The Annals of Probability
1986, Vol. 14, No. 2, 582–597

A UNIFORM CENTRAL LIMIT THEOREM FOR SET-INDEXED PARTIAL-SUM PROCESSES WITH FINITE VARIANCE

BY KENNETH S. ALEXANDER AND RONALD PYKE[1]

University of Washington

Given a class $\mathscr{A}$ of subsets of $[0,1]^d$ and an array $\{X_{\mathbf{j}}: \mathbf{j} \in \mathbb{Z}_+^d\}$ of independent identically distributed random variables with $EX_{\mathbf{j}} = 0$, $EX_{\mathbf{j}}^2 = 1$, the (unsmoothed) partial-sum process S_n is given by $S_n(A) := n^{-d/2}\sum_{\mathbf{j} \in nA} X_{\mathbf{j}}$, $A \in \mathscr{A}$. If for the metric $\rho(A, B) = |A \,\Delta\, B|$ the metric entropy with inclusion $N_I(\varepsilon, \mathscr{A}, \rho)$ satisfies $\int_0^1 (\varepsilon^{-1} \log N_I(\varepsilon, \mathscr{A}, \rho))^{1/2}\, d\varepsilon < \infty$, then an appropriately smoothed version of the partial-sum process converges weakly to the Brownian process indexed by $\mathscr{A}$. This improves on previous results of Pyke (1983) and of Bass and Pyke (1984) which require stronger conditions on the moments of $X_{\mathbf{j}}$.

1. Introduction. The purpose of this paper is to establish a uniform central limit theorem for partial-sum processes indexed by large families of sets when only the second moment is assumed to be finite. The context of the problem is as follows. Let $\{X_{\mathbf{j}}: \mathbf{j} \in \mathbb{Z}_+^d\}$ be an array of real valued random variables (r.v.) indexed by the d-dimensional positive integer lattice. For any bounded $B \in \mathscr{B}^d$, the class of all Borel sets in $\mathbb{R}^d$, define

$$S(B) = \sum_{\mathbf{j} \in B} X_{\mathbf{j}}.$$

Let $\mathscr{A}$ be a family of Borel subsets of the unit cube $I^d := [0,1]^d$ and define the normalized *partial-sum process* $S_n := \{S_n(A): A \in \mathscr{A}\}$ by

$$S_n(A) = n^{-d/2} S(nA); \tag{1.1}$$

where $nA = \{nx: x \in A\}$. The normalization used here is determined by the assumption made throughout that the $X_{\mathbf{j}}$'s are independent and identically distributed (iid) with $EX_{\mathbf{1}} = 0$ and $EX_{\mathbf{1}}^2 = 1$. (Here $\mathbf{1}$ denotes $(1, \ldots, 1) \in \mathbb{Z}_+^d$.) Our aim is to study the weak convergence of processes like S_n. However, to avoid difficulties that arise when the lattice points in a set A are not in some sense representative of A (e.g., when the boundary of A weaves excessively in and out of lattice points), it is necessary to consider an appropriate smoothed version of the partial-sum process as follows. For $B \in \mathscr{B}^d$, define

$$X(B) = \sum_{\mathbf{j}} |B \cap C_{\mathbf{j}}| X_{\mathbf{j}},$$

where $C_{\mathbf{j}}$ is the unit cube $(\mathbf{j} - \mathbf{1}, \mathbf{j}]$ and $|\cdot|$, or λ, is used to denote Lebesgue

Received July 1984; revised April 1985.

[1]Research supported in part by National Science Foundation Grants MCS-83-11686 and MCS-82-02861, respectively.

AMS 1980 *subject classifications*. Primary 60F05; secondary 60B10.

Key words and phrases. Partial-sum processes, metric entropy, weak convergence, set-indexed processes, Gaussian processes.

measure. The appropriate smoothed partial-sum process indexed by $\mathscr{A}$ is then given by

$$Z_n(A) = n^{-d/2}X(nA) = n^{-d/2}\sum_{\mathbf{j}} b_{n\mathbf{j}}(A)X_{\mathbf{j}} \tag{1.2}$$

for $A \in \mathscr{A}$, where

$$b_{n\mathbf{j}}(A) = |(nA) \cap C_{\mathbf{j}}|. \tag{1.3}$$

In these partial-sum processes, the r.v. $X_{\mathbf{j}}$ may be interpreted as a random measurement associated to either the location point $\mathbf{j}$ or to the $\mathbf{j}$th element $C_{\mathbf{j}}$ of a given partition. In the smoothed case, the total random mass $X_{\mathbf{j}}$ is viewed as being uniformly spread over the cube $C_{\mathbf{j}}$. Comments about possible generalizations of this set-up are included below in Section 4.

The first weak convergence results for multi-dimensionally indexed partial-sum processes were given for the case when $\mathscr{A}$ was equal to $\mathscr{I}^d := \{(0, \mathbf{x}]: \mathbf{x} \in I^d\}$, the set of all lower-left orthants. They were by Wichura (1969) under a finite variance condition, and earlier by Kuelbs (1968) under additional moment restrictions. These results reduce in dimension one to the original invariance principle of Donsker (1951).

In Pyke (1983) the weak convergence was derived for the smoothed partial-sum processes Z_n when they are indexed by rather large families $\mathscr{A}$. This uniform central limit theorem required however that the $X_{\mathbf{j}}$'s satisfy a moment condition which becomes more restrictive as the size of $\mathscr{A}$ increases. Specifically, it is assumed in Pyke (1983) that $E|X_{\mathbf{j}}|^s < \infty$ for some $s > 2(1 + r)/(1 - r)$ where r is the exponent of entropy defined in the following section. By contrast, the theorem proved in this paper requires only that the second moment of the $X_{\mathbf{j}}$'s be finite. With respect to moment assumptions then, this result can not be improved.

The reader is referred to other recent papers concerning partial-sum processes; to Bass and Pyke (1984) for a law of the iterated logarithm and uniform central limit theorem for independent arrays obtained via a Skorokhod-type embedding; to Goldie and Greenwood (1984) in the case of arrays of dependent r.v.'s; to Morrow and Philipp (1984) for invariance principles and rates of convergence in the independence case; and to Alexander (1984) for independent arrays indexed by Vapnik–Červonenkis classes. Related results by Ossiander and Pyke (1984) and Ossiander (1984) study limit laws for arrays that have random locations as well as random masses.

The outline of this paper is as follows. In Section 2 we introduce notation, state the assumptions that we impose on $\mathscr{A}$, and establish the convergence of the finite-dimensional distributions of Z_n. In Section 3 it is shown that the image laws of $\{Z_n: n \geq 1\}$ are tight, thereby establishing the desired uniform central limit theorem. Several remarks about extensions and open questions are included in Section 4.

