The Annals of Probability
1989, Vol. 17, No. 1, 266–291

STRONG APPROXIMATION FOR MULTIVARIATE EMPIRICAL AND RELATED PROCESSES, VIA KMT CONSTRUCTIONS

BY PASCAL MASSART

Université Paris–Sud

Let P be the Lebesgue measure on the unit cube in $\mathbb{R}^d$ and Z_n be the centered and normalized empirical process associated with n independent observations with common law P.

Given a collection of Borel sets $\mathscr{S}$ in $\mathbb{R}^d$, it is known since Dudley's work that if $\mathscr{S}$ is not too large (e.g., either $\mathscr{S}$ is a Vapnik–Červonenkis class (VČ class) or $\mathscr{S}$ fulfills a suitable "entropy with bracketing" condition), then (Z_n) may be strongly approximated by some sequence of Brownian bridges indexed by $\mathscr{S}$, uniformly over $\mathscr{S}$ with some rate b_n.

We apply the one-dimensional dyadic scheme previously used by Komlós, Major and Tusnády (KMT) to get as good rates of approximation as possible in the above general multidimensional situation. The most striking result is that, up to a possible power of $\log(n)$, b_n may be taken as $n^{-1/2d}$ which is the best possible rate, when $\mathscr{S}$ is the class of Euclidean balls (this is the KMT result when $d = 1$ and the lower bounds are due to Beck when $d \geq 2$). We also obtain some related results for the set-indexed partial-sum processes.

1. Introduction.

Definitions and notation. Let P be the Lebesgue measure on $[0,1]^d$ and $x_1, x_2, \dots$ be a sequence of independent random variables with common distribution P defined on a "rich enough" probability space $(\Omega, \mathscr{X}, \Pr)$.

Let P_n denote the empirical measure associated with $(x_1, \dots, x_n)$, $P_n = 1/n\sum_{i=1}^n \delta_{x_i}$, $P_0 = 0$. We call the empirical process the centered and normalized process $Z_n = \sqrt{n}(P_n - P)$.

NOTATION. Given a set $\mathscr{T}$ and a function $f\colon \mathscr{T} \to \mathbb{R}$, let $\bigvee_{\mathscr{T}} f$ denote the supremum of f over $\mathscr{T}$.

Let $\mathscr{L}_P^2(\mathbb{R}^d)$ be given the metric $\rho_P\colon (f, g) \to (P((f-g)^2) - (P(f-g))^2)^{1/2}$ and $\mathscr{S}$ be a collection of Borel sets in $\mathbb{R}^d$, which is totally bounded for ρ_P.

DEFINITION 1. A Brownian bridge indexed by $\mathscr{S}$ is a centered Gaussian process indexed by $\mathscr{S}$ with covariance function $(S, S') \to P(S \cap S') - P(S)P(S')$.

We say that the strong invariance principle (respectively, strong Kiefer invariance principle) holds for $\mathscr{S}$ with rate (b_n) if there exists some sequence $(B_n)_{n \geq 1}$ of versions (respectively, independent versions) of a Brownian bridge indexed by $\mathscr{S}$ that are almost surely uniformly continuous on $(\mathscr{S}, \rho_P)$ such that

$$\bigvee_{\mathscr{S}} |Z_n - B_n| = O(b_n) \quad \text{a.s.,}$$

Received March 1987; revised April 1988.

AMS 1980 *subject classifications*. Primary 60F17; secondary 60G50, 60G60, 62E20.

Key words and phrases. Brownian bridge, empirical process, invariance principle, metric entropy, Poissonization, set-indexed process, Vapnik–Červonenkis class.

respectively,

$$\bigvee_{\mathscr{S}}\left|Z_n - \frac{1}{\sqrt{n}}\sum_{j=1}^{n} B_j\right| = O(b_n) \quad \text{a.s.}$$

In what follows the sequences (b_n) of the above definition will always decrease to zero, but for Kiefer type approximations, denoting from now on by L the function $x \to \log(\max(x, e))$, the case where $b_n = o(\sqrt{LLn})$ is also of interest (because when the strong Kiefer invariance principle holds with such a rate, any kind of law of the iterated logarithm which holds for partial sums of Gaussian processes may then be transferred to the empirical process).

Let us review the main attempts of getting as good rates of approximation as possible in the strong invariance principle for empirical processes.

Some bibliography. Let $\mathscr{Q}^{(d)}$ denote the class of quadrants in $\mathbb{R}^d$, that is,

$$\mathscr{Q}^{(d)} = \left\{\prod_{j=1}^{d}\,]-\infty, u_j];\ u \in \mathbb{R}^d\right\}.$$

The sharpest known results about this classical class are the following:

1.1. The KMT results [see KMT (1975) and Tusnády (1977)]. The strong invariance principle holds for $\mathscr{Q}^{(1)}$ with rate $Lnn^{-1/2}$ which is the best possible one. Moreover the strong invariance principle (respectively, Kiefer invariance principle) holds for $\mathscr{Q}^{(2)}$ (respectively, for $\mathscr{Q}^{(1)}$) with rate $Ln^2 n^{-1/2}$.

1.2. Results of Csörgő and Révész (1975). For any $d \geq 3$, the strong invariance principle (respectively, Kiefer invariance principle) holds for $\mathscr{Q}^{(d)}$ (respectively, for $\mathscr{Q}^{(d-1)}$) with rate $Ln^{3/2}n^{-1/2(d+1)}$ (respectively, with rate $Ln^2 n^{-1/2(d+1)}$). Note that Révész [see Lemma 11 of Révész (1976a)] showed that the same rate $Ln^{3/2}n^{-1/6}$ remains valid when $d = 2$ and replacing $\mathscr{Q}^{(2)}$ with the VČ class of convex polygons with no more than K vertices.

Some other classes of sets were also studied by Révész [see Révész (1976a), (1976b)].

1.3. Révész's work on classes of sets with smooth boundaries. For the class of some Borel sets of $\mathbb{R}^d$ with $(d + \rho)$-regular boundaries (with $0 < \rho \leq 1$), Révész proved that the strong invariance principle (respectively, Kiefer invariance principle) holds with rate $n^{-\rho/12(d+1)+\varepsilon}$ (respectively, $n^{-\rho/12(d+1)+2\rho+\varepsilon}$).

Let us finally mention:

1.4. Some general results about rates of convergence. Dudley and Philipp (1983) gave entropy conditions on a class $\mathscr{S}$ for the strong Kiefer invariance principle to hold with rate b_n, where b_n is some negative power of either n or Ln (in that framework $\mathbb{R}^d$ may be replaced with a general measurable space and P is just any probability law in the Vapnik–Červonenkis case), but of course, in such a general situation, the rates b_n are far from being optimal.

Note that, refining their method, we proved [see Massart (1986)] that b_n may be taken as $n^{-1/(18+20d(\mathscr{S})+\varepsilon)}$ in the case where $\mathscr{S}$ is a VČ class with real density $d(\mathscr{S})$ [in the sense of Assouad (1983)] fulfilling some mild measurability condition.

Statement of results. First we need some more definition and notations. From now on $\mathscr{S}$ is a class of Borel sets of $\mathbb{R}^d$.

NOTATION. Given a norm $|\cdot|$ on $\mathbb{R}^d$, let A^ε denote the set $A^\varepsilon = \{y \in \mathbb{R}^d: |y - z| < \varepsilon \text{ for some } z \in A\}$ for any $\varepsilon \in]0, 1]$.

We shall need the following smoothness condition on the boundaries of the sets belonging to $\mathscr{S}$.

DEFINITION 2. The class $\mathscr{S}$ is said to fulfill the uniform Minkowski condition [condition (UM)] if and only if there exists some constant K such that

$$P\big((\partial S)^\varepsilon\big) \leq K\varepsilon \quad \text{for any } \varepsilon \in]0,1] \text{ and any } S \in \mathscr{S}.$$

COMMENTS. Of course the above definition does not depend on the given norm on $\mathbb{R}^d$. Two classes of interest fulfill the uniform Minkowski condition:

(a) The class of convex sets [this is essentially the Steiner–Minkowski theorem; see Dudley (1982), page 62].

(b) The class of Borel sets with uniformly Lipschitz boundaries in the sense of Dudley (1974).

Condition UM will play an essential part in our forthcoming constructions. Assuming condition UM is some kind of loss of generality with respect to the framework of Dudley and Philipp (1983), but however it turns out that many interesting geometric examples of P-Donsker classes fulfill condition UM.

We need to introduce some reasonable growth conditions on $\mathscr{S}$. We shall make two kinds of assumption on $\mathscr{S}$ in what follows. We denote the first one by $H(0)$: $\mathscr{S}$ is a VČ class. We recall that this means

$$\sup\{n \in \mathbb{N} \,|\, |A \cap \mathscr{S}| = 2^n \text{ for some set } A \text{ with cardinality } n\} < \infty,$$

where $A \cap \mathscr{S} = \{A \cap S, S \in \mathscr{S}\}$ and $|E|$ denotes from now on the cardinality of the set E. Moreover $\mathscr{S}$ fulfills the mild measurability condition ($\mathscr{M}$): There exist some Suslin space Y and some mapping T from Y onto $\mathscr{S}$ such that $(x, y) \to \mathbb{1}_{T(y)}(x)$ is measurable on $\mathbb{R}^d \times Y$.

We denote the second one [which is the entropy with bracketing condition used in Dudley (1978) and Dudley and Philipp (1983)], by

$$H(\zeta), \text{ with } 0 < \zeta < 1: N_{[\ ]}(\varepsilon, \mathscr{S}) \leq \exp(K\varepsilon^{-\zeta})$$

for some constant K and any ε in $]0, 1]$, where $N_{[\ ]}(\varepsilon, \mathscr{S})$ denotes the minimal cardinality of a collection $\mathscr{S}(\varepsilon)$ of Borel sets such that for any S in $\mathscr{S}$ there exist $S^-(\varepsilon), S^+(\varepsilon)$ in $\mathscr{S}(\varepsilon)$ with $S^-(\varepsilon) \subseteq S \subseteq S^+(\varepsilon)$ and $P(S^+(\varepsilon) \setminus S^-(\varepsilon)) \leq \varepsilon$.

EXAMPLE 1. The class $\mathscr{Q}^{(d)}$, the class of balls in $\mathbb{R}^2$ (associated with any norm), the class of balls in $\mathbb{R}^d$ associated with some polynomial norm (in particular an Euclidean norm is convenient), the class of ellipsoids and the class of half spaces do all fulfill assumption $H(0)$. [See Assouad (1983) for many examples and properties of VČ classes.]

EXAMPLE 2. In another connection the class of convex sets in $\mathbb{R}^2$ fulfills $H(1/2)$ [see Dudley (1982)] and the class of Borel sets whose boundaries have uniformly bounded derivatives of orders up to α for some positive α with $\alpha > d - 1$ [in the sense of Dudley (1974) or Révész (1976b)] fulfills $H((d-1)/\alpha)$.

It follows from the above comment that any of these classes fulfills the uniform Minkowski condition, too.

Moreover it is worth noticing that if $\mathscr{S}$ fulfills $H(\zeta)$ for some $0 \le \zeta < 1$ and condition UM, then so does the class of sets resulting from at most k Boolean operations with arguments in $\mathscr{S}$ for some fixed k.

Throughout the paper d is an integer greater than or equal to 2; moreover, unless we give some other specifications, the space Ω is assumed to be rich enough in the following sense: There exists a random variable, defined on Ω, with uniform distribution over $[0, 1]$, which is independent of the observations.

