The Annals of Statistics
1975, Vol. 3, No. 2, 350-362

CONSISTENCY IN GENERALIZED ISOTONIC REGRESSION[1]

BY TIM ROBERTSON AND F. T. WRIGHT

The University of Iowa

Suppose T is a partially ordered set and that associated with each t in T we have a distribution with "location parameter" $m(t)$. In this paper we discuss consistency properties of estimates of $m(\cdot)$ which are isotonic with respect to a partial order. The results extend results in the literature, some of which are contained in Brunk (1970) (Estimation of isotonic regression in *Nonparametric Techniques in Statistical Inference*, Cambridge University Press, 177–195), Cryer, *et al.* (1972) (Monotone median regression in *Ann. Math. Statist.* **43**), Hanson, *et al.* (1973) (On consistency in monotone regression in *Ann. Statist.* **1**), and Robertson and Wright (1973) (Multiple isotonic median regression in *Ann. Statist.* **1**).

1. Introduction. Recently, in the literature, several estimation problems have been considered which fall within the following framework. T is an index set with partial ordering $\ll$ and $m(\cdot)$ is a real-valued function on T which is isotonic with respect to $\ll$. $\{(t_n, X_n)\}$ is a sequence where $t_n \in T$; $n = 1, 2, \cdots$ and $\{X_n\}$ is a sequence of independent random variables. ($m(\cdot)$ is to be thought of as a regression function and a value of X_i as an observation at t_i. The distribution of X_i depends on t_i at least through $m(t_i)$.) We wish to estimate $m(\cdot)$.

In Section 2 we investigate consistency properties of certain estimators $\hat{m}_n(\cdot)$ of $m(\cdot)$ which satisfy our order restriction. The techniques used to prove these results are like some of those in Hanson, Pledger and Wright (1973), Cryer, Robertson, Wright and Casady (1972), and Robertson and Wright (1973). In Section 3 we discuss the application of these results to several examples and discuss the relationship of these results to those in the literature.

2. Consistency. The estimator of $m(\cdot)$ generally depends on (t_1, x_1), $(t_2, x_2), \cdots, (t_n, x_n)$ through another sequence of functions, $\{M_n\}$. We assume that M_n is real-valued and its domain is the Cartesian product of n copies of $T \times R$ where R is the set of real numbers. We assume that M_n has four properties which we enumerate with letters of the alphabet.

(a) If $t_1, t_2, \cdots, t_n$ are fixed elements of T then $M_n[(t_1 x_1), (t_2, x_2), \cdots, (t_n, x_n)]$, thought of as a function of $(x_1, x_2, \cdots, x_n)$, is a Borel measurable function from R^n to R.

(b) M_n is symmetric in the sense that if $p(\cdot)$ is any permutation of the integers from 1 through n then $M_n[(t_{p(1)}, x_{p(1)}), (t_{p(2)}, x_{p(2)}), \cdots, (t_{p(n)}, x_{p(n)})] \equiv M_n[(t_1, x_1), (t_2, x_2), \cdots, (t_n, x_n)]$.

Received September 1972; revised September 1973.

[1] This research was supported by the National Science Foundation under Grant GP-30966X.

AMS 1970 *subject classifications.* Primary 60F05, 60F15; Secondary 62G05.

Key words and phrases. Isotonic regression, consistency, exponential rates of convergence.

(c) With $t_1, t_2, \cdots, t_n$ fixed elements of T and c a constant we have $M_n[(t_1, x_1), (t_2, x_2), \cdots, (t_n, x_n)] - c = M_n[(t_1, x_1 - c), (t_2, x_2 - c), \cdots, (t_n, x_n - c)]$.

(d) If $t_1, t_2, \cdots, t_n$ are fixed elements of T and $x_i \leqq y_i$; $i = 1, 2, \cdots, n$ then $M_n[(t_1, x_1), (t_2, x_2), \cdots, (t_n, x_n)] \leqq M_n[(t_1, y_1), (t_2, y_2), \cdots, (t_n, y_n)]$.

In some of the examples discussed in Section 3, $M_n[(t_1, x_1), (t_2, x_2), \cdots, (t_n, x_n)]$ is a solution to a certain optimization problem. We do not mean for (a) to imply in these problems that the solution is unique but only that we have a procedure for selecting one of the solutions so that $M_n[(t_1, X_1), (t_2, X_2), \cdots, (t_n, X_n)]$ is measurable and thus $\hat{m}_n(t)$ is a random variable for each t.

If $\{j(1), j(2), \cdots, j(n)\}$ is the set of all indices for which a certain condition $c(j)$ is satisfied then we let $M\{X_j; c(j) \text{ holds}\}$ denote the random variable, $M_n[(t_{j(1)}, X_{j(1)}), (t_{j(2)}, X_{j(2)}), \cdots, (t_{j(n)}, X_{j(n)})]$. Condition (b) ensures that this symbol is well defined. Finally, we remark that (c) and (d) are satisfied by location statistics.

Let $\mathscr{L}$ be the complete σ-lattice of subsets of T induced by our partial ordering (i.e., $L \in \mathscr{L}$ if and only if $x \ll y$ and $x \in L$ imply that $y \in L$). We shall refer to members of $\mathscr{L}$ as upper layers and use the symbol L, exclusively, to denote upper layers. If $t_j \in \{t_1, t_2, \cdots, t_n\}$ we define our estimator $\hat{m}_n(\cdot)$ at t_j by:

$$\hat{m}_n(t_j) = \max\nolimits_{L \ni t_j} \min\nolimits_{L' \not\ni t_j} M\{X_i; i \leqq n, t_i \in L - L'\} . \tag{2.1}$$

It is easy to see that $t_i \ll t_j$ implies $\hat{m}_n(t_i) \leqq \hat{m}_n(t_j)$ and one can usually extend $\hat{m}_n(\cdot)$ to all of T in several ways so that $\hat{m}_n(\cdot)$ satisfies various "smoothness" conditions as well as being isotonic.

In most of the examples which have been considered in the literature, M_n has the following additional property:

(e) If $t_1, t_2, \cdots, t_{m+n} \in T$ then $M_{m+n}[(t_1, x_1), (t_2, x_2), \cdots, (t_{m+n}, x_{m+n})]$ is between (not necessarily strictly) $M_m[(t_1, x_1), (t_2, x_2), \cdots, (t_m, x_m)]$ and $M_n[(t_{m+1}, x_{m+1}), (t_{m+2}, x_{m+2}), \cdots, (t_{m+n}, x_{m+n})]$.

