The Annals of Probability
1978, Vol. 6, No. 1, 118-127

EMPIRICAL DISCREPANCIES AND SUBADDITIVE PROCESSES

BY J. MICHAEL STEELE

University of British Columbia

If X_i, $i = 1, 2, \cdots$ are independent and identically distributed vector-valued random variables with distribution F, and S is a class of subsets of R^d, then necessary and sufficient conditions are given for the almost sure convergence of $(1/n)D_n^S = \sup_{A \in S} |(1/n) \sum 1_A(X_i) - F(A)|$ to zero. The criteria are defined by combinatorial entropies which are given as the time constants of certain subadditive processes. These time constants are estimated, and convergence results for $(1/n)D_n^S$ obtained, for the classes of algebraic regions, convex sets, and lower layers. These results include the solution to a problem posed by W. Stute.

1. Introduction. One of the most useful tools of probability theory developed in recent years is the theory of subadditive stochastic processes introduced by Hammersley and Welsh [10] and perfected by Kingman [11, 12] and Hammersley [9]. The first objective of this paper is to provide three new examples of such processes. The underlying motivation for studying these specific processes is that they provide an effective and unified approach to convergence theorems for the empirical discrepancy function.

To set the problem precisely, we let X_i, $i = 1, 2, \cdots$ be a stationary sequence of random variables which are defined on a probability space $(\Omega, P, \mathscr{F})$ and which take values in R^d, $d \geqq 1$. The common distribution function of the X_i will be denoted by F; thus for any measurable $A \subset R^d$ we have $F(A) = P(X_i \in A)$. For most of the results obtained here, independence of the X_i will be assumed, but, as indicated in the following, some results only require ergodicity of the X_i process.

Now let S be any class of measurable subsets of R^d. For any $A \in S$ we denote by $1_A(x)$ the function on R^d which is 1 if $x \in A$ and 0 otherwise. The fundamental object of concern here is the generalized discrepancy function defined by

$$D_n^S = \sup_{A \in S} \left| \sum_{i=1}^n 1_A(X_i) - nF(A) \right| . \tag{1.1}$$

The most classical choice of S is

$$S = \{(-\infty, y_1] \times (-\infty, y_2] \times \cdots \times (-\infty, y_d] : (y_1, y_2, \cdots, y_d) \in R^d\} ,$$

and in this case $(1/n)D_n^S$ is the very well studied Kolmogorov–Smirnov statistic (see, for example, Durbin [4]). Beyond these classical regions there are, of course, numerous choices of S of interest. In particular, for S the class of convex Borel sets the first work is due to R. Ranga Rao [14] and there are refinements by W. Phillip [13], P. Gaenssler [8], and W. Stute [17, 18].

Received October 20, 1976; revised April 18, 1977.

AMS 1970 *subject classifications.* Primary 60F15; Secondary 60G99.

Key words and phrases. Empirical distribution, subadditive processes, entropy, convex sets, lower layers, algebraic regions.

The study of general classes S was also begun in R. Ranga Rao [14], and, this time, followed by Topsøe [19]. The approach which is taken here is along the completely different lines introduced by Vapnik and Chervonenkis [20] and [21]. Their analysis was based on the combinatorial, rather than the topological, properties of S. The present work is concerned with perfecting and applying the method of Vapnik and Chervonenkis by means of subadditive ergodic theory.

Since in most cases of interest S is uncountable, the function D_n^S is not a priori measurable. Still, for all classes which we consider, D_n^S is measurable and for all results quoted here it will be assumed that D_n^S is measurable.

The main result of this paper is Theorem 4.2 which gives two necessary and sufficient conditions for the a.s. convergence of $(1/n)D_n^S$ to zero. This result is preceded by two brief sections which introduce the quantities necessary to define the convergence criteria. The fifth section applies the main result to the class of algebraic regions of R^d. The sixth section specializes S to the class of convex sets, and thus obtains a solution to a problem posed by W. Stute [17]. The final section similarly estimates the "time constant" of the subadditive process associated to the class of lower layers and consequently obtains as corollaries results due to Blum [1] and DeHardt [5, 6].

