The Annals of Probability
1979, Vol. 7, No. 1, 109–127

A STRONG LAW FOR VARIABLES INDEXED BY A PARTIALLY ORDERED SET WITH APPLICATIONS TO ISOTONE REGRESSION[1]

BY F. T. WRIGHT

University of Missouri-Rolla

In studying the asymptotic properties of certain isotone regression estimators, one is led to consider the maximum of sums of independent random variables indexed by a partially ordered set. An index set which is a sequence of β dimensional vectors, $\{t_k\}_{k=1}^{\infty}$, and the usual partial order on R_β, the β dimensional reals, are considered here. The random variables are assumed to satisfy a condition equivalent to a finite first moment in the identically distributed case and are assumed to be centered at their means. For $A \subset R_\beta$, let $S_n(A)$ denote the sum of those random variables with indices $t_k \in A$ and $k \leqslant n$. It is shown that if the sequence $\{t_k\}$ satisfies a certain condition, then the maximum, over all upper layers U in R_β, of $S_n(U)/n$ converges almost surely to zero. As a corollary to this result one obtains the strong consistency of this isotone regression estimator. If the sequence $\{t_k\}$ is a realization of a sequence of independent, identically distributed, β dimensional random vectors and if the probability induced by such a vector is discrete, absolutely continuous or a mixture of the two, then the condition on the sequence $\{t_k\}$ is satisfied almost surely. Some nondiscrete, singular induced probabilities of interest in these regression problems are considered also.

1. Introduction. Let $\{t_k\}$ be a sequence of points in $\Delta = x_{i=1}^{\beta}(a_i, b_i)$ with $-\infty \leqslant a_i < b_i \leqslant \infty$ for $i = 1, 2, \cdots, \beta$, and let $\{X_k\}$ be a sequence of independent random variables which are centered at their means. The variable X_k is to be thought of as associated with the point t_k for $k = 1, 2, \cdots$. For $A \subset \Delta$ let

$$S_n(A) = \Sigma_{\{k\,:\,k \leqslant n,\, t_k \in A\}} X_k$$

with $\Sigma_\varnothing = 0$. If $\nu_n(A) = \text{card}\,\{k : k \leqslant n, t_k \in A\} \to \infty$ as $n \to \infty$ and the random variables X_k are identically distributed, then the strong law of large numbers ensures that $S_n(A)/\nu_n(A) \to 0$ a.s. It is of interest to consider $M_n = \max_{A \in \mathcal{Q}} S_n(A)$ where $\mathcal{Q}$ is a collection of subsets of Δ. Under what conditions does $M_n/n \to 0$ a.s.? Some restriction must be placed on $\mathcal{Q}$; for if the t_k's are all distinct and $\mathcal{Q}$ is the power set of Δ, then $M_n = \Sigma_{k=1}^{n} X_k^+$ where $X_k^+ = \max(X_k, 0)$. (To see this, fix ω in the underlying probability space and consider $A_\omega = \{t_k : X_k(\omega) \geqslant 0\}$.) In this case, assuming the X_k are identically distributed, $M_n/n \to E(X_1^+)$ a.s. and $E(X_1^+) > 0$ unless the X_k are degenerate at zero.

Received November 19, 1976; revised January 25, 1978.

[1]This research was sponsored by the National Science Foundation under Grant Number MCS 76-06949.

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

Key words and phrases. Strong law of large numbers, partially ordered sets, isotone regression and strong consistency.

The regression application which has motivated this investigation suggests that we consider for $\mathcal{C}$ the collection of all upper layers, which we now define. Let $\ll$ be the usual partial order on R_β, the β dimensional reals, that is, $(s_1, s_2, \cdots, s_\beta) \ll (t_1, t_2, \cdots, t_\beta)$ provided $s_i \leqslant t_i$ for $i = 1, 2, \cdots, \beta$. Let $U \subset \Delta$. U is called an upper layer if $t \in U$ whenever $t \in \Delta$, $s \ll t$ and $s \in U$; we reserve the symbol U, with or without subscripts, for upper layers and $\mathcal{U}$ for the collection of upper layers. For the remainder of this discussion we take $\mathcal{C} = \mathcal{U}$. For this choice of $\mathcal{C}$, the case $\beta = 1$ has been considered. Theorem 6.1 of Brunk (1958) shows that $M_n/n \to 0$ a.s., if the X_k's satisfy an rth order Kolmogorov condition with $r \geqslant 1$. Lemma 1 of Hanson, Pledger and Wright (1973) shows that the variables need not have finite second moments. In particular, they have shown that

$$(1) \qquad F(y) \to 0 \quad \text{as} \quad y \to \infty \quad \text{and} \quad I = \int_0^\infty y|dF(y)| < \infty$$

suffices where $F(y) = \sup_k P[|X_k| \geqslant y]$ for $y \geqslant 0$. Integrals like the one in condition (1) are to be interpreted as Lebesgue-Stieltjes integrals with respect to the measure determined by the nondecreasing function F. (A sequence of random variables is said to satisfy (1) if its tail function F does.) There are sequences of variables which satisfy the rth order Kolmogorov condition and not (1) and vice versa. However, we use condition (1) because, in the identically distributed case, it is equivalent to $E|X_1| < \infty$. Of course, implicit in Lemma 1 of Hanson et al. (1973) is the apparently well-known result:

THEOREM 1. *If $\{X_k\}$ is a sequence of independent random variables which are centered at their means and satisfy* (1), *then $n^{-1}\sum_{k=1}^n X_k \to 0$ a.s.*

In the case being discussed, $\beta = 1$, Makowski (1973) has given some law of the iterated logarithm type results for M_n. To see that the results mentioned above apply to M_n a sequence of order preserving permutations needs to be constructed from $\{t_k\}$. This is done on page 886 of Makowski (1973).

If $\beta \geqslant 2$, conditions must be imposed on the sequence $\{t_k\}$ to obtain the desired convergence. Consider the following example: let $\beta = 2$, let $\Delta = R_2$, let the variable X_k be identically distributed and let the $t_k = (t_k^{(1)}, t_k^{(2)})$ be distinct and lie on the negative 45° line; that is, $t_k^{(2)} = -t_k^{(1)}$. For $t \in \Delta$, let $U(t) = \{s \in \Delta : t \ll s\}$. Then for any n and any nonempty A_n^*, a subset of $A_n = \{t_1, t_2, \cdots, t_n\}$, the upper layer $\bigcup_{t \in A_n^*} U(t)$ contains A_n^* but none of the elements in $A_n - A_n^*$. Hence, $M_n = \sum_{k=1}^n X_k^+$ as before and so M_n/n does not converge to zero unless the X_k's are degenerate.

In Section 2, assuming (1), conditions are imposed on $\{t_k\}$ which ensure that $M_n/n \to 0$ a.s. and these results are applied in Section 3 to give the strong consistency of a certain regression estimator. In the regression problems considered, the $t_k = (t_k^{(1)}, t_k^{(2)}, \cdots, t_k^{(\beta)})$ are to be thought of as values of the so-called independent variable and will be referred to as observation points. In some situations they may be the values of a sequence of β-dimensional random vectors,

$\{T_k\}$. Hence, if the T_k are independent and identically distributed as P, a probability measure on R_β, we would want the conditions imposed on the sequence of observation points to hold for $\{T_k\}$ with probability one, at least for a large collection of probability measures P. In Section 2, it is shown that the desired strong law holds for $\{t_k\}$ obtained in this manner if P is discrete, absolutely continuous or a mixture of discrete and absolutely continuous probabilities. The two-dimensional example given above shows that the strong law cannot hold for such $\{t_k\}$ with P an arbitrary singular probability. However, certain nondiscrete, singular probabilities of interest in the regression setting are considered. The primary emphasis here is on stochastic sequences of observation points, that is, sequences $\{t_k\}$ which are realizations of a sequence of random vectors $\{T_k\}$. However, the results will be stated first for deterministic sequences and then the conditions imposed on $\{t_k\}$ will be shown to hold almost surely for certain $\{T_k\}$.

We now describe the isotone regression estimator which has motivated this investigation. Suppose that for each $t \in \Delta \subset R_\beta$ there is a univariate probability distribution $D(t)$ with mean $\mu(t)$ and suppose that $\mu(t)$ is isotone with respect to $\ll$, that is, $\mu(s) \leqslant \mu(t)$ if $s \ll t$. Let $w(t)$ be a positive weight function defined on Δ, for $k = 1, 2, \cdots$, let Y_k have distribution $D(t_k)$ with $t_k \in \Delta$, and let $Y_1, Y_2, \cdots$ be independent. The variable Y_k is to be thought of as an observation taken at the observation point t_k and $w(t_k)$ is the weight to be associated with this observation. Based on the first n observations, Brunk (1958) proposed the following estimator of $\mu(t_k)$ with $k \leqslant n$:

$$\hat{\mu}_n(t_k) = \max_{\{U : t_k \in U\}} \min_{\{U' : t_k \notin U'\}} Av_n(U - U'), \tag{2}$$

where

$$Av_n(B) = \frac{\sum_{\{j \leqslant n : t_j \in B\}} w(t_j) Y_j}{\sum_{\{j \leqslant n : t_j \in B\}} w(t_j)}$$

for $B \cap \{t_1, t_2, \cdots, t_n\} \neq \varnothing$. This estimator is isotone on $\{t_1, t_2, \cdots, t_n\}$ and we let $\hat{\mu}_n$ denote any extension to Δ which is isotone on Δ.