2. Notation, assumptions, and finite-dimensional convergence. Define the pseudometric d_λ on $\mathscr{A}$ by $d_\lambda(A, B) = \lambda(A \,\Delta\, B) = |A \,\Delta\, B|$ where λ and $|\cdot|$

are both used to denote Lebesgue measure. We assume that *with respect to* d_λ, $\mathscr{A}$ *is totally bounded with inclusion and has a convergent entropy integral.* That is, first, for every $\varepsilon > 0$ there exists a finite collection (called an *ε-net*) $\mathscr{A}(\varepsilon)$ of measurable sets such that $A \in \mathscr{A}$ implies $A^{(1)} \subset A \subset A^{(2)}$ and $d_\lambda(A^{(1)}, A^{(2)}) \leqq \varepsilon$ for some $A^{(1)}$, $A^{(2)}$ in $\mathscr{A}(\varepsilon)$. Second, the number of pairs $A^{(1)}$, $A^{(2)}$ in $\mathscr{A}(\varepsilon)$, which we assume to be the minimum possible and which we denote by

$$N_I(\varepsilon, \mathscr{A}, d_\lambda) := \min\{k \geqq 1\colon \text{There exist measurable sets } A_i^{(1)}, A_i^{(2)}, 1 \leqq i \leqq k \text{ such that for every } A \in \mathscr{A} \text{ there is some } i \text{ such that } |A_i^{(2)} \setminus A_i^{(1)}| \leqq \varepsilon \text{ and } A_i^{(1)} \subset A \subset A_i^{(2)}\}$$

satisfies

$$\int_0^1 \left(\varepsilon^{-1} \log N_I(\varepsilon, \mathscr{A}, d_\lambda)\right)^{1/2} d\varepsilon < \infty. \tag{2.1}$$

Note that (2.1) is equivalent to

$$\int_0^1 \left(\log N_I(\varepsilon^2, \mathscr{A}, d_\lambda)\right)^{1/2} d\varepsilon < \infty,$$

an alternate form which has been used by some authors.

Define the *exponent of metric entropy* of $\mathscr{A}$, denoted r, by $r := \inf\{s > 0\colon \log N_I(\varepsilon, \mathscr{A}, d_\lambda) = O(\varepsilon^{-s}) \text{ as } \varepsilon \to 0\}$. If $r < 1$, then (2.1) holds.

Examples of index families which satisfy our metric entropy assumptions include the following. If $\mathscr{C}^d$ denotes the convex subsets of I^d, then it is shown in Dudley (1974), that $r = (d-1)/2$; noting that the use of entropy with inclusion does not affect the computation for convex sets. Now, let $\mathscr{J}(\alpha, d, M)$, for $\alpha > 0, M > 0$, denote the class of sets introduced in Dudley (1974), whose boundaries are images of α-differentiable mappings of the $(d-1)$-sphere into I^d, with all derivatives of orders up to α uniformly bounded by M. Then, $r = (d-1)/\alpha$; cf. Dudley (1974). A related family of sets with α-smooth boundaries, denoted $\mathscr{R}(\alpha, d, M)$, was proposed by Révész (1976) and shown there to satisfy $r = (d-1)/\alpha$ as well. Notice that in these last two examples, $r < 1$ if and only if $d < \alpha + 1$. Thus as the dimension increases, the smoothness of the sets as measured by the differentiability of their boundaries must also increase.

Some examples of "small" classes of sets are $\mathscr{I}^d$, the set of intervals on lower orthants defined above; $\mathscr{P}^{d,m}$, the family of all polygonal regions in I^d with no more than m vertices; and $\mathscr{E}^d$, the set of all ellipsoidal regions in I^d. For all of these, $r = 0$; see Erickson (1981) for $\mathscr{P}^{d,m}$ and Gaenssler (1983) for $\mathscr{E}^d$. An important class of sets that includes these last three examples are those known now as Vapnik–Červonenkis classes. For these, it is known (Dudley, 1978) that $N(\varepsilon, \mathscr{A}, d_\lambda) \leqq A\varepsilon^{-v}$ for some A and $v > 0$, where N is the (usual) metric entropy, like N_I but without the requirement of inclusion. In reasonably regular examples it may be possible to show a similar bound holds for N_I, in which case $r = 0$. Alexander (1984) proved necessary and sufficient conditions for the uniform central limit theorem when $\mathscr{A}$ is a Vapnik–Červonenkis class, and showed $N_I(\varepsilon, \mathscr{A}, d_\lambda) < \infty$ for all $\varepsilon > 0$ is a sufficient condition.

As we study the weak convergence of the smoothed partial-sum processes Z_n, it is clear that the limiting process $Z := \{Z(A)\colon A \in \mathscr{A}\}$ must necessarily be the Brownian process indexed by $\mathscr{A}$ which has mean zero and

$$\operatorname{cov}(Z(A), Z(B)) = |A \cap B|, \qquad A, B \in \mathscr{A}. \tag{2.2}$$

That this is so follows from the convergence of the finite-dimensional distributions which we now state and derive.

THEOREM 2.1. *If* $\{X_{\mathbf{j}}\colon \mathbf{j} \in \mathbb{Z}_+^d)$ *are iid with mean zero and variance* 1, *then the finite-dimensional distributions of* $\{Z_n(B)\colon B \in \mathscr{B}^d \cap I^d\}$ *converge weakly to those of* $\{Z(B)\colon B \in \mathscr{B}^d \cap K^d\}$.

PROOF. Let $B_1, \dots, B_m$ be any Borel subsets of I^d and let $a_1, \dots, a_m$ be any real numbers. We consider the asymptotic distribution of the general linear combination

$$\begin{aligned} V_n &:= \sum_{i=1}^{n} a_i Z_n(B_i) = n^{-d/2} \sum_{i=1}^{m} a_i \sum_{\mathbf{j}} \left|(nB_i) \cap C_{\mathbf{j}}\right| X_{\mathbf{j}} \\ &= n^{-d/2} \sum_{\mathbf{j}} \gamma_{n,\mathbf{j}} X_{\mathbf{j}} \end{aligned}$$

in which

$$\gamma_{n,\mathbf{j}} = \sum_{i=1}^{m} a_i \left|(nB_i) \cap C_{\mathbf{j}}\right|.$$

Note that $|\gamma_{n,\mathbf{j}}| \leqq \sum_{i=1}^m |a_i| := M < \infty$.