If we assume the averaging property (e) in addition to (a) through (d) then, using an argument similar to the one described in Robertson and Wright (1973), one can show that $\hat{m}_n(t_j) = \bar{m}_n(t_j)$ where

$$\bar{m}_n(t_j) = \min\nolimits_{L' \not\ni t_j} \max\nolimits_{L \ni t_j} M\{X_i; i \leqq n, t_i \in L - L'\} .$$

However, it is easy to see, by considering trimmed means (Example 4) that $\hat{m}_n(\cdot)$ and $\bar{m}_n(\cdot)$ are not necessarily equal if we do not assume the averaging property, (e). On the other hand, $\bar{m}_n(\cdot)$ is isotone and our consistency theorems also apply to $\bar{m}_n(\cdot)$.

Before presenting our consistency results we mention an example which has received considerable attention in the literature. Suppose the distribution at t is normal with unknown mean, $m(t)$, and known variance $\sigma^2(t)$ and $m(t)$ is known to be isotone with respect to $\ll$. If we choose

$$M_n[(t_1, x_1), (t_2, x_2), \cdots, (t_n, x_n)] = [\textstyle\sum_{i=1}^n (1/\sigma^2(t_i))]^{-1} \cdot [\sum_{i=1}^n (x_i/\sigma^2(t_i))]$$

then the resulting $\hat{m}_n(\cdot)$ provides the maximum likelihood estimator of $m(\cdot)$ (cf. Example 2.4 of Barlow, Bartholomew, Bremner and Brunk (1972)). It is a simple matter to verify conditions (a) through (e) in this situation.

We first consider the case where $\ll$ is a linear ordering on T (i.e., for all $s, t \in T$, $s \ll t$ or $t \ll s$). This includes the reals, and in fact any subset of real numbers, with the usual ordering. For n fixed we assume the distinct observation points are labeled so that $s_1 \ll s_2 \ll \cdots \ll s_k$ and we extend $\hat{m}_n$ to all of T as follows: $\hat{m}_n(t) = \hat{m}_n(s_i)$ for $s_i \ll t \ll s_{i+1} \neq t$ and $i = 1, \cdots, k-1$; $\hat{m}_n(t) = \hat{m}_n(s_1)$ for $t \ll s_1$; and $\hat{m}_n(t) = \hat{m}_n(s_k)$ for $s_k \ll t$. The proofs of the following two theorems will be given simultaneously.

THEOREM 2.1. *Let $M_n(\ )$ satisfy conditions* (a) *through* (d); *let T be a linearly ordered set; let j be such that for each $\varepsilon > 0$*

(2.2) *there exist s_b and s_a with $s_b \ll t_j \ll s_a$ and* $\max(m(t_j) - m(s_b), m(s_a) - m(t_j)) < \varepsilon$

and

$$\operatorname{card}\{k : t_j \ll t_k \ll s_a\} = \operatorname{card}\{k : s_b \ll t_k \ll t_j\} = \infty\,; \tag{2.3}$$

and for each $\varepsilon > 0$

$$P[\sup_{N \leqq n} |M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))]| \geqq \varepsilon] \to 0 \tag{2.4}$$

as $N \to \infty$ uniformly for all one-to-one functions, f, mapping $\{1, 2, \cdots\}$ onto $\{1, 2, \cdots\}$. Then $\hat{m}_n(t_j) \to_p m(t_j)$.

Through the remainder of this section, f will always be used to denote one-to-one mappings from $\{1, 2, \cdots\}$ onto $\{1, 2, \cdots\}$.

THEOREM 2.2. *If we assume the hypotheses of Theorem* 2.1 *with* (2.3) *replaced by*

$$\liminf_{n\to\infty} \operatorname{card}\{k \leqq n : t_j \ll t_k \ll s_a\}/n > 0 \quad \text{and} \quad \liminf_{n\to\infty} \operatorname{card}\{k \leqq n : s_b \ll t_k \ll t_j\}/n > 0 \tag{2.5}$$

and if in addition we assume for each $\varepsilon > 0$, there exist positive constants C and ρ with $\rho < 1$ such that

$$P[|M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))]| \geqq \varepsilon] \leqq C\rho^n \tag{2.6}$$

for all n and f, then for each $\varepsilon > 0$ there exist positive constants C_1 and ρ_1 with $\rho_1 < 1$ such that

$$P[|\hat{m}_n(t_j) - m(t_j)| \geqq \varepsilon] \leqq C_1 \rho_1^{\,n}\,.$$

PROOFS. Fix $\varepsilon > 0$ and fix j so that t_j has the properties guaranteed in the hypotheses of Theorems 2.1 and 2.2. Choose $s_a \in T$ so that $t_j \ll s_a$, $m(s_a) - m(t_j) < \varepsilon$, and $\operatorname{card}\{k : t_j \ll t_k \ll s_a\} = \infty$ or $\liminf_{n\to\infty} \operatorname{card}\{k \leqq n : t_j \ll t_k \ll s_a\}/n > 0$ depending on the theorem considered. For $n \geqq j$,

$$\hat{m}_n(t_j) - m(t_j) \leqq \max_{r \ll t_j} M\{X_k : k \leqq n, r \ll t_k \ll s_a\} - m(s_a) + \varepsilon\,. \tag{2.7}$$

Using (c), (d) and the fact that m is isotonic we see that the right-hand side of (2.7) can be rewritten

$$\max\nolimits_{r \ll t_j} M\{X_k - m(s_a): k \leqq n, r \ll t_k \ll s_a\} + \varepsilon \leqq \max\nolimits_{r \ll t_j} M\{X_k - m(t_k): k \leqq n, r \ll t_k \ll s_a\} + \varepsilon .$$

For each $n \geqq j$, let f be chosen so that f restricted to $\{1, 2, \cdots, n\}$ is a permutation of $\{1, 2, \cdots, n\}$ and $t_j = t_{f(1)}$; $t_{f(1)} \ll t_{f(2)} \ll \cdots \ll t_{f(l_n)} \ll s_a$; $t_j \gg t_{f(l_n+1)} \gg \cdots \gg t_{f(r_n)}$ and $t_j \neq t_{f(l_n+1)}$; and $s_a \ll t_{f(r_n+1)} \ll \cdots \ll t_{f(n)}$ and $s_a \neq t_{f(r_n+1)}$. We observe that f may depend on n and we define $Y_{k,n} = X_{f(k)} - m(t_{f(k)})$ for $k = 1, 2, \cdots$. Then

$$\hat{m}_n(t_j) - m(t_j) \leqq \max\nolimits_{l_n \leq k \leq n} M_k[(t_{f(1)}, Y_{1,n}), (t_{f(2)}, Y_{2,n}), \cdots, (t_{f(k)}, Y_{k,n})] + \varepsilon . \tag{2.8}$$

