ANNALES DE L'I. H. P., SECTION B

P. DOUKHAN
P. MASSART
E. RIO

Invariance principles for absolutely regular empirical processes

Annales de l'I. H. P., section B, tome 31, nº 2 (1995), p. 393-427
<http://www.numdam.org/item?id=AIHPB_1995__31_2_393_0>

Ann. Inst. Henri Poincaré,

Vol. 31, n° 2, 1995, p. 393-427 *Probabilités et Statistiques*

Invariance principles for absolutely regular empirical processes

by

P. DOUKHAN, P. MASSART and E. RIO

U.R.A., C.N.R.S. 743, Département de Mathématiques,
Bât. 425, Université de Paris Sud, 91405 Orsay Cedex, France.

ABSTRACT. – Let $(\xi_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of random variables with marginal distribution P. Let $Z_n = n^{-1/2}\sum_1^n(\delta_{\xi_i} - P)$ denote the centered and normalized empirical measure. Assume that the sequence $(\beta_n)_{n\geq 0}$ of β-mixing coefficients of $(\xi_i)_{i\in\mathbb{Z}}$ satisfies the summability condition $\sum_{n\geq 0}\beta_n < +\infty$. Define the mixing rate function $\beta(.)$ by $\beta(t) = \beta_{[t]}$ if $t \geq 1$, and $\beta(t) = 1$ otherwise. For any numerical function f, we denote by Q_f the quantile function of $|f(\xi_0)|$. We define a new norm for f by

$$\|f\|_{2,\beta} = \left[\int_0^1 \beta^{-1}(u)[Q_f(u)]^2\,du\right]^{1/2},$$

where β^{-1} denotes the càdlàg inverse of the monotonic function $\beta(.)$. This norm coincides with the usual $\mathcal{L}_2(P)$-norm in the independent case. We denote by $\mathcal{L}_{2,\beta}(P)$ the class of numerical functions with $\|f\|_{2,\beta} < +\infty$. Let $\mathcal{F}$ be a class of functions, $\mathcal{F} \subset \mathcal{L}_{2,\beta}(P)$. In a recent paper, the authors have shown that the finite dimensional convergence of $\{Z_n(f) : f \in \mathcal{F}\}$ to a Gaussian random vector holds.

The main result of this paper is that a functional invariance principle in the sense of Donsker holds for $\{Z_n(f) : f \in \mathcal{F}\}$ if the entropy with bracketing of $\mathcal{F}$ in the sense of Dudley with respect to $\|\,.\,\|_{2,\beta}$ satisfies some integrability condition. This result generalizes Ossiander's theorem (1987) for independent observations.

1980 Mathematics Subject Classification (1985 Revision) : 60 F 17, 62 G 99.

Key words: Central limit theorem, β-mixing processes, Donsker invariance principle, strictly stationary sequences, empirical processes, entropy with bracketing.

RÉSUMÉ. – Soit $(\xi_i)_{i\in\mathbb{Z}}$ une suite strictement stationnaire de variables aléatoires de loi marginale P. Soit $Z_n = n^{-1/2}\sum_1^n(\delta_{\xi_i} - P)$ la mesure empirique normalisée et centrée. On suppose que la suite $(\xi_i)_{i\in\mathbb{Z}}$ est β-mélangeante, de coefficients de β-mélange β_n en série sommable. On lui associe la fonction de mélange $\beta(.)$ définie par $\beta(t) = \beta_{[t]}$ si $t \geq 1$ et $\beta(t) = 1$ sinon. Pour toute fonction numérique f, on note Q_f la fonction de quantile de $|f(\xi_0)|$. Nous définissons une nouvelle norme pour f par

$$\|f\|_{2,\beta} = \left[\int_0^1 \beta^{-1}(u)[Q_f(u)]^2\,du\right]^{1/2},$$

où β^{-1} désigne l'inverse càdlàg de la fonction décroissante $\beta(.)$. Cette norme coïncide avec la norme usuelle de $\mathcal{L}_2(P)$ dans le cas indépendant. On note alors $\mathcal{L}_{2,\beta}(P)$ l'espace des fonctions numériques de norme $\|.\|_{2,\beta}$ finie. Soit $\mathcal{F}$ une classe de fonctions de $\mathcal{L}_{2,\beta}(P)$. Dans un article récent, les auteurs ont montré la convergence vers un vecteur gaussien des marginales de $\{Z_n(f) : f \in \mathcal{F}\}$ de dimension finie.

Dans cet article, nous démontrons un principe d'invariance fonctionnel au sens de Donsker pour $\{Z_n(f) : f \in \mathcal{F}\}$ sous une condition d'intégrabilité de l'entropie avec crochets de $\mathcal{F}$ par rapport à la norme $\|.\|_{2,\beta}$. Ce théorème généralise celui d'Ossiander (1987) pour des variables indépendantes.

1. INTRODUCTION

Let $(\xi_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of random elements of a Polish space $\mathcal{X}$, with common distribution P. Let us denote by P_n the empirical probability measure $P_n = \sum_{i=1}^n \delta_{\xi_i}$ and then by Z_n the centered and normalized empirical measure $Z_n = \sqrt{n}(P_n - P)$. Let $\mathcal{F}$ be a class of functions of $\mathcal{L}_2(P)$.

Independent observations

Let us assume the variables $(\xi_i)_{i\in\mathbb{Z}}$ to be independent. The theory of uniform central limit theorems (CLT) for the empirical process

$\{Z_n(f), f \in \mathcal{F}\}$ was historically motivated by the classical example $\mathcal{F} = \{\mathbb{I}_{]-\infty,t]}, t \in \mathbb{R}\}$ to justify the heuristic of Doob for explaining the convergence of the Kolmogorov-Smirnov statistics. The first result in this context is due to Donsker (1952). The first attempts at generalizing this theorem concerned classes of sets, that is when $\mathcal{F} = \{\mathbb{I}_C; C \in \mathcal{C}\}$. Dudley (1966) extended Donsker's result to the multivariate empirical distribution function, which corresponds to the case where $\mathcal{C}$ is the class of quadrants (actually, Dudley also provided a more rigorous formulation of Donsker's theorem). The problem is that in the multidimensional situation, several classes of sets are candidate to extend the class of intervals of the real line. In that spirit, several new results were established in the seventies. Bolthausen (1978) proved the uniform CLT for the class of convex subsets of the unit square, Sun (1976) and Révész (1976) studied the CLT for classes of subsets with α-differentiable boundaries.

In his landmark paper, Dudley (1978) has unified these various results: he gave a precise definition of the uniform CLT property in terms of weak convergence of the empirical process in the (generally non separable) Banach space $\mathcal{L}_\infty(\mathcal{F})$ to the Gaussian process with the same covariance structure and almost sure uniformly continuous (with respect to the covariance pseudo-metric) sample paths. The existence of such a regular version of the Gaussian process is now refered to as the P-pregaussian property and the existence of a uniform CLT as the P-Donsker property. He also provided two different sufficient conditions on $\mathcal{C}$ for the P-Donsker property to hold. The first one is of combinatorial type, namely $\mathcal{C}$ is a Vapnik-Chervonenkis class of sets. In that case, the uniform CLT is universal in the sense that it holds for all P. The second one is a condition of integrability on the metric entropy with inclusion of $\mathcal{C}$ [with respect to the $\mathcal{L}_2(P)$-metric]. Since that time the theory of CLT's for empirical processes has been developped mainly in the direction of unbounded classes of functions $\mathcal{F}$. The P-Donsker property is equivalent to the fact that $\mathcal{F}$ is totally bounded and Z_n is asymptotically equicontinuous [this tightness criterion may be found in Dudley (1984)]. So the main challenge is to provide minimal geometric conditions on the class $\mathcal{F}$ ensuring the tightness of Z_n. Giné and Zinn (1984) pointed out that $\mathcal{F}$ is a P-Donsker class if $\mathcal{F}$ is P-pregaussian and satisfies some extra condition ensuring that the empirical process Z_n may be controlled on the small balls. Two main techniques have been proposed to solve this problem. The guiding idea has been Dudley's criterion (1967) which ensures the stochastic equicontinuity of a Gaussian process $\{G(t), t \in T\}$ if the metric entropy function $H(\cdot, T, d)$ with respect to the pseudo-metric $d(s,t) = \sqrt{\text{Var}\,(G(s) - G(t))}$ satisfies the integrability condition

$$\int_0^1 \sqrt{H(u, T, d)}du < +\infty. \tag{1.1}$$

Specializing to the Gaussian process with the same covariance structure as Z_n, we have $T = \mathcal{F}$ and $d(f, g) = d_P(f, g) = \sqrt{\text{Var}_P(f - g)}$. Kolchinskii (1981) and Pollard (1982) proposed independently a sufficient condition analogous to (1.1) but involving the universal entropy function which is defined as the envelope function of the $H(\cdot, \mathcal{F}, d_Q)$ when Q is any probability measure. On the other hand, Pollard (1982) was able to relax the boundedness condition on $\mathcal{F}$, assuming only that $\sup_{f \in \mathcal{F}} |f|$ is bounded by some function F in $\mathcal{L}_2(P)$. Another approach is to consider entropy with bracketing, which is the natural extension to functions of entropy with inclusion, used by Dudley (1978) for classes of sets. It has taken quite a long time before one realizes that the $\mathcal{L}_2(P)$-norm was still the right norm to consider in order to measure the size of the brackets. In fact the first results in that direction were involving the $\mathcal{L}_1(P)$-norm [*see* for instance Dudley (1984)]. Using a delicate adaptive truncation procedure, Ossiander (1987) proved that a sufficient condition for $\mathcal{F}$ to be a P-Donsker class is that

$$\int_0^1 \sqrt{H_{[\,]}(u, \mathcal{F}, d_P)}du < +\infty,$$

where $H_{[\,]}(\cdot, \mathcal{F}, d_P)$ denotes the entropy with bracketing function of $\mathcal{F}$ with respect to d_P. Next, Andersen *et al.* (1988) have weakened the bracketing condition [controling the size of the brackets in a weak $\mathcal{L}_2(P)$-space] and also used majorizing measure instead of metric entropy [actually it is known since Talagrand (1987) that the P-pregaussian property may be characterized in terms of the existence of a majorizing measure].