Write $\beta_n(\mathbf{t}) = \gamma_{n,\mathbf{j}}^2$ if $\mathbf{t} \in n^{-1}C_{\mathbf{j}}$. Then

$$\sigma_n^2 := \operatorname{var}(V_n) = \int_{I^d} \beta_n \, d\lambda$$

can be written as a Lebesgue integral. Since

$$\beta_n(\mathbf{t}) = \left\{ \sum_{i=1}^{m} a_i \frac{|B_i \cap n^{-1}C_{\mathbf{j}}|}{|n^{-1}C_{\mathbf{j}}|} \right\}^2 \quad \text{for } \mathbf{t} \in n^{-1}C_{\mathbf{j}}, \tag{2.3}$$

it follows from the d-dimensional Lebesgue density theorem [cf. Zaanen (1958, page 148)] applied to the ratios in (2.3) that

$$\beta_n \to \left\{ \sum_{i=1}^{m} a_i 1_{B_i} \right\}^2, \quad \text{a.e.} -\lambda.$$

Since β_n is nonnegative and bounded by M^2, the Lebesgue dominated convergence theorem shows that the limiting variance exists and equals

$$\begin{aligned} \sigma^2 &:= \lim_{n\to\infty} \sigma_n^2 = \int_{1^d} \left\{ \sum_{i=1}^{m} a_i 1_{B_i} \right\}^2 d\lambda \\ &= \sum_{i,k=1}^{m} a_i a_k |B_i \cap B_k| = \operatorname{var}\left(\sum_{i=1}^{m} a_i Z(B_i) \right). \end{aligned}$$

If $\sigma^2 = 0$, the asymptotic normality is trivially true. If $\sigma^2 > 0$, the Lindeberg condition for the central limit theorem is clearly satisfied since

$$\lim_{n \to \infty} E\left\{ n^{-d} \sum_{\mathbf{j}} \gamma_{n,\mathbf{j}}^2 X_{\mathbf{j}}^2 1_{[\gamma_{n,\mathbf{j}} |X_{\mathbf{j}}| > \varepsilon n^{d/2}]} \right\} = 0 \quad \text{for all } \varepsilon > 0$$

by boundedness of $|\gamma_{n,\mathbf{j}}|$ and finiteness of $EX_{\mathbf{1}}^2$. □

Let $C(\mathscr{A}) = C(\mathscr{A}, d_\lambda)$ denote the set of all continuous real-valued functions defined on $(\mathscr{A}, d_\lambda)$. It is clear that each $Z_n \in C(\mathscr{A})$ and it is known, Dudley (1973), that under our assumptions the Brownian process Z can be assumed to be $C(\mathscr{A})$-valued as well. The latter fact requires only that (2.1) hold with N (the usual metric entropy) in place of N_I. The stronger concept of total boundedness with inclusion is required here for our proof of the weak convergence, however. We do not know whether the weaker form of (2.1) would suffice. It is easy to construct examples of $\mathscr{A}$ in which (2.1) holds for N but not for N_I; e.g., take $\mathscr{A}$ to be the family of all subsets of I^d having zero Lebesgue measure. We do not know of any natural restrictions on $\mathscr{A}$ under which (2.1) holds when it holds with N replacing N_I.

It is convenient to introduce for any $\delta > 0$ the set of δ-caps,

$$\mathscr{C}_\delta := \{A \setminus B \colon A, B \in \mathscr{A} \text{ and } |A \setminus B| \leqq \delta\}.$$

Since

$$N_I(\varepsilon, \mathscr{C}_\delta, d_\lambda) \leqq N_I(\varepsilon/2, \mathscr{A}, d_\lambda)^2,$$

$\mathscr{C}_\delta$ also satisfies (2.1), so Z may be extended to $\mathscr{C}_\delta$. For any real-valued function f defined on S let $\| f \|_S := \sup_{s \in S} |f(s)|$ denote the supremum norm.

Here now is our main theorem.

THEOREM 2.2. *Let* $\{X_{\mathbf{j}} \colon \mathbf{j} \in \mathbb{Z}_+^d\}$ *be iid with* $EX_{\mathbf{j}} = 0$, $EX_{\mathbf{j}}^2 = 1$, *and suppose* $\mathscr{A}$ *satisfies the metric entropy condition* (2.1). *Then* Z_n *converges weakly to* Z.

3. Tightness and weak convergence. In view of the finite-dimensional result of Theorem 2.1, the weak convergence of the Z_n-processes will be proved once it is shown that their image laws are tight. Since each Z_n is continuous with respect to d_λ, the appropriate modulus of continuity is

$$\omega(Z_n, \delta) := \sup\{|Z_n(B) - Z_n(C)|;\, B, C \in \mathscr{A}, |B \Delta C| < \delta\}.$$

Consequently, $\omega(z_n, \delta) \leqq 2\|Z_n\|_{\mathscr{C}_\delta}$. Thus, in order to establish tightness, it suffices [cf. Billingsley (1968, page 55)] to show that for all $\varepsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P\left[\|Z_n\|_{\mathscr{C}_\delta} > \varepsilon\right] = 0. \tag{3.1}$$

Our approach to verifying (3.1) involves *symmetrization*, *truncation*, *stratification*, and *Gaussian domination*. It is fairly clear what ideas are referred to by the first three of these terms. The fourth concept refers to the idea of constructing a Gaussian process, with a related covariance structure to that of Z_n, which dominates Z_n stochastically in the sense of the expected value of the norm $\|\cdot\|_{\mathscr{C}_S}$. This idea was introduced in the related context of empirical processes by Giné and Zinn (1984). To be useful, however, the constructed Gaussian process to which Z_n is to be compared must be one for which the analogous statement to (3.1) is known to hold. The following three lemmas summarize the key steps in this approach. The ideas for the first two of these are contained in Lemma 2.9 of Giné and Zinn (1984). The first result is a simple consequence of Jensen's inequality. Let E^* denote upper expectation, i.e.,

$$E^*f := \inf\{Eg\colon g \geqq f,\, g \text{ measurable}\}.$$

LEMMA 3.1. *Let $\{f_j\colon j \in T\}$ be a finite set of real-valued functions defined on a space S. Let $\{V_j\colon j \in T\}$ be nonnegative random variables with $EV_j = 1$. Then*

$$\left\| \sum_{j\in T} f_j \right\|_S \leqq E^* \left\| \sum_{j\in T} V_j f_j \right\|_S .$$

PROOF. Use $|\sum_{j\in T}(EV_j)f_j(s)| \leqq E|\sum_{j\in T}V_j f_j(s)|$ for each $s \in S$. The upper expectation is needed only since the assumptions do not ensure the measurability of the arbitrary supremum $\|\cdot\|_S$. □

COROLLARY 3.1. *If in addition to the assumptions of Lemma 3.1, $\{\varepsilon_j\colon j \in T\}$ is a set of random variables independent of $\{V_j\colon j \in T\}$, then*

$$E\left\| \sum_{j\in T} \varepsilon_j f_j \right\|_S \leqq E\left\| \sum_{j\in T} \varepsilon_j V_j f_j \right\|_S , \tag{3.2}$$

provided these suprema are measurable.

PROOF. Apply Lemma 3.1 conditionally given $\{\varepsilon_j\colon j \in T\}$. □

With a little more care, the measurability assumption in Corollary 3.1 can be removed.

The application to be made below of this result is in the case of the ε_j's being ± 1 r.v.'s and the V_j's being the absolute values of Gaussian r.v.'s. The symmetrized partial-sum process will have the form of the sum on the left-hand side of (3.2) while the sum on the right-hand side will generate the desired Gaussian process. This in turn will be compared to a second Gaussian process about which the property (3.1) is known to hold. This step will require the following two results, the first of which is essentially stated in Giné and Zinn (1984) and based on a result of Fernique (1974).