THEOREM 1 (Strong approximation of the empirical process). *Let $\mathscr{S}$ fulfill the uniform Minkowski condition and assumption $H(\zeta)$ for some $0 \le \zeta < 1$. We set $a_n(\zeta, d) = \sqrt{Ln}\, n^{-1/2d}$ when $\zeta = 0$ and $a_n(\zeta, d) = n^{-(1-\zeta)/2d}$ when $\zeta > 0$.*

Then there exist some constants C, Λ and θ (depending only on $\mathscr{S}$) and, for any integer n, some version $Z^{(n)}$ of a Brownian bridge indexed by $\mathscr{S}$ which is almost surely uniformly continuous on $(\mathscr{S}, \rho_P)$ such that for any positive t

$$\Pr{}^*\Big(\bigvee_{\mathscr{S}} |Z_n - Z^{(n)}| \ge a_n(\zeta, d)(t + CLn)\Big) \le \Lambda e^{-\theta t},$$

where $\Pr^*$ *stands for the outer measure associated with* Pr.

An immediate consequence of Theorem 1 is

COROLLARY 1. *Let $\mathscr{S}$ fulfill condition UM and assume that $H(\zeta)$ holds for some $0 \le \zeta < 1$, then the strong invariance principle holds for $\mathscr{S}$ with rate $Lna_n(\zeta, d)$, where $a_n(\zeta, d)$ is defined as in Theorem* 1.

The corresponding results concerning Kiefer type approximations are stated below.

THEOREM 2 (Kiefer-type approximation of the empirical process). *Let $\mathscr{S}$ fulfill the uniform Minkowski condition and assumption $H(\zeta)$ for some $0 \le \zeta < 1$. Then, there exist some sequence $(B_j)_{j \ge 1}$ of independent versions of a Brownian bridge indexed by $\mathscr{S}$ which are almost surely uniformly continuous on $(\mathscr{S}, \rho_P)$*

and some constants C, Λ and θ (depending only on $\mathscr{S}$) such that for any positive t

$$\Pr{}^*\left(\bigvee_{\mathscr{S}}\left|Z_n - \frac{1}{\sqrt{n}}\sum_{j=1}^{n} B_j\right| \geq b_n(\zeta, d)(t + CLn)^{3/2}\right) \leq \Lambda e^{-\theta t},$$

where $b_n(\zeta, d) = a_n(\zeta, d+1)$ whenever $d \geq 3$ or $\zeta = 0$ [with $a_n(\cdot,\cdot)$ defined as in Theorem 1] and $b_n(\zeta, 2) = n^{-(1-\zeta)/(2(3+\zeta))}$ if $\zeta > 0$.

COROLLARY 2. *Let $\mathscr{S}$ fulfill condition UM and assume that $H(\zeta)$ holds for some $0 \leq \zeta < 1$. Then the strong Kiefer invariance principle holds for $\mathscr{S}$ with rate $(Ln)^{3/2}b_n(\zeta, d)$, where $b_n(\cdot,\cdot)$ is defined in Theorem 2.*

COMMENT. Corollaries 1 and 2 generalize and improve on 1.2 in the case where $\zeta = 0$. Moreover, Corollary 1 is optimal up to a power of Ln in the case where either $\mathscr{S}$ is the collection of Euclidean balls or $\mathscr{S}$ is the collection of half spaces and $d = 2$ [see Beck (1985) for the lower bounds].

In another connection, Corollaries 1 and 2 generalize 1.3 and improve on 1.3 when $d \leq 6$ [in that case, our result could be refined by using the oscillation control of Révész (1976b) instead of that of Alexander (1984) which we use in Lemma 1 below]. Note, in particular, that the strong invariance principle holds with rate $Lnn^{-1/8}$ for the class of convex sets in $\mathbb{R}^2$ as well as for the class of Borel sets with twice differentiable boundaries in $\mathbb{R}^2$.

It is worth noticing that Borisov has already used [see Borisov (1982)] KMT constructions to prove multidimensional invariance principles; the rates that he got are less efficient than those given in Theorem 1, but hold for more general P (for instance, for all P the rate $Ln^2 n^{-1/2(2d-1)}$ is valid for the class $\mathscr{Q}^{(d)}$). Also, as far as the class $\mathscr{Q}^{(d)}$ is concerned (in the case where P is uniform), the rate $Ln^d n^{-1/2}$ has been announced by Borisov in (1981) (unfortunately no proof of this striking result has been yet published as far as we know); as a matter of fact, in that case, our corollaries could easily be improved. However, since there is no hope of obtaining a rate such as $Ln^\alpha n^{-1/2}$ when using the classical Poissonization technique, we shall not detail this improvement here.

Sections 2, 3 and 4 are devoted to the proofs of Theorems 1 and 2. In Section 5 we use the results of Section 3 to prove an analogue of Theorem 1 for set-indexed partial sum processes [a good introduction to the study of set-indexed processes is given by Pyke (1984)].

Until the end of Section 4, $\mathscr{S}$ is supposed to fulfill condition UM and assumption $H(\zeta)$ for some $0 \leq \zeta < 1$.

2. Approximation with finite regular classes and Poissonization. Given an integer ν, let $\mathscr{B}_\nu^{(d)}$ be the σ-algebra generated by the cubes

$$C_\nu^{(i)} = \left\{y \in \mathbb{R}^d/(i_j - 1)\nu^{-1} < y_j \leq i_j\nu^{-1} \text{ for all } j\right\}, \qquad i \in \mathbb{Z}_+^d \cap [0, \nu]^d,$$

where $\mathbb{Z}_+$ denotes the set of strictly positive integers. When there is no confusion to be feared we shall simply write $\mathscr{B}_\nu$ instead of $\mathscr{B}_\nu^{(d)}$.

NOTATION. For any Borel set B in $\mathscr{B}_\nu$, we call the *perimeter* of B the hypersurface area of its boundary. Let $s_{d-1}(B)$ denote this quantity. Given a Borel set A, let A_ν be the set of $\mathscr{B}_\nu$ composed with those atoms of $\mathscr{B}_\nu$ which intersect A. The essential part (i.e., the oscillation control) of Lemma 1 derives from the Bernstein type inequalities which were proved by Alexander (1984) and by Massart (1986).

LEMMA 1. *There exists a map* Π_ν: $\mathscr{S} \to \mathscr{B}_\nu$ *such that, denoting by* $\mathscr{S}_\nu$ *the range of* Π_ν, *the following three properties hold*:

(2.1) *Moderate growth of the entropy.*

$$|\mathscr{S}_\nu| \le \exp\left(C_0 \nu^{\zeta}\right) \quad \text{if } 0 < \zeta < 1 \text{ or } |\mathscr{S}_\nu| \le C_0 \nu^r \text{ if } \zeta = 0$$

for some constants C_0 *and* r *depending on* $\mathscr{S}$.

(2.2) *The uniform perimetric property.* $\bigcup_\nu \mathscr{S}_\nu$ *fulfills condition UM on the one hand and* $\bigvee_{\mathscr{S}_\nu} s_{d-1} \le K$ *for some constant* K *depending only on* $\mathscr{S}$ *on the other hand.*

(2.3) *Oscillation control. Given* $\eta > 0$, *an upper bound for*

$$\Pr{}^*\left(\bigvee_{\mathscr{S}} |Z_n - Z_n \circ \Pi_\nu| > \nu^{-(1-\zeta)/2} T \right)$$

is, for any positive T, *given by*

$$\Lambda \exp(-\theta T) \quad \text{if } 0 < \zeta < 1,$$
$$\Lambda n^{\alpha} \exp(-\theta T) \quad \text{if } \zeta = 0,$$

whenever $\nu^{-(1+\zeta)} n \ge \eta$ *for some constants* Λ, α *and* θ *depending on* $\mathscr{S}$ *and* η.

PROOF. For any positive ε, let $N(\varepsilon, \mathscr{S})$ denote the minimal cardinality of the range of a mapping $\Pi(\varepsilon)$: $\mathscr{S} \to \mathscr{S}$ such that $P(\Pi(\varepsilon) S \triangle S) \le \varepsilon$ for any S in $\mathscr{S}$.

Then it is easy to see that $N(\varepsilon, \mathscr{S}) \le N_{[\ ]}(\varepsilon, \mathscr{S})$ on the one hand and it follows from Dudley (1978) that, if $\zeta = 0$, $N(\varepsilon, \mathscr{S}) \le C\varepsilon^{-r}$ for some constants C and r on the other hand [more about this property is given in Assouad (1983)].

Given $\Pi(\varepsilon)$ as above with minimal cardinality range, we define $\Pi^\nu = \Pi(\nu^{-1})$ and then $\Pi_\nu S = (\Pi^\nu S)_\nu$ for any S in $\mathscr{S}$. Clearly (2.1) holds.

From now on we choose to work with the supremum norm on $\mathbb{R}^d$ $|y| = \sup_{1 \le i \le d} |y_i|$.

We show a little more than (2.2), namely we show that (2.2) holds when replacing $\mathscr{S}_\nu$ with $\tilde{\mathscr{S}}_\nu = \{S_\nu;\ S \in \mathscr{S}\}$. To prove this, we note that, given $\varepsilon = u\nu^{-1}$ and $B \in \mathscr{B}_\nu$, the quantity $2(((1 + 2u)^d - 1)/2d)\nu^{-d}$ [respectively, $((1 - (1 - 2u)/2d)\nu^{-d})$ represents the maximal (respectively, minimal) contribution to $P((\partial B)^\varepsilon)$ of one of the elementary faces (of a cube) composing the boundary of B. The number of such faces is equal to $\nu^{d-1} s_{d-1}(B)$, so upper and lower bounds for $P((\partial B)^\varepsilon)$ are, respectively, given by

$$2\left(\left((1 + 2u)^d - 1\right)/2d\right)\nu^{-1} s_{d-1}(B) \le 2(1 + 2u)^{d-1} u \nu^{-1} s_{d-1}(B) \quad \text{for } u > 0, \tag{2.4}$$

$$\left(\left(1 - (1 - 2u)^d\right)/2d\right)\nu^{-1} s_{d-1}(B) \ge (1 - 2u)^{d-1} u \nu^{-1} s_{d-1}(B) \quad \text{for } 0 < u \le 1/2. \tag{2.5}$$

Now, the following inclusion is elementary:

$$(\partial A_\nu)^\varepsilon \subset (\partial A)^{\varepsilon+\nu^{-1}} \quad \text{for any } A \subset \mathbb{R}^d. \tag{2.6}$$

Then, because $\mathscr{S}$ fulfills conditions UM we get, using (2.5),

$$\bigvee_{\tilde{\mathscr{S}}_\nu} s_{d-1} \le K \quad \text{for some constant } K. \tag{2.7}$$

Moreover, using either (2.4) and (2.7) when $u \le 1$ or (2.6) and condition UM for $\mathscr{S}$ when $u > 1$, we get that $\bigcup_\nu \tilde{\mathscr{S}}_\nu$ fulfills condition UM. Thus (2.2) holds.

As a consequence of Corollary 2.4 of Alexander (1984) and of Proposition 3.7 of Massart (1986), both applied to the classes $\{S \setminus \Pi^\nu S\colon S \in \mathscr{S}\}$ and $\{\Pi^\nu S \setminus S\colon S \in \mathscr{S}\}$, we get that (2.3) holds when replacing Π_ν with Π^ν.

