The Annals of Probability
1981, Vol. 9, No. 2, 323–329

THE EMPIRICAL DISCREPANCY OVER LOWER LAYERS AND A RELATED LAW OF LARGE NUMBERS[1]

By F. T. Wright

University of Missouri-Rolla

Let $\{X_k\}$ be a sequence of independent random variables which are centered at their means; let $\{T_k\}$ be an i.i.d. sequence of β-dimensional random vectors with common distribution μ; and let $\{X_k\}$ and $\{T_k\}$ be independent. With $\mathscr{L}$ the collection of lower layers, a necessary and sufficient condition for the almost sure convergence of $\sup_{L\in\mathscr{L}}|\sum_{k=1}^n \chi_L(T_k)/n - \mu(L)|$ to zero is given. In addition, this condition on μ is shown to imply that $\sup_{L\in\mathscr{L}}|\sum_{k=1}^n X_k\chi_L(T_k)|/n \to 0$ a.s. provided the X_k satisfy a first moment-like condition. Rates of convergence are also investigated.

1. Introduction. Let $\{t_k\}$ be a sequence of points in R_β, the β-dimensional reals; let $\{X_k\}$ be a sequence of independent random variables (We think of X_k as being associated with t_k for $k = 1, 2, \ldots$); let $\mathscr{A}$ be a nonempty collection of subsets of R_β; let $\chi_A(\cdot)$ denote the indicator of the set A; and let

$$S_n(A) = \textstyle\sum_{k=1}^n X_k\chi_A(t_k), \quad \text{for} \quad A \subset R_\beta \quad \text{and} \quad M_n^{\mathscr{A}} = \sup_{A\in\mathscr{A}}|S_n(A)|/n.$$

The collection $\mathscr{A}$ of primary interest here is $\mathscr{L}$, the collection of lower layers. A Borel set L is a lower layer provided $x \in L$ and $y \le x$ imply that $y \in L$, where $\le$ is the usual coordinate-wise ordering on R_β. This choice of $\mathscr{A}$ is motivated by the study of the consistency properties of an isotone regression estimator (see Wright (1979)). We will consider conditions on $\{t_k\}$ and $\{X_k\}$ which imply that $M_n^{\mathscr{L}} \to 0$ a.s. We express the conditions on $\{X_k\}$ in terms of $F(y) \equiv \sup_k P[|X_k| \ge y]$. It is well known that $\sum_{k=1}^n X_k/n \to 0$ a.s. provided

$$E(X_k) = 0 \text{ for } k = 1, 2, \ldots, F(y) \to 0 \text{ as } y \to \infty \text{ and } \int_0^\infty y|dF(y)| < \infty. \tag{1}$$

For the case $\beta = 1$, Brunk (1958) proved that $M_n^{\mathscr{L}} \to 0$ a.s. for any sequence $\{t_k\}$ provided the random variables $\{X_k\}$ have zero means and satisfy the rth order Kolmogorov condition for some $r \ge 1$. Hanson, et al. (1973) have shown that these assumptions on the sequence $\{X_k\}$ can be replaced by (1). For $\beta \ge 2$, Wright (1979) gave an example to show that some conditions must be imposed on $\{t_k\}$ if $M_n^{\mathscr{L}}$ is to converge to zero. The conditions studied in that paper are most easily understood in the case in which $\{t_k\}$ is the realization of a random sequence. So we assume throughout this note that $\{T_k\}$ is an i.i.d. sequence of β-dimensional random vectors with $\{X_k\}$ and $\{T_k\}$ independent. Denoting the common distribution of the T_k by μ, he has shown that $M_n^{\mathscr{L}} \to 0$ a.s. provided $\{X_k\}$ satisfies (1) and the continuous singular part of μ vanishes. Smythe (1980) extended this result and gave a very nice proof using a theorem due to Steele (1978) concerning the empirical discrepancy. Smythe has shown that the condition on μ can be weakened to assuming that μ_c, the continuous part of μ, does not charge the boundary of any lower layer. (Brunk, et al. (1956) have shown that the Lebesgue measure of the boundary of a lower layer is zero.) This

Received March 21, 1979; revised September 1979.