LEMMA 3.2. *Let $\{Y_i(t), t \in S\}$, $i = 1, 2$, be centered Gaussian processes indexed by a countable set S, such that $0 \in \text{range}(Y_1)$ a.s. Suppose that for all $s, t \in S$,*

$$E(Y_1(t) - Y_1(s))^2 \leqq E(Y_2(t) - Y_2(s))^2.$$

Then

$$E\|Y_1\|_S \leqq 2E\|Y_2\|_S.$$

The requirement $0 \in \text{range}(Y_1)$ is needed to pass from the bounds on $E\sup(Y_1)$ and $E\sup(-Y_1)$ in Giné and Zinn (1984) to a bound on $E\|Y_1\|_S$. For our applications the index set S is a class of sets containing the empty set, so this requirement is always satisfied.

The final dominating process will be based on the Brownian process Z defined as for (2.2). The second result that is required is the one that provides the convergence to zero of $\|Z\|_{\mathscr{C}_\delta}$ for the dominating Brownian process as required for (3.1). Here, $N(\varepsilon, \mathscr{A}, d_\lambda)$ denotes usual metric entropy for $\mathscr{A}$ under d_λ. Clearly $N \leqq N_I$.

LEMMA 3.3. *There exists a universal constant K such that*

$$E\|Z\|_{\mathscr{C}} \leqq K\int_0^\delta \left(\varepsilon^{-1}\log N(\varepsilon, \mathscr{C}, d_\lambda)\right)^{1/2} d\varepsilon + K\delta$$

for all $\delta > 0$ and all classes $\mathscr{C}$ with $\|\lambda\|_{\mathscr{C}} \leqq \delta$.

PROOF. This follows directly from the estimates on $P[\omega(Z, \delta) > x]$ that are given in Theorem 2.1 of Dudley (1973). It is also a special case of Theorem 1.1 of Pisier (1983). □

The first step of the proof of tightness is one of symmetrization. Without loss of generality the $X_{\mathbf{j}}$'s may be assumed to be symmetric r.v.'s. This is an immediate consequence of the following symmetrization inequality that dates back at least to Vapnik and Červonenkis (1971). (Their result is stated for empirical processes, but the same technique applies here.) The proof uses the fact that

$$\operatorname{var} Z_n(A) \leqq |A| \leqq \delta \quad \text{for } A \in \mathscr{C}_\delta.$$

LEMMA 3.4. *Let $\{X_{\mathbf{j}}': \mathbf{j} \in \mathbb{Z}_+^d\}$ be an independent copy of $\{X_{\mathbf{j}}: \mathbf{j} \in \mathbb{Z}_+^d\}$, and let Z_n' be its corresponding smoothed partial-sum process. Then for every $\delta > 0$ and $M > 0$,*

$$P\left[\|Z_n\|_{\mathscr{C}_\delta} > M\right] \leqq 2P\left[\|Z_n - Z_n'\|_{\mathscr{C}_\delta} > M - (2\delta)^{1/2}\right].$$

Once we assume that the $X_{\mathbf{j}}$'s are symmetric, we may take them to have the particular form $X_{\mathbf{j}} = \varepsilon_{\mathbf{j}} Y_{\mathbf{j}}$ where $Y_{\mathbf{j}} \stackrel{L}{=} |X_{\mathbf{j}}|$ and $\{\varepsilon_{\mathbf{j}}: \mathbf{j} \in \mathbb{Z}_+^d\}$ is an independent array of independent r.v.'s with $P[\varepsilon_{\mathbf{j}} = 1] = P[\varepsilon_{\mathbf{j}} = -1] = \frac{1}{2}$.

The second step in the proof of Theorem 2.2 is to use the second moment assumption to permit us to truncate the summands $X_{\mathbf{j}}$. Since $EX_{\mathbf{1}}^2 < \infty$, there exists $\{\eta_n: n \geqq 1\}$ with $\eta_n \to 0$ such that

$$\lim_{n\to\infty} n^d P\left[X_{\mathbf{1}}^2 > \eta_n^2 n^d\right] = 0. \tag{3.3}$$

This permits truncation at $\eta_n n^{d/2}$, but we will need a slightly greater truncation. To do this, fix constants M, $\delta > 0$ and define

$$\gamma_n := \inf\{\gamma > 0: n^d P[X_1^2 > \gamma^2 n^d] < M\delta\eta_n^{-1}\} \wedge \eta_n. \tag{3.4}$$

Consider now the two partial-sum components of Z_n defined by

$$Z_n' := n^{-d/2}\sum_{\mathbf{j}} b_{n\mathbf{j}} X_{\mathbf{j}} 1_{[|X_{\mathbf{j}}| > \eta_n n^{d/2}]}$$

and

$$Z_n'' := n^{-d/2}\sum_{\mathbf{j}} b_{n\mathbf{j}} X_{\mathbf{j}} 1_{[\gamma_n n^{d/2} < |X_{\mathbf{j}}| \le \eta_n n^{d/2}]}.$$

By their construction it follows that

$$P\left[\|Z_n'\|_{\mathscr{C}_8} > M\right] \leqq P\left[|X_{\mathbf{j}}| > \eta_n n^{d/2} \text{ for some } \mathbf{j} \in nI^d\right] = o(1)$$

by (3.2) and, using Chebyshev's inequality, if $\gamma_n < \eta_n$ then

$$P\left[\|Z_n''\|_{\mathscr{C}_8} > M\right] \leqq M^{-1}E\|Z_n''\|_{\mathscr{C}_8} \leqq M^{-1}\eta_n \sum_{j \leqq n\mathbf{1}} P\left[|X_{\mathbf{j}}| > \gamma_n n^{d/2}\right] \leqq \delta.$$

It therefore follows that to prove (3.1) it remains to show that (3.1) holds when Z_n is replaced by the truncated partial-sum process

$$Z_n^T := Z_n - Z_n' - Z_n'' = n^{-d/2}\sum_{\mathbf{j}} b_{n\mathbf{j}} X_{\mathbf{j}} 1_{[|X_{\mathbf{j}}| \leqq \gamma_n n^{d/2}]}.$$

The third step in the proof is that of stratification, in which the interval $(0, \gamma_n n^{d/2}]$ is partitioned into sub-intervals $(a_k, a_{k+1}]$ with the levels $\{a_k\}$ defined for a fixed $\beta \in (0,1)$ by

$$a_k := \min\{x \geqq 0: P[|X_1| > x] \leqq \beta^{3k}\}.$$

Define

$$k_n := \max\{k: a_k < \gamma_n n^{d/2}\};$$

$$\begin{aligned} &J_k := (a_k, a_{k+1}], \quad 0 \leqq k < k_n; \quad J_{k_n} := (a_{k_n}, \gamma_n n^{d/2}]; \\ &p_k := \beta^{3k}, \quad \theta_k := a_{k+1} p_k^{1/2}, \quad 0 \leqq k \leqq k_n. \end{aligned} \tag{3.5}$$

