The Annals of Statistics
1988, Vol. 16, No. 2, 741–750

MONOTONE NONPARAMETRIC REGRESSION

BY HARI MUKERJEE

University of California, Davis

In monotone regression procedures one utilizes only the monotonicity of the regression function. In nonparametric regression one utilizes only the assumed smoothness. The analytic and asymptotic properties of the estimator are superior in the latter case; however, monotonicity is not guaranteed. We study a hybrid procedure that produces monotone estimators with properties similar to those of nonparametric regression estimators.

1. Introduction. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. bivariate random variables with common joint density f. Let g be the marginal density of X_1, $h(x) = \int y f(x, y)\,dy$, and let $m(x) = E(Y_1 | X_1 = x] = h(x)/g(x)$ be the regression function of Y_1 on X_1. Our problem is to estimate m.

In many applications the regression function is believed to be nondecreasing (the nonincreasing case is similar) and it is natural to try to incorporate this prior information into the estimation procedure. Brunk (1958) proposed the following least-squares estimator:

Suppose $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ is the order statistic corresponding to $\{X_i\}$ with $Y_{(j)}$ being the "observation" at the "observation point" $X_{(j)}$. Then Brunk's estimator $m_n^*(X_{(i)})$ is given by

$$m_n^*(X_{(i)}) = \max_{s \le i} \min_{t \ge i} \sum_{j=s}^{t} Y_{(j)}/(t - s + 1). \tag{1.1}$$

Let $S_0 = 0$, $S_k = \sum_{j=1}^{k} Y_{(j)}$, $1 \le k \le n$ and let $\{(k, S_k): 0 \le k \le n\}$ be the cumulative sum diagram (CSD) of $\{Y_{(j)}\}$. Let G be the greatest convex minorant (GCM) of the CSD on $[0, n]$, i.e., G is the supremum of all convex functions on $[0, n]$ whose graphs lie below the CSD. Then an equivalent definition of the estimator is given by

$$m_n^*(X_{(i)}) = \text{the lhs slope of } G \text{ at } i, \qquad 1 \le i \le n. \tag{1.2}$$

Simple algorithms for computations are discussed in detail in Barlow, Bartholomew, Bremner and Brunk (1972). The fitted function m_n^*, defined so far only on $\{X_i\}$, is called the isotonic regression function, and it corresponds to some random local smoothing or averaging over the so-called level sets—subintervals of the order statistics of $\{X_i\}$ over which m_n^* has the same value. This function could be extended to any nondecreasing function m_n^* (we use the same notation for m_n^* and its restriction to $\{X_i\}$) agreeing with (1.1) on $\{X_i\}$.

The function m_n^* has been studied extensively. Barlow, Bartholomew, Bremner and Brunk (1972) contains a comprehensive account of the subject. Brunk (1955)

Received June 1986; revised June 1987.

AMS 1980 *subject classifications*. Primary 62G05; secondary 60F05, 60F15.

Key words and phrases. Isotonic regression, nonparametric regression, uniform strong consistency, asymptotic normality.

has shown that m_n^*, restricted to $\{X_i\}$, is the maximum likelihood estimator when $Y_1 - m(X_1)$ has a $N(0, \sigma^2)$ distribution. Brunk (1958) and Hanson, Pledger and Wright (1973) establish strong uniform consistency of m_n^* on closed, bounded intervals. Brunk (1970) has shown that $n^{1/3}[m_n^*(x) - m(x)]$ has an asymptotic distribution with a density related to the solution of a heat equation when $m'(x) > 0$. Groeneboom (1985) has found the solution in a closed form. Wright (1982) has extended this result to the cases where $|m(x + h) - m(x)| = A|h|^{\alpha}(1 + o(1))$ as $h \to 0$ for positive A and α.

In spite of the availability of these statistical properties of isotonic regression, it has not proven to be popular in applications. This is mainly because it has several drawbacks from the user's viewpoint:

(i) The extension of m_n^*, although nondecreasing, is arbitrary, and is usually not very smooth.

(ii) Frequently, there are too many "flat spots" in isotonic regression, corresponding to the level sets where m_n^* is constant.

(iii) The asymptotic distribution of m_n^* has a norming only by $n^{1/3}$ at points of positive slope.