For $\beta \geqslant 2$, the consistency results obtained for $\hat{\mu}_n$ are a strengthening of previous results. The rate of convergence given in Wright (1976) could be combined with the Borel-Cantelli lemma to give the almost sure consistency of $\hat{\mu}_n$. However, the result giving the desired rate of convergence requires $\int_0^\infty y^2 |dF(y)| < \infty$ rather than (1). More significantly, for stochastic observation points, which are independent and identically distributed as P, the results given here apply to a much larger class of P's than those in Hanson et al. (1973), Robertson and Wright (1975) and Wright (1976). To show that the rates of convergence given for $\hat{\mu}_n$ in those last three papers still hold under the assumptions on $\{t_k\}$ considered here, the rate at which $P[M_n/n \geqslant \varepsilon]$ converges to zero is considered in Section 2.

Smythe (1973, 1974) has considered strong laws for sums of random variables indexed by partially ordered sets. In the earlier work, the index set was taken to be the positive integer lattice on R_β and in the later work the set is assumed to be a "local lattice." (See his Definition 2.1.) In either case, there are only a finite number of elements in the lattice less than or equal to any given element. This is too restrictive for our purposes, because the t_k's must be dense in Δ if the regression estimator is to be consistent.

2. A strong law. Throughout this section we assume that the X_k are independent random variables, which are centered at their means and satisfy (1). With $S_n(\cdot)$ and M_n defined as in Section 1 we impose conditions on the sequence $\{t_k\}$ which imply that $M_n/n \to 0$ a.s. For $A \subset \Delta$ define

$$M_n(A) = \max_U S_n(U \cap A).$$

PROPOSITION 2. *Let A be a fixed subset of Δ. If for each $\varepsilon > 0$ there exists a set $B \subset \Delta$ with* $\limsup_{n\to\infty} \nu_n(B)/n \leqslant \varepsilon$ *and* $M_n(A \cap B^c)/n \to 0$ a.s., *then* $M_n(A)/n \to 0$ a.s.

PROOF. First observe that $E|X_k| = \int_0^\infty x|dP[|X_k| \geqslant x]| \leqslant I$ and that $I < \infty$ by (1). (Recall, $I = \int_0^\infty x|dF(x)|$ and that integrals of the form $\int g(x)|dF(x)|$ are taken with respect to the Lebesgue-Stieltjes measure determined by the nondecreasing function-F.) Hence, $F^*(x) = \sup_k P[||X_k| - E|X_k|| \geqslant x] \leqslant F(x - I)$ for $x \geqslant I$ and so F^* satisfies (1). (For future reference observe that, if the sequence $\{X_k\}$ satisfies (1) and $\{l_k\}$ is bounded, then $\{X_k + l_k\}$ also satisfies (1).) With $\varepsilon > 0$, fixed,

$$M_n(A \cap B)/n \leqslant (\nu_n(B)/n)(\nu_n(B))^{-1}\Sigma_{\{k\,:\,k<n,\, t_k \in B\}}|X_k|.$$

If $\nu_n(B) \to \infty$ as $n \to \infty$, then applying the strong law given in Theorem 1, we obtain $\limsup_{n\to\infty} M_n(A \cap B)/n \leqslant I\varepsilon$ a.s. If $\nu_n(B)$ is bounded then $\limsup_{n\to\infty} M_n(A \cap B)/n = 0$ a.s. Since $M_n(A) \leqslant M_n(A \cap B) + M_n(A \cap B^c)$ and since ε was arbitrary the desired result follows.

It is convenient to consider those points t for which $\limsup_{n\to\infty} \nu_n(\{t\})/n$ is positive or zero separately. For stochastic observation points, which are independent and identically distributed as P, this first type of points would correspond to the possible values in the discrete part of P. Let $B_d = \{t_k\colon \limsup_{n\to\infty} \nu_n(\{t_k\})/n > 0\}$. At first it might seem that no conditions need to be placed on the elements of B_d since $\nu_n(\{t\}) \to \infty$ for each $t \in B_d$ (which together with (1) implies that the strong law holds at each such t). However, for infinitely many n there may be a substantial number of t_k with $k \leqslant n$ which occur only once in $t_1, t_2, \cdots, t_n$. If the points $\{t_k\}$ are such that any subset of $\{t_k\colon k = 1, 2, \cdots\}$ can be included in an upper layer without including an element of $\{t_k : k = 1, 2, \cdots\}$ not in the given subset, then $M_n(B_d)/n$ need not converge to zero. In fact, in the example that is given below $\liminf_{n\to\infty} \nu_n(\{t_k\})/n > 0$ for $k = 1, 2, \cdots$ but $M_n/n = M_n(B_d)/n$ does not converge to zero in probability.

EXAMPLE. Let $\{s_k\}$ be a sequence of distinct points with s_j and s_k unordered for each j and k with $j \neq k$. (For instance, let $\beta = 2$, $\Delta = R_2$, and let each s_k be on the negative 45° line.) Let $\{t_k\}$ be a sequence of points with $t_j \in \{s_k : k = 1, 2, \cdots\}$

and $\lim\inf_{n\to\infty}\nu_n(\{s_j\})/n > 0$ for $j = 1, 2, \cdots$. (If $\{T_k\}$ is a sequence of independent, identically distributed random vectors with support $\{s_k : k = 1, 2, \cdots\}$, then, with probability one, $\{T_k\}$ is such a sequence.) From $\{t_k\}$ a new sequence is constructed with $t'_{2j-1} = t_j$ for $j = 1, 2, \cdots$. Next the t'_{2j} are chosen to be distinct for $j = 1, 2, \cdots$ with $t'_{2j} \in \{s_k : k = 1, 2, \cdots\} - \{t_1, t_2, \cdots, t_2\alpha\}$ for $2^\alpha < 2j \leqslant 2^{\alpha+1}$ and $\alpha = 0, 1, 2, \cdots$. Letting $\nu'_n(A)$ denote the function card $\{j \leqslant n : t'_j \in A\}$ defined for $A \subset R_\beta$, we see that $\lim\inf_{n\to\infty}\nu'_n(\{s_j\})/n > 0$ for $j = 1, 2, \cdots$ and that card $\{j \leqslant n : \nu'_n(\{t'_j\}) = 1\}/n \geqslant \frac{1}{4}$ for $n = 2^{\alpha+1}$ and $\alpha = 0, 1, \cdots$. Hence, $M_2\alpha + 1\ (B_d)/2^{\alpha+1} \geqslant \sum_{k=1}^{2^{\alpha-1}} X_2^+\alpha_{+2k}/2^{\alpha+1} \to_p E(X_1^+)/4$ if the $\{X_k\}$ are independent and identically distributed.

The crucial elements in the example were the ordering relationship among the observation points and the fact that

$$\lim\sup_{n\to\infty}\text{card}\{j \leqslant n : \nu'_n(\{t'_j\}) \leqslant M\}/n > 0$$

for some M positive. (Observe that $\inf_{1\leqslant k\leqslant m}E(X_1 + X_2 + \cdots + X_k)^+ > 0$ for $\{X_k\}$ independent, indentically aistributed, nondegenerate and centered at their means.) Thus this example provides a partial converse for the next theorem.

THEOREM 3. *If the sequence* $\{t_k\}$ *satisfies the following*:

(3) $\quad \text{card}\{k \leqslant n : t_k \in B_d \text{ and } \nu_n(\{t_k\}) \leqslant M\}/n \to 0 \qquad \text{for} \quad M = 1, 2, \cdots,$

then $M_n(B_d)/n \to 0$ a.s.

PROOF. If card $(B_d) < \infty$ the conclusion is obvious and so we suppose card $(B_d) = \infty$ and show that $\lim\sup_{n\to\infty}M_n(B_d)/n$ can be made arbitrarily small with probability one. Let $\varepsilon > 0$, let M and K be positive integers to be chosen later, let $M^* = M \cdot K$ and let $B_{M^*,n} = \{t_k : k \leqslant n, \nu_n(\{t_k\}) \leqslant M^*\}$. Since $M_n(B_d)$ is bounded above by the sum of

$$V_n = \sum_{t\in B_d\cap B_{M^*,n}}(S_n(t))^+ \text{ and } W_n = \sum_{t\in B_d\cap B^c_{M^*,n}}(S_n(t))^+,$$

we consider V_n/n and W_n/n separately. Setting $S_n^*(A) = \sum_{\{k\leqslant n : t_k\in A\}}(X_k^+ - E(X_k^+))$ for $A \subset R_\beta$ and recalling that $0 \leqslant E(X_k^+) \leqslant E|X_k| \leqslant I = \int_0^\infty y|dF(y)|$, we have

$$\begin{aligned} V_n/n &\leqslant S_n^*(B_d \cap B_{M^*,n})/n + I \cdot \text{card}\{j \leqslant n : t_j \in B_d \cap B_{M^*,n}\}/n \\ &= S_n^*(B_d)/n - S_n^*(B_d \cap B^c_{M^*,n})/n + o(1). \end{aligned}$$

Clearly $\{X_k^+\}$ satisfies (1) since $\{X_k\}$ does, and since $E(X_k^+)$ is bounded $\{X_k^+ - E(X_k^+)\}$ also satisfies (1). So $S_n^*(B_d)/n \to 0$ a.s. by Theorem 1. By relabelling and applying Theorem 1 again, we will show that the second term in the above expression converges to zero almost surely. Let n_1 be the smallest integer for which $B_d \cap B^c_{M^*,n_1} \neq \varnothing$; then there exist $1 \leqslant k_1 < k_2 < \cdots < k_{M^*} < k_{M^*+1} = n_1$ with $t_{k_i} = t_{n_1}$ for $i = 1, 2, \cdots, M^*$. Set $Z_i = X_{k_i}^+ - E(X_{k_i}^+)$ for $i = 1, 2, \cdots, M^* + 1$. Denote by n_2 the next integer after n_1 at which another summand (or summands) occurs in $S_n^*(B_d \cap B^c_{M^*,n})$, and if $t_{n_2} = t_{n_1}$ denote this summand by Z_{M^*+2} (or if $t_{n_2} \neq t_{n_1}$ denote these summands by $Z_{M^*+2}, Z_{M^*+3}, \cdots, Z_{2M^*+2}$). Continuing this process we obtain an infinite sequence of independent random variables which are centered at their means and satisfy (1). Next observe that $|S_n^*(B_d \cap B_{M^*,n}^c)|/n \leqslant |S_n^*(B_d \cap B_{M^*,n}^c)|/\nu_n(B_d \cap B_{M^*,n}^c)$ and that the latter is a subsequence of $\Sigma_{i=1}^m Z_i/m$ which converges almost surely to zero. Hence, $V_n/n \to 0$ a.s.