In order to control $Z_n \circ \Pi_\nu - Z_n \circ \Pi^\nu$ uniformly over $\mathscr{S}$, note that $A \subset A_\nu$ and $(A_\nu \setminus A) \subset (\partial A)^{\nu^{-1}}$ for any subset A of $]0,1]^d$, so since $\mathscr{S}$ fulfills condition UM, $P(\Pi_\nu S \triangle \Pi^\nu S) \le K'\nu^{-1}$ for any S in $\mathscr{S}$ and some constant K'. Thus, using Bernstein's inequality [see Bennett (1962)], we get

$$\Pr\Big(\bigvee_{\mathscr{S}} |Z_n \circ \Pi_\nu - Z_n \circ \Pi^\nu| > T\nu^{-(1-\zeta)/2}\Big)$$

$$\le 2N(\nu^{-1}, \mathscr{S})\exp\big(-T^2\nu^\zeta/(2K'^2 + \eta^{-1/2}T)\big),$$

yielding (2.3). Thus the proof of Lemma 1 is complete. □

Let π be a random variable independent of $(x_j)_{j\ge1}$, such that π has the Poisson law on $\mathbb{N}$ with parameter n [recall that this means $\Pr(\pi = k) = e^{-n}(n^k/k!)$]. Then it is well known [see for instance Gaenssler (1983)] that πP_π is a Poisson point process with intensity measure nP.

Moreover, in the spirit of Csörgő and Révész (1975), the following Poissonization lemma is available.

Lemma 2 (Poissonization). *Let $\mathscr{C}$ be a finite collection of Borel sets of $\mathbb{R}^d$. For any positive U and T, we have*

$$\Pr\left(\bigvee_{\mathscr{C}} \left|\sqrt{\frac{\pi}{n}}Z_\pi - Z_n\right| > T\right) \le 2\exp\left(-\frac{U}{4}(U \wedge \sqrt{n})\right) + 2|\mathscr{C}|\exp\left(-\frac{2T\sqrt{n}}{U}\right).$$

Proof. We have

$$\Pr\Big(\bigvee_{\mathscr{C}} |\sqrt{\pi/n}\,Z_\pi - Z_n| > T\Big)$$

$$\le \Pr\big(|\pi - n| > U\sqrt{n}\big)$$

$$+ \sum_{|k-n|\le U\sqrt{n}} \Pr\Big(\bigvee_{\mathscr{C}} |\sqrt{k/n}\,Z_k - Z_n| > T\Big)\Pr(\pi = k),$$

so, using Chernoff's inequality, we get

$$\Pr\Big(\bigvee_{\mathscr{C}} |\sqrt{\pi/n}\, Z_\pi - Z_n| > T\Big) \le 2\exp\Big(-n\Big(h\Big(\frac{U}{\sqrt{n}}\Big) \wedge h\Big(\frac{-U}{\sqrt{n}}\Big)\Big)\Big) + \max_{|k-n|\le U\sqrt{n}} \Pr\Big(\bigvee_{\mathscr{C}} |\sqrt{k/n}\, Z_k - Z_n| > T\Big),$$

where $h(x) = (1+x)\log(1+x) - x$ when $x > -1$ and $h(x) = +\infty$ when $x \le -1$. Now, given $A \in \mathscr{C}$, $|(kP_k - nP_n)(A)|$ has the binomial distribution $\mathscr{B}(|k-n|, P(A))$. Then, using the inequality of Hoeffding (1963) leads to

$$\max_{|k-n|\le U\sqrt{n}} \Pr\Big(\bigvee_{\mathscr{C}} |\sqrt{k/n}\, Z_k - Z_n| > T\Big) \le 2|\mathscr{C}| \max_{|k-n|\le U\sqrt{n}} \exp\Big(-\frac{2nT^2}{|k-n|}\Big) \le 2|\mathscr{C}|\exp\Big(-\frac{2T^2\sqrt{n}}{U}\Big).$$

To complete the proof of Lemma 2, it is enough to note that the following elementary inequality holds for any x: $h(x) \ge (|x|/4)(|x| \wedge 1)$. □

How does the proof of Theorem 1 *work*? Lemmas 1 and 2 mean that in order to approximate Z_n by some suitable Brownian bridge uniformly over $\mathscr{S}$, it is enough to approximate the centered and normalized Poisson process $\xi_n = (\pi P_\pi - nP)/\sqrt{n}$, uniformly over the subcollection $\Pi_\nu(\mathscr{S})$ of $\mathscr{B}_\nu$ by some suitable Wiener process W.

Let us explain briefly how our method works in the VČ case and the difference between our approach and that of Csörgő and Révész. The method of Csörgő and Révész (1975) or Révész (1976a), to perform the above construction, would consist of constructing W *independently* on each atom of $\mathscr{B}_\nu$ with a rate of approximation which is about $Ln/\sqrt{n}$, then getting

$$\bigvee_{\Pi_\nu(\mathscr{S})} |\xi_n - W| = O\Big(\frac{\nu^{d/2}(Ln)^{3/2}}{\sqrt{n}}\Big)$$

with great probability, because each set of $\mathscr{B}_\nu$ is composed of no more than ν^d atoms and the errors are independent on disjoint atoms. [The cardinality of $\Pi_\nu(\mathscr{S})$ is so small in the VČ case that all works as if it was bounded.] The key idea in our approach consists in counting "intervals" rather than atoms. In fact, in the KMT constructions, the rate of approximation is still about $Lnn^{-1/2}$ on the intervals, moreover the uniform perimetric property [see (2.2) above] means that each set of $\Pi_\nu(\mathscr{S})$ is composed of no more than $O(\nu^{d-1})$ intervals.

The main difficulty is that, with this new construction, the errors are not independent on disjoint sets anymore, but we claim that this dependence is weak enough to allow an evaluation such as

$$\bigvee_{\Pi_\nu(\mathscr{S})} |\xi_n - W| = O\Big(\frac{\nu^{(d-1)/2}(Ln)^{3/2}}{\sqrt{n}}\Big)$$

with great probability.

The purpose of Section 3 (which is the main part of this paper) is to give a precise sense to this claim.

3. A KMT-type inequality for set-indexed partial sums. First we need some notations.

NOTATION. Given a finite subset A of $\mathbb{N}$ and a sequence $u \in \mathbb{R}^{\mathbb{N}}$ we set $u[A] = \sum_{i \in A} u_i$. Moreover, we say that two intervals of $\mathbb{N}$ are strictly disjoint if their union is not an interval. Of course A may be partitioned into strictly disjoint intervals (which we call the components of A). We denote by $c(A)$ the number of components of A.

In this section, $(\Omega, \mathscr{X}, \Pr)$ simply denotes an appropriate probability space.

THEOREM 3. *Let F be a probability law on $\mathbb{R}$ with $\int x\,dF(x) = 0$ and $\int x^2\,dF(x) = 1$ and such that we have*

$$\text{for some } t_0 > 0, \qquad \int e^{tx}\,dF(x) < \infty \quad \text{for } |t| < t_0.$$

A sequence $Y_1, Y_2, \dots$ of independent standard normal random variables and a sequence $X_1, X_2, \dots$ of independent random variables with common law F may be constructed in such a way that, for any integer n and any collection $\mathscr{A}$ of subsets of $\mathbb{N} \cap]0, n]$, the following inequality holds for any positive x, y and χ with $c(A) \le \chi$ for any A in $\mathscr{A}$:

$$(3.1)\qquad \begin{aligned} &\Pr\Big(\sup_{A \in \mathscr{A}} |X[A] - Y[A]| > \sqrt{\chi x}\,(Ln + \sqrt{y}) + x(1 + Ln)\Big) \\ &\quad \le \Lambda n^{\alpha}\big(e^{-\theta y} + |\mathscr{A}| e^{-\theta x}\big), \end{aligned}$$

where α, Λ and θ are some constants depending only on F.

COMMENTS. The notation being that of Theorem 3, suppose that $|\mathscr{A}| = O(n^{\beta})$ for some β and take $x = y = O(Ln)$. Then Theorem 3 gives

$$\sup_{A \in \mathscr{A}} |X[A] - Y[A]| = O\big(\sqrt{\chi}\,(Ln)^{3/2} + (Ln)^2\big)$$

with great probability. As a crude application of Theorem 1 of KMT (1976), one would get that the same result holds with χLn instead of $\sqrt{\chi}\,(Ln)^{3/2} + (Ln)^2$. So our result improves on this crude application of KMT's theorem whenever $Ln \le \chi$ (in the forthcoming application χ will be some power of n). Note that all works as if the errors were strictly independent on disjoint intervals.

Theorem 3 is really a multivariate analogue of Theorem 1 of KMT (1976): This idea will be illustrated in Section 5 below. In view of applying Theorem 3 to approximate empirical processes, we shall take F as the centered Poisson law with parameter 1, as explained at the end of Section 2.

PROOF OF THEOREM 3. For the sake of simplicity we first assume the regularity condition of Theorem 1 of KMT (1975) is fulfilled, that is:

(a) $\int_{\mathbb{R}} |R(t + iu)|^p\,du < \infty$ for some $p \ge 1$ and all t, $|t| < t_0$, where $R(z) = \int_{\mathbb{R}} e^{zx}\,dF(x)$ for arbitrary complex z with $|\mathrm{Re}\,z| < t_0$.

Next, we shall explain how to get rid of this condition following the lines of the proof of Theorem 1 of KMT (1975).

So first the sequence $(X_n)_{n\geq 1}$ is constructed exactly as in KMT (1975). We shall not explicitly recall this construction but our notation will be as close to that of Theorem 1 of KMT (1975) as possible, so that we hope it will be easy for the reader to refer to KMT (1975) when necessary.

Throughout the proof the intervals $]k, k']$ have to be interpreted as subsets of $\mathbb{Z}_+$. There is no loss of generality in assuming that $n = 2^N$, which we shall do in the sequel.

$\mathbb{R}^n$ is given the canonical inner product which we denote by $(\cdot|\cdot)$. The functions $\mathbb{1}_B$, $B \subset]0, n]$ will be considered as vectors of $\mathbb{R}^n$.

The dyadic orthogonal expansion. Given $I_{j,k} =]k2^j, (k+1)2^j]$, we set $e_{j,k} = \mathbb{1}_{I_{j,k}}$. Then, let $\tilde{e}_{j,k} = e_{j-1,2k} - e_{j-1,2k+1}$ and $\tilde{e}_j = e_{j,1}$.

It is easy to verify that the family

$$\left\{\tilde{e}_{j,k}, 1 \leq j < N, 1 \leq k < 2^{N-j}\right\} \cup \left\{\tilde{e}_j, 0 \leq j < N\right\} \cup \left\{e_{0,0}\right\}$$

is an orthogonal basis of $\mathbb{R}^n$ with furthermore $(\tilde{e}_{j,k}|\tilde{e}_{j,k}) = (\tilde{e}_j, \tilde{e}_j) = 2^j$. The orthogonal expansion of a vector $\mathbb{1}_B$, $B \subset]0, n]$ with respect to the above basis has the form

$$\begin{aligned} \mathbb{1}_B &= |B \cap \{1\}| e_{0,0} + \sum_{0 \leq j < N} |B \cap I_{j,1}| 2^{-j} \tilde{e}_j \\ &\quad + \sum_{1 \leq j < N} \sum_{1 \leq k < 2^{N-j}} 2^{-j}\left(|B \cap I_{j-1,2k}| - |B \cap I_{j-1,2k+1}|\right) \tilde{e}_{j,k}. \end{aligned} \tag{3.2}$$

Following KMT (1975) we then set $\tilde{U}_{j,k} = (X|\tilde{e}_{j,k})$, $U_{j,k} = (X|e_{j,k})$ and $\tilde{U}_j = (X|\tilde{e}_j)$, where $X = (X_1, \ldots, X_n)$. The variables $\tilde{V}_{j,k}$, $V_{j,k}$ and $\tilde{V}_j$ are defined from $Y = (Y_1, \ldots, Y_n)$ in the same way.

