The Annals of Probability
1993, Vol. 21, No. 2, 759–790

STRONG APPROXIMATION FOR SET-INDEXED PARTIAL SUM PROCESSES VIA KMT CONSTRUCTIONS I.

BY EMMANUEL RIO

CNRS, Orsay

Let $(X_i)_{i \in \mathbb{Z}_+^d}$ be an array of independent identically distributed zero-mean random vectors with values in $\mathbb{R}^k$. When $E(|X_1|^r) < +\infty$, for some $r > 2$, we obtain the strong approximation of the partial sum process $(\Sigma_{i \in \nu S} X_i\colon S \in \mathscr{S})$ by a Gaussian partial sum process $(\Sigma_{i \in \nu S} Y_i\colon S \in \mathscr{S})$, uniformly over all sets in a certain Vapnik–Chervonenkis class $\mathscr{S}$ of subsets of $[0,1]^d$.

The most striking result is that both an array $(X_i)_{i \in \mathbb{Z}_+^d}$ of i.i.d. random vectors and an array $(Y_i)_{i \in \mathbb{Z}_+^d}$ of independent $N(0, \operatorname{Var} X_1)$-distributed random vectors may be constructed in such a way that, up to a power of $\log \nu$,

$$\sup_{S \in \mathscr{S}} \left| \sum_{i \in \nu S} (X_i - Y_i) \right| = O(\nu^{(d-1)/2} \vee \nu^{d/r}) \quad \text{a.s.},$$

for any Vapnik–Chervonenkis class $\mathscr{S}$ fulfilling the uniform Minkowsky condition.

From a 1985 paper of Beck, it is straightforward to prove that such a result cannot be improved, when $\mathscr{S}$ is the class of Euclidean balls.

1. Introduction. This paper focuses on the asymptotic properties of partial sum processes indexed by sets in Euclidean spaces. These processes are determined by an array $(X_i)_{i \in \mathbb{Z}_+^d}$ of random vectors. Throughout, we assume that these vectors have values in $\mathbb{R}^k$. If $\mathscr{S}$ is any collection of subsets of $[0,1]^d$ we define the partial sum process $\{X(\nu S)\colon S \in \mathscr{S}\}$ by $X(\nu S) = \Sigma_{i \in \nu S} X_i$, where we use the convention that $\Sigma_{i \in \varnothing} X_i = 0$, the null vector of $\mathbb{R}^k$. When $d = 1$ and $\mathscr{S} = \{[0,t],\ 0 \le t \le 1\}$, that is, when $(X_i)_{i \ge 1}$ is a sequence of i.i.d. $\mathbb{R}$-valued random variables with a finite r-th moment, Komlós, Major and Tusnády [(1975, 1976), KMT] proved that a sequence $(Y_i)_{i \ge 1}$ of i.i.d. Gaussian variables may be constructed in such a way that, denoting by Y the partial sum process associated with $(Y_i)_{i \ge 1}$,

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

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

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

It is worth noticing that the rates of strong approximation appearing above are optimal. This comes from Breiman's remark [Breiman (1967)] when the r-th

Received December 1990; revised October 1991.
AMS 1991 *subject classifications*. 60F17, 62G99.
Key words and phrases. Central limit theorem, set-indexed process, partial sum process, invariance principle, Vapnik–Chervonenkis class, metric entropy, random measure.

moment is finite and from Erdös and Rényi (1970) [see Csörgő and Révész (1981)] when the moment-generating function is finite. Recently, Einmahl (1989) extended these results to $\mathbb{R}^k$-valued random vectors.

At the same time, several authors studied functional laws of the iterated logarithm and uniform central limit theorems for multidimensionally indexed partial sum processes ($d \geq 2$). The reader is referred to the above papers concerning partial sum processes and to the following: Bass and Pyke (1984) for a law of the iterated logarithm and uniform central limit theorem for independent arrays obtained via a Skorohod-type embedding; Bass (1985) for a functional law of the iterated logarithm and Alexander and Pyke (1986) for a uniform central limit theorem and partial sum processes indexed by large families of sets when only the second moment is assumed to be finite; Alexander (1987) for independent arrays indexed by Vapnik–Chervonenkis classes; Morrow and Philipp (1986) for invariance principles and rates of convergence in the independent case for i.i.d. random vectors with a finite r-th moment in the more general setting of entropy without inclusion and Banach space valued random vectors. However, these rates are not explicit because of the techniques used by Morrow and Philipp. On the other hand, Massart (1989) obtained recently the rate $\nu^{-1/2}(\log \nu)^{3/2}$ in the strong approximation for $\mathbb{R}$-valued partial sum processes indexed by Vapnik–Chervonenkis classes fulfilling the uniform Minkowski condition, via K.M.T. constructions. However, he had to assume the existence of the moment-generating function. In Section 3, we shall prove that Massart's result is, up to a power of $\log \nu$, optimal when $\mathscr{S}$ is the class of Euclidean balls. Here, our aim is to strengthen and to unify the results obtained by Massart, by Morrow and Philipp and by Bass and Pyke to almost sure invariance principles with optimal rates of convergence.

We are interested in independent arrays of $\mathbb{R}^k$-valued random vectors with a finite r-th moment ($r \geq 2$) indexed by Vapnik–Chervonenkis (VC) classes of sets. In the forthcoming paper II we shall study classes whose entropy with inclusion satisfies some integrability condition. Note that, in this case, it is necessary to consider a smoothed version of the partial sum process. We mention in advance that we obtain an almost sure invariance principle with an optimal rate of convergence for any $r \geq 2$.

Now, we discuss further the scope of results and the related literature. In Sections 3 and 4, using the extension made by Einmahl (1989) to the multidimensional case of KMT's results, we give a new multidimensionally indexed ($d \geq 2$) embedding based on Rosenblatt's multidimensional quantile transformation. The method is much closer to the method based on Skorohod-type embeddings previously used by Bass and Pyke than to the techniques used by Morrow and Philipp. However, for each S in $\mathscr{S}$, we obtain much better upper bounds on $\mathbb{P}(|X(\nu S) - Y(\nu S)| \geq t)$ than those of Bass and Pyke (1984). The order of magnitude of this bound depends mainly on the smoothness of the boundary ∂S. So we shall need an extra condition on boundaries of elements of the class $\mathscr{S}$. Let λ be the Lebesgue measure on $\mathbb{R}^d$. Given a norm $|\cdot|$ on $\mathbb{R}^d$

and a subset S of $\mathbb{R}^d$, we set

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

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

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

When $\delta = 1$, this condition is the uniform Minkowsky condition previously used by Massart (1989) and Bass and Pyke (1984). When $\mathscr{S}$ is a VC class fulfilling (1.1), our upper bounds and the combinatorial properties of the VC classes yield

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

For example, note that Morrow and Philipp obtained an almost sure error term of the order of $O(\nu^{d/2}(\log \nu)^{-\sigma_1})$.

In Section 5, starting from Beck's results [Beck (1985)], we prove that such a result cannot be improved when $\delta = 1$, $r > 2d(d-1)^{-1}$ and $\mathscr{S}$ is the class of Euclidean balls. On the other hand, according to Breiman's remark, this result is optimal when $r < 2d(d-\delta)^{-1}$. Finally, the Appendix is devoted to the proof of a Gaussian approximation lemma, based on multivariate quantile transformations.

2. Definitions and results. Throughout, the probability space Ω satisfies the following usual condition, due to Dudley and Philipp (1983). There exists an atomless random variable, defined on Ω, which is independent of the observations. For any subset B of $\mathbb{R}^d$, define

$$X(B) = \sum_{i \in B} X_i.$$

If $\mathscr{S}$ is any family of subsets of $[0,1]^d$, let $\nu\mathscr{S} = \{\nu S \cap \mathbb{Z}^d: S \in \mathscr{S}\}$, where $\nu S = \{\nu x: x \in S\}$. In order to get nice asymptotic properties for a normalized version of the partial sum process $\{X(\nu S): S \in \mathscr{S}\}$, we need to have some reasonable growth conditions on $\nu\mathscr{S}$ when $\nu \to +\infty$. So, we shall assume that $\mathscr{S}$ is a Vapnik–Chervonenkis class. We recall that this means

$$\begin{aligned} D(\mathscr{S}) &= \sup\{n \in \mathbb{N}: \#(A \cap \mathscr{S}) \\ &\qquad = 2^n \text{ for some set } A \text{ with } \#A = n\} < \infty, \end{aligned} \tag{2.1}$$

where $A \cap \mathscr{S} = \{A \cap S: S \in \mathscr{S}\}$ and $\#E$ denotes from now on the cardinality of E. We call $D(\mathscr{S})$ the density of $\mathscr{S}$. See Dudley (1978) or Assouad (1983) for many examples and properties of such classes.

When $E(|X_1|^r) < +\infty$ for some r large enough, we obtain a strong invariance principle with an error term depending only on the class $\mathscr{S}$. Let us now state the related result.

THEOREM 1. *Let $\mathcal{S}$ be a family of subsets of* $[0,1]^d$. *Assume that $\mathcal{S}$ is a VC class satisfying* (1.1) *for some* $0 < \delta \le 1$ *and* $\delta < d$. *Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, satisfying*

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

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

$$\sup_{S\in\mathcal{S}} |X(\nu S) - Y(\nu S)| = O\big(\nu^{(d-\delta)/2}(\log\nu)^{1/2}\big) \quad a.s.$$

COMMENTS. The construction of $(Y_i)_i$ does not depend on $\mathcal{S}$. Note that no smoothing 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, Theorem 1 means that, in some sense, the rate of convergence in Alexander's central limit theorem is of the order of $(\nu^{-\delta}\log\nu)^{1/2}$.

Note that, when $d = 1$, Theorem 1 still holds when $\delta < 1$. When $\delta = d = 1$ and Q has a finite moment-generating function, the results of KMT (1976) prove that the rate of approximation is of the order of $O(\log\nu)$ a.s.

When $d > 1$ and $\mathcal{S}$ is the class of Euclidean balls, we obtain the following lower bounds on the approximation. Let F and G be two different probability laws on $\mathbb{R}$ with finite variance, and let $W(F,G)$ denote the Wasserstein-type distance between F and G, which is precisely defined in Section 5 [cf. Zolotarev (1983) for probability metrics].

THEOREM 2. *Let F and G be different probability laws on $\mathbb{R}$ with a finite variance. 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 distribution functions F and G, and let $\mathcal{S}$ denote the class of intersections of closed Euclidean balls with the unit cube. Then there exists some positive constant $c(d)$ depending only on d such that the two following inequalities hold*:

$$\text{(a)} \qquad E\Big(\nu^{1-d}\sup_{S\in\mathcal{S}}(X(\nu S) - Y(\nu S))^2\Big) \ge (c(d)W(F,G))^2;$$

$$\text{(b)} \quad \liminf_{\nu\to+\infty} \nu^{(1-d)/2}\sup_{S\in\mathcal{S}}|X(\nu S) - Y(\nu S)| \ge c(d)W(F,G) \quad a.s.$$

Now, assume that the moment of Q is between 2 and $2d(d-\delta)^{-1}$. Then the rate of convergence does not depend on $\mathcal{S}$ anymore. However, in order to get a strong invariance principle when only the second moment of Q is assumed to be finite, we have to put an additional condition on $\mathcal{S}$. Define $Lx = \log(x \vee e)$ and $LLx = L(Lx)$ and let ψ be a mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$

such that the following hold:

(2.2)
(i) $\int\psi(|x|)\,dQ(x) < 1/2$, 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.

Note that, when Q has a finite second moment, such a mapping exists [see Major (1976)]. Throughout, ψ^{-1} denotes the inverse function of ψ.

THEOREM 3. *Let $\mathscr{S}$ be a family of subsets of $[0,1]^d$. Assume that $\mathscr{S}$ is a VC class satisfying* (1.1) *for some $0 < \delta \le 1$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, and let ψ be a mapping satisfying* (2.2) *for some $r < 2d(d-\delta)^{-1}$. 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*

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

Moreover, if we assume that

(2.3)
$$\mathscr{V} = \bigcup_{\nu\in\mathbb{N}} \nu\mathscr{S} \quad \text{is a VC class},$$

then, setting $\varphi(x) = \psi^{-1}(x)\sup(1, (x^{-1}LLx)^{1/2}\psi^{-1}(x))$, we have

(b)
$$\sup_{p\le\nu}\sup_{S\in\mathscr{S}} |X(pS) - Y(pS)| = o_P(\psi^{-1}(\nu^d))$$

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

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

COMMENTS. This result is a generalization of Einmahl's results [Einmahl (1987)] to multidimensionally indexed partial sum processes. The rates of approximation appearing in Theorem 3 are exactly the same as in Theorems 2 and 3 of Einmahl (1987).

Clearly, it is enough 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. [See Major (1976).]