2. The Vapnik–Chervonenkis theorem. We will first observe that there is a direct relationship between the combinatorial structure of S and the convergence in probability of $(1/n)D_n^S$. If T is any set we denote by $|T|$ the cardinality of T. Now if W is a finite subset of R^d then there are only finitely many distinct sets $W \cap A$ for $A \in S$, and we let $\Delta^S(W)$ denote the number of such sets. More succinctly, we have

$$\Delta^S(W) = |\{W \cap A : A \in S\}| . \tag{2.1}$$

To solidify the understanding of this definition we note that for $S = \{(-\infty, y) : y \in R\}$ and W any k element subset of R one has $\Delta^S(W) = k + 1$.

Now if X_i, $i = 1, 2, \cdots$ are i.i.d. with distribution F the Vapnik–Chervonenkis entropy $V(F, S)$ is defined by

$$V(F, S) = \lim_{n\to\infty} \frac{1}{n} E \log \Delta^S(X_1, X_2, \cdots, X_n) . \tag{2.2}$$

With this definition we now have the *main theorem of Vapnik and Chervonenkis*:

If X_i, $i = 1, 2, \cdots$ are i.i.d. then $(1/n)D_n^S$ converges to zero in probability if and only if $V(F, S) = 0$.

In addition to this weak law for $(1/n)D_n^S$, Vapnik and Chervonenkis also provide a strong law under much more stringent conditions on $\Delta^S(X_1, X_2, \cdots, X_n)$, (see e.g., Theorem 3, [20]). A principal corollary of the present work is that one actually has the strong law under precisely the same circumstance as the weak law. That is, we prove that $(1/n)D_n^S$ *converges to zero with probability one if and only if* $V(F, S) = 0$.

3. Combinatorial entropy. The Vapnik–Chervonenkis entropy is clearly a

useful measure of the complexity of S, but it is not at all easy to compute. For this reason we introduce a new type of combinatorial entropy which is much more easily computed, but which is still effective in the study of D_n^S.

If W' is a finite subset of R^d, we say that W' is *shattered* by S provided that *every* subset of W' can be written as $W' \cap A$ for some $A \in S$. Next we let $k^S(W)$ equal the cardinality of the largest subset $W' \subset W$ which is shattered by S. This set function can also be expressed in terms of $\Delta^S(W)$ as follows:

$$k^S(W) = \max\{k : \Delta^S(W') = 2^k, W' \subset W, |W'| = k\}. \tag{3.1}$$

The combinatorial entropy $K(F, S)$ is now defined by the limit

$$K(F, S) = \liminf_{n\to\infty} \frac{1}{n} Ek^S(X_1, X_2, \cdots, X_n). \tag{3.2}$$

In the next section we will prove that $(1/n)D_n^S$ *converges* a.s. *to zero if and only if* $K(F, S) = 0$. In later sections $K(F, S)$ will be estimated for several specific classes, in particular for the class of convex subsets of R^d.

4. Three subadditive processes. After Kingman [11, 12] we call a family Y_{st} $(s < t)$ of doubly integer-indexed random variables a subadditive process provided three conditions hold:

S_1. Whenever $s < t < u$, $Y_{su} \leqq Y_{st} + Y_{tu}$;

S_2. The joint distributions of the process $(Y_{s+1,t+1})$ are the same as those of $(Y_{s,t})$;

S_3. The expectations $g_t = E(Y_{0t})$ exist and satisfy $g_t \geqq -At$ for some $A \geqq 0$ and all $t \geqq 1$.

The most valuable consequence of these properties is brought out by Kingman's theorem which says that for any subadditive process there is a random variable ξ with finite mean such that

$$\xi = \lim_{t\to\infty} Y_{st}/t \tag{4.1}$$

with probability one and in mean.

Perhaps the easiest, but most crucial, observation of this section is that the generalized discrepancy, the Vapnik–Chervonenkis entropy, and the combinatorial entropy are all essentially related to subadditive processes.

THEOREM 4.1. *If X_i, $i = 1, 2, \cdots$, is a stationary sequence of random variables then each of the following is a subadditive process*:

$$D_{st}^S = \sup_{A\in S} |\textstyle\sum_{i=s+1}^{t} 1_A(X_i) - (t - s)F(A)| \tag{4.2}$$

$$\log \Delta_{st}^S = \log \Delta^S(X_{s+1}, X_{s+2}, \cdots, X_t) \tag{4.3}$$

$$k_{st}^S = k^S(X_{s+1}, X_{s+2}, \cdots, X_t). \tag{4.4}$$

PROOF. The stationarity property S_2 follows directly from the stationarity of the X_i in all three cases. Also, since each of the processes is positive and bounded we note that S_3 is valid. Only S_1 will require individual checking.