Given $B \subset]0, n]$, it follows from (3.2) and the identities $X[B] = (X|\mathbb{1}_B)$ and $Y[B] = (Y|1_B)$, that $X[B] - Y[B]$ is a linear combination of the elementary differences $(\tilde{U} - \tilde{V}) - s$. So the problem is first to control the differences $(\tilde{U} - \tilde{V}) - s$, second to handle some of their linear combinations.

Some technical lemmas. Exactly as in KMT (1975), the differences $(\tilde{U} - \tilde{V}) - s$ are controlled via the following crucial estimate which is proved in KMT (1975).

LEMMA 3. *If F fulfills condition* (a), *then there are positive constants C_1, C_2 and ε such that*

$$|\tilde{U}_j - \tilde{V}_j| < C_1 2^{-j} \tilde{U}_j^2 + C_2 \quad \textit{if } |\tilde{U}_j| < \varepsilon 2^j,$$

$$|\tilde{U}_{j,k} - \tilde{V}_{j,k}| < C_1 2^{-j}\left(\tilde{U}_{j,k}^2 + U_{j,k}^2\right) + C_2 \quad \textit{if } |\tilde{U}_{j,k}| < \varepsilon 2^j \textit{ and } |U_{j,k}| < \varepsilon 2^j,$$

where the $U - s$ and $V - s$ are defined above.

We also need the following large deviation lemma. Let $\overline{H}(a, r)$ be the class of random variables Z such that

$$E(\exp(tZ)) \leq r \quad \text{for all } |t| \leq a.$$

We denote by $H(a, r)$ the class of the $Z - EZ$'s with Z in $\overline{H}(a, r)$.

LEMMA 4. *Let $(T_i)_{i \in I}$ be a finite family of independent elements of $H(a, r)$ and $(w_i)_{i \in I}$ be a family of real numbers with $|w_i| \leq 1$ for all $i \in I$. Setting $S(w) = \sum_{i \in I} w_i T_i$ we have, for all v such that $\sum_{i \in I} w_i^2 \leq v$,*

(i) *for* $0 \leq \xi \leq 2r^2/a$, $\Pr(|S(w)| \geq v\xi) \leq 2\exp(-v(a\xi/2r)^2)$; *for* $2r^2/a < \xi$, $\Pr(|S(w)| \geq v\xi) \leq 2\exp(-va\xi/2)$;

(ii) *moreover* $(S^2(w)/v)\mathbb{1}_{|S(w)| \leq 2vr^2/a}$ *belongs to* $\overline{H}(a^2/8r^2, 3)$.

(3.3) REMARK. (i) Note that, using Lemma 4(i), the following inequality holds for *all* positive ξ:

$$\Pr(|S(w)| \geq v\xi) \leq 2\exp\left(-v\left(\frac{a}{2r}\right)^2 \frac{\xi^2}{1+\xi}\right) \quad \text{whenever } a \leq 2r^2.$$

(ii) Given $K_1, \dots, K_J$ in $\overline{H}(a, r)$ and some positive $\lambda_1, \dots \lambda_J$, we have

$$\sum_{j=1}^{J} \lambda_j K_j \in \overline{H}\left(\frac{a}{\sum_{j=1}^{J} \lambda_j}, r\right)$$

because $x \to e^{tx}$ is a convex function.

PROOF OF LEMMA 4. Let $Z \in \overline{H}(a, r)$ and Z' be an independent copy of Z. We define $\tilde{Z} = Z - Z'$. Then $\tilde{Z}$ is in $\overline{H}(a, r^2)$.

As $\tilde{Z}$ is a symmetric random variable, we have for all $|t| \leq a$

$$E(\exp(t\tilde{Z})) = 1 + \sum_{k>0} \frac{t^{2k}}{2k!} E(\tilde{Z}^{2k}) \leq 1 + \left(\frac{t}{a}\right)^2 E(\exp(a\tilde{Z}) - 1),$$

so

$$E(\exp(t\tilde{Z})) \leq \exp\left(\left(\frac{tr}{a}\right)^2\right)$$

and it follows from Jensen's inequality that the same inequality holds for $E(\exp(t(Z - EZ))$.

Thus, $\log E(\exp(tT)) \leq (rt/a)^2$ whenever $|t| \leq a$ for all $T \in H(a, r)$.

Therefore $\log E(\exp(tS(w)) \leq v(rt/a)^2$ whenever $|t| \leq a$ and we get (i) using the classical Cramér–Chernoff calculation.

To bound the Laplace transform of

$$K = \frac{S^2(w)}{v} \mathbb{1}_{|S(w)| \leq 2vr^2/a}.$$

we use (i). Then, integrating by parts, we get

$$E(\exp(tK)) = \int_0^{4vr^2/a} te^{tu}\Pr\left(|S(w)| > \sqrt{vu}\right) du + 1 \le 1 + 2\int_0^{\infty} te^{tu-(a^2u/(4r^2))}\, du$$

so

$$E(\exp(tK)) \le \frac{2t}{\left(a^2/(4r^2)\right) - t} + 1 \le 3,$$

whenever $|t| \le a^2/8r^2$, and the proof of Lemma 4 is complete. □

End of the proof of Theorem 3 *in case* (*a*). We first note that X_j belongs to $H(a_0, 3)$ for all j with a_0 depending only on F. We may always assume that $a_0 \le 1$. In the same way, we shall assume that the constants ε, C_1 and C_2 of Lemma 3 fulfill the conditions $\varepsilon \le 1$, $C = C_1 = C_2 \ge 1$.

We finally set $\alpha = a_0^2/216C$ and note that α depends only on F.

We define $\tilde{Z}_j = (\tilde{U}_j - \tilde{V}_j)\mathbb{1}_{(|\tilde{U}_j| < \varepsilon 2^j)}$ and

$$\tilde{Z}_{j,k} = \left(\tilde{U}_{j,k} - \tilde{V}_{j,k}\right)\mathbb{1}_{(|\tilde{U}_{j,k}| < \varepsilon 2^j,\ |U_{j,k}| < \varepsilon 2^j)}$$

and note that

(3.4) the variables $\tilde{Z} - s$ all belong to $\overline{H}(\alpha, 3)$

[for instance $|\tilde{Z}_j| \le C(2^{-j}\tilde{U}_j^2\mathbb{1}_{(|\tilde{U}_j| < 2^{j+1}9/a_0^2)} + 1)$ because of Lemma 3 and then conclude that $\tilde{Z}_j$ belongs to $\overline{H}(a_0^2/144C, 3)$ using Lemma 4(ii), Remark 3.3(ii) and the fact that $1 \in \overline{H}(a_0^2/72, 3)$].

We fix α_0 and β_0 such that

$$2^{3N+8}\exp\left(-(\alpha_0 N + \beta_0)\left(\frac{a_0\varepsilon}{12}\right)^2\right) \le 1 \tag{3.5}$$

and we assume that $x \ge 4$ and $y \ge \alpha_0 N + \beta_0$. We define the integer M by

$$\frac{y}{4} < 2^M \le \frac{y}{2}.$$

We start by regularizing the subsets of $]0, 2^N]$ at the scale M.

Regularization at the scale M. Given an integer k, let $k(M)$ be either the first multiple of 2^M which is greater than or equal to k if $M < N$ or equal to 2^N if $M \ge N$.

Now, given an interval $B =]k, k']$, let $B(M) =]k(M), k'(M)]$ be the regularization of B at the scale M.

Given a subset A of $]0, n]$, we finally define the regularization $A(M)$ of A as the union of the regularizations of the components of A. The regularization has two elementary properties:

$$|A \triangle A(M)| \le c(A)2^{M+1}, \tag{3.6}$$

$$c(A(M)) \le c(A). \tag{3.7}$$

Now we pass to the technical part of the proof of Theorem 3.

Estimation of $\sup_{A\in\mathscr{A}}|X[A]-Y[A]|$. Given $B\subseteq\mathbb{N}$, we set $\gamma_j(B)=2^{-j}|B\cap I_{j,1}|$ and $\gamma_{j,k}(B)=2^{-j}(|B\cap I_{j-1,2k}|-|B\cap I_{j-1,2k+1}|)$. Since the extremities of the components of $A(M)$ are multiples of 2^M and $A(M)\subset[2^M,2^N]$, we have $\gamma_j(A(M))=\gamma_{j,k}(A(M))=0$ for $j<M$ and $\gamma_{M,k}(A(M))=0$ for any $A\in\mathscr{A}$. Thus, using (3.2) with $B=A(M)$, we get the decomposition

$$X[A]-Y[A]=D_1(A)+D_2(A)+D_3(A),$$

where

$$\begin{aligned}D_1(A)&=(X[A\setminus A(M)]-X[A(M)\setminus A])\\&\quad+(Y[A(M)\setminus A]-Y[A\setminus A(M)]),\\ D_2(A)&=\sum_{M\le j<N}\gamma_j(A(M)(\tilde{U}_j-\tilde{V}_j)\ \text{ if } M<N \text{ or } D_2(A)=0 \text{ if } M\ge N\end{aligned}$$

and

$$D_3(A)=\sum_{M<j<N}\ \sum_{1\le k<2^{N-j}}\gamma_{j,k}(A(M))(\tilde{U}_{j,k}-\tilde{V}_{j,k})$$
$$\text{if } M<N-1 \text{ or } D_3(A)=0 \text{ if } M\ge N-1.$$

Defining the event Θ by

$$\Theta=\Big\{|\tilde{U}_j|<\varepsilon 2^j \text{ for all } M\le j<N,\ |\tilde{U}_{j,k}|<\varepsilon 2^j,$$
$$|U_{j,k}|<\varepsilon 2^j \text{ for all } M<j<N,\ 0<k<2^{N-j}\Big\}$$

if $M<N$ or $\Theta=\Omega$ if $M\ge N$ and recalling that $4\le x$, the above decomposition leads to

$$\Pr\Big(\sup_{A\in\mathscr{A}}|X[A]-Y[A]|>\sqrt{\chi x}(N+\sqrt{y})+(N+1)x\Big)$$
$$\le\mathbb{P}_1+\mathbb{P}_2+\mathbb{P}_3+\Pr(\Theta^c),$$

where

$$\mathbb{P}_1=|\mathscr{A}|\Big(\bigvee_{\mathscr{A}}\Pr\Big(|D_1|>\sqrt{\chi xy}+\frac{x}{4}\Big)\Big),$$

$$\mathbb{P}_2=|\mathscr{A}|\Big(\bigvee_{\mathscr{A}}\Pr\Big(\Big(|D_2|>\frac{x}{2}\Big)\cap\Theta\Big)\Big)$$

and

$$\mathbb{P}_3=|\mathscr{A}|\Big(\bigvee_{\mathscr{A}}\Pr\big((|D_3|>N(\sqrt{\chi x}+x)+1)\cap\Theta\big)\Big).$$