When $\mathscr{S}$ is contraction closed, that is, $t\mathscr{S} \subset \mathscr{S}$ for any $0 < t < 1$ (this condition is fulfilled by many classes of interest), the class $\mathscr{V}$ defined in Theorem 3 is a VC class with $D(\mathscr{V}) = D(\mathscr{S})$. [To see this, note that, for any subset A of $\mathbb{R}^d$ with cardinality $D(\mathscr{S}) + 1$, there exists a positive integer p such that $A \cap \mathscr{V} = A \cap p\mathscr{S}$. Hence, $|A \cap \mathscr{V}| < 2^{D(S)+1}$.]

Theorem 3 provides a rate of the order $\nu^{-d/2}\psi^{-1}(\nu^d)$ in Alexander's central limit theorem [Alexander (1987), Corollary 4.4]. From Breiman's remark, we believe that this rate is optimal.

Before discussing further our results, we give a consequence of Theorems 1 and 3 which was mentioned in the introduction.

COROLLARY 1. *Let $\mathscr{S}$ be a family of subsets of* $[0,1]^d$. *Assume that $\mathscr{S}$ is a VC class fulfilling* (1.1) *for some δ in* $[0,1[$. *Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, satisfying*

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

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

$$\sup_{S \in \mathscr{S}} |X(\nu S) - Y(\nu S)| = O_P\big(\nu^{(d-\delta)/2}(\log \nu)^{1/2} + \nu^{d/r}\big).$$

Furthermore, if $\mathscr{S}$ satisfies (2.3), *the strong invariance principle holds with the above rate of approximation.*

COMMENT. When $r = 2d(d-\delta)^{-1}$, using our construction method, we are able to prove that the rate of approximation is of the order of $\nu^{d/r}(\log \nu)^{3/2}$ a.s.

Theorem 3(b) is a weak invariance principle in the sense of Philipp (1980) while Theorem 3(c) is a strong invariance principle where the function $x \to x^2 LLx$ plays a fundamental role. In fact, we can derive two different results from Theorem 3(c) according to the monotonicity of the function $x \to \psi(x)(x^2 LLx)^{-1}$.

COROLLARY 2. *Let $\mathscr{S}$ be a family of subsets of* $[0,1]^d$. *Assume that $\mathscr{S}$ is a VC class satisfying* (1.1) *for some* $0 < \delta \leq 1$ *and* (2.3). *Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, and let ψ be a mapping satisfying* (2.2) *for some* $r < 2d(d-\delta)^{-1}$ *and the condition* $x \to \psi(x)(x^2 LLx)^{-1}$ *is nondecreasing. 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*

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

From Breiman's remark [Breiman (1967)], it follows that this result is optimal.

On the other hand, when $x \to \psi(x)(x^2 LLx)^{-1}$ is nonincreasing, Theorem 3 yields the following Strassen-type invariance principle.

COROLLARY 3. *Let $\mathscr{S}$ be a family of subsets of* $[0,1]^d$. *Assume that $\mathscr{S}$ is a VC class satisfying* (1.1) *for some* $0 < \delta \le 1$ *and* (2.3). *Let Q be a law on* $\mathbb{R}^k$ *with mean zero and positive definite covariance, and let ψ be a mapping satisfying* (2.2) *and the condition* $x \to \psi(x)(x^2 LLx)^{-1}$ *is nonincreasing. 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*

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

COMMENTS. When $\mathscr{S}$ contains the class of lower-left orthants and when only the second moment is assumed to be finite, this result is optimal [see Major (1976)]. Moreover, when $\psi(x) = x^2(LLx)^\alpha$ for some $0 < \alpha < 1$, the power of $LL\nu$ cannot be improved [see Einmahl (1987), Theorem 4].

Now, from Corollary 3, we can derive a functional law of the iterated logarithm (LIL). More precisely, let $(X_i)_{i \in \mathbb{Z}^d}$ be an array of independent random vectors with common law Q such that $\int x \, dQ(x) = 0$ and $\operatorname{Var} Q = I_k$. Let $\mathscr{S}$ be a family of Borelian subsets of $[0,1]^d$ satisfying the assumptions of Corollary 3. For any function F from $\mathscr{S}$ into $\mathbb{R}^k$, let

$$\|F\|_{\mathscr{S}} = \sup_{S \in \mathscr{S}} |F(S)|,$$

where $|x|$ denotes the Euclidean norm of x, and let $\mathscr{K}$ be the subset of functions from $\mathscr{S}$ to $\mathbb{R}^k$ given by

$$\mathscr{K} = \left\{ F\colon \text{for some } f\colon I^d \to \mathbb{R}^k \text{ with } \int_{I^d} |f(t)|^2 \, dt \le 1, \right.$$
$$\left. F(S) = \int_S f(t) \, dt \text{ for all } S \in \mathscr{S} \right\}.$$

When $\mathscr{S}$ contains the class of lower-left orthants, f is uniquely defined. So, we shall assume that $\mathscr{S}$ contains the class of lower-left orthants. Then the approximating Gaussian process constructed in Corollary 3 satisfies the conditions of Theorem 3.1 of Bass and Pyke (1984). So, the following result holds.

LAW OF THE ITERATED LOGARITHM. $((2\nu^d LL\nu)^{-1/2} X(\nu S)\colon S \in \mathscr{S})$ *is relatively compact with respect to the metric* $\|\cdot\|_{\mathscr{S}}$, *and the set of limit points is exactly* $\mathscr{K}$ *a.s.*

COMMENTS. This result is new, as far as we know. Note that Bass (1985) has proved such a law for smoothed partial sum processes indexed by classes having an *integrable entropy with inclusion*.

Now we prove Theorems 1 and 3. The proofs of these theorems are based on the methods of a common probability space previously introduced by Komlós, Major and Tusnády. So, we first describe our method of construction of the two arrays of independent random vectors.

3. Construction of the arrays. Throughout this section, Q is a law on $\mathbb{R}^k$ with mean zero, finite variance and positive definite covariance matrix. We may without loss of generality assume that $\operatorname{Var} Q = I_k$. $(X_i)_{i \in \mathbb{Z}_+^d}$ denotes an array of independent random vectors with common law Q, and ψ is any mapping satisfying (2.2).

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 $N(0, I_k)$. Then, by means of a one-to-one mapping σ from $\mathbb{Z}_+^d$ onto $\mathbb{Z}_+$, we will turn the sequences so defined into arrays.

We need to recall the following lemma, due to Skorohod [see Dudley and Philipp (1983) for a proof of this result].

LEMMA [Skorohod (1976)]. *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, and let U be a random variable with uniform distribution over* $[0, 1]$, *defined on Ω, which is independent of V. Then there exists a measurable map Ψ from* $[0, 1] \times R_2$ *to R_1 such that $(\Psi(U, V), V)$ has law Q.*

From Skorohod's lemma it follows that it suffices to construct the sequence $(X_i)_{i \geq 1}$ from a Gaussian sample in another probability space. In order to define the r.v.'s X_i, we define partial sums of the r.v.'s X_i from the corresponding Gaussian partial sums by means of multivariate quantile transformations [see Major (1978) for more about the properties of such transformations]. But, in order to get nice asymptotic properties for these transformations, we have to put additional conditions on the law of the non-Gaussian random vectors. More precisely, we need to work with smoothed random vectors.

In order to avoid these technical difficulties, we shall use an argument of Sakhanenko (1984), which consists of iterating the same construction method, that turns a sequence of independent standard Gaussian vectors into a sequence $(Z_i)_{i > 0}$ of independent random vectors such that, for each $i \geq 1$, Z_{2i} has law $N(0, I_k)$ and Z_{2i-1} has law Q. This argument allows us to transform partial sums of smoothed random vectors. Let us now describe more precisely our method of construction.

Let $(Y_i)_{i > 0} = (Y_i^0)_{i > 0}$ be a sequence of independent standard normal random vectors. Suppose that there also exists a sequence $(\delta_l)_{l \geq 0}$ of independent random variables having uniform distribution on $[0, 1]$ and being independent of the sequence $(Y_i^0)_{i > 0}$. By means of a construction method which shall be explained later [cf. $(*)$], we define a sequence $(Z_i)_{i > 0} = (Z_i^0)_{i > 0}$ of independent random vectors from $(Y_i^0)_{i > 0}$ and δ_0 such that the following hold:

1. The random vectors $(Z_{2i}^0)_{i > 0}$ are $N(0, I_k)$-distributed.
2. The random vectors $(Z_{2i-1}^0)_{i > 0}$ have law Q.

[This means that $(Z_i^0)_{i>0}$ is a deterministic measurable function of $(Y_i^0)_{i>0}$ and δ_0, which shall be explained later.]

Now we define the sequence $(Y_i^1)_{i>0}$ of independent Gaussian r.v.'s by $Y_{2i-1}^1 = 0$ and $Y_{2i}^1 = Z_{2i}^0$ for each positive i, and we set $Z_{2i-1}^0 = X_{2i-1}$.

Clearly, the random vectors $(X_{2i-1})_{i>0}$ are independent with common law Q. It remains to define the random vectors $(X_{2i})_{i>0}$. By Skorohod's lemma, there exists a sequence $(Z_i^1)_{i>0}$ of independent $\sigma\{\delta_1, Y_i^1\colon i > 0\}$-measurable r.v.'s such that the following hold:

1. $Z_{2i-1}^1 = 0$ a.s. for any positive i.
2. $(Y_{2i}^1, Z_{2i}^1)_{i>0}$ has the same law as $(Y_i^0, Z_i^0)_{i>0}$.

So, by induction, for each integer l, there exists a sequence $(Y_i^l, Z_i^l)_{i>0}$ such that the following hold:

1. For any positive i, Z_i^l is $\sigma\{\delta_l, Y_i^l\colon i > 0\}$-measurable.
2. The sequence $(Z_{2^l i}^l, Y_{2^l i}^l)_{i>0}$ has the same law as the sequence $(Z_i^0, Y_i^0)_{i>0}$, and $(Z_i^l, Y_i^l) = (0,0)$ a.s. for any $i \notin 2^l\mathbb{N}$.
3. For any positive i, $Y_{2^{l+1}i}^{l+1} = Z_{2^{l+1}i}^l$.

Then, for each nonnegative integer l, for each odd integer i, we set $X_{2^l i} = Z_{2^l i}^l$. Clearly, the sequence so defined will be a sequence of independent random vectors with common law Q [see Einmahl (1989)]. Now, it remains to explain the method of construction of the sequences $(Z_i^0)_{i>0}$ and $(Y_i^0)_{i>0}$ in a common probability space. By Skorohod's lemma again, it suffices to construct the sequence $(Y_i^0)_{i>0}$ from $(Z_i^0)_{i>0}$. Our construction method uses the dyadic scheme previously introduced in KMT (1975). However, if one wants to use the dyadic scheme exactly as in KMT, the main technical difficulty is that one cannot perform only a truncation at the beginning of the construction, because this technique would not provide optimal rates of convergence, as illustrated by the work of Bass and Pyke (1984). So, we shall adapt the technique of adaptive truncations initiated by Bass (1985) in his paper on the functional LIL for partial sum processes to the dyadic scheme of KMT.

(*) CONSTRUCTION OF THE TWO SEQUENCES. Throughout the construction, the intervals $]l, m]$ have to be interpreted as subsets of the set of positive integers $\mathbb{Z}_+$. $\ell^2(\mathbb{Z}_+)$ 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 $]l, m]$. We want to define the finite sequence $(Y_i)_{2^L < i \le 2^{L+1}}$ as a deterministic function of $(Z_i)_{2^L < i \le 2^{L+1}}$. Here, it will be convenient to define a dyadic orthogonal basis, as Massart (1989) does.

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 $\tilde{e}_{j,p} = e_{j,p} - 2e_{j-1,2p}$. Clearly $\{e_{L,1}, \tilde{e}_{j,p}\colon\ 0 < j \le L,\ 2^{L-j} \le p < 2^{1+L-j}\}$ is an orthogonal basis of $\ell^2(]2^L, 2^{L+1}])$. So, in order to construct the sequence $(Y_i)_{2^L < i \le 2^{L+1}}$, it suffices to construct the random vectors $Y(e_{L,1})$ and $Y(\tilde{e}_{j,p})$, where $Y(f) = \sum_i f(i) Y_i$ for any function f of $\ell^2(\mathbb{Z}^+)$ with finite support. Now we set $V_{L,1} = Y(e_{L,1})$ and $\tilde{V}_{j,p} = Y(\tilde{e}_{j,p})$.

In order to construct the random vectors $V_{L,1}$ and $\tilde{V}_{j,p}$ from the sequence $(Z_i)_{2^L < i \le 2^{L+1}}$, we now introduce a nonincreasing filtration $(\mathscr{F}_{j,L})_{0<j\le L}$ of σ-fields, related to different levels of truncations at each stage of the dyadic scheme.

