot|x),\ x \in J\}$ satisfying* (3.10). *Let $\hat{l}_n(\cdot)$ be based on a family $\{\hat{F}_n(\cdot|x),\ x \in J\}$ satisfying* (3.7). *Then*

$$\sup_{x \in J}|\hat{l}_n(x) - l(x)| = O(Z_n) \quad a.s.,\ n \to \infty. \tag{3.11}$$

PROOF. For each $x \in J$, we apply Lemma 3.1 with F and G given by $F(\cdot|x)$ and $\hat{F}_n(\cdot|x)$, respectively. Combining these results, we obtain

$$\sup_{x \in J}|\hat{l}_n(x) - l(x)| \le \|J_0\|_\infty Z_n \sup_{x \in J}\left[F^{-1}(v_1 + Z_n|x) - F^{-1}(v_0 - Z_n|x)\right]. \tag{3.12}$$

By (3.7) and (3.10), the third factor on the right-hand side of (3.12) is a.s. bounded above by $(a_1 - a_0)$ for all large n. Thus (3.11) follows. □

Let us now consider the special case that $l(x)$ is estimated by $l_n(x)$ based on the family $\{F_n(\cdot|x),\ x \in J\}$ considered in Section 3.3. We have in this case the following explicit rate.

COROLLARY 3.1. *Let $l(\cdot)$ be defined by* (3.6), *with $J_0(\cdot)$ satisfying* (3.8) *and $\{F(\cdot|x),\ x \in J\}$ satisfying* (3.10). *Let $l_n(x)$ be based on the family $\{F_n(\cdot|x),\ x \in J\}$ considered in Theorem* 3.2 *and assume the conditions of that theorem. Then*

$$\sup_{x \in J} |l_n(x) - l(x)| = O\big(\max\{\Delta_n^{-1/2}, h_n^{\alpha}\}\big) \quad a.s.,\ n \to \infty.$$

3.5. *M-smoothing.* Continuing the discussion in Example 6 of Section 1, we establish here, for a fixed $\psi(\cdot)$ function, a uniform strong convergence rate for $r_{\psi n}$. We apply the results of Section 2 by taking $\beta_t(y) = \psi(y - t)$, $y \in \mathbb{R}$, for $t \in I = \mathbb{R}$. In this case (1.2) becomes

$$d_t(x) = \int_{\mathbb{R}} \psi(y - t) f(x, y)\, dy \tag{3.13}$$

and (1.3c) becomes, for a fixed kernel $K(\cdot)$,

$$d_{tn}(x) = (nh_n)^{-1} \sum_{i=1}^{n} \psi(Y_i - t) K\left(\frac{x - X_i}{h_n}\right). \tag{3.14}$$

Clearly, we may characterize $r(x)$ and $r_{\psi n}(x)$ as the solutions, with respect to t, of the equations

$$d_t(x) = 0, \tag{3.15a}$$

$$d_{tn}(x) = 0, \tag{3.15b}$$

respectively. Accordingly, we shall reduce the problem of strong convergence of $r_{\psi n}(x)$ to $r(x)$, uniformly in x, to an application of the strong convergence of $d_{tn}(x)$ to $d_t(x)$, uniformly in x and t, as given by Theorem 2.2.

To apply Theorem 2.2, we satisfy the representation (2.38) for $\{\beta_t,\ t \in \mathbb{R}\}$ by taking $q_1 = q_2 = -1$, $\gamma_{t1}(y) = \max\{0, -\psi(y - t)\}$ and $\gamma_{t2}(y) = \min\{0, -\psi(y - t)\}$ and adopting the following assumptions:

(3.16a) $\psi(\cdot)$ is bounded, antisymmetric, monotone (incr.) and continuous;

$$0 < m_1 \le f_0(x) \le M_1 < \infty, \qquad x \in J; \tag{3.16b}$$

(3.16c) the conditional densities $f(\cdot|y)$, $y \in \mathbb{R}$, are ulL-α on J, uniformly in $y \in \mathbb{R}$, with $0 < \alpha \le 1$.