The proof of S_1 for D_{st}^S is just the triangle inequality. To prove S_1 for $\log \Delta_{st}^S$ we first note that for any $A \in S$ one has

$$A \cap \{X_{s+1}, X_{s+2}, \cdots, X_t, X_{t+1}, \cdots, X_u\} = (A \cap \{X_{s+1}, X_{s+2}, \cdots, X_t\}) \cup (A \cap \{X_{t+1}, X_{t+2}, \cdots, X_u\}) . \tag{4.5}$$

But (4.5) shows that there are fewer sets of the form $A \cap \{X_{s+1}, X_{s+2}, \cdots, X_u\}$ than there are pairs of sets $(A \cap \{X_{s+1}, X_{s+2}, \cdots, X_t\}, A \cap \{X_{t+1}, X_{t+2}, \cdots, X_u\})$. Since the number of such pairs is $\Delta_{st}^S \Delta_{tu}^S$, (4.5) proves that

$$\Delta_{su}^S \leqq \Delta_{st}^S \Delta_{tu}^S . \tag{4.6}$$

On taking logarithms one has S_1 for $\log \Delta_{st}^S$.

Now to prove S_1 for k_{st}^S suppose that W is the largest shattered subset of $\{X_{s+1}, X_{s+2}, \cdots, X_{t+1}, \cdots, X_u\}$. Then $W_1 = W \cap \{X_{s+1}, X_{s+2}, \cdots, X_t\}$ and $W_2 = W \cap \{X_{t+1}, X_{t+2}, \cdots, X_n\}$ are also shattered. Hence we have

$$k_{st}^S = |W| = |W_1| + |W_2| \leqq k_{st}^S + k_{tu}^S \tag{4.7}$$

and the proof of Theorem 2.1 is complete.

To state the next result in a symmetric fashion we introduce a constant $D(F, S)$ defined as follows:

$$D(F, S) = \liminf_{n\to\infty} \frac{1}{n} E \sup_{A \in S} \left| \sum_{i=1}^n 1_A(X_i) - F(A) \right| . \tag{4.8}$$

THEOREM 4.2. *If X_i, $i = 1, 2, \cdots$ is a stationary ergodic sequence then the following limits hold almost surely and in mean*:

$$\lim_{n\to\infty} \frac{1}{n} D_{0n}^S = D(F, S) \tag{4.9}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \Delta_{0n}^S = V(F, S) \tag{4.10}$$

$$\lim_{n\to\infty} \frac{1}{n} k_{0n}^S = K(F, S) . \tag{4.11}$$

Moreover, if the X_i are i.i.d. and one of the constants $D(F, S)$, $V(F, S)$, or $K(F, S)$ is zero then they are all zero.

PROOF. The existence of the limits on the left side of equations (4.9), (4.10), and (4.11) is an immediate consequence of Kingman's subadditive ergodic theorem and Theorem 2.1. Further since these limits are unchanged by the shift $(X_i) \to (X_{i+1})$ the ergodicity of the (X_i) process implies the limits are constants. These constants are then identified as above by noting that all the variables above are bounded by 1 and then applying the dominated convergence theorem.

To prove the second half of the theorem we first recall that if the X_i are i.i.d. the main theorem of Vapnik and Chervonenkis says that $n^{-1}D_{0n}^S$ converges to zero in probability if and only if $V(F, S) = 0$. Hence by the convergence in

mean of $n^{-1}D_{0n}^S$ we have that $D(F, S) = 0$ if and only if $V(F, S) = 0$. The theorem is completed once one proves that $V(F, S) = 0$ if and only if $K(F, S) = 0$ and that will be accomplished by the following lemma.

LEMMA 4.3. *If we define* $h(x) = -\log x^x(1-x)^{1-x}$ *for* $0 < x \leqq \frac{1}{2}$ *and* $h(x) = 1$ *for* $\frac{1}{2} \leqq x \leqq 1$ *we have the inequality*

$$K(F, S) \leqq V(F, S) \leqq h(K(F, S)) . \tag{4.12}$$

Also we have $V(F, S) = 0$ *if and only if* $K(F, S) = 0$.