By definition $P[|X_1| > a_k] \leqq p_k$. Using the representations $X_{\mathbf{j}} = \varepsilon_{\mathbf{j}} Y_{\mathbf{j}}$, we stratify Z_n^T according to

$$Z_n^T := \sum_{k \leqq k_n} \theta_k \nu_{nk}, \tag{3.6}$$

where

$$\nu_{nk}(A) := \left(n^d p_k\right)^{-1/2}\sum_{\mathbf{j}} b_{n\mathbf{j}}(A)\varepsilon_{\mathbf{j}} a_{k+1}^{-1} Y_{\mathbf{j}} 1_{[Y_{\mathbf{j}} \in J_k]}.$$

Observe that $a_{k+1}^{-1} Y_{\mathbf{j}} \leqq 1$ when $Y_{\mathbf{j}} \in J_k$ and that the expected number of $\mathbf{j} \in nI^d$ with $Y_{\mathbf{j}} \in J_k$ is at most $n^d p_k$. Hence the partial-sum process ν_{nk} should behave somewhat like a symmetrized empirical measure for the uniform law on I^d based on a sample of size $[n^d p_k]$. The only difference is that the random locations of the sample are restricted to the lattice points $\mathbf{j}/n \in I^d$, without replacement, and that the masses at those points are not exactly 1 but are $\leqq 1$. The representation (3.6) contains the underlying theme of this "vertical" stratification, namely, Z_n^T is written as a weighted sum of pseudo-empirical processes with coefficients that are square-summable. The finiteness of $\sum_{k=0}^{\infty}\theta_k^2$ follows easily from the finiteness of $EX_1^2 = \int_0^1 \bar{F}^{-1}(x)\,dx$, where $\bar{F}(x) = P[X_1^2 > x]$.

Define

$$H(\varepsilon) := \log N_I(\varepsilon, \mathscr{A}, d_\lambda) \quad \text{and} \quad f(\varepsilon) := \left(\varepsilon^{-1}H(\varepsilon)\right)^{1/2}.$$

Since the only facts we really use about $H(\varepsilon)$ are that it gives an upper bound for the number $\log N_I(\varepsilon, \mathscr{A}, d_\lambda)$ and that f is integrable, we may, by increasing H a little if necessary, assume that

(3.7) $\quad H$ is continuous and strictly decreasing, and $H(\varepsilon) \geqq 1 + \log \varepsilon^{-1}$.

Therefore the function f^{-1} is well defined. In order to study the supremum over $\mathscr{C}_\delta$ of each of the empirical-like processes ν_{nk}, the first step is to introduce a finite net of subsets that is suitable for approximating ν_{nk} at $A \in \mathscr{C}_\delta$ by its value at a set which is in the net and close to A. Set $\delta_{nk} := f^{-1}((n^d p_k)^{1/2}/4)$, so

$$n^d p_k = 16H(\delta_{nk})/\delta_{nk}. \tag{3.8}$$

In accordance with our assumptions of total boundedness with inclusion, let $\mathscr{D}_{nk}^{(i)}$, $i = 1, 2$, be finite nets of cardinality less than $\exp(2H(\delta_{nk}))$ such that for each $A \in \mathscr{C}_\delta$ there exists $D_{nk}^{(i)}(A) \in \mathscr{D}_{nk}^{(i)}$ satisfying

$$D_{nk}^{(1)}(A) \subset A \subset D_{nk}^{(2)}(A) \quad \text{and} \quad \left|D_{nk}^{(2)}(A) \setminus D_{nk}^{(1)}(A)\right| \leqq 2\delta_{nk}.$$

Recall that each element in $\mathscr{C}_\delta$ is a cap of the form $A_2 \setminus A_1$ for $A_1, A_2 \in \mathscr{A}$. Thus the $2\delta_{nk}$-net for the caps can be formed from caps $A \setminus B$ of pairs A, B in the δ_{nk}-net for $\mathscr{A}$. Hence the "2" in the exponent of the cardinality. Now by the additivity of each ν_{nk} write (3.6) as

$$\begin{aligned} Z_n^T(A) &= \sum_{k \leqq k_n} \theta_k \nu_{nk}\left(D_{nk}^{(1)}(A)\right) + \sum_{k \leqq k_n} \theta_k \nu_{nk}\left(A \setminus D_{nk}^{(1)}(A)\right) \\ &:= Z_n^{(1)}(A) + Z_n^{(2)}(A). \end{aligned} \tag{3.9}$$

First show that the "error" terms, $Z_n^{(2)}$, are negligible in the sense of (3.1). To this end set

$$\beta_{njk}(A) := b_{nj}\left(D_{nk}^{(2)}(A) \setminus D_{nk}^{(1)}(A)\right).$$

Then

$$\begin{aligned} \left|\nu_{nk}\left(A \setminus D_{nk}^{(1)}(A)\right)\right| &\leqq \left(n^d p_k\right)^{-1/2} \sum_{\mathbf{j}} \beta_{njk}(A) 1_{[Y_\mathbf{j} \in J_k]} \\ &\leqq \left(n^d p_k\right)^{-1/2} \sum_{\mathbf{j}} \beta_{njk}(A)\left\{1_{[Y_\mathbf{j} \in J_k]} - p_k'\right\} \\ &\quad + 2\delta_{nk}\left(n^d p_k\right)^{1/2}, \end{aligned} \tag{3.10}$$

where $p'_k := P[Y_j \in J_k] \leqq p_k$. The right-hand term is equal to $8(\delta_{nk}H(\delta_{nk}))^{1/2}$, so by (3.10), Bernstein's inequality [cf. Bennett (1962)], (3.8) and (3.7),

$$
\begin{aligned}
&P\Big[\sup_{\mathscr{C}_\delta} \big|\nu_{nk}\big(A \setminus D_{nk}^{(1)}(A)\big)\big| > 16\big(\delta_{nk}H(\delta_{nk})\big)^{1/2}\Big] \\
&\quad\leqq \#\big(\mathscr{D}_{nk}^{(1)}\big) \max_{\mathscr{C}_\delta} P\Big[n^{-d/2}\sum_{\mathbf{j}} \beta_{n\mathbf{j}k}(A)\big\{1_{[Y_{\mathbf{j}} \in J_k]} - p'_k\big\} \\
&\qquad\qquad > 8\big(\delta_{nk}H(\delta_{nk})\big)^{1/2} p_k^{1/2}\Big] \\
&\quad\leqq \exp\left(4H(\delta_{nk}) - \frac{64\delta_{nk}H(\delta_{nk})p_k}{2(2\delta_{nk}p_k + \delta_{nk}p_k/3)}\right) \\
&\quad\leqq \exp(-4H(\delta_{nk})) \\
&\quad\leqq \delta_{nk} \leqq \big(\delta_{nk}H(\delta_{nk})\big)^{1/2}.
\end{aligned} \tag{3.11}
$$