There are other nonparametric regression procedures, which are competitors of isotonic regression procedures and provide smooth estimators with faster convergence rates and asymptotic distributions that are normal. Possibly the three most popular ones are those based on kernels, k-nearest neighbors (k-NN) and smoothing splines. The articles by Stone (1977), Wegman and Wright (1983) and Silverman (1985) contain substantial references. Although these estimators are not necessarily nondecreasing [the spline method is an exception in that the existence of monotone smoothing splines can be demonstrated [Wright and Wegman (1980)], but no computing algorithms exist], their usage has gained much popularity, even when m is believed to be nondecreasing, because of their advantages over the isotonic regression mentioned previously. In this paper we initiate the study of a "marriage"—first "isotonize" the raw data on $\{X_i\}$, then smooth the resulting isotonic regression function using an appropriate kernel. Judicious choice of kernels yields smooth nondecreasing estimators with asymptotic properties similar to those in nonparametric regression.

The idea of improving on convergence rates of isotonic estimators by averaging over larger neighborhoods than the typical (random) ones used by the isotonic regression procedure is not new. Barlow and van Zwet (1971) and Wright (1982) first grouped the data and then isotonized the resulting estimator to get improved convergence rates. Friedman and Tibshirani (1984) first smoothed the data by the k-NN method, restricted the estimator to the "observation points" $\{X_i\}$ and then isotonized the resulting function. Since the smoothing in the second stage requires averaging typically over smaller intervals than in the first, it is not clear how to attack the problem of asymptotics in this case. It may be noted that in the procedures described in this paragraph the group-averaging or the k-NN smoothing is performed first and isotonization is done last. Thus the objections (i) and (ii) to the isotonic regression procedures mentioned previously still remain valid against these procedures.

2. The estimator. Let m_n^* be defined on $\{X_i\}$ by (1.1) and let its extension to the reals be given by

$$
\begin{aligned}
& m_n^*(x) = m_n^*(X_{(i)}), \quad \text{for } X_{(i)} \le x < X_{(i+1)}, \qquad 1 \le i \le n-1, \\
(2.1) \qquad & m_n^*(x) = m_n^*(X_{(1)}), \quad \text{for } x < X_{(1)}, \\
& m_n^*(x) = m_n^*(X_{(n)}), \quad \text{for } x \ge X_{(n)}.
\end{aligned}
$$

Let the kernel k be a log-concave density and let the bandwidth sequence $\{b_n\}$ be a positive sequence converging to 0. Let

$$
\begin{aligned}
& h_n(x) = (nb_n)^{-1} \sum_{i=1}^{n} k[(x - X_i)/b_n] m_n^*(X_i), \\
(2.2) \qquad & g_n(x) = (nb_n)^{-1} \sum_{i=1}^{n} k[(x - X_i)/b_n].
\end{aligned}
$$

Since k is log-concave its support is an interval. If the support is the entire real line, then $g_n(x) > 0$ for all x; if not, $g_n(x)$ may be positive only on a finite number of disjoint intervals, not including or including their endpoints depending on whether the kernel has a positive value or is equal to 0 at the corresponding endpoints of its support. We define our estimator of $m(x)$ by:

(2.3) $m_n(x) = h_n(x)/g_n(x)$ on $\{x\colon g_n(x) > 0\}$ and, if necessary, it is extended (i) first by closure, (ii) then by linear interpolation in "gaps" inside $(X_{(1)}, X_{(n)})$ and (iii) as a continuous function with constant values in each of the semiinfinite outer intervals.

When k is differentiable we estimate $m'(x)$ by the left-hand derivative of m_n at x, which exists everywhere.

REMARK 2.1. We have introduced the assumption of log-concavity of the kernel in the definition of our estimator to guarantee isotonicity of m_n. These kernels include the uniform, the normal, the Laplacian and all concave densities as well as the C_∞-density with compact support given by

$$k(x) = C \exp[1/(x^2 - 1)] I(|x| < 1).$$

From the definition of m_n, to establish its monotonicity, it is sufficient to do so on $\{x\colon g_n(x) > 0\}$, and this monotonicity, as pointed out by a referee, is a consequence of the monotone likelihood ratio property of $k_n(y|\theta) = b_n^{-1}k[(y - \theta)/b_n]$, which is log-concave. Let θ have the prior distribution given by $P(\theta = X_i) = 1/n$, $1 \le i \le n$. Then $m_n(y) = E[m_n^*(\theta)|Y = y]$ (k_n is the conditional density of Y) is nondecreasing because m_n^* is nondecreasing and the posterior distribution of θ given $Y = y$ has the monotone likelihood ratio property as well [Lemma 2, page 74, in Lehmann (1959)]. The smoothness of m_n on an interval is a direct consequence of the smoothness of k and positivity of g_n on the interval.