In considering W_n, we denote by $s_1, s_2, \cdots$ the distinct points in B_d, in the order in which they become elements in $B_{M^*,n}^c$; that is, if n_1 is the smallest integer n for which $B_d \cap B_{M^*,n}^c \neq \varnothing$ then $s_1 = t_{n_1}$, etc. For $i = 1, 2, \cdots$, let $j_1^{(i)} < j_2^{(i)} < \cdots$ be the integers for which $t_{j_\alpha^{(i)}} = s_i$ and set $Z_{\alpha+1}^{(i)} = (X_{j_{\alpha M+1}^{(i)}} + \cdots + X_{j_{(\alpha+1)M}^{(i)}})^+/M$ for $\alpha = 0, 1, \cdots$ and $i = 1, 2, \cdots$. For i with $s_i \in B_{M^*,n}^c$, let $\delta_n^{(i)}$ be the largest integer δ for which $j_\delta^{(i)} \leqslant n$ and let $\alpha_n^{(i)}$ be the largest integer α for which $\alpha M \leqslant \delta_n^{(i)}$; then

$$W_n \leqslant \Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}\left\{M\Sigma_{\alpha=1}^{\alpha_n^{(i)}} Z_\alpha^{(i)} + \Sigma_{\alpha=\alpha_n^{(i)}+1}^{\delta_n^{(i)}} X_{j_\alpha}^{+}(i)\right\}$$

where the last sum is zero if $\alpha_n^{(i)} = \delta_n^{(i)}$. Applying Lemma 6 of Hanson et al. (1973), we choose M large enough so that $E(Z_\alpha^{(i)}) \leqslant \varepsilon/2$ for all i and α. Since

$$\Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}} M\alpha_n^{(i)}/n \leqslant \Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}} \delta_n^{(i)}/n \leqslant 1,$$

W_n/n is bounded above by the sum of

$$M\Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}\Sigma_{\alpha=1}^{\alpha_n^{(i)}}\left(Z_\alpha^{(i)} - E(Z_\alpha^{(i)})\right)/n + \varepsilon/2 \tag{4}$$

and

$$\Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}\Sigma_{\alpha=\alpha_n^{(i)}+1}^{\delta_n^{(i)}} X_{j_\alpha^{(i)}}^{+}/n. \tag{5}$$

Clearly $\{Z_\alpha^{(i)} - E(Z_\alpha^{(i)})\}$ satisfies (1) since $E(Z_\alpha^{(i)})$ is bounded and $P[|Z_\alpha^{(i)}| \geqslant y] \leqslant MF(y)$ and so Theorem 1 can be applied to show that the limit superior of (4) is less than or equal to $\varepsilon/2$ almost surely. In considering (5), we observe that $\Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}(\delta_n^{(i)} - \alpha_n^{(i)}) \leqslant M$ card $\{i : s_i \in B_{M^*,n}^c\}$, $n \geqslant KM$ card $\{i : s_i \in B_{M^*,n}^c\}$, and as we have seen before $E(X_k^+) \leqslant$ I. Hence, if K is chosen so that $I/K \leqslant \varepsilon/2$, then (5) is bounded by

$$\Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}\Sigma_{\alpha=\alpha_n^{(i)}+1}^{\delta_n^{(i)}}\left(X_{j_\alpha^{(i)}}^{+} - E\left(X_{j_\alpha^{(i)}}^{+}\right)\right)/n + \varepsilon/2$$
$$= S_n^*(B_d \cap B_{M^*,n}^c)/n - \Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}\Sigma_{\alpha=1}^{\alpha_n^{(i)}}\left(X_{j_\alpha^{(i)}}^{+} - E\left(X_{j_\alpha^{(i)}}^{+}\right)\right)/n + \varepsilon/2.$$

We have already seen that $S_n^*(B_d \cap B_{M^*,n}^c)/n \to 0$ a.s. For fixed i, $\alpha_n^{(i)}$ is increasing in n and $s_i \in B_{M^*,n}^c$ for some n_0 implies $s_i \in B_{M^*,n}^c$ for each $n \geqslant n_0$. So

$$\Sigma_{\{i\,:\,s_i \in B_{M^*,n}^c\}}\Sigma_{\alpha=1}^{\alpha_n^{(i)}}\left(X_{j_\alpha^{(i)}}^{+} - E\left(X_{j_\alpha^{(i)}}^{+}\right)\right)/n \to 0 \text{ a.s.}$$

or $\limsup_{n\to\infty} W_n/n \leqslant \varepsilon$ and the proof is completed.

COROLLARY 4. *If* $\{T_k\}$ *is a sequence of independent, identically distributed, β-dimensional random vectors, then* $\{T_k\}$ *satisfies* (3) *almost surely.*

PROOF. Choose $\Delta = R_\beta$. For almost all ω in the underlying probability space the set $B_d = \{t : \limsup_{n\to\infty}$ card $\{k \leqslant n : T_k(\omega) = t\}/n > 0\}$ is the collection of points in R_β which are assigned positive probability under P, the probability induced by T_1. Fix a positive integer M. Let C denote the support of the discrete part of P and for an arbitrary $\varepsilon > 0$ let C_ε be a finite subset of C with $P(C - C_\varepsilon) \leqslant \varepsilon$. Since card $\{k \leqslant n : T_k(\omega) = t\}/n \to P(\{t\})$ a.s. for each $t \in C$, with probability one there exists an n_0 (possibly depending on ω) with card $\{k \leqslant n : T_k(\omega) = t\} > M$ for each $t \in C_\varepsilon$ and $n \geqslant n_0$. So for almost all ω and $n \geqslant n_0$

$$\text{card}\quad \{k \leqslant n : T_k \in B_d \quad \text{and card} \quad \{j \leqslant n : T_j(\omega) = T_k(\omega) \leqslant M\}\}/n$$
$$\leqslant \text{card} \quad \{k \leqslant n : T_k(\omega) \in C - C_\varepsilon\}/n.$$

The proof is completed by noting that the last expression converges almost surely to $P(C - C_\varepsilon) \leqslant \varepsilon$.

In considering points not in B_d, we partition R_β into disjoint rectangles and group these rectangles into "chains." A chain has the property that a set composed of one element from each rectangle in the chain is linearly ordered and so one-dimensional results can be applied. For $i = 1, 2, \cdots, \beta$, let $\{x(j, i)\}_{j=0}^{\eta_i}$ partition (a_i, b_i); in particular, let $b_i = x(0, i) > x(1, i) > x(2, i) > \cdots > x(\eta_i, i) = a_i$ for $i = 1, 2, \cdots, \beta$. Let $I(j, i) = [x(j, i), x(j - 1, i))$ for $j = 1, 2, \cdots, \eta_i - 1$ and let $I(\eta_i, i) = (a_i, x(\eta_i - 1, i))$. For $i = 1, 2, \cdots, \beta$, let $j(i)$ be an index with possible values $1, 2, \cdots, \eta_i$ and let $I(j(1), j(2), \cdots, j(\beta)) = X_{i=1}^{\beta} I(j(i), i)$. This is the desired partition of Δ. A collection of sets of the form $\{I(j(1) + \tau, j(2) + \tau, \cdots, j(\beta) + \tau)\}_{\tau=0}^{\mu-1}$ is called a chain provided $j(i) \in \{1, 2, \cdots, \eta_i\}$ for $i = 1, 2, \cdots, \beta$; $j(i_0) = 1$ for some i_0; and $\mu - 1 = \min\{\tau : j(i) + \tau = \eta_i\}$. With λ the number of chains we arbitrarily label the chains $1, 2, \cdots, \lambda$ and for $k = 1, 2, \cdots, \lambda$ we relabel the sets in the kth chain as follows: set $C_{k, \tau+1} = I(j(1) + \tau, j(2) + \tau, \cdots, j(\beta) + \tau)$ for $\tau = 0, 1, \cdots, \mu_{k-1}$ where μ_k is the number of rectangles in the kth chain. One of the hypotheses of the next result involves a sequence of such partitions, $\{C_{k,\tau}^{(\alpha)} : \tau = 1, 2, \cdots, \mu_k^{(\alpha)}, k = 1, 2, \cdots, \lambda^{(\alpha)}\}_{\alpha=1}^{\infty}$. We will denote the αth element in the sequence by $\mathcal{C}^{(\alpha)}$ and its partition points by $x^{(\alpha)}(j, i)$.

While the sequence $\{t_k\}$ need not be stochastic, it is convenient to pose the restrictions on this sequence in terms of a probability measure Q defined on the Borel subsets of Δ.

THEOREM 5. *Suppose that the sequence $\{t_k\}$ satisfies the following:*

(6) *for each $\varepsilon > 0$ there is a set D with $\limsup_{n\to\infty} \nu_n(D)/n \leqslant \varepsilon$, a probability measure Q, a constant M, a sequence of constants $c_n \to 0$ as $n \to \infty$ and a sequence of partitions $\{\mathcal{C}^{(\alpha)}\}_{\alpha=1}^{\infty}$, with $x^{(\alpha)}(\eta_i - j, i) \to a_i$ as $\alpha \to \infty$ for each fixed j and i, for which the following hold with $Q_{k,\tau}^{(\alpha)} = Q(C_{k,\tau}^{(\alpha)} - B_d)$,*

(a) $\nu_n(C_{k,\tau}^{(\alpha)} - B_d - D)/n \leqslant Q_{k,\tau}^{(\alpha)} + c_n$ *for all* k, τ *and* α *and*

(b) $\sum_{k=1}^{\lambda^{(\alpha)}} \mu_k^{(\alpha)} \max_\tau Q_{k,\tau}^{(\alpha)} \leqslant M$ *for all* α.