A DYADIC FILTRATION. Let us define the increasing sequence $(M_j)_{j\ge 0}$ by $\psi(M_j) = 2^{j+1}$ for any nonnegative integer j, let $\overline{Q}_j$ be the distribution of $X\mathbb{1}_{|X|\le M_j}$ and let Q_j be the conditional distribution of X, given $(|X| \le M_j)$, where X is a r.v. with law Q. With B_j denoting the random set of odd integers i such that $|Z_i| > M_j$, for any positive integers j and p, we set

$$U_{j,p}^0 = \sum_{i \in I_{j,p} \setminus B_j} Z_i, \tag{3.1}$$

and we define $\mathscr{F}_{j,L}$, for any $0 < j \le L$, by

$$\mathscr{F}_{j,L} = \sigma\left(B_j, |X_i|: i \in B_j, U_{j,p}^0: 2^{L-j} \le p < 2^{1+L-j}\right).$$

Clearly, $(\mathscr{F}_{j,L})_{0<j\le L}$ is a nonincreasing filtration. So, if we define the r.v.'s $V_{L,1}$ and $\tilde{V}_{j,p}$ in such a way that:

1. $V_{L,1}$ is $\mathscr{F}_{L,L}$-measurable with law $N(0, 2^L I_k)$,
2. for each j in $]1, L]$, the random vectors $\tilde{V}_{j,p}$ are $\mathscr{F}_{j-1,L}$-measurable and, given $\mathscr{F}_{j,L}$, conditionally independent with law $N(0, 2^j I_k)$,

the random vectors $V_{L,1}$ and $\tilde{V}_{j,p}$ will be independent with a normal distribution. In order to construct these Gaussian r.v.'s, we need further notation. So, we set

$$\tilde{U}_{j,p}^0 = U_{j,p}^0 - 2 \sum_{i \in I_{j-1,2p} \setminus B_j} Z_i. \tag{3.2}$$

Clearly, $\tilde{U}_{j,p}^0$ is $\mathscr{F}_{j-1,L}$-measurable. We also set

$$b_{j,p} = \#(B_j \cap I_{j,p}) \quad \text{and} \quad \tilde{b}_{j,p} = b_{j,p} - 2b_{j-1,2p}. \tag{3.3}$$

Now we want to define $\tilde{V}_{j,p}$ from $\mathscr{F}_{j,L}$ and $\tilde{U}_{j,p}^0$. Clearly, the r.v.'s $(Z_i)_{i\notin B_j}$, given $\{B_j: Z_i | i \in B_j\}$, are conditionally independent with conditional distribution Q_j when i is odd and $N(0, I_k)$ when i is even. Hence, conditionally given $\mathscr{F}_{j,L}$, the r.v.'s $\{\tilde{U}_{j,p}^0: 2^{L-j} \le p < 2^{1+L-j}\}$ are independent, and, for each p, the conditional law of $\tilde{U}_{j,p}^0$ has a smooth and strictly positive density on $\mathbb{R}^k$. Furthermore, this density depends only on $b_{j,p}$, $\tilde{b}_{j,p}$ and $U_{j,p}^0$. So, we may define $\tilde{V}_{j,p}$ as a function of $(b_{j,p}, \tilde{b}_{j,p}, U_{j,p}^0, \tilde{U}_{j,p}^0)$.

In order to define $\tilde{V}_{j,p}$, we define a random vector $\tilde{W}_{j,p}$ from $(U_{j,p}^0, \tilde{U}_{j,p}^0)$ such that, given B_j, the random vectors $U_{j,p}^0$ and $\tilde{W}_{j,p}$ are uncorrelated. So, we set

$$\tilde{W}_{j,p} = \tilde{U}_{j,p}^0 + \tilde{b}_{j,p} \operatorname{Var} Q_j \left(\operatorname{Var}\left(U_{j,p}^0 | B_j\right)\right)^{-1} U_{j,p}^0, \tag{3.4a}$$

where $\mathrm{Var}(\cdot|B_j)$ is the conditional variance, given B_j. We also set

$$\tilde{W}^0_{j,p} = \left(\mathrm{Var}\left(\tilde{W}_{j,p}|B_j\right)\right)^{-1/2}\tilde{W}_{j,p}. \tag{3.4b}$$

Definition of $\tilde{V}_{j,p}$. Given $(b_{j,p}, \tilde{b}_{j,p})$, $\tilde{V}_{j,p}$ is the multivariate conditional quantile transformation of $\tilde{W}^0_{j,p}$, for given $U^0_{j,p}$.

This transformation will be precisely defined and studied in Appendix A.
Now, for each nonnegative integer j, we set

$$\Gamma_j = \frac{1}{\sqrt{2}}\left(I_k + \mathrm{Var}\,\overline{Q}_j\right)^{1/2}. \tag{3.5}$$

Definition of $V_{L,1}$. $V_{L,1}$ is the multivariate quantile transformation of $\Gamma_L^{-1}U^0_{L,1}$.

By definition, $V_{L,1}$ is $\mathscr{F}_{L,L}$-measurable with law $N(0, 2^L I_k)$. For each $j > 1$, the random vectors $(\tilde{W}^0_{j,p})_{p>0}$ are, conditionally given $\mathscr{F}_{j,L}$, independent with a conditional smooth and strictly positive density depending only on $b_{j,p}$, $\tilde{b}_{j,p}$ and $U^0_{j,p}$. Hence the random vectors $(\tilde{V}_{j,p})_{p>0}$ are $\mathscr{F}_{j-1,L}$-measurable, and, given $\mathscr{F}_{j,L}$, conditionally independent with common law $N(0, 2^j I_k)$. It remains now to define Y_1, Y_2, and the random vectors $(\tilde{V}_{1,p})_{p>0}$. Here, we may assume that the probability space is rich enough to contain a sequence $(Y_i')_{i>0}$ of independent standard normal random vectors independent of the sequence $(Z_i)_{i>0}$, and we set $Y_1 = Y_1'$, $Y_2 = Y_2'$ and $\tilde{V}_{1,p} = Y'_{2p+2} - Y'_{2p+1}$, for any positive p. Then, the sequence defined above is a sequence of independent standard normal random vectors. Moreover, the following nice property holds.

Property 1. For each positive j, the random vectors $(\tilde{U}^0_{j,p}, \tilde{V}_{j,p})_{p>0}$ are independent and identically distributed.

We will now turn the sequences $(Y_i)_{i>0}$ and $(X_i)_{i>0}$ so constructed into arrays. $\mathbb{Z}^d$ being given the usual sum, product and order, we define a subset J of $\mathbb{N}^d$ by

$$J = \left\{(j_1, j_2, \dots, j_d) \in \mathbb{N}^d \text{ such that } j_1 \le j_2 \le \cdots \le j_d \le j_1 + 1\right\}.$$

Clearly the map from J onto $\mathbb{N}$ which maps $(j_1, \dots, j_d)$ onto $j_1 + j_2 + \cdots + j_d$ is one-to-one. For each integer j, let $(j_1, j_2, \dots, j_d)$ be the unique element of J such that $j = j_1 + j_2 + \cdots + j_d$. Let R_j be the lattice of integers multiples of $(2^{j_1}, 2^{j_2}, \dots, 2^{j_d})$. Let us define the box $C'_{j,p}$ for any p in R_j by [here $\mathbb{1} = (1, 1, \dots, 1) \in \mathbb{R}^d$]

$$C'_{j,p} = \left\{x \in \mathbb{Z}^d,\ p + \mathbb{1} \le x \le p + (2^{j_1}, \dots, 2^{j_d})\right\}. \tag{3.6}$$

Let us now state a lemma which provides a one-to-one mapping having some nice geometrical properties with respect to the dyadic boxes defined above.

LEMMA 0. *There exists a one-to-one mapping σ from $\mathbb{Z}_+^d$ onto $\mathbb{Z}_+$, mapping the boxes $C'_{N,0}$ on the intervals $]0, 2^N]$ and the boxes $(C'_{j,p})_{p \in R_j}$ on the intervals $I_{j,q} =]q2^j, (q+1)2^j]$.*

PROOF. It will be more convenient to define σ^{-1}. Let x be any positive integer. One can write $x - 1$ in radix-2:

$$x - 1 = a_0 + 2.a_1 + \cdots + 2^k a_k,$$

where $k = \lfloor \log(x-1)/\log 2 \rfloor$, floor brackets designating the integral part. Let us now define $\sigma^{-1}(x) = (x_1, \dots, x_d)$ by

$$x_i - 1 = \sum_{l \geq 0} a_{d(l+1)-i} 2^l, \tag{3.7}$$

for any i in $[1, d]$, where we use the convention that $a_l = 0$ if $l > k$.

Clearly, σ^{-1} is a one-to-one map. We now prove that, for any nonnegative N, σ^{-1} maps $]0, 2^N]$ on the box $C'_{0,N}$. Let x be any element of $]0, 2^N]$. One can write $x - 1$ in radix-2 with N digits at most (i.e., $k < N$). Let $(N_1, \dots, N_d)$ be the element of J such that $N_1 + \cdots + N_d = N$. By definition of J, $j \geq N_i$ iff $jd - i \geq N$. Hence, for any i in $[1, d]$, $x_i - 1$ can be written in radix-2 with N_i digits at most, that is, $x_i - 1 < 2^{N_i}$. It follows that $\sigma^{-1}(x)$ belongs to $C'_{0,N}$. Since σ^{-1} is one-to-one, σ^{-1} maps $]0, 2^N]$ onto $C'_{0,N}$. Therefore, σ^{-1} is a one-to-one mapping from $\mathbb{Z}_+$ onto $\mathbb{Z}_+^d$.

Let j be any positive integer. For any $q \geq 0$, x belongs to $I_{j,q}$ iff $x = q.2^j + y$ for some y in $]0, 2^j]$. Let $(y_1, \dots, y_d) = \sigma^{-1}(y)$. Then, using the definition of σ, it is easy to see that $\sigma^{-1}(x) = \sigma^{-1}(1 + q.2^j) + \sigma^{-1}(y)$. Clearly, $\sigma^{-1}(1 + q.2^j)$ belongs to the lattice R_j, and $\sigma^{-1}(y)$ is an element of the box $C'_{0,j}$ defined in (3.6). Hence, $\sigma^{-1}(I_{j,q}) \subset C'_{j,p}$ for some p in R_j. Since σ^{-1} is one-to-one, the equality holds and Lemma 0 follows. □

Then, we set $Y_i = Y_{\sigma(i)}$ and $X_i = X_{\sigma(i)}$ for any $i \in \mathbb{Z}_+^d$. Clearly, the array $(X_i)_{i \in \mathbb{Z}_+^d}$ so defined is an array of independent random vectors with common law Q.

4. Upper bounds for the construction. For any class $\mathscr{V}$ of subsets of $\mathbb{Z}_+^d$, we set $\sigma\mathscr{V} = \{\sigma(V) \colon V \in \mathscr{V}\}$. For each positive integer ν, let $\mathscr{S}_\nu = \bigcup_{p \leq \nu} p\mathscr{S}$ and $\mathscr{A}_\nu = \sigma\mathscr{S}_\nu$. By Lemma 0, $\mathscr{A}_\nu$ is a class of subsets of $]0, 2^{Nd}]$, where N is the smallest integer such that $2^N \geq \nu$. With the above definitions,

$$\sup_{p \leq \nu} \sup_{S \in \mathscr{S}} |X(pS) - Y(pS)| = \sup_{A \in \mathscr{A}_\nu} |X(A) - Y(A)|, \tag{4.1}$$

where $X(A) = \sum_{\sigma(i) \in A} X_i$. So, henceforward we work with the sequences $(X_i)_{i>0}$ and $(Y_i)_{i>0}$ defined in Section 3.

In order to control the random vector $Y(A) - X(A)$, uniformly on the class $\mathscr{A}_\nu$, it will be convenient to use the orthogonal basis previously introduced in the construction. Let $\tilde{e}_j = e_{j,1}$. We define the orthogonal systems $\mathscr{B}_0$ and $\mathscr{B}_j$ by

$$\mathscr{B}_0 = \{\tilde{e}_j : 0 \le j < Nd\} \cup \{e_{0,0}\} \quad \text{and} \quad \mathscr{B}_j = \{\tilde{e}_{j,p} : 0 < p < 2^{Nd-j}\},$$

where N is the smallest integer such that $2^N \ge \nu$. Then $\mathscr{B} = \bigcup_{j=0}^{Nd-1} \mathscr{B}_j$ is an orthogonal basis of $l^2(]0, 2^{Nd}])$. Now, let Π_j be the orthogonal projection on the space generated by $\bigcup_{l=1}^{j} \mathscr{B}_l$. If f is any mapping from $]0, 2^{Nd}]$ to $\mathbb{R}$, let $X(f) = \sum_i f(i) X_i$.