We see that the $\mathcal{L}_2(P)$-norm plays a crucial role in this theory in the independent case. In fact, on the one hand the finite dimensional convergence of Z_n to a Gaussian random vector holds whenever $\mathcal{F}$ is included in $\mathcal{L}_2(P)$ and on the other hand the geometry of $\mathcal{F}$ with respect to the $\mathcal{L}_2(P)$-norm characterizes the P-pregaussian property and is involved in the description of tightness of the empirical process.

Weakly dependent observations

There are several notions of weak dependence. It is quite natural when dealing with empirical processes to consider mixing type conditions which are defined in terms of coefficients measuring the asymptotic independence between the σ-fields of the past and the future generated by $(\xi_i)_{i\in\mathbb{Z}}$. Of course different mixing coefficients have been introduced, corresponding to different ways of measuring the asymptotic independence. To our knowledge, all these coefficients are nondecreasing with respect to these σ-fields. Hence, $(f(\xi_i))_{i\in\mathbb{Z}}$ has smaller mixing coefficients than the initial sequence.

Let us recall the definitions of some classical mixing coefficients. Rosenblatt (1956) introduced the strong mixing coefficient $\alpha(\mathcal{A},\mathcal{B})$ between two σ-fields $\mathcal{A}$ and $\mathcal{B}$:

$$\alpha(\mathcal{A},\mathcal{B}) = \sup_{(A,B)\in\mathcal{A}\times\mathcal{B}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|. \tag{1.2}$$

The absolutely regular mixing coefficient $\beta(\mathcal{A},\mathcal{B})$ was defined by Volkonskii and Rozanov (1959) [*see* also Kolmogorov and Rozanov (1960)]. Namely

$$2\beta(\mathcal{A},\mathcal{B}) = \sup \sum_{(i,j)\in I\times J} |\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|, \tag{1.3}$$

where the supremum is taken over all the finite partitions $(A_i)_{i\in I}$ and $(B_j)_{j\in J}$ respectively $\mathcal{A}$ and $\mathcal{B}$ measurable. This coefficient is also called β-mixing coefficient between $\mathcal{A}$ and $\mathcal{B}$. Ibragimov (1962) introduced the φ-mixing coefficient between $\mathcal{A}$ and $\mathcal{B}$:

$$\varphi(\mathcal{A},\mathcal{B}) = \sup_{(A,B)\in\mathcal{A}\times\mathcal{B}} |\mathbb{P}(B\mid A) - \mathbb{P}(B)|. \tag{1.4}$$

The following relations between these coefficients hold:

$$2\alpha(\mathcal{A},\mathcal{B}) \le \beta(\mathcal{A},\mathcal{B}) \le \varphi(\mathcal{A},\mathcal{B}) \le 1. \tag{1.5}$$

The Markov chains for instance are geometrically φ-mixing under the Döblin recurrence condition [Doob (1953)] and β-mixing under the milder Harris recurrence condition if the underlying space is finite [Davydov (1973)]. Mokkadem (1990) obtains sufficient conditions for $\mathbb{R}^d$-valued polynomial autoregressive processes to be Harris recurrent and geometrically absolutely regular. He also shows that these processes may fail to be φ-mixing.

As in the independent case, the first results for weakly dependent observations were concerned with the (possibly multivariate) empirical distribution function. Berkes and Philipp (1977) [resp. Philipp and

Pinzur (1980)] generalized Donsker's theorem for the univariate (resp. d-dimensional) empirical distribution function of a strongly mixing stationary sequence with $\alpha_n = O(n^{-a})$, $a > 4+2d$. Next, Yoshihara (1979) weakened the strong mixing assumption of Berkes and Philipp: he proved that, if the strong mixing coefficients $(\alpha_n)_{n>0}$ of the sequence $(\xi_i)_{i\in\mathbb{Z}}$ satisfy $\alpha_n = O(n^{-a})$, for some $a > 3$, then the uniform CLT for the univariate distribution function holds. Dhompongsa (1984) extended Yoshihara's result to the multivariate empirical distribution function. For $\mathbb{R}^d$-valued random variables, he obtained the uniform CLT for the multivariate empirical distribution function under the condition $\alpha_n = O(n^{-a})$, $a > d+2$, which is the best known result for the class of quadrants in the strong mixing case.

Doukhan, León and Portal (1987) studied classes of functions embedded in Hilbertian spaces in the strong mixing case. Massart (1987) proved a uniform CLT for strongly or uniformly mixing empirical processes in a more general framework: if the class $\mathcal{F}$ has a finite dimension d of $\mathcal{L}_1(P)$-entropy with bracketing, and if the strong mixing coefficients satisfy $\alpha_n = O(n^{-a})$ for some $a > 32d+3$, then the uniform CLT holds. Moreover, nonpolynomial covering numbers are allowed in the geometrically mixing case. Some related results in the setting of polynomial bracketing covering numbers [with respect to $\mathcal{L}_2(P)$] may be found in Andrews and Pollard (1994). Looking carefully at these papers, we see that, for classes $\mathcal{F}$ with finite dimension of entropy, some polynomial rate of decay of the strong mixing coefficients is required and that this rate depends on the entropy dimension of $\mathcal{F}$.

By contrast, Arcones and Yu (1994) have established a uniform CLT for absolutely regular empirical processes indexed by Vapnik-Chervonenkis subgraph classes of functions under some polynomial rate of decay of the β-mixing coefficients not depending on the entropy dimension of the class. This makes feasible the existence of a uniform CLT for β-mixing empirical processes which does not require a geometrical decay of the mixing coefficients. Actually since β-mixing allows decoupling (*see* Berbee's lemma in section 3 below), it is much easier to work with this notion rather than with strong mixing. Moreover it still covers a wide class of examples (from this point of view, the φ-mixing may happen to be too restrictive). Hence, in this paper, we shall deal with β-mixing, which seems to be a good compromise.

Our approach consists in exhibiting a norm which plays the role of the $\mathcal{L}_2(P)$-norm in the independent setting. At a first glance, one could think that the Hilbertian pseudonorm associated with the limiting covariance of Z_n should be a good candidate. It has two main defaults. First, the existence of a limiting variance for $Z_n(f)$ does not imply the CLT. Secondly it does not allow bracketing. We will provide a norm depending on P and on the mixing structure of the sequence of observations, which coincides with the usual $\mathcal{L}_2(P)$-norm in the independent case. Denoting by $\mathcal{L}_{2,\beta}(P)$ the so-defined normed space, the finite dimensional convergence of Z_n to some Gaussian vector with covariance function Γ holds on $\mathcal{L}_{2,\beta}(P)$. Γ is the limiting covariance of Z_n and is majorized by the square of the $\mathcal{L}_{2,\beta}(P)$-norm. Moreover, this norm allows bracketing and we will obtain a generalization of Ossiander's theorem by simply measuring the size of the brackets with the $\mathcal{L}_{2,\beta}(P)$-norm instead of the $\mathcal{L}_2(P)$-norm.

2. STATEMENT OF RESULTS

Throughout the sequel, the underlying probability space $(\Omega, \mathcal{T}, \mathbb{P})$ is assumed to be rich enough in the following sense: there exists a random variable U with uniform distribution over $[0,1]$, independent of the sequence $(\xi_i)_{i\in\mathbb{Z}}$.

For any numerical integrable function f, we set $E_P(f) = \int_{\mathcal{X}} f(x)P(dx)$. For any $r \geq 1$, let $\mathcal{L}_r(P)$ denote the class of numerical functions on $(\mathcal{X}, P)$ such that $\|f\|_r = [E_P(|f|^r)]^{1/r} < +\infty$.

Since $(\xi_i)_{i\in\mathbb{Z}}$ is a strictly stationary sequence, the mixing coefficients $(\beta_n)_{n>0}$ of the sequence $(\xi_i)_{i\in\mathbb{Z}}$ are defined by $\beta_n = \beta(\mathcal{F}_0, \mathcal{G}_n)$, where $\mathcal{F}_0 = \sigma(\xi_i : i \leq 0)$ and $\mathcal{G}_n = \sigma(\xi_i : i \geq n)$. $(X_i)_{i\in\mathbb{Z}}$ is called a β-mixing sequence if $\lim_{n\to+\infty} \beta_n = 0$. Examples of such sequences may be found in Davydov (1973), Bradley (1986) and Doukhan (1994).