Then $M_n(\Delta - B_d)/n \to 0$ a.s.

Before the proof is given we wish to comment on condition (6) and prove a lemma. It should first be noted that $x^{(\alpha)}(\eta_i - j, i) \to a_i$ for each i and j implies that

$\eta_i^{(\alpha)} \to \infty$ as $\alpha \to \infty$ for $i = 1, 2, \cdots, \beta$ and, as we shall see in the proof of the theorem, this in turn means that for any $t \in \Delta$ the length of the chain containing t tends to infinity with α. This with 6(b) forces Q to be continuous except possibly on B_d. Condition 6(a) essentially requires that these special rectangles contain no more observation points than if the observation points were generated by a sequence of independent random vectors each distributed as Q. However, as we have already seen this is not sufficient. Certain continuous singular measures (for example, the continuous probabilities whose supports are contained in the negative 45° line in the plane) do not yield suitable sequences of observation points. Condition 6(b) is clearly not optimal since it also excludes some continuous singular probabilities whose supports are contained in linearly ordered sets, such as a uniform distribution over the line segment joining (0, 0) and (0, 1) in the plane. However, it will be shown that 6(b) is satisfied almost surely if the observation points are generated stochastically by a probability whose continuous singular part vanishes and some continuous singular probabilities will be considered in Theorem 8 and Corollaries 9 and 10.

Let $\pi_n = (i(1, n), i(2, n), \cdots, i(n, n))$ be a permutation of the first n integers for $n = 1, 2, \cdots$. If the nth permutation is obtained by assigning a place to the integer n between two of the successive values or at the end or the beginning of the permutation π_{n-1} for $n = 1, 2, \cdots$, then the sequence has been called an order preserving sequence.

LEMMA. *Let $\{W_n\}_{n=1}^{\infty}$ be a sequence of independent random variables which are centered at their means and satisfy* (1). *Let $\{\pi_n\}_{n=1}^{\infty}$ be a sequence of order preserving permutations and set $S_{j,n} = \sum_{\nu=1}^{j} W_{i(\nu,n)} (S_{0,n} \equiv 0)$, $T_{n,n} = \max_{j=0,1,\cdots,n} S_{j,n}$ and $V_N = \max_{n=1,2,\cdots,N} T_{n,n}$. For each $\delta > 0$, there exists a positive integer N_0, depending only on δ and F, for which $|E(V_N/N)| \leqslant \delta$ for all $N \geqslant N_0$.*

PROOF. Consider the first quadrant in the plane and place W_1 on the line $y = 1$ at a point $(x_1, 1)$ with $x_1 > 0$. Next place W_2 on the line $y = 2$ at a point $(x_2, 2)$ with $x_2 > 0$ and $x_2 < x_1$ if $i(1, 2) > i(2, 2)$ or $x_2 > x_1$ if $i(1, 2) < i(2, 2)$. Continue this process until $W_1, W_2, \cdots, W_N$ have placed on the lines $y = 1, y = 2, \cdots, y = N$ with $x_1, x_2, \cdots, x_n$ ordered according to the permutation π_N. Drawing the vertical lines through the N points (x_i, i) for $i = 1, 2, \cdots, N$ forms an $N \times N$ grid with the random variables W_i at some of the grid points. Assign random variables degenerate at 0 to the other grid points and with $x_{(1)} < x_{(2)} < \cdots < x_{(N)}$ the ordered x_i's, denote the grid point $(x_{(i)}, j)$ by (i, j) and the associated random variable by $W_{(i,j)}$ for $i, j = 1, 2, \cdots, N$. For each $n \leqslant N$ and $k \leqslant n$, $S_{k,n}$ can be written as $\sum_{(i',j') \ll (i,j)} W_{(i',j')}$ for some $(i, j) \ll (N, N)$ and $V_N = \max_{(i,j) \ll (N,N)} \sum_{(i',j') \ll (i,j)} W_{(i',j')}$. The inequality given by Gabriel (1974) then shows that

$$P[V_N \geqslant N\gamma] \leqslant 9\gamma^{-1} E|S_{N,N}|/N.$$

Lemma 6 of Hanson et al. (1973) shows that there exists an N_0 depending only on ε

and F for which $E|S_{N,N}|/N \leqslant \varepsilon$ for $N \geqslant N_0$. The conclusion of this lemma follows just as inequality (30) of Lemma 6 follows from (20) of Lemma 3 of Hanson et al. (1973).

PROOF OF THEOREM 5. Because of Proposition 2 it suffices to show that with ε a fixed, arbitrary positive number, $M_n(\Delta - (B_d \cup D))/n \to 0$ a.s. If there are only a finite number of indices k for which $t_k \in \Delta - (B_d \cup D)$, the proof is trivial. Otherwise, we consider the subsequence of points in $\Delta - (B_d \cup D)$ and, for simplicity, denote it by $\{t_k\}$ also. With α fixed but to be chosen later, we select a partition, $\mathcal{C}^{(\alpha)}$, from the sequence in (6). (In the argument that follows many of the quantities depend on α; however, since α is fixed, this dependency will not be shown.)

With k fixed, those of the first n elements in $\{t_j\}$ which are in the kth chain are grouped into "strings." In forming these strings t_{j_1} and t_{j_2} with $j_1 \neq j_2$ will be considered different elements in $\{t_j\}$ even though $t_{j_1} = t_{j_2}$ and a string will contain at most one element from each rectangle in a chain. Such a grouping will be referred to as a string at time n in chain k. If none of the first n elements are in chain k, there is one string in chain k at time n and it is empty. Denote by $l(j)$ the jth smallest index of an element of the sequences $\{t_n\}$ which is in the kth chain. (There may be a finite or countably infinite number of such indices, but in either case, $l(j)$ increases with j and the collection of all such $l(j)$ is $\{i : t_i \in C_{k,\tau}$ for some $\tau = 1, 2, \cdots, \mu_k\}$.) Let $\tau(j)$ denote the index of the rectangle in the kth chain which contains $t_{l(j)}$, that is, $t_{l(j)} \in C_{k,\tau(j)}$. At time n, $n < l(1)$, there is one string in the kth chain and it is empty and at n, $l(1) \leqslant n < l(2)$, there is one string in chain k and it contains one element $t_{l(1)}$. If $t_{l(2)} \not\in C_{k,\tau(1)}$ then at time n, $l(2) \leqslant n < l(3)$, chain k has one string with elements $t_{l(1)}$ and $t_{l(2)}$; but if $t_{l(2)} \in C_{k,\tau(1)}$ there are two strings in chain k at time n, $l(2) \leqslant n < l(3)$, the first contains $t_{l(1)}$ and the second contains $t_{l(2)}$. With the requirement that once an element is placed in a string it remains in that string, we now describe the strings inductively. At time n there are $\gamma_n(k) = \max_\tau \nu_n(C_{k,\tau})$ strings in the kth chain. Suppose that at time $n = l(m) - 1$ the $\gamma_n(k)$ strings in chain k have been determined. If one of these strings does not contain an element from $C_{k,\tau(m)}$ then place $t_{l(m)}$ in the string; but if each of these strings contains an element from $C_{k,\tau(m)}$ then a new string is started with one element, $t_{l(m)}$. Let $\nu_{k,j}^{(n)}$ denote the number of elements in the jth string in the kth chain at time n and let $G_{k,j}^{(n)}$ denote the collection of these $\nu_{k,j}^{(n)}$ elements. Let $\nu_{k,j} = \lim_{n\to\infty} \nu_{k,j}^{(n)}$, observe that $\nu_{k,j} \leqslant \mu_k$ and set

$$Z_{k,j} = \max_n \max_U S_n(U \cap G_{k,j}^{(n)}).$$

Hence, $M_n(\Delta - (B_d \cup D))/n$ is bounded above by

$$n^{-1}\Sigma_{k,j}^{(n)}(Z_{k,j} - E(Z_{k,j})) \tag{7}$$

$$+ n^{-1}\Sigma_{k,j}^{(n)} E(Z_{k,j}) \tag{8}$$

where $\Sigma_{k,j}^{(n)} = \Sigma_{\{k,j\,:\, G_{k,j}^{(n)} \neq \varnothing,\, k=1,2,\cdots,\lambda,\, j=1,2,\cdots,\gamma_n(k)\}}$.

The number of summands in (7) is less than or equal to the number of strings, $G_{k,j}^{(n)}$, which are nonempty at time n, which in turn is less than or equal to n. Hence, expression (7) converges to zero almost surely if $\{Z_{k,j} - E(Z_{k,j})\}$ satisfies (1). However, $|E(Z_{k,j})|$ is bounded by

$$(9) \qquad \Sigma\left\{_{\nu \leqslant n\,:\, t_\nu \in G_{k,j}^{(n)}}\right\} E|X_\nu| \leqslant I\mu_k \leqslant I \cdot \max_{1 \leqslant k \leqslant \lambda} \mu_k$$

and similarly,

$$(10) \qquad P\left[\,|Z_{k,j}| \geqslant x\right] \leqslant \mu_k F(x/\mu_k).$$

Since there are only a finite number of μ_k's, $\{Z_{k,j}\}$ satisfies (1) and, as was noted earlier, this and (9) imply that $\{Z_{k,j} - E(Z_{k,j})\}$ satisfies (1).