3. Consistency of the estimator and its derivative. To prove consistency of m_n, we will use some known properties of m_n^*. Let G be the marginal *df* of X_1 and let S be the support of G. For $B \subset S$ let

$$L_B(t) = \sup_{x \in B} P\{|Y_1 - m(x)| \geq t | X_1 = x\}, \qquad t \geq 0. \tag{3.1}$$

Consider the assumption

$$L_B(t) \to 0, \quad \text{as } t \to \infty \quad \text{and} \quad \int_0^\infty t^s |dL_B(t)| < \infty, \quad \text{for some } s > 0. \tag{3.2}$$

For any interval I and $\varepsilon > 0$ let I_ε denote the open ε-neighborhood of I, and let $N_n(I) = \#\{X_i\colon X_i \in I, 1 \leq i \leq n\}$. The following result is due to Brunk (1958) and Hanson, Pledger and Wright (1973).

THEOREM 3.1. *Let I be any closed, bounded interval. Assume that for some* $\varepsilon > 0$,

$$P\Big\{\liminf_n N_n(J)/n > 0\Big\} = 1, \quad \textit{for all intervals } J \subset I_\varepsilon, \tag{3.3}$$

condition (3.2) *holds with $B = I_\varepsilon$ and $s = 1$, and m is continuous on I_ε. Then*

$$\sup_{x \in I} |m_n^*(x) - m(x)| \to 0 \quad a.s. \tag{3.4}$$

Note that I is bounded away from the endpoints of I_ε. This is necessary because the bias in m_n^* at the extreme order statistics may be very large. However, the following result due to Makowski (1973) and Wright (1981) controls the growth rate of $|m_n^* - m|$.

THEOREM 3.2. *If* (3.2) *holds with $s = 2$, then*

$$\max_{1 \leq j \leq n} \left| \sum_{i=1}^{j} \left[m_n^*(X_{(i)}) - m(X_{(i)})\right] \right| = O\big([n \log\log n]^{1/2}\big) \quad a.s. \tag{3.5}$$

In kernel procedures it is necessary to make certain assumptions about the kernel, many of which are automatically satisfied by our log-concave density. Since $k(x) = \exp[C(x)]$ for some function C concave on its support, $k(x)$ is initially nondecreasing and then nonincreasing as x goes from $-\infty$ to ∞. Since k is also a density, it is bounded. Moreover, if k is positive on the entire right tail, then $C(x) \downarrow -\infty$ as $x \uparrow \infty$ and $C'(x)$ (with some arbitrary but fixed determination on the interior of the support of C), which is nonincreasing, must be less than or equal to $-M$ for some $M > 0$ for all x sufficiently large, showing k decays at least as fast as a negative exponential on the right tail. A similar result holds if k is positive on the entire left tail.

We now state and prove our main results.

THEOREM 3.3. *Let I be any closed, bounded interval with $I_\varepsilon \subset S$ for some $\varepsilon > 0$. Assume that m is continuous on I_ε and that* (3.2) *holds with $B = I_\varepsilon$ and $s = 1$. If k has compact support and $b_n \to 0$, then*

$$\sup_{x \in I} |m_n(x) - m(x)| \to 0 \quad a.s. \tag{3.6}$$

If k has infinite support, then (3.6) *holds under the additional assumptions that $s = 2$ in* (3.2), $|m| \le M < \infty$ *and $nb_n^r \to 0$ for some $r > 0$.*

PROOF. Devroye (1978) showed that if J is a compact subset of S, then $\inf_{x \in J} P\{X_1 \in [x - \delta, x + \delta]\} > 0$ for all $\delta > 0$. Fix $0 < \beta < \varepsilon$. By SLLN condition (3.3) holds with I_ε replaced by $\bar{I}_\beta$, the closure of I_β and thus Theorem 3.1 implies

$$\sup_{x \in \bar{I}_\beta} |m_n^*(x) - m(x)| \to 0 \quad \text{a.s.} \tag{3.7}$$

By monotonicity of m_n and m and continuity of m, it is sufficient to prove pointwise consistency. Fix $x \in I$. If $g_{n(k)}(x) = 0$ for some subsequence $\{n(k)\}$, then, from the definition of m_n, it is easy to see that $m_{n(k)}(x) \to m(x)$ on $\{m_{n(k)}^*(x) \to m(x)\}$. Thus we assume $m_n(x) = h_n(x)/g_n(x)$ for all n without loss of generality.