It is readily seen that these yield (A.1)–(A.7), with $\lambda = \infty$ in (A.7), and thus from Theorem 2.2 and Remark 2.3(v) we immediately have

LEMMA 3.2. *Let $d_t(\cdot)$ and $d_{tn}(\cdot)$ be given by* (3.13) *and* (3.14). *Assume* (3.16) *and let $\{h_n\}$ satisfy* (3.1) *with $\lambda = \infty$. Then, for some constant A^*, we have a.s.*

$$\sup_{t \in \mathbb{R}} \sup_{x \in J} |d_{tn}(x) - d_t(x)| \le A^* \max\{\Delta_n^{-1/2}, h_n^{\alpha}\}, \quad \textit{all large } n. \tag{3.17}$$

For our result on $r_{\psi n}(\cdot)$, we shall also require

$$\inf_{x \in J} \left| \int \psi(y - r(x) + \varepsilon)\, dF(y|x) \right| \ge q_0 |\varepsilon|, \quad \text{for } |\varepsilon| \le \delta, \tag{3.18}$$

where δ and q_0 are some positive constants. [This assumption is also used by Härdle and Luckhaus (1984); see their discussion.]

THEOREM 3.4. *Under the conditions of Lemma* 3.2 *and also assuming* (3.18), *we have a.s.*

$$\sup_{x \in J} |r_{\psi n}(x) - r(x)| \le B^* \max\{\Delta_n^{-1/2}, h_n^\alpha\}, \quad \textit{all large } n, \tag{3.19}$$

with $B^* = 2A^*/m_1 q_0$.

PROOF. By the monotonicity of ψ and the definition of $r_{\psi n}(x)$ as solution of (3.15b), we have, for $\varepsilon > 0$,

$$r_{\psi n}(x) > r(x) + \varepsilon \Rightarrow d_{r(x)+\varepsilon, n}(x) > 0. \tag{3.20}$$

Now

$$d_{r(x)+\varepsilon, n}(x) \le d_{r(x)+\varepsilon}(x) + \sup_{t \in \mathbb{R}} |d_{tn}(x) - d_t(x)|. \tag{3.21}$$

Also, by monotonicity of $\psi(\cdot)$ and the identity $d_{r(x)}(x) = 0$, the function $d_{r(x)+\varepsilon}(x)$ is nonpositive and by (3.16b) and (3.18) has magnitude $\ge m_1 q_0 \varepsilon$, for $0 < \varepsilon < \delta$. That is, for $0 < \varepsilon < \delta$,

$$d_{r(x)+\varepsilon}(x) \le -m_1 q_0 \varepsilon. \tag{3.22}$$

Combining (3.20), (3.21) and (3.22), we have, for $0 < \varepsilon < \delta$,

$$r_{\psi n}(x) > r(x) + \varepsilon \Rightarrow \sup_{t \in \mathbb{R}} |d_{tn}(x) - d_t(x)| > m_1 q_0 \varepsilon.$$

With a similar inequality proved for the case $r_{\psi n}(x) < r(x) - \varepsilon$, we obtain, for $0 < \varepsilon < \delta$,

$$\sup_{x \in J} |r_{\psi n}(x) - r(x)| > \varepsilon \Rightarrow \sup_{r \in \mathbb{R}} \sup_{x \in J} |d_{tn}(x) - d_t(x)| > m_1 q_0 \varepsilon. \tag{3.23}$$

It readily follows that (3.23) and (3.17) imply (3.19). □

4. Proof of Lemma 2.2. Put

$$a_n = \Delta_n^{-1/2} c_n = n^{-1/2} (c_n \log n)^{1/2}. \tag{4.1}$$

We first reduce $\sup_{t \in I}$ in (2.7) to a maximum over a finite set. By (A.3), (A.5)–(A.7) and the monotone convergence theorem, the function $g(t) = E\gamma_t(Y)$ is nondecreasing and continuous in t with finite limits $g(t_*)$ and $g(t^*)$ as $t \to t_*$ and t^*. Let us partition I by finite points $t_1 < t_2 < \cdots < t_{N_n}$ such that $g(t_1) - g(t_*) \le a_n$, $g(t^*) - g(t_{N_n}) \le a_n$ and $g(t_j) - g(t_{j-1}) \le a_n$ for $2 \le j \le N_n$. Clearly, we may arrange that

$$N_n \le 2(g(t^*) - g(t_*))/a_n. \tag{4.2}$$

Also, for fixed x and z, the functions $G_{tn}(x+z) - G_{tn}(x)$ and $G_t(x+z) - G_t(x)$ are monotone in t and, by (A.3) and (A.6) a.s., these functions for all x and z have finite limits as $t \to t_*, t^*$.

