The Annals of Probability
1993, Vol. 21, No. 3, 1706–1727

STRONG APPROXIMATION FOR SET-INDEXED PARTIAL-SUM PROCESSES, VIA KMT CONSTRUCTIONS II

BY EMMANUEL RIO

CNRS, Orsay

Let $(X_i)_{i \in \mathbb{Z}_+^d}$ be an array of zero-mean independent identically distributed random vectors with values in $\mathbb{R}^k$ with finite variance, and let $\mathscr{S}$ be a class of Borel subsets of $[0,1]^d$. If, for the usual metric, $\mathscr{S}$ is totally bounded and has a convergent entropy integral, we obtain a strong invariance principle for an appropriately smoothed version of the partial-sum process $\{\Sigma_{i \in \nu S} X_i \colon S \in \mathscr{S}\}$ with an error term depending only on $\mathscr{S}$ and on the tail distribution of X_1. In particular, when $\mathscr{S}$ is the class of subsets of $[0,1]^d$ with α-differentiable boundaries introduced by Dudley, we prove that our result is optimal.

1. Introduction. In this paper, we continue the research started in our previous paper I. So, the purpose of this paper is to establish strong invariance principles for partial sum processes indexed by a family of subsets of the unit cube $[0,1]^d$. As motivation and potential application for our results, we refer the reader to Pyke's review (1984). The context of the problem is as follows. Let $(X_i)_{i \in \mathbb{Z}_+^d}$ be an array of i.i.d. $\mathbb{R}^k$-valued random vectors with mean zero and finite variance. If $\mathscr{S}$ is any collection of Borel subsets of $[0,1]^d$, we define the smoothed partial-sum process $\{X(\nu S) \colon S \in \mathscr{S}\}$ by

$$X(\nu S) = \sum_{i \in \mathbb{Z}_+^d} \lambda([i - \mathbb{1}, i] \cap \nu S) X_i, \tag{1.0}$$

where $\mathbb{1} = (1, \dots, 1)$, $[i - \mathbb{1}, i]$ is the unit cube with upper right vertex i, and λ is the Lebesgue measure on $\mathbb{R}^d$.

When $d = 1$ and $\mathscr{S} = \{[0,t], t \in [0,1]\}$, where $(X_i)_{i \geq 1}$ is a sequence of i.i.d. $\mathbb{R}$-valued random variables with a finite rth moment, Komlós, Major and Tusnády (1975, 1976) proved that a sequence $(Y_i)_{i \geq 1}$ of i.i.d. Gaussian variables may be constructed in such a way that, denoting by Y_ν the partial-sum process associated with $(Y_i)_{i \geq 1}$,

$$\sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = o(\nu^{1/r}) \quad \text{a.s.}$$

Moreover, if the moment-generating function of X_1 is finite in a neighborhood of 0,

$$\sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = O(\log \nu) \quad \text{a.s.}$$

Received December 1990; revised January 1992.
AMS 1991 *subject classifications*. 60F17, 62G99.
Key words and phrases. Central limit theorem, set-indexed process, partial-sum process, invariance principle, metric entropy with inclusion, multivariate empirical processes.

It is worth noticing that the rates of strong approximation appearing above are optimal: This comes from Breiman's remark (1967) when the rth moment is finite and from Bartfai (1966) when the moment-generating function is finite. Recently, Einmahl (1987, 1989) extended these results to $\mathbb{R}^k$-valued random vectors and to more general moment conditions.

Our aim is to obtain optimal rates of convergence in strong invariance principles for multidimensionally-indexed partial-sum processes. The general approach is analogous to that introduced by Komlós, Major and Tusnády (KMT). In the spirit of Massart (1989), the methods are then extended to obtain strong invariance principles for multidimensionally-indexed processes. Recall that Massart (1989) obtained optimal rates in the strong invariance principle for real-valued set-indexed partial-sum processes, when $\mathscr{S}$ is not too large (i.e., $\mathscr{S}$ is a Vapnik–Chervonenkis class or $\mathscr{S}$ fulfills a suitable condition of entropy with inclusion). However, he had to assume the existence of the moment-generating function of the random variables. Recently, we generalized Massart's results for Vapnik–Chervonenkis classes of sets to $\mathbb{R}^k$-valued random vectors and weaker moment conditions. Here, using the modification of KMT's dyadic scheme previously introduced in Part I, we study the rates of convergence in the strong invariance principle for partial sum processes indexed by classes $\mathscr{S}$ of sets having an *integrable entropy with inclusion* and fulfilling some uniform smoothness condition on the boundaries of elements of $\mathscr{S}$. Due to the largeness of these classes, it is necessary to consider a smoothed version of the partial sum process. In this case, Bass (1985) and Alexander and Pyke (1986) obtained recently a functional LIL and an uniform CLT, when only the second moment of X_1 is assumed to be finite. In a slightly more general context, the uniform central limit theorem of Bass and Pyke (1984), and the strong invariance principles with rates of convergence of Morrow and Philipp (1986) required that X_1 satisfy a moment condition which becomes more restrictive as the size of $\mathscr{S}$ increases. By contrast, the strong invariance principle proved in this paper requires only that the second moment of X_1 be finite.

Now, we discuss further the scope of results: We need an extra condition on the boundaries of the elements of $\mathscr{S}$. Given a norm $|\cdot|$ on $\mathbb{R}^d$ and a subset S of $\mathbb{R}^d$, we set

$$(\partial S)^{\epsilon} = \{y \in \mathbb{R}^d \colon |y - z| < \epsilon \text{ for some } z \in \partial S\}$$

and we make the following standing assumption on $\mathscr{S}$:

$$\sup_{S \in \mathscr{S}} \lambda\big((\partial S)^{\epsilon}\big) \le K\epsilon^{\delta} \quad \text{for any } \epsilon \in]0,1], \text{ for some } \delta \in]0,1]. \tag{1.1}$$

When $\delta = 1$, this condition is the uniform Minkowsky condition previously used by Massart (1989) and Bass and Pyke (1984). Let $H(\epsilon)$ be the entropy with inclusion of $\mathscr{S}$ related to the pseudometric d_λ on $\mathscr{S}$ defined in Section 2. Let us define the entropy integral function $I(\cdot)$ by

$$I(x) = \int_0^x (H(u)/u)^{1/2}\, du.$$

This function is an upper bound on the modulus of continuity of the Brownian process W indexed by $\mathscr{S}$ with mean zero and covariance $\mathrm{Cov}(W(A), W(B)) = \lambda(A \cap B)\mathrm{Var}\, X_1$, which is the limit of the normalized processes $\{\nu^{-d/2}X(\nu S)$: $\nu \in \mathscr{S}\}$ as $\nu \to +\infty$ [see Dudley (1973) for more about Gaussian processes].

When $\mathscr{S}$ satisfies (1.1) and has an integrable entropy with inclusion, and $E(|X_1|^r) < +\infty$ for some large enough r, we obtain an almost sure strong invariance principle with rate of convergence $O(I(\nu^{-\delta}))$. Moreover, when $\mathscr{S}$ is the class of sets with α-differentiable boundaries introduced in Dudley (1974), we prove in Section 4 that this result is optimal.

On the other hand, when $\mathscr{S}$ is not too large and r is closer to 2, we prove that the rate of approximation is of the order of $o(\nu^{d(1/r-1/2)})$ and this result still is optimal.

2. Definitions and results. Throughout this paper, the probability space Ω is assumed to be such that there exists an atomless real-valued random variable, defined on Ω, which is independent of the observations. Let $\mathscr{S}$ be a family of Borel subsets of the unit cube $[0,1]^d$ satisfying the smoothness condition (1.1) for some positive δ. In order to get nice asymptotic properties for a normalized version of the smoothed partial-sum process $\{X(\nu S)$: $S \in \mathscr{S}\}$ defined by (1.0), we have to put some additional conditions on $\mathscr{S}$. Let us define the pseudometric d_λ by $d_\lambda(A, B) = \lambda(A \,\Delta\, B)$, where λ is the Lebesgue measure on $\mathbb{R}^d$. We shall assume that $\mathscr{S}$ is *totally bounded with inclusion* and has a *convergent entropy integral* with respect to d_λ. This means that, first, for every positive ε there exists a finite collection (called an ε-net) $\mathscr{S}(\varepsilon)$ such that for any S in $\mathscr{S}$, there exits S^+ and S^- in $\mathscr{S}(\varepsilon)$ with $S^- \subset S \subset S^+$ and $d_\lambda(S^-, S^+) \le \varepsilon$, and second, that the minimal cardinality of such a collection $\mathscr{S}(\varepsilon)$ which we denote by $N_I(\varepsilon, \mathscr{S})$ satisfies

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

Define

$$H(\varepsilon) = \log N_I(\varepsilon, \mathscr{S}) \quad \text{and} \quad I(x) = \int_0^x \left(\varepsilon^{-1} N_I(\varepsilon, \mathscr{S})\right)^{1/2} d\varepsilon.$$

We may, by enlarging the class $\mathscr{S}$ a little if necessary, assume that $H(\varepsilon) > |\log \varepsilon|$ for any ε in $]0,1[$. In order to get a strong invariance principle when only the second moment of Q is assumed to be finite, we shall assume that $\mathscr{S}$ is *contraction closed*. This means that

$$\text{for all } t \in]0,1[, \quad \text{for all } S \in \mathscr{S}, \qquad tS \in \mathscr{S}. \tag{2.2}$$

In many cases of interest, there is no loss of generality in making this assumption, because if the approximating collections $\mathscr{S}(\varepsilon)$ satisfy (1.1) for some constants δ and K, and if $\mathscr{S}^* = \{tS: S \in \mathscr{S}, 0 \le t \le 1\}$ has an entropy function $H^*(\varepsilon)$, then it is obvious that $\mathscr{S}^*$ is totally bounded with inclusion and satisfies $H^*(\varepsilon) = O(H(\varepsilon))$.