Let $0 < \eta < \beta$ be arbitrary. If k has compact support and $b_n \to 0$, then $k[(x - X_i)/b_n] = 0$ if $|x - X_i| > \eta$ and n is large enough. Thus (3.6) follows from (3.7) and the continuity of m. Now suppose k has infinite support and the assumptions of the theorem hold. Using the first result, it is sufficient to show that

$$\begin{gathered} (nb_n)^{-1} \sum_{|x - X_i| > \eta} k[(x - X_i)/b_n] |m_n^*(X_i) - m(x)| \to 0 \quad \text{a.s.}, \\ (nb_n)^{-1} \sum_{|x - X_i| > \eta} k[(x - X_i)/b_n] |m(x)| \to 0 \quad \text{a.s.} \end{gathered} \tag{3.8}$$

Using Theorem 3.2 and the bound of m, we have

$$\begin{aligned} |m_n^*(X_i) - m(x)| &\le |m_n^*(X_i) - m(X_i)| + |m(X_i)| + |m(x)| \\ &\le C(n \log\log n)^{1/2} + 2M, \quad \text{for some } C > 0. \end{aligned} \tag{3.9}$$

Let $Q(t) = \sup\{k(u): |u| \ge t\}$, $t \ge 0$. Then $t^p Q(t) \to 0$ as $t \to \infty$ for all real p since k is log-concave. Thus, for any $p \ge 0$,

$$\begin{aligned} &(nb_n)^{-1} \sum_{|x - X_i| > \eta} k[(x - X_i)/b_n] n^p \\ &\le (nb_n)^{-1} Q(\eta/b_n) n^{p+1} \\ &= \eta^{-rp-1} (nb_n^r)^p (\eta/b_n)^{rp+1} Q(\eta/b_n) \to 0, \quad \text{since } nb_n^r \to 0. \end{aligned} \tag{3.10}$$

The results in (3.8) now follow from (3.9) and (3.10), which completes the proof of the theorem. □

In the remainder of this section we assume m, g and k are differentiable. Let

$$g_n^*(x) = (nb_n)^{-1} \sum_{i=1}^n [(X_i - x)/b_n] k'[(x - X_i)/b_n].$$

Note that g_n^* is the usual kernel density estimator using a kernel whose value at t is $-tk'(t)$ and which integrates to 1. In proving consistency of m_n', we will use the convergences of g_n, g_n^* and g_n' to g, g and g', respectively. There is a vast literature on the consistency of kernel density estimators using a variety of techniques and noncomparable assumptions. We follow Silverman (1978), who uses the least restrictive bandwidths.

For any real-valued function δ on the reals, consider the assumptions

(3.11)
(a) δ is uniformly continuous (with modulus of continuity w_δ) and of bounded variation $V(\delta)$;
(b) $\int_0^1 [\log(1/u)]^{1/2}\, d\gamma(u) < \infty$, where $\gamma(u) = [w_\delta(u)]^{1/2}$;
(c) $\int |x \log|x|\,|^{1/2}\, d\delta(x) < \infty$;
(d) $\int |\delta(x)|\, dx < \infty$ and $\delta(x) \to 0$ as $|x| \to \infty$; and
(e) $\int \delta(x)\, dx = 1$.

Silverman (1978) showed that $g_n \to g$ a.s. uniformly for any kernel k obeying (3.11) if g is uniformly continuous, $b_n \to 0$ and $\log(1/b_n)/nb_n \to 0$ and that $g_n' \to g'$ a.s. uniformly if k' obeys (3.11)(a)–(d), k obeys (3.11)(e), g' is uniformly continuous, $b_n \to 0$ and $\log(1/b_n)/nb_n^3 \to 0$. For our log-concave density kernel k, (3.11)(d) is automatically satisfied if $\delta(t) = k(t)$, $-tk'(t)$ or $k'(t)$ and (3.11)(c) and (e) are satisfied if $\delta = k$. Thus we have

THEOREM 3.4. *Assume* (3.11)(a) *and* (b) *hold with* $\delta(t) = k(t)$, (3.11)(a)–(c) *hold with* $\delta(t) = k'(t)$ *and* $\delta(t) = -tk'(t)$, g' *is uniformly continuous,* $b_n \to 0$ *and* $\log(1/b_n)/nb_n^3 \to 0$. *Then* g_n, g_n^* *and* g_n' *converge a.s. uniformly to* g, g, *and* g', *respectively.*

We now prove consistency of m_n'. Even for log-concave kernels their derivatives may have arbitrarily large "spikes" in the tails. Assumption (3.13) controls their growth rate.