PROOF. Since $2^{k_{0n}^S} \leqq \Delta_{0n}^S$ the first inequality of (4.12) follows by taking logarithms and limits. To prove the second inequality we recall a result due to N. Sauer [15] (which is only a little sharper than Lemma 1, page 266, [20]). According to Sauer's theorem there must exist a shattered k element subset of $X_1, X_2, \cdots, X_n$ if $\Delta_{0n}^S \geqq \sum_{j=0}^{k-1} \binom{n}{j} + 1$. Hence for any $0 < \alpha < \frac{1}{2}$ we have that $k_{0n}^S < [\alpha n]$ implies

$$\Delta_{0n}^S \leqq \sum_{j=0}^{[\alpha n]} \binom{n}{j} \leqq (\alpha l)\binom{l}{\alpha l} . \tag{4.13}$$

By Stirling's formula one then obtains

$$V(F, S) \leqq h(\alpha) \tag{4.14}$$

for all α such that $K(F, S) < \alpha$. Finally by the continuity of $h(x)$ one obtains (4.12) and the fact that $V(F, S) = 0$ if and only if $K(F, S) = 0$.

The proof of Lemma 4.2 completes the proof of Theorem 4.2 and, even though the result has been obtained effortlessly, we will note in succeeding sections that it contains numerous known results concerning D_{0n}^S as well as new ones.

Simplest among the conclusions of Theorem 4.2 is the fact that $(1/n)D_{0n}^S$ always converges a.s., even when the limit is not zero. In the present context this is not as surprising as it would be if one approached the study of D_{0n}^S without subadditive ergodic theory. Otherwise, it would have been noted long ago.

To demonstrate the applications of the other conclusions of Theorem 4.2 we will consider specific classes in the following sections and show how $K(F, S)$ can be computed.

5. An application to algebraic regions. The first special class which we consider is the class of polynomial regions. To describe these we first let $P_{m,d}$ denote the set of all real polynomials in d variables of degree not greater than m. For example,

$$P_{2,3} = \{\sum a_{ijk}x^iy^jz^k,\ a_{ijk} \text{ real},\ 0 \leqq i + j + k \leqq 2\} .$$

Now if $g \in P_{m,d}$ then g defines a subset A_g of R^d in a natural way as follows,

$$A_g = \{v \in R^d : g(v) \geqq 0\} . \tag{5.1}$$

Finally, we call the class of sets $\tilde{P}_{m,d} = \{A_g : g \in P_{m,d}\}$ the class of mth degree polynomial regions of R^d. One should note, for example, that $\tilde{P}_{2,3}$ contains all the ellipsoids in R^3 as well as many other regions.

An especially important class of algebraic regions are the half-spaces $H = \tilde{P}_{1,d}$. For these regions the calculation of Δ_{0n}^H is essentially a classical result due to Schäfli (see [3] and [16]), which we now note.

PROPOSITION 5.1. *If $y_1, y_2, \cdots, y_n$ are elements of R^τ and if $H = P_{1,\tau}$ then*

$$\Delta^H(y_1, y_2, \cdots, y_n) \leqq 2 \sum_{k=0}^{\tau} \binom{n-1}{k}, \tag{5.2}$$

and equality holds in (5.2) *provided the y_i are in general position* (i.e., *any τ or fewer of the y_i are linearly independent*).

That an estimate for Δ_{0n}^S with $S = \tilde{P}_{m,d}$ follows from (5.2) was first observed by T. Cover [3], whose brief proof we quote.

PROPOSITION 5.2. *For $S = \tilde{P}_{m,d}$ one has*

$$\Delta_{0n}^S \leqq 2 \sum_{k=0}^{\tau} \binom{m-1}{k} \quad \text{where} \quad \tau = \binom{m+d}{d}. \tag{5.3}$$

PROOF. The idea is to associate to each element of $S = \tilde{P}_{m,d}$ an element of $H = \tilde{P}_{\tau,1}$. If $y \in R^d$ and $y = (y(1), y(2), \cdots, y(d))$ we define $\Phi(y) \in R^\tau$ by

$$\Phi(y) = (1, y(1), \cdots, y(d), y(1)^2, \cdots, y(d)^2, \cdots, y(i)y(j), \cdots, y(d)^m). \tag{5.4}$$

Now the proof is completed by noting that

$$\Delta^S(y_1, y_2, \cdots, y_n) \leqq \Delta^H(\phi(y_1), \phi(y_2), \cdots, \phi(y_n)) \leqq 2 \sum_{k=0}^{\tau} \binom{m-1}{k}.$$

The preceding estimates naturally yield the following consequence.