We now complete the proof of Theorem 2.1. By (2.4) there exists an N such that

$$P[\sup\nolimits_{N \leq n} |M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))]| \geqq \varepsilon] < \varepsilon$$

for all f. Since $l_n \to \infty$ we may choose n_0 such that $l_n \geqq N$ for $n \geqq n_0$. If $n \geqq n_0$, $P[\max_{l_n \leq k \leq n} M_k[(t_{f(1)}, Y_{1n}), (t_{f(2)}, Y_{2n}), \cdots, (t_{f(k)}, Y_{kn})] > \varepsilon] < \varepsilon$ and so $P[\hat{m}_n(t_j) - m(t_j) > 2\varepsilon] < \varepsilon$. Similarly it can be shown that $P[\hat{m}_n(t_j) - m(t_j) < -2\varepsilon] < \varepsilon$. So $\hat{m}_n(t_j) \to_p m(t_j)$.

For the proof of Theorem 2.2 we note that since $l_n > 0$ for all $n \geqq j$ and $\liminf_{n\to\infty} \operatorname{card}\{k \leqq n: t_j \ll t_k \ll s_a\}/n$ is positive there exists a positive constant d such that $l_n \geqq d \cdot n$ for all $n \geqq j$. Using (2.7) and the hypotheses of Theorem 2.2, $P[\hat{m}_n(t_j) - m(t_j) > 2\varepsilon] \leqq C \sum_{k=[d\cdot n]}^{\infty} \rho^k \leqq C[\rho(1-\rho)]^{-1}(\rho^d)^n$ for $n \geqq j$ and it is clear that one can choose a C^* so that this probability is bounded above by $C^*(\rho^d)^n$ for all n. The quantity $P[\hat{m}_n(t_j) - m(t_j) < -2\varepsilon]$ is treated similarly.

We now prove three corollaries which give more global results.

COROLLARY 2.3. *Let the hypotheses of Theorem* 2.1 *hold and let* $T' \subset T$. *If for each* $\varepsilon > 0$ *there exsists a finite set of observation points each satisfying* (2.2) *and* (2.3), *say* $\tau_1 \ll \tau_2 \ll \cdots \ll \tau_l$ *such that*

(2.9) *if* $t \in T'$ *then* $\tau_1 \ll t \ll \tau_l$ *and if* $t \in T'$, $\tau_i \ll t \ll \tau_{i+1}$ *and* $\tau_i \neq t \neq \tau_{i+1}$ *with* $i = 1, \cdots, l-1$, *then* $\max(m(t) - m(\tau_i), m(\tau_{i+1}) - m(t)) < \varepsilon$,

then $\sup_{t \in T'} |\hat{m}_n(t) - m(t)| \to_p 0$.

PROOF. Let $\varepsilon > 0$ be fixed and let $\tau_1 \ll \cdots \ll \tau_l$ be as in the hypotheses of the theorem. If $t \in T'$, $\tau_i \ll t \ll \tau_{i+1}$ and $\tau_i \neq t \neq \tau_{i+1}$, then since $\hat{m}_n$ is isotone $|\hat{m}_n(t) - m(t)| \leqq \max(|\hat{m}_n(\tau_i) - m(\tau_i)|, |\hat{m}_n(\tau_{i+1}) - m(\tau_{i+1})|) + \varepsilon$. Hence $\sup_{t \in T'} |\hat{m}_n(t) - m(t)| \leqq \max_{1 \leq i \leq l} |\hat{m}_n(\tau_i) - m(\tau_i)| + \varepsilon$. The conclusion follows immediately.

COROLLARY 2.4. *Let the hypotheses of Theorem* 2.2 *hold and let* $T' \subset T$. *If for each* $\varepsilon > 0$ *there exists a finite set of observation points satisfying* (2.2), (2.5) *and*

(2.9), *then for each* $\varepsilon > 0$ *there exist positive constants* C_1 *and* ρ_1 *with* $\rho_1 < 1$ *such that*

$$P[\sup_{t\in T'} |\hat{m}_n(t) - m(t)| \geqq \varepsilon] \leqq C_1 \rho_1{}^n .$$

PROOF. The proof is similar to the proof of Corollary 2.3.

If one is interested in a T' which cannot be "covered" with a finite number of observation points as in Corollary 2.4, then the following result may be of use. This would be the case if $T = T' = (-\infty, \infty)$ with the usual ordering and $m(t)$ were not bounded. We now assume the underlying probability space is complete.

COROLLARY 2.5. *Let the hypotheses of Theorem* 2.2 *hold and let* $T' \subset T$. *Assume that for each* $\varepsilon > 0$ *there exists a countable set of observation points satisfying* (2.2) *and* (2.5), *say* $\tau_0, \tau_1, \tau_{-1}, \tau_2, \tau_{-2}, \cdots$ *such that if* $t \in T'$ *then* $\tau_i \ll t \ll \tau_{i+1}$ *for some* i *and if* $t \in T'$, $\tau_i \ll t \ll \tau_{i+1}$ *and* $\tau_i \neq t \neq \tau_{i+1}$ *then* $\max(m(t) - m(\tau_i), m(\tau_{i+1}) - m(t)) < \varepsilon$. *Then* $P[\hat{m}_n(t) \to m(t)$ *for all* $t \in T'] = 1$.

PROOF. For each $\varepsilon > 0$ let $\{\tau_i(\varepsilon)\}_{i=-\infty}^{\infty}$ be the sequence of points hypothesized in Corollary 2.5. As in the proof of Corollary 2.3, if $\tau_i \ll t \ll \tau_{i+1}$, $\tau_i \neq t \neq \tau_{i+1}$ and $t \in T'$ then $|\hat{m}_n(t) - m(t)| \leqq \max(|\hat{m}_n(\tau_i) - m(\tau_i)|, |\hat{m}_n(\tau_{i+1}) - m(\tau_{i+1})|) + \varepsilon$. Hence $[\hat{m}_n(t) \to m(t)$ for all $t \in T'] \supset \bigcap_v \bigcap_i [\hat{m}_n(\tau_i(v^{-1})) \to m(\tau_i(v^{-1}))]$ and the desired conclusion follows immediately since the conclusion of Theorem 2.2 implies $\hat{m}_n(t_j) \to_{\text{a.s.}} m(t_j)$.