For any function bounded by 1, the control of $X(f) - Y(f)$ depends mainly on the inner products $(\Pi_j f | \Pi_j f)$. From now on, for convenience, we shall confuse the class $\mathscr{A}_\nu$ with the class of indicator functions of the elements of $\mathscr{A}_\nu$. Then the uniform control on $\mathscr{A}_\nu$ of the above inner products is ensured via the geometrical assumption (1.1) on the boundaries of elements of $\mathscr{S}$ and the perimetric properties of the mapping σ.

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

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

PROOF. First, we note that, for any function f taking its values in $[0, 1]$,

$$\Pi_j f - f = -\sum_{l=1}^{j} 2^{-l} (e_l | f) e_l - \sum_{p>0} 2^{-j} (e_{j,p} | f) e_{j,p}. \tag{4.2}$$

Hence, $f - \Pi_j f$ takes its values in $[-1, 0]$ and the first assertion of Lemma 1 holds true.

Let $[i - \mathbb{1}, i]$ denote the unit cube of $\mathbb{R}^d$ with upper-right vertex i. For each integer p, we define the closed subset $C_{j,p}$ of $\mathbb{R}^d$ from $I_{j,p} =]p2^j, (p+1)2^j]$ by

$$C_{j,p} = \bigcup_{\sigma(i) \in I_{j,p}} [i - \mathbb{1}, i].$$

For each element f of $\mathscr{A}_\nu$, there exists an integer m smaller than ν and an element S of the family $\mathscr{S}$ such that $f = \mathbb{1}_{\sigma(mS)}$. If the boundary of mS does not meet the box $C_{j,p}$, then f is a constant function on the interval $I_{j,p}$ and, using (4.1), we get $\Pi_j f(i) = 0$ for any i in $I_{j,p}$. Now, we may assume that $\mathbb{R}^d$ is provided with the norm of the supremum. Recall that $\sigma^{-1}(I_{j,p})$ is exactly a dyadic box $C'_{j,q}$ [cf. (3.6)] for some q in R_j. So, if the boundary of $m\mathscr{S}$ meets the box $C_{j,p}$, $C_{j,p}$ is included in the Borel set $(m\,\partial S)^a$, where $a = 2^{jd}$. Recall that the interiors of the boxes $C_{j,p}$ are disjoint. Hence the cardinality of the set of integers p such that $\Pi_j(f)(i) \ne 0$ for some i in $I_{j,p}$ is no more than $2^{-j}\lambda((m\,\partial S)^a)$. We complete the proof by combining this inequality and (1.1). □

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

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

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

$$\sup_{f \in \mathscr{A}_\nu} |X(f) - Y(f)| \le \sum_{i=1}^{n} |X_i - \tilde{X}_i| + \sup_{f \in \mathscr{A}_\nu} |\tilde{X}(f) - Y(f)|. \tag{4.3}$$

So, it will be enough to control each of the terms on the right-hand side. First, the control of the sequence $\sum_{i=1}^n |X_i - \tilde{X}_i|$ is ensured via the following lemma. The proof will be carried out in Appendix B, being straightforward and using only the moment assumptions (2.2)(i) and (2.2)(ii).

LEMMA 2. $\sum_{i=1}^n |X_i - \tilde{X}_i| = o(\psi^{-1}(n))$ *a.s.*

In order to obtain an exponential bound on the random vector $Y(f) - \tilde{X}(f)$, we will use the dyadic decomposition previously introduced in KMT (1975). If f is any function of $\ell^2(]0, 2^{Nd}])$, we set $\gamma_j(f) = 2^{-j}(f|\tilde{e}_j)$ and $\gamma_{j,p}(f) = 2^{-j}(f|\tilde{e}_{j,p})$. Then the orthogonal expansion of the function f with respect to $\mathscr{B}$ has the following form:

$$f = f(1)e_{0,0} + \sum_{0 \le j < Nd} \gamma_j(f)\tilde{e}_j + \sum_{\substack{0<j<Nd \\ 0<p<2^{Nd-j}}} \gamma_{j,p}(f)\tilde{e}_{j,p}.$$

We now introduce further notation. Let the random sequences $(\bar{\xi}_i^j)_{i>0}$ and $(\xi_i^j)_{i>0}$ be defined by $\bar{\xi}_i^j = \mathbb{1}_{(M_j < |\bar{X}_i| \le M_{j+1})} \bar{X}_i$ and $\xi_i^j = \bar{\xi}_i^j - \mathbb{E}(\bar{\xi}_i^j)$. Let

$$\tilde{U}_{j,p} = \tilde{X}(\tilde{e}_{j,p}) - \sum_{l \ge j} \xi^l(\tilde{e}_{j,p}), \qquad \tilde{U}_j = \tilde{X}(\tilde{e}_j), \qquad \tilde{V}_j = Y(\tilde{e}_j)$$

and

$$\tilde{V}_{j,p} = Y(\tilde{e}_{j,p}).$$

Let

$$D_j(f) = \sum_{0<p<2^{Nd-j}} \gamma_{j,p}(f)\left(\tilde{V}_{j,p} - \tilde{U}_{j,p}\right) - \xi^j(f),$$

for any positive j, and

$$D_0(f) = \sum_{j=0}^{Nd-1} \gamma_j(f)\left(\tilde{V}_j - \tilde{U}_j\right) + f(1)\left(Y_1 - \tilde{X}_1\right).$$

Noting then that $\tilde{X}(\tilde{e}_{j,p}) = \tilde{U}_{j,p} + \sum_{l \geq j} \xi^l(\tilde{e}_{j,p})$, we get

$$\begin{aligned} Y(f) - \tilde{X}(f) &= D_0(f) + \sum_{j=1}^{Nd-1} \sum_{0<p<2^{Nd-j}} \gamma_{j,p}(f)\big(\tilde{V}_{j,p} - \tilde{U}_{j,p}\big) \\ &\quad - \sum_{l=1}^{Nd-1} \sum_{j \leq l} \gamma_{j,p}(f)\xi^l(\tilde{e}_{j,p}) \qquad (4.4) \\ &= D_0(f) + \sum_{j=1}^{Nd-1} D_j(\Pi_j f). \end{aligned}$$

Now, for any function f with values in $[-1, 1]$, $|\gamma_j(f)| \leq 1$. Hence, setting $D_0 = |Y_1 - \tilde{X}_1| + \sum_{j=0}^{Nd-1} |\tilde{U}_j - \tilde{V}_j|$, we have $|D_0(f)| \leq D_0$. Now, let $D_j = \sup_{A \in \mathscr{A}_\nu} |D_j(\Pi_j f)|$. Clearly,

$$\sup_{f \in \mathscr{A}_\nu} \left| Y(f) - \tilde{X}(f) \right| \leq \sum_{j=0}^{Nd-1} D_j. \qquad (4.5)$$

The control of D_0 and D_j is based on the following normal approximation lemma, which is a straightforward consequence of Einmahl's results on multivariate transformations of smoothed partial sums of smoothed random vectors (the proof is carried out in Appendix A). Let us define the random vectors $\tilde{T}^0_{j,p}$ and T^0_j [the upper index 0 is related to the construction—see $(*)$—and Γ_j is defined by (3.5)] by

$$\tilde{T}^0_{j,p} = \Gamma_j \tilde{V}_{j,p} - \tilde{U}^0_{j,p} \quad \text{and} \quad T^0_j = \Gamma_j \tilde{V}_j - U^0_{j,1} + \mathbb{E}\big(U^0_{j,1}\big).$$

Then the following control on the above random vectors is available.

Lemma 3. *There exists a positive constant c_1 and a summable sequence $(\alpha_j)_{j \geq 0}$ of positive numbers each bounded by $1/2$ such that the following two inequalities hold*:

$$\text{(a)} \qquad \mathbb{E}\Big(\exp\Big(c_1\big(|\log \alpha_j| \psi^{-1}(2^j)\big)^{-1} |T^0_j|\Big)\Big) \leq 3;$$

$$\text{(b)} \qquad \mathbb{E}\Big(\exp\Big(c_1\big(|\log \alpha_j| \psi^{-1}(2^j)\big)^{-1} |\tilde{T}^0_{j,p}|\Big)\Big) \leq 3.$$

Now, Lemma 3 and Property 1 of the construction allow us to prove exponential bounds on the r.v.'s $D_j(f)$.

Proposition 1. *Let $n = 2^{Nd}$. There exist a constant c_2 and a summable sequence $(\beta_j)_{j \geq 1}$ of positive numbers each bounded by $1/2$ such that, for all positive t and u,*

$$\text{(4.6a)} \quad \mathbb{P}\big(D_0 \geq c_2(\psi^{-1}(n)t + \varphi(n)u)\big) \leq 4k\big((\beta_{Nd})^t + \exp(-2u^2 LLn)\big)$$

and, for any mapping g from $\mathbb{Z}_+$ into $[-1,1]$, for any positive $v \geq 2^{-j}(g|g)$,

$$\mathbb{P}\Big(\big|D_j(g)\big| \geq c_2\sqrt{v}\left(\psi^{-1}(2^j)t + \varphi(2^j)u\right)\Big)$$

$$\leq 4k\exp\Big(t^2\big(1+v^{-1/2}t\big)^{-1}\log\beta_j\Big) \tag{4.6b}$$

$$+ 4k\exp\big(-2u^2\log(1+j)\big).$$

REMARK. An immediate consequence of (4.6a) is

$$D_0 = O(\varphi(n)) \text{ a.s. and } D_0 = O_P(\psi^{-1}(n)). \tag{4.7}$$

PROOF OF PROPOSITION 1. We prove only (4.6b). The proof of (4.6a) uses the same arguments and will be omitted. In order to prove Proposition 1, we need the following large deviation lemma. Let a and r be positive reals and let $\overline{H}(a,r)$ be the class of real-valued random variables Z such that $\mathbb{E}(\exp(tZ)) \leq r$ for all $|t| \leq a$. We denote by $H(a,r)$ the class of random variables $Z - \mathbb{E}(Z)$, with Z in $\overline{H}(a,r)$.

LEMMA [Massart (1989)]. *Let a be a positive real, let $(r_i)_{i\in I}$ be a finite family of positive reals and let $(T_i)_{i\in I}$ be a family of independent random elements of $H(a,r_i)$. If $(w_i)_{i\in I}$ is any family of real numbers each bounded by 1, setting $T(w) = \sum_{i\in I} w_i T_i$, we have, for all positive v such that $\sum_{i\in I} w_i^2(r_i^2-1) \leq v$, for all t in $]0,a[$,*

$$\log\big(\mathbb{E}(\exp(tT(w)))\big) \leq vt^2a^{-2}.$$

Hence the usual Cramér–Chernoff calculation yields, for any positive u,

$$\mathbb{P}(|T(w)| \geq u/a) \leq 2\exp\big(-u^2(4v+u)^{-1}\big).$$

REMARK. This lemma is slightly more general than Massart's. However, the proof of this result is exactly the same as in Massart's paper.

Now, let us introduce the following notation. For any l in $[0, j-1]$, define the random sequence $\overline{X}^l$ by $\overline{X}^l = \mathbb{1}_{2^l(2\mathbb{N}+1)}\overline{X}$ and the random sequence $\tilde{X}^l$ from the already defined sequence $\overline{X}^l$ by $\tilde{X}_i^l = \overline{X}_i^l - E(\overline{X}_i^l)$ for any positive integer i. We also set $\tilde{V}_{j,p}^l = Y^l(\tilde{e}_{j,p})$, where Y^l is the sequence of Gaussian random vectors already defined in Section 3, and

$$\tilde{U}_{j,p}^l = Y^{l+1}\big(\tilde{e}_{j,p}\big) + \tilde{X}^l\big(\tilde{e}_{j,p}\mathbb{1}_{|X|\leq M_{j-l}}\big),$$

where $(|X| \leq M_{j-l})$ denotes the set of integers i such that $|X_i| \leq M_{j-l}$. Let

$$\tilde{U}_{j,p}^{j-1} = \tilde{X}\big(\tilde{e}_{j,p}\mathbb{1}_{2^{j-1}\mathbb{N}\cap(|X|\leq M_1)}\big) \quad\text{and}\quad \tilde{T}_{j,p}^l = \Gamma_{j-l}\tilde{V}_{j,l}^l - \tilde{U}_{j,p}^l,$$

and define the sequences $(\overline{\eta}_i^j)_{i>0}$ and $(\eta_i^j)_{i>0}$ by

$$\eta_i^j = \overline{\eta}_i^j - E\big(\overline{\eta}_i^j\big) \quad\text{and}\quad \overline{\eta}_i^j = \mathbb{1}_{(M_{(j-l-1)^+ +1} < |\overline{X}_i| \leq M_j)}\overline{X}_i,$$

for any integer i in $2^l(2\mathbb{N}+1)$. Let $\Delta_j = I_k - \Gamma_j$. By definition of the above random vectors,

$$\tilde{V}_{j,p} - \tilde{U}_{j,p} = \sum_{l=0}^{j-1} \left(\tilde{T}^l_{j,p} + \Delta_{j-l}\tilde{V}^l_{j,p}\right) - \eta^j(\tilde{e}_{j,p}).$$