Letting I_n denote the set $\{t_*, t_1, \dots, t_{N_n}, t^*\}$ and I_n^* the set $\{(t_*, t_1), (t_1, t_2), \dots, (t_{N_n}, t^*)\}$, we therefore have, for arbitrary $t \in I$,

$$\begin{aligned} &|G_{tn}(x+z) - G_{tn}(x) - [G_t(x+z) - G_t(x)]| \\ &\quad \le \max_{t \in I_n} |G_{tn}(x+z) - G_{tn}(x) - [G_t(x+z) - G_t(x)]| \\ &\qquad + \max_{(s,t) \in I_n^*} |G_t(x+z) - G_t(x) - [G_s(x+z) - G_s(x)]|. \end{aligned} \tag{4.3}$$

Now, by nonnegativity of $\{\gamma_t,\ t \in I\}$ [by (A.3)], for $s < t$, the function $G_t(x) - G_s(x)$ is nonnegative and nondecreasing in x, so that for $(s, t) \in I_n^*$ we have

$$(4.4)\qquad |G_t(x) - G_s(x)| \le G_t(\infty) - G_s(\infty) = g(t) - g(s) \le a_n, \quad \text{all } x.$$

It follows from (4.3) and (4.4) that

$$(4.5)\qquad V_n \le \max_{t \in I_n} \sup_{x \in J} V_{tn}(x, c_n) + 2a_n.$$

Next we reduce $\sup_{x \in J}$ to a maximum over a finite set. In this case, we first transform to a supremum over a finite interval, as follows. Define

$$\tilde{G}_{tn}(v) = n^{-1} \sum_{i=1}^{n} \gamma_t(Y_i) I\{F_0(X_i) \le v\}$$

and $\tilde{G}_t(v) = E\tilde{G}_{tn}(v)$, $0 \le v \le 1$, where F_0 denotes the (continuous) df of X. Then $G_{tn}(x) = \tilde{G}_{tn}(F_0(x))$ a.s., $G_t(x) = \tilde{G}_t(F_0(x))$ and, by (A.2), $|F_0(x + z) - F_0(x)| \le M_1|z|$. Define

$$\tilde{V}_{tn}(v, \delta) = \sup_{|u| \le \delta} |\tilde{G}_{tn}(v + u) - \tilde{G}_{tn}(v) - [\tilde{G}_t(v + u) - \tilde{G}_t(v)]|.$$

Then $V_{tn}(x, c_n) \le \tilde{V}_{tn}(F_0(x), M_1 c_n)$ a.s. and hence (4.5) yields

$$(4.6)\qquad V_n \le \max_{t \in I_n} \sup_{v \in F_0(J)} \tilde{V}_{tn}(v, M_1 c_n) + 2a_n \quad \text{a.s.}$$

We now partition the interval $[0,1]$ by $v_0 = 0$ and $v_k = k[2/M_1 c_n]^{-1}$ for $1 \le k \le [2/M_1 c_n]$, where $[\cdot]$ denotes greatest integer part. Let $|u| \le M_1 c_n$. For v and $v + u$ in the same subinterval $[v_k, v_{k+1}]$, we easily find

$$|\tilde{G}_{tn}(v + u) - \tilde{G}_{tn}(v) - [\tilde{G}_t(v + u) - \tilde{G}_t(v)]| \le 2\tilde{V}_{tn}(v_k, M_1 c_n),$$

and for v and $v + u$ in $[v_k, v_{k+1}]$ and $[v_j, v_{j+1}]$, respectively, with $k < j$, we find

$$\begin{aligned} &|\tilde{G}_{tn}(v + u) - \tilde{G}_{tn}(v) - [\tilde{G}_t(v + u) - \tilde{G}_t(v)]| \\ &\quad \le \tilde{V}_{tn}(v_j, M_1 c_n) + 2\tilde{V}_{tn}(v_{k+1}, M_1 c_n). \end{aligned}$$

It follows that

$$(4.7)\qquad V_n \le 3 \max_{t \in I_n} \max_{v \in \tilde{J}_n} \tilde{V}_{tn}(v, M_1 c_n) + 2a_n \quad \text{a.s.},$$

with $\tilde{J}_n = \{0, [2/M_1 c_n]^{-1}, 2[2/M_1 c_n]^{-1}, \dots, 1\}$.