We now need to recall the definitions of entropy and entropy with bracketing.

Entropy. – Given a metric set (V, d), let $\mathcal{N}(\delta, V, d)$ be the minimum of

$$\{n \in \mathbb{N} : \exists S_n = \{x_1, ..., x_n\} \subset S, \text{ s.t. } \forall x \in S, d(x, S_n) \leq \delta\}.$$

The entropy function $H(\delta, V, d)$ is the logarithm of $\mathcal{N}(\delta, V, d)$.

Bracketing. – Let V be some linear subspace of the space of numerical functions on $(\mathcal{X}, P)$. Assume that there exists some application $\Lambda : V \to \mathbb{R}^+$ such that, for any f and any g in V,

$$|f| \leq |g| \text{ implies } \Lambda(f) \leq \Lambda(g). \tag{2.1}$$

Assume that $\mathcal{F} \subset V$. A pair $[f, g]$ of elements of V such that $f \leq g$ is called a bracket of V. $\mathcal{F}$ is said to satisfy (A.1) if, for any $\delta > 0$, there exists a finite collection $S(\delta)$ of brackets of V such that

$$\text{for all } f \in \mathcal{F}, \text{ there exists } [g, h] \text{ in } S(\delta) \text{ such that } g \leq f \leq h \text{ and } \Lambda(h - g) \leq \delta. \tag{2.2}$$

The bracketing number $\mathcal{N}_{[]}(\delta, \mathcal{F})$ of $\mathcal{F}$ with respect to (V, Λ) is the minimal cardinality of such collections $S(\delta)$. The entropy with bracketing $H_{[]}(\delta, \mathcal{F}, \Lambda)$ is the logarithm of $\mathcal{N}_{[]}(\delta, \mathcal{F})$. When Λ is a norm, denoting by d_Λ the corresponding metric, it follows from (2.1) and (2.2) that

$$H(\delta, \mathcal{F}, d_\Lambda) \leq H_{[]}(\delta, \mathcal{F}, \Lambda). \tag{2.3}$$

We now define a new norm, which emerges from a covariance inequality due to Rio (1993).

The $\mathcal{L}_{2,\beta}(P)$-spaces

Notations. – Throughout the sequel, we make the convention that $\beta_0 = 1$. If $(u_n)_{n \geq 0}$ is a nonincreasing sequence of nonnegative real numbers, we denote by $u(.)$ the rate function defined by $u(t) = u_{[t]}$. For any nonincreasing function ψ, let ψ^{-1} denote the inverse function of ψ,

$$\psi^{-1}(u) = \inf \{t : \psi(t) \leq u\}.$$

For any f in $\mathcal{L}_1(P)$, we denote by Q_f the quantile function of $|f(\xi_0)|$, which is the inverse of the tail function $t \to \mathbb{P}(|f(\xi_0)| > t)$.

Let $(\alpha_n)_{n \geq 0}$ denote the sequence of strong mixing coefficients of $(\xi_i)_{i \in \mathbb{Z}}$. It follows from (1.5) that $\alpha^{-1}(u) \leq \beta^{-1}(2u)$. Hence, by Theorem 1.2 in Rio (1993), the following result holds.

PROPOSITION 1 [Rio, 1993]. – *Assume that the β-mixing coefficients of $(\xi_i)_{i \in \mathbb{Z}}$ satisfy the summability condition*

$$\sum_{n \geq 0} \beta_n < +\infty. \tag{2.4}$$

Let $\mathcal{L}_{2,\beta}(P)$ denote the class of numerical functions f such that

$$\|f\|_{2,\beta} = \sqrt{\int_0^1 \beta^{-1}(u)[Q_f(u)]^2 \, du} < +\infty. \tag{2.5}$$

Then, for any f *in* $\mathcal{L}_{2,\beta}(P)$,

$$\sum_{t\in\mathbb{Z}} |\mathrm{Cov}\,(f(\xi_0), f(\xi_t))| \leq 4\|f\|_{2,\beta}^2,$$

and denoting by $\Gamma(f,f)$ *the sum of the series* $\sum_{t\in\mathbb{Z}} \mathrm{Cov}\,(f(\xi_0), f(\xi_t))$, *we have:*

$$(2.6)\qquad \begin{cases} \mathrm{Var}\, Z_n(f) \leq 4\|f\|_{2,\beta}^2 \\ \text{and} \\ \lim_{n\to+\infty} \mathrm{Var}\, Z_n(f) = \Gamma(f,f) \leq 4\|f\|_{2,\beta}^2. \end{cases}$$

Remark. – Note that $\sum_{n\geq 0} \beta_n = \int_0^1 \beta^{-1}(u)du$. So, under condition (2.4), $\mathcal{L}_{2,\beta}(P)$ contains the space $\mathcal{L}_\infty(P)$ of bounded functions. Moreover, we will prove in section 6 the following basic lemma.

Lemma 1. – *Assume that the sequence of* β*-mixing coefficients of* $(\xi_i)_{i\in\mathbb{Z}}$ *satisfies condition* (2.4). *Then,* $\mathcal{L}_{2,\beta}(P)$*, equipped with* $\|.\|_{2,\beta}$ *is a normed subspace of* $\mathcal{L}_2(P)$ *satisfying* (2.1), *and for any* f *in* $\mathcal{L}_{2,\beta}(P)$, $\|f\|_2 \leq \|f\|_{2,\beta}$.

We now define the corresponding weak space as follows. Let $B(t) = \int_0^t \beta^{-1}(u)du$. For any measurable numerical function f, we set

$$(2.7)\qquad \Lambda_{2,\beta}(f) = \sup_{t\in]0,1]} Q_f(t)\sqrt{B(t)}.$$

Throughout the sequel, $\Lambda_{2,\beta}(P)$ denotes the space of measurable functions f on $\mathcal{X}$ such that $\Lambda_{2,\beta}(f) < +\infty$. Clearly $\mathcal{L}_{2,\beta}(P) \subset \Lambda_{2,\beta}(P)$ and

$$(2.8)\qquad \Lambda_{2,\beta}(f) \leq \|f\|_{2,\beta} \quad \text{for any } f \in \mathcal{L}_{2,\beta}(P).$$

Now we compare the so defined spaces with the Orlicz spaces of functions, and with the so-called weak Orlicz spaces of functions.

Comparison with Orlicz spaces

Let Φ be the class of increasing functions

$$\Phi = \Big\{\phi : \mathbb{R}^+ \to \mathbb{R}^+ : \phi \nearrow \text{ convex and differentiable}, \phi(0) = 0,\ \lim_{+\infty} \frac{\phi(x)}{x} = \infty\Big\}.$$

For any $\phi \in \Phi$, we denote by $\mathcal{L}_{\phi,2}(P)$ the space of functions

$$\mathcal{L}_{\phi,2}(P) = \{h : \mathcal{X} \to \mathbb{R} \text{ s.t. } E_P(\phi(th^2)) < +\infty \text{ for some } t > 0\}.$$

$\mathcal{L}_{\phi,2}(P)$ is equipped with the following norm :

$$\|h\|_{\phi,2} = \inf \left\{ c > 0 : E_P\left(\phi\left(\left|\frac{h}{c}\right|^2\right)\right) \leq 1 \right\}.$$

This norm is the usual Orlicz norm associated with the function $x \to \phi(x^2)$.

Let us now define the corresponding weak Orlicz spaces. For any ϕ in Φ, let $\Lambda_\phi(P)$ be the space of measurable functions f on $\mathcal{X}$ such that

$$\sup_{u>0} [uP(\phi(f^2) > u)] = \sup_{t>0} [\phi(t^2)P(|f| > t)] < +\infty,$$

or equivalently such that

$$\Lambda_\phi(f) = \sup_{u\in]0,1]} ([\phi^{-1}(1/u)]^{-1/2} Q_f(u)) < +\infty.$$

Of course, by Markov's inequality, $\Lambda_\phi(h) \leq \|h\|_{\phi,2}$.

The purpose of the following lemma is to relate the $\mathcal{L}_{2,\beta}$-norm with $\|.\|_{\phi,2}$ and $\Lambda_\phi(.)$.

LEMMA 2. – *For any element ϕ of Φ, let define the dual function ϕ^* by $\phi^*(y) = \sup_{x>0} [xy - \phi(x)]$. If*

$$\text{(2.9a)} \qquad \int_0^1 \phi^*(\beta^{-1}(u))\, du < +\infty,$$

then

$$\text{(a)} \qquad \|f\|_{2,\beta} \leq \|f\|_{\phi,2} \sqrt{1 + \int_0^1 \phi^*(\beta^{-1}(u))\, du}\,,$$

and so $\mathcal{L}_{\phi,2}(P) \subset \mathcal{L}_{2,\beta}(P)$. If

$$\text{(2.9b)} \qquad \int_0^1 \phi^{-1}(1/u)\beta^{-1}(u) du < +\infty\,,$$

then

$$\text{(b)} \qquad \|f\|_{2,\beta} \leq \Lambda_\phi(f) \sqrt{\int_0^1 \phi^{-1}(1/u)\beta^{-1}(u) du}\,,$$

and so $\Lambda_\phi(P) \subset \mathcal{L}_{2,\beta}(P)$.