We remark that in Theorem 2.1, if $\{X_i - m(t_i)\}$ is a sequence of i.i.d. random variables then

$$M_n[(t_1, X_1 - m(t_1)), (t_2, X_2 - m(t_2)), \cdots, (t_n, X_n - m(t_n))] \to_{\text{a.s.}} 0$$

implies that (2.4) holds uniformly for all such f. The hypotheses regarding the convergence of $M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))]$ will be discussed further in connection with specific functions $M_n(\cdot)$ in Section 3.

Theorem 2 of Hanson *et al.* shows that a condition stronger than (2.2) is needed to obtain strong consistency in the mean regression case.

In Brunk (1970), Hanson, *et al.* and Cryer, *et al.* consistency results were obtained when T is a finite interval of real numbers and M_n is either the mean or median function. If $T = (a, b)$ with $a < b$, m is continuous and the sequence $\{t_i\}$ is dense in T, then Theorem 2.1 and Corollary 2.3 hold with $T' = [c, d]$ where $a < c < d < b$. Furthermore, if every nondegenerate sub-interval of (a, b) eventually contains at least some positive proportion of the observation points, then Theorem 2.2 and Corollary 2.4 hold with T' as above.

The results of this section can also be applied to $T = (-\infty, \infty)$ with the usual ordering. Conditions (2.2) and (2.3) hold if m is continuous and $\{t_i\}$ is dense. Furthermore, if each nondegenerate interval eventually contains at least some positive proportion of the observation points, then Theorem 2.2 and Corollary 2.4 hold with T' any finite interval of real numbers. Also Corollary 2.5 holds with $T' = (-\infty, \infty)$. One might wonder if there exist sequences which satisfy $\liminf_{n\to\infty} \text{card}\{k \leqq n : a \leqq t_k \leqq b\}/n > 0$ for all $a < b$. It is possible to actually construct such sequences; however, the following argument would appear to be of interest. Let $F(x)$ be a distribution function which assigns positive probability to every nondegenerate interval. If $U_1, U_2, \cdots$ is a sequence of independent random variables each having distribution function $F(x)$ and if $F_n(x)$ is the empirical distribution function based on $U_1, \cdots, U_n$ then $\sup_{-\infty<b<\infty} |F_n(b) - F(b)|$ converges almost surely to zero and so $\sup_{-\infty<a<b<\infty} |F_n(b) - F_n(a-) - F[a, b]|$ converges almost surely to zero. Hence there exists a null set outside of which $\text{card}\{k \leqq n : a \leqq U_k \leqq b\}/n \to F[a, b] > 0$ for all $a < b$.

While Theorem 2.2 does show that $m_n(t_j) \to_{\text{a.s.}} m(t_j)$, the hypotheses are apparently more restrictive than necessary for this conclusion. It seems clear that Theorem 2.2 could be modified to give strong consistency by replacing (2.6) with

$$\sum_{n=1}^{\infty} \sup_f P[\sup_{N\leqq n}|M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))]| \geqq \varepsilon] < \infty \qquad \text{for all} \quad \varepsilon > 0\,,$$

but it is not clear if such a result would be even reasonably tight. However, considering the mean regression case it appears that the results given thus far in this section are in some sense tight. In Brunk (1958, 1970) consistency results were obtained by considering permuted sums of random variables. It would be interesting to know if his results could be extended to more general functions and thereby obtain strong consistency results with weaker hypotheses than in Theorem 2.2.

Our consistency results when the ordering is not linear will be confined to E_β with the ordering $(x_1, x_2, \cdots, x_\beta) \ll (y_1, y_2, \cdots, y_\beta)$ if and only if $x_i \leqq y_i$ for $i = 1, 2, \cdots, \beta$ and $\beta \geqq 2$. We consider the two cases $T = E_\beta$ and $T = \prod_{i=1}^{\beta} (0, 1)$. Assume that $s_1, s_2, \cdots, s_k$ are the distinct observation points among $t_1, t_2, \cdots, t_n$. We let $\hat{m}_n(t)$ be any of the isotone extensions of $\hat{m}_n$ from $\{s_1, s_2, \cdots, s_k\}$ to all of T. Before stating the results for E_β we prove a lemma.

Note that points $0 < x_1 < x_2 < \cdots < x_\nu < 1$ determine a partition of (0, 1) as follows: $I_0 = (0, x_1)$, $I_j = [x_j, x_{j+1})$ for $j = 1, \cdots, \nu - 1$ and $I_\nu = [x_\nu, 1)$. Also $I(i(1), i(2), \cdots, i(\beta)) = \prod_{j=1}^{\beta} I_{i(j)}$ partitions $T = \prod_{j=1}^{\beta} (0, 1)$ into $(\nu + 1)^\beta$ disjoint sets. In a similar fashion $x_1 < x_2 < \cdots < x_\nu$ determines a partition of E_β. We let $\mathscr{R}$ denote the collection of sets of the form $R = \sum_{i(1)=0}^{\nu} \sum_{i(2)=0}^{\nu} \cdots \sum_{i(\beta-1)=0}^{\nu} \sum_{i(\beta)=n(i(1),\cdots,i(\beta-1))}^{\nu+1} I(i(1), i(2), \cdots, i(\beta))$ where $0 \leqq n(i(1), \cdots, i(\beta-1)) \leqq \nu + 1$ and $I(i(1), i(2), \cdots, i(\beta - 1), \nu + 1) = \emptyset$ for any $i(j) = 0, \cdots, \nu$ and $j = 1, \cdots, \beta - 1$.

Let L be an upper layer and for each $i(j) = 0, \cdots, \nu$ and $j = 1, \cdots, \beta - 1$ define $n(i(1), \cdots, i(\beta - 1))$ to be the smallest $i(\beta)$ for which $(x_{i(1)}, x_{i(2)}, \cdots, x_{i(\beta)}) \in L$ and $n(i(1), \cdots, i(\beta - 1)) = \nu + 1$ if no such $i(\beta)$ exists or if $i(j) = 0$ for some $j = 1, \cdots, \beta - 1$. Let $R(L)$ be the member of $\mathscr{R}$ associated with this function $n(\cdot)$. Corresponding to each $R(L)$ there is a function $n(\cdot) \geqq 1$ and based on it we define a function $n'(i(1), \cdots, i(\beta-1)) = n(i(1)+1, \cdots, i(\beta-1)+1)-1$ if $i(j) < \nu$ for $j = 1, 2, \cdots, \beta-1$ and $n'(i(1), \cdots, i(\beta-1)) = 0$, otherwise. Let $F(R(L))$ denote the member of $\mathscr{R}$ corresponding to $n'(\cdot)$. It can be seen that $R(L)$, $F(R(L)) \in \mathscr{L}$ for each $L \in \mathscr{L}$ but we will not use this fact in our arguments.