THEOREM 3.5. *Let* I *be any closed, bounded interval with* $I_\varepsilon \subset S$ *for some* $\varepsilon > 0$. *Assume that* g *is bounded away from* 0 *on* I_ε, m' *is continuous on* I_ε, *the conditions of Theorem* 3.4 *hold and that* (3.2) *holds with* $B = I_\varepsilon$ *and* $s = 1$. *If* k *has compact support, then*

$$\sup_{x \in I} |m_n'(x) - m'(x)| \to 0 \quad a.s. \tag{3.12}$$

If k *has infinite support, then* (3.12) *holds under the additional assumptions that* $s = 2$ *in* (3.2), m *and* m' *are bounded and*

$$nb_n^{2r} \log\log n \to 0 \quad and \quad t^{r+2} \sup\{|k'(u)|: |u| \geq t\} \to 0, \quad as\ |t| \to \infty, \tag{3.13}$$

for some $r > 0$.

PROOF. Let E be the set of sample sequences for which the conclusions of Theorems 3.3 and 3.4 hold. Then $P(E) = 1$ from the conditions of this theorem for both compact and noncompact kernels. Henceforth we assume all sample sequences to be in E. From our assumption on g,

$$\liminf_n \inf_{x \in I_\varepsilon} g_n(x) > 0.$$

Thus, for all n large enough, $m_n(x) = h_n(x)/g_n(x)$ for all $x \in I_\varepsilon$, and we assume n is at least this large. Since $m_n'(x) = h_n'(x)/g_n(x) - h_n(x)g_n'(x)/[g_n(x)]^2$, it is sufficient to show that $h_n' \to mg' + m'g$ a.s. uniformly on I.

If k has compact support, then $k'[(x - X_i)/b_n] = 0$ if $|x - X_i| > \eta$ and n is large enough for every $\eta > 0$. Write

$$\begin{aligned} m_n^*(X_i) &= [m_n^*(X_i) - m(X_i)] + [m(X_i) - m(x)] + m(x) \\ &= [m_n^*(X_i) - m(X_i)] + [m'(x)(X_i - x) + o(1)] + m(x), \end{aligned}$$

where $o(1) \to 0$ as $\eta \to 0$ uniformly in $x \in I_\beta$ for some $0 < \beta < \varepsilon$. From (3.7), Theorem 3.4 and the aforementioned, we have

$$(nb_n^2)^{-1} \sum_{i=1}^n k'[(x - X_i)/b_n][m_n^*(X_i) - m(X_i)] \to 0,$$

$$(nb_n^2)^{-1} \sum_{i=1}^n k'[(x - X_i)/b_n][m(X_i) - m(x)] \to m'(x)g(x)$$

and

$$(nb_n^2)^{-1} \sum_{i=1}^n k'[(x - X_i)/b_n]m(x) \to m(x)g'(x),$$

almost surely and uniformly in $x \in I$.

If k has infinite support, then a.s. uniform convergence of

$$(nb_n^2)^{-1} \sum_{|x - X_i| > \eta} k'[(x - X_i)/b_n]\{|m_n^*(X_i) - m(X_i)| + |m(X_i)|\}$$

to 0 for every $\eta > 0$ follows from the same method used to prove (3.8) using assumption (3.13). Thus the consistency result follows from the first half of the theorem. □

4. Asymptotic normality. We will prove asymptotic normality of $(nb_n)^{1/2}[m_n(x_0) - m(x_0)]$ at a point x_0 using the rectangular kernel $k_0(x) = I(|x| \le 1/2)$. Assume $v(x) = \mathrm{Var}[Y_1|X_1 = x] < \infty$. Schuster's (1972) theorem on asymptotic normality of kernel regression estimators implies

THEOREM 4.1. *If $g(x_0) > 0$, g', m', v', g'' and m'' exist and are bounded in a neighborhood of x_0, $E|Y_1|^3 < \infty$, $nb_n^3 \to \infty$ and $nb_n^5 \to 0$, then*

$$(4.1)\quad (nb_n)^{1/2}\left[\sum_{i=1}^n I(|x_0 - X_i| \le b_n/2)Y_i / \sum_{i=1}^n I(|x_0 - X_i| \le b_n/2) - m(x_0)\right]$$

converges in distribution to $N(0, v(x_0)/g(x_0))$.