The main result

In a recent paper, Doukhan, Massart and Rio (1994) prove that, for any finite subset $\mathcal{G} \subset \mathcal{L}_{2,\beta}(P)$, the convergence of $\{Z_n(g) : g \in \mathcal{G}\}$ to a Gaussian random vector with covariance function Γ holds true. Moreover, a counterexample shows that fidi convergence may fail to hold if $\mathcal{G} \not\subset \mathcal{L}_{2,\beta}(P)$. So, it is quite natural to assume that $\mathcal{F} \subset \mathcal{L}_{2,\beta}(P)$. Now the question raises whether an integrability condition on the metric entropy with bracketing in $\mathcal{L}_{2,\beta}(P)$ is sufficient to imply the uniform CLT. In fact the answer is positive, as shown by the following theorem.

THEOREM 1. – *Let $(\xi_i)_{i\in\mathbb{Z}}$ be a strictly stationary and β-mixing sequence of random variables with common marginal distribution P. Assume that the sequence $(\beta_n)_{n\geq 0}$ satisfies* (2.4). *Let $\mathcal{F}$ be a class of functions f, $\mathcal{F} \subset \mathcal{L}_{2,\beta}(P)$. Assume that the entropy with bracketing with respect to $\|.\|_{2,\beta}$, which we denote by $H_\beta(\delta, \mathcal{F})$, satisfies the integrability condition*

$$(2.10) \qquad \int_0^1 \sqrt{H_\beta(u, \mathcal{F})}\, du < +\infty.$$

Then,

(i) *the series* $\sum_{t\in\mathbb{Z}} \mathrm{Cov}\,(f(\xi_0), f(\xi_t))$ *is absolutely convergent over $\mathcal{F}$ to a nonnegative quadratic form $\Gamma(f, f)$, and*

$$(\Gamma(f,f))^{1/2} = \|f\|_\Gamma \leq 2\|f\|_{2,\beta}.$$

(ii) *There exists a sequence $(Z^{(n)})_{n>0}$ of Gaussian processes indexed by $\mathcal{F}$ with covariance function Γ and a.s. uniformly continuous sample paths such that*

$$\sup_{f\in\mathcal{F}} |Z_n(f) - Z^{(n)}(f)| \to_P 0 \quad \text{as } n \to +\infty.$$

Applications of Theorem 1

1. *Orlicz spaces.* – Let ϕ be some element of Φ such that the mixing coefficients β_n satisfy

$$\sum_{n>0} (\phi')^{-1}(n)\beta_n < +\infty \qquad \text{(S.1)}$$

and suppose that $\mathcal{F} \subset \mathcal{L}_{\phi,2}(P)$. Then (2.9*a*) holds [*see* Rio, 1993]. Hence it follows from Lemma 2 that $\mathcal{F}$ satisfies the assumptions of Theorem 1 if the metric entropy with bracketing of $\mathcal{F}$ in $\mathcal{L}_{\phi,2}(P)$ fulfills the integrability condition

$$\int_0^1 \sqrt{H_{[\,]}(t, \mathcal{F}, \|.\|_{\phi,2})}dt < +\infty. \tag{2.11}$$

For example, if $\phi(x) = x^r$, $\mathcal{L}_{\phi,2}(P) = \mathcal{L}_{2r}(P)$, $\|.\|_{\phi,2}$ is the usual norm $\|.\|_{2r}$ and (S.1) means that the series $\sum_{n>0} n^{1/(r-1)}\beta_n$ is convergent.

As an application of (2.11), we can derive the following striking result. Assume that the mixing coefficients satisfy $\beta_k = O(b^k)$ for some b in $]0, 1[$. There exists some $s > 0$ such that (S.1) holds with

$$\phi(x) = h(sx) = (1 + sx)\log\,(1 + sx) - sx. \tag{2.12}$$

We notice that $\mathcal{L}_{\phi,2}(P)$ is the space $L_{2\log^+}(P)$ of numerical functions f such that $E_P(f^2 \log^+ |f|) < \infty$ and it is equipped with a norm which is equivalent to the usual Orlicz norm in this space. Hence, by Lemma 2 and Theorem 1, the uniform CLT for the empirical process holds as soon as the entropy with bracketing of $\mathcal{F}$ in $L_{2\log^+}(P)$ satisfies the usual integrability condition.

2. *Weak Orlicz spaces.* – Let ϕ be some element of Φ such that $x \to x^{-r}\phi(x)$ is nondecreasing for some $r > 1$. Then, (2.9*b*) is equivalent to the summability condition of Herrndorf (1985) on the mixing coefficients:

$$\sum_{n>0} \phi^{-1}(1/\beta_n)\beta_n < +\infty \tag{S.2}$$

It follows from Lemma 2 that $\mathcal{F} \subset \Lambda_\phi(P)$ satisfies the assumptions of Theorem 1 if the entropy with bracketing of $\mathcal{F}$ with respect to $\Lambda_\phi(.)$ verifies

$$\int_0^1 \sqrt{H_{[\,]}(t, \mathcal{F}, \Lambda_\phi)}\, dt < +\infty. \tag{2.13}$$

Some calculations (*cf.* Rio, 1993) show that (S.2) is stronger than (S.1). For example, if $\phi(x) = x^r$ for some $r > 1$, $\Lambda_\phi(P)$ is the usual weak $\mathcal{L}_{2r}(P)$-space $\Lambda_{2r}(P)$, equipped with the usual weak norm $\Lambda_{2r}(.)$ and (S.2) is equivalent to the convergence of the series $\sum_{n>0} \beta_n^{1-1/r}$, while (S.1) holds iff the series $\sum_{n>0} n^{1/(r-1)}\beta_n$ is convergent, which is a weaker condition. We refer the reader to application 3 of Theorem 1 in Doukhan, Massart and Rio (1994), for more about comparisons with the previous conditions of Ibragimov (1962) and Herrndorf (1985).

3. *Conditions involving the envelope function of $\mathcal{F}$.* – Let $\mathcal{G} \subset \mathcal{L}_\infty(P)$ be a class of functions satisfying the entropy condition

$$\int_0^1 \sqrt{H(t, \mathcal{G}, \|.\|_\infty)}\, dt < +\infty.$$

Let F be some element of $\mathcal{L}_{2,\beta}(P)$ satisfying $F \geq 1$ and $\mathcal{F} = \{gF : g \in \mathcal{G}\}$. Both Theorem 1 and the elementary inequality $\|gF\|_{2,\beta} \leq \|g\|_\infty \|F\|_{2,\beta}$ imply the uniform CLT.

4. *Conditions involving the $\mathcal{L}_2(P)$-entropy of $\mathcal{F}$.* – In this subsection, we consider the following problem: given a class $\mathcal{F} \subset \mathcal{L}_2(P)$, with metric entropy with bracketing $H_2(.)$ with respect to the metric induced by $\|.\|_2$, we want to find a condition on the mixing coefficients and on the $\mathcal{L}_2(P)$-entropy implying (2.10). Throughout application 4, we assume that the envelope function of $\mathcal{F}$ is in $\Lambda_{2r}(P)$, for some $r \in]1, +\infty]$, where we make the convention that $\Lambda_\infty(P) = \mathcal{L}_\infty(P)$. We shall prove in appendix C that condition (2.10) is satisfied if any of the three following conditions is fulfilled:

$$(2.15) \qquad \begin{cases} \beta_n = O(n^{-b}) \quad \text{with} \quad b > r/(r-1), \\ \text{and} \\ \displaystyle\int_0^1 v^{-r/(b(r-1))} H_2^{1/2}(v) dv < +\infty, \end{cases}$$

$$(2.16) \qquad \begin{cases} H_2(u) = O(u^{-2\zeta}) \quad \text{with } \zeta \in]0, 1[, \\ \text{and} \\ \displaystyle\sum_{n>0} n^{-1/2} \beta_n^{(1-\zeta)(r-1)/(2r)} < +\infty, \end{cases}$$

$$(2.17) \qquad H_2(u) = O(|\log u|) \quad \text{and} \quad \sum_{n>0} \frac{1}{n} \sqrt{\frac{\sum_{k \geq n} \beta_k^{1-1/r}}{\log n}} < +\infty.$$

In particular, (2.15) and (2.16) are satisfied if $\beta_n = O(n^{-b})$, $H_2(u) = O(u^{-2\zeta})$ with $b(1-\zeta) > r/(r-1)$. Also (2.17) holds whenever $\beta_n = O(n^{-1}(\log n)^{-2-\delta})$ and $\mathcal{F}$ is the class of quadrants, which improves in this special case on Corollary 3 of Arcones and Yu (1994).