In order to set the stage for an application of Bernstein's inequality, we now replace $\tilde{V}_{tn}(v, M_1 c_n)$ by an analogue given by replacing $\tilde{G}_{tn}$ and $\tilde{G}_t$ by analogues based on truncation of $\{\gamma_t(Y_i)\}$. Put

$$(4.8)\qquad Q_n = M_\lambda a_n^{-1/(\lambda - 1)}, \qquad n \ge 1,$$

and

$$H_{tn}(v) = n^{-1} \sum_{i=1}^{n} \gamma_t(Y_i) I\{\gamma_t(Y_i) \le Q_n\} I\{F_0(X_i) \le v\},$$

define $V_{tn}^*(v, \delta)$ by substitution of H_{tn} for $\tilde{G}_{tn}$ and EH_{tn} for $\tilde{G}_t$ in the definition of $\tilde{V}_{tn}(v, \delta)$ and define

$$V_n^* = \max_{t \in I_n} \max_{v \in \tilde{J}_n} V_{tn}^*(v, M_1 c_n).$$

Then (4.7) yields

$$V_n \le 3V_n^* + 3a_n(2/3 + W_n + \theta_n) \quad \text{a.s.,} \tag{4.9}$$

where

$$W_n = a_n^{-1} \sup_{t\in I_n} \sup_{v\in \tilde{J}_n} \sup_{|u|\le M_1 c_n} |\tilde{G}_{tn}(v+u) - \tilde{G}_{tn}(v) - [H_{tn}(v+u) - H_{tn}(v)]|$$

and

$$\theta_n = a_n^{-1} \sup_{t\in I_n} \sup_{v\in \tilde{J}_n} \sup_{|u|\le M_1 c_n} |\tilde{G}_t(v+u) - \tilde{G}_t(v) - [EH_{tn}(v+u) - EH_{tn}(v)]|.$$

Note that $W_n \equiv 0$ and $\theta_n \equiv 0$ in the case $\lambda = \infty$.

Using monotonicity of γ_t in t [by (A.3)] and (A.6) and noting that $a_n^{-1} = (Q_n/M_\lambda)^{\lambda-1}$, we readily obtain

$$\begin{aligned} M_\lambda^{\lambda-1} W_n &\le Q_n^{\lambda-1} n^{-1} \sum_{i=1}^n \gamma_{t^*}(Y_i) I\{\gamma_{t^*}(Y_i) > Q_n\} \\ &\le n^{-1} \sum_{i=1}^n [\gamma_{t^*}(Y_i)]^\lambda I\{\gamma_{t^*}(Y_i) > Q_n\}. \end{aligned} \tag{4.10}$$

For fixed Q, we have by (A.7) and the classical SLLN,

$$n^{-1} \sum_{i=1}^n [\gamma_{t^*}(Y_i)]^\lambda I\{\gamma_{t^*}(Y_i) > Q\} \to E\{[\gamma_{t^*}(Y)]^\lambda I\{\gamma_{t^*}(Y) > Q\}\} \quad \text{a.s.} \tag{4.11}$$

Thus, since the right-hand side of (4.11) a.s. dominates $\limsup_n W_n$ and $\to 0$ as $Q \to \infty$, we have

$$W_n \to 0 \quad \text{a.s., } n \to \infty. \tag{4.12}$$

Also, we see via (4.10) that

$$\theta_n \le EW_n \to 0, \qquad n \to \infty. \tag{4.13}$$

By (4.9), (4.12) and (4.13), it suffices for (2.8a) to show

$$V_n^* = O(a_n) \quad \text{a.s., } n \to \infty. \tag{4.14}$$

We shall establish this and (2.8b) as well, by developing a suitable upper bound for $P\{V_n^* > B_0 a_n\}$, for appropriate choice of B_0. We write

$$P\{V_n^* \ge B_0 a_n\} \le \sum_{t\in I_n} \sum_{v\in \tilde{J}_n} P\{V_{tn}^*(v, M_1 c_n) \ge B_0 a_n\}, \tag{4.15}$$

with B_0 to be specified later, and we estimate the terms of this summation by an adaptation of the proof of Lemma 2.2 of Serfling (1982).