Control of $\mathbb{P}_1$. We use Lemma 4 with $a=a_0$ and $v=\max(\chi y,x/16)$. Given $A\in\mathscr{A}$, note that $|A\triangle A(M)|\le v$ because of (3.6), so inequality (i) of Lemma 4 means when $\xi=\sqrt{x/v}$,

$$\Pr\big(|X[A\setminus A(M)]-X[A(M)\setminus A]|>\sqrt{x\max(\chi y,x/16)}\big)$$
$$\le 2\exp\big(-x(a_0/6)^2\big)$$

whenever $\xi \le 18/a_0$. But the latter condition is always fulfilled because of our choice of v and because $9/a_0 > 8$. We finally note that the same inequality holds for the Gaussian term because of course the $Y-s$ belong to $H(a_0, 3)$. Thus

$$\mathbb{P}_1 \le 4|\mathscr{A}|\exp\left(-\left(\frac{a_0}{6}\right)^2 x\right).$$

Control of $\Pr(\Theta^c)$. (We may assume that $M < N$.) The number of variables $\tilde{U}_j, \tilde{U}_{j,k}$ and $U_{j,k}$ is less than 2^{N+1}, so we get as a direct application of Lemma 4(i) with $v = 2^j$,

$$\Pr(\Theta^c) \le 4 \sum_{M \le j \le N} 2^{N-j}\exp\left(-2^j\left(\frac{a_0\varepsilon}{6}\right)^2\right) \le 2^{N+2}\exp\left(-y\left(\frac{a_0\varepsilon}{12}\right)^2\right).$$

Control of $\mathbb{P}_2$. (We may assume that $M < N$.) We first note that on the event Θ we have, given $A \in \mathscr{A}$,

$$D_2(A) = \sum_{M \le j < N} \gamma_j(A(M))\tilde{Z}_j.$$

But from (3.4), the $\tilde{Z}-s$ belong to $\overline{H}(\alpha, 3)$. Moreover $|\gamma_j(A(M))| \le 1$, so crudely, $E(\exp(tD_2(A))) \le 3^N$ for $t = \pm\alpha$ yielding the estimate,

$$\mathbb{P}_2 \le 2|\mathscr{A}|3^N\exp\left(-\alpha\frac{x}{2}\right).$$

Control of $\mathbb{P}_3$. We *fix* an integer j. The variable $U_{j,k}$ is constructed as a function of $(\tilde{V}_{j+1,[k/2]}, \dots, \tilde{V}_J)$, where J is that integer for which $2^J \le k2^j < 2^{J+1}$ holds.

Now the variables $\tilde{V}_{j'}, \tilde{V}_{j'',k}$ are orthogonal when j', j'' and k vary, so they are independent (because, of course, of the Gaussian property). Hence, the sequence $(U_{j,k})_{k\ge1}$ and $(\tilde{V}_{j,k})_{k\ge1}$ are independent. But each of these sequences is composed with independent random variables when k varies.

Since $\tilde{U}_{j,k}$ is constructed as a function of $U_{j,k}$ and $\tilde{V}_{j,k}$, we deduce that $(U_{j,k}; \tilde{U}_{j,k}; \tilde{V}_{j,k})$ is a sequence of independent random variables when k varies.

Thus the variables $\tilde{Z}_{j,k}$, as defined in (3.4), are independent when k varies and belong to $\overline{H}(\alpha, 3)$.

Moreover the coefficients $\gamma_{j,k}(A(M))$ have the nice properties that they vanish when no extremity of a component of $A(M)$ belongs to $I_{j,k}$ and that they are absolutely bounded by 1. Recalling furthermore that (3.7) holds, we may apply Lemma 4(i) with $a = \alpha$, $v = 2\max(\chi, x)$ and $\xi = \sqrt{x/2v}$, getting, for each fixed j and A,

$$\left|\sum_{1 \le k < 2^{N-j}} \gamma_{j,k}(A(M))\left(\tilde{Z}_{j,k} - E\tilde{Z}_{j,k}\right)\right| \le \sqrt{x\max(\chi, x)}$$

except on an event with probability less than $2\exp(-x(\alpha/6)^2/2)$ whenever $\xi \le 18/\alpha$; but the latter condition is ensured via our choice of v and the fact that $9/\alpha > 1$.

Now $E(\tilde{U}_{j,k} - \tilde{V}_{j,k}) = 0$ and $E(\tilde{U}_{j,k} - \tilde{V}_{j,k})^2 \le 2^{j+2}$, so the Cauchy–Schwarz inequality yields

$$|E\tilde{Z}_{j,k}| \le 2(\Pr(\Theta^c))^{1/2} 2^{j/2}.$$

Then, recalling that $|\gamma_{j,k}(A(M))| \le 1$ once more, we get using the above estimate of $\Pr(\Theta^c)$ and (3.5),

$$\left| \sum_{M<j<N} \sum_{1 \le k < 2^{N-j}} \gamma_{j,k}(A(M)) E\tilde{Z}_{j,k} \right| \le 8 \cdot 2^N (\Pr(\Theta^c))^{1/2} \le 1.$$

Noting that $\tilde{U}_{j,k} - \tilde{V}_{j,k} = \tilde{Z}_{j,k}$ on Θ, we finally get

$$\mathbb{P}_3 \le 2N|\mathscr{A}| \exp\left(-\frac{x}{2}\left(\frac{\alpha}{6}\right)^2 \right).$$

Collecting the above estimates of $\mathbb{P}_1$, $\mathbb{P}_2$, $\mathbb{P}_3$ and $\Pr(\Theta^c)$ yields for all positive x, y,

$$\Pr\left(\sup_{A \in \mathscr{A}} |X[A] - Y[A]| > \sqrt{\chi x}\,(N + \sqrt{y}) + (N+1)x \right)$$
$$\le 4 \cdot 3^N |\mathscr{A}| \exp\left(-\frac{x}{2}\left(\frac{\alpha}{6}\right)^2 \right) + 2^{3N+8} \exp\left(-y\left(\frac{\alpha_0 \varepsilon}{12}\right)^2 \right),$$

completing the proof of Theorem 3 in case (a).

Now let us sketch the proof of Theorem 1 in the general case when the smoothness condition (a) is not assumed. We follow the approach used by KMT to prove Theorem 1 of KMT (1976).

Their proof is based on two special cases:

1. F has an absolute continuous component.
2. F is concentrated on a finite interval.

CASE 1. The construction that they make in that case leads to estimates for the differences $(\tilde{U} - \tilde{V}) - s$ which are just the same as in Lemma 3, so our proof in case (a) carries over in case (1) without any difficulty.

CASE 2. The situation is here slightly different. In fact, KMT showed that the estimates of Lemma 3 do not remain valid.

Anyway the new "control" variables $W - s$ introduced on page 48 of KMT (1976) have the key property that for each fixed j, the variables $(\tilde{U}_{j,k}; U_{j,k}; \tilde{V}_{j,k}; W_{j,k})$ are constructed in such a way that they are still independent when k varies. Thus our proof in case (a) carries over in that case, too.

Now we pass to the general case. We shall need the following measure-theoretic lemma due to Berkes and Philipp [Lemma A1 of Berkes and Philipp (1979)].

LEMMA 5. *Given three Polish spaces R_1, R_2 and R_3, let Q_1 and Q_2 be probability laws which are, respectively, defined on $R_1 \times R_2$ and $R_2 \times R_3$. If Q_1 and Q_2 have the same marginal distribution on R_2, then there exists a distribu-*

tion Q on $R_1 \times R_2 \times R_3$ whose marginals on $R_1 \times R_2$ and $R_2 \times R_3$ are, respectively, Q_1 and Q_2.

We write F as $F = pF_1 + (1-p)F_2$ for some $p \in]0,1]$ with F_1 concentrated on a finite interval. By Theorem 3 in case (2), we may consider a sequence $(X^{(1)})$ of i.i.d. random variables with common law F_1 and a sequence $(Y^{(1)})$ of i.i.d. random variables with the corresponding common normal distribution $\varphi(m_1, \sigma_1)$, such that (3.1) holds for $(X^{(1)}, Y^{(1)})$.

Now, given a sequence $(X^{(2)})$ of i.i.d. random variables with common law F_2 which is independent of $(X^{(1)})$ and $(Y^{(1)})$ and a sequence (ε) of Bernoulli trials with probability of success p which is independent of $(X^{(1)})$, $(Y^{(1)})$ and $(X^{(2)})$, define $\mu(m) = \sum_{j=1}^{m}\varepsilon_j$ for any integer m and then the sequences (X) and (X') via their partial sums $S_m = S^{(1)}_{\mu(m)} + S^{(2)}_{m-\mu(m)}$, respectively, $S'_m = T^{(1)}_{\mu(m)} + S^{(2)}_{m-\mu(m)}$ for any integer m, where $(S^{(1)})$, $(S^{(2)})$ and $(T^{(1)})$ denote, respectively, the partial sum processes associated with $(X^{(1)})$, $(X^{(2)})$ and $(Y^{(1)})$.

Then (X) is a sequence of i.i.d. random variables with common law F and (X') is a sequence of i.i.d. random variables with common law $p\varphi(m_1, \sigma_1) + (1-p)F_2$. Moreover, setting $\mu(A) = \{\mu(j)\colon\ j \in A\}$ for any A in $\mathscr{A}$ and next $\mu(\mathscr{A}) = \{\mu(A)\colon\ A \in \mathscr{A}\}$, we clearly have

$$\sup_{A \in \mathscr{A}} |X[A] - X'[A]| \leq \sup_{A^* \in \mu(\mathscr{A})} |X^{(1)}[A^*] - Y^{(1)}[A^*]|. \tag{3.8}$$

Now we work conditionally on (ε). Note that $\mu(\mathscr{A})$ has the same characteristics as $\mathscr{A}$, that is, $\mu(\mathscr{A})$ is a class of subsets of $]0, n]$ with $\sup_{A^* \in \mu(\mathscr{A})} c(A^*) \leq \chi$ and $|\mu(\mathscr{A})| \leq |\mathscr{A}|$.

So since (3.1) holds for $(X^{(1)})$, $(Y^{(1)})$ and $\mu(\mathscr{A})$, conditionally on (ε) we get from (3.8) and the above remarks on the characteristics of $\mu(\mathscr{A})$,

$$\Pr\left(\left\{\sup_{A \in \mathscr{A}} |X[A] - X'[A]| > \sqrt{\chi x}\left(Ln + \sqrt{y}\right) + x(1 + Ln)\right\} \Big| (\varepsilon)\right)$$
$$\leq \Lambda n^{\alpha}\left(e^{-\theta y} + |\mathscr{A}| e^{-\theta x}\right).$$

Now we may take the expectation with respect to (ε) in the above inequality, thus getting (3.1) for (X), (X') and $\mathscr{A}$.

In another connection, since $p\varphi(m_1, \sigma_1) + (1-p)F_2$ has an absolute continuous component, we may apply Theorem 3 in case (1). So, consider a sequence (X'') of i.i.d. random variables with common law $p\varphi(m_1, \sigma_1) + (1-p)F_2$ and a sequence (Y) of i.i.d. standard normal random variables such that (3.1) holds for (X''), (Y) and $\mathscr{A}$.