The outline of the paper is as follows: in section 3, the technique of blocking for mixing processes is applied to establish an upper bound on the mean of the supremum of Z_n over a finite class of bounded functions. Next, in section 4, we show how both the upper bound of section 3 and a generalization of Ossiander's method for proving tightness of the empirical process yield the stochastic equicontinuity of $\{Z_n(f) : f \in \mathcal{F}\}$ under the assumptions of Theorem 1. In section 5, we weaken the bracketing conditions, in the spirit of Andersen *et al.* (1988). The stochastic equicontinuity of Z_n is ensured in the bounded case by the following result, which is in fact the crucial technical part of the paper. In what follows, for any positive (non necessarily measurable) random element X, $\mathbb{E}(X)$ denotes the smallest expectation of the (measurable) random variables majorizing X.

THEOREM 2. – *Let σ be a positive number and let $\mathcal{F}_\sigma \subset \mathcal{L}_{2,\beta}$ be a class of functions satisfying the condition $\|f\|_{2,\beta} \leq \sigma$ for any f in $\mathcal{F}_\sigma$. Suppose that $\mathcal{F}_\sigma$ fulfills* (2.10) *and that for some $M \geq 1$, $|f| \leq M$ for any function f in $\mathcal{F}_\sigma$. Then there exists a constant K, depending only on $\theta = (\sum_{n\geq 0} \beta_n)^{1/2}$, such that, for any positive integer q,*

$$\mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n(f)|) \leq K\Big(\varphi(\sigma) + \frac{Mq\varphi^2(\sigma)}{\sigma^2\sqrt{n}} + \sqrt{n}M\beta_q\Big). \tag{2.18}$$

3. A MAXIMAL INEQUALITY FOR β-MIXING PROCESSES INDEXED BY FINITE CLASSES OF FUNCTIONS

In order to establish the stochastic equicontinuity of the empirical process $\{Z_n(f) : f \in \mathcal{F}\}$, we need to control the mean of the supremum of the empirical process $Z_n(.)$ over a finite class of bounded functions. This control is performed via the approximation of the original process by conveniently defined independent random variables. The main argument is Berbee's coupling lemma.

LEMMA [Berbee (1979)]. – *Let X and Y be two r.v.'s taking their values in Borel spaces S_1 and S_2 respectively and let U be a r.v. with uniform distribution over $[0,1]$, independent of (X,Y). Then, there exists a random variable $Y^* = f(X,Y,U)$, where f is a measurable function from $S_1 \times S_2 \times [0,1]$ into S_2, such that:*

- *Y^* is independent of X and has the same distribution as Y.*
- *$\mathbb{P}(Y \neq Y^*) = \beta$, where β denotes the β-mixing coefficient between the σ-fields generated by X and Y respectively.*

The following lemma will be used repeatedly in the proof of Theorem 2. In fact, this lemma is the only thing one has to know about β-mixing sequences in order to prove Theorem 2.

LEMMA 3. – *Let a and δ be positive reals and $\mathcal{G}$ be any finite subclass of $\mathcal{L}_{2,\beta}(P)$ satisfying the following assumptions:*
(i) *For any g in $\mathcal{G}$, $\mathbb{E}(g(\xi_0)) = 0$.*
(ii) *For any g in $\mathcal{G}$, $\|g\|_\infty \leq a$ and $\|g\|_{2,\beta} \leq \delta$.*
Let $L(\mathcal{G}) = \max(1, \log|\mathcal{G}|)$, where $|\mathcal{G}|$ denotes the cardinality of $\mathcal{G}$. There exists some universal positive constant C such that, for any q in $[1, n]$,

$$\mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n(g)|) \leq C\Big(\delta\sqrt{L(\mathcal{G})} + aq\,\frac{L(\mathcal{G})}{\sqrt{n}} + a\beta_q\sqrt{n}\Big).$$

Proof. – The proof of Lemma 3 is mainly based on Bernstein's exponential inequalities for independent r.v.'s and on Berbee's lemma. Let us now give a Corollary of Berbee's lemma, whose proof only uses the fact that a countable product of Polish spaces is a Polish space (the proof will be omitted, being elementary).

PROPOSITION 2. – *Let $(X_i)_{i>0}$ be a sequence of random variables taking their values in a Polish space $\mathcal{X}$. For any integer $j > 0$, let $b_j = \beta(\sigma(X_j), \sigma(X_i : i > j))$. Then, there exists a sequence $(X_i^*)_{i>0}$ of* independent *random variables such that, for any positive integer j, X_j^* has the same distribution as X_j and $\mathbb{P}(X_j \neq X_j^*) \leq b_j$.*

Invoke now Proposition 2 to construct a sequence $(\xi_i^*)_{i>0}$ of real-valued r.v.'s such that the random vectors $Y_k = (\xi_{qk+1}, \ldots, \xi_{q(k+1)})$ and $Y_k^* = (\xi_{qk+1}^*, \ldots, \xi_{q(k+1)}^*)$ fulfill the conditions below:

- For any $k > 0$, Y_k^* and Y_k have the same distribution and $\mathbb{P}(Y_k \neq Y_k^*) \leq \beta_q$.
- the r.v.'s $(Y_{2k}^*)_{k>0}$ are independent, the r.v.'s $(Y_{2k-1}^*)_{k>0}$ are independent.

Now, let $Z_n^* = n^{-1/2}\sum_{i=1}^{n}(\delta_{\xi_i^*} - P)$ denote the normalized empirical measure associated with the r.v.'s ξ_i^*:

$$\mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n(g)|) \leq \mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n^*(g)|) + \mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n(g) - Z_n^*(g)|). \tag{3.1}$$

First, we give an upper bound on the error term due to this substitution. Clearly,

$$\sqrt{n}\,|Z_n(g) - Z_n^*(g)| \leq 2\|g\|_\infty \sum_{i=1}^{n} \mathbb{1}_{(\xi_i \neq \xi_i^*)}.$$

Now, recall that, for any positive integer i, $\mathbb{P}(\xi_i \neq \xi_i^*) \leq \beta_q$. It follows that

$$\mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n(g) - Z_n^*(g)|) \leq 2a\beta_q\sqrt{n}. \tag{3.2}$$

It remains to handle the first term on right hand in (3.1). Let

$$Y_k^*(g) = \sum_{i=qk+1}^{q(k+1)\wedge n} g(\xi_i^*).$$

The random variables $Y_k^*(g)$ are centered and each bounded by qa. Moreover, it follows from Proposition 1 that

$$q^{-1}\mathrm{Var}\, Y_k^*(g) \leq 4\|g\|_{2,\beta}^2 \leq 4\delta^2$$

for any g in $\mathcal{G}$. So, recalling that the r.v.'s $(Y_{2k}^*)_{k>0}$ are independent, and that the corresponding r.v.'s with odd index also are independent, and applying twice Bernstein's inequality [*see* Pollard (1984), p. 193] we get: there exists some positive constant c, such that for any positive λ,

$$\mathbb{P}(|Z_n^*(g)| \geq \lambda) \leq 4\exp(-c\inf((\lambda/\delta)^2, \lambda\sqrt{n}/(qa))). \tag{3.3}$$

It follows that

$$\begin{aligned}\mathbb{P}(\sup_{g\in\mathcal{G}} |Z_n^*(g)| \geq \lambda) &\leq 4\exp(-(c\inf((\lambda/\delta)^2, \lambda\sqrt{n}/(qa)) - L(\mathcal{G}))^+)\\ &\leq 4\exp(-(c(\lambda/\delta)^2 - L(\mathcal{G}))^+)\\ &\quad + 4\exp(-(c\lambda\sqrt{n}/(qa) - L(\mathcal{G}))^+).\end{aligned}$$

Let λ_0 and λ_1 be the positive numbers defined by the equations

$$c(\lambda_0/\delta)^2 = L(\mathcal{G}) \quad \text{and} \quad c\lambda_1\sqrt{n} = qaL(\mathcal{G}). \tag{3.4}$$

Integrating the above inequality yields

$$\begin{aligned}&\mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n^*(g)|)\\ &\leq 4(\lambda_0+\lambda_1) + 4\int_0^\infty \exp\left(-\frac{c\lambda^2}{\delta^2}\right)d\lambda + 4\int_0^\infty \exp\left(-\frac{c\lambda\sqrt{n}}{qa}\right)d\lambda\\ &= 4(\lambda_0+\lambda_1) + 4\delta\sqrt{\frac{\pi}{c}} + \frac{4\sqrt{n}}{cqa}.\end{aligned} \tag{3.5}$$

Both (3.4) and (3.5) ensure that

$$\mathbb{E}(\sup_{g\in\mathcal{G}} |Z_n^*(g)|) \leq C\left(\delta\sqrt{L(\mathcal{G})} + aq\frac{L(\mathcal{G})}{\sqrt{n}}\right), \tag{3.6}$$

for some positive constant C, which, together with (3.2), implies Lemma 3.

4. PROOF OF THEOREM 1

We first notice that we may assume $E_P(f) = 0$ for any f in $\mathcal{F}$. This follows from the elementary inequality below (proof omitted).