Now, let us state our main result, which provides an invariance principle with an error term depending only on the moment of the r.v.'s X_i and on the entropy with inclusion of $\mathscr{S}$. As Einmahl does, let us introduce more general moment functions. Let $Lx = \log(x \vee e)$ and $LLx = L(Lx)$. Let ψ be a mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$ such that:

(2.3)
(i) $\int \psi(|x|)\, dQ(x) < +\infty$, and $x^{-2}\psi(x)$ is a one to one continuous, increasing mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$.
(ii) There exists $r > 2$ such that $x^{-r}\psi(x)$ is nonincreasing.
(iii) Furthermore, if there does not exist $s < 4$ such that $x^{-s}\psi(x)$ is nonincreasing, then $(x^2 LLx)^{-1}\psi(x)$ is nondecreasing.

Throughout, ψ^{-1} denotes the inverse function of ψ. Note that, when Q has a finite second moment, such a mapping ψ exists [see Major (1976)]. We also need to introduce the following notations. Let $H_1(\varepsilon) = \varepsilon^{-1}H(\varepsilon)$ and let H_1^{-1} be the inverse of H_1, that is, $H_1^{-1}(x) = y$ if and only if $H(y^+) \le xy \le H(y^-)$ for any positive y. For any positive nondecreasing function f, let

$$b(x, f) = \int_1^x f(u^{-1}x)\big(H_1^{-1}(u) + u^{-1/2}I(u^{-\delta/d})\big)\, du. \tag{2.4}$$

Our main result is the following.

THEOREM 1. *Let $\mathscr{S}$ be a family of Borel subsets of $[0,1]^d$. Assume that $\mathscr{S}$ is a class satisfying* (2.1), (2.2) *and* (1.1) *for some δ in* $]0,1]$. *Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance matrix, and let ψ be a mapping satisfying* (2.3). *Let $(X_i)_{i\in\mathbb{Z}_+^d}$ be an array of independent random vectors with common law Q. Then, there exists an array $(Y_i)_{i\in\mathbb{Z}_+^d}$ of independent $N(0, \operatorname{Var} Q)$-distributed random vectors such that*

(a) $$\sup_{S\in\mathscr{S}} |X(\nu S) - Y(\nu S)| = O_P\big(b(\nu^d, \psi^{-1})\big),$$

where $b(\cdot)$ is defined by (2.4), *and setting*

$$\varphi(x) = \psi^{-1}(x)\sup\Big(1, \big(x^{-1}LLx\big)^{1/2}\psi^{-1}(x)\Big),$$

(b) $$\sup_{S\in\mathscr{S}} |X(\nu S) - Y(\nu S)| = O\big(b(\nu^d, \varphi)\big) \quad a.s.$$

COMMENTS. When $E(|X_1|^2 LL(|X_1|)) < +\infty$, there exists a mapping ψ such that $(x^2LLx)^{-1}\psi(x)$ is increasing. Then, $\varphi(x) = O(\psi^{-1}(x))$ as $x \to +\infty$ and (b) of Theorem 1 holds with rate $O(b(\nu^d, \psi^{-1}))$ a.s.

Note that $\sqrt{u}\, H_1^{-1}(u) = O(I(H_1^{-1}(u)))$ and $H_1^{-1}(u) = o(u^{-1/2})$ as $u \to +\infty$. So, when $d > 1$ or $\delta \le 1/2$, we have

$$H_1^{-1}(u) + u^{-1/2}I(u^{-\delta/d}) = O\big(u^{-1/2}I(u^{-\delta/d})\big). \tag{2.5a}$$

On the other hand, when $d = 1$ and $\delta > 1/2$, it is easily seen that $\mathscr{S}$ is totally bounded with inclusion and that its log-entropy $H(\cdot)$ satisfies $H(\varepsilon) = O(\varepsilon^{1-1/\delta}|\log \varepsilon|)$ and then

$$(2.5b)\qquad H_1^{-1}(u) + u^{-1/2}I(u^{-\delta}) = O\big(u^{-1/2}I\big(u^{-\delta}|\log u|^{\delta}\big)\big).$$

Theorem 1 does not provide a uniform CLT when only the second moment is assumed to be finite in the general case: For example, if $\psi(x) = x^2(Lx)^{\beta}$, the uniform CLT requires $\beta > 2$. So, in order to get optimal results, we need to put additional conditions on $\mathscr{S}$ and on the law Q. Hence, throughout, we assume that either Q has a finite rth moment for some $r > 2$ or $\mathscr{S}$ has an entropy exponent ζ in [0, 1[. Recall that this means

$$(2.6)\qquad \limsup_{\varepsilon \to 0} |\log(H(\varepsilon))/\log \varepsilon| = \zeta.$$

If $\mathscr{S}$ has an exponent of entropy ζ in [0, 1[and if the moment of Q is between 2 and $2d(d - \delta(1-\zeta))^{-1}$, Theorem 1 and (2.5) provide a strong invariance principle with a rate depending only on ψ. The following corollary then generalizes results of Einmahl (1987, 1989) to the multidimensional case.

COROLLARY 1. *Let $\mathscr{S}$ be a family of Borel subsets of* $[0,1]^d$. *Assume that $\mathscr{S}$ is a class satisfying* (2.1), (2.2) *and* (1.1) *for some δ in*]0, 1]. *Assume $\mathscr{S}$ to be totally bounded with inclusion and to have an exponent of entropy ζ in* [0, 1[. *Let Q be a law on $\mathbb{R}^k$ with zero-mean and positive definite covariance, and let ψ be a mapping satisfying* (2.3) *for some $r < 2d(d - \delta(1-\zeta))^{-1}$. Let $(X_i)_{i \in \mathbb{Z}_+^d}$ be an array of independent random vectors with common law Q. Then, there exists an array $(Y_i)_{i \in \mathbb{Z}_+^d}$ of independent $N(0, \mathrm{Var}\, Q)$-distributed random vectors such that*

$$(a)\qquad \sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = o_P(\psi^{-1}(\nu^d))$$

and, if we assume furthermore that $x^{-1/r}\varphi(x)$ is nondecreasing,

$$(b)\qquad \sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = o(\varphi(\nu^d)) \quad a.s.$$

COMMENTS. Clearly, this is sufficient to obtain a construction of the arrays such that (a) and (b) hold with respective rates $O_P(\psi^{-1}(\nu^d))$ and $O(\varphi(\nu^d))$ a.s.

Note that Corollary 1 provides a rate of the order of $\nu^{-d/2}\psi^{-1}(\nu^d)$ in the uniform CLT of Alexander and Pyke (1986).

Part (a) of Corollary 1 is a weak invariance principle in the sense of Philipp (1980) while (b) is a strong invariance principle where the function $x \to x^2 LLx$ plays an important role. In fact, Corollary 1 yields two different results according to the monotonicity of the function $x \to \psi(x)(x^2 LLx)^{-1}$. When $x \to \psi(x)(x^2 LLx)^{-1}$ is nondecreasing, $\varphi = O(\psi^{-1})$ and (b) of Corollary 1 holds with the error term $o(\psi^{-1}(\nu^d))$. From Breiman's remark, this result is optimal. When $x \to \psi(x)(x^2 LLx)^{-1}$ is nonincreasing, Corollary 1 yields the following Strassen's type invariance principle:

$$(\nu^d LL\nu)^{-1/2} \sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = o\big(\nu^{2d}/\psi(\nu^d)\big) \quad \text{a.s.}$$

We refer to Corollary 3 in part I [Rio (1993)] for more comments about this result. The above result and Theorem 3.1 of Bass and Pyke (1984) yield a functional law of the iterated logarithm [see Bass and Pyke (1984)]. However, the functional LIL has been proved by Bass (1985), by means of entirely different methods.

On the other hand, when the moment of Q is large enough, the rate of approximation depends only on $\mathscr{S}$. Moreover, this rate of approximation is related to the modulus of continuity $I(\varepsilon)$ of the standard Brownian process indexed by $\mathscr{S}$. However, we need to put an additional condition when the moment is smaller than $2d(d-\delta)^{-1}$. Let $0 \le \zeta \le 1$. The class $\mathscr{S}$ is said to satisfy $\mathscr{H}(\zeta)$ if $\mathscr{S}$ satisfies (2.1), (2.2) and if the following holds:

$$\mathscr{H}(\zeta) \qquad \liminf_{\varepsilon \to 0} |\log(H(\varepsilon))/\log \epsilon| \ge \zeta.$$

Then, the following result holds.