LEMMA 2.6. *If L is an upper layer $R(L) \subset L \subset F(R(L))$ and the number of $I(i(1), \cdots, i(\beta))$ in $F(R(L)) - R(L)$ is $O((\nu+1)^{\beta-1})$ as $\nu \to \infty$.*

PROOF. If $(y_1, \cdots, y_\beta) \in R(L)$ then $(y_1, \cdots, y_\beta) \in I(i(1), \cdots, i(\beta))$ with $i(j) > 0$ for $j = 1, \cdots, \beta-1$ and $\nu + 1 > i(\beta) \geqq n(i(1), \cdots, i(\beta-1)) > 0$. Hence $(x_{i(1)}, x_{i(2)}, \cdots, x_{i(\beta-1)}, x_{n(i(1),\cdots,i(\beta-1))}) \in L$ and $y_j \geqq x_{i(j)}$ for $j = 1, \cdots, \beta$. Since L is an upper layer, $(y_1, \cdots, y_\beta) \in L$.

Let $(y_1, \cdots, y_\beta) \in L$ and $(y_1, \cdots, y_\beta) \in I(i(1), \cdots, i(\beta))$. If $i(j) = \nu$ for some $j = 1, \cdots, \beta-1$ then $(y_1, \cdots, y_\beta) \in F(R(L))$. So we assume $i(j) < \nu$ for $j = 1, \cdots, \beta-1$ and show that $n'(i(1), \cdots, i(\beta-1)) \leqq i(\beta)$ which implies $(y_1, \cdots, y_\beta) \in F(R(L))$. Suppose $i(\beta) < n'(i(1), \cdots, i(\beta-1)) = n(i(1)+1, \cdots, i(\beta-1)+1)-1$ or $i(\beta)+1 \leqq n(i(1)+1, \cdots, i(\beta-1)+1)-1 \leqq \nu$. Then $y_j \leqq x_{i(j)+1}$ for $j = 1, \cdots, \beta$ and so $(x_{i(1)+1}, \cdots, x_{i(\beta)+1}) \in L$ but this contradicts the definition of $n(\cdot)$.

Since $R(L) \subset L \subset F(R(L))$ we see that $n'(i(1), \cdots, i(\beta-1)) \leqq n(i(1), \cdots, i(\beta-1))$ for all $i(j) = 0, \cdots, \nu$ and $j = 1, \cdots, \beta-1$. The number of $I(i(1), \cdots, i(\beta))$ in $F(R(L)) - R(L)$ is bounded above by

$$\begin{aligned}
&\sum_{i(1)=0}^{\nu} \cdots \sum_{i(\beta-1)=0}^{\nu} \sum_{i(\beta)=n'(i(1),\cdots,i(\beta-1))}^{n(i(1),\cdots,i(\beta-1))} 1 \\
&\quad \leqq \sum_{i(1)=0}^{\nu} \cdots \sum_{i(\beta-1)=0}^{\nu} n(i(1), \cdots, i(\beta-1)) \\
&\qquad - \sum_{i(1)=0}^{\nu-1} \cdots \sum_{i(\beta-1)=0}^{\nu-1} [n(i(1)+1, \cdots, i(\beta-1)+1)-1] \\
&\quad = \nu^{\beta-1} + (\nu+1)\, \mathrm{card}\, \{(i(1), \cdots, i(\beta-1)) \colon 0 \leqq i(j) \leqq \nu \\
&\qquad\quad \text{for } j = 1, \cdots, \beta-1, \text{ and } i(j) = 0 \\
&\qquad\quad \text{for some } j = 1, \cdots, \beta-1\}, \\
&\quad = \nu^{\beta-1} + (\nu+1)[(\nu+1)^{\beta-1} - \nu^{\beta-1}].
\end{aligned}$$

The desired conclusion follows from the fact that $1 - (1 - 1/x)^{\beta-1} \sim (\beta-1)/x$.

THEOREM 2.7. *Let M_n satisfy* (a) *through* (d) *and let $m(t)$ be continuous. If for every nondegenerate rectangle $J \subset T$,* $\liminf_{n\to\infty} \mathrm{card}\, \{k \leqq n \colon t_k \in J\}/n > 0$; *if there exists a sequence $c_n \to 0$ such that for every positive integer ν there are points $x_1 < x_2 < \cdots < x_\nu$ which partition T as above and an integer $n(\nu)$ such that for $n \geqq n(\nu)$ and $i(j) = 0, \cdots, \nu$ and $j = 1, \cdots, \beta$*

$$\mathrm{card}\, \{k \leqq n \colon t_k \in I(i(1), \cdots, i(\beta))\}/n \leqq (\nu+1)^{-\beta} + c_n; \tag{2.10}$$

and if for each $\varepsilon > 0$ there exist positive constants C and ρ with $\rho < 1$ such that (2.6) *holds for all n and f, then for each j and $\varepsilon > 0$ there exist positive constants C_1 and ρ_1 with $\rho_1 < 1$ such that*

$$P[|\hat{m}_n(t_j) - m(t_j)| \geqq \varepsilon] \leqq C_1 \rho_1^{\,n}.$$

PROOF. Fix j and let $\varepsilon > 0$ and $n \geqq j$. Choose s_a with each coordinate of s_a strictly greater than the corresponding coordinate of t_j and with $m(s_a) - m(t_j) < \varepsilon$. Set $L_0^c = \{t \in T: t \ll s_a\}$. Using (c), (d) and an argument like that given in the first part of the proof of Theorem 3.1 of Robertson and Wright (1973), it can be shown that

$$\hat{m}_n(t_j) - m(t_j) \leqq \max_{L \ni t_j} M\{X_k - m(t_k): k \leqq n, t_k \in L - L_0\} + \varepsilon .$$

With ν fixed and to be chosen later, partition T into the $(\nu + 1)^\beta$ sets $I(i(1), \cdots, i(\beta))$. Let $\mathscr{R}_0 = \{R(L): L \in \mathscr{L}\}$, let $F(R(L))$ be defined as in Lemma 2.6 and let $J = L_0^c \cap \{t: t_j \ll t\}$. We observe that $F(R(L))$ depends on L only through $R(L)$ and hence

$$\begin{aligned}\hat{m}_n(t_j) - m(t_j) &\leqq \max_L M\{X_k - m(t_k): k \leqq n, t_k \in (L - L_0) \cup J\} + \varepsilon \\ &\leqq \max_{R \in \mathscr{R}_0} \max_{A \subset F(R) - R} M\{X_k - m(t_k): k \leqq n, t_k \in (R \cup A - L_0) \cup J\} + \varepsilon .\end{aligned}$$