Noting that expression (8) is nonnegative, the proof is completed by showing that its limit superior can be made arbitrarily small by choosing the partition fine enough. Fix $\delta > 0$. One can choose a sequence of random variables satisfying the assumptions in the lemma and a sequence of order preserving permutations so that a given $Z_{k,j}$ can be written in the same form as the V_N considered in the lemma with $N = \nu_{k,j}$. If $\nu_{k,j} < \mu_k$ then by adding random variables degenerate at zero $Z_{k,j}$ can be written in the same form as V_N with $N = \mu_k$. Applying the lemma, there exists an integer $N_0 = N_0(\delta, F)$ for which $E(Z_{k,j}/\mu_k) \leqslant \delta$ for each (k,j) with $\mu_k \geqslant N_0$. Expression (8) is bounded above by

$$(11) \qquad n^{-1}\Sigma_{\{k\,:\,\mu_k \geqslant N_0\}}\, \mu_k \Sigma_{j=1}^{\gamma_k(n)} E(Z_{k,j}/\mu_k)$$
$$+ n^{-1}\Sigma_{\{k\,:\,\mu_k < N_0\}} \Sigma_{j=1}^{\gamma_k(n)} E(Z_{k,j}).$$

The first expression in (11) is bounded by

$$\delta \Sigma_{\{k\,:\,\mu_k \geqslant N_0\}}\, \mu_k \gamma_k(n)/n \leqslant \delta \Sigma_{\{k\,:\,\mu_k \geqslant N_0\}}\, \mu_k(\max_\tau Q_{k,\tau} + c_n)$$
$$\leqslant \delta M + \delta c_n \pi_{i=1}^{\beta} \eta_i$$

and since $E(Z_{k,j}) \leqslant I\mu_k$, the second expression in (11) is bounded by

$$N_0 I \Sigma_{\{k\,:\,\mu_k < N_0\}} \gamma_k(n)/n \leqslant N_0 I \Sigma_{\{k\,:\,\mu_k < N_0\}} (\max_\tau Q_{k,\tau} + c_n)$$
$$\leqslant N_0 I Q\Big(\bigcup{}_{\{k,\tau\,:\,\mu_k < N_0\}} C_{k,\tau}\Big) + N_0 I c_n \operatorname{card}\{k : \mu_k < N_0\}.$$

Now $\lim_{\alpha \to \infty} \bigcup_{\{k,\tau\,:\,\mu_k < N_0\}} C_{k,\tau} = \varnothing$, for if $t = (t^{(i)}, t^{(2)}, \cdots t^{(\beta)}) \in \Delta$ then there is an α_0 for which $x^{(\alpha)}(\eta_i - N_0, i) < t^{(i)}$ for $i = 1, 2, \cdots, \beta$ and $\alpha \geqslant \alpha_0$. So for $\alpha \geqslant \alpha_0$, t is a member of a chain which has length greater than N_0. Select α large enough so that $Q(\bigcup_{\{k,\tau\,:\,\mu_k < N_0\}} C_{k,\tau}) < \delta/N_0$ and then the limit superior (as $n \to \infty$) of expression (8) is less than or equal to $\delta(M + I)$. The proof is completed.

Combining Theorems 3 and 5 we obtain

THEOREM 6. *If the sequence* $\{t_k\}$ *satisfies* (3) *and* (6), *then* $M_n/n \to 0$ *a.s.*

COROLLARY 7. *Let* $\{T_k\}$ *be a sequence of independent, identically distributed,* β*-dimensional random vectors defined on some probability space and let* P *denote their common probability distribution. If* $P = aP_d + (1 - a)P_{ac}$ *where* $0 \leqslant a \leqslant 1$, P_d *is discrete and* P_{ac} *is absolutely continuous, then* $\{T_k\}$ *satisfies* (6) *almost surely.*

PROOF. Choose $\Delta = R_\beta$. If $a > 0$, then for almost all ω in the underlying probability space, the set B_d is the support of P_d and if $a = 0$, then $B_d = \varnothing$ almost surely.

If $a = 1$, then (6) holds almost surely with $D = \varnothing$, $Q = P_d$, $c_n \equiv 0$, $M = 0$ and $\{\mathcal{C}^{(\alpha)}\}$ any sequence of partitions satisfying $\lim_{\alpha\to\infty} x^{(\alpha)}(\eta_i - j, i) = -\infty$ for each fixed j and i. Hence, suppose $a < 1$ and consider ε fixed with $0 < \varepsilon < 1 - a$. Choose γ_1 and γ_2 positive reals with $P([x_{i=1}^{\beta}[-\gamma_1, \gamma_1)]^c) \leqslant \varepsilon/2$ and $P_{ac}\{t : p(t) > \gamma_2\} \leqslant \varepsilon/2$ where p is the density of P_{ac}. Set $D = [x_{i=1}^{\beta}[-\gamma_1, \gamma_1)]^c \cup \{t : p(t) > \gamma_2\} - B_d$. Since $P(D) \leqslant \varepsilon$, the law of large numbers ensures that $\lim\sup_{n\to\infty} \nu_n(D)/n \leqslant \varepsilon$ almost surely. Set $Q(\cdot) = P(\cdot|(B_d \cup D)^c)$ and let $\{\mathcal{C}^{(\alpha)}\}$ be the sequence of partitions determined by $x(0, i) = \infty$; $x(j, i) = \gamma_1 + \alpha - j$ for $j = 1, 2, \cdots, \alpha$; $x(j, i) = \gamma_1 - 2(j - \alpha)\gamma_1/\alpha$ for $j = \alpha + 1, \alpha + 2, \cdots, 2\alpha$; $x(j, i) = -\gamma_1 - (j - 2\alpha)$ for $j = 2\alpha + 1, 2\alpha + 2, \cdots, 3\alpha - 1$; and $x(3\alpha, i) = -\infty$ for $i = 1, 2, \cdots, \beta$. If for some $i = 1, 2, \cdots, \beta$, $j(i) \leqslant \alpha$ or $j(i) \geqslant 2\alpha + 1$ then $I(j(1), j(2), \cdots, j(\beta)) \subset D \cup B_d$ and so $Q(I(j(1), j(2), \cdots, j(\beta))) = 0$. Otherwise,

$$Q(I(j(1), j(2), \cdots, j(\beta))) =$$
$$P_{ac}(I(j(1), j(2), \cdots, j(\beta)) \cap D^c)/P_{ac}(D^c) \leqslant \left(1 - \varepsilon(1 - a)^{-1}\right)^{-1}\gamma_2(2\gamma_1/\alpha)^{\beta}.$$

Recalling the definition of a chain, it is clear that $\lambda \leqslant \beta(3\alpha)^{\beta-1}$ and $\mu_k \leqslant 3\alpha$ for $k = 1, 2, \cdots, \lambda$. Hence, (6b) holds with $M = \beta(1 - \varepsilon(1 - a)^{-1})^{-1}\gamma_2(6\gamma_1)^{\beta}$.

In considering (6a), we note that $\nu_n(B_d \cup D)^c \to \infty$ almost surely and if $\hat{T}_i(\omega) = T_{j_i}(\omega)$, where j_i is the ith smallest of the positive integers k for which $T_k(\omega) \not\in B_d \cup D$, then $\{\hat{T}_i\}_{i=1}^{\infty}$ are independent and identically distributed as Q. Since Q is continuous, the law of the iterated logarithm for empirical distribution functions proved by Kiefer (1961) shows that there exists a constant c_β for which

$$\lim\sup_{n\to\infty}(2n/\ln\ln(n))^{\frac{1}{2}}\sup_{k,\tau,\alpha}\left|\operatorname{card}\{i \leqslant n : \hat{T}_i \in \mathcal{C}_{k,\tau}^{(\alpha)}\}/n - Q_{k,\tau}^{(\alpha)}\right| \leqslant c_\beta$$

almost surely. So for almost all ω there is a constant $c(\omega)$ such that (6a) holds with $c_n = c(\omega)c_\beta\{\ln\ln(\nu_n(B_d^c \cap D^c))/\nu_n(B_d^c \cap D^c)\}^{\frac{1}{2}}$. (Notice that in (6) c_n may depend on the sequence $\{t_k\}$.)

Corollaries 4 and 7 show that if the T_k are independent and identically distributed as P, with the continuous singular part of P identically zero, then $\{T_k\}$ satisfies (3) and (6). However, the assumption on the continuous singular part of P is only needed for (6b).

As we have seen, condition (6) excludes certain stochastic sequences of observation points which might be of interest and so we prove a theorem designed to treat some of these cases. Let $1 \leqslant \beta' \leqslant \beta$ and let $A \subset X_{i=\beta'}^{\beta}(a_i, b_i)$ with A linearly ordered with respect to the coordinate-wise ordering being considered here. If $\beta' > 1$ let $\Delta_{\beta'} = X_{i=1}^{\beta'-1}(a_i, b_i)$ and if $\beta' = 1$ we agree $\Delta_{\beta'} \times A = A$. For $\beta' > 1$ a sequence of partitions of $\Delta_{\beta'} \times A$ like the one used on Δ in Theorem 5 is needed.

Let $x(0, \beta') > x(1, \beta') > \cdots > x(\eta_{\beta'}, \beta')$ be elements of A with

$$A = U_{j=1}^{\eta_{\beta'}} I(j, \beta') \quad \text{where} \quad I(j, \beta') = \{x \in A : x(j, \beta') \leqslant x < x(j-1, \beta')\}$$

for $j = 1, 2, \cdots, \eta_{\beta'} - 1$ and $I(\eta_{\beta'}, \beta') = \{x \in A : x(\eta_{\beta'}, \beta') < x < x(\eta_{\beta'} - 1, \beta')\}$. Let $x(0, i) > x(1, i) > \cdots > x(\eta_i, i)$ partition (a_i, b_i) for $i = 1, \cdots \beta' - 1$ and with $j(i)$ an index for $i = 1, 2, \cdots \beta'$ with $j(i) = 1, 2, \cdots, \eta_i$, form the sets $I(j(i), i)$ as in the discussion before Theorem 5. Then set $I(j(1), j(2), \cdots, j(\beta')) = X_{i=1}^{\beta'} I(j(i), i)$. As was done previously, these sets can be grouped into chains, with the elements of the kth chain denoted by $C_{k,\tau}$ for $\tau = 1, 2, \cdots, \mu_k$ and $k = 1, 2, \cdots, \lambda$.