COROLLARY 2. *Let $d > 1$ and let $\mathscr{S}$ be a family of Borel subsets of the unit cube satisfying $\mathscr{H}(\zeta)$ for some ζ in* [0, 1] *and* (1.1) *for some δ in*]0, 1]. *Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite variance matrix, satisfying*

$$\int_{\mathbb{R}^k} |x|^r \, dQ(x) < +\infty \quad \text{for some } r > 2d(d - \delta(1-\zeta))^{-1}.$$

Let $(X_i)_{i \in \mathbb{Z}_+^d}$ be an array of independent random vectors with common law Q. Then, there exists an array $(Y_i)_{i \in \mathbb{Z}_+^d}$ of independent $N(0, \operatorname{Var} Q)$-distributed random vectors such that

$$\nu^{-d/2} \sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = O\big(I(\nu^{-\delta})\big) \quad a.s.$$

COMMENTS. When $\mathscr{S}$ satisfies the uniform Minkowsky condition (i.e., $\delta = 1$), Corollary 2 means that in some sense, the rate of convergence in the uniform CLT of Alexander and Pyke (1986) is of the order of $I(\nu^{-1})$. This result improves a previous result by Massart (1989).

Before discussing further our results, we give a consequence of Corollaries 1 and 2.

COROLLARY 3. *Let $d > 1$ and let $\mathscr{S}$ be a family of Borel subsets of* $[0,1]^d$ *satisfying* (2.2). *Assume that $\mathscr{S}$ is totally bounded with inclusion and has an entropy $H(\cdot)$ satisfying $H(\varepsilon) = O(\varepsilon^{-\zeta})$ as $\varepsilon \to 0$ for some ζ in*]0, 1[. *Assume that $\mathscr{S}$ fulfills* (1.1) *with $\delta = 1$. Let Q be a law on $\mathbb{R}^k$ with mean zero and* *positive definite covariance matrix, satisfying*

$$\int_{\mathbb{R}^k} |x|^r \, dQ(x) < +\infty \quad \text{for some } r > 2 \text{ such that } 2d(d-1+\zeta)^{-1} \neq r.$$

Let $(X_i)_{i \in \mathbb{Z}_+^d}$ *be an array of independent random vectors with common law* Q. *Then, there exists an array* $(Y_i)_{i \in \mathbb{Z}_+^d}$ *of independent* $N(0, \operatorname{Var} Q)$*-distributed random vectors such that*

$$\sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = O(\nu^{d/r} + \nu^{(d-1+\zeta)/2}) \quad a.s.$$

COMMENT. Among examples of index families which satisfy the conditions of Corollary 3 for some positive ζ, let us consider the following. If $\mathscr{C}^2$ denotes the class of convex subsets of $[0,1]^2$, Bronshtein (1976) has shown that $\zeta = 1/2$. Now, if $d > 1$ and if $J(\alpha, d, M)$ denotes the class of sets introduced in Dudley (1974), whose boundaries are images of α-differentiables mappings of the $(d-1)$-sphere into $\mathbb{R}^d$, with all derivatives of order up to α uniformly bounded by M, then $\zeta = (d-1)/\alpha$. In these cases, we prove that our result cannot be improved. Let us now state the related result.

THEOREM 2. *Let F and G be different probability laws on $\mathbb{R}$ with mean zero and finite fourth-moment. Let* $(X_i)_{i \in \mathbb{Z}_+^d}$ *and* $(Y_i)_{i \in \mathbb{Z}_+^d}$ *be two arrays of i.i.d. random variables with respective laws F and G, and let $\mathscr{S}$ denote either the class of convex subsets of $[0,1]^2$ or the class $J(\alpha, d, M)$. Then, there exists some positive constant $C(F, G)$ such that*

$$\liminf_{\nu} \nu^{(1-\zeta-d)/2} \sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| \geq C(F, G) \quad a.s.$$

Before proving Theorem 1, we give another Corollary in the unidimensional case, which is a by-product of (2.5) and of Theorem 1.

COROLLARY 4. *Let $d = 1$, let $\delta > 1/2$, and let $\mathscr{A}_\delta$ be the class of Borel subsets of the unit interval fulfilling* (1.1) *with δ. Let Q be a law on $\mathbb{R}^k$ with mean zero, with positive definite covariance, satisfying $\int_{\mathbb{R}^k} |x|^r \, dQ(x) < +\infty$ for some $r > (1-\delta)^{-1}$. Let $(X_i)_{i \in \mathbb{Z}_+}$ be a sequence of independent random vectors with common law Q. Then, there exists a sequence $(Y_i)_{i \in \mathbb{Z}_+}$ of independent* $N(0, \operatorname{Var} Q)$*-distributed random vectors such that*

$$\sup_{S \in \mathscr{A}(\delta)} |X(\nu S) - Y(\nu S)| = O\big(\nu^{1-\delta} (\log \nu)^{\delta}\big) \quad a.s.$$

Now, we prove Theorem 1. The proof of this theorem is based on the methods of a common probability space previously introduced by Komlós, Major, and Tusnády. Here, our method of construction of the two arrays of independent random vectors is exactly the same as in Part I. Yet, we need only recall some of its basic properties.

3. Strong approximation. Throughout this section, Q is a law on $\mathbb{R}^k$ with zero-mean and finite variance. $(X_i)_{i\in\mathbb{Z}_+^d}$ denotes an array of independent random vectors with common law Q, and ψ is any mapping satisfying (2.3).

In order to construct the two arrays $(X_i)_{i\in\mathbb{Z}_+^d}$ and $(Y_i)_{i\in\mathbb{Z}_+^d}$ on our rich enough space, we first construct two sequences $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ of independent identically distributed random vectors with respective distributions Q and $N(0,\operatorname{Var} X_1)$. This construction is exactly the same as in I. And second, by means of the one to one mapping σ from $\mathbb{Z}_+^d$ onto $\mathbb{Z}_+$ which was defined in I, we will turn the so defined sequences into arrays.

So, let σ be the one to one mapping defined in part I [Rio (1993), Lemma 0, Section 3] and set $Y_i = Y_{\sigma(i)}$ and $X_i = X_{\sigma(i)}$ for any $i\in\mathbb{Z}_+^d$. Let ν be any positive integer and let N be the smallest integer such that $2^N\geq\nu$. We define the class $\mathscr{A}_\nu$ of functions of $\ell^2(\mathbb{Z}^+)$ with support included in $]0,2^{Nd}]$ as follows: If B is any subset of $\mathbb{Z}_+^d$, let σ^*B be the mapping from $\mathbb{Z}_+$ to $[0,1]$ defined by

$$\sigma^*B = f \quad\text{iff}\quad f(\sigma i) = \lambda([i-\mathbb{1}, i]\cap B) \quad\text{for any } i\in\mathbb{Z}_+^d$$

and, for each integer ν, define the class $\mathscr{A}_\nu$ by $\mathscr{A}_\nu = \{\sigma^*B\colon B\in\nu\mathscr{S}\}$. When $\mathscr{S}$ is contraction closed, $(\mathscr{A}_\nu)_{\nu>0}$ is a nondecreasing sequence of families of elements of $\ell^2(\mathbb{Z}_+)$. Now, given a function f in $\ell^2(\mathbb{Z}_+)$ with finite support and a sequence $(u_i)_{i>0}$ of vectors of $\mathbb{R}^k$, we set $u(f) = \sum_{i>0} f(i)u_i$. Clearly,

$$\sup_{S\in\mathscr{S}} |X(\nu S) - Y(\nu S)| = \sup_{f\in\mathscr{A}_\nu} |X(f) - Y(f)|, \tag{3.1}$$

where X and Y denote either the sequences or the arrays. So, throughout the sequel, we work with the sequences $(X_i)_{i>0}$ and $(Y_i)_{i>0}$ defined in our previous paper. Throughout, the intervals $]l,m]$ have to be interpreted as subsets of $\mathbb{Z}_+$. $\ell^2(\mathbb{Z}_+)$ is given the canonical inner product, which we denote by $(\cdot\mid\cdot)$, and $\ell^2(]l,m])$ denotes the subspace of $\ell^2(\mathbb{Z}_+)$ of functions with support included in $]l,m]$. Let $I_{j,p} =]p2^j,(p+1)2^j]$, and let $e_{j,p}$ be the characteristic function of $I_{j,p}$. For any positive integers p and j, we set $\tilde e_{j,p} = e_{j,p} - 2e_{j-1,2p}$. Let $\tilde e_j = e_{j,1}$ and define the orthogonal systems $\mathscr{B}_0$ and $\mathscr{B}_j$ by

$$\mathscr{B}_0 = \{\tilde e_j\colon 0\leq j<Nd\}\cup\{e_{0,0}\} \quad\text{and}\quad \mathscr{B}_j = \{\tilde e_{j,p}\colon 0<p<2^{Nd-j}\}.$$

$\mathscr{B} = \bigcup_{j=0}^{Nd-1}\mathscr{B}_j$ is an orthogonal basis of $\ell^2(]0,Nd])$. Let Π_j be the orthogonal projector on the space generated by $\bigcup_{l=1}^{j}\mathscr{B}_l$. For any function bounded by 1, the control of $X(f)-Y(f)$ depends mainly on the quantities $(\Pi_j f\mid\Pi_j f)$. Exactly as in part I, the uniform control on $\mathscr{A}_\nu$ of the above inner products is insured via (1.1) and the perimetric properties of σ.