Hence $P[\hat{m}_n(t_j) - m(t_j) > 2\varepsilon] \leqq \sum_{R \in \mathscr{R}_0} \sum_{A \subset F(R) - R} P[M\{X_k - m(t_k): k \leqq n, t_k \in (R \cup A - L_0) \cup J\} > \varepsilon]$ for n sufficiently large. Now there exists a positive constant c' such that card $\{k \leqq n: t_k \in J\} \geqq c'n$ for n sufficiently large and hence for each $R \in \mathscr{R}_0$ and $A \subset F(R) - R$, using (2.6) we have $P[M\{X_k - m(t_k): k \leqq n, t_k \in (R \cup A - L_0) \cup J > \varepsilon\} \leqq C_\rho^{c'n}$. The cardinality of $\mathscr{R}_0$ is a constant depending only on ν, say d_ν. According to Lemma 2.6 we may choose a constant d so that the number of $I(i(1), \cdots, i(\beta))$ in $F(R) - R$ is bounded above by $d(\nu + 1)^{\beta-1}$ for all ν and $R \in \mathscr{R}_0$. Using (2.10) we see that for sufficiently large n, $P[\hat{m}_n(t_j) - m(t_j) > 2\varepsilon] \leqq C \cdot d_\nu \cdot 2^{dn(\nu+1)^{-1} + dnc_n(\nu+1)^{\beta-1}} \cdot \rho^{c'n} = C \cdot d_\nu \exp\{n(d(\nu + 1)^{-1} \ln(2) + dc_n(\nu + 1)^{\beta-1} \ln(2) + c' \ln(\rho))\}$. If we choose ν so that $d(\nu + 1)^{-1} \ln(2) < -(c' \ln(\rho))/4$ then for n large enough that $dc_n \cdot (\nu + 1)^{\beta-1} \ln(2) < -(c' \ln(\rho))/4$ we have $P[\hat{m}_n(t_j) - m(t_j) > 2\varepsilon] \leqq C \cdot d_\nu \rho_1^n$ with $\rho_1 = \rho^{c'/2}$. A similar treatment of $P[\hat{m}_n(t_j) - m(t_j) < -2\varepsilon]$ completes the proof.

We comment that the β-dimensional analogue of (27) of Hanson, *et al.* could be substituted for (2.10) in Theorem 2.7 and the proof given holds with only minor modifications. We have chosen (2.10) because it can be shown that "almost all" sequences $\{t_j\}$ generated from certain distributions satisfy this property, but it is not clear that they satisfy (27) of Hanson, *et al.* in the case $\beta = 2$.

Let $F(u)$ be an absolutely continuous distribution function which assigns positive probability to every nondegenerate sub-interval of $(-\infty, \infty)$ let $U_{11}, U_{12}, \cdots, U_{1\beta}, U_{21}, \cdots, U_{2\beta}, \cdots$ be a sequence of i.i.d. random variables each having distribution function $F(u)$. (The case where $F(u)$ has support (0, 1) is similar.) Denote the $j/(\nu + 1)$ quantile of F by x_j for $j = 1, \cdots, \nu$. So for each ν this determines a partition $I(i(1), \cdots, i(\beta))$ of T. Using the results by Kiefer (1961) and the fact that the probability of a rectangle $\prod_{i=1}^{\beta} (a_i, b_i]$ can be expressed as sums and differences of values of the associated distribution function, we see that there exists a constant C_β such that

$$\begin{aligned}P[\lim_{n\to\infty} \sup (2n/\ln \ln (n))^{\frac{1}{2}} \sup_{i(j),\nu} |\text{card}\, \{k \leqq n: (U_{k1}, \cdots, U_{k\beta}) \\ \in I(i(1), \cdots, i(\beta)) - (\nu + 1)^{-\beta}| \leqq C_\beta] = 1 .\end{aligned}$$

So almost surely the sequence $t_k = (U_{k1}, \cdots, U_{k\beta})$ satisfies (2.10). A similar argument shows that almost surely $\liminf_{n\to\infty} \text{card}\,\{k \leqq n: (U_{k1}, \cdots, U_{k\beta}) \in J\}/n > 0$ for each nondegenerate rectangle J.

We now prove two corollaries which give more global consistency results in E_β.

COROLLARY 2.8. *Let $T' = \prod_{j=1}^{\beta} [a, b]$ where $-\infty < a < b < \infty$ in the case $T = E_\beta$ or $0 < a < b < 1$ in the case $T = \prod_{j=1}^{\beta} (0, 1)$. If the hypotheses of Theorem 2.7 hold then there exist positive constants C_1 and ρ_1 with $\rho_1 < 1$ such that for each $\varepsilon > 0$, $P[\sup_{t\in T'} |\hat{m}_n(t) - m(t)| > \varepsilon] \leqq C_1 \rho_1^{\,n}$.*

PROOF. See the initial part of the proof of Theorem 5 of Hanson, *et al.*

COROLLARY 2.9. *If the hypotheses of Theorem 2.7 hold, then $P[\hat{m}_n(t) \to m(t)$ for all $t \in T] = 1$.*

PROOF. The result follows immediately from $[\hat{m}_n(t) \to m(t)$ for all $t \in T] = [\hat{m}_n(t_j) \to m(t_j)$ for all $j]$.

3. Examples. Suppose $w(\cdot)$ is a weight function defined on T such that $w(t) \geqq \delta > 0$ for each $t \in T$. In the normal regression problem discussed in Section 2, $w(t) = [\sigma^2(t)]^{-1}$. Define the weighted empirical distribution function, $F_n(\cdot)$, by

$$F_n(x) = \sum_{i=1}^{n} w(t_i) \cdot [\sum_{j=1}^{n} w(t_j)]^{-1} \cdot I_{(-\infty,x]}(X_i)\,.$$

In our examples M_n will depend on the t_i's and X_i's through $F_n(\cdot)$ and verification of (a) through (d) is routine.