THEOREM 8. *Suppose that there is a positive integer $\beta' \leqslant \beta$, a linearly ordered set $A \subset X_{i=\beta'}^{\beta}(a_i, b_i)$, and if $\beta' > 1$ suppose that the sequence $\{t_k\}$ satisfies* (6) *with $\{\mathcal{C}^{(\alpha)}\}_{\alpha=1}^{\infty}$ a sequence of partitions of $\Delta_{\beta'} \times A$, then $M_n((\Delta_{\beta'} \times A) - B_d)/n \to 0$ a.s.*

PROOF. If $\beta' = 1$, then by ignoring those $t_k \in B_d$ the desired conclusion follows from the case $\beta = 1$ discussed in the introduction, and in fact $M_n(A)/n \to 0$ a.s. So we assume $\beta' > 1$ and observe that for U an upper layer in Δ, $(\Delta_{\beta'} \times A) \cap U$ is an upper layer in the space $\Delta_{\beta'} \times A$ with the partial order it inherits from R_β. The remainder of the proof is like the one given for Theorem 5.

It is clear that the cylinder considered in Theorem 8 need not have its base a subset of the product of (a_i, b_i) for $i = \beta', \beta' + 1, \cdots, \beta$ but the result has been stated this way since other cylinders could be transformed into such by relabelling. Also if $\Delta_{\beta'_i} \times A_i$ for $i = 1, 2, \cdots, m$ are disjoint cylinders with the hypothesis of Theorem 8 satisfied for each one, then $M_n(U_{i=1}^m(\Delta_{\beta'_i} \times A_i) - B_d)/n \to 0$ a.s. The following example illustrates this remark.

EXAMPLE. Let $\beta = 3$, let $\Delta = R_3$ and let $P = (P_1 + P_2)/2$ where P_1 is induced by $T = (T^{(1)}, T^{(2)}, T^{(3)})$ with $T^{(2)} \equiv 0$ and $(T^{(1)}, T^{(3)})$ absolutely continuous and P_2 is induced by T with $T^{(3)} \equiv 0$ and $(T^{(1)}, T^{(2)})$ absolutely continuous. Set $A_1 = \{(0, x) : x \in R\}$ and $A_2 = \{(x, 0) : x \neq 0\}$ and note that $R \times A_1$ and $R \times A_2$ are disjoint. If the T_k are independent and identically distributed as P, then for almost all ω, the sequence $\{t_k = T_k(\omega)\}$ satisfies the hypothesis of Theorem 8 for both of the cylinders $R \times A_i$, has $B_d = \emptyset$, and is contained in $(R \times A_1) \cup (R \times A_2)$. To obtain the desired sequence of partitions of $R \times A_2$, one should first partition R_2 as in the proof of Corollary 7 using the marginal distribution of $(T^{(1)}, T^{(2)})$ under P_2 instead of the P considered there and then for each rectangle $I(j(1), j(2))$ obtained, construct a new rectangle $I(j(1), j(2)) \times \{0\}$. These new rectangles provide the desired partition of $R \times A_2$ for α odd. (If α is odd then none of the partitions of the sides has zero as an endpoint. This is important since $(0, 0) \notin A_2$.) The desired partition of $R \times A_1$ is obtained similarly. So for such sequences $\{t_k\}$, $M_n = M_n(U_{i=1}^2(R \times A_i) - B_d)$ and we have $M_n/n \to 0$ a.s.

COROLLARY 9. *Suppose that the sequence* $\{t_k\}$ *satisfies* (3) *and the following*:

(12) *there is a positive integer* $\beta' \leqslant \beta$; *there is a sequence of disjoint, linearly ordered sets* $A_i \subset X_{i=\beta'}^{\beta}(a_i, b_i)$ *with* $\liminf_{n\to\infty} \nu_n(B_d \cup \bigcup_{i=1}^{\infty}(\Delta_{\beta'} \times A_i))/n = 1$; *for each* $\delta > 0$, *there is an integer* $m = m(\delta)$ *with* $\limsup_{n\to\infty} \nu_n(\bigcup_{i=m}^{\infty}(\Delta_{\beta'} \times A_i) - B_d)/n \leqslant \delta$; *and if* $\beta' > 1$, *there is for each* $i = 1, 2, \cdots$ *a sequence of partitions of* $\Delta_{\beta'} \times A_i$, $\{\mathcal{C}_i^{(\alpha)}\}_{\alpha=1}^{\infty}$, *which satisfies* (6).

Then, $M_n/n \to 0$ *a.s.*

PROOF. Because of assumption (12), Proposition 2 and Theorem 3 it suffices to show that $M_n(\bigcup_{i=1}^{m}(\Delta_{\beta'} \times A_i) - B_d)/n \to 0$ a.s. for $m = 1, 2, \cdots$. This is exactly what was shown in the discussion after Theorem 8.

COROLLARY 10. *If the* T_k *are independent and identically distributed and if* $T_1' = (T_1^{(\beta'+1)}, \cdots, T_1^{(\beta)})$ *is discrete for some* $1 \leqslant \beta' < \beta$ *and* $(T_1^{(1)}, \cdots, T_1^{(\beta')})$ *given* T_1' *is absolutely continuous, then* $\{T_k\}$ *almost surely satisfies* (3) *and* (12) *with* $B_d = \varnothing$.

PROOF. Let $(a_i, b_i) = R$ for $i = 1, 2, \cdots, \beta$. Since the discrete part of the probability induced by T_1 vanishes, $B_d = \varnothing$ almost surely and consequently (3) holds almost surely. Let $s_1, s_2, \cdots$ denote the distinct values of T_1' which occur with positive probability and let $A_i = R \times \{s_i\}$. The A_i are disjoint, linearly ordered sets and by the strong law of large numbers $\nu_n(\bigcup_i(\Delta_{\beta'} \times A_i))/n \to 1$ a.s. (as before, we agree that $\Delta_{\beta'} \times A_i = A_i$ if $\beta' = 1$). For each $\delta > 0$, let $m = m(\delta)$ be chosen so that $P_{T_1'}(\{s_m, s_{m+1}, \cdots\}) \leqslant \delta$ where $P_{T_1'}$ is the probability induced by T_1', then $\nu_n(\bigcup_{i=m}^{\infty}(\Delta_{\beta'} \times A_i))/n \to P_{T_1'}(\{s_m, s_{m+1}, \cdots\})$ a.s. With i fixed, the desired sequence of partitions is constructed as in the example following Theorem 8. First $R_{\beta'}$ is partitioned as in the proof of Corollary 7 with the distribution of $(T_1^{(1)}, \cdots, T_1^{(\beta')})$ given T_1' used rather than P and with $\hat{T}_k(\omega)$ the kth $T_j(\omega)$ with $T_j' = s_i$ and $(T_j^{(1)}, \cdots, T_j^{(\beta')}) \notin D$. The desired partition of $\Delta_{\beta'} \times \{s_i\}$ is obtained by taking the cartesian product of each rectangle in the partition above with $\{s_i\}$.

We now wish to consider the weak rate of convergence of M_n/n, that is, the rate at which $P[M_n/n \geqslant \varepsilon]$ converges to zero. The rates that will be established are generalizations of the following known results.

THEOREM 11. *If the sequence* $\{X_k\}$ *satisfies*

(13) $$\lim_{y\to\infty} y^r F(y) = 0$$

for some $r > 1$, *then for each* $\varepsilon > 0$

$$P\big[|n^{-1}\Sigma_{k=1}^{n} X_k| \geqslant \varepsilon\big] = o(n^{-r+1}).$$

PROOF. This is a special case of Theorem 2 of Franck and Hanson (1966) and was proved in the identically distributed case by Baum and Katz (1965).

THEOREM 12. *If the sequence* $\{X_k\}$ *satisfies*

(14) $$F(y) \leqslant 0(\exp(-cy))$$

for some $c > 0$, *then for each* $\varepsilon > 0$ *there exists a* $\rho < 1$ *with*

$$P\left[\left|n^{-1}\Sigma_{k=1}^{n} X_k\right| \geqslant \varepsilon\right] \leqslant 2\rho^n.$$

PROOF. This is a special case of Theorem A of Hanson (1967) and was proved in the identically distributed case by Cramér (1938).

THEOREM 13. *Suppose the sequence* $\{t_k\}$ *satisfies* (3) *and either* (6) *or* (12). *If the sequence* $\{X_k\}$ *satisfies* (13) *for some* $r > 1$, *then for each* $\varepsilon > 0$

$$P\left[M_n/n \geqslant \varepsilon\right] = o(n^{-r+1}).$$

If, in addition, the sequence $\{X_k\}$ *satisfies* (14) *for some* $c > 0$, *then for each* $\varepsilon > 0$ *there exists a* $\rho < 1$ *and a* $c^* > 0$ *for which*

$$P\left[M_n/n \geqslant \varepsilon\right] \leqslant c^*\rho^n.$$

PROOF. It suffices to establish the desired rates for $P[M_n(B_d)/n \geqslant \varepsilon/2]$ and $P[M_n(\Delta - B_d)/n \geqslant \varepsilon/2]$. The first step is to establish an analogue to Proposition 2.