To sum the terms on the right side of (3.11)—which must be done several times in this proof—we use (3.8) and the metric entropy condition (2.1). Observe that by (3.4), $n^d p_{k_n} \geqq M\delta\eta_n^{-1} \to \infty$; therefore $\delta_{nk_n} \to 0$ by (3.8). This was the purpose of the second truncation, removing Z''_n, made above. Set $q_{nk} := (n^d p_k)^{1/2}/4 = f(\delta_{nk})$; then

$$
\begin{aligned}
\sum_{k \leqq k_n} \big(\delta_{nk}H(\delta_{nk})\big)^{1/2} &= \sum_{k \leqq k_n} q_{nk} f^{-1}(q_{nk}) \\
&= q_{nk_n} f^{-1}(q_{nk_n}) + \big(1 - \beta^{3/2}\big)^{-1} \\
&\quad \times \sum_{k \leqq k_n} (q_{nk} - q_{n,k+1}) f^{-1}(q_{nk}) \\
&\leqq q_{nk_n} f^{-1}(q_{nk_n}) + \big(1 - \beta^{3/2}\big)^{-1} \int_{q_{nk_n}}^{\infty} f^{-1}(x)\,dx \\
&\leqq \big(1 - \beta^{3/2}\big)^{-1} \int_0^{\delta_{nk_n}} f(\varepsilon)\,d\varepsilon \to 0 \quad \text{as } n \to \infty.
\end{aligned} \tag{3.12}
$$

Therefore

$$\sum_{k \leqq k_n} 16\theta_k\big(\delta_{nk}H(\delta_{nk})\big)^{1/2} \to 0 \quad \text{as } n \to \infty,$$

so by (3.6), (3.9), (3.11), and (3.12), one has that

$$P\Big[\|Z_n^{(2)}\|_{\mathscr{C}_\delta} > M\Big] \to 0 \quad \text{as } n \to \infty.$$

It remains now to consider $Z_n^{(1)}$, the simple function aproximating term. However, it must be modified slightly before we can show its uniform convergence in probability to zero. The problem is that some of the approximating sets $D_{nk}^{(1)}(A)$ may be too close together to permit us to obtain a suitable Gaussian approximation. To avoid this, let $\mathscr{C}_{nk} \subset \mathscr{C}_\delta$ be a maximal subset satisfying $|C_1 \Delta C_2| \geqq 2\delta_{nk}$ for all $C_1 \neq C_2$ in $\mathscr{C}_{nk}$. Then for each $A \in \mathscr{C}_\delta$, there is a member of $\mathscr{C}_{nk}$, denote it by $C_{nk}(A)$, which satisfies

(3.13) $\quad |C_{nk}(A)\,\Delta A| < 2\delta_{nk}$ and therefore $|C_{nk}(A)\,\Delta D_{nk}^{(1)}(A)| < 4\delta_{nk}$.

Now partition $Z_n^{(1)}$ as follows:

$$Z_n^{(1)}(A) = \sum_{k \leq k_n} \theta_k \nu_{nk}(C_{nk}(A)) + \sum_{k \leq k_n} \theta_k \{\nu_{nk}(D_{nk}^{(1)}(A)) - \nu_{nk}(C_{nk}(A))\}$$

$$:= Z_n^{(3)}(A) + Z_n^{(4)}(A).$$

By (3.13), a similar proof to that used on $Z_n^{(2)}$ suffices to show that

$$P\Big[\|Z_n^{(4)}\|_{\mathscr{C}_\delta} > M\Big] \to 0 \quad \text{as } n \to \infty;$$

simply use $\beta'_{njk} := b_{nj}(D_{nk}^{(1)}(A)\,\Delta C_{nk}(A))$ in placel of β_{njk}. It therefore remains to study the simple function approximation $Z_n^{(3)}$.

Write

$$Z_n^{(3)}(A) = \sum_{\mathbf{j}} \varepsilon_{\mathbf{j}} U_{n\mathbf{j}}(A)$$

with

$$U_{n\mathbf{j}}(A) := \sum_{k \leq k_n} n^{-d/2} b_{n\mathbf{j}}(C_{nk}(A)) Y_{\mathbf{j}} 1_{[Y_{\mathbf{j}} \in J_k]}.$$

Let $\{g_{\mathbf{j}}\colon \mathbf{j} \in \mathbb{Z}_+^d\}$ be independent standard normal r.v.'s independent of $\{\varepsilon_{\mathbf{j}}\colon \mathbf{j} \in \mathbb{Z}_+^d\}$. Let E_ε and E_g denote expectation with respect to the $\varepsilon_{\mathbf{j}}$'s and $g_{\mathbf{j}}$'s respectively; that is, conditional on the $Y_{\mathbf{j}}$'s. Let $\mu = 1/E|g_1|$. Since $\varepsilon_{\mathbf{j}}|g_{\mathbf{j}}| \overset{L}{=} g_{\mathbf{j}}$, set $V_{\mathbf{j}} = \mu|g_{\mathbf{j}}|$ in Corollary 3.1 to obtain

$$E_\varepsilon \|Z_n^{(3)}\|_{\mathscr{C}_\delta} \leq \mu E_g \|Z_n^g\|_{\mathscr{C}_\delta}, \tag{3.14}$$

where

$$Z_n^g(A) := \sum_{\mathbf{j}} g_{\mathbf{j}} U_{n\mathbf{j}}(A).$$

Notice that measurability of the norm is not a difficulty here since the supremum is essentially over the finite set $\bigcup_{k \leq k_n} \mathscr{C}_{nk}$.

Conditionally given $\{Y_{\mathbf{j}}\colon \mathbf{j} \in \mathbb{Z}_+^d\}$, Z_n^g is a Gaussian process. We wish to compare these Gaussian processes, via Lemma 3.2, to a "known" Gaussian process based on Z. Let $G^{(k)}$, $k \geq 1$, be the independent copies of Z, and define

$$G_n(A) := \sum_{k \leq k_n} 2\theta_k G^{(k)}(C_{nk}(A)),$$

$$Q_{nk}(A) := (n^d p_k)^{-1} \sum_{\mathbf{j}} b_{n\mathbf{j}}(A) 1_{[Y_{\mathbf{j}} \in J_k]},$$

$$\mathscr{E}_{nk} := \{E\,\Delta F\colon E \neq F \in \mathscr{C}_{nk}\}.$$

For $A, B \in \mathscr{C}_\delta$,

$$E_g(Z_n^g(A) - Z_n^g(B))^2 = \sum_{\mathbf{j}} (U_{n\mathbf{j}}(A) - U_{n\mathbf{j}}(B))^2$$
$$\leqq \sum_{k \leqq k_n} n^{-d} \sum_{\mathbf{j}} \left(b_{n\mathbf{j}}(C_{nk}(A)) - b_{n\mathbf{j}}(C_{nk}(B))\right)^2 a_{k+1}^2 1_{[Y_{\mathbf{j}} \in J_k]}$$
$$\leqq \sum_{k \leqq k_n} \theta_k^2 Q_{nk}(C_{nk}(A)\, \Delta C_{nk}(B)).$$

Also

$$E(G_n(A) - G_n(B))^2 = \sum_{k \leqq k_n} 4\theta_k^2 |C_{nk}(A)\, \Delta C_{nk}(B)|.$$