THEOREM 4.2. *Assume that the conditions of Theorem 4.1 hold and that $|m(x_0) - m(x)| = A|x_0 - x|^\alpha(1 + o(1))$ for some $A > 0$ and $1 \le \alpha < 2$, where $o(1) \to 0$ as $|x_0 - x| \to 0$. If $n(b_n/\log\log n)^{2\alpha+1} \to \infty$, then*

$$(4.2)\qquad (nb_n)^{1/2}[m_n(x_0) - m(x_0)] \to_d N(0, v(x_0)/g(x_0)).$$

PROOF. From the conditions on g' and g'', Nadaraya's (1965) result implies that $g_n = (nb_n)^{-1}\sum_{i=1}^n I(|x_0 - X_i| \le b_n/2) \to g$ a.s. uniformly in a neighborhood of x_0 if $nb_n^2/\log n \to \infty$ and $b_n \to 0$. Since $g(x_0) > 0$ we have $m_n(x_0) = h_n(x_0)/g_n(x_0)$ for almost all sample sequences if n is large enough. Thus (4.2) is the same as (4.1) with $m_n^*(X_i)$ replacing Y_i for large n with probability 1. Since $g_n(x_0) \to g(x_0)$ a.s., an equivalent form of (4.1) is

$$(nb_n)^{-1/2} \sum_{i=1}^n I(|x_0 - X_i| \le b_n/2)Y_i - (nb_n)^{1/2}m(x_0)g(x_0) \to_d N(0, v(x_0)g(x_0))$$

and it is sufficient to show that

$$(nb_n)^{-1/2} \sum_{i=1}^n I(|x_0 - X_i| \le b_n/2)[Y_i - m_n^*(X_i)] \to_p 0. \tag{4.3}$$

For each $c > 0$ and $n \ge 1$ let $c_n = cn^{-1/(2\alpha+1)}$. For $s \le t$ let $A(s, t) = \sum_{j=1}^n I(s \le X_j \le t)Y_j/\sum_{j=1}^n I(s \le X_j \le t)$,

$$\overline{m}_{nc}(X_i) = \max_{X_i - c_n \le s \le X_i} \min_{X_i \le t \le X_i + c_n} A(s, t),$$

and extend it the same way as in (2.1) using $\overline{m}_{nc}(X_i)$ instead of $m_n^*(X_i)$. Using assumptions considerably weaker than ours, it can be shown, by a slight modification (in the extension of $\overline{m}_{nc}$) of Wright's (1981) lemma, that for all x in a neighborhood of x_0,

$$\lim_{c\to\infty} \limsup_n P[m_n^*(x) \ne \overline{m}_{nc}(x)] = 0. \tag{4.4}$$

Let $L_n = x_0 - b_n/2 - c_n$ and $U_n = x_0 + b_n/2 + c_n$. For $X_i \in [L_n, U_n]$ define

$$m_{nc}^*(X_i) = \max_{L_n \le s \le X_i} \min_{X_i \le t \le U_n} A(s, t)$$

and extend it to $[L_n, U_n]$ the same way as in (2.1) with $m_{nc}^*(X_i)$ replacing $m_n^*(X_i)$. From (4.4) it is clear that

$$P[m_{nc}^*(x_0 - b_n/2) \ne m_n^*(x_0 - b_n/2) \text{ or } m_{nc}^*(x_0 + b_n/2) \ne m_n^*(x_0 + b_n/2)]$$

can be made arbitrarily small by choosing n and c large enough. The same can be said about

$$P\left\{ \sum_{i=1}^n I(|x_0 - X_i| \le b_n/2)|m_n^*(X_i) - m_{nc}^*(X_i)| \ne 0\right\},$$