[1] Supported in part by the National Science Foundation under Grant Number MCS 78-01514.

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

Key words and phrases. Glivenko-Cantelli theorem, lower layers, isotone regression, maxima of partial sums.

result is clearly not optimal since, in the case $\beta = 2$, the distribution μ which is uniform on $\{0\} \times (0, 1)$ assigns positive probability to the boundary of the lower layer $(-\infty, 0] \times R$, but the desired result should hold since $\{0\} \times (0, 1)$ is linearly ordered. In fact, Theorem 8 of Wright (1979) shows that $M_n^{\mathscr{L}} \to 0$ a.s. in this situation if $\{X_k\}$ satisfies (1).

By modifying the work of Dehardt (1971), we obtain a sufficient condition for this convergence which is optimal for these random $\{T_k\}$. In particular, it is shown that the condition on μ can be weakened to assuming that neither μ nor any of its marginal probabilities have continuous parts which charge any strictly decreasing graph. This condition on μ is also shown to be necessary and sufficient for the convergence to zero of the empirical discrepancy over lower layers. The rate of convergence of $P[M_n^{\mathscr{L}} \geq \epsilon]$ is also investigated.

2. Results. The empirical discrepancy of $T_1, T_2, \ldots, T_n$ and μ over $\mathscr{A}$ is defined by

$$D_n^{\mathscr{A}} = \sup_{A \in \mathscr{A}} | \mu_n(A) - \mu(A) |, \tag{2}$$

where $\mu_n(A) = \sum_{k=1}^n \chi_A(T_k)/n$. Examining the proof given for Theorem 1.1 of Smythe (1980), it is clear that for $\{X_k\}$ satisfying (1), $D_n^{\mathscr{A}} \to 0$ a.s. implies that $M_n^{\mathscr{A}} \to 0$ a.s. (For example, if $\mathscr{C}$ denotes the convex sets in R_β, then this observation combined with Theorem 6.1 of Steele (1978) shows that $M_n^{\mathscr{C}} \to 0$ a.s. if $\sup_{A \in C} \mu_c(\partial A) = 0$.) So we wish to find an optimal condition on μ which implies $D_n^{\mathscr{L}} \to 0$ a.s. This will be accomplished by establishing an analogue to Dehardt's generalization of the Glivenko-Cantelli Theorem.

Let $\mathscr{M}$ be a collection of real valued, Borel functions defined on R_β and set

$$I_n^{\mathscr{M}} = \sup_{f \in \mathscr{M}} \left| \int f \, d\mu_n - \int f \, d\mu \right|.$$

With $\mathscr{M}$ the collection of all functions which are uniformly bounded by a fixed constant and are monotone in each variable separately (such functions may be nondecreasing in some of the variables and nonincreasing in the others), Theorem 2 of Dehardt (1971) states that $I_n^{\mathscr{M}} \to 0$ a.s. if and only if μ_c does not charge any strictly monotone graph in R_β. There is a slight error in this result, but it can easily be corrected. A Borel set B in R_k is said to be a decreasing graph (d.g.) provided $B \cap (-\infty, b] \subset \partial(-\infty, b]$ for each $b \in B$. (Here $(-\infty, b]$ denotes the set $\{t: t \in R_k, t \leq b\}$.) Furthermore B is said to be a strictly decreasing graph (s.d.g.) provided $B \cap (-\infty, b] = \{b\}$ for each $b \in B$. Of course, these really are decreasing or strictly decreasing graphs with respect to the ordering $s \leq t$ provided $s_i \leq t_i$ for $i = 1, 2, \ldots, k$. A monotone graph (strictly monotone graph) is a decreasing graph (strictly decreasing graph) with respect to this coordinate-wise partial order or one of the other $2^k - 1$ partial orders obtained by reversing one or more of the inequalities $s_i \leq t_i$.