Thus on the event

$$L_n := [Q_{nk}(A) \leqq 4|A| \text{ for all } A \in \mathscr{E}_{nk} \text{ and } k \leqq k_n]$$

one obtains

$$E_g(Z_n^g(A) - Z_n^g(B))^2 \leqq E(G_n(A) - G_n(B))^2.$$

Therefore by Lemma 3.2, on L_n,

$$E_g \|Z_n^g\|_{\mathscr{C}_\delta} \leqq 2E\|G_n\|_{\mathscr{C}_\delta}. \tag{3.15}$$

To bound $E\|G_n\|_{\mathscr{C}_\delta}$, begin by defining

$$G_{nk} := \sum_{l=k}^{k_n} 2\theta_l G^{(l)}, \qquad G_n'(A) := G_{no}(C_{no}(A)),$$
$$W_{nk}(A) := G_{nk}(C_{nk}(A)) - G_{nk}(C_{n,k-1}(A)),$$
$$v_{nk} := 2\left(\sum_{l=k}^{k_n} \theta_l^2\right)^{1/2}, \qquad v := \left(\sum_{l \geqq 0} \theta_l^2\right)^{1/2}.$$

Observe that, by changing the order of summation,

$$\begin{aligned} G_n(A) &= \sum_{k \leqq k_n} 2\theta_k G^{(k)}(C_{no}(A)) + \sum_{k \leqq k_n} \sum_{i=1}^{k} 2\theta_k \\ &\qquad \times \{G^{(k)}(C_{ni}(A)) - G^{(k)}(C_{n,i-1}(A))\} \\ &= G_n'(A) + \sum_{k=1}^{k_n} W_{nk}(A), \end{aligned} \tag{3.16}$$

and note that G_{nk} has the same law as $v_{nk} G^{(no)}$. Therefore

$$E\|G_n'\|_{\mathscr{C}_\delta} \leqq E\|G_{no}\|_{\mathscr{C}_\delta} \leqq vE\|Z\|_{\mathscr{C}_\delta}. \tag{3.17}$$

To bound $E\|W_{nk}\|_{\mathscr{C}_\delta}$, first observe that $W_{nk}(A)$ is Gaussian, with $EW_{nk}^2(A) \leqq 8v^2\delta_{nk}$ by (3.13). Therefore for all $t \geqq (16\beta^{-3}H(\delta_{nk}))^{1/2}$,

$$\begin{aligned} P\Big[\|W_{nk}\|_{\mathscr{C}_\delta} > t\big(8v^2\delta_{nk}\big)^{1/2}\Big] &\leqq \#(\mathscr{C}_{nk})\#(\mathscr{C}_{n,k-1})t^{-1}\exp\big(-t^2/2\big) \\ &\leqq \exp\big(-\big(t^2 - 8H(\delta_{n,k-1})\big)/2\big) \\ &\leqq \exp\big(-\big(t^2 - 8\beta^{-3}H(\delta_{nk})\big)/2\big) \\ &\leqq \exp\big(-t^2/4\big). \end{aligned}$$

It follows that

$$E\|W_{nk}\|_{\mathscr{C}_\delta} \leqq \big(16v^2\delta_{nk}\big)^{1/2}\Big(\big(16\beta^{-3}H(\delta_{nk})\big)^{1/2} + 4\Big),$$

so $\sum_{k=1}^{k_n} E\|W_{nk}\|_{\mathscr{C}_\delta} \to 0$ as $n \to \infty$ by (3.12). Combining this with (3.16) and (3.17), we see that

$$E\|G_n\|_{\mathscr{C}_\delta} \leqq vE\|Z\|_{\mathscr{C}_\delta} + o(1).$$

Hence using (3.14) and (3.15), one has for large n that

$$P\Big[\|Z_n^{(3)}\|_{\mathscr{C}_\delta} > M\Big] \leqq P(L_n^c) + 2M^{-1}\mu vE\|Z\|_{\mathscr{C}_\delta} + o(1).$$

By Lemma 3.3, then, the proof will be finished when we show that

$$P(L_n^c) \to 0 \quad \text{as } n \to \infty.$$

Much as ν_{nk} behaves like a symmetrized empirical process, Q_{nk} behaves like an unsymmetrized empirical measure. In fact, by Bernstein's inequality and (3.8), since $EQ_{nk}(A) \leqq |A|$ and $|A| \geqq 2\delta_{nk}$ for all $A \in \mathscr{E}_{nk}$,

$$\begin{aligned} P(L_n^c) &\leqq \sum_{k \leqq k_n} \#(\mathscr{E}_{nk}) \max_{\mathscr{E}_{nk}} P\big[Q_{nk}(A) > 4|A|\big] \\ &\leqq \sum_{k \leqq k_n} \big(\#(\mathscr{D}_{nk}^{(1)})\big)^2 \max_{\mathscr{E}_{nk}} P\Big[n^{-d}\sum_{\mathbf{j}} b_{n\mathbf{j}}(A)\big(1_{[Y_{\mathbf{j}} \in J_k]} - p_k'\big) > 3|A|p_k\Big] \\ &\leqq \sum_{k \leqq k_n} \exp(4H(\delta_{nk})) \max_{\mathscr{E}_{nk}} \exp\big(-9|A|n^d p_k/4\big) \\ &\leqq \sum_{k \leqq k_n} \exp(-68H(\delta_{nk})) \end{aligned}$$

which approaches 0 as $n \to \infty$, as in (3.11) and (3.12). □

4. Remarks. 1. With only minor modifications, the results of this paper hold for independent arrays that are not iid. If $\{X_{n\mathbf{j}}\colon n \geqq 1, \mathbf{j} \leqq n\mathbf{1}\}$ is a "pyramidal array" of random variables that are independent for each fixed n and satisfy $EX_{n\mathbf{j}} = 0$ and $\sigma_{n\mathbf{j}}^2 := EX_{n\mathbf{j}}^2 < \infty$, construct the partial-sum process as $Z_n(A) = \sum_{\mathbf{j}} b_{n\mathbf{j}}(A)X_{n\mathbf{j}}$. The proof of tightness can be carried through with little change if we assume

(4.1) the distribution functions $F_n(t) := \min_{\mathbf{j} \leqq n\mathbf{1}} P[n^{d/2}|X_{n\mathbf{j}}| \leqq t]$ have second moments which are bounded in n.