from definition (1.1) of isotonic regression. We also have $\sum_{i=1}^n I(|x_0 - X_i| \le b_n/2 + c_n)[Y_i - m_{nc}^*(X_i)] = 0$ [Barlow, Bartholomew, Bremner and Brunk (1972), page 27]. Thus to prove (4.3) it is sufficient to show that

$$U_n = (nb_n)^{-1/2} \sum_{i=1}^n I(b_n/2 < |x_0 - X_i| \le b_n/2 + c_n)Y_i \to_p 0 \tag{4.5}$$

and

$$V_n = (nb_n)^{-1/2} \sum_{i=1}^n I(b_n/2 < |x_0 - X_i| \le b_n/2 + c_n)[Y_i - m_{nc}^*(X_i)] \to_p 0. \tag{4.6}$$

$E(U_n^2) = (nb_n)^{-1} 2nc_n g(x_0)[m^2(x_0) + v(x_0)](1 + o(1)) \to 0$ since $(b_n/c_n)^{2\alpha+1} = n(b_n/c)^{2\alpha+1} \to \infty$ by assumption. This proves (4.5). Theorem 3.2 implies that V_n is almost surely bounded by $(nb_n)^{-1/2}\{D[nc_n g(x_0)]\log\log[nc_n g(x_0)]\}^{1/2}$ for some $D > 0$ and this bound converges to 0 since $(b_n/c_n \log\log n)^{2\alpha+1} = n(b_n/c \log\log n)^{2\alpha+1} \to \infty$ by assumption. This proves (4.6) and completes the proof of the theorem. □

REMARK 4.3. The key to the proof of Theorem 4.2 is the equality of the unweighted sums $\sum_{i=1}^n I(|x_0 - X_i| \le b_n/2 + c_n) m_{nc}^*(X_i)$ and $\sum_{i=1}^n I(|x_0 - X_i| \le b_n/2 + c_n) Y_i$, which follows from the properties of isotonic regression. It is reasonable to hope that if an arbitrary log-concave kernel k is approximated by a step function k_s and the preceding result is applied to each step, the sum of the differences

$$\{k[(x_0 - X_i)/b_n] - k_s[(x_0 - X_i)/b_n]\}[Y_i - m_{nc}^*(X_i)],$$

over each step will be small. Unfortunately, this does not follow from the isotonicity of m_{nc}^* and m. It appears that the desirable smallness can only be proven to be true in probability and will require a more precise knowledge of the joint distribution of $\{m_{nc}^*(X_i)\}$ than is known at present.

Acknowledgments. The author wishes to acknowledge his gratitude to Neil Willits for pointing out the relevance of log-concave densities to our problem and to a referee for helpful suggestions resulting in substantial improvement in proofs and presentation.

REFERENCES

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression.* Wiley, New York.

BARLOW, R. E. and VAN ZWET, W. R. (1971). Comparison of several nonparametric estimators of the failure rate function. In *Operations Research and Reliability, Proc. of the NATO Conference at Turin, 1969.* 375–399. Gordon and Breach, New York.

BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26** 607–616.

BRUNK, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29** 437–454.

BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* 177–195. Cambridge Univ. Press, New York.

DEVROYE, L. P. (1978). The uniform convergence of nearest neighbor regression function estimation and their application in optimization. *IEEE Trans. Inform. Theory* **IT-24** 142–151.

FRIEDMAN, J. and TIBSHIRANI, R. (1984). The monotone smoothing of scatterplots. *Technometrics* **26** 242–250.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 539–555. Wadsworth, Monterey, Calif.

HANSON, D. L., PLEDGER, G. and WRIGHT, F. T. (1973). On consistency in monotonic regression. *Ann. Statist.* **1** 401–421.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

MAKOWSKI, G. (1973). Laws of the iterated logarithm for permuted random variables and regression applications. *Ann. Statist.* **1** 872–877.

NADARAYA, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theory Probab. Appl.* **10** 186–190.

SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184.

SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. Roy. Statist. Soc. Ser. B* **47** 1–52.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

WEGMAN, E. J. and WRIGHT, I. W. (1983). Splines in statistics. *J. Amer. Statist. Assoc.* **78** 351–365.

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448.

WRIGHT, F. T. (1982). Monotone regression estimates for grouped observations. *Ann. Statist.* **10** 278–286.

WRIGHT, I. W. and WEGMAN, E. J. (1980). Isotonic, convex and related splines. *Ann. Statist.* **8** 1023–1035.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616