For $1 \leq k \leq \beta$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq \beta$, let $\mu_{i_1, i_2, \ldots, i_k}$ denote the distribution of $(T_1^{(i_1)}, T_1^{(i_2)}, \ldots, T_1^{(i_k)})$. It can be shown that $I_n^{\mathscr{M}} \to 0$ a.s. if and only if for each $k \in \{1, 2, \ldots, \beta\}$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq \beta$, $(\mu_{i_1, i_2, \ldots, i_k})_c$ does not charge any s.m.g. in R_k. Dehardt (1970) has shown that $I_n^{\mathscr{M}} \to 0$ a.s. implies that μ_c does not charge any s.m.g. in R_β. However, if f is a monotone function defined on R_k, then $f^*(t_1, t_2, \ldots, t_\beta) = f(t_{i_1}, t_{i_2}, \ldots, t_{i_k})$ is a monotone function on R_β and so it follows that this new condition is also necessary. (The needed modifications of Dehardt's proof of sufficiency are given in the proof of Theorem 2.) We now give an example to show that this new condition is in fact stronger than the original one.

EXAMPLE. Let $\beta = 3$, let $T_1^{(1)}$ and $T_1^{(3)}$ be independent with $T_1^{(1)}$ a continuous random variable and let $T_1^{(2)} \equiv -T_1^{(1)}$. Now μ_{12} is continuous and $\mu_{12}(B_0) = 1$ where $B_0 = \{(t_1, t_2): t_1 = -t_2\}$ is a s.d.g. in R_2. Furthermore, any s.m.g. contained in $B_0 \times R$, the support of μ, has no more than one point with a given first coordinate. Consequently, μ does not charge any s.m.g.

To show that $D_n^{\mathscr{L}}$ converges to zero we need only consider the indicator functions of lower layers, which are nonincreasing in each variable. So throughout the remainder of this note we take $\mathscr{M}$ to be the collection of all functions which are uniformly bounded by a fixed constant and are nonincreasing in each variable. It would seem reasonable to expect that $I_n^{\mathscr{M}} \to 0$ a.s. if and only if

(3) for each $k \in \{1, 2, \ldots, \beta\}$ and $1 \le i_1 < i_2 < \cdots < i_k \le \beta$,

$$(\mu_{i_1, i_2, \ldots, i_k})_c \text{ does not charge any s.d.g. in } R_k.$$

This is in fact the case. It is interesting to note that the proof Dehardt gives for the necessity of his condition shows that $D_n^{\mathscr{L}} \to 0$ a.s. (and consequently $I_n^{\mathscr{M}} \to 0$ a.s.) implies that μ_c does not charge any s.d.g. in R_β. Furthermore, if L is a lower layer in R_k, then $L^* = \{(t_1, t_2, \ldots, t_\beta): (t_{i_1}, t_{i_2}, \ldots, t_{i_k}) \in L\}$ is a lower layer in R_β and so (3) is also necessary for $D_n^{\mathscr{L}} \to 0$ a.s.

THEOREM 1. *Let $\{T_k\}$ be* i.i.d. *with distribution μ. $I_n^{\mathscr{M}} \to 0$ a.s. if and only if (3) holds. Furthermore, if $\{X_k\}$ satisfies (1) and is independent of $\{T_k\}$ and if (3) holds, then $M_n^{\mathscr{L}} \to 0$ a.s.*

PROOF. For the first claim the necessity has already been shown and the proof of the sufficiency is much like that given for Theorem 2 of Dehardt (1971). However, a stronger result will be proved in the next theorem.

As was noted earlier, the first conclusion combined with the proof given by Smythe (1980) for Theorem 1.1 establishes the second conclusion.

Considering the collection $\{f: f(t) = \chi_L(t), L \in \mathscr{L}\}$, it is seen that the first part of the theorem extends the work of Blum (1955), who has shown that if μ is absolutely continuous, then $D_n^{\mathscr{L}} \to 0$ a.s. To see that this result extends Steele's theorem, note that for B a s.d.g. in R_k with $k \le \beta$, $B^* = \{(b_1, b_2, \ldots, b_\beta): (b_{i_1}, b_{i_2}, \ldots, b_{i_k}) \in B\}$ is a d.g. in R_β; for B^* a d.g. in R_β, $B^* \subset \partial L$ with L the lower layer $\cup_{b \in B^*}(-\infty, b]$; and $(\mu_{i_1, i_2, \ldots, i_k})_c \le (\mu_c)i_1, i_2, \ldots, i_k$. So if μ_c does not charge the boundary of any lower layer then (3) holds. Furthermore, the second part of Theorem 1 extends the work of Smythe (1980) and Wright (1979). This condition on μ is optimal in the sense that it is also necessary if the $\{X_k\}$ are identically distributed and nondegenerate. For, if $M_n^{\mathscr{L}} \to 0$ a.s., then