The values $a_k = F_n^{-1}(1 - \beta^{3k})$ and $\theta_k = a_k p_k^{1/2}$ to be used in the proof will now depend on n, but (4.1) ensures that $\Sigma_k \theta_k^2$ remains bounded in n. The assumption (4.1) also implies a Lindeberg-type condition, and the proof of finite-dimensional convergence goes through as before under one additional assumption: define

$$\gamma_n(\mathbf{t}) = n^d \sigma_{n\mathbf{j}}^2 \quad \text{for } \mathbf{t} \in n^{-1} C_{\mathbf{j}}.$$

Observe that $\gamma_n = d\Lambda_n / d\lambda$, the density of the variance measure Λ_n defined by $\Lambda_n(B) = \Sigma_{\mathbf{j}} b_{n\mathbf{j}}(B) \sigma_{n\mathbf{j}}^2$. We assume that $\{\gamma_n\}$ converges in L^1 to a density γ of a measure Λ, and this is equivalent to saying that $\Lambda_n(B) \to \Lambda(B)$ uniformly for Borel sets B [cf. Pyke (1983, Section 3)]. The process Z will then be the mean zero Gaussian process with covariance

$$EZ(A)Z(B) = \int_{A \Delta B} \gamma(\mathbf{t})\, d\mathbf{t} = \Lambda(A \Delta B). \tag{4.2}$$

This is summarized as follows:

THEOREM 4.1. *Let $\{X_{n\mathbf{j}}\colon n \geq 1, \mathbf{j} \leq n\mathbf{1}\}$ be a pyramidal array, independent for fixed n, with $EX_{n\mathbf{j}} = 0$ and $EX_{n\mathbf{j}}^2 < \infty$, and suppose $\mathscr{A}$ satisfies the metric entropy condition* (2.1). *Let Z_n and γ_n be as above. Suppose* (4.1) *holds and γ_n converges in L^1 to some function γ. Then Z_n converges weakly to a sample-continuous Gaussian process Z with covariance given by* (4.2).

When $\mathscr{A}$ is a Vapnik–Červonenkis class, the non-iid case, in greater generality than the above, presents essentially no difficulties beyond the iid case–see Alexander (1984) where necessary and sufficient conditions for the uniform CLT are obtained in this case.

2. The smoothing in (1.2) and (1.3) of the partial-sum process has been done to ensure that the entropy assumption (2.1) alone, without additional restrictions (e.g., on the smoothness of the boundaries of the sets), is sufficient to ensure that the uniform CLT holds. The particular smoothing used also insures that the processes are in $C(\mathscr{A})$, a space in which the questions of weak convergence are more easily formulated and studied. Also, smoothing ensures that $\operatorname{var}(Z_n(A) - Z_n(B))$ is related to $\lambda(A \Delta B)$, which is necessary if metric entropy considerations are to be a useful tool. Smoothing is, of course, natural, since in many applications the measurement $X_{\mathbf{j}}$ located at the grid point $\mathbf{j}$ does in reality represent a measure of some quantity present over the corresponding region $C_{\mathbf{j}}$. In any event, without smoothing, it is known that there are some important families $\mathscr{A}$ such that far too many sub-sums of the random variables $\{X_{\mathbf{j}}\colon \mathbf{j} \leq n\mathbf{1}\}$, would be of the form $S_n(A)$ for some $A \in \mathscr{A}$, cf. Erickson (1981) and Pyke (1983). Under additional restrictions on $\mathscr{A}$ it would be possible to weaken or even drop the smoothing operation. For example, if the sets in $\mathscr{A}$ have structural restrictions on their boundaries, such as is true for convex sets, then smoothing would not be necessary. Furthermore, the particular partition $\{C_{\mathbf{j}}\}$ used in this paper for the smoothing can also be generalized. Rather than distributing the mass $X_{\mathbf{j}}$ uniformly over a cube $C_{\mathbf{j}}$ it could be distributed according to other measures over the $\mathbf{j}$th set of a partition. These types of generalizations are used in Morrow and

Philipp (1984). A particular nonuniform smoothing is used in Bass and Pyke (1985) in situations in which the limiting process is a non-Gaussian Lévy process. It should be noted that no smoothing is required when $\mathscr{A}$ is a Vapnik–Červonenkis class; this is shown in Alexander (1984).

3. Although we have considered only the case of set-indexed partial-sum processes, similar results may also be obtained for partial-sum processes indexed by functions. If F is a family of real-valued functions defined on I^d, define $S_n = \{S_n(f): f \in \mathscr{F}\}$ where

$$S_n(f) = \int f\, dS_N = n^{-d/2} \sum_{\mathbf{j} \leq n\mathbf{1}} f(\mathbf{j}/n) X_{\mathbf{j}}.$$

The smoothed versions would be defined in terms of $\int f\, dZ_n$.

4. In a forthcoming paper, Bass (1984) will derive a functional law of the iterated logarithm (LIL) for partial-sum processes. His methods give an alternate approach to our uniform CLT which does not involve comparison with a dominating Gaussian process. It would be of interest to know whether or not our method can be refined to yield the sharper estimates necessary for the LIL.

Acknowledgments. We greatly appreciate the helpful comments of the referee and the associate editor.

REFERENCES

ALEXANDER, K. S. (1984). Central limit theorems for stochastic processes under random entropy conditions. To appear.

BASS, R. F. (1984). A law of the iterated logarithm for set-indexed partial-sum processes with finite variance. To appear.

BASS, R. F. and PYKE, R. (1984). Functional law of the iterated logarithm and uniform central limit theorem for partial-sum processes indexed by sets. *Ann. Probab* **12** 13–34.

BASS, R. F. and PYKE, R. (1985). The space $D(A)$ and weak convergence for set-indexed processes. *Ann. Probab.* **13** 860–884.

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

DONSKER, M. D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6** 1–12.

DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.

DUDLEY, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approx. Theory* **10** 227–236.

ERICKSON, R. V. (1981). Lipschitz smoothness and convergence with applications to the central limit theorem for summation processes. *Ann. Probab.* **9** 831–851.

FERNIQUE, X. (1974). Regularité des trajectoires des functions aléatoires gaussiennes. *Lecture Notes in Math.* **480** 1–96.

GAENSSLER, P. (1983). Limit theorems for empirical processes indexed by classes of sets allowing a finite-dimensional parametrization. *Probab. Math. Statist.* **4** 1–12.

GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.

GOLDIE, C. M. and GREENWOOD, P. (1984). Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes. *Ann. Probab.* To appear.

KUELBS, J. (1968). The invariance principle for a lattice of random variables. *Ann. Math. Statist.* **39** 382–389.

MORROW, G. J. and PHILIPP, W. (1984). Invariance principles for partial sum processes and empirical processes indexed by sets. To appear.

OSSIANDER, M. (1984). Weak convergence of partial-sum processes indexed by points in a metric space. Ph.D. dissertation, Univ. Washington.

OSSIANDER, M. and PYKE, R. (1984). Lévy's Brownian motion as a set-indexed process and related central limit theorems. Technical Report 42, Dept. Math., Univ. Washington.

PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Lecture Notes in Math.* **995** 123–154.

PYKE, R. (1983). A uniform central limit theorem for partial-sum processes indexed by sets. In *Probability, Statistics, and Analysis.* (J. F. C. Kingman and G. E. H. Reuter, eds.) 219–240. Cambridge Univ. Press.

RÉVÉSZ, P. (1976). Three theorems of multivariate empirical processes. *Lecture Notes in Math.* **566** 106–126.

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

WICHURA, M. J. (1969). Inequalities with applications to the weak convergence of random processes with multidimensional time parameters. *Ann. Math. Statist.* **40** 681–687.

ZAANEN, A. C. (1958). *An Introduction to the Theory of Integration.* North Holland, Amsterdam.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195