We set

$$D^l_{j,1}(g) = \Delta_{j-l} \sum_{0<p<2^{Nd-j}} \gamma_{j,p}(g)\tilde{V}^l_{j,p}, \qquad D_{j,1}(g) = \sum_{l=0}^{j-1} D^l_{j,1}(g)$$

and

$$D^l_{j,2}(g) = \sum_{0<p<2^{Nd-j}} \gamma_{j,p}(g)\tilde{T}^l_{j,p}, \qquad D_{j,2}(g) = \sum_{l=0}^{j-1} D^l_{j,2}(g).$$

By definition of $D_j(g)$,

$$D_j(g) = D_{j,1}(g) + D_{j,2}(g) - \sum_{0<p<2^{Nd-j}} \gamma_{j,p}(g)\eta^j(\tilde{e}_{j,p}) - \xi^j(g). \tag{4.8}$$

Relation (4.6b) of Proposition 1 follows from classical Cramér–Chernoff calculations. So we have to bound the Laplace transforms of each component of the above random vectors. So, throughout the proof of Proposition 1, we may assume w.l.o.g. that $k=1$. Let h_2 be the Laplace transform of $D_{j,2}(g)$ and $h_{2,l}$ be the Laplace transform of $D^l_{j,2}(g)$. By convexity of $x \to \exp(tx)$, for any sequence $(u_l)_{l\in[0,j[}$ of positive numbers such that $u_0 + \cdots + u_{j-1} = 1$, for any real x,

$$h_2(x) \le \sup_{l\in[0,j[} h_{2,l}(x/u_l). \tag{4.9}$$

Recall that the sequences $(\tilde{T}^l_{j,p})_{p>0}$ and $(\tilde{T}^0_{j-l,p})_{p>0}$ have the same joint distribution. So, for any l in $[0,j[$, it follows from Property 3.1 and from Lemma 3 that the r.v.'s $|\tilde{T}^l_{j,p}|$ are independent random elements of $\overline{H}(c_1|\log\alpha_{j-l}|/\psi^{-1}(2^{j-l}),3)$. Furthermore, since $\mathscr{B}$ is a dyadic orthogonal basis of $\ell^2(]0,2^{Nd}])$, the real numbers $\gamma_{j,p}(g)$ are each bounded by 1, and $\sum_p \gamma^2_{j,p}(g) \le v$. So, by Massart's lemma,

$$\log h_{2,l}(x) \le c_3 v\left(\psi^{-1}(2^{j-l})/|\log\alpha_{j-l}|\right)^2 x^2,$$

for any x satisfying $|x|\psi^{-1}(2^{j-l}) \le c_1|\log\alpha_{j-l}|$. Now we set $u_l = u_0(1+l)2^{-l/r}$, where u_0 is the positive number such that $u_0 + \cdots + u_{j-1} = 1$. Let the sequence $(\beta_j)_{j>0}$ be defined by

$$\beta_j = \sup_{l\le j} \alpha^{l+1}_{j-l}.$$

Clearly, the sequence $(\beta_j)_{j>0}$ so defined fulfills the conditions of Proposition 1. It follows from (4.9), (2.2)(ii) and the above upper bound on $h_{2,l}$ that

$$\log h_2(x) \le c_4 v\left(\psi^{-1}(2^j)/|\log\beta_j|\right)^2 x^2$$

for any x fulfilling $|c_4 x|\psi^{-1}(2^j) \le |\log \beta_j|$, where c_4 is some positive constant. The usual Cramér–Chernoff calculation then yields

$$(4.10)\quad \mathbb{P}\Big(\big|D_{j,2}(g)\big| \ge c_5\sqrt{v}\,\psi^{-1}(2^j)t\Big) \le 2k\exp\Big(t^2\big(1+v^{-1/2}t\big)^{-1}\log\beta_j\Big).$$

Now we pass to the control of the r.v. $\sum_p \gamma_{j,p}(g)\eta^j(\tilde{e}_{j,p}) + \xi^j(g)$. First, for each positive p, we decompose this random vector as follows: By definition of the sequence η_j,

$$\sum_{0<p<2^{Nd-j}} \gamma_{j,p}(g)\eta^j(\tilde{e}_{j,p}) = \sum_{l=1}^{j-1}\sum_{0<p<2^{Nd-j}} \gamma_{j,p}(g)\xi^{j-l}\big(\mathbb{1}_{2^l\mathbb{N}}\tilde{e}_{j,p}\big).$$

We now prove that, for each $j>0$, the random vectors $(\bar{\xi}_i^j)_{i>0}$ are independent and such that, for some sequence $(a_j)_{j>0}$ of positive reals satisfying $\sum_{j>0} a_j \le 1/2$,

$$(4.11)\qquad \mathbb{E}\Big(\exp\Big(c_5|\log a_j|\,|\bar{\xi}_i^j|\Big/\psi^{-1}(2^j)\Big)\Big) \le 1 + 2^{-j}.$$

Proof of (4.11). Clearly, the random vectors $\bar{\xi}_i^j$ are independent. Moreover, by (2.2), there exists some sequence $(a_j)_{j\ge1}$ of positive numbers satisfying the above condition and such that $\mathbb{P}(\bar{\xi}_i^j \ne 0) \le 2^{-j}a_j$. Since $|\bar{\xi}_i^j| \le M_{j+1}$ a.s. [recall that $M_{j+1} = \psi^{-1}(2^{j+2})$], we get

$$\mathbb{E}\Big(\exp\Big(|\log a_j|\,|\bar{\xi}_i^j|/M_{j+1}\Big)\Big) \le 1 + p_j\big(a_j^{-1} - 1\big) \le 1 + 2^{-j},$$

and (4.11) follows immediately from the above inequality. Next, by (4.11) and the first part of Massart's lemma, the random variables $\xi^{j-l}(\mathbb{1}_{2^l\mathbb{N}}\tilde{e}_{j,p})$ are independent elements of the class $H(c_5|\log a_{j-l}|/\psi^{-1}(2^{j-l}), 3)$. So, using exactly the same arguments as in the proof of (4.10), we get, for any positive t,

$$\mathbb{P}\left(\left|\sum_{0<p<2^{Nd-j}} \gamma_{j,p}(g)\eta^j(\tilde{e}_{j,p})\right| \ge c_6\sqrt{v}\,\psi^{-1}(2^j)t\right) \le 2k\exp\Big(t^2\big(1+v^{-1/2}t\big)^{-1}\log\beta_j\Big)$$

for *another* sequence $(\beta_j)_{j>0}$ of positive reals fulfilling the conditions of Proposition 1. Since $\ell^2(]0,2^{Nd}])$ is equipped with the canonical inner product, it also follows from (4.11) and Massart's lemma that

$$\mathbb{P}\Big(\big|\xi^j(g)\big| \ge c_6\sqrt{v}\,\psi^{-1}(2^j)t\Big) \le 2k\exp\Big(t^2\big(1+v^{-1/2}t\big)^{-1}\log a_j\Big).$$

Both (4.10) and the two above inequalities then yield, for any positive t,

$$(4.12)\qquad \begin{aligned}&\mathbb{P}\Big(\big|D_j(g) - D_{j,1}(g)\big| \ge c_7\sqrt{v}\,\psi^{-1}(2^j)t\Big)\\ &\quad\le 4k\exp\Big(t^2\big(1+v^{-1/2}t\big)^{-1}\log\beta_j\Big),\end{aligned}$$

for some constant c_7 large enough, for *another* sequence $(\beta_j)_{j>0}$ of positive reals fulfilling the conditions of Proposition 1.

It remains only to bound the Laplace transform of $D_{j,1}(g)$. Now, for each l, $D_{j,1}^l(g)$ is a centered Gaussian random vector fulfilling

$$\operatorname{Var}\big(D_{j,1}^l(g)\big) \le 2^{-l}\|\Delta_{j-l}\|^2(g|g).$$

where $\|\cdot\|$ denotes the matrix norm associated to the usual Euclidean norm on $\mathbb{R}^k$. Then an easy calculation [see Einmahl (1987), (5.3)] ensures that

$$\|\Delta_j\| = O\Big(2^{-j}\big(\psi^{-1}(2^j)\big)^2\Big).$$

Hence, we have

$$\big(\operatorname{Var} D_{j,1}^l(g)\big)^{1/2} = O\Big(\sqrt{v}\,.2^{(j-l)/2}\big(\psi^{-1}(2^{j-l})\big)^2\Big).$$

Let h_1 be the Laplace transform of $D_{j,1}(g)$ and let $h_{1,l}$ be the Laplace transform of $D_{j,1}^l(g)$. By convexity of $x \to \exp(tx)$, for any sequence $(u_l)_{l\in[0,j[}$ of positive numbers such that $u_0 + \cdots + u_{j-1} = 1$, for any real x,

$$h_1(x) \le \sup_{l\in[0,j[} h_{1,l}(x/u_l). \tag{4.13}$$

Now, by (2.2)(iii), either

$$\big(\psi^{-1}(x)\big)^2/\sqrt{x} = O\big(\psi^{-1}(x)/\sqrt{LLx}\big) \quad \text{or} \quad x^{-s'}\big(\psi^{-1}(x)\big)^2/\sqrt{x} \quad \text{is increasing}$$

for some positive real s'. In the first situation, setting $u_l = u_0(1+l)2^{-l/r}$ and using (4.13), we get that $h_1(x) \le c.vx^2(\psi^{-1}(x))^2/LLx$, for any positive x, for some positive constant c_7. In the second situation, setting $u_l = u_0(1+l)2^{-ls'}$, we get $h_1(x) \le c.vx^2(\varphi(x))^2/LLx$. Now, using these inequalities, the usual Cramér–Chernoff calculation and inequality (4.12), we get (4.6b) of Proposition 1. □

Remark. Let $D_j'(g) = D_j(g) + \xi^j(g)$. As a by-product of the proof of Proposition 1, we have the following: There exist some constant c_8 and some sequence $(\beta_j')_{j>0}$ of positive reals satisfying $\sum_{j>0}\beta_j' \le 1/2$ such that, for any function g fulfilling $|\gamma_{j,p}(g)| \le 1$ for any positive integer p, for any real x in $[0, 1/c_8]$,

$$\mathbb{E}\Big(\exp\big(x\big|D_j'(g)\big|\,|\log\beta_j'|/\varphi(2^j)\big)\Big) \le 2k\exp\Big(c_8x^2\sum_{p>0}\gamma_{j,p}^2(g)\Big). \tag{4.14}$$

Proof of Theorem 1. By Proposition 1, with $u = t^2(1+tv^{-1/2})^{-1}$ and $\psi(x) = c.x^r$, there exists some constant c_9 such that

$$\mathbb{P}\Big(\big|D_j(g)\big| \le c_9 2^{j/r}\big(1 + u\sqrt{v} + u^2\big)\Big) \le 8k\exp(-u^2). \tag{4.15}$$

Let us now apply (4.15) with $g = \Pi_j f$. By Lemma 1, one can choose $v = 2K2^{(Nd-j)(d-\delta)/d}$. Hence, by (4.4), there exists some constant c_{10} such that, for any f in $\mathscr{A}_\nu$,

$$\mathbb{P}\Big(\sup_{f\in\mathscr{A}_\nu}|Y(f)-\tilde{X}(f)|\ge c_{10}\big(\nu^{d/r}(1+u^2)+\nu^{(d-\delta)/2}u\big)\Big) \le 8kNd\#\mathscr{A}_\nu\exp(-u^2). \tag{4.16}$$

It remains to majorize the cardinality of $\mathscr{A}_\nu$. Recall that, for any finite subset B of $[0,1]^d$ with cardinality $q\ge 2$, $\#(B\cap\mathscr{S})\le q^{D(\mathscr{S})}$, where $D(\mathscr{S})$ is the entire density defined by (2.1). [This assertion follows from the Vapnik–Chervonenkis lemma; see Assouad (1983), Section 1.9.] Hence,

$$\#\mathscr{A}_\nu\le\sum_{p=1}^{\nu}\#\big(p^{-1}\mathbb{Z}_+^d\cap\mathscr{S}\big)\le\nu^{1+dD(\mathscr{S})}.$$

Now, the end of the proof is straightforward, using (4.16) with $u^2=(dD(\mathscr{S})+3)\log\nu$ and the Borel–Cantelli lemma. □

PROOF OF THEOREM 3. We will prove only (c). The proofs of (a) and (b) use the same arguments and will be omitted. Theorem 3(c) follows clearly from (4.3), Lemma 2, (4.5), and (4.7) and from the following proposition.