By (4.1) it is seen that

$$a_n/c_n \to 0, \qquad n \to \infty. \tag{4.16}$$

Now define

$$w_n = \left[\frac{2Q_n c_n}{a_n} + 1\right],$$

with $[\cdot]$ denoting greatest integer part. Fix v and put

$$\eta_{n,r} = v + \frac{rM_1 c_n}{w_n}, \quad \text{for } r = -w_n, -w_n + 1, \dots, w_n.$$

Note that $\eta_{n,r+1} - \eta_{n,r} = M_1 c_n / w_n$. Defining

$$\xi_{tnr} = |H_{tn}(\eta_{n,r}) - H_{tn}(v) - [EH_{tn}(\eta_{n,r}) - EH_{tn}(v)]|,$$

we have, by monotonicity of $H_{tn}(v)$ and $EH_{tn}(v)$ as functions of v, that

$$V_{tn}^*(v, M_1 c_n) \le \max_{-w_n \le r \le w_n} \xi_{tnr} + \max_{-w_n \le r \le w_{n-1}} |EH_{tn}(\eta_{n,r+1}) - EH_{tn}(\eta_{n,r})|.$$

Now

$$\begin{aligned} E[H_{tn}(\eta_{n,r+1}) - H_{tn}(\eta_{n,r})] &\le Q_n P\{\eta_{n,r} < F_0(X) \le \eta_{n,r+1}\} \\ &= Q_n(\eta_{n,r+1} - \eta_{n,r}) \\ &\le M_1 Q_n c_n / w_n \\ &\le M_1 a_n / 2 \le M_1 a_n, \end{aligned}$$

so that

$$\begin{aligned} P\{V_{tn}^*(v, M_1 c_n) \ge B_0 a_n\} &\le P\Big\{\max_{-w_n \le r \le w_n} \xi_{tnr} \ge (B_0 - M_1) a_n\Big\} \\ &\le \sum_{r=-w_n}^{w_n} P\{\xi_{tnr} \ge (B_0 - M_1) a_n\}. \end{aligned}$$

By Bernstein's inequality [Uspensky (1937)],

$$P\{\xi_{tnr} \ge (B_0 - M_1) a_n\} \le 2\exp(-\delta_{n,r}),$$

where

$$\delta_{n,r} = \frac{(B_0 - M_1)^2 n^2 a_n^2}{2n\sigma_{tnr}^2 + \frac{2}{3}(B_0 - M_1) Q_n n a_n}$$

and $\sigma_{tnr}^2 = \mathrm{Var}\{Z_{tnr}\}$, with

$$Z_{tnr} = \gamma_t(Y) I\{\gamma_t(Y) \le Q_n\} I\{v < F_0(X) \le \eta_{n,r}\}.$$

Applying (A.1), we obtain

$$\begin{aligned} \sigma_{tnr}^2 &\le EZ_{tnr}^2 \\ &\le \int\int \gamma_t^2(y) I\{\gamma_t(y) \le Q_n\} I\{v < F_0(x) \le v + M_1 c_n\} f(x, y)\, dx\, dy \qquad (4.17) \\ &\le M_0 M_1 c_n. \end{aligned}$$

By (4.8) and restriction (iii) on $\{c_n\}$ in the hypothesis of the lemma, we obtain

$$Q_n a_n = M_\lambda a_n^{(\lambda-2)/(\lambda-1)} = M_\lambda \left(\frac{c_n \log n}{n}\right)^{(\lambda-2)/2(\lambda-1)} \le M_\lambda c_n. \qquad (4.18)$$

By (4.1), (4.17) and (4.18), we thus have

$$\delta_{n,r} \ge B_0^* \log n, \qquad (4.19)$$

with

$$B_0^* = \frac{(B_0 - M_1)^2}{2M_0M_1 + \frac{2}{3}(B_0 - M_1)M_\lambda}. \tag{4.20}$$

Since (4.19) holds uniformly in r, $r = -w_n, -w_n + 1, \dots, w_n$, we have

$$P\{V_{tn}^*(v, M_1c_n) \geq B_0 a_n\} \leq 6w_n n^{-B_0^*}. \tag{4.21}$$

And since (4.21) holds uniformly in $t \in I_n$ and $v \in \tilde{J}_n$, (4.15) yields

$$P\{V_n^* \geq B_0 a_n\} \leq 6(N_n + 2)\tilde{N}_n w_n n^{-B_0^*}, \tag{4.22}$$

where $\tilde{N}_n$ denotes the cardinality of the set $\tilde{J}_n$. By (4.1) and (4.2) we find

$$N_n + 2 \leq 2(M_\lambda + 1)\left(\frac{n}{c_n \log n}\right)^{1/2}. \tag{4.23a}$$

Also, using the restriction $c_n \leq 1$,

$$\tilde{N}_n \leq [2/M_1c_n] + 1 \leq (2/M_1 + 1)c_n^{-1}. \tag{4.23b}$$

By (4.1) and (4.8), we have

$$w_n \leq \frac{2Q_nc_n}{a_n} + 1 = 2M_\lambda c_n\left(\frac{n}{c_n \log n}\right)^{\lambda/2(\lambda-1)} + 1.$$