Suppose that for each $\delta > 0$ there is a set B with $\lim\sup_{n\to\infty} \nu_n(B)/n \leqslant \delta$. Observe that (14) implies (13) which in turn implies (1) and so, in either case being considered, $E(X_k^+) \leqslant I < \infty$. Furthermore, if $\{X_k\}$ satisfies (13) then so does $\{X_k^+ - E(X_k^+)\}$ and if $\{X_k\}$ satisfies (14) then so does $\{X_k^+ - E(X_k^+)\}$. For future reference we note that, if $\{X_k\}$ satisfies (13) ((14)) and $\{l_k\}$ is bounded then $\{X_k + l_k\}$ also satisfies (13) ((14)). For $A \subset \Delta$, $M_n(A \cap B)/n$ is bounded above by

$$\Sigma_k a_{n,k} X_k^+ \leqslant \Sigma_k a_{n,k}(X_k^+ - E(X_k^+)) + I\nu_n(B)/n$$

where $a_{n,k} = 1/n$ if $k \leqslant n$ and $t_k \in B$ and $a_{n,k} = 0$ otherwise. For n sufficiently large, $I\nu_n(B)/n \leqslant \varepsilon/2$. If (13) holds then Theorem 2 of Franck and Hanson (1966) shows that $P[\Sigma_k a_{n,k}(X_k^+ - E(X_k^+)) \geqslant \varepsilon/2] = o(n^{-r+1})$ for each $\varepsilon > 0$, and if (14) holds then Theorem A of Hanson (1967) shows that for each $\varepsilon > 0$ there exists a $\rho < 1$, for which $P[\Sigma_k a_{n,k}(X_k^+ - E(X_k^+)) \geqslant \varepsilon/2] \leqslant \rho^n$. Hence, if $A \subset \Delta$ and if for each $\delta > 0$ there exists a set $B \subset \Delta$ with $\lim\sup_{n\to\infty} \nu_n(B)/n \leqslant \delta$, then for each $\varepsilon > 0$, $P[M_n(A)/n \geqslant \varepsilon]$ converges to zero at the desired rate if $P[M_n(A \cap B^c)/n \geqslant \varepsilon]$ does for each $\varepsilon > 0$ and each such B.

Next $P[M_n(B_d)/n \geqslant \varepsilon]$ is shown to converge at the desired rate for each $\varepsilon > 0$. The proof of this part is very similar to the proof of Theorem 3. If card $(B_d) < \infty$, then the desired conclusion follows by applying either Theorem 11 or 12 at each point at B_d. So we suppose that card $(B_d) = \infty$. It is sufficient to show that for each $\varepsilon > 0$, $P[V_n/n \geqslant \varepsilon]$ and $P[W_n/n \geqslant 2\varepsilon]$ converge at the specified rates. It has been shown that $V_n/n \leqslant S_n^*(B_d \cap B_{M^*,n})/n + o(1)$ and so for n sufficiently large we consider $P[S_n^*(B_d \cap B_{M^*,n})/n \geqslant \varepsilon/2]$. Since $S_n^*(B_d \cap B_{M^*,n})/n = \Sigma_k a_{n,k}(X_k^+ - E(X_k^+))$ with $a_{n,k} = 1/n$ if $t_k \in B_d \cap B_{M^*,n}$ and $a_{n,k} = 0$ otherwise, an argument like the one at the first of this proof shows that $P[V_n/n \geqslant \varepsilon]$ converges at the desired rate.

Since W_n/n is bounded above by the sum of (4) and (5) and since (5) is bounded by $\Sigma_{\{i\,:\,s_i \in B^c_{M^*,n}\}}\Sigma_{\alpha=\alpha_n^{(i)}+1}^{\delta_n^{(i)}}(X^+_{j_\alpha^{(i)}} - E(X^+_{j_\alpha^{(i)}}))/n + \varepsilon/2$ we need to show that

$$P\left[\Sigma_{\{i\,:\,s_i \in B^c_{M^*,n}\}}\Sigma_{\alpha=1}^{\alpha_n^{(i)}}\left(Z_\alpha^{(i)} - E\left(Z_\alpha^{(i)}\right)\right)/n \geqslant \varepsilon/(2M)\right]$$

and

$$P\left[\Sigma_{\{i\,:\,s_i \in B^c_{M^*,n}\}}\Sigma_{\alpha=\alpha_n^{(i)}+1}^{\delta_n^{(i)}}\left(X^+_{j_\alpha^{(i)}} - E\left(X^+_{j_\alpha^{(i)}}\right)\right)/n \geqslant \varepsilon/2\right]$$

converge to zero properly. These conclusions follow from Theorem 2 of Franck and Hanson (1966) and Theorem A of Hanson (1967) provided $\{Z_\alpha^{(i)} - E(Z_\alpha^{(i)})\}$ and $\{X_k^+ - E(X_k^+)\}$ satisfy (13) or (14), depending on which conclusion is desired. It has already been shown that the needed assumption holds for the latter. Since $P[|Z_\alpha^{(i)}| \geqslant y] \leqslant M \cdot F(y)$ and $E(Z_\alpha^{(i)}) < \varepsilon/2$ the same is true for $\{Z_\alpha^{(i)} - E(Z_\alpha^{(i)})\}$.

Finally, we consider the rate of convergence of $P[M_n(\Delta - B_d)/n \geqslant \varepsilon]$ assuming (6) or (12). If we assume (6) the argument is like that given for Theorem 5 and if we assume (12) the argument is like that given for Corollary 9. Since these two arguments are very similar we give only the former. Clearly, it suffices to show that $P[M_n(\Delta - (B_d \cup D))/n \geqslant \varepsilon]$ converges at the appropriate rate for each $\varepsilon > 0$ and for an arbitrary D as specified in (6). Again the proof is trivial if there are only a finite number of k for which $t_k \in \Delta - (B_d \cup D)$ because for each k, $P[|X_k|/n \geqslant \varepsilon]$ has the appropriate rate of convergence. So we assume there are an infinite number and denote them by $\{t_k\}$. Choose a partition and group the elements $t_1, t_2, \cdots, t_n$ into strings as in the proof of Theorem 5. For $\varepsilon > 0$ fixed, choose a partition so that for n sufficiently large (8) is less than or equal to $\varepsilon/2$. Then we need only show that $P[|n^{-1}\Sigma_{k,j}^{(n)}(Z_{k,j} - E(Z_{k,j}))| \geqslant \varepsilon/2]$ converges at the desired rate. However, (9) shows that $E(Z_{k,j})$ is bounded and (10) shows that $\{Z_{k,j}\}$ satisfies (13) ((14)) provided $\{X_k\}$ satisfies (13) ((14)). Again appealing to Theorem 2 of Franck and Hanson (1966) and Theorem A of Hanson (1967), the proof is completed.

3. Consistency results. Suppose that $\mu(t)$ is an isotone regression function with domain $\Delta \subset R_\beta$, that $w(t)$ is a positive weight function and that $\hat{\mu}_n(t)$ is the estimator based on the first n observations, $Y_1, Y_2, \cdots, Y_n$, described in Section 1 (cf. (2)). Assume that there exist constants w_1 and w_2 such that

(15) $$0 < w_1 \leqslant w(t) \leqslant w_2 < \infty \qquad \text{for each} \quad t \in \Delta.$$

THEOREM 14. *Let* $\mu(\cdot)$ *be continuous, let* $\{t_k\}$ *satisfy* (3) *and* (6), *and with* k_0 *a positive integer suppose that*

(16) $\lim \inf_{n\to\infty} \nu_n(J)/n > 0$ *for each nondegenerate rectangle* J *which contains* t_{k_0}.

If $\{Y_k - \mu(t_k)\}$ *satisfies* (1) *then* $\hat{\mu}(t_{k_0}) \to \mu(t_{k_0})$ *a.s.*; *if* $\{Y_k - \mu(t_k)\}$ *satisfies* (13) *for some* $r > 1$, *then for each* $\varepsilon > 0$

$$P\left[\left|\hat{\mu}_n(t_{k_0}) - \mu(t_{k_0})\right| \geqslant \varepsilon\right] = o(n^{-r+1});$$

and if $\{Y_k - \mu(t_k)\}$ *satisfies* (14) *for some* $c > 0$, *then for each* $\varepsilon > 0$ *there exist* *positive constants* c^* *and* $\rho < 1$ *for which*

$$P\big[|\hat{\mu}_n(t_{k_0}) - \mu(t_{k_0})| \geqslant \varepsilon\big] \leqslant c^*\rho^n.$$

PROOF. Consider $\varepsilon > 0$ and choose $t_0 \in \Delta$ with $t_0^{(i)} > t_{k_0}^{(i)}$ for $i = 1, 2, \cdots, \beta$ and $\mu(t_0) - \mu(t_{k_0}) < \varepsilon/3$. Let $U^*(t_0)$ denote the upper layer which is the complement of $\{s \in \Delta : s \ll t_0\}$. Then

$$\hat{\mu}_n(t_{k_0}) - \mu(t_{k_0}) \leqslant \max_{\{U : t_{k_0} \in U\}} Av_n(U - U^*(t_0)) - \mu(t_{k_0})$$
$$\leqslant \max_{\{U : t_{k_0} \in U\}} Av_n(U - U^*(t_0)) - \mu(t_0) + \varepsilon/3.$$

Letting $X_k = w(t_k)(Y_k - \mu(t_k))$ and $D^* = \{s : t_{k_0} \ll s \ll t_0\}$, the above expression is bounded above by

$$w_1^{-1}\max_{\{U : t_{k_0} \in U\}} |S_n(U - U^*(t_0))|/\nu_n(D^*) + \varepsilon/3$$

and by assumption $\nu_n(D^*) \geqslant \delta n$ for some $\delta > 0$ and for sufficiently large n. Hence, for sufficiently large n, the latter expression is bounded by

$$(w_1\delta)^{-1}\big[\max_{\{U : t_{k_0} \in U\}} |S_n(U \cup U^*(t_0))|/n + |S_n(U^*(t_0))|/n\big] + \varepsilon/3$$
$$\leqslant (w_1\delta)^{-1}\big[\max(M_n, M_n^{(-)})/n + |S_n(U^*(t_0))|/n\big] + \varepsilon/3,$$

where $M_n^{(-)}$ is defined as M_n with X_k replaced by $-X_k$. Because of (15), $\{X_k\}$ and $\{-X_k\}$ satisfy (1), (13) or (14) provided $\{Y_k - \mu(t_k)\}$ satisfies (1), (13) or (14) respectively. Theorem 1, Theorem 2 of Franck and Hanson (1966) and Theorem A of Hanson (1967) show that $S_n(U^*(t_0))/n \to 0$ a.s. and that $P[|S_n(U^*(t_0))|/n \geqslant \varepsilon]$ converges at the specified rate. Applying Theorems 6 and 13 to M_n and $M_n^{(-)}$ and noting that $\mu(t_{k_0}) - \hat{\mu}_n(t_{k_0})$ can be bounded similarly completes the proof.