PROPOSITION 2. *Almost surely in* $\{(j,N)\in\mathbb{Z}_+^2: 0<j<Nd\}$,

$$\sup_{f\in\mathscr{A}_{2^N}}|D_j(\Pi_j f)|=O\big(\varphi(2^j)2^{(Nd-j)(d-\delta)/2d}\sqrt{Nd-j}\big).$$

PROOF. Clearly, it is sufficient to prove Proposition 2 for each of the components of $D_j(\Pi_j f)$. So, throughout the proof, we may assume without loss of generality that $k=1$. Let

$$D_j'(g)=D_j(g)+\xi^j(g)=\sum_{p>0}\gamma_{j,p}(g)\big(\tilde{V}_{j,p}-\tilde{U}_{j,p}\big). \tag{4.17}$$

Clearly, $D_j(\Pi_j f)=D_j'(f)-\xi^j(\Pi_j f)$. First, we prove that

$$\sup_{f\in\mathscr{A}_{2^N}}|D_j'(f)|=O\big(\varphi(2^j)2^{(Nd-j)(d-\delta)/(2d)}\sqrt{Nd-j}\big)\quad\text{a.s. in }(j,N). \tag{4.18}$$

PROOF OF (4.18). For convenience, let $\nu=2^N$. Let $D'(A)=D'(\mathbb{1}_A)$ for any subset A of $]0,2^{Nd}]$. In order to prove (4.18), it will be necessary to use the entropy properties of VC classes of sets. So, we need to recall some well-known results on VC classes of sets.

Let P be a probability law on $\mathbb{R}_+^d$ with finite support and let $\mathscr{V}$ be a VC class of subsets of $\mathbb{R}^d$ with entire density $D(\mathscr{V})$ [cf. (2.1)]. Take on $\mathscr{V}$ the usual pseudometric d_p associated with P: For any (S,S') in $\mathscr{V}\times\mathscr{V}$, $d_P(S,S')=P(S\,\Delta\,S')$, where Δ denotes the symmetric difference, and let $N(\varepsilon,\mathscr{V},P)$ denote the minimal cardinality of a collection $\mathscr{V}(\varepsilon)$ of elements of $\mathscr{V}$ such that for any S in $\mathscr{V}$ there exists $S(\varepsilon)$ in $\mathscr{V}(\varepsilon)$ with $d_P(S,S(\varepsilon))\le\varepsilon$. Then $\log N(\varepsilon,\mathscr{V},P)$ is called a metric entropy. When $\mathscr{V}$ is a VC class, the following nice result holds.

LEMMA [Dudley (1978)]. *There exists a constant C depending only on $D(\mathscr{V})$ such that for any probability P with finite support, for any ε in* $]0, 1/2]$,

$$N(\varepsilon, \mathscr{V}, P) \le C|\varepsilon^{-1}\log\varepsilon|^{D(\mathscr{V})}.$$

Now, from the assumption (2.3) and from the definition of $\mathscr{A}_\nu$, it follows that $\mathscr{A}_\nu$ is a VC class of subsets of $\mathbb{Z}_+ \cap\,]0, 2^{Nd}]$ satisfying $D(\mathscr{A}_\nu) \le D(\mathscr{V})$, where $\mathscr{V}$ is the VC class of subsets of $\mathbb{R}^d$ defined in (2.3). Let $D = D(\mathscr{V}) + 1$. Applying Dudley's lemma to the uniform distribution on $]0, 2^{Nd}] \cap \mathbb{Z}^+$, we obtain that, for each j in $]0, Nd[$, there exists a finite net $\mathscr{A}_{j,N}$ of elements of $\mathscr{A}_\nu$ such that the following hold:

(4.19)
(i) $\#\mathscr{A}_{j,N} \le C2^{D(Nd-j)}$ for some positive constant C.
(ii) For each $A \in \mathscr{A}_\nu$, there exists A_j in $\mathscr{A}_{j,N}$ such that $\#(A \,\Delta\, A_j) \le 2^j$.

Let the neighborhood $\mathscr{U}_{j,N}$ of the diagonal of $\mathscr{A}_\nu \times \mathscr{A}_\nu$ be defined by

$$\mathscr{U}_{j,N} = \{(A, A') \in \mathscr{A}_\nu \times \mathscr{A}_\nu : \#(A \,\Delta\, A') \le 2^j\}. \tag{4.20}$$

Clearly, we have

$$\sup_{A\in\mathscr{A}_{2^N}} |D_j'(A)| \le \sup_{A\in\mathscr{A}_{j,N}} |D_j'(A)| + \sup_{(A,A')\in\mathscr{U}_{j,N}} |D_j'(A) - D_j'(A')|. \tag{4.21}$$

Let

$$v_l = 2K.2^{l(1-\delta/d)}. \tag{4.22}$$

Using Lemma 1, inequality (4.14) and (4.19)(i), we get for any $x < 1/c_8$,

$$\begin{aligned}&\mathbb{E}\left(\exp\left(x \sup_{A\in\mathscr{A}_{j,N}} |D_j'(A)\log\beta_j'|/\varphi(2^j)\right)\right)\\ &\qquad\le 2C.2^{D(Nd-j)}\exp\left(2c_8 v_{Nd-j}x^2\right).\end{aligned} \tag{4.23}$$

The usual Cramér–Chernoff calculation then yields

$$\begin{aligned}&\mathbb{P}\left(\sup_{A\in\mathscr{A}_{j,N}} |D_j'(A)| \ge c_{11}\varphi(2^j)x\sqrt{v_{Nd-j}}\right)\\ &\qquad\le 2.2^{D(Nd-j)}\exp\left(\frac{x^2\log\beta_j'}{1 + xv_{Nd-j}^{-1/2}}\right).\end{aligned} \tag{4.24}$$

Applying (4.24) with $x = c\sqrt{Nd - j}$ for some c large enough, we obtain

$$\mathbb{P}\left(\sup_{A\in\mathscr{A}_{j,N}} |D_j'(A)| \ge c_{11}\varphi(2^j)x(Nd-j)^{1/2}v_{Nd-j}^{1/2}\right) \le (\beta_j')^{Nd-j}.$$

Using the Borel–Cantelli lemma, we then get

$$(4.25)\qquad \sup_{A\in\mathscr{A}_{j,N}} |D_j'(A)| = O\big(\varphi(2^j)2^{(Nd-j)(d-\delta)/(2d)}\sqrt{Nd-j}\big) \quad \text{a.s. in } (j,N).$$

Using (4.21), the proof of (4.18) will be achieved iff we prove that

$$(4.26)\qquad \sup_{(A,A')\in\mathscr{U}_{j,N}} |D_j'(A) - D_j'(A')| = O\big(\varphi(2^j)(Nd-j)\big) \quad \text{a.s. in } (j,N).$$

PROOF OF 4.26. By definition of $\mathscr{U}_{j,N}$, for any (A, A') in $\mathscr{U}_{j,N}$, the following inequality holds:

$$\sum_{0<p<2^{Nd-j}} |\gamma_{j,p}(\mathbb{1}_A - \mathbb{1}_{A'})| \le 1.$$

It follows that

$$(4.27)\qquad \sup_{(A,A')\in\mathscr{U}_{j,N}} |D_j'(A) - D_j'(A')| \le \sup_{0<p<2^{Nd-j}} \big|\tilde{V}_{j,p} - \tilde{U}_{j,p}\big|.$$

Now, using (4.27), inequality (4.14) with $g = \tilde{e}_{j,p}$ and the same arguments as in the proof of (4.25), we obtain (4.26). Hence (4.18) holds. □

Second, we prove that

$$(4.28)\qquad \sup_{A\in\mathscr{A}_{2^N}} |\xi^j(\Pi_j\mathbb{1}_A)| = O\big(\psi^{-1}(2^j)v_{Nd-j}^{1/2}\sqrt{Nd-j}\big) \quad \text{a.s. in } (j,N).$$

PROOF OF 4.28. Using (4.19) and the definition of $\mathscr{U}_{j,N}$, it is easily seen that

$$(4.29)\qquad \begin{aligned} \sup_{A\in\mathscr{A}_{2^N}} |\xi^j(\Pi_j\mathbb{1}_A)| \le{}& \sup_{A\in\mathscr{A}_{j,N}} |\xi^j(\Pi_j\mathbb{1}_A)| \\ &+ \sup_{(A,A')\in\mathscr{U}_{j,N}} \big|\xi^j\big((\mathrm{Id}-\Pi_j)(\mathbb{1}_A - \mathbb{1}_{A'})\big)\big| \\ &+ \sup_{(A,A')\in\mathscr{U}_{j,N}} |\xi^j(A) - \xi^j(A')|. \end{aligned}$$

The control of the first term on the right-hand side uses (4.11), Massart's lemma, (4.19)(i) and the same arguments as in the proof of (4.25), so it will not be detailed.

We now control the second term on the right-hand side in (4.29). By (4.2) and by definition of the sequence ξ^j,

$$\xi^j\big((\mathrm{Id}-\Pi_j)(\mathbb{1}_A)\big) = \sum_{0<p<2^{Nd-j}} 2^{-j}(e_{j,p}|\mathbb{1}_A)\xi^j(e_{j,p}).$$

Note then that, for any (A, A') in $\mathscr{U}_{j,N}$,

$$\sum_{0<p<2^{Nd-j}} \left| e_{j,p}(\mathbb{1}_A - \mathbb{1}_{A'}) \right| \leq 1,$$

where $e_{j,p} = \mathbb{1}_{]p2^j,(p+1)2^j]}$. Hence,

$$\sup_{(A,A')\in\mathscr{U}_{j,N}} \left| \xi^j\big((\mathrm{Id} - \Pi_j)(\mathbb{1}_A - \mathbb{1}_{A'})\big) \right| \leq \sup_{0<p<2^{Nd-j}} \left| \xi^j(e_{j,p}) \right|. \tag{4.30}$$

Now, it follows from (4.11) and Massart's lemma that the r.v. $\xi^j(e_{j,p})$ belongs to the class $H(c|\log a_j|/\psi^{-1}(2^j), 3)$ for some positive constant c, where a_j is defined in the proof of (4.11). Hence, applying Markov's inequality to $\exp(c|\log a_j|\xi^j(e_{j,p})/\psi^{-1}(2^j))$, we get

$$\mathbb{P}\Big(c\left|\xi^j(e_{j,p})\right| \geq 2(Nd-j)\psi^{-1}(2^j)\Big) \leq 2.(a_j/2)^{Nd-j}. \tag{4.31}$$

Combining (4.30) and (4.31) with the Borel–Cantelli lemma, we then obtain

$$\begin{aligned} &\sup_{(A,A')\in\mathscr{U}_{j,N}} \left| \xi^j\big((\mathrm{Id} - \Pi_j)(\mathbb{1}_A - \mathbb{1}_{A'})\big) \right| \\ &\qquad = O\big(\psi^{-1}(2^j)(Nd-j)\big) \quad \text{a.s. in } (j,N). \end{aligned} \tag{4.32}$$

It remains only to give an upper bound on the third term on the right-hand side in (4.29). This is the purpose of the following lemma.

LEMMA 4. *Almost surely in (j, N),*

$$\sup_{(A,A')\in\mathscr{U}_{j,N}} \left| \xi^j(\mathbb{1}_A - \mathbb{1}_{A'}) \right| = O\Big(\psi^{-1}(2^j) v_{Nd-j}^{1/2} \sqrt{Nd-j}\Big).$$

PROOF. As the first step of the proof, note that there exists some positive constant c_{14} such that, for any (A, A') in $\mathscr{U}_{j,N}$,

$$\mathbb{P}\big(\left|\xi^j(\mathbb{1}_A - \mathbb{1}_{A'})\right| \geq c_{14}\psi^{-1}(2^j)\big) \leq 1/2.$$

[This inequality follows from (4.14), (4.20) and Massart's lemma.] Hence, by Lévy's symmetrization inequality [see Pollard (1984), page 14, for a proof], if $\tilde{\xi}^j$ is an independent copy of the sequence ξ^j, there exists some constant c_{14} such that

$$\begin{aligned} &\mathbb{P}\left(\sup_{(A,A')\in\mathscr{U}_{j,N}} \left| \xi^j(\mathbb{1}_A - \mathbb{1}_{A'}) \right| \geq t + c_{14}\psi^{-1}(2^j) \right) \\ &\qquad \leq 2\mathbb{P}\left(\sup_{(A,A')\in\mathscr{U}_{j,N}} \left| (\xi^j - \tilde{\xi}^j)(\mathbb{1}_A - \mathbb{1}_{A'}) \right| \geq t \right). \end{aligned}$$