But Lemma 5 allows us to consider that $(X') = (X'')$, so that (3.1) holds for (X), (Y) and $\mathscr{A}$ and the proof of Theorem 3 is now complete. □

4. Strong approximations for empirical processes. A centered Gaussian process with covariance function $(S, S') \to P(S \cap S')$ is a Wiener process. As $\mathscr{S}$ has the integrable entropy property,

$$\int_0^1 \left(\varepsilon^{-1} \log N(\varepsilon, \mathscr{S})\right)^{1/2} d\varepsilon < \infty \tag{4.1}$$

(see the proof of Lemma 1), it follows from a well-known theorem of Dudley (1967) that there exists a uniformly continuous Wiener process indexed by $\mathscr{S}$, with respect to the pseudometric $(A, B) \to P(A \Delta B)$. In order to construct such a process on our rich enough space Ω, we shall need the following lemma of Skorohod (1976).

LEMMA 6. *Given two Polish spaces R_1, R_2 and a random variable V from Ω to R_2 with law q, let Q be a probability law on $R_1 \times R_2$ with marginal distribution q on R_2. If there exists a random variable U defined on Ω which is independent of V and whose distribution is the Lebesgue measure on* $[0,1]$, *then there is a random variable Y from Ω to R_1 such that (Y, V) has law Q.*

We now pass to the

PROOF OF THEOREM 1. Given a fixed integer n of the form $n = \nu^d$, we define the one-to-one mapping τ: $\mathbb{Z}_+ \cap [0, \nu^d] \to (\mathbb{Z}_+ \cap [0, \nu])^d$ by $\tau(j) =$ the jth point of $(\mathbb{Z}_+ \cap [0, \nu])^d$ with respect to the lexicographical ordering.

Now let π be a $\mathbb{N}$-valued random variable independent of $(x_j)_{j \geq 1}$ and whose distribution is the Poisson law with parameter n. We first recall that the elementary cubes $C_\nu^{(i)}$ were defined at the beginning of Section 2 as

$$C_\nu^{(i)} = \left\{ y \in \mathbb{R}^d / (i_j - 1)\nu^{-1} < y_j \leq i_j \nu^{-1} \text{ for all } j \right\}, \qquad i \in \mathbb{Z}_+^d \cap [0, \nu]^d.$$

Then, since πP_π is a Poisson point process, $\pi P_\pi(C_\nu^{(\tau(1))}), \dots, \pi P_\pi(C_\nu^{(\tau(n))})$ are independent random variables with common law the Poisson law with parameter 1 (denoted by F_1). It follows from Theorem 3 and Lemma 5 that a family $(X_j)_{1 \leq j \leq n}$ of i.i.d. random variables such that $X + 1$ has distribution F_1 and a Wiener process W which is uniformly continuous on $(\mathscr{S}, \rho_P)$ may be constructed in such a way that (3.1) holds when setting $Y_j = n^{1/2} W(C_\nu^{(\tau(j))})$.

Using Lemma 6, we may assume that W is constructed on Ω and that $X_j = \pi P_\pi(C_\nu^{(\tau(j))}) - 1$.

Given Π_ν as in Lemma 1, we define

$$\mathscr{A} = \left\{ \tau^{-1}(\Pi_\nu S),\, S \in \mathscr{S} \right\} \cup \left\{ [0, \nu^d] \cap \mathbb{Z}_+ \right\},$$

where for each B in $\mathscr{B}_\nu$, $\tau^{-1}(B)$ denotes (abusively) the set $\{ j \in \mathbb{Z}_+ : C_\nu^{(\tau(j))} \subset B \}$.

Now, given B in $\mathscr{B}_\nu$ and $\bar{i} = (i_1, \dots, i_{d-1})$, we define

$$B_{\bar{i}} = B \cap \left(\prod_{j=1}^{d-1} \left] (i_j - 1)\nu^{-1}, i_j \nu^{-1} \right] \times [0,1] \right).$$

Then $c(\tau^{-1}(B_{\bar{i}}))$ is less than or equal to the number of elementary faces (of cubes) composing the boundary of $B_{\bar{i}}$ that are parallel to the hyperplane $y_d = 0$; the number of such faces (for all possible values of $\bar{i}$) is not greater than or equal to $\nu^{d-1} s_{d-1}(B)$, so, since $c(\tau^{-1}(B)) \leq \sum_{\bar{i}} c(\tau^{-1}(B_{\bar{i}}))$, we get

$$\sup_{A \in \mathscr{A}} c(A) \leq \nu^{d-1} \sup_{S \in \mathscr{S}} s_{d-1}(\Pi_\nu(S)),$$

so that, using property (2.2), we may choose χ in (3.1) as $\chi = K\nu^{d-1}$ for some constant K.

Thus, defining the Brownian bridge associated with W by $Z = W - W(]0,1]^d)P$, (3.1) yields (note that in the sequel, the same letters Λ, α, C and θ will denote some constants which may be different at different steps of the proof)

$$\Pr\Big(\bigvee_{\mathscr{S}} \big|\sqrt{\pi/n}\, Z_\pi \circ \Pi_\nu - Z \circ \Pi_\nu\big| > 2\sqrt{Kx}\,(Ln + \sqrt{y})\nu^{-1/2} + 2x(1 + Ln)\nu^{-d/2}\Big) \le \Lambda n^\alpha\big(e^{-\theta y} + (1 + |\mathscr{S}|)\, e^{-\theta x}\big) \tag{4.2}$$

for all positive x, y.

Now we write the difference $Z_n - Z$ as

$$Z_n - Z = D_1 + D_2 + D_3 + D_4,$$

where $D_1 = Z_n - Z_n \circ \Pi_\nu$, $D_2 = Z_n \circ \Pi_\nu - \sqrt{\pi/n}\, Z_\pi \circ \Pi_\nu$, $D_3 = \sqrt{\pi/n}\, Z_\pi \circ \Pi_\nu - Z \circ \Pi_\nu$ and $D_4 = Z \circ \Pi_\nu - Z$.

D_1, D_2, D_3 and D_4 are, respectively, controlled with the help of (2.3), Lemma 2, (4.2) and an analogue of (2.3). In fact, as far as D_4 is concerned, it is clear that (2.3) remains valid when replacing Z_n with Z, because the inequalities on which (2.3) is based still hold for Z [this was developed in Massart (1986) on the one hand; on the other hand note that, in the Gaussian case, the exponent θT in (2.3) could even be replaced with θT^2].

Control of D_1 and D_4. Applying (2.3) with $\eta = 1$ and

$$T = t/4 + CLn,$$

we get (since $\nu^{-(1-\zeta)/2} = n^{-(1-\zeta)/2d}$)

$$\Pr{}^*\Big(\bigvee_{\mathscr{S}} |D_1| > n^{-(1-\zeta)/2d}\Big(\frac{t}{4} + CLn\Big)\Big) \le \Lambda e^{-\theta t}.$$

It follows from the above remark that this inequality remains valid when D_1 is replaced by D_4.

Control of D_2. Applying Lemma 2 with $\mathscr{C} = \mathscr{S}_\nu$, $T = n^{-1/4}\nu^{\zeta/2}(t/4 + CLn)$ and $U = 2(t/4 + CLn)$, we get [using (2.1)]

$$\Pr\Big(\bigvee_{\mathscr{S}} |D_2| > n^{-(1/4)+(\zeta/2d)}\Big(\frac{t}{4} + CLn\Big)\Big) \le \Lambda e^{-\theta t}.$$

Control of D_3. We use inequality (4.2) and property (2.1), taking x and y in (4.2) of the form: $y = aLn + bt$ and $y = y + a\nu^\zeta$.

We get, via an appropriate choice of constants a and b,

$$\Pr\Big(\bigvee_{\mathscr{S}} |D_3| > a_n(\zeta, d)\Big(\frac{t}{4} + CLn\Big)\Big) \le \Lambda e^{-\theta t}.$$

We complete the proof of Theorem 1 in the case where $n^{1/d}$ is integral by collecting the above estimates of D_1, D_2, D_3 and D_4. In the general case, let ν be the integer such that $\nu^d \leq n < (\nu + 1)^d$. As an immediate application of the Hoeffding-type inequalities of Massart [(1986), Theorem 3.3.1°)a)] when $\zeta = 0$ and of Alexander [(1984), Corollary 2.4] when $\zeta > 0$, we get for all positive T,

$$\Pr^*\left(\bigvee_{\mathscr{S}}\left|\sum_{j=\nu^d+1}^{n}\left(\delta_{x_j} - P\right)\right| > \sqrt{d(\nu+1)^{d-1}}\,T\right) \leq \Lambda e^{-T^2},$$

which precisely means that the empirical process does not move too fast between time ν^d and time n. The proof may then be easily completed by using (3.1) for Z_ν^d. □

We now pass to the proof of Theorem 2. Some of the arguments that we shall use below were already used by Dudley and Philipp (1983), Csörgő and Révész (1975) and Massart (1986).

PROOF OF THEOREM 2. For sake of simplicity, we make the complete calculations only in the case where $d \geq 3$ or $\zeta = 0$.

Given the sequence $n_j = [j^\rho]$, $j \in \mathbb{N}$ with $\rho = (d+1)/(1-\zeta)$, define the partition $\{H_j,\ j \in \mathbb{Z}_+\}$ of $\mathbb{Z}_+$, by $H_j =]n_{j-1}, n_j]$ and denote by m_j the length of H_j, $m_j = n_j - n_{j-1}$.

Now we want to build a filtration with the σ-algebras $\mathscr{B}_\nu$, $\nu \in \mathbb{Z}_+$. So, given the sequence $(\mu_j)_{j\in\mathbb{Z}_+}$ such that $m_j 2^{-d} < 2^{d\mu_j} \leq m_j$, we rename by $\mathscr{B}_j$ the σ-algebra previously denoted by $\mathscr{B}_{2^{\mu_j}}$. With this new definition we have $\mathscr{B}_j \subset \mathscr{B}_{j+1}$.

Also, given Π as in Lemma 1 and a Borel subset A of $]0,1]^d$, Π_j and A_j will stand for the quantities previously denoted by $\Pi_{2^{\mu_j}}$ and $A_{2^{\mu_j}}$ as well. Note that the same remark as in the proof of Theorem 1 concerning the constants Λ, θ, C and K, remains valid here. We make the construction of Theorem 1 independently on each H_j (using this time product spaces when applying Lemma 5 or Lemma 6), so a sequence $(Z^{(j)})_{j\in\mathbb{Z}_+}$ of almost surely uniformly continuous over $(\mathscr{S}, \rho_P)$ independent Brownian bridges may be constructed on Ω, in such a way that property (4.3) holds.

Given the mapping τ_j^{-1}: $\mathscr{B}_j \to [0, 2^{d\mu_j}] \cap \mathbb{Z}_+$ which preserves the lexicographical ordering (defined as τ^{-1} in the proof of Theorem 1) for any collection $\mathscr{C} \subset \mathscr{B}_j$ with $c(\tau_j^{-1}(C)) \leq \chi$ for any C in $\mathscr{C}$ and all positive x, y, U and T, an upper bound for the quantity

$$\Pr\left(\bigvee_{\mathscr{C}}\left|\frac{1}{\sqrt{m_j}}\sum_{k\in H_j} X_k - Z^{(j)}\right| > 2\left(\sqrt{\chi x}\left(Lm_j + \sqrt{y}\right) + \left(1 + Lm_j\right)x\right)2^{-d\mu_j/2} + T\right) \tag{4.3}$$

is given by

$$2\exp\left(-\frac{U}{4}\left(U \wedge \sqrt{m_j}\right)\right) + 2|\mathscr{C}|\exp\left(-\frac{T^2\sqrt{m_j}}{U}\right) + \Lambda m_j^{\alpha}\left(e^{-\theta y} + (1 + |\mathscr{C}|)\, e^{-\theta x}\right),$$

where X_k denotes $\delta_{x_k} - P$ for all k.