$$M_n^\Delta = \sup\{S_n(L - L'): L' \subset L \text{ with } L, L' \in \mathscr{L}\}/n \to 0 \text{ a.s.}$$

Any Borel subset B' of a s.d.g. B can be written in the form $L - L'$ with $L' \subset L$ and $L, L' \in \mathscr{L}$, in particular $B' = (\cup_{b \in B'}(-\infty, b]) - (\cup_{b \in B'}((-\infty, b] - \{b\}))$. Let D^c be the set of points $t \in R_\beta$ with $\mu(\{t\}) > 0$ and note that the T_k which have values in D are distinct with probability one. So for any s.d.g. B in R_β, $M_n^\Delta \ge \sum_{k=1}^n X_k^+ \chi_{B \cap D}(T_k)/n$ a.s. If $\mu_c(B) > 0$ then M_n^Δ does not converge to zero a.s. Finally, note that if $M_n^{\mathscr{L}} \to 0$ a.s., then, with $1 \le k < \beta$, $1 \le i_1 < i_2 < \cdots < i_k \le \beta$, and $\mathscr{L}(k)$ the lower layers in R_k,

$$\sup_{L \in \mathscr{L}(k)} \left| \sum_{j=1}^n X_j \chi_L((T_j^{(i_1)}, T_j^{(i_2)}, \ldots, T_j^{(i_k)}) \right| / n \to 0 \text{ a.s.},$$

since $L^* = \{(t_1, t_2, \ldots, t_\beta): (t_{i_1}, t_{i_2}, \ldots, t_{i_k}) \in L\}$ is a lower layer if L is.

Wright (1979) investigated the rate of convergence of $P[M_n^{\mathscr{L}} \ge \epsilon]$ to zero since this gives an indication of the rate of consistency of the isotone regression estimator studied there. It is of interest to establish the rates given there under this less restrictive condition on μ. These results are analogues of known results giving rates of convergence in the law of large numbers. If, for some $r > 1$, $\{X_k\}$ satisfies

(4) $$E(X_k) \equiv 0, \text{ and } y^r F(y) \to 0 \text{ as } y \to \infty,$$

then $p(n, \epsilon) = P[|n^{-1} \sum_{k=1}^n X_k| \ge \epsilon] = o(n^{-r+1})$ for each $\epsilon > 0$. 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). If $\{X_k\}$ satisfies

(5) $$E(X_k) \equiv 0 \text{ and } F(y) \le O(\exp(-cy)) \text{ for } y \ge 0 \text{ and some } c > 0,$$

then $p(n, \epsilon)$ converges to zero exponentially; that is, for each $\epsilon > 0$ there are constants $M(\epsilon) > 0$ and $\rho(\epsilon)$ with $0 \leq \rho(\epsilon) < 1$ for which $p(n, \epsilon) \leq M(\epsilon)(\rho(\epsilon))^n$ for $n = 1, 2, \ldots$. This is a special case of Theorem A of Hanson (1967) and was proved in the identically distributed case by Cramér (1938).

THEOREM 2. *Let* $\{T_k\}$ *be* i.i.d. *as* μ *and suppose that* (3) *holds. Then* $q(n, \epsilon) = P[I_n^{\mathscr{M}} \geq \epsilon]$ *converges exponentially to zero for each* $\epsilon > 0$. *Furthermore, if* $\{X_k\}$ *satisfies* (4) *and is independent of* $\{T_k\}$, *then* $r(n, \epsilon) = P[M_n^{\mathscr{L}} \geq \epsilon] = o(n^{-r+1})$ *for each* $\epsilon > 0$ *and if* $\{X_k\}$ *satisfies* (5) *rather than* (4), *then* $r(n, \epsilon)$ *converges exponentially to zero for each* $\epsilon > 0$.