COROLLARY 5.1. *If X_i, $i = 1, 2, \cdots$ is a sequence of* i.i.d. *random variables with any distribution, and if $S = \tilde{P}_{m,d}$ then* $\lim_{n\to\infty} (1/n) D_{0n}^S = 0$ a.s.

Of course, to prove the above one just notes that Δ_{0n}^S is dominated by a polynomial and thus $\lim_{n\to\infty} (1/n) \log \Delta_{0n}^S = 0$ a.s. One should note that the corollary contains the Glivenko–Cantelli theorem for half spaces due to Wolfowitz [22, 23].

6. Convex sets and Stute's problem. By applying our main result, Theorem 4.2, to the class of convex Borel sets, we can now answer a question posed by W. Stute [17, page 168].

To state Stute's problem we first note that for any distribution F, there are unique finite measures F_0 and F_1 such that $F = F_0 + F_1$, F_1 is atomic, and F_0 is non-atomic. Next if C is the class of convex Borel subsets of R^d and $A \in C$ we write ∂A for the boundary of A. It was proved by R. Ranga Rao [14] that

$$\sup_{A \in C} F_0(\partial A) = 0 \tag{6.1}$$

is a sufficient condition for $D(F, C) = 0$. To provide a comparable necessary condition, W. Stute [18] introduced $e(A)$, the set of extreme points of A and proved that $D(F, C) = 0$ implies

$$\sup_{A \in C} F_0(e(A)) = 0. \tag{6.2}$$

The problem considered of "considerable interest" by W. Stute ([17, page 168]) is to find conditions for $D(F, C) = 0$ which are stronger than (6.2). The following

theorem proves a new proof of R. Ranga Rao's sufficient condition, W. Stute's necessary condition, and an answer to W. Stute's problem.

THEOREM 6.1. *Let X_i, $i = 1, 2, \cdots$, be a sequence of ergodic stationary random variables with values in R^d and distribution F. For C, the class of convex Borel sets, one has*

$$\sup_{A \in C} F_0(e(A)) \leqq K(F, C) \leqq \sup_{A \in C} F_0(\partial A) . \tag{6.3}$$

PROOF. First note that any finite set of distinct elements of $e(A)$ is shattered by C. Now if D^e is the union of all the atoms of F, then by the continuity of F on D the set $\{X_i : X_i \in D, i = 1, 2, \cdots, n\}$ contains only distinct points with probability one. Hence we have for any $A \in C$

$$k^C(X_1, X_2, \cdots, X_n) \geqq \sum_{i=1}^n 1_{D \cap e(A)}(X_i) . \tag{6.4}$$

By the ergodic theorem (6.4) yields

$$\lim_{n \to \infty} n^{-1} k_{0n}^C \geqq F(e(A) \cap D) = F_0(e(A)) ,$$

and taking the supremum yields the first inequality in (6.3).

To prove the second inequality we require a simple truncation. Letting B_m denote the ball about 0 of radius m we define $k(m)$ by

$$k(m) = \lim_{n \to \infty} n^{-1} k^C(\{X_1, X_2, \cdots, X_n\} \cap B_m \cap D) . \tag{6.5}$$

By subadditivity we have

$$K(F, C) \leqq k(m) + F(B_m{}^c) + \lim_{n \to \infty} n^{-1} k^C(\{X_1, X_2, \cdots, X_n\} \cap D^e) . \tag{6.6}$$

If $A_1, A_2, \cdots$ denote the atoms of F and $G_k = \bigcup_{i \geqq k} A_i$, then for any $k \geqq 1$

$$n^{-1} k^C(\{X_1, \cdots, X_n\} \cap D^e) \leqq n^{-1} k + n^{-1} \sum_{i=1}^n 1_{G_k}(X_i) . \tag{6.7}$$

The ergodic theorem yields for all fixed k,

$$\lim_{n \to \infty} n^{-1} k^C(\{X_1, X_2, \cdots, X_n\} \cap D^e) \leqq F(G_k) .$$