Of course we may assume that $Z^{(j)} = m_j^{-1/2}\sum_{k \in H_j} B_k$, where the $B - s$ are themselves independent and regular Brownian bridges. To begin with the technical part of the proof of Theorem 2, we fix an integer n. Then, given J such that $n \in H_J$, we set $J_0 = [J^{d/(d+\zeta)}]$. We define $\Delta_j = \sum_{k \in H_j}(X_k - B_k)$ for all j. Now we decompose the difference between the partial sums as

$$\sum_{k=1}^{n} (X_k - B_k) = D_1 + D_2 + D_3 + D_4,$$

where

$$D_1 = \sum_{k=1}^{n_{J-1}} \left(X_k - X_k \circ \Pi_{J_0}\right) - \sum_{k=1}^{n_{J-1}} \left(B_k \circ \Pi_{J_0} - B_k\right),$$

$$D_2 = \sum_{k=1+n_{J-1}}^{n} X_k - \sum_{k=1+n_{J-1}}^{n} B_k,$$

$$D_3 = \sum_{j=1}^{J_0} \Delta_j \circ \Pi_{J_0}$$

and

$$D_4 = \sum_{j=1+J_0}^{J} \Delta_j \circ \Pi_{J_0} \quad \text{if } J_0 < J \text{ or } D_4 = 0 \text{ if } J = J_0.$$

Control of D_1. As was already mentioned in the proof of Theorem 1, inequality (2.3) holds for a Brownian bridge as well.

So, since $\sqrt{n_J}\,2^{-\mu_{J_0}(1-\zeta)/2} \le a_1 n^{1/2 - 1/(2\rho)}$ for some positive constant a_1, applying (2.3) twice with $T = t/8a_1 + CLn$, we get

$$\Pr{}^*\left(\bigvee_{\mathscr{S}} |D_1| > \sqrt{n}\, n^{-(1-\zeta)/2(d+1)}\left(\frac{t}{4} + CLn\right)\right) \le \Lambda e^{-\theta t}.$$

Control of D_2.

$$\bigvee_{\mathscr{S}} |D_2| \le \bigvee_{\mathscr{S}} \left|\sum_{k=1+n_{J-1}}^{n} X_k\right| + \bigvee_{\mathscr{S}} \left|\sum_{k=1+n_{J-1}}^{n} B_k\right|.$$

On the one hand we use as at the end of the proof of Theorem 1 the Hoeffding-type inequalities of Massart when $\zeta = 0$ or of Alexander when $\zeta > 0$ to bound the first term of the above sum, and on the other hand we use the Fernique inequality [see Fernique (1970)] to bound the second term. Thus the following inequality is valid for all positive T:

$$\Pr{}^*\left(\bigvee_{\mathscr{S}}|D_2| > \sqrt{m_J}\,T\right) \le \Lambda e^{-T^2/4}.$$

Noting that $\sqrt{m_J} \le a_2 n^{1/2-1/2\rho}$ for some positive constant a_2, we apply the above inequality with $T = \sqrt{t}/4a_2$ and get

$$\Pr{}^*\left(\bigvee_{\mathscr{S}}|D_2| > n^{1/2-1/2\rho}\frac{\sqrt{t}}{4}\right) \le \Lambda e^{-\theta t}.$$

Control of D_3. We define $\tilde{\mathscr{S}}_{j,k} = \{S_j,\ S \in \mathscr{S}_k\}$ for all $j \le k$. So for all $j \le J_0$, we have

$$\bigvee_{\mathscr{S}_{J_0}}|\Delta_j| \le \sup_{S\in\mathscr{S}_{J_0}}\left|\Delta_j(S_j \setminus S)\right| + \bigvee_{\tilde{\mathscr{S}}_{j,J_0}}|\Delta_j|. \tag{4.4}$$

To bound the first term in (4.4), note that via (2.2), we have

$$\bigcup_{l\in\mathbb{Z}_+}\mathscr{S}_l \quad \text{fulfills condition UM.} \tag{4.5}$$

Thus $P(S_j \setminus S) \le K2^{-\mu_j}$ for all $S \in \mathscr{S}_{J_0}$. So using Bernstein's inequality on the one hand and a classical inequality for Gaussian variables on the other, we get as in the proof of (2.3) that for all positive T

$$\Pr\left(\sup_{S\in\mathscr{S}_{J_0}}\left|\Delta_j(S_j\setminus S)\right| > m_{J_0}^{1/2-(1-\zeta)/2d}(T + CLn)\right) \le \Lambda e^{-\theta T}.$$

Now, note that the following bounds are available:

$$\sqrt{J_0}\, m_{J_0}^{\zeta/2d} \le a_3 n^{1/2(d+1)(\zeta+d(1-\zeta)/(d+\zeta))} \le a_3 n^{1/2(d+1)}, \tag{4.6}$$

$$m_{J_0}^{1/2-(1-\zeta)/2d} \le a_3' n^{1/2-(2-\zeta)/2(d+1)}. \tag{4.7}$$

So, using the above estimate of $\sup_{S\in\mathscr{S}_{J_0}}|\Delta_j(S_j\setminus S)|$ with $T = t/12a_3'a_3$, we get

$$\Pr\left(\sup_{S\in\mathscr{S}_{J_0}}\left|\Delta_j(S_j\setminus S)\right| > n^{1/2-(2-\zeta)/2(d+1)}\left(\frac{t}{12a_3} + CLn\right)\right) \le \Lambda e^{-\theta t}. \tag{4.8}$$

To bound the second term in (4.4), note that, setting $\tilde{\mathscr{S}} = \bigcup_{j\le k}\tilde{\mathscr{S}}_{j,k}$, we have $\bigvee_{\tilde{\mathscr{S}}} s_{d-1} \le K$ because of (4.5). Also $|\tilde{\mathscr{S}}_{j,J_0}| \le |\mathscr{S}_{J_0}|$ for all $j \le J_0$.

Then we use (4.3) with $\mathscr{C} = \tilde{\mathscr{S}}_{j,J_0}$, $\chi = K2^{\mu_j(d-1)}$, $y = aLn + bt$, $x = y + a2^{\mu_{J_0}\zeta}$, $T = m_j^{-1/4}2^{\zeta\mu_{J_0}/2}(t/12a_3'a_3 + CLn)$ and $U = 2(t/12a_3'a_3 + CLn)$.

Via an appropriate choice of the constants a and b, we get

$$\Pr\left(\bigvee_{\tilde{\mathscr{S}}_{j,J_0}}|\Delta_j| > Ln^{\varepsilon(\zeta)}n^{1/2-(2-\zeta)/2(d+1)}\left(\frac{t}{6a_3} + CLn\right)\right) \le \Lambda e^{-\theta t},$$

where $\varepsilon(0) = 1/2$ and $\varepsilon(\zeta) = 0$ if $\zeta > 0$. This bound, together with (4.8), yields

$$\Pr(\Theta^c) \le \Lambda e^{-\theta t}, \tag{4.9}$$

where, setting $M_n(t) = Ln^{\varepsilon(\zeta)} n^{1/2-(2-\zeta)/2(d+1)}(t/4a_3 + CLn)$, Θ denotes the event

$$\Theta = \left(\max_{1 \le j \le J_0} \bigvee_{\mathscr{S}_{J_0}} |\Delta_j| \le M_n(t) \right).$$

Of course, possibly enlarging C once more, we may always assume that this estimate is also valid:

$$\Pr(\Theta^c) \le J_0^{-2} m_{J_0}^{-1}. \tag{4.10}$$

Now we pass to the control of D_3 itself. Defining

$$z_j = \Delta_j \mathbb{1}_{(\bigvee_{\mathscr{S}_{J_0}} |\Delta_j| \le M_n(t))}$$

we have, using Hoeffding's inequality and (2.1),

$$\Pr\left(\bigvee_{\mathscr{S}_{J_0}} \left| \sum_{j=1}^{J_0} (z_j - E(z_j)) \right| > (t + CLn)^{1/2} \sqrt{J_0}\, 2^{\zeta\mu_{J_0}/2} M_n(t) \right) \le \Lambda e^{-\theta t}. \tag{4.11}$$

However, the term

$$\bigvee_{\mathscr{S}_{J_0}} \left| \sum_{j=1}^{J_0} E(z_j) \right|$$

is negligible; more precisely, it is bounded by 1 because using the Cauchy–Schwarz inequality and (4.10), we have

$$\bigvee_{\mathscr{S}_{J_0}} |E(z_j)| \le (\Pr(\Theta^c))^{1/2} \bigvee_{\mathscr{S}_{J_0}} \left(E\left(\Delta_j^2\right)\right)^{1/2} \le J_0^{-1},$$

so that (4.11) leads, via (4.6), to

$$\Pr\left(\bigvee_{\mathscr{S}_{J_0}} \left| \sum_{j=1}^{J_0} z_j \right| > (t + CLn)^{1/2} \left(\frac{t}{4} + CLn \right) a_n(\zeta, d+1) \right) \le \Lambda e^{-\theta t}. \tag{4.12}$$

Now (4.9) and (4.12) yield

$$\Pr\left(\bigvee_{\mathscr{S}} |D_3| > (t + CLn)^{1/2} \left(\frac{t}{4} + CLn \right) a_n(\zeta, d+1) \right) \le \Lambda e^{-\theta t}.$$

Control of D_4. (We may assume that $J_0 < J$.) As for the control of D_3, we first bound a generic term of the sum, namely, $\Delta_j \circ \Pi_{J_0}$.

Since, $(\mathscr{B}_j)_{j \in \mathbf{Z}_+}$ is a filtration, we may apply (4.3) with $\mathscr{C} = \mathscr{S}_{J_0}$, $\chi = K 2^{(d-1)\mu_{J_0}}$, $y = aLn + bt$, $x = y + a 2^{\zeta\mu_{J_0}}$, $T = m_j^{-1/4} 2^{\zeta\mu_{J_0}/2}(t/8a_4' a_4 + CLn)$ and $U =$

$2(t/8a_4'a_4 + CLn)$, where a_4' and a_4 are some positive constants for which

$$m_J^{1/4} 2^{\zeta\mu_{J_0}/2} \le a_4' n^{(d+3\zeta)/4(d+1)},$$
$$\sqrt{J}\, m_{J_0}^{\zeta/2d} \le a_4 n^{1/2(d+1)}.$$

[Note that $(d + 3\zeta)/4(d + 1) \le 1/2 - (2 - \zeta)/2(d + 1)$ when $d \ge 3$ or $\zeta = 0$.]

Via an appropriate choice of the constants a and b, we get (in the case where $d \ge 3$ or $\zeta = 0$)

$$\Pr\left(\max_{J_0 < j \le J} \bigvee_{\mathscr{S}} |\Delta_j \circ \Pi_{J_0}| > L n^{\varepsilon(\zeta)} n^{1/2-(2-\zeta)/2(d+1)} \left(\frac{t}{4a_4} + CLn \right) \right) \le \Lambda e^{-\theta t}.$$

Then using the same argument as for the control of D_3, we have

$$\Pr\left(\bigvee_{\mathscr{S}} |D_4| > (t + CLn)^{1/2} \left(\frac{t}{4} + CLn \right) a_n(\zeta, d + 1) \right) \le \Lambda e^{-\theta t}$$

when $d \ge 3$ or $\zeta = 0$.