PROOF. We first show that $q(n, \epsilon)$ converges to zero exponentially for each $\epsilon > 0$. The proof is patterned after that given for Theorem 2 of Dehardt (1971) and begins by proving several lemmas. Dehardt has shown that with $\delta(.,.)$ the Lévy metric on $\mathscr{M}$, $\mathscr{M}(A) = \{f_A: f \in \mathscr{M}\}$ is compact if A is a bounded rectangle and f_A is the restriction of f to A, and that $f_n, f \in \mathscr{M}$ and $\delta(f_n, f) \to 0$ imply that $f_n(t) \to f(t)$ at each continuity point of f. (See his Lemmas 7 and 8.) Brunk, et al. (1956), have shown that the discontinuities of an $f \in \mathscr{M}$ lie on a countable number of decreasing graphs. So if μ does not charge any decreasing graph, then $T_A(f) = \int_A f\,d\mu$ is continuous in the metric δ for each Borel set $A \subset R_\beta$.

LEMMA 1. *If* $\{A_k\}$ *is a partition of* R_β *and*

$$q(n, \epsilon, A_k) = P\left[\sup_{f\in\mathscr{M}}\left|\int_{A_k} f\,d\mu_n - \int_{A_k} f\,d\mu\right| \geq \epsilon\right]$$

converges to zero exponentially for each $\epsilon > 0$, *and each* $k = 1, 2, \ldots$, *then so does* $q(n, \epsilon)$.

PROOF. Fix $\epsilon > 0$, let C denote the uniform bound on the elements in $\mathscr{M}$ and choose N so that $\mu(\cup_{k>N} A_k) \leq (6C)^{-1}\epsilon$. Clearly $q(n, \epsilon)$ is bounded above by $\sum_{k=1}^{N} q(n, (2N)^{-1}\epsilon, A_k) + q(n, \epsilon/2, \cup_{k>N} A_k)$ and the proof is completed by showing that the last term converges to zero exponentially. But

$$\left|\int_{\cup_{k>N}A_k} f\,d(\mu_n - \mu)\right| \leq C\,|\mu_n(\cup_{k>N}A_k) - \mu(\cup_{k>N}A_k)| + \epsilon/3,$$

and so $q(n, \epsilon/2, \cup_{k>N}A_k) \leq P[(|\mu_n(\cup_{k>N}A_k) - \mu(\cup_{k>N}A_k)| \geq (6C)^{-1}\epsilon]$ which converges to zero exponentially since $X_k = \chi_{\cup_{j>N}A_j}(T_k) - \mu(\cup_{j>N}A_j)$, $k + 1, 2, \ldots$ satisfy (5). (Any uniformly bounded sequence of random variables which are centered at their means satisfies (5).)

LEMMA 2. *If* $q(n, \epsilon, A)$ *converges to zero exponentially for each* $\epsilon > 0$ *and for each bounded rectangle* A, *then so does* $q(n, \epsilon)$.

PROOF. This is a corollary to Lemma 1.

LEMMA 3. *If* μ *does not charge any decreasing graph in* R_β, *then* $q(n, \epsilon)$ *converges to zero exponentially for each* $\epsilon > 0$.

PROOF. We consider an arbitrary bounded rectangle A and recall that $\mathscr{M}(A)$ is compact and that $T_A(\cdot)$ is continuous. For a fixed $\epsilon > 0$, $\mathscr{M}(A)$ is covered by neighborhoods $N(f)$ such that $g \in N(f)$ implies that $|\int_A (f - g)\,d\mu| < \epsilon/4$. Let $N(f_1), N(f_2), \ldots, N(f_\alpha)$ be a finite subcovering and note that there are "smallest" and "largest" elements in the closure of $N(f_i)$ for each i. Denote these by g_{i1} and g_{i2} for $i = 1, 2, \ldots, \alpha$. So for any $g \in N(f_i)$,

$g_{i1}(t) \le g(t) \le g_{i2}(t)$ for each $t \in A$. Notice that $\int_A (g_{i2} - g_{i1})\, d\mu \le \epsilon/2$ and $q(n, \epsilon, A)$ is bounded above by

$$\sum_{j=1}^{2} \sum_{i=1}^{\alpha} P\left[\,\left|\int_A g_{ij}\, d(\mu_n - \mu)\right| \ge (4\alpha)^{-1}\epsilon\right].$$

Each of the terms in this sum converge to zero exponentially since $X_k = g_{ij}(T_k)\chi_A(T_k) - \int_A g_{ij}\, d\mu$, $k = 1, 2, \ldots$ satisfy (5).