Now, we will use the following representation for the symmetric random variables $\xi^j - \tilde{\xi}^j$. If the probability space is rich enough there exists some sequence $(B_i)_{i>0}$ of i.i.d. random variables with Bernoulli distribution $B(2^{1-j})$ and some sequence $(m_i)_{i>0}$ of symmetric i.i.d. random variables each bounded by $2M_{j+1}$ with $\mathbb{P}(m_j \neq 0) \le a_j$ [the sequence $(a_j)_{j>0}$ is defined in (4.11)] satisfying: The sequence $(m_i)_{i>0}$ is independent of the sequence $(B_i)_{i>0}$ and $\xi_i^j - \tilde{\xi}_i^j 1 = B_i m_i$. Let B denote the random set of integers i such that $B_i = 1$. We want to majorize the Laplace transform of the random variable defined in Lemma 4. First, we will control the conditional Laplace transform, given the set B. Since $\mathscr{A}_\nu$ is a Vapnik–Chervonenkis class with $D(\mathscr{A}_\nu) = D - 1$, we have

$$\#\{(B \times B) \cap \mathscr{U}_{j,N}\} \le (\#B)^{2D}. \tag{4.33}$$

Let $n(B) = \sup_{(A, A') \in \mathscr{U}_{j,N}} \#(B \cap A \,\Delta\, A')$, and let

$$D_j^* = \sup_{(A, A') \in \mathscr{U}_{j,N}} \left|(\xi^j - \tilde{\xi}^j)(\mathbb{1}_A - \mathbb{1}_{A'})\right|.$$

From (4.33) it follows that

$$\begin{aligned} \mathbb{E}\big(\exp(xD_j^*)\big) &\le \sum_{(C, C') \in (B \times B) \cap \mathscr{U}_{j,N}} \mathbb{E}(\exp(x(m(C) - m(C')))) \\ &\le 2k(\#B)^{2D} \exp\big(n(B) a_j\big(\cosh(2M_{j+1}x) - 1\big)\big). \end{aligned} \tag{4.34}$$

Let $n_j(B) = \sup_{0 < p < 2^{Nd-j}} \#(I_{j,p} \cap B)$. Using Lemma 1 and (4.20), we get

$$n(B) \le 4v_{Nd-j} n_j(B) \quad \text{and} \quad \#B \le 2^{Nd-j} n_j(B). \tag{4.35}$$

Let $t(x) = \exp(4v_{Nd-j} a_j(\cosh(2M_{j+1}x) - 1))$. From (4.34) and (4.35) it follows that

$$\mathbb{E}\big(\exp(xD_j^*)\big) \le 2k.2^{2D(Nd-j)} \mathbb{E}\Big(n_j(B)^{2D} (t(x))^{n_j(B)}\Big). \tag{4.36}$$

Now, let $n_{j,p}(B) = \#(I_{j,p} \cap B)$. Clearly, the random variables $n_{j,p}(B)$ have as common distribution the binomial distribution $B(2^j, 2^{1-j})$. Since $n_j(B)$ is the supremum of these random variables, it follows that

$$\mathbb{E}\big(\exp(xD_j^*)\big) \le 2k.2^{(1+2D)(Nd-j)} \mathbb{E}\Big(n_{j,1}(B)^{2D} (t(x))^{n_{j,1}(B)}\Big). \tag{4.37}$$

Let $t = t(x)$. For any nonnegative integer n,

$$n^{2D} t^n \le \frac{\partial^{2D} t^{n+2D}}{\partial t^{2D}}.$$

Then, using (4.37), we get

$$\begin{aligned} \mathbb{E}\big(\exp(xD_j^*)\big) &\le 2k.2^{(1+2D)(Nd-j)} \frac{\partial^{2D}}{\partial t^{2D}} \Big(t^{2D}\big(1 + 2^{1-j}(t-1)\big)^{2^j}\Big) \\ &\le k(2D)!2^{(1+2D)(2+Nd-j)} t^{2D} \exp(2t) \\ &\le k((2D)!)^2 2^{(1+2D)(2+Nd-j)} \exp(3t). \end{aligned}$$

Hence, setting $x = x_j = (2M_{j+1})^{-1} \operatorname{arg\,cosh}(1 + (Nd - j)/(4v_{Nd-j} a_j))$, we get

$$\log \mathbb{E}\big(\exp(xD_j^*)\big) \le c(D) + (4 + 2D)(Nd - j), \tag{4.38}$$

for some constant $c(D)$ depending only on D. By Markov's inequality applied to $\exp(xD_j^*)$, it follows that

(4.39) $\mathbb{P}\big(D_j^* \geq x_j^{-1}(Nd - j)(4 + 2D)\log(e/a_j)\big) \leq \exp(c(D)).a_j^{Nd-j}$.

Noting that $\log(e/a_j) = O(\arg\cosh(1 + e/a_j))$ and that, for any a in $]0, b]$,

$$(\arg\cosh(1 + a))^{-1}\arg\cosh(1 + b) \leq \sqrt{a^{-1}b}\,,$$

and using inequality (4.39) and the Borel–Cantelli lemma, we obtain Lemma 4.

Proposition 2, (4.3), (4.5), (4.7) and Lemma 2 then yield Theorem 3(c). □

5. Lower bounds for the approximation. In this section, starting from a paper of Beck (1985) on lower bounds on the approximation of the multivariate empirical process indexed by the class of Euclidean balls, we prove that our Theorem 1 is nearly optimal. So, $\mathscr{S}$ shall be the class

$$\mathrm{BALL}(d) = \{G \cap [0,1]^d\colon G \text{ is an arbitrary Euclidean closed ball of radius } r,\ r \leq 1\},$$

which was previously used by Beck (1985).

Now, we define the following Wasserstein-type distance $W(F, G)$ of the distributions F and G. Let $\mathscr{L}(F, G)$ denote the class of random vectors on $\mathbb{R}^2$ with respective marginals F and G. We set

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

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

$$W^2(F, G) = \int_0^1 (F^{-1}(u) - G^{-1}(u))^2\,du. \tag{5.1}$$

Now we prove Theorem 2. Let M_ν be defined by

$$M_\nu = \sup_{S\in\mathscr{S}}\Big|\sum_{i\in\nu S}(X_i - Y_i)\Big|.$$

Throughout this section, $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$. Define an empirical measure μ associated with the arrays $(X_i)_{i\in\mathbb{Z}_+^d}$ and $(Y_i)_{i\in\mathbb{Z}_+^d}$ by

$$\mu = \sum_{i\in]0,\nu]^d}(X_i - Y_i)\delta_{\nu^{-1}i}.$$

As Beck does, we set

$$\chi_r(y) = \mathbb{1}_{|y|\leq r}, \qquad g(r,t) = \hat{\chi}_r(t) \quad \text{and} \quad h(\rho,t) = \frac{2}{\rho}\int_{\rho/2}^{\rho}|g(r,t)|^2\,dr,$$

for any ρ in $[0,1]$ [note that $h(\rho,t)$ depends only on ρ and $|t|$]. Using the Parseval–Plancherel identity, we get

$$\frac{2}{\rho}\int_{\rho/2}^{\rho} dr\int_{\mathbb{R}^d}|(\chi_r * \mu)(x)|^2\,dx = \int_{\mathbb{R}^d} h(\rho,t)|\hat{\mu}(t)|^2\,dt.$$

According to Beck's calculus [Beck (1985), inequality 29], there exists some positive constant c depending only on d such that, for any t in $\mathbb{R}^d$, $h(1,t) \geq c^2\nu^{d-1}h(1/(2\nu), t)$. Hence, using (5.1) and the above inequality,

$$2\int_{1/2}^{1} dr \int_{\mathbb{R}^d} |(\chi_r * \mu)(x)|^2 \, dx \geq \frac{c^2}{\nu} \sum_{i \in]0,\nu]^d} (X_i - Y_i)^2.$$

From this it follows that there exists a ball $B(x,r)$ with radius $r \in [1/2, 1]$ such that

$$\left(\sum_{i \in \nu(B(x,r) \cap]0,1]^d)} (X_i - Y_i) \right)^2 \geq \frac{c^2}{\nu} \sum_{i \in]0,\nu]^d} (X_i - Y_i)^2. \tag{5.2}$$

Now, by definition of $W(F,G)$, $E((X_i - Y_i)^2) \geq W^2(F,G)$ for any i in $\mathbb{Z}_+^d$. So, the first part of Theorem 2 holds.

Let F_ν be the empirical distribution function of $(X_i)_{i \in]0,\nu]^d}$, and let G_ν be the corresponding empirical distribution function associated with Y. By (5.1),

$$M_\nu^2 \geq c^2\nu^{d-1} \int_0^1 \left(F_\nu^{-1}(u) - G_\nu^{-1}(u)\right)^2 du \quad \text{a.s.}$$

We then complete the proof of Theorem 2 by combining the Glivenko–Cantelli theorem, the above inequalities and standard arguments of measure theory. □

APPENDIX A

Multivariate quantile transformations. Let $X = (X_1, \ldots, X_s)$ be a random vector on some P-space with law Q_0. Our aim is to define a random vector $Y = (Y_1, \ldots, Y_s)$ from X such that the following hold:

1. Y is a $N(0, I_s)$-distributed random vector.
2. For each $l < s$, for given $(X_1, \ldots, X_{l-1})$, the random vector $(Y_l, \ldots, Y_s)$ is conditionally $N(0, I_{1+s-l})$-distributed.

Let F_1 denote the distribution function of X_1, and, for any l in $]1, s]$, let $F_l(\cdot | X_1, \ldots, X_{l-1})$ denote the conditional distribution function of X_l, for given $(X_1, \ldots, X_{l-1})$. Define the random vector $U = (U_1, \ldots, U_s)$ from $(X_1, \ldots, X_l)$ by $U_1 = F_1(X_1)$ and, for each $l > 1$, $U_l = F_l(X_l | X_1, \ldots, X_{l-1})$.

Clearly, U_l is a measurable function of $(X_1, \ldots, X_l)$. Moreover, if we assume that Q_0 is absolutely continuous with respect to the Lebesgue measure and has a strictly positive and continuous density, then, conditionally given $(X_1, \ldots, X_{l-1})$, the random vector $(U_l, \ldots, U_s)$ has a uniform distribution over I^{s+1-l}, for each $l > 0$. Now, let Φ be the distribution function of a standard normal, and let $Y = (Y_1, \ldots, Y_l)$ be the random vector defined from U by $Y_l = \Phi^{-1}(U_l)$ for any $l > 0$. Clearly Y satisfies the above conditions. Henceforth, we call the so-defined transformation the *multivariate quantile transformation*. The conditional transformation of $(X_l, \ldots, X_s)$ given $(X_1, \ldots, X_{l-1})$ is called the *multivariate conditional quantile transformation* of $(X_l, \dots, X_s)$ given $(X_1, \dots, X_{l-1})$. Now, we recall some recent results on multivariate quantile transformations of partial sums of smoothed random vectors.

In fact, the proof of Lemma 3 is based on Einmahl's results on the Gaussian approximation of a sum of independent vectors, via Rosenblatt's transformation. However, we need to modify slightly his main result. We consider a sequence $\xi_1, \dots, \xi_m$ of independent mean zero random vectors with values in $\mathbb{R}^s$ such that

$$\mathrm{Var}(\xi_1 + \cdots + \xi_m) = V_m I_s \quad \text{and} \quad \mathbb{E}\big(\exp|t\xi_p|\big) < +\infty,$$

for any t in $\mathbb{R}$, for any $p \le m$. Let $\alpha_0 > 0$ be the positive number such that

$$\text{(A.1)} \qquad \sum_{p=1}^{m} \alpha_0 \mathbb{E}\big(|\xi_p|^3 \exp(\alpha_0 |\xi_p|)\big) = V_m, \quad \text{and let} \quad \alpha = \alpha_0 \wedge (1/2).$$

THEOREM [Einmahl (1989)]. *Let $\xi_1, \dots, \xi_m$ be centered random vectors with values in $\mathbb{R}^s$ satisfying the above conditions and let $S_m = \xi_1 + \cdots + \xi_m$. Furthermore, assume that there exists some positive v such that the Gaussian law $N(0, vV_m I_s)$ divides the law of S_m. Let Y be the standard Gaussian r.v. obtained from S_m via the multivariate quantile transformation. Then, there exists some positive constant $C(v)$ depending only on v and s such that, if $V_m \ge C(v)\alpha^{-2}$, the following holds true*:

$$\text{(A.2)} \qquad \big|S_m - \sqrt{V_m}\, Y\big| \le C(v)\alpha^{-1}\left(\frac{|S_m|^2}{V_m} + 1\right),$$

provided that $|S_m| \le C(v)\alpha V_m$.

REMARK. This theorem is exactly Einmahl's Theorem 7 [Einmahl (1989), Section 3]. Note that Einmahl had to assume $V_m \ge c\alpha^{-2} \log \alpha^{-1}$ [cf. Einmahl (1989), page 43] but this condition comes from $\exp(-\frac{3}{8}c^2\alpha^2 V_m) \le \beta_m$: here, $\beta_m = 1/(2\alpha\sqrt{V_m})$. Hence, (A.2) still holds when $m \ge c\alpha^{-2}$. Moreover, the condition $v^{(s+1)/2} \ge Q_m(2c\alpha)$ in Einmahl's result is ensured by $Q_m(2c\alpha) \le \exp(-V_m \alpha^2 c)$.