LEMMA 1 [Rio (1993)]. *Assume that $\mathscr{S}$ is a class of subsets of the unit cube fulfilling* (1.1) *for some constants $0<\delta\leq 1$ and $K\geq 1$. Then, for any element f in $\mathscr{A}_\nu$, $\Pi_j f$ has values in $[-1,1]$ and*

$$\#\{p\in\mathbb{N}\colon \Pi_j f(i)\neq 0 \text{ for some } i\in I_{j,p}\}\leq 2K\nu^{d-\delta}2^{-j(1-\delta/d)}.$$

Now, we pass to the control of the random vector $X(f) - Y(f)$. Let $(\bar{X}_i)_{i>0}$ and $(\tilde{X}_i)_{i>0}$ be the sequences defined by

$$\bar{X}_{2^l i} = \mathbb{1}_{(|X_{2^l i}| M_L)} X_{2^l i} \quad \text{and} \quad \tilde{X}_i = \bar{X}_i - \mathbb{E}(\bar{X}_i), \tag{3.2}$$

for any integer l, for any odd integer i in $[2^L, 2^{L+1}[$. Let $n = 2^{Nd}$, where N is the smallest integer such that $2^N \geq \nu$. Clearly,

$$\sup_{f \in \mathscr{A}_\nu} |X(f) - Y(f)| \leq \sum_{i=1}^{n} |X_i - \tilde{X}_i| + \sup_{f \in \mathscr{A}_\nu} |\tilde{X}(f) - Y(f)|. \tag{3.3}$$

First, by Lemma 2 in Part I [Rio (1993)], $\Sigma_{i=1}^n |X_i - \tilde{X}_i| = o(\psi^{-1}(n))$ a.s. Second, by (4.4) in Part I,

$$Y(f) - \tilde{X}(f) = D_0(f) + \sum_{j=1}^{Nd-1} D_j(\Pi_j f), \tag{3.4}$$

where the random vectors $D_j(f)$ are defined just before equation (4.4) in part I. Let D_j be the r.v.'s defined in part I, just before (4.5) (recall that $D_j = \sup_{\mathscr{A}_\nu} |D_j(\Pi_j f)|$ if $j > 0$). Then,

$$\sup_{f \in \mathscr{A}_\nu} |Y(f) - \tilde{X}(f)| \leq \sum_{j=0}^{Nd-1} D_j. \tag{3.5}$$

First, as in Part I, equation (4.7), $D_0 = O(\varphi(n))$ a.s. and $D_0 = O_P(\psi^{-1}(n))$. Second, we will use (4.6b) of Proposition 1 in Part I and an oscillation lemma to control each of the random variables D_j. In order to prove Theorem 1, it is necessary to use the entropy properties of $\mathscr{A}_\nu$. For any Borel subset B of $\mathbb{R}_+^d$, let

$$\lambda(B) = \sum_{i \in \mathbb{Z}_+} \sigma^* B(i), \tag{3.6}$$

where λ denotes Lebesgue measure. For any Borel sets A and B, $A \subset B$ implies $\sigma^* A \leq \sigma^* B$. Hence, the ℓ^1-entropy with bracketing $N(\varepsilon, \mathscr{A}_\nu, P)$ of $\mathscr{A}_\nu$ with respect to the uniform probability P on $\mathbb{Z}^+ \cap\,]0, 2^{Nd}]$ (recall N is the smallest integer such that $2^N \geq \nu$) of the family $\mathscr{A}_\nu$ satisfies

$$N(\varepsilon, \mathscr{A}_\nu, P) \leq N_I(\varepsilon, \mathscr{S}, \lambda). \tag{3.7}$$

Recall that the ℓ^1-entropy with bracketing related to P is the minimal cardinality of a collection $\mathscr{A}(\varepsilon)$ of functions such that, for any f in $\mathscr{A}_\nu$, there exist f^+ and f^- in $\mathscr{A}(\varepsilon)$ such that $f^- \leq f \leq f^+$ and $P(f^+ - f^-) \leq \varepsilon$.

Throughout, we work with the family $\mathscr{A}_\nu$. In order to prove Theorem 1, it will be necessary to use Proposition 1 and a *chaining argument*. If furthermore the entropy with bracketing is integrable, one can use a restricted chaining argument [see Pollard (1984), page 159, for an example]. By means of a restricted chaining, in the spirit of Bass (1985), we now control each of the random variables D_j. Here we state a general oscillation lemma.

Let $(\mathscr{X}, \mu)$ be a finite measure space. Furthermore, assume that μ is a probability measure, and let $\mathscr{G}$ be the space of measurable functions taking their values in $[-1,1]$. Let $\mathscr{F}$ be a family of elements of $\mathscr{G}$ satisfying:

(3.8)
(i) $\mathscr{F}$ contains the null function.
(ii) $\mathscr{F}$ has an integrable $\mathscr{L}^1(\mu)$-entropy with bracketing, which we denote by $N(\varepsilon, \mathscr{F})$.

Throughout this section, $H(\varepsilon)$ is any nonincreasing function such that $H(0) = 0$, e^H is with values in $\mathbb{Z}_+$ and $H(\varepsilon) \geq \log N(\varepsilon, \mathscr{F})$. Define $I(x)$ from $H(\cdot)$ exactly as in section 2 (i.e., $I(x) = \int_0^x \sqrt{H(u)/u}\, du$) and set $H_1(u) = u^{-1}H(u)$. Let us define the subset $\mathscr{U}_\epsilon$ of $\mathscr{F} \times \mathscr{F}$ by

$$\mathscr{U}_\epsilon = \{(f,g) \in \mathscr{F} \times \mathscr{F} \text{ such that } \|f-g\|_{1,\mu} \leq \epsilon\}.$$

Let A be a random linear process on $L^\infty(\mathscr{X})$ with values in $\mathbb{R}$. A is said to satisfy (3.9) if and only if there exist a linear positive process A^* and some constants Λ, θ and $a \leq 1$ such that:

(3.9)
(i) for any $g \in \mathscr{G}$, $|A(g)| \leq A^*(|g|)$.
(ii) For any positive ϵ, for any positive t, for any $(f,g) \in \mathscr{U}_\epsilon$, $\mathbb{P}(A(f-g) \geq (a \wedge \epsilon)t) \leq \Lambda \exp(-2\theta(\epsilon \wedge a)t^2(1+t)^{-1})$.
(iii) For any positive ϵ, for any positive g in $\mathscr{G}$ satisfying $\|g\|_{1,\mu} \leq \epsilon$, for any $t \geq 1$, $\mathbb{P}(A^*(g) \geq \epsilon t) \leq \Lambda \exp(-\theta\epsilon t)$.

When the random process A satisfies the above conditions, the following lemma holds.

LEMMA 2. *Let $\mathscr{F}$ be a class of measurable functions from $\mathscr{X}$ to* $[-1,1]$ *satisfying* (3.8) *and let A be a linear process fulfilling* (3.9). *Then, for any positive x,*

$$\mathbb{P}\left(\sup_{f\in\mathscr{F}} A(f) > 7\theta^{-1/2}I(a) + 2H_1^{-1}\left(\frac{\theta}{3}\right) + 2x + 7\sqrt{ax}\right) \leq \Lambda\left(1 + \frac{\pi^2}{6}\right)\exp(-\theta x).$$

PROOF. By (3.8), for each positive ϵ, there exists a family $\mathscr{F}_\epsilon^*$ of positive elements of $\mathscr{G}$ and a family $\mathscr{F}_\epsilon$ of elements of $\mathscr{F}$ satisfying:

1. For any $f^* \in \mathscr{F}_\epsilon^*$, $\|f^*\|_{1,\mu} \leq \epsilon$.
2. For any f in $\mathscr{F}$ there exists some $(f_\epsilon, f_\epsilon^*)$ in $\mathscr{F}_\epsilon \times \mathscr{F}_\epsilon^*$ such that $|f - f_\epsilon| \leq f_\epsilon^*$, and the minimal cardinality of such pairs is no more than $\exp(H(\epsilon))$.

Now, we may, by increasing H a little, assume that $\exp H$ is left-continuous and takes entire values. For any positive x, let $\varepsilon(x)$ be the positive number such that

$$\varepsilon(x) - 3\theta^{-1}H^+(\varepsilon(x)) \leq x \leq \varepsilon(x) - 3\theta^{-1}H(\varepsilon(x)). \tag{3.10}$$

By definition of both processes A and A^*,

$$\sup_{f\in\mathscr{F}} A(f) \le \sup_{f\in\mathscr{F}_{\varepsilon(x)}} A(f) + \sup_{f^*\in\mathscr{F}^*_{\varepsilon(x)}} A^*(f^*). \tag{3.11}$$

1. *If* $x \ge 1$. It is obvious that

$$\sup_{f\in\mathscr{F}} A(f) \le 2A^*(1).$$

So, by (iii) of (3.9), Lemma 2 holds.