LEMMA 4. *Let B be a Borel subset of R_β with $\mu(B) > 0$ and let μ^* be defined by $\mu^*(A) = \mu(A \cap B)/\mu(B)$. If $q^*(n, \epsilon) = P[\sup_{f \in \mathscr{M}} | \int f\, d\mu_n^* - \int f\, d\mu^* | \ge \epsilon]$ converges to zero exponentially for each $\epsilon > 0$ with $\mu_n^*(A) = \sum_{k=1}^{n} \chi_A(T_k^*)/n$ and $\{T_k^*\}$* i.i.d. *as μ^*, then $q(n, \epsilon, B)$ converges to zero exponentially for each $\epsilon > 0$.*

PROOF. If $\mu(B) = 1$, then the result is clear. For $k = 1, 2, \ldots$ define T_k^* to be the value of T_j where j is the kth index for which $T_j \in B$ and note that $\{T_k^*\}$ is i.i.d. as μ^*. Now let $\int f\, d\mu_n^* = 0$ if $n = 0$ and observe that $\sup_{f \in \mathscr{M}} | \int_B f\, d\mu_n - \int_B f\, d\mu |$ is bounded above by

$$\sup_{f \in \mathscr{M}} \mu(B) \left| \frac{\mu_n(B)}{\mu(B)} \int f\, d\mu^*_{n \cdot \mu_n(B)} - \int f\, d\mu^* \right| \le$$

$$\sup_{f \in \mathscr{M}} \left| \int f\, d\mu^*_{n \cdot \mu_n(B)} - \int f\, d\mu^* \right| + C\,|\,\mu_n(B) - \mu(B)\,|,$$

where C is the uniform bound on the functions in $\mathscr{M}$. Clearly, $P[C\,|\,\mu_n(B) - \mu(B)\,| \ge \epsilon/2]$ converges at the desired rate and the probability that the first term exceeds $\epsilon/2$ is bounded by $\max_{k \ge n\mu(B)/2} q^*(k, \epsilon) + P[\mu_n^*(B) < \mu(B)/2]$. Both of these terms behave as specified.

For $A \subset R_k$ with $k < \beta$ and $1 \le i_1 < i_2 < \cdots < i_k \le \beta$, define $A^* = \{(a_1, a_2, \ldots, a_\beta)\colon (a_{i_1}, a_{i_2}, \ldots, a_{i_k}) \in A\}$. We call A^* a β-dimensional extension of A.

LEMMA 5. *If B is a* d.g. *in R_β, then there is a* s.d.g. *B_0 in R_β and a countable collection $\{B_n^*\}_{n=1}^{\infty}$ of β-dimensional extension of* s.d.g.'s *(with possibly different k and $1 \le i_1 < \cdots < i_k \le \beta$) for which $B \subset B_0 \cup \bigcup_{n=1}^{\infty} B_n^*$.*

PROOF. Since the only d.g.'s in R_1 are singletons the statement is clearly true for $\beta = 1$ and the proof proceeds by induction. For an arbitrary β, choose B_0 to be $\{b \in B$: there does not exist $b^* \in B$ with $b^* \le b$ or $b \le b^*\}$ and set $B' = B - B_0$. Clearly, B_0 is a s.d.g. in R_β and $B' = \{b \in B$: there is a $b^* \in B$ with $b^* \le b$ /or $b^* \ge b$ and $b_i = b_i^*$ for at least one $i\}$. Because of the inductive hypothesis it suffices to show that B' is contained in a countable union of β-dimensional extensions of $\beta - 1$ dimensional d.g.'s.