CLAIM 1. – *Let* $\theta = (\sum_{n\geq 0} \beta_n)^{1/2}$. *For any* f *in* $\mathcal{L}_{2,\beta}(P)$,

$$\|f - E_P(f)\|_{2,\beta} \leq \|f\|_{2,\beta} + \theta\|f\|_1 \leq \|f\|_{2,\beta}(1 + \theta)$$

Throughout the proof, $\mathcal{L}_{2,\beta}$ stands for $\mathcal{L}_{2,\beta}(P)$. (i) follows from Proposition 1. To prove (ii), we know from Doukhan, Massart, Rio (1994) that fidi convergence of Z_n holds. Moreover (2.3) and (i) ensure that the integral entropy condition (2.10) imply that $(\mathcal{F}, \Gamma)$ is pregaussian via Dudley's criterion. It is well known [*see* for instance Pollard (1990)] that (ii) will be ensured by the asymptotic stochastic equicontinuity of Z_n with respect to the $\mathcal{L}_{2,\beta}$-metric. This will follow easily from the fundamental inequality below, which is in fact a corollary of Theorem 2.

THEOREM 3. – *Let* σ *be a positive nunber and let* $\mathcal{F}_\sigma \subset \mathcal{L}_{2,\beta}$ *be a class of functions satisfying the condition* $\|f\|_{2,\beta} \leq \sigma$ *for any* f *in* $\mathcal{F}_\sigma$. *Suppose that* $\mathcal{F}_\sigma$ *fulfills* (2.10) *and let*

$$\varphi(\sigma) = \int_0^\sigma \sqrt{H_\beta(u, \mathcal{F}_\sigma)}\, du.$$

Let $B : \mathbb{R}^+ \to [0, B(1)]$ *be the application defined by* $B(x) = \int_0^x \beta^{-1}(t)dt$ *[note that* $\beta^{-1}(t) = 0$ *if* $t > 1$*] and, for any measurable function* h, *let*

$$\delta_h(\varepsilon) = \sup_{t\leq\varepsilon} Q_h(t)\sqrt{B(t)}.$$

Let F *be some positive function satisfying the following conditions:* $F \geq |f|$ *for any* f *in* $\mathcal{F}$, *and* $\lim_{\varepsilon\to 0} \delta_F(\varepsilon) = 0$. *Then there exists some positive constant* A *depending only on* $B(1) = \theta^2$ *such that, for any positive integer* n,

$$\mathbb{E}(\sup_{f\in\mathcal{F}_\sigma} |Z_n(f)|) \leq A\varphi(\sigma)\left[1 + \frac{\delta_F(1 \wedge \varepsilon(\sigma, n))}{\sigma}\right], \tag{4.1}$$

where $\varepsilon(\sigma, n)$ *is the unique solution of the equation*

$$x^2/B(x) = \varphi^2(\sigma)/(n\sigma^2).$$

Remark. – Since B is a nondecreasing concave function, $x \to x^2/B(x)$ is an increasing and continuous application from $\mathbb{R}^+$ onto $\mathbb{R}^+$. Hence the equation defining $\varepsilon(\sigma, n)$ has an unique solution, and $\lim_{n \to +\infty} \varepsilon(\sigma, n) = 0$. Furthermore, the bracketing condition implies the existence of an envelope function F belonging to $\mathcal{L}_{2,\beta}$, which in turn ensures that $\lim_{\varepsilon \to 0} \delta_F(\varepsilon) = 0$, via (2.8).

It follows from the above remark that, if n is large enough

$$\mathbb{E}(\sup_{f \in \mathcal{F}_\sigma} |Z_n(f)|) \leq 2A\varphi(\sigma).$$

Hence, applying Theorem 3 to $\mathcal{F}_\sigma = \{f - g : f, g \in \mathcal{F} \text{ and } \|f - g\|_{2,\beta} \leq \sigma\}$, we obtain the asymptotic stochastic equicontinuity of $\{Z_n(f) : f \in \mathcal{F}\}$, which completes the proof of Theorem 1.

Proof of Theorem 2. – In order to prove Theorem 2 we shall use a chaining argument with adaptive truncatures. This technique was initiated by Bass (1985) in the context of set-indexed partial sum processes and then used by Ossiander (1987) and next by Andersen *et al.* (1988) in order to prove uniform central limit theorems for function-indexed empirical processes. Contrary to Ossiander and Andersen *et al.*, we do not force the brackets to be decreasing all along the chaining. This may be done using a new chaining decomposition ([1]).

We need to impose some additional monotonicity condition on the entropy function H_β. The following claim is a direct consequence of Lemma 6, which is elementary but of independent interest. This lemma will be proved in section 6.

CLAIM 2. – *There exists a nonincreasing function $H(.)$, majorizing $H_\beta(., \mathcal{F}_\sigma)$, such that $x \to x^4 H(x)$ is nondecreasing and*

$$\int_0^\sigma \sqrt{H(t)}\, dt \leq 4\varphi(\sigma).$$

Since Theorem 2 is trivial when $\varphi(\sigma)/\sigma \geq 2^{-6}\sqrt{n}$ (in that case Theorem 2 holds with $K = 2^{13}$) we shall assume that condition

$$\varphi(\sigma)/\sigma < 2^{-6}\sqrt{n} \tag{4.2}$$

([1]) The idea of using such a decomposition was given to us by D. Pollard (private communication).

is satisfied. We define $\delta_0 = \sigma$ and then, for any integer k, $\delta_k = 2^{-k}\delta_0$. Now, since (A1) holds for $\mathcal{F}$, for each nonnegative integer k, we may choose a covering of $\mathcal{F}$ by brackets $B_{j,k} = [g_{j,k}, h_{j,k}], 1 \leq j \leq J_k$, with $\|h_{j,k} - g_{j,k}\|_{2,\beta} \leq \delta_k$ and $J_k \leq \exp(H(\delta_k))$. Of course we may assume that $|h_{j,k} - g_{j,k}| \leq 2M$. Then in each bracket $B_{j,k}$ we fix a point $f_{j,k}$ belonging to $\mathcal{F}$. We now define a mapping ψ_k from $\mathcal{F}$ to $[1, J_k]$ by $\psi_k f = \min\{j \in [1, J_k] : f \in B_{j,k}\}$. Finally we set $\Pi_k f = f_{\psi_k f,k}$ and $\Delta_k f = h_{\psi_k f,k} - g_{\psi_k f,k}$. It comes from these definitions that

$$|f - \Pi_k f| \leq \Delta_k f, \tag{4.3}$$

with

$$\|\Delta_k f\|_{2,\beta} \leq \delta_k. \tag{4.4}$$

We need to define some more functions and parameters. For any positive δ we set $\mathbb{H}(\delta) = \sum_{\delta_k \geq \delta} H(\delta_k)$. This function $\mathbb{H}$ will be usefull because the mapping $(\psi_0, \cdots, \psi_k)$ ranges in a finite set with cardinality less than or equal to $\exp(\mathbb{H}(\delta_k))$. The choice of parameters that we propose hereunder will tend to make the three terms of Lemma 3 of the same order of magnitude when applied to the control of the different terms of decomposition (4.8). We define $q(\delta) = \min\{s \in \mathbb{N}^* : \beta(s)/s \leq \mathbb{H}(\delta)/n\}$ and the parameter $\varepsilon(\delta)$ by $n\varepsilon(\delta) = ((q(\delta) - 1) \vee 1)\mathbb{H}(\delta)$. Let

$$q_k = q(\delta_{k+1}), \qquad \varepsilon_k = \varepsilon(\delta_{k+1}) \tag{4.5}$$

and

$$b_k = 2\sqrt{n}\delta_k(\mathbb{H}(\delta_{k+1}))^{-1/2}. \tag{4.6}$$

We note that both sequences $(b_k)_k$ and $(q_k)_k$ are nonincreasing. Moreover, it follows from the definitions of q and ε that $\beta^{-1}(\varepsilon(\delta)) \geq (q(\delta) - 1) \vee 1$. So $n\varepsilon(\delta) \leq \beta^{-1}(\varepsilon(\delta))\mathbb{H}(\delta)$, which means that

$$\varepsilon_k \leq \beta^{-1}(\varepsilon_k)\mathbb{H}(\delta_{k+1})/n. \tag{4.7a}$$

Also $q(\delta) \leq 2((q(\delta) - 1) \vee 1)$, thus

$$q_k \leq \frac{2n\varepsilon_k}{\mathbb{H}(\delta_{k+1})}. \tag{4.7b}$$

Let

$$N = \min\{k \geq 0 : \delta_k \leq 2^6\varphi(\sigma)/\sqrt{n}\}.$$

We note that, because of (4.2), $N \geq 1$. We will check at the end of the proof of Theorem 2 that $\mathbb{H}(\delta_{N+1}) \leq n$, which implies that $\varepsilon_k \leq 1$ for all $k \leq n$. Finally for any $f \in \mathcal{F}$, we set

$$\nu(f) = [\min\{k \geq 0 : q_k \Delta_k f > b_k\}] \wedge N.$$

Let I denote the identity operator. Starting from

$$I = \Pi_0 + (I - \Pi_N) + \sum_{k=1}^{N} (\Pi_k - \Pi_{k-1})$$

and noting that

$$\sum_{k=0}^{N-1} (\Pi_{k+1} - \Pi_k) \mathbb{1}_{\nu < k+1} = \sum_{k=0}^{N-1} (I - \Pi_k) \mathbb{1}_{\nu = k} + (\Pi_N - I) \mathbb{1}_{\nu < N}$$

we get

$$I = \Pi_0 + \sum_{k=0}^{N} (I - \Pi_k) \mathbb{1}_{\nu = k} + \sum_{k=1}^{N} (\Pi_k - \Pi_{k-1}) \mathbb{1}_{\nu \geq k}.$$