Since $F(G_k) \to 0$ as $k \to \infty$, (6.6) simplifies to

$$K(F, C) \leqq k(m) + F(B_m{}^c) . \tag{6.8}$$

To estimate $k(m)$ we use the Blaschke selection theorem (see Eggleston [7, page 64]) which allows us to choose for each r a finite subclass S_r of convex subsets of B_m with the following property:

If E is a convex subset of B_m, there is an $A \in S_r$ such that

$$\partial E \subset \{x : \inf_{y \in \partial A} |x - y| \leqq 1/r\} .$$

We note that the set $\{x : \inf_{y \in \partial A} |x - y| < \delta\}$ is a "thickened boundary" of A and we will denote it by $T(A, \delta)$. Now we have a basic observation:

$$k^C(\{X_1, X_2, \cdots, X_n\} \cap B_m \cap D) \leqq \max_{E \in B_m} \sum_{i=1}^n 1_{C \cap e(E)}(X_i) . \tag{6.9}$$

The proof of (6.9) consists in noting that for a set of points to be shattered they must be extreme points of a convex set. Now by the choice of S_r, the inequality

(6.9) implies that

$$k^C(\{X_1, X_2, \cdots, X_n\} \cap B_m \cap D) \leqq \max_{A \in S_r} \sum_{i=1}^n 1_{D \cap T(A,1/r)}(X_i) \tag{6.10}$$

which by the ergodic theorem yields

$$k(m) \leqq \max_{A \in S_r} F_0(T(A, 1/r)) . \tag{6.11}$$

To take advantage of (6.11) we choose for each r an $A_r \in S_r$ such that

$$k(m) \leqq F_0(T(A_r, 1/r)) \qquad \text{for all} \quad r . \tag{6.12}$$

Again by Blaschke's theorem there is a convex $A' \subset B_m$ such that one has the following:

Given any δ and any R there is an $r = r(\delta, R)$ such that $r \geqq R$ and

$$T(A_r, 1/r) \subset T(A', \delta) .$$

Hence we have

$$\lim_{r \to \infty} F_0(T(A_r, 1/r)) \leqq F_0(T(A', \delta)) , \tag{6.13}$$

and by (6.12) and (6.6) this yields

$$K(F, C) \leqq F(B_m{}^c) + F_0(T(A', \delta)) . \tag{6.14}$$

The countable additivity of F_0 gives $\lim_{\delta \to 0} F_0(T(A', \delta)) = F_0(\partial A')$ so (6.14) implies the second inequality in (6.3), since $F(B_m{}^c)$ can be made as small as we like. This completes the proof of Theorem 6.1.

Now we need to check that Theorem 6.1 actually provides a solution to Stute's problem. We already know from Theorem 4.2 that $K(F, C) = 0$ is necessary and sufficient for $D(F, C) = 0$. But by the inequality of Theorem 6.1 we also have $K(F, C) = 0$ is in a very strict sense a stronger necessary condition than

$$\sup_{A \in C} F_0(e(\partial A)) = 0 \qquad \text{for} \quad D(F, C) = 0$$

One further consequence of Theorem 6.1 is that in the case where all subflats have measure zero one can prove that

$$\sup_{A \in C} F_0(e(\partial A)) = \sup_{A \in C} F_0(\partial A)$$

so the inequalities in (6.3) become equalities. The Theorem 6.1 is then an instance of computing the "time constant" of a subadditive process, and the calculation of such constants was pointed out by Kingman [13] to be a basic problem in the theory of subadditive processes.

7. The class of lower layers. There is an interesting class of sets introduced by J. Blum [1] which is quite different from the class of convex sets but for which a twin result to Theorem 6.1 can be proved. A Borel set $A \subset R^d$ is called a lower layer provided for all $(y_1, y_2, \cdots, y_d) \in A$ we also have

$$\{(y_1', y_2', \cdots, y_d') : y_i' \leqq y_i, i = 1, 2, \cdots, d\} \subset A . \tag{7.1}$$

To fortify ones understanding of (7.1) it is useful to check that in R^2 the set of points which lie below a decreasing function forms a lower layer.