For $i \in \{1, 2, \ldots, \beta\}$ and $t \in R_\beta$ we let $t_{(i)}$ denote $(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots t_\beta)$ where, of course, t_i represents the ith coordinate of t. Define the nonincreasing function on $R_{\beta-1}$ by

$$f_i(t) = \sup\{b_i\colon b \in B' \text{ and } b_{(i)} \ge t\} \qquad (\sup \phi = -\infty).$$

(This function is not necessarily finite.) We show that if $b \in B'$, then for some i, $b_{(i)}$ is a discontinuity point of f_i. Suppose that $b, b^* \in B$ with $b_{i_\alpha} = b_{i_\alpha}^*$ for $\alpha = 1, 2, \ldots, k$ and $b_j < b_j^*$ for $j \notin \{i_\alpha\colon \alpha = 1, 2, \ldots, k\}$ with $k < \beta$. Choose $i \in \{1, 2, \ldots, \beta\} - \{i_\alpha\colon \alpha = 1, 2, \ldots, k\}$ and note that for $a \le b_{(i)}$, $f_i(a) \ge f_i(b_{(i)}) \ge b_i^*$. But for $b_{(i)} < c$ ($s < t$ means $s_i < t_i$ for each i), $f_i(c) \le b_i$ since $b_{(i)}^+ \ge c$ with $b^+ \in B'$ implies that $b_{(i)} < b_{(i)}^+$ and so $b_i^+ \le b_i$. Hence, $b_{(i)}$ is a point of discontinuity of f_i since $b_i^* > b_i$. (Similarly, $b_{(i)}^*$ can be shown to be a point of discontinuity of f_i.) The proof of the lemma is completed by noting that Brunk, et al. (1956), have shown that the discontinuities of each f_i are contained in a countable number of d.g.'s.

LEMMA 6. *If $\beta = 1$, then $q(n, \epsilon)$ converges to zero exponentially for each $\epsilon > 0$.*

PROOF. Let A_j denote the atoms of μ and partition R_1 by $(\cup_j A_j)^c$, $A_1, A_2, \ldots$ It is clear that $q(n, \epsilon, A_j)$ converges at the proper rate for each $\epsilon > 0$ and for each j. Define μ^* as in Lemma 4 with $B = (\cup_j A_j)^c$ and observe that μ^* does not charge any decreasing graph since the decreasing graphs in R_1 are singletons. Applying Lemmas 1, 3 and 4, the proof is completed.

Before considering arbitrary β, we restate Dehardt's definition of a flat. A set of the form $\{t_1, \ldots, t_\beta): t_{j_1} = c_1, \ldots t_{j_{\beta-i}} = c_{\beta-i}\}$ is called an i-dimensional flat for $i = 0, 1, \ldots, \beta - 1$. Let $A_{0,j}$ denote the 0-dimensional flats in R_β which are charged by μ (these are the atoms of μ), let $G_0 = \cup_j A_{0,j}$, let $A_{1,j}$ be the 1-dimensional flats in R_β for which $\mu(A_{1,j} \cap G_0^c) > 0$, let $G_1 = \cup_j A_{1,j}$, let $A_{2,j}$ be the 2-dimensional flats in R_β for which $\mu(A_{2,j} \cap (G_0 \cup G_1)^c) > 0$ and repeating this process let $A_{\beta-1,j}$ be the β-1-dimensional flats for which $\mu(A_{\beta-1,j} \cap (G_0 \cup \ldots \cup G_{\beta-2})^c) > 0$.