We complete the proof of Theorem 2 by collecting the above estimates of D_1, D_2, D_3 and D_4 and noting that the case where $d = 2$ and $\zeta > 0$ may be handled using the same method as above and choosing this time $\rho = (3 + \zeta)/(1 - \zeta)$ and $J_0 = [J^{1/(1+\zeta)}]$. □

5. Strong approximation for set-indexed partial-sum processes. In this section $\mathscr{S}$ is a class of Borel subsets of $[0,1]^d$ and Ω is an appropriate probability space. Theorem 4 is a multivariate analogue of Theorem 1 of KMT (1976).

THEOREM 4. *Let F be a probability law on* $\mathbb{R}$ *such that*

$$\int x \, dF(x) = 0,$$

$$\int x^2 \, dF(x) = 1$$

and

$$\int e^{tx} \, dF(x) < \infty \quad \text{for } |t| < t_0.$$

Assume that $\mathscr{S}$ *is a VČ class fulfilling condition UM.*

Then an array $\{Y_i;\ i \in \mathbb{Z}_+^d\}$ *of independent standard normal random variables and an array* $\{X_i;\ i \in \mathbb{Z}_+^d\}$ *of independent random variables with common law F may be constructed in such a way that for any positive t and any integer* ν

$$\Pr\left(\sup_{S \in \mathscr{S}} \nu^{-d/2} \left| \sum_{i \in \nu S} (X_i - Y_i) \right| > \sqrt{\frac{L\nu}{\nu}} (t + CL\nu) \right) \le \Lambda e^{-\theta t}$$

for some constants Λ, C *and* θ *depending on* $\mathscr{S}$ *and F.*

In particular,

$$\sup_{S \in \mathscr{S}} \nu^{-d/2} \left| \sum_{i \in \nu S} (X_i - Y_i) \right| = O\left(\frac{(L\nu)^{3/2}}{\sqrt{\nu}} \right) \quad a.s.$$

COMMENTS. The notation for these comments is that of Theorem 4.

(a) Note that the above construction of $\{X_i;\ i \in \mathbb{Z}_+^d\}$ does not depend on $\mathscr{S}$.

(b) Note that no smoothing of the partial-sum process is required in the above result; this is not surprising in view of the central limit theorem (Corollary 4.4) of Alexander (1987). From that point of view, our Theorem 4 means that, in some sense under a strong moment condition, the rate of convergence in Alexander's central limit theorem is of the order of $L\nu^{3/2}\nu^{-1/2}$. Of course this rate is related to that of Theorem 1 in the VČ case (setting $n = \nu^d$) because of the exponential moment condition. For the sake of conciseness, the case where $\mathscr{S}$ fulfills $H(\zeta)$ for some $0 < \zeta < 1$ will not be considered in this paper. In that case some smoothing of the partial-sum process is necessary [as pointed out by Bass and Pyke (1984)] and our method also leads to an analogue of Theorem 1 for the smoothed partial-sum process, this time with rate $L\nu\nu^{-(1-\zeta)/2}$ [see Massart (1987)]. Note that, though these rates of convergence are much more efficient than those given by Morrow and Philipp (1986) in their Theorem 1 or Theorem 5, our moment assumption is more restrictive than that of Morrow and Philipp. Nevertheless, we believe that in the spirit of the work of KMT (1976) and Major (1976), our method could lead to good results under weaker moment conditions, too.

PROOF OF THEOREM 4. Let ∂_d denote the boundary of the unit cube $[-1,1]^d$, that is, the unit "sphere" $\partial_d = \{y \in \mathbb{R}^d;\ |y| = 1\}$.

We define a one-to-one mapping φ from $\mathbb{Z}_+$ onto $\mathbb{Z}_+^d$ as follows: Given an integer n, let ν be the first integer greater than or equal to $n^{1/d}$ and then let $\varphi(n)$ be the $(n - (\nu - 1)^d)$th point of $\mathbb{Z}_+^d \cap (\nu\partial_d)$ with respect to the lexicographical ordering (our construction consists in "pealing" $\mathbb{Z}_+^d$ sphere by sphere).

Now we fix an integer ν. For any $B \in \mathscr{B}_\nu$ we set (with a comprehensive confusion in the notation)

$$\varphi^{-1}(B) = \bigcup_{C_\nu^{(i)} \subset B} \varphi^{-1}(i).$$

Next, "reading" a given sphere $\mathbb{Z}_+^d \cap (\mu\partial_d)$ following the lexicographical ordering consists in shattering this sphere in at most $d\mu^{d-2}$ lines (each line is a segment which is parallel to one of the coordinate axes). Now following one of these lines, each time we meet an extremity of a component of $\varphi^{-1}(B)$, we also cross an elementary face of the boundary of B at the same time, except maybe for the last point on the line.

Thus the number of elementary faces of the boundary of B that we cross when pealing all the spheres $(\mu\partial_d) \cap \mathbb{Z}_+^d$ with $1 \le \mu \le \nu$ being less than

$\nu^{d-1}s_{d-1}(B)$, we have

$$2c(\varphi^{-1}(B)) \le \nu^{d-1}s_{d-1}(B) + d\sum_{\mu\le\nu}\mu^{d-2},$$

getting the following nice property for φ:

$$c(\varphi^{-1}(B)) \le \frac{\nu^{d-1}}{2}(s_{d-1}(B) + d) \quad \text{for any } B \in \mathscr{B}_\nu. \tag{5.1}$$

Given $S \in \mathscr{S}$, we set $S_\nu^* = \bigcup_{i\in\mathbb{Z}_+^d \cap \nu S} C_\nu^{(i)}$. Let $\mathscr{S}_\nu^* = \{S_\nu^*,\ S \in \mathscr{S}\}$. We claim that $\mathscr{S}_\nu^*$ has the two properties

$$|\mathscr{S}_\nu^*| \le C_0\nu^{rd}, \tag{5.2}$$

$$c(\varphi^{-1}(S_\nu^*)) \le K\nu^{d-1} \quad \text{for any } S \text{ in } \mathscr{S} \tag{5.3}$$

for some constants C_0, K and r depending on $\mathscr{S}$.

In fact $\mathscr{S}_\nu^*$ has the same cardinality as the class $(\nu^{-1}\mathbb{Z}_+^d) \cap \mathscr{S}$. So, since $\mathscr{S}$ is a VČ class and $|\nu^{-1}\mathbb{Z}_+^d \cap [0,1]^d| = \nu^d$, the Vapnik–Červonenkis lemma [see Assouad (1983)] yields (5.2).

In order to prove (5.3), note that $S_\nu^* \triangle S \subset (\partial S)^{\nu^{-1}}$, so $(\partial S_\nu^*)^{(1/2)\nu^{-1}} \subset (\partial S)^{(3/2)\nu^{-1}}$. Then since $\mathscr{S}$ fulfills condition UM, we get [using (2.5)] that $\bigvee_{\mathscr{S}_\nu^*} s_{d-1} \le K'$ for some constant K'. Thus (5.1) leads to (5.3).

Defining the sequence $(Y_n)_{n\in\mathbb{Z}_+}$ from the array $\{Y_j;\ j \in \mathbb{Z}_+^d\}$ by $Y_n = Y_{\varphi(n)}$, we may apply Theorem 3 and consider the sequence $(X_n)_{n\in\mathbb{Z}_+}$ constructed in Theorem 3. We define the array $\{X_j;\ j \in \mathbb{Z}_+^d\}$ by $X_j = X_{\varphi^{-1}(j)}$.

Using (5.2) and (5.3), (3.1) leads to Theorem 4 when setting $\mathscr{A} = \{\varphi^{-1}(S_\nu^*);\ S \in \mathscr{S}\}$, $\chi = K\nu^{d-1}$, $n = \nu^d$ and $x = y = at + bL\nu$ via an appropriate choice of the constants a and b. [Note that the constants Λ and θ in Theorem 4 are not necessary the same as in (3.1).] □

Acknowledgments. I wish to express my gratitude to my thesis advisor Professor Jean Bretagnolle and to the referee who concurred to simplify the proof of the main theorem of this paper, making it much more concise and readable.

REFERENCES

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067.

Alexander, K. S. (1987). Central limit theorems for stochastic processes under random entropy conditions. *Probab. Theory Related Fields* **75** 351–378.

Alexander, K. S. and Pyke, R. (1986). A uniform central limit theorem for set-indexed partial-sum processes with finite variance. *Ann. Probab.* **14** 582–597.

Assouad, P. (1983). Densité et dimension. *Ann. Inst. Fourier* (*Grenoble*) **33** (3) 233–282.

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.

Beck, J. (1985). Lower bounds on the approximation of the multivariate empirical process. *Z. Wahrsch. verw. Gebiete* **70** 289–306.

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

BERKES, I. and PHILIPP, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54.

BORISOV, I. S. (1981). On the accuracy of the approximation of empirical random fields. *Theory Probab. Appl.* **26** 632–633.

BORISOV, I. S. (1982). Rate of convergence in invariance principle in linear spaces. *Applications to Empirical Measures. Lecture Notes in Math.* **1021** 45–58. Springer, Berlin.

CSÖRGŐ, M. and RÉVÉSZ, P. (1975). A new method to prove Strassen-type laws of invariance principle. II. *Z. Wahrsch. verw. Gebiete* **31** 261–269.

DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330.

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

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929. Correction **7** (1979) 909–911.

DUDLEY, R. M. (1982). A course on empirical processes. *Ecole d'Eté de Probabilités de Saint-Flour XII—1982. Lecture Notes in Math.* **1097** 1–142. Springer, Berlin.

DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Gebiete* **62** 509–552.

FERNIQUE, X. (1970). Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris Ser. A* **270** 1698–1699.

GAENSSLER, P. (1983). *Empirical Processes.* IMS, Hayward, Calif.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df. I. *Z. Wahrsch. verw. Gebiete* **32** 111–131.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent rv's and the sample df. II. *Z. Wahrsch. verw. Gebiete* **34** 33–58.

MAJOR, P. (1976). The approximation of partial sums of independent rv's. *Z. Wahrsch. verw. Gebiete* **35** 213–220.

MASSART, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.* **22** 381–423.

MASSART, P. (1987). Quelques problèmes de vitesse de convergence pour des processus empiriques. Thèse de doctorat d'Etat, Université de Paris-Sud, Orsay.

MORROW, E. J. and PHILIPP, W. (1986). Invariance principles for partial sum processes and empirical processes indexed by sets. *Probab. Theory Related Fields* **73** 11–42.

PYKE, R. (1984). Asymptotic results for empirical and partial-sum processes: A review. *Canad. J. Statist.* **12** 241–264.

RÉVÉSZ, P. (1976a). On strong approximation of the multidimensional empirical process. *Ann. Probab.* **4** 729–743.

RÉVÉSZ, P. (1976b). Three theorems of multivariate empirical process. *Empirical Distributions and Processes. Lecture Notes in Math.* **566** 106–126. Springer, Berlin.

SKOROHOD, A. V. (1976). On a representation of random variables. *Theory Probab. Appl.* **21** 628–632.

TUSNÁDY, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Period. Math. Hungar.* **8** 53–55.

UNIVERSITÉ PARIS–SUD
U.A. CNRS 743 "STATISTIQUE APPLIQUÉE"
MATHÉMATIQUES
BÂT. 425
91405 ORSAY CEDEX
FRANCE