2. *If* $x < 1$. Then there exists some $\varepsilon(x)$ in $[0, 1[$ satisfying (3.10). Clearly, it is enough to control the two random variables on the right-hand side in (3.11). First, by (iii) of (3.9) and by definition of $\varepsilon(x)$, it is easily seen that

$$\mathbb{P}\Bigl(\sup_{f^*\in\mathscr{F}^*_{\varepsilon(x)}} A^*(f^*) \ge \varepsilon(x)\Bigr) \le \Lambda\exp(-\theta x). \tag{3.12}$$

It remains now to give an upper bound on $\varepsilon(x)$. Here, it is sufficient to prove that

$$\varepsilon(x) \le \varepsilon = x + H_1^{-1}(\theta/3). \tag{3.13}$$

Clearly, (3.13) follows from $\varepsilon - 3\theta^{-1}H(\varepsilon) \ge x$. Now, by monotonicity of H, we have

$$3\theta^{-1}H(\varepsilon) < 3\theta^{-1}H\bigl(H_1^{-1}(\theta/3)\bigr) \le H_1^{-1}(\theta/3).$$

So, (3.13) holds true, and collecting the above inequalities, we get

$$\mathbb{P}\Bigl(\sup_{f^*\in\mathscr{F}^*_{\varepsilon(x)}} A^*(f^*) \ge x + H_1^{-1}(\theta/3)\Bigr) \le \Lambda\exp(-\theta x). \tag{3.14}$$

It remains to control the random variable $\sup_{f\in\mathscr{F}\varepsilon(x)} A(f)$. Set $h(t) = 2t^2(1+t)^{-1}$. First, if $x > a$, using the convexity of h and the definition of $\varepsilon(x)$, it is easily seen that

$$\mathbb{P}\Bigl(\sup_{f\in\mathscr{F}_{\varepsilon(x)}} A(f) \ge \varepsilon(x)\Bigr) \le \Lambda\exp(-\theta x).$$

Then, we complete the proof of Lemma 2 by collecting (3.13), (3.14) and the above inequality. Second, when $x \le a$, we need to use a restricted chaining argument. Let $\varepsilon_0 = \varepsilon(x)$. For any integer j, we set $\varepsilon_j = 2^j\varepsilon_0$, and we define the sequence $\mathscr{F}_j$ of approximating nets associated with the sequence $(\varepsilon_j)_j$ as follows: $\mathscr{F}_0 = \mathscr{F}_{\varepsilon_0}$ and, for any natural j, $\mathscr{F}_{j+1} = \mathscr{F}_j$ if and only if $H(\varepsilon_{j+1}) = H(\varepsilon_j)$, and $\mathscr{F}_{j+1} = \mathscr{F}_{\varepsilon_j+1}$ otherwise. Clearly, the so defined collection $\mathscr{F}_j$ is an ε_j-net. Hence, there exist mappings ϕ_j from $\mathscr{F}_j$ to $\mathscr{F}_{j+1}$ such that

$$\phi_j = Id_{\mathscr{F}_j} \quad \text{iff} \quad \mathscr{F}_j = \mathscr{F}_{j+1}, \quad \text{and for any } (j, f), \quad \|f - \phi_j(f)\|_{1,\mu} \le \varepsilon_{j+1}.$$

Now, let $l = \sup\{j\epsilon\mathbb{N}\colon \varepsilon_j \le a\}$. For any f_0 in $\mathscr{F}_0$, define the mappings $(f_j)_{j>0}$ by $f_{j+1} = \phi_j(f_j)$ and let $J_0 = \{j < l$ such that $\mathscr{F}_j \ne \mathscr{F}_{j+1}\} \cup \{l\}$. For any f_0 in

$\mathscr{F}_0$, one can write

$$f_0 = (f_0 - \phi_0(f_0)) + \cdots + (f_j - \phi_j(f_j)) \cdots + (f_{l-1} - f_l) + f_l.$$

Starting from the above equality, it is easily seen [cf. Pollard (1984), page 142] that, for any sequence $(t_j)_{j\geq 0}$ of positive numbers,

$$\mathbb{P}\left(\sup_{f_0\in\mathscr{F}_0} A(f_0) \geq \sum_0^l t_j\right) \leq \sum_{i\in J_0} |\mathscr{F}_j| \sup_{f\in\mathscr{F}_j} \mathbb{P}\big(A(f-\phi_j(f)) \geq t_j\big). \tag{3.15}$$

Now, we set $\varepsilon_{l+1} = a$. By (ii) of (3.9), for any j in J_0, for any positive number u_j, for every f in $\mathscr{F}_j$,

$$\mathbb{P}\big(A(f-\phi_j(f)) \geq \varepsilon_{j+1}u_j\big) \leq \Lambda \exp\big(-\theta\varepsilon_{j+1}h(u_j)\big).$$

So, setting $t_j = u_j\varepsilon_{j+1}$, and choosing u_j such that $3H(\varepsilon_j) + \theta x = \theta\varepsilon_j h(u_j)$, we get

$$|\mathscr{F}_j| \exp\big(-\theta\varepsilon_{j+1}h(t_j)\big) \leq |\mathscr{F}_j|^{-2}\exp(-\theta x).$$

From this it follows that

$$\mathbb{P}\left(\sup_{f_0\in\mathscr{F}_0} A(f_0) \geq \sum_0^l t_j\right) \leq \Lambda\frac{\pi^2}{6}\exp(-\theta x). \tag{3.16}$$

It remains to bound $\sum_j t_j$. First we note that $u_j \leq 1$ for each natural j. Now, it is obvious that, for all $u \leq 1$, $h(u) \leq \sqrt{u}$. Hence, using the definition of u_j and the above remarks, we get

$$u_j \leq 2\varepsilon_j\big(3H_1(\varepsilon_j)/\theta\big)^{1/2} + 2(\varepsilon_j x)^{1/2}. \tag{3.17}$$

Together, (3.16) and (3.17) imply that

$$\begin{aligned}&\mathbb{P}\left(\sup_{f\in\mathscr{F}_0} A(f) \geq 7\theta^{-1/2}I(a) + H_1^{-1}\left(\frac{\theta}{3}\right) + x + 7\sqrt{ax}\right)\\ &\quad\leq \Lambda\frac{\pi^2}{6}\exp(-\theta x)\end{aligned} \tag{3.18}$$

and from here on the rest of the proof of Lemma 2 is straightforward. □

To complete the proof of Theorem 1, we now control each of the random variables D_j. Clearly, there is no loss of generality in assuming that $k = 1$. Define a majorizing positive linear random process $D_j^*(\cdot)$ associated with $D_j(\Pi_j\,\cdot\,)$ by

$$D_j^*(f) = \sum_{0<p<2^{Nd-j}} 2^{-j}(f|e_{j,p})\Big(|\tilde{U}_{j,p} - \tilde{V}_{j,p}| + \big|\xi^j(e_{j,p})\big|\Big) + \sum_{i=1}^n f(i)|\xi_i^j|,$$

where the r.v.'s $U_{j,p}$ and ξ_i^j are defined in Part I [Rio (1993)], before (4.4). It is easily seen that, for any f in $\ell^1(]0,n])$, $D_j(\Pi_j f) \leq D_j^*(|f|)$. In order to apply Lemma 2, it then is sufficient to prove that, for any mapping g from $\mathbb{Z}_+$ into

$[-1,1]$ and any positive $v \geq 2^{-j}(g|g)$,

$$\begin{aligned}(3.19)\qquad &\mathbb{P}\big(D_j^*(g) - \mathbb{E}(D_j^*(g)) \geq c_2\sqrt{v}\,(\psi^{-1}(2^j)t + \varphi(2^j)u)\big)\\ &\leq 4k\exp\Big(t^2(1+v^{-1/2}t)^{-1}\log\beta_j\Big)\\ &\quad + 4k\exp(-2u^2\log(1+j)).\end{aligned}$$

The proof of (3.19) will be omitted, since it uses exactly the same arguments as the proof of (4.6b) of Proposition 1 in Part I [Rio (1993)]. Let P be the uniform law on $]0,n]$. Clearly,

$$(3.20)\qquad \mathbb{E}\big(D_j^*(g)\big) = O\big(2^{Nd-j}\psi^{-1}(2^j)\|g\|_{1,P}\big).$$

We now complete the proof of (b) of Theorem 1. The proof of (a) of Theorem 1 will be omitted, since it uses the same arguments as in the proof of (b). We may, without loss of generality, assume, by increasing β_j a little, that $\beta_j \geq (1+j)^{-2}$. Define

$$A_j(g) = \big(4c_2\varphi(2^j)\big)^{-1}2^{j-Nd}D_j(\Pi_j g)$$

and

$$A_j^*(g) = \big(4c_2\varphi(2^j)\big)^{-1}2^{j-ND}D_j^*(g).$$

Collecting Lemma 1, (4.6b) of Proposition 1 in Part I, (3.19) and (3.20), it is easily seen that the so defined processes A_j and A_j^* satisfy the assumption (3.8) of Lemma 2 with $a = 1 \wedge 2K2^{(j-Nd)\delta/d}$ and $\theta = -c_3 2^{Nd-j}\log\beta_j$. Hence, by Lemma 2, setting $u_j = 2^{Nd-j}$, there exists some positive constant c_4, such that for any positive t,

$$\mathbb{P}\Big(\Big(D_j \geq c_4\varphi\big(u_j^{-1}n\big)\Big(\sqrt{u_j}\,I\big(u_j^{-\delta/d}\big)\Big) + u_jH_1^{-1}(u_j) + u_j^{(d-\delta)/2d}\sqrt{t} + t\Big)\Big) \leq c_4\beta_j^t.$$

Now, the end of the proof is straightforward, setting $t = t_j = Nd - j$ in the above inequalities and using the Borel–Cantelli lemma. □

4. Lower bounds for strong approximations. In this section, starting from techniques previously initiated by Beck (1985, 1987) for lower bounds in the theory of irregularities of distribution, we prove that Corollary 3 is optimal when $\mathscr{S}$ is either the class $J(\alpha, d, M)$ introduced in Dudley (1974) or the class $\mathscr{C}^2$ of convex subsets of $[0,1]^2$. In fact, we derive Theorem 2 from the more general result which is stated below.