The proof of Theorem 2 is an induction on β. The result is valid for R_1 and so we suppose it is valid for R_j with $j = 1, 2, \ldots, \beta - 1$. By Lemma 1 it is sufficient to consider $q(n, \epsilon, G_0)$, $g(n, \epsilon, G_1 \cap G_0^c), \ldots, q(n, \epsilon, G_{\beta-1} \cap (G_0 \cup \cdots \cup G_{\beta-1})^c)$. Since $q(n, \epsilon\ A_{0,j})$ behaves properly for each j, we appeal to Lemma 1 to show that $q(n, \epsilon, G_0)$ also does. In considering $q(n, \epsilon, G_k \cap (G_0 \cup \cdots \cup G_{k-1})^c)$, we note that the $A_{k,j}$ which make up G_k could have been chosen to be subsets of k-dimensional flats and disjoint. So applying Lemma 1 again, we need to show that $q(n, \epsilon, A_{kj} \cap (G_0 \cup \cdots \cup G_{k-1})^c)$ behaves as specified. Define $\mu_{k,j}^*(A) = \mu(A \cap A_{kj} \cap (G_0 \cup \cdots \cup G_{k-1})^c)/\mu(A_{k,j} \cap (G_0 \cup \cdots \cup G_{k-1})^c)$ and note that since $\mu_{k,j}^*$ has support contained in the k-dimensional set $A_{k,j}$ it may be viewed as a probability on the Borel subsets of R_k. So according to Lemma 4 we need to show that the desired conclusion holds for this probability, but this follows from the inductive hypothesis if $\mu_{k,j}^*$, viewed as a probability on R_k, satisfies the k-dimensional analogue of (3). To show that μ_{kj}^* satisfies the k-dimensional analogue of (3), note that μ_{kj}^*, viewed as a probability on R_k, is continuous, its marginals are continuous, and if $1 \le \alpha \le k$ and B is a s.d.g. in R_α then $B \times \{(c_{k+1}, \ldots, c_\beta)\}$ is a s.d.g. in $R_{\beta-k+\alpha}$.

For the last term, define $\mu^*(A) = \mu(A \cap (G_0 \cup \cdots \cup G_{\beta-1})^c)/\mu((G_0 \cup \cdots \cup G_{\beta-1})^c)$ and note that μ^* as well as all of its marginals are continous. Furthermore, appealing to Lemma 5, we see that $\mu^*(B) = 0$ for any d.g. B in R_β, which according to Lemma 3 yields the desired conclusion.

With straightforward modifications, the proof of Theorem 1.1 of Smythe (1980) can be used to show that $r(n, \epsilon)$ converges to zero at the proper rate. So the proof of Theorem 2 is completed.

It is interesting to note that even though $P[D_n^{\mathscr{L}} \ge \epsilon]$ converges to zero exponentially, $D_n^{\mathscr{L}}$ does not obey the law of the iterated logarithm for $\beta > 2$. This latter result is due to Steele (1977) and the question is still open for $\beta = 2$.

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] BLUM, J. R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* **26** 527–529.

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

[4] BRUNK, H. D., EWING, G. M. and UTZ, W. R. (1956). Some Helly theorems for monotone functions. *Proc. Amer. Math. Soc.* **7** 776–783.

[5] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilitiés. *Actualities Sci. Indust.* **736**.

[6] DEHARDT, J. (1970). A necessary condition for Glivenko-Cantelli convergence in E_n. *Ann. Math. Statist.* **41** 2177–2178.

[7] DEHARDT, J. (1971). Generalizations of the Glivenko-Cantelli theorem. *Ann. Math. Statist.* **43** 2050–2055.

[8] 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.

[9] 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.
[10] HANSON, D. L., PLEDGER, GORDON and WRIGHT, F. T. (1973). On consistency in monotonic regression. *Ann. Statist.* **1** 401–421.
[11] SMYTHE, R. T. (1980). Maxima of partial sums and a monotone regression estimator. *Ann. Probability.* **6** 630–635.
[12] STEELE, J. MICHAEL (1977). Limit properties of random variables associated with a partial ordering on R^d. *Ann. Probability* **5** 395–403.
[13] STEELE, J. MICHAEL (1978). Empirical discrepancies and subadditive processes. *Ann. Probability* **6** 118–127.
[14] WRIGHT, F. T. (1979). A strong law for variables indexed by a partially ordered set with applications to isotone regression. *Ann. Probability* **7** 109–127.

DEPARTMENT OF MATHEMATICS
326 MATHEMATICS-COMPUTER SCIENCE BLDG.
UNIVERSITY OF MISSOURI-ROLLA
ROLLA, MISSOURI 65401