There are several comments that need to be made concerning Theorem 14.

COMMENT 1. While condition (16) was not assumed in establishing the convergence of M_n/n, Hanson et al. (1973) have shown that a condition on the placement of the t_k's is needed if $\hat{\mu}_n$ is to be strongly consistent. (See their Theorem 2 and the discussion after Theorem 6.) Condition (16) requires that, in the limit, the rectangles containing t_{k_0} contain a positive proportion of the observation points.

Suppose $t_k = T_k(\omega)$ and the T_k are independent and identically distributed as P. If $\beta = 1$, (16) holds almost surely for each k_0. This follows from the Glivenko-Cantelli theorem and the fact that

$$P\{x : P[x, x + \varepsilon) = 0 \text{ or } P(x - \varepsilon, x] = 0 \text{ for some } \varepsilon > 0\} = 0.$$

Since Hanson et al. (1973) have shown that assumptions (3) and (6) can be omitted in Theorem 6 if $\beta = 1$, condition (16) is the only assumption that needs to be made concerning the t_k's in order for $\hat{\mu}_n(t_{k_0})$ to be strongly consistent in this case. If $\beta \geqslant 2$ some restrictions must be placed on P for (16) to hold almost surely. For if $\beta = 2$ and P is continuous with probability concentrated on the negative 45° line, then for almost all ω, any k_0 and any rectangle with $T_{k_0}(\omega)$ as the upper or lower vertex there is only one observation point in the rectangle. We will discuss a version of Theorem 14 for the continuous singular case later. If the continuous singular part of P vanishes and if the boundary of the support of the absolutely continuous part of P has Lebesgue measure zero, then (16) holds almost surely for each k_0. To see this, let $(P)_{ac}$ denote the absolutely continuous part of P, let A_0 be the union of the interior of the support of $(P)_{ac}$ and the support of the discrete part of P, and let $\{G_i\}_{i=1}^{\infty}$ be a countable basis for R_β. Set

$$N_1 = \bigcup_{k=1}^{\infty}\{\omega : T_k(\omega) \notin A_0\}$$

$$N_2 = \bigcup_{i=1}^{\infty}\{\omega : \text{card}\{k \leqslant n : T_k(\omega) \in G_i\}/n \nrightarrow P(G_i)\} \qquad \text{and}$$

$$N_3 = \bigcup_{\{t : P(\{t\})>0\}}\{\omega : \text{card}\{k \leqslant n : T_k(\omega) = t\}/n \nrightarrow P(\{t\})\}.$$

If t is in the interior of the support of $(P)_{ac}$ and J is a nondegenerate rectangle containing t, then there is a G_i contained in J and in the support of $(P)_{ac}$ with $P(G_i) > 0$. Hence, (16) holds for each k_0 if $\omega \in (N_1 \cup N_2 \cup N_3)^c$.

COMMENT 2. For $\beta \geqslant 2$ the strong consistency results given here improve on those in the literature in that they have weaker moment assumptions and the conditions imposed on the sequence $\{t_k\}$ are satisfied almost surely by a larger class of stochastic observation points. For $\beta \geqslant 2$ the results which yield strong consistency in Hanson et al. (1973), Robertson and Wright (1975) and Wright (1976) require at least finite second moments in the identically distributed case, but the results given here only require finite first moments in this case. In these three papers, the conditions imposed on $\{t_k\}$ essentially place a uniform upper bound on the density of the observation points. In the case of stochastic observation points, this excludes probabilities whose discrete part does not vanish or whose absolutely continuous part has an unbounded density.

COMMENT 3. For $-\infty \leqslant a_i < a_i^* < b_i^* < b_i \leqslant \infty$ the convergence of $\sup_{t \in \times_{i=1}^{\beta}[a_i^*, b_i^*]} |\hat{\mu}_n(t) - \mu(t)|$ has been investigated in the three papers mentioned in Comment 2. The same techniques could be applied here to obtain these more global types of results.

COMMENT 4. In Theorem 14 the continuity assumption can be omitted for points in B_d. The proof given for Theorem 14 with $t_0 = t_{k_0}$ establishes the following result: if $\liminf_{n\to\infty} \nu_n(\{t_{k_0}\})/n > 0$, if (3) and (6) hold and if $\{Y_n - \mu(t_n)\}$ satisfies (1), then $\hat{\mu}_n(t_{k_0}) \to \mu(t_{k_0})$ a.s. If $\{Y_n - \mu(t_n)\}$ satisfies (13) or (14) then the rates of convergence given in Theorem 14 also hold. Furthermore, if the $\{t_k\}$ are generated by a discrete probability P, the results just stated apply at each point in the support of P.

COMMENT 5. Consistency results can also be obtained in the case discussed in Corollary 9. Suppose that $\{t_k\}$ satisfies (3) and (12) and that $\{Y_n - \mu(t_n)\}$ satisfies (1). We first consider points in B_d. If $\lim \inf_{n\to\infty}\nu_n(\{t_{k_0}\})/n > 0$, then $\hat{\mu}_n(t_{k_0}) \to \mu(t_{k_0})$ a.s. This is essentially the same observation as was made in Comment 4, except there (3) and (6) were assumed to obtain $M_n/n \to 0$ and here (3) and (12) serve that purpose. Next we consider points in $\Delta_{\beta'} \times A_i$ for one of the linearly ordered sets A_i. If $(t_{k_0}^{(\beta')}, \cdots, t_{k_0}^{(\beta)}) \in A_i$ is the limit of a sequence of points in A_i converging from above, as well as the limit of a sequence of points in A_i converging from below; if μ restricted to $\Delta_{\beta'} \times A_i$ is continuous; and if $\lim \inf_{n\to\infty}\nu_n(J \times [a, b])/n > 0$ for each nondegenerate rectangle $J \subset \Delta_{\beta'}$ and for each $a, b \in A_i$ with $a < b$, then $\hat{\mu}_n(t_{k_0}) \to \mu(t_{k_0})$ a.s. If (13) or (14) is assumed in place of (1) the rates of convergence obtained in Theorem 14 also hold in this case. The proof is like the one given for Theorem 14 except that t_0 is chosen so that $(t_0^{(\beta')}, \cdots, t^{(\beta)}) \in A_i$.

COMMENT 6. In the case $\beta = 1$, Makowski (1973) established a law of the iterated logarithm for M_n and applied it to obtain a strong rate of convergence for $\hat{\mu}_n(t) - \mu(t)$. This strong rate established by Makowski (1973) states that with probability one, $\lim \sup_{n\to\infty}(n/\ln\ln(n))^{\frac{1}{4}}|\hat{\mu}_n(t) - \mu(t)| \leqslant c$ for some constant c. Can similar results be obtained for $\beta > 1$?

Acknowledgment. The author is indebted to the referee and the Associate Editor for their helpful comments. The proof of the lemma and part of Comment 1 are due to the referee.

REFERENCES

[1] BAUM, L. E. and KATZ, MELVIN (1965). Convergence rates in the law of large numbers. *Trans. Amer. Math. Soc.* **120** 108–123.

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

[3] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference.* (Madan Lal Puri, Ed.,) 177–195. Cambridge Univ. Press.

[4] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilitiés. *Actualities Sci. Indus.* No. 736.

[5] FRANCK, W. E. and HANSON, D. L. (1966). Some results giving rates of convergence in the law of large numbers for weighted sums of independent random variables. *Trans. Amer. Math. Soc.* **124** 347–359.

[6] GABRIEL, JEAN-PIERRE (1974). Loi des grands nombres, séries et martingales à deux indices. *C.R. Acad. Sci. Paris Ser. A* **279** 169–171.

[7] HANSON, D. L. (1967). Some results relating moment generating functions and convergence rates in the law of large numbers. *Ann. Math. Statist.* **38** 742–750.

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

[9] KIEFFER, 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.

[10] MAKOWSKI, G. G. (1973). Laws of the iterated logarithm for permuted random variables and regression applications. *Ann. Statist.* **1** 872–887.

[11] MAKOWSKI, G. G. (1975). A rate of convergence for a nondecreasing regression estimator. *Bull. Inst. Math. Acad. Sinica* **3** 61–64.

[12] ROBERTSON, TIM and WRIGHT, F. T. (1975). Consistency in generalized isotonic regression. *Ann. Statist.* **3** 350–362.

[13] SMYTHE, R. T. (1973). Strong laws of large numbers for r-dimensional arrays of random variables. *Ann. Probability* **1** 164–170.

[14] SMYTHE, R. T. (1974). Sums of independent random variables on partially ordered sets. *Ann. Probability* **2** 906–917.

[15] WRIGHT, F. T. (1976). Algebraic rates of convergence for isotonic regression functions. *Bull. Inst. Math. Acad. Sinica* **4** 281–287.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MISSOURI
ROLLA, MISSOURI 65401