THEOREM 3. *Let $2 < r \leq 4$ and let F and G be different probability laws on $\mathbb{R}$ with mean zero and finite rth-moment and let $\mathscr{S} = \{S \cap [0,1]^d\colon S \in J(\alpha, d, M)\}$. Let $(X_i)_{i\in\mathbb{Z}_+^d}$ and $(Y_i)_{i\in\mathbb{Z}_+^d}$ be two arrays of i.i.d. random variables with respective laws F and G. Then there exist some positive constants*

C(F, G) and C such that

$$\mathbb{P}\Big(\nu^{(1-d)(1+1/\alpha)/2} \sup_{S\in\mathscr{S}} |X(\nu S) - Y(\nu S)| \leq C(F,G)\Big) \leq C\nu^{(1-r/2)(\alpha+d-1)/\alpha}.$$

Before proving Theorem 3, we still give another theorem whose proof also uses the Fourier analysis techniques initiated by Beck. Let λ_0 be the Lebesgue measure on the unit cube and let $(x_i)_{i>0}$ be a sequence of i.i.d. random vectors with distribution λ_0. The following result is a partial converse of Massart's results (1989) on the strong approximation of the multivariate empirical bridge. Let $P_n = n^{-1}\Sigma_1^n \delta_{x_i}$ denote the empirical measure associated with $(x_1, \dots, x_n)$. We call the empirical bridge the centered and normalized process $Z_n = \sqrt{n}(P_n - \lambda_0)$.

THEOREM 4. *Let $\mathscr{S} = J(\alpha, d, M)$. There exists a positive universal constant $C_0(d)$ such that, for any positive integer n, for any standard Brownian bridge $Z^{(n)}$ on the unit cube indexed by $\mathscr{S}$ which is almost surely continuous on $(\mathscr{S}, d_\lambda)$,*

$$\mathbb{E}\Big(\sup_{S\in\mathscr{S}} \big|Z_n(S) - Z^{(n)}(S)\big|\Big) \geq C_0(d) n^{(d-1-\alpha)/(2d\alpha)}.$$

COMMENTS. When $\mathscr{S}$ is the class of convex subsets of $\mathbb{R}^2$, Theorems 3 and 4 still hold with $\alpha = 2$. In the special case of Lebesgue measure, Theorem 4 improves lower bounds for an arbitrary probability law obtained by Borisov (1985).

PROOF OF THEOREM 3. Throughout, $\mathbb{R}^d$ is provided with the usual sum and product. Let μ be a signed Borel measure on $[0,1]^d$, absolutely continuous with respect to λ. We will give a lower bound on the variable $\sup_{S\in\mathscr{S}}|\mu(S)|$. Let the test function ϱ: $\mathbb{R}^{d-1} \to \mathbb{R}^d$ be defined by $\varrho(x) = 0$ if and only if $|x| \geq 1$ and

$$2\sqrt{d}\,\varrho(x) = \exp\Big(-(1-|x|^2)^{-1}\Big) \quad \text{iff} \quad |x| < 1$$

and let $\mathscr{C} = \{(x_1, \dots, x_d) \in \mathbb{R}^d\colon\ |x_d| < \varrho(x_1, \dots, x_{d-1})\}$. Set $\mathscr{C}_A = \nu^{-1}(A, \dots, A, 1)\mathscr{C}$ and let χ_A denote the characteristic function of $\mathscr{C}_A$. $O(d)$ is the group of proper orthogonal transformations in $\mathbb{R}^d$ and dv is the volume element of the invariant measure normalized such that $\int_{O(d)} dv = 1$. Let

$$\Delta(\mu, A) = \int_{O(d)} dv \int_{\mathbb{R}^d} |\mu * v.\chi_A(x)|\, dx$$

where $v.\chi_A(x) = \chi_A(v^{-1}x) = \mathbb{1}_{v\mathscr{C}_A}$. As the first step, we give the following lower bound on the variable $\sup_{S\in\mathscr{S}}|\mu(S)|$.

LEMMA 3. *Let $A = \nu^{1-1/\alpha}$ and let μ be a signed measure with support included in $[0,1]^d$. Then, there exists a positive constant C_0 depending only on* *d, such that*

$$\sup_{S \in J(\alpha, d, M)} |\mu(S)| \geq C_0 (A^{-1}\nu)^{d-1} \Delta(\mu, A).$$

PROOF. Let ε be any mapping from $\mathbb{Z}^{d-1}$ into $\{-1, 1\}$ and define the C^∞ function Ψ_ε of $\mathbb{R}^{d-1}$ into $\mathbb{R}$ by

$$\Psi_\varepsilon(x) = \nu^{-1} \sum_{p \in \mathbb{Z}^{d-1}} \varepsilon(p) \varrho(\nu A^{-1}(x - 2p)).$$

Clearly $\Psi_\varepsilon \in F(\mathbb{R}^{d-1}, \alpha)$ and $\|\Psi_\varepsilon\|_\alpha \leq \|\rho\|_\alpha$. Now, we define from the functions Ψ_ε a family of subsets of $[0, 1]^d$ with α-smooth boundaries. So, we set

$$\mathscr{C}(\varepsilon, y) = \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_d - y_d < \Psi_\varepsilon(x_1 - y_1, \dots, x_{d-1} - y_{d-1})\}.$$

It is easily seen that, if M is large enough, for all $v \in O(d)$, there exists some $S(\varepsilon, y, v)$ in $J(\alpha, d, M)$ such that $S(\varepsilon, y, v) \cap [0, 1]^d = v.\mathscr{C}(\varepsilon, y) \cap [0, 1]^d$. So, recalling that the support of μ is included in $[0, 1]^d$, we get

$$\begin{aligned} \sup_{S \in \mathscr{S}} |\mu(S)| &\geq \tfrac{1}{2} \sup_{\varepsilon} \big(\mu(v.\mathscr{C}(\varepsilon, y)) - \mu(v.\mathscr{C}(-\varepsilon, y))\big) \\ &\geq \tfrac{1}{2} \sum_{p \in \mathbb{Z}^{d-1}} \big|\mu\big(v.(y + 2\nu^{-1}A(p, 0) + \mathscr{C}_A)\big)\big|. \end{aligned} \tag{4.1}$$

Let $y = (y_1, \dots, y_d)$. Then $|y_d| > 1 + d$, $y + 2\nu^{-1}A(p, 0) + \mathscr{C}_A$ does not intersect the Euclidean ball $B(0, d)$. The terms on the right-hand side then are null. Hence, integrating, we get

$$2^{d+1}(1 + d) \sup_{S \in \mathscr{S}} |\mu(S)| \geq (A^{-1}\nu)^{d-1} \iint_{O(d) \times \mathbb{R}^d} |\mu(v.(x + \mathscr{C}_A))| \, dv \, dx$$

and from here on the rest of the proof is straightforward. □

Now, let us define the quadratic functional $D(\mu, A)$ by

$$D(\mu, A) = \iint_{O(d) \times \mathbb{R}^d} |\mu * v.\chi_A(x)|^2 \, dv \, dx.$$

The signed measure μ is said to satisfy the condition $\mathscr{K}(A, C)$ if and only if the following holds:

$$\mathscr{K}(A, C) \qquad \Delta(\mu, A) \geq A^{(1-d)/2} C D(\mu, A).$$

If μ is a random measure, we set $\Gamma(C) = \{\omega \in \Omega : \mu(\omega) \text{ satisfies } \mathscr{K}(A, C)\}$. We shall see later that, when μ is a certain empirical measure associated with the difference of the two partial-sum processes X and Y,

$$1 - \mathbb{P}(\Gamma(C)) = O(\nu^{(1-r/2)(\alpha+d-1)/\alpha}). \tag{4.2}$$

Then, Theorem 3 is a straightforward consequence of the following proposition:

PROPOSITION 1. *There exists some positive constant $c(d)$ such that, for all signed measures μ,*

$$\nu^d D(\mu, A) \geq c(d) A^{d-1} \int_{[-\pi,\pi]^d} |\hat{\mu}(\nu t)|^2 \, dt.$$

PROOF. By the Parseval–Plancherel identity,

$$(2\pi)^d D(\mu, A) = \int_{\mathbb{R}^d} |\hat{\mu}(t)|^2 \, dt \int_{O(d)} |\hat{\chi}_A(vt)|^2 \, dv.$$

From this it follows that

$$(2\pi)^d D(\mu, A) \geq \int_{[-\pi,\pi]^d} |\hat{\mu}(\nu t)|^2 \, dt \inf_{t \in [-\pi,\pi]^d} \int_{O(d)} \nu^d |\hat{\chi}_A(v.\nu t)|^2 \, dv.$$

Now, let

$$g(A, |t|) = \int_{O(d)} \nu^{2d} |\hat{\chi}_A(v.\nu t)|^2 \, dv.$$

Proposition 1 follows from the next lemma.