Now, using Einmahl's theorem, we prove that $T_m = |S_m - \sqrt{V_m}\, Y|$ belongs to $H(c\alpha, 3)$, for some constant c depending only on v. Clearly, if $V_m \ge C(v)\alpha^{-2}$,

$$|T_m| \le C(v)\alpha^{-1}\left(1 + \frac{|S_m|^2}{V_m}\right)\mathbb{1}_{(|S_m| \le C(v)\alpha V_m)} + |T_m|\,\mathbb{1}_{(|S_m| > C(v)\alpha V_m)}.$$

Integrating by parts we obtain that $m^{-1}|S_m|^2 \mathbb{1}_{|S_m| \le C(v)\alpha V_m}$ belongs to $\overline{H}(c_1, 3)$ [see Massart (1989), Lemma 4, Section 3]. Now, from the definition of α, it follows that

$$\mathbb{E}\big(\exp(t|S_m)\big) \le \exp\big(2|V_m t|^2\big), \quad \text{for any } |t| \le \alpha,$$

where $(t|S_n)$ denotes the Euclidean inner product on $\mathbb{R}^k$. Hence, the classical Cramer–Chernoff calculation yields

$$\mathbb{P}(|S_m| > C(v)\alpha V_m) \le 2s\exp(-c_2 V_m \alpha^2),$$

and, by the Cauchy–Schwarz inequality, for any $t \le \alpha/2$,

$$\begin{aligned}\mathbb{E}\Big(\exp\big(t|S_m|\,\mathbb{1}_{|S_m|>C(v)\alpha V_m}\big)\Big) &\le 1 + \exp\left(-c_3 V_m \frac{\alpha^2}{2}\right)\Big(\mathbb{E}\big(\exp(2t|T_m|)\big)\Big)^{1/2}\\ &\le 1 + \exp\left(V_m\left(8t^2 - c_3\frac{\alpha^2}{2}\right)\right).\end{aligned}$$

Hence, $|S_m|\,\mathbb{1}_{|S_m|>C(v)\alpha V_m}$ belongs to $\overline{H}((\alpha/4)\sqrt{c_3}, 3)$, and using the convexity of $x \to \exp(x)$,

$$\text{(A.3)} \qquad |T_m| \in \overline{H}(c\alpha, 3) \quad \text{for some } c > 0.$$

Now, we prove Lemma 3(b). The proof of (a) uses the same arguments and will be omitted. From now on we work conditionally given B_j [B_j is defined just before (3.1)]. For given B_j, one can write $(U_{j,p}^0, \tilde{U}_{j,p}^0)$ as a sum of 2^{j-1} independent random vectors with values in $\mathbb{R}^{2k}$. We set

$$W_{j,p}^0 = \Big(\mathrm{Var}\big(U_{j,p}^0|B_j\big)\Big)^{-1/2} U_{j,p}^0 \quad \text{and} \quad S_{j,p} = 2^{j/2}\Big(W_{j,p}^0, \tilde{W}_{j,p}^0\Big).$$

Recall that $\tilde{W}_{j,p}^0$ is defined by (3.4) and $U_{j,p}^0$ is defined by (3.1). Clearly, $2^{-j/2}\tilde{V}_{j,p}$ is the multivariate conditional quantile transform of the last k components of $S_{j,p}$ given the k first components. Therefore, if α satisfies the condition (A.1) of Einmahl's theorem, we have

$$\text{(A.4)} \qquad \Big|2^{j/2}\tilde{W}_{j,p}^0 - \tilde{V}_{j,p}\Big| \in \overline{H}(c\alpha, 3).$$

Moreover, it is easily seen that for any B_j, there exists some symmetric matrix Γ depending on p and B_j, satisfying $I_{2k} \le 4\Gamma \le 16 I_{2k}$, where $\le$ denotes the usual partial order relation in the space of symmetric real matrices (i.e., $A \le B$ iff $B - A$ is a symmetric positive matrix), and such that

$$S_{j,p} = \Gamma\Big(U_{j,p}^0, \tilde{U}_{j,p}^0\Big).$$

Hence, if ζ_j is the positive number such that

$$2\zeta_j \int_{\mathbb{R}^k} |x|^3 \exp(|\zeta_j x|)\, d\overline{Q}_j(x) = 1,$$

then the random variable $|2^{j/2}\tilde{W}_{j,p}^0 - \tilde{V}_{j,p}|$ is an element of $\overline{H}(c_4\zeta_j, 3)$, for some universal constant c_4 (here the constant v of Theorem 4 satisfies $v \ge 1/4$). Now, the following lower bound on ζ_j is the main technical tool for the proof of Lemma 3.

LEMMA 5. *There exists a sequence* $(\alpha_j)_{j>0}$ *of positive numbers each bounded by* $1/2$ *and a positive constant* c_5 *such that*

$$\sum_{j>0} \alpha_j < +\infty \quad \text{and} \quad c_4\zeta_j \geq c_5\big(\psi^{-1}(2^j)\big)^{-1}|\log \alpha_j|. \tag{A.5}$$

Before proving (A.5), we conclude the proof of Lemma 3. For the sake of simplicity (throughout the sequel, p is a fixed positive integer), we write

$$\tilde{V}_{j,p} = \tilde{V} = 2^{j/2}Y, \qquad U^0_{j,p} = U, \qquad \tilde{U}^0_{j,p} = \tilde{U} \quad \text{and} \quad \tilde{W}_{j,p} = \tilde{W}.$$

Now, recall that we work conditionally given B_j: The matrices $\operatorname{Var} U$, $\operatorname{Var} \tilde{U}$, and $\operatorname{Cov}(U, \tilde{U})$ will refer to conditional variances, given B_j. For convenience, we also set $A = (\operatorname{Var} U)^{-1}\operatorname{Cov}(U, \tilde{U})$. By definition, $\tilde{W} = \tilde{U} - AU$. Hence, the following decomposition holds:

$$\tilde{U} - \Gamma_j\tilde{V} = AU + \Big((\operatorname{Var} \tilde{W})^{1/2} - 2^{j/2}\Gamma_j\Big)Y + \Big(\tilde{W} - (\operatorname{Var} \tilde{W})^{1/2}Y\Big),$$

where Γ_j is defined by (3.5). By convexity of $x \to \exp(x)$, it suffices to control each of the terms on the right-hand side. First we note that the above-defined matrices are elements of the commutative ring generated by $\operatorname{Var} Q_j$. Hence, we have

$$\operatorname{Var} \tilde{W} = \operatorname{Var} \tilde{U} - A\operatorname{Cov}(U, \tilde{U}).$$

Moreover, by definition of $(U, \tilde{U})$, $\operatorname{Var} U \geq 2^{j-1}I_k$, and $\operatorname{Cov}(U, \tilde{U}) = -\tilde{b}_{j,p}\operatorname{Var} Q_j$ [$b_{j,p}$ and $\tilde{b}_{j,p}$ are defined in (3.3)]. Hence, we have

$$\|A\| \leq 2^{2-j}b_{j,p} \leq 2, \tag{A.6}$$

from which it follows that, for any B_j, $\|\operatorname{Var} \tilde{W}\| \leq 3.2^j$. Hence, by (A.5), we get

$$\Big|\tilde{W} - (\operatorname{Var} \tilde{W})^{1/2}Y\Big| \in \overline{H}(c_4\zeta_j/3, 3). \tag{A.7}$$

On the other hand, the r.v. $T_2 = (2^{j/2}\Gamma_j - (\operatorname{Var} \tilde{W})^{1/2})Y$ is conditionally Gaussian, given B_j. Hence, to control this r.v. it will be enough to bound the norm of the above matrix. Here, a few calculations prove that

$$\Big\|2^{j/2}\Gamma_j - (\operatorname{Var} \tilde{W})^{1/2}\Big\| \leq 4\sqrt{2^{1-j}}(1 + b_{j,p}).$$

From this it follows that, for any B_j,

$$|T_2| \leq 4|Y| + Y^2 + b_{j,p}.$$

From the above inequality, it can easily be seen that there exists some universal constant c_6 such that $|T_2|$ belongs to the class $\overline{H}(c_6, 3)$. It remains to control the random vector AU. Let $V^1 = \sum_{i \in I_{j,p} \cap 2\mathbb{N}} Z^0_i$. V^1 is a Gaussian vector with law $N(0, 2^{j-1}I_k)$. Moreover, by definition of $\tilde{U}$ and by (A.6),

$$|AU| \leq (2M_j + 1)b_{j,p} + 2^{1-j}|V^1|^2.$$

Now, recall that $b_{j,p}$ has a binomial law $B(2^{j-1}, p_j)$. Hence $|AU|$ belongs to the class $\overline{H}(c_7|\log a_j|/M_j, 3)$ for some universal positive constant c_7. Then, we complete the proof of Lemma 3 by collecting the above inequalities and (A.7). □

PROOF OF LEMMA 5. First, we note that $\alpha t^3 \exp(\alpha t) \le t^2(\exp(2\alpha t) - 1)$, for any positive t. From this it follows that there exists some constant C_0 depending only on ψ and on k such that

$$\int_{\mathbb{R}^k} \alpha|x|^3 e^{|\alpha x|}\, d\overline{Q}_j(x) \le C_0\left(1 + \sum_{l=1}^{j} a_{l-1}(\exp(2\alpha M_l) - 1)\right), \tag{A.8}$$

where the reals a_{l-1} are defined by (4.11) and $M_l = \psi^{-1}(2^{l+1})$. Hence, to prove Lemma 5, it is sufficient to prove that there exist some positive constants C_1 and C_2 such that

$$\sum_{l=1}^{j} a_{l-1}\left(\alpha_j^{-C_1 M_l/M_j} - 1\right) \le C_2, \tag{A.9}$$

for some sequence $(\alpha_j)_{j>0}$ satisfying the conditions of Lemma 3. Now, recall that there exists some $\varepsilon > 0$ satisfying $M_l \le M_j 2^{(l-j)\varepsilon}$. Hence, there exists some positive constant C_3 such that

$$\sum_{l=1}^{j} \sqrt{M_l M_j^{-1}} \le C_3 \quad \text{and} \quad \sqrt{M_l M_j^{-1}}\,(1 + j - l) \le C_3.$$

So, setting $\alpha_j = \sup_{l<j} a_l^{j-l}$ and using the convexity of $x \to e^x$, we get

$$\sum_{l=1}^{j} a_{l-1}\left(\alpha_j^{M_l/(C_3 M_j)} - 1\right) \le \sum_{l=1}^{j} (1 - a_l)\sqrt{M_l M_j^{-1}} \le C_3.$$

Hence, (4.9) holds true and the proof of Lemma 5 is complete. □

APPENDIX B

PROOF OF LEMMA 2. By Kronecker's lemma, Lemma 2 follows if

$$\sum_{i=1}^{+\infty} \frac{|\tilde{X}_i - X_i|}{\psi^{-1}(i)} < +\infty \quad \text{with probability 1.}$$

So it is sufficient to prove that

$$\sum_{i=1}^{+\infty} \frac{\mathbb{E}\left(|\tilde{X}_i - X_i|\right)}{\psi^{-1}(i)} < +\infty.$$

Now, if i' is the greatest odd divisor of i, it can easily be seen that

$$\mathbb{E}\left(|X_i - \tilde{X}_i|\right) \le 2\int_{|x|>\psi^{-1}(i')} |x|\, dQ(x).$$

Hence, we have

$$\sum_{i=1}^{+\infty} \frac{\mathbb{E}(|X_i - \tilde{X}_i|)}{\psi^{-1}(i)} \le 2 \sum_{i \text{ odd}} \int_{|x| > \psi^{-1}(i)} |x|\, dQ(x) \sum_{l \ge 0} \frac{1}{\psi^{-1}(i2^l)}.$$

By (2.2)(ii), there exists some positive constant c such that $\sum_{l \ge 0} 1/\psi^{-1}(i2^l) \le c/\psi^{-1}(i)$. Moreover, it is well known that the series

$$\sum_{i=1}^{+\infty} \int_{\mathbb{R}^k} \mathbb{1}_{|x| \ge \psi^{-1}(i)} |x|\, dQ(x)$$

is convergent iff $\int_{\mathbb{R}^k} \psi(|x|)\, dQ(x)$ is finite. Hence, the proof of Lemma 2 is complete. □

REFERENCES

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** 233–282.

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.

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

CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.

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

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

EINMAHL, U. (1987). Strong invariance principles 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 KMT constructions. *Ann. Probab.* **17** 266–291.

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.

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. Trudy Inst. Mat.* 4–49. Nauka Sibirsk. Otdel, Novosibirsk. (In Russian.)

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

STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226.

ZOLOTAREV, V. M. (1983). Probability metrics. *Theory Probab. Appl.* **28** 278–302.

URA D 0 743 CNRS
STATISTIQUE APPLIQUÉE
BÂT. 425, MATHÉMATIQUE
91405 ORSAY CEDEX
FRANCE