Let k be an integer $1 \leq k \leq N - 1$. Since $b_{k-1} \geq b_k$, we have

$$\{\nu(f) \geq k, q_k \Delta_k f > b_{k-1}\} = \{\nu(f) = k, q_k \Delta_k f > b_{k-1}\}.$$

Plugging this relation in the decomposition above produces the following relation where the summations have to be taken as 0 if $N = 1$

$$\begin{aligned} (4.8) \quad I = & \Pi_0 + (I - \Pi_0) \mathbb{1}_{\nu = 0} \\ & + \sum_{k=1}^{N-1} (I - \Pi_{k-1}) \mathbb{1}_{\nu = k, q_k \Delta_k > b_{k-1}} + (I - \Pi_{N-1}) \mathbb{1}_{\nu \geq N} \\ & + \sum_{k=1}^{N-1} (I - \Pi_k) \mathbb{1}_{\nu = k, q_k \Delta_k \leq b_{k-1}} \\ & + \sum_{k=1}^{N-1} (\Pi_k - \Pi_{k-1}) \mathbb{1}_{\nu \geq k, q_k \Delta_k \leq b_{k-1}}. \end{aligned}$$

The following inequality derives straightforwardly from (4.8)

$$\mathbb{E}(\sup_{f \in \mathcal{F}} |Z_n(f)|) \leq \mathbb{E}_1 + \mathbb{E}_2 + \mathbb{E}_3 + \mathbb{E}_4 + \mathbb{E}_5 + \mathbb{E}_6$$

where

$$\mathbb{E}_1 = \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n(\Pi_0 f)|)$$
$$\mathbb{E}_2 = \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[(f - \Pi_0 f)\mathbb{1}_{\nu(f)=0}]|)$$
$$\mathbb{E}_3 = \sum_{k=1}^{N-1} \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[(f - \Pi_{k-1} f)\mathbb{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}]|)$$
$$\mathbb{E}_4 = \sum_{k=1}^{N-1} \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[(f - \Pi_k f)\mathbb{1}_{\nu(f)=k, q_k\Delta_k f \le b_{k-1}}]|)$$
$$\mathbb{E}_5 = \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[(f - \Pi_{N-1} f)\mathbb{1}_{\nu(f)\ge N}]|)$$
$$\mathbb{E}_6 = \sum_{k=1}^{N-1} \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[(\Pi_k f - \Pi_{k-1} f)\mathbb{1}_{\nu(f)\ge k, q_k\Delta_k f \le b_{k-1}}]|).$$

Control of $\mathbb{E}_1$

We note that Π_0 ranges in the finite set of functions $\mathcal{F}_0$. As a direct application of Lemma 3 and Claim 2 we get that for any integer q

$$\mathbb{E}_1/C \le \sigma(H(\sigma))^{1/2} + 2\,\frac{MqH(\sigma)}{\sqrt{n}} + 2\sqrt{n}M\beta_q$$
$$\le \sigma(H(\sigma))^{1/2} + 2^5\,\frac{Mq\varphi^2(\sigma)}{\sigma^2\sqrt{n}} + 2\sqrt{n}M\beta_q.$$

In order to control the expectations $\mathbb{E}_i, i = 2,3,4$ we will need to relate the $\mathcal{L}_1$-norm of a function which is truncated from below to the $\mathcal{L}_{2,\beta}$-norm of the nontruncated function. This is what is done in the following lemma.

LEMMA 4. – *Let h be some function in $\mathcal{L}_1(P)$. Then, for any ε in $]0,1]$, the following inequality holds*

$$(a) \qquad ||h\mathbb{1}_{h>Q_h(\varepsilon)}||_1 \le \frac{2\varepsilon\,\delta_h(\varepsilon)}{\sqrt{B(\varepsilon)}}.$$

In particular, if

$$(\mathrm{i}) \qquad \Lambda_{2,\beta}(h) = \delta_h(1) \le \delta \qquad \text{and} \qquad a\sqrt{\varepsilon\beta^{-1}(\varepsilon)} \ge \delta,$$

then

$$(b) \qquad ||h\mathbb{1}_{h>a}||_1 \le 2\,\delta\sqrt{\frac{\varepsilon}{\beta^{-1}(\varepsilon)}}.$$

Proof of Lemma 4. – Clearly, for all $t \leq \varepsilon$,

$$Q_h(t) \leq \frac{\delta}{\sqrt{B(t)}}. \tag{4.9}$$

Now $Q_h^{-1}(Q_h(\varepsilon)) \leq \varepsilon$. So,

$$\|h\mathbb{1}_{h>Q_h(\varepsilon)}\|_1 = \int_0^{Q_h^{-1}(Q_h(\varepsilon))} Q_h(t)\,dt \leq \int_0^{\varepsilon} Q_h(t)\,dt,$$

thus, applying (4.9) and the concavity of $B(.)$ provides:

$$\|h\mathbb{1}_{h>Q_h(\varepsilon)}\|_1 \leq \delta_h(\varepsilon)\int_0^{\varepsilon}\frac{dt}{\sqrt{B(t)}} \leq \delta_h(\varepsilon)\sqrt{\frac{\varepsilon}{B(\varepsilon)}}\int_0^{\varepsilon}\frac{dt}{\sqrt{t}},$$

proving (*a*). Noticing that $\varepsilon\beta^{-1}(\varepsilon) \leq B(\varepsilon)$, condition (i) implies, via (4.9) that $a \geq Q_h(\varepsilon)$. Hence, (*b*) follows from (*a*).

We can use this lemma to produce a result which is crucial for what follows.

CLAIM 3. – *For any integer $k \leq N-1$, we have*

$$\sup_{f\in\mathcal{F}} \|\Delta_k f\mathbb{1}_{q_k\Delta_k f>b_k}\|_1 \leq \frac{2\delta_k}{\sqrt{n}}(\mathbb{H}(\delta_{k+1}))^{1/2}.$$

Proof of Claim 3. – We wish to apply Lemma 4 with $a = b_k/q_k$, $\delta = \delta_k$ and $\varepsilon = \varepsilon_k$. We have to check that condition (i) is satisfied. By (4.6) and (4.7*b*) we have

$$a\sqrt{\varepsilon\beta^{-1}(\varepsilon)} \geq \sqrt{\frac{\beta^{-1}(\varepsilon_k)\mathbb{H}(\delta_{k+1})}{n\varepsilon_k}}\delta_k,$$

thus, it follows from (4.7*a*) that $a\sqrt{\varepsilon\beta^{-1}(\varepsilon)} \geq \delta_k$, hence condition (i) is fulfilled. Therefore

$$\|\Delta_k f\mathbb{1}_{q_k\Delta_k f>b_k}\|_1 \leq 2\delta_k\sqrt{\frac{\varepsilon_k}{\beta^{-1}(\varepsilon_k)}},$$

so, using (4.7*a*) again provides

$$\|\Delta_k f\mathbb{1}_{q_k\Delta_k f>b_k}\|_1 \leq \frac{2\delta_k}{\sqrt{n}}(\mathbb{H}(\delta_{k+1}))^{1/2}.$$

Also, it is usefull to notice that $|g| \leq h$ implies

$$|Z_n(g)| \leq |Z_n(h)| + 2\sqrt{n}\|h\|_1 \tag{4.10}$$

Control of $\mathbb{E}_2$

Since $\{\nu(f) = 0\} = \{q_0\Delta_0 f > b_0\}$, (4.3) and (4.10) yield

$$\mathbb{E}_2 \leq \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[\Delta_0 f \mathbb{1}_{q_0\Delta_0 f > b_0}]|) + 2\sqrt{n}\sup_{f\in\mathcal{F}} \|\Delta_0 f \mathbb{1}_{q_0\Delta_0 f > b_0}\|_1;$$

but (2.1) together with Claim 1 and (4.4) ensure that

$$\|\Delta_0 f \mathbb{1}_{q_0\Delta_0 f > b_0} - E[\Delta_0 f \mathbb{1}_{q_0\Delta_0 f > b_0}]\|_{2,\beta} \leq (1+\theta)\sigma,$$

so applying Lemma 3 with $a = 2M$ and $\delta = (1+\theta)\sigma$, we get for any integer q

$$\mathbb{E}\big(\sup_{f\in\mathcal{F}} |Z_n[\Delta_0 f \mathbb{1}_{q_0\Delta_0 f > b_0}]|\big) \leq C\left((1+\theta)\sigma(H(\sigma))^{1/2} + \frac{2MqH(\sigma)}{\sqrt{n}} + 2\sqrt{n}M\beta_q\right).$$

Hence, using Claim 3 with $k = 0$, for any integer q , we have:

$$\mathbb{E}_2 \leq C\left((1+\theta)\sigma(H(\sigma))^{1/2} + 2^5\frac{Mq\varphi^2(\sigma)}{\sigma^2\sqrt{n}} + 2\sqrt{n}M\beta_q\right) + 4\sigma(\mathbb{H}(\delta_1))^{1/2}.$$