EXAMPLE 3.1. Let

$$M_n[(t_1, X_1), (t_2, X_2), \cdots, (t_n, X_n)] = \int x\, dF_n(x) = (\sum_{j=1}^{n} w(t_j))^{-1} \sum_{j=1}^{n} w(t_j) X_j\,.$$

Van Eeden (1957), Brunk (1958, 1970) and Hanson *et al.* discuss consistency properties of this estimator. We assume for this example that $E(X_n) = m(t_n)$ for all n and observe that if $F(y) \to 0$ as $y \to \infty$ and $\int_0^\infty y|dF(y)| < \infty$ where $F(y) = \sup_n P[w(t_n)|X_n - m(t_n)| \geqq y]$ then Lemma 3 of Hanson, *et al.* shows (2.4) holds since $w(t)$ is bounded away from zero. Notice that if $w(t)$ is also bounded above and $\{X_n - m(t_n)\}$ is a sequence of identically distributed random variables this requirement is equivalent to $E|X_1 - m(t_1)| < \infty$. Also we observe that if $F(y) \to 0$ as $y \to \infty$ and $\int_0^\infty e^{\tau y}|dF(y)| < \infty$ for some $\tau > 0$ with F defined as above then Lemma 5 of this same paper ensures that (2.6) holds.

In this case where $T = [0, 1]$ with the usual ordering we see that Corollary 2.3 essentially contains the first conclusion of Theorem 1 of Hanson *et al.* and extends it to obtain weak consistency when $w(t)$ is not necessarily constant, and the observation points are dense. Furthermore, in this case, Brunk (1970) obtained strong consistency assuming that every nondegenerate sub-interval eventually contains some positive proportion of the observation points. His argument was modified in Hanson *et al.* to weaken the moment requirements on the sequence $\{X_n\}$. It is interesting to observe that part two of Theorem 1 of Hanson *et al.* can be extended to the case where $w(t)$ is not constant by using the argument given by Brunk (1970) for Theorem 4.1 and Lemma 2 of Hanson *et al.* applied to $\{w(t_n)[X_n - m(t_n)]\}$. Specifically this shows that $\sup_{a \leqq t \leqq b} |\hat{m}_n(t) - m(t)|$ converges almost surely to zero if $0 < a < b < 1$, every nondegenerate interval eventually contains at least some positive proportion of the observation points, $F(y) \to 0$ as $y \to \infty$ and $\int_0^\infty y\, dF(y) < \infty$. As a final comment in this case we note that Corollary 2.4 essentially contains Theorem 4 of Hanson *et al.* and Corollary 2.5 gives consistency results when $T = (-\infty, \infty)$.

The primary contribution of this paper to the isotonic mean regression problem is the extension of Theorem 6 of Hanson *et al.* to Euclidean space with dimension higher than 2 and the substitution of (2.10) for their condition (27). We conclude this example by mentioning two unresolved questions concerning consistency in the isotonic mean regression problem when T is not linearly ordered. Can strong consistency be obtained with conditions like a first moment on the X_i's? Are the rates presented in Theorem 5 of Hanson *et al.* tight?

EXAMPLE 3.2. Let $M_n[(t_1, X_1), (t_2, X_2), \cdots, (t_n, X_n)] = \min\{x : F_n(x) \geqq \frac{1}{2}\}$. M_n is a weighted sample median. The $\hat{m}_n$ derived from this M_n would seem to be of interest because it provides a solution to the following minimization problem. Let $s_1, s_2, \cdots, s_k$ be the distinct elements among $t_1, t_2, \cdots, t_n$ and label the X's among the first n taken at an observation point s_i by $X_{i,1}, X_{i,2}, \cdots X_{i,n_i}$ with $i = 1, 2, \cdots, k$. It is argued in Robertson and Wright (1974) that $\hat{m}_n(\ )$ minimizes $\sum_{i=1}^k w_i \sum_{j=1}^{n_i} |X_{ij} - m(s_i)|$ among all isotone m. This can be interpreted as follows: $\hat{m}_n$ is one of the closest isotone functions to the data in this weighted l_1 sense.

Let F_j denote the distribution function of X_j and assume that

(3.1) for each $\varepsilon > 0$ there is a $\gamma > 0$ such that $\inf_j F_j(m(t_j) + \varepsilon) - \frac{1}{2} \geqq \gamma$ and $\frac{1}{2} - \sup_j F_j(m(t_j) - \varepsilon) \geqq \gamma$

and that $w(t)$ is bounded above as well as away from zero. We now show that under these conditions (2.6) holds. Fix f and consider

$$P[M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))] > \varepsilon]$$
$$= P[\textstyle\sum_{i=1}^n w(t_{f(i)}) I_{(-\infty,\varepsilon]}(X_{f(i)} - m(t_{f(i)})) / \sum_{i=1}^n w(t_{f(i)}) < \frac{1}{2}]\,.$$

Let $Z_{f(i)} = I_{(-\infty,\varepsilon]}(X_{f(i)} - m(t_{f(i)}))$. This probability can be bounded above by $P[\sum_{i=1}^n w(t_{f(i)})(Z_{f(i)} - EZ_{f(i)}) < -\delta\gamma n]$. Since $w(t_{f(i)})(Z_{f(i)} - EZ_{f(i)})$ are bounded above uniformly in f, Lemma 5 of Hanson *et al.* may be used to complete this argument. The term $P[M_n[(t_{f(1)}, X_{f(1)} - m(t_{f(1)})), \cdots, (t_{f(n)}, X_{f(n)} - m(t_{f(n)}))] < -\varepsilon]$ is handled similarly.

In this setting with the above assumptions, Corollaries 2.4 and 2.8 provide generalizations of Corollary 2.4 of Cryer *et al.* and Corollary 3.2 of Robertson and Wright (1973) in that weighted medians and Euclidean spaces of dimension greater than two are considered.

Percentiles other than the median also fall within this framework. See Casady

(1972) and Section 4 of Robertson and Wright (1973). In fact, the smallest and largest order statistic satisfy the requirements placed on M_n.

EXAMPLE 3.3. Let $M_n[(t_1, X_1), \cdots, (t_n, X_n)]$ be the midrange of $X_1, \cdots, X_n$, that is, the average of the smallest and largest item. The arguments given in Chapter 3 of Ubhaya (1971) for Lemma 2 and Theorem 3 of Section 4 show that the $\hat{m}_n$ corresponding to this M_n provide a solution to the following minimization problem: minimize $\max_{1 \leqq i \leqq k, 1 \leqq j \leqq n_i} |X_{ij} - m(s_i)|$ among all isotone m, where k, n_i and X_{ij} were defined in Example 3.2. It should be noted that the solution is not necessarily unique, but this solution would seem to be of particular interest in this setting because consistency properties of the midrange have been studied extensively.