LEMMA 4. *There exists some positive constant $c(d)$ such that, for any $\rho \leq \pi\sqrt{d}$, $g(A, \rho) \geq c(d) A^{d-1}$.*

PROOF. Let $t = (0, \dots, 0, \rho)$. From the rotational symmetry of $\mathscr{C}_A$, it follows that, if ds denotes the invariant measure on S^{d-1} normalized such that $\int_{S^{d-1}} ds = 1$,

$$g(A, \rho) = \int_{S^{d-1}} \nu^{2d} |\hat{\chi}_A(\rho\nu s)|^2 \, ds.$$

Let $e_d = (0, \dots, 0, 1)$, and let $V_\epsilon = \{x \in S^{d-1}\colon |x - e_d| < \epsilon\}$. Obviously,

$$g(A, \rho) \geq \int_{V_\epsilon} \nu^{2d} |\hat{\chi}_A(\rho\nu s)|^2 \, ds.$$

Now, we have to show that, when ϵ is small enough, there exists some positive constant c_1 such that, for any s in V_ϵ, $\nu^d \Re e \hat{\chi}_A(\rho\nu s) \geq c_1 A^{d-1}$. Clearly,

$$\nu^d \Re e \hat{\chi}_A(\rho\nu s) = \int_{\nu\mathscr{C}_A} \cos(\rho s | x) \, dx.$$

Now, it is easily seen that, for any s in V_ϵ, for any x in $\nu\mathscr{C}_A$ and any $\rho \leq \pi\sqrt{d}$, $(\rho s | x) \leq \pi/6 + \epsilon A \pi\sqrt{d}$. Hence, if one chooses ϵ such that $\epsilon A\sqrt{d} = 1/6$,

$$\nu^d \Re e \hat{\chi}_A(\rho\nu s) \geq \tfrac{1}{2}\lambda(\nu\mathscr{C}_A).$$

Recall that $A \geq 1$. So, the measure of V_ϵ is greater than $c_2 A^{1-d}$. Then, the proof of Lemma 4 is achieved by noting that $\lambda(\mathscr{C}_A) \geq c_3 A^{d-1}$ and by collecting the above inequalities. □

Define now the empirical measure μ associated with the difference of the two partial-sum processes. Let

$$\mu_X = \Bigg(\sum_{p \in]0,\nu]^d} \mathbb{1}_{\nu^{-1}]p - \mathbb{1}, p]} X_p \Bigg) \nu^d . \lambda_0 . .$$

and define μ_Y from Y in the same way. Let $\mu = \mu_X - \mu_Y$. Then

$$(2\pi\nu)^d |\hat{\mu}(\nu t)|^2 = \prod_1^d \left(\frac{2}{t_l} \sin \frac{t_l}{2} \right)^2 \left| \sum_{p \in]0,\nu]^d} (X_p - Y_p) e^{ip.t} \right|^2 .$$

Hence, for each $t \in [-\pi, \pi]^d$,

$$(2\pi\nu)^d |\hat{\mu}(\nu t)|^2 \geq (2/\pi)^d \left| \sum_{p \in]0,\nu]^d} (X_p - Y_p) e^{ip.t} \right|^2 .$$

So, by Proposition 1 and by the Parseval–Plancherel identity,

$$D(\mu, A) \geq (2/\pi)^d c(d) A^{d-1} \nu^{-d} \sum_{p \in]0,\nu]^d} (X_p - Y_p)^2 .$$

Let $\mathscr{L}(F, G)$ denote the class of $\mathbb{R}^2$-valued r.v.'s with respective marginals F and G, and define the Wasserstein distance $W(F, G)$ by

$$W^2(F, G) = \inf_{(X,Y) \in \mathscr{L}(F,G)} \mathbb{E}\big((X - Y)^2\big).$$

From a result of Bartfai [see Major (1978) for a proof], it follows that

$$W^2(F, G) = \int_0^1 (F^{-1}(u) - G^{-1}(u))^2 \, du .$$

Let F_ν (resp., G_ν) be the empirical distribution function of $(X_p)_{p \in]0,\nu]^d}$ [resp., $(Y_p)_{p \in]0,\nu]^d}$.] Using the above identity, we get

$$D(\mu, A) \geq (2/\pi)^d c(d) A^{d-1} W^2(F_\nu, G_\nu). \tag{4.3}$$

Now, let $d(F, G)$ denote the Lévy distance of the laws F and G; it is easily seen that for any laws F and G, $W^2(F, G) \geq d^3(F, G)$. Clearly,

$$d(F_\nu, G_\nu) \geq d(F, G) - d(F, F_\nu) - d(G, G_\nu).$$

Let $b = d(F, G)/4$. Starting from the Dvoretzky–Kiefer–Wolfowitz inequality [Massart (1990)], we get

$$\mathbb{P}\big(d(F, F_\nu) > b\big) \leq 2 \exp(-2nb^2).$$

Hence, there exists some positive constant c_4 depending only on F and G such that

$$\mathbb{P}\big(W^2(F_\nu, G_\nu) \geq c_4\big) \leq 4 \exp(-nc_4). \tag{4.4}$$

So, by (4.4) and (4.3),

$$\mathbb{P}\big(D(\mu, A) \geq c_5 A^{d-1}\big) \leq 4\exp(-nc_5) \tag{4.5}$$

for some positive constant c_5. Hence, the proof of Theorem 3 will be achieved if we prove that μ satisfies (4.2) for some positive constant C.

PROOF OF 4.2. For any signed measure β, define $k(\beta, v, x) = |\beta * v.\chi_A(x)|$. For convenience, we write $k^2(\beta, v, x) = (k(\beta, v, x))^2$. With the above notation, for any positive a,

$$\begin{aligned} 2a\Delta(\mu, A) &\geq D(\mu, A) - \iint_{O(d)\times\mathbb{R}^d} k^2(\mu, v, x)\,\mathbb{1}_{k(\mu, v, x) > 2a}\,dv\,dx \\ &\geq D(\mu, A) - 4\iint_{O(d)\times\mathbb{R}^d}\big(k^2(\mu_X, v, x)\,\mathbb{1}_{k(\mu_X, v, x) > a} \\ &\qquad + k^2(\mu_Y, v, x)\,\mathbb{1}_{k(\mu_Y, v, x) > a}\big)\,dv\,dx. \end{aligned}$$

So, in order to prove (4.2), it is sufficient to prove that there exists some constant C_1 such that, setting $a = C_1 A^{(d-1)/2}$,

$$\begin{aligned} &\mathbb{P}\bigg(16\iint_{O(d)\times\mathbb{R}^d} k^2(\mu_X, v, x)\,\mathbb{1}_{k(\mu_X, v, x) > a}\,dv\,dx \geq c_5 A^{d-1}\bigg) \\ &\qquad = O\big(\nu^{(1-r/2)(\alpha+d-1)/\alpha}\big). \end{aligned} \tag{4.6}$$

PROOF OF 4.6. For all $(x, v) \in \mathbb{R}^d \times O(d)$, there exists a family of positive numbers $(\alpha_p)_p$, each bounded by 1, such that $k(\mu_X, v, x) = |\Sigma_p \alpha_p X_p|$. Moreover, the cardinality of the set $\{p: \alpha_p \neq 0\}$ is smaller than $2.3^d(1 + d)A^{d-1}$. Hence, by the Marcinkiewicz–Zygmund inequality [see Meyer (1972)],

$$\mathbb{E}\big(k^r(\mu_X, v, x)\big) \leq c_6 A^{(d-1)r/2}.$$

From this and from Markov's inequality, it follows that

$$\mathbb{E}\big(k^2(\mu_X, v, x)\,\mathbb{1}_{k(\mu_X, v, x) > a}\big) \leq c_6 a^{2-r} A^{(d-1)r/2}.$$

Hence, if $a = C_1 A^{d-1}$ and if C_1 is large enough,

$$\mathbb{E}\bigg(32\iint_{O(d)\times\mathbb{R}^d} k^2(\mu_X, v, x)\,\mathbb{1}_{k(\mu_X, v, x) > a}\,dv\,dx\bigg) \leq c_5 A^{d-1}. \tag{4.7}$$

Now, let the random variables $\bar{Z}(v, x)$ and $Z(v, x)$ be defined by

$$\bar{Z}(v, x) = k^2(\mu_X, v, x)\,\mathbb{1}_{k(\mu_X, v, x) > a} \quad\text{and}\quad Z(v, x) = \bar{Z}(v, x) - \mathbb{E}(Z(v, x)).$$

We also set $\nu B = (A, \dots, A, 0) + (1 + d)(1, \dots, 1)$. Clearly, for all (v, x) in $O(d) \times \mathbb{R}^d$, the random variables $(Z(v, x + 2Bp))_{p \in \mathbb{Z}^d}$ are independent and with finite $(r/2)$th moment. Moreover, if $y \notin (\partial[0, 1]^d)^{(1+A)/\nu}$, then $Z(v, x) = 0$ almost surely. So, by the Marcinkiewicz–Zygmund inequality, there exists some positive constant c_6 such that

$$\mathbb{E}\left|\sum_{p\in\mathbb{Z}^d} Z(v, x + 2pB)\right|^{r/2} \leq c_6 \nu^d A^{(r-2)(d-1)/2}.$$

Hence, by Jensen's inequality,

$$\begin{aligned}\mathbb{E}\left|A^{1-d}\iint_{O(d)\times\mathbb{R}^d} Z(v, x)\, dv\, dx\right|^{r/2} &\leq c_7 \nu^{d(1-r/2)} A^{(d-1)(r-2)/2} \\ &\leq c_8 \nu^{(1-r/2)(1+(d-1)/\alpha)}.\end{aligned}$$

Finally, (4.6) follows from (4.7) and Markov's inequality applied to the above random variable, therefore completing the proof of Theorem 3. □

Proof of Theorem 4. Let $B_n = \sqrt{n}\, Z^{(n)}$ and $U_n = \sqrt{n}\, Z_n = \sum_1^n \delta_{x_p} - n\lambda_0$. Define the real number ν by $\nu^d = n$ and define A from ν exactly as in Lemma 3. Let $\mu = U_n - B_n$. Using the same arguments as in the proof of Lemma 3, it can easily be proved that

$$\mathbb{E}\Big(\sup_{S\in J(\alpha, d, M)} |\mu(S)|\Big) \geq C_0 (A^{-1}\nu)^{d-1} \mathbb{E}(\Delta(\mu, A)).$$

Define the functional $D_4(\mu, A)$ by

$$D_4(\mu, A) = \iint_{O(d)\times\mathbb{R}^d} |\mu * v.\chi_A(x)|^4\, dv\, dx.$$

By Hölder's inequality,

$$\mathbb{E}(\Delta(\mu, A)) \geq \mathbb{E}(D(\mu, A))^{3/2} \mathbb{E}(D_4(\mu, A))^{-1/2}.$$

Moreover, it is straightforward to prove that there exists a positive universal constant c_7 such that $\mathbb{E}(D_4(\mu, A)) \leq c_7 A^{2(d-1)}$. Hence, Theorem 4 is a consequence of the following proposition:

Proposition 2. $\mathbb{E}(D(\mu, A)) \geq C_1 A^{d-1}$, *for some positive constant* C_1.