The class of lower layers in R^d will be denoted by L, and we need a notion of extreme point for $A \in L$. By analogy to the convex case we say x is an extreme point of $A \in L$ provided $A \backslash \{x\} \in L$. There will be no confusion in denoting the set of extreme points of A by $e(A)$. If F_0 is again used to denote the non-atomic part of F, then the following result holds:

THEOREM 7.2. *For X_i, $i = 1, 2, \cdots$ a stationary ergodic sequence with values in R^d and distribution F,*

$$\sup_{A \in L} F_0(e(A)) \leqq K(F, L) \leqq \sup_{A \in L} F_0(\partial A) . \tag{7.2}$$

The proof goes so nearly like that of Theorem 6.1 that it can be omitted safely. The crucial observation is that a set of points $\{x_1, x_2, \cdots, x_n\} \subset R^d$ is shattered if and only if $\{x_1, x_2, \cdots, x_n\} \subset e(A)$ for some $A \in L$. Also one needs to know that there is a compactness theorem due to Brunk, Ewing, and Utz [2] which plays the same role for L that Blaschke's theorem plays for C.

As a corollary to Theorem 7.1 we obtain a sufficient condition (due to Blum [1]) and a necessary condition (due to DeHardt [5, 6]) for a Glivenko–Cantelli theorem for the class of lower layers.

COROLLARY 7.2. *For X_i, $i = 1, 2, \cdots$* i.i.d. *one has $D(F, L) = 0$ if $\sup_{A \in L} F_0(\partial A) = 0$, and $\sup_{A \in L} F_0(e(A)) = 0$ if $D(F, L) = 0$.*

REFERENCES

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

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

[3] COVER, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications to pattern recognition. *IEEE Trans. on Elec. Comp.* **EC-14** 326–334.

[4] DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Bristol.

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

[6] DEHARDT, J. (1971). Generalizations of the Glivenko–Cantelli theorem. *Ann. Math. Statist.* **42** 2050–2055.

[7] EGGLESTON, H. G. (1958). *Convexity*. Cambridge University Press, Cambridge.

[8] GAENSSLER, P. (1974). Around the Glivenko–Cantelli theorem. Colloquia Mathematica Societatis Jánor Bolyai **11** *Limit Theorems of Probability Theory* (P. Révész, ed.) Keszthely (Hungary).

[9] HAMMERSLEY, J. M. (1974). Postulates for Subadditive Processes. *Ann. Probability* **2** 652–680.

[10] HAMMERSLEY, J. M. and WELSH, D. J. A. (1965). First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. *Bernoulli–Bayes–Laplace Anniversary Volume.* Springer-Verlag, Berlin, 61–110.

[11] KINGMAN, J. F. C. (1968). The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc.* **B30** 499–510.

[12] KINGMAN, J. F. C. (1973). Subadditive ergodic theory. *Ann. Probability* **1** 883–909.

[13] PHILLIP, W. (1973). Empirical distribution function and uniform distribution mod 1. *Diophantine Approximation and its Application.* Academic Press, New York.

[14] RAO, R. RANGA (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

[15] SAUER, N. (1972). On the density of families of sets. *J. Comb. Theory (A)* **13** 145–147.

[16] SCHLÄFLI, L. (1950). *Gesammelte Mathematische Abhandlungen I.* Verlag Birkhaüser, Bern, 209–212.

[17] STUTE, W. (1975). A necessary condition for the convergence of the isotrope discrepancy. No. 16 RUB Preprint, Ruhr-Universität.

[18] STUTE, W. (1976). On a generalization of the Glivenko–Cantelli theorem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 167–175.

[19] TOPSØE, F. (1970). On the Glivenko–Cantelli theorem. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 239–250.

[20] VAPNIK, V. N. and CHERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probability Appl.* **16** 264–280.

[21] VAPNIK, V. N. and CHERVONENKIS, A. YA. (1974). *Teoriya Raspoznavaniya Obrazov.* Nauka, Moscow.

[22] WOLFOWITZ, J. (1954). Generalizations of the theorem of Glivenko–Cantelli. *Ann. Math. Statist.* **25** 131–138.

[23] WOLFOWITZ, J. (1960). Convergence of the empiric distribution function on half-spaces. *Contributions to Probability and Statistics.* Stanford University Press.

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF BRITISH COLUMBIA
2075 WESBROOK PLACE
VANCOUVER, B.C.
CANADA V6T 1W5

Added in proof. My thanks are due Rick Greer and Ron Pyke for their kind and careful comments on an earlier draft of this paper.