Using the restriction (iii) on $\{c_n\}$, we easily obtain

$$c_n\left(\frac{n}{c_n \log n}\right)^{\lambda/2(\lambda-1)} \geq c_n^{-2/(\lambda-2)} \geq 1.$$

Thus

$$w_n \leq (2M_\lambda + 1)c_n\left(\frac{n}{c_n \log n}\right)^{\lambda/2(\lambda-1)}. \tag{4.23c}$$

Putting

$$L_\lambda = 12(M_\lambda + 1)(2/M_1 + 1)(2M_\lambda + 1) \tag{4.24}$$

and combining (4.22) and (4.23), we obtain

$$P\{V_n^* \geq B_0 a_n\} \leq L_\lambda\left(\frac{n}{c_n \log n}\right)^{(2\lambda-1)/2(\lambda-1)} n^{-B_0^*}. \tag{4.25}$$

Again using the restriction (iii) on $\{c_n\}$, we find $c_n^{-1} \leq (n/\log n)^{\lambda/(\lambda-2)}$, whence (4.25) yields

$$P\{V_n^* \geq B_0 a_n\} \leq L_\lambda\left(\frac{n}{\log n}\right)^{(2\lambda-1)/(\lambda-2)} n^{-B_0^*}. \tag{4.26}$$

Now, for given λ and for given real $\kappa > 0$, the constant B_0^* can be made to satisfy

$$B_0^* \geq \kappa + \frac{2\lambda - 1}{\lambda - 2}$$

by taking B_0 sufficiently large in (4.20). Let $B_{\kappa,\lambda}$ denote such a determination of B_0. Then (4.26) yields [using $(2\lambda - 1)/(\lambda - 2) \geq 2$ for $\lambda > 2$],

$$P\{V_n^* \geq B_{\kappa,\lambda}a_n\} \leq L_\lambda(\log n)^{-2}n^{-\kappa}. \tag{4.27}$$

In particular, taking $\kappa = 2$ in (4.27) and applying the Borel–Cantelli lemma, we obtain (4.14), thus establishing (2.8a).

To obtain (2.8b), we take $\lambda = \infty$ and apply (4.9) with $W_n = 0$ and $\theta_n = 0$ to write, for any $B \geq 2$,

$$P\{V_n \geq Ba_n\} \leq P\{V_n^* \geq \tfrac{1}{3}(B-2)a_n\}. \tag{4.28}$$

For each real $\kappa > 0$, define $B_\kappa = 3B_{\kappa,\infty} + 2$. Then (4.27) and (4.28) yield

$$P\{V_n \geq B_\kappa a_n\} \leq L_\infty(\log n)^{-2}n^{-\kappa}, \tag{4.29}$$

which, recalling the definition (4.1) of a_n, yields (2.8b). □

Acknowledgment. The authors thank one of the referees for the suggestions leading to Theorem 2.3(ii) and all referees for careful reading of the manuscript.

REFERENCES

CARROLL, R. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10** 1224–1233.

COLLOMB, G. (1980). Estimation non paramétrique de probabilités conditionnelles. *C. R. Acad. Sci. Paris Sér. A–B* **291** 427–430.

COLLOMB, G. (1981). Estimation non-paramétrique de la régression: Reveu bibliographique. *Internat. Statist. Rev.* **49** 75–93.

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

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

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

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

SERFLING, R. J. (1982). Properties and applications of metrics on nonparametric density estimators. In *Nonparametric Statistical Inference* (B. V. Gnedenko, M. L. Puri and I. Vincze, eds.) 859–873. North-Holland, Amsterdam.

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

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

STUTE, W. (1986). On almost sure convergence of conditional empirical distribution functions. *Ann. Statist.* **14** 891–901.

TSYBAKOV, E. A. (1983). Robust estimates of a function. *Problems Inform. Transmission* **18** 39–52.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

INSTITUT FÜR GESELLSCHAFTS- UND
WIRTSCHAFTWISSENSCHAFTEN
UNIVERSITÄT BONN
ADENAUERALLEE 24-26
D-5300 BONN 1
WEST GERMANY

LIMBURGS UNIVERSITAIR CENTRUM
UNIVERSITAIRE CAMPUS
B-3610 DIEPENBEEK
BELGIUM

DEPARTMENT OF MATHEMATICAL SCIENCES
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218