Proof. First, we note that, necessarily, the Gaussian process $B_n(v.(\mathscr{C}_A + x))$ is almost surely uniformly continuous on $O(d)\times\mathbb{R}^d$ and with compact support. So, we may employ the theory of Fourier transforms. By the Parseval–Plancherel identity,

$$(2\pi)^d D(\mu, A) = \int_{\mathbb{R}^d} |\hat{\mu}(t)|^2\, dt \int_{O(d)} |\hat{\chi}_A(vt)|^2\, dv.$$

Hence, by Lemma 4,

$$(2\pi)^d D(\mu, A) \geq c(d) A^{d-1} \nu^{-d} \int_{[-\pi,\pi]^d} |\hat{\mu}(\nu t)|^2\, dt.$$

Now, let $\kappa(x)$ be the Fourier transform of $\mathbb{1}_{[-\pi,\pi]^d}$ and let $\kappa_\nu(x) = \kappa(\nu x)$. By the Parseval–Plancherel identity,

$$\mathbb{E}(D(\mu, A)) \geq c(d) A^{d-1} \int_{\mathbb{R}^d} \mathbb{E}\big((\mu * \kappa_\nu(x))^2\big)\, dx. \tag{4.8}$$

Let $F_{n,x}$ (resp., $G_{n,x}$) denote the distribution of $U_n * \kappa_\nu(x)$ [resp., $B_n * \kappa_\nu(x)$]. By definition of the Wasserstein distance,

$$\int_{\mathbb{R}^d} \mathbb{E}\big((\mu * \kappa_\nu(x))^2\big)\, dx \geq \int_{\mathbb{R}^d} W^2(F_{n,x}, G_{n,x})\, dx.$$

Hence, Proposition 2 follows from:

LEMMA 5. *There exists some positive constant C_2 such that*

$$\liminf_{n \to +\infty} \int_{\mathbb{R}^d} W^2(F_{n,x}, G_{n,x})\, dx \geq C_2.$$

PROOF. Since $G_{n,x}$ is a Gaussian distribution, by Fatou's lemma, it suffices to prove that, for all x in $]0,1[^d$, $F_{n,x}$ converges in distribution to some non-Gaussian law. Define

$$\overline{X}(n,x) = \sum_1^n \kappa\big(\nu(x - x_p)\big) \quad \text{and} \quad X(n,x) = \overline{X}(n,x) - \mathbb{E}\big(\overline{X}(n,x)\big).$$

By definition, $X(n,x)$ has distribution $F_{n,x}$. Since $\kappa = (\sin \pi t/\pi t)^{\otimes d}$, an elementary computation gives

$$\mathbb{E}\big(\overline{X}(n,x)\big) = \int_{\nu[x - \mathbb{1}, x]} \kappa(t)\, dt = (2\pi)^d + o(1). \tag{4.9}$$

It only remains now to study the convergence in distribution of $\overline{X}(n,x)$. Now, by Lévy's theorem, $\overline{X}(n,x)$ converges in distribution if and only if $\mathbb{E}(\exp(i\xi\overline{X}(n,x))$ converges to a continuous function. Clearly,

$$\mathbb{E}\big(\exp\big(i\xi\overline{X}(n,x)\big)\big) = \left(1 + \frac{1}{n}\int_{\nu[x - \mathbb{1}, x]} \exp(i\xi\kappa(t))\, dt\right)^n.$$

Since $\kappa \in L^2(\mathbb{R}^d)$, it follows from (4.9) that the integral $\int_{\nu[x - \mathbb{1}, x]} \exp(i\xi\kappa(t))\, dt$ is semiconvergent. Hence, for any x in $]0,1[^d$, $X(n,x)$ converges weakly to a non-Gaussian infinitely divisible distribution, therefore completing the proof of Theorem 4. □

When $\mathscr{S}$ is the class $\mathscr{C}_2$ of convex subsets of $\mathbb{R}^2$, we may apply the same techniques to provide lower bounds on the approximation. Let $\mathscr{P}_m$ be the regular m-gon with vertices $\exp(i2k\pi/m)$ and let $\mathscr{D}_m = \mathscr{P}_{2m} \setminus \mathscr{P}_m$. Let $C(\mathscr{D}_m) = \{T_1, \dots, T_m\}$ be the set of convex components of the interior of $\mathscr{D}_m$. Then, for any signed measure μ, any v in $O(2)$, and any (τ, x) in $[0,1] \times \mathbb{R}$,

$$2 \sup_{S \in \mathscr{C}_2} |\mu(S)| \geq \sum_{T \in C(\mathscr{D}_m)} |\mu(\tau v T + x)|. \tag{4.10}$$

Let us show (4.10) in the case $x = 0$, $\tau = 1$, $v = Id$. Define $\tilde{T}_i$ by $\tilde{T}_i = T_i$ if and only if $\mu(T_i) > 0$ and $\tilde{T}_i = \varnothing$ otherwise. Let $C_1 = \overline{\mathscr{P}}_m \cup \bigcup_{i=1}^{m} \tilde{T}_i$ and $C_2 = \overline{\mathscr{P}}_m \cup \bigcup_{i=1}^{m} (T_i \setminus \tilde{T}_i)$. Clearly, C_1 and C_2 are elements of $\mathscr{C}_2$ and

$$\mu(C_1) - \mu(C_2) = \sum_{i=1}^{m} |\mu(T_i)|,$$

establishing therefore (4.10). Then, choosing m of the order of $\sqrt{\nu}$, and using the same techniques of *rotation discrepancy* previously introduced by Beck (1987), one can prove the corresponding lower bounds for the approximation of partial-sum or multivariate empirical processes indexed by $\mathscr{C}_2$. □

REFERENCES

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.

BÁRTFAI, P. (1966). Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Versteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. *Studia Sci. Math. Hungar.* **1** 161–168.

BASS, R. F. (1985). Law of the iterated logarithm for partial-sum processes with finite variance. *Z. Wahrsch. Verw. Gebiete* **70** 591–608.

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.

BECK, J. (1987). Irregularities of distribution I. *Acta Math.* **159** 1–49.

BORISOV, I. S. (1985). Upper and lower estimates of the convergence rate in the invariance principle. In *Proceedings of the Fourth Vilnius Conference I* 251–259. VNU Science Press.

BREIMAN, L. (1967). On the tail behavior of sums of independent random variables. *Z. Wahrsch. Verw. Gebiete* **9** 20–25.

BRONSHTEIN, E. M. (1976). ε-Entropy of convex sets and functions. *Siberian Math. J.* **17** 393–398.

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.

EINMAHL, U. (1987). Strong invariance principle for partial sums of independent random vectors. *Ann. Probab.* **15** 1419–1440.

EINMAHL, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68.

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** 35–58.

MAJOR, P. (1976). Approximation of partial sums of i.i.d.r.v.'s when the summands have only two moments. *Z. Wahrsch. Verw. Gebiete* **35** 221–229.

MAJOR, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* **8** 487–517.

MASSART, P. (1989). Strong approximation for multivariate empirical and related processes, via K.M.T. constructions. *Ann. Probab.* **17** 266–291.

MASSART, P. (1990). About the constant in the DKW inequality. *Ann. Probab.* **18** 1269–1283.

MEYER, P. A. (1972). *Martingales and Stochastics Integrals I. Lecture Notes in Math.* **284**. Springer, Berlin.

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

PHILIPP, W. (1980). Weak and L^p-invariance principles for sums of B-valued random variables. *Ann. Probab.* **8** 68–82.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, Berlin.

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

RIO, E. (1993). Strong approximation for set-indexed partial sum processes, via KMT constructions I. *Ann. Probab.* **21** 759–790.

SAKHANENKO, A. I. (1984). Rate of convergence in the invariance principle for variables with exponential moments that are not identically distributed. In *Limit Theorems for Sums of Random Variables* (A. A. Borovkov, ed.). *Trudy Inst. Mat.* (*Novosibirsk*) **3** 4–49. Nauka Sibirsk. Otdel, Novosibirsk. (in Russian).

URA 743 CNRS
STATISTIQUE APPLIQUEE
BAT. 425 MATHEMATIQUE
91405 ORSAY CEDEX
FRANCE