If $X_n - m(t_n)$ is a sequence of independent, identically distributed, symmetric, bounded random variables then it is well known that (2.6) holds and hence we have conditions on the sequence $\{t_n\}$ which ensure that $\hat{m}_n$ is consistent in this case. However, if the random variables are not bounded, but instead we assume $P[X_1 - m(t_1) \geqq y] \sim M \int_y^\infty \exp\{-Cx^p\}\, dx$ with $p > 1$, then as a consequence of Theorem 4.5 (with $k = 1$) of Barndorff-Nielsen (1963) we know (2.4) holds and so $\hat{m}_n$ is weakly consistent for m if m and $\{t_n\}$ satisfy the conditions of Theorem 2.1 or Corollary 2.3.

We conclude with an example of a function M_n which does not have the averaging property, (e).

EXAMPLE 3.4. We consider the trimmed mean as an estimator of the mean of a truncated distribution. We make several simplifying assumptions. Take $w(t) \equiv 1$ and assume that F_t is the distribution function associated with t and $0 < \alpha < \frac{1}{2}$. We assume that F_t is symmetric (not necessarily about zero), that it is continuous at its αth and $1 - \alpha$th percentiles and that these percentiles are unique. Let $m(t)$ be the mean of the distribution at t truncated at these percentiles ($m(t)$ is, in our specialized case, the point of symmetry). Suppose that $\{X_n - m(t)\}$ are independent and identically distributed and let $X_{n1} \leqq X_{n2} \leqq \cdots \leqq X_{nn}$ be the order statistics corresponding to $X_1, X_2, \cdots, X_n$. We define: $M_n[(t_1, X_1), (t_2, X_2), \cdots, (t_n, X_n)] = n(\alpha)^{-1} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{ni}$ where $n(\alpha) = n - 2[n\alpha]$ ($[n\alpha]$ denotes the largest integer in $n\alpha$). It is easy to construct examples illustrating that M_n does not satisfy condition (e) and that the associated $\hat{m}_n(\cdot)$ and $\bar{m}_n(\cdot)$ are not necessarily equal.

The probability in (2.4) and (2.6) reduces to considering $P[n(\alpha)^{-1}|\sum_{i=1}^n Z_i \times I_{[U_n, V_n]}(Z_i)| \geqq \varepsilon]$ where $Z_1, Z_2, \cdots$ are independent each having distribution which is symmetric about zero and U_n and V_n are the $[n\alpha] + 1$th and $n - [n\alpha]$th order statistics of $Z_1, Z_2, \cdots, Z_n$, respectively. Let ζ_α be the common $(1 - \alpha)$th percentile of the distribution at t centered at zero (i.e., shifted by $m(t)$). The probability of interest is bounded by

$$P[n(\alpha)^{-1} \cdot |\textstyle\sum_{i=1}^n Z_i \cdot \{I_{[U_n, V_n]}(Z_i) - I_{[-\zeta_\alpha, \zeta_\alpha]}(Z_i)\}| \geqq \varepsilon/2] + P[n(\alpha)^{-1} \cdot |\textstyle\sum_{i=1}^n Z_i \cdot I_{[-\zeta_\alpha, \zeta_\alpha]}(Z_i)| \geqq \varepsilon/2]\,.$$

The second term converges exponentially fast to zero (i.e., there exists constant C and ρ such that $0 < \rho < 1$ and the probability is bounded by $C \cdot \rho^n$) using well-known results about sums of independent, uniformly bounded random variables with zero mean. The first term is bounded by

$$P[n(\alpha)^{-1} \cdot \sum_{i=1}^n |Z_i| \cdot I_{[-\zeta_\alpha, \zeta_\alpha] \triangle [U_n, V_n]}(Z_i) \geqq \varepsilon/2] .$$

Suppose $\delta > 0$ and intersect this event with the event $[U_n \in (-\zeta_\alpha - \delta, -\zeta_\alpha + \delta)$, $V_n \in (\zeta_\alpha - \delta, \zeta_\alpha + \delta)]$ and its complement. The probability of the complement converges exponentially fast to zero using well-known properties of (U_n, V_n). The probability of the intersection of the two events in question is bounded above by

$$P[n(\alpha)^{-1} \cdot \sum_{i=1}^n |Z_i| \cdot I_{(-\zeta_\alpha - \delta, -\zeta_\alpha + \delta) \cup (\zeta_\alpha - \delta, \zeta_\alpha + \delta)}(Z_i) \geqq \varepsilon/2] .$$

Now, using the continuity of $F_t(\cdot)$ at the percentiles we can choose δ sufficiently small that the mean of each term in the sum is bounded by $\varepsilon/4$. The exponential convergence of this probability to zero then follows from the previously mentioned results about sums of independent random variables. Thus (2.6) is satisfied and Theorems 2.1 and 2.2 give consistency results for isotonized trimmed means.

Our assumptions in this example are unnecessarily restrictive. It seems clear that the symmetry and continuity assumptions on F_t and the assumption that $\{X_n - m(t_n)\}$ are identically distributed may be relaxed. However, the uniqueness of the percentiles does seem to be necessary.

REFERENCES

[1] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

[2] Barndorff-Nielsen, O. (1963). On the limit behavior of extreme order statistics. *Ann. Math. Statist.* **34** 992–1002.

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

[4] Brunk, H. D. (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference.* Cambridge Univ. Press. 177–195.

[5] Casady, Robert J. (1972). Monotone percentile regression estimation: large sample and other properties, Ph. D. thesis, Univ. of Iowa, Iowa City.

[6] Cryer, J. D., Robertson, Tim, Wright, F. T. and Casady, R. J. (1972). Monotone median regression. *Ann. Math. Statist.* **43** 1459–1469.

[7] Hanson, D. L., Pledger, Gordon and Wright, F. T. (1973). On consistency in monotonic regression. *Ann. Statist.* **1** 401–421.

[8] Kiefer, J. (1961). On large deviations of the empiric distribution function of vector random variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.

[9] Robertson, Tim and Wright, F. T. (1973). Multiple isotonic median regression. *Ann. Statist.* **1** 422–432.

[10] Robertson, Tim and Wright, F. T. (1974). A norm reducing property for isotonized Cauchy mean value functions. *Ann. Statist.* **2** 1302–1307.

[11] Ubhaya, Vasant A. (1971). Isotonic and convex optimization. Technical Report, Operations Research Center, Univ. of California, Berkeley.

[12] VAN EEDEN, C. (1957). Maximum likelihood estimation of partially or completely ordered parameters, I and II. *Indag. Math.* **19** 128–136, 201–211.

DEPARTMENT OF STATISTICS
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52240