Control of $\mathbb{E}_3$

By (4.3) and (4.10) we have

$$\mathbb{E}_3 \leq \sum_{k=1}^{N-1} \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[\Delta_{k-1} f \mathbb{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}]|) + \sum_{k=1}^{N-1} 2\sqrt{n}\sup_{f\in\mathcal{F}} \|\Delta_{k-1} f \mathbb{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}\|_1.$$

To control the first term of the above inequality we notice that on the one hand (2.1), Claim 1 and (4.4) ensure that

$$\|\Delta_{k-1} f \mathbb{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}} - E[\Delta_{k-1} f \mathbb{1}_{\nu(f)=k, q_k\Delta_k f > b_{k-1}}]\|_{2,\beta} \leq (1+\theta)\delta_{k-1},$$

and on the other hand that $\nu(f) = k$ implies $q_{k-1}\Delta_{k-1}f \leq b_{k-1}$. Hence, applying Lemma 3 with $q = q_{k-1}$, $a = b_{k-1}/q_{k-1}$ and $\delta = (1+\theta)\delta_{k-1}$ and considering (4.5), we get as an upper bound for the first term

$$C\sum_{k=1}^{N-1}(1+\theta)\,\delta_{k-1}(\mathbb{H}(\delta_k))^{1/2} + \frac{2b_{k-1}\mathbb{H}(\delta_k)}{\sqrt{n}},$$

which is in turn bounded via (4.6) by

$$C(5+\theta)\sum_{k=1}^{N-1}\delta_{k-1}(\mathbb{H}(\delta_k))^{1/2}.$$

To bound the second term we note that

$$\{\nu(f) = k, q_k\Delta_k f > b_{k-1}\} \subset \{q_{k-1}\Delta_{k-1}f \leq b_{k-1} < q_k\Delta_k f\}.$$

Hence, since $(q_k)_k$ and $(b_k)_k$ are both nonincreasing sequences

$$||\Delta_{k-1}f\mathbb{1}_{\nu(f)=k,q_k\Delta_k f>b_{k-1}}||_1 \leq ||\Delta_k f\mathbb{1}_{q_k\Delta_k f>b_k}||_1.$$

So, we may now apply Claim 3 and get

$$\mathbb{E}_3 \leq C(5+\theta)\sum_{k=1}^{N-1}\delta_{k-1}(\mathbb{H}(\delta_k))^{1/2} + 4\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_{k+1}))^{1/2}.$$

Control of $\mathbb{E}_4$

By (4.3) and (4.10) we have

$$\mathbb{E}_4 \leq \sum_{k=1}^{N-1}\mathbb{E}(\sup_{f\in\mathcal{F}}|Z_n[\Delta_k f\mathbb{1}_{\nu(f)=k,q_k\Delta_k f\leq b_{k-1}}]|) + \sum_{k=1}^{N-1}2\sqrt{n}\sup_{f\in\mathcal{F}}\ ||\Delta_k f\mathbb{1}_{\nu(f)=k,q_k\Delta_k f\leq b_{k-1}}||_1.$$

Majorizing the first term requires the same arguments as in the control of $\mathbb{E}_3$ above. Namely, we first notice that

$$||\Delta_k f\mathbb{1}_{\nu(f)=k,q_k\Delta_k f\leq b_{k-1}} - E[\Delta_k f\mathbb{1}_{\nu(f)=k,q_k\Delta_k f\leq b_{k-1}}]||_{2,\beta} \leq (1+\theta)\delta_k,$$

and then apply Lemma 3 with with $q = q_k$, $a = b_{k-1}/q_k$ and $\delta = (1+\theta)\delta_k$ and considering (4.5), we get as an upper bound for the first term

$$C\sum_{k=1}^{N-1}(1+\theta)\delta_k(\mathbb{H}(\delta_k))^{1/2} + \frac{b_{k-1}\mathbb{H}(\delta_k)}{\sqrt{n}} + \frac{b_{k-1}\mathbb{H}(\delta_{k+1})}{\sqrt{n}}.$$

Now, by Claim 2, we have $H(\delta_{k+1}) \leq 16H(\delta_k)$, hence $\mathbb{H}(\delta_{k+1}) \leq 17\mathbb{H}(\delta_k)$, so via (4.6) the above upper bound of the first term becomes

$$C(73+\theta)\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_k))^{1/2}.$$

To majorize the second term we note that

$$\{\nu(f) = k, q_k\Delta_k f \leq b_{k-1}\} \subset \{b_k < q_k\Delta_k f\},$$

so that we are back to the same quantity that we have already majorized when controling $\mathbb{E}_3$. Therefore

$$\mathbb{E}_4 \leq C(73+\theta)\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_k))^{1/2} + 4\sum_{k=1}^{N-1}\delta_k(\mathbb{H}(\delta_{k+1}))^{1/2}.$$

Control of $\mathbb{E}_5$

By (4.3) and (4.10) we have

$$\mathbb{E}_5 \leq \mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n[\Delta_{N-1}f\mathbb{I}_{\nu(f)\geq N}]|) + 2\sqrt{n}\sup_{f\in\mathcal{F}} \|\Delta_{N-1}f\|_1.$$

Since $\nu(f) \geq N$ implies that $q_{N-1}\Delta_{N-1}f \leq b_{N-1}$, we can apply Lemma 3 with $q = q_{N-1}$, $a = b_{N-1}/q_{N-1}$ and $\delta = (1+\theta)\delta_{N-1}$, which, considering (4.5), provides the following upper bound for the first term of the above inequality:

$$C(1+\theta)\delta_{N-1}(\mathbb{H}(\delta_{N-1}))^{1/2} + \frac{2b_{N-1}\mathbb{H}(\delta_N)}{\sqrt{n}},$$

which is in turn bounded via (4.6) by $C(5+\theta)\delta_{N-1}(\mathbb{H}(\delta_N))^{1/2}$. It remains to control the second term which is easily done because $\|\Delta_{N-1}f\|_1 \leq$

$\|\Delta_{N-1}f\|_2 \le \|\Delta_{N-1}f\|_{2,\beta} \le \delta_{N-1}$, thus, using the definition of N and Claim 2, we obtain

$$\mathbb{E}_5 \le C(5+\theta)\delta_{N-1}(\mathbb{H}(\delta_N))^{1/2} + 16\varphi(\sigma).$$

Control of $\mathbb{E}_6$

The control of $\mathbb{E}_6$ is performed exactly as the control of the first term of $\mathbb{E}_4$. We note that $\nu(f) \ge k$ implies $q_{k-1}\Delta_{k-1}f \le b_{k-1}$. Because of (2.1), Claim 1, (4.3) and (4.4) we can apply Lemma 3 with $q = q_k$, $a = 4b_{k-1}/q_k$ and $\delta = 3(1+\theta)\delta_k$ which gives

$$\mathbb{E}_6/C \le \sum_{k=1}^{N-1} 3(1+\theta)\delta_k(\mathbb{H}(\delta_k))^{1/2} + \frac{4b_{k-1}\mathbb{H}(\delta_k)}{\sqrt{n}} + \frac{4b_{k-1}\mathbb{H}(\delta_{k+1})}{\sqrt{n}}.$$

Arguing as in the control of $\mathbb{E}_4$ we obtain

$$\mathbb{E}_6/C \le \sum_{k=1}^{N-1}(307+3\theta)\delta_k(\mathbb{H}(\delta_k))^{1/2}.$$

End of the proof of Theorem 2. – Collecting the above inequalities yields for any integer q

$$\begin{aligned}\mathbb{E}(\sup_{f\in\mathcal{F}} |Z_n(f)|) \le & K'\sum_{k\ge 0}\delta_k(\mathbb{H}(\delta_k))^{1/2} \\ & + 16\varphi(\sigma) + 2^6 CM\left(\frac{q\varphi^2(\sigma)}{\sigma^2\sqrt{n}} + \sqrt{n}\beta_q\right),\end{aligned}$$

where $K' = (390+6\theta)C + 16$. To finish the proof, it remains to majorize the series above by the integral of the function H. We have

$$\begin{aligned}\sum_{k\ge 0}\delta_k(\mathbb{H}(\delta_k))^{1/2} &\le \sum_{k\ge 0}\delta_k\Big(\sum_{j\le k}(H(\delta_j))^{1/2}\Big) \\ &\le \sum_{j\ge 0}(H(\delta_j))^{1/2}\Big(\sum_{k\ge j}\delta_k\Big) \\ &\le 2\sum_{j\ge 0}\delta_j(H(\delta_j))^{1/2} \le 4\int_0^\sigma \sqrt{H(t)}\,dt.\end{aligned}$$

Hence, by Claim 2, (4.1) holds with $K = 16(K' + 1)$. Note also that, as a byproduct, we have:

$$\delta_{N+1}\mathbb{H}^{1/2}(\delta_{N+1}) \leq 16\varphi(\sigma) \tag{4.11}$$

Proof of Theorem 3. – Let us denote $\varepsilon(\sigma, n)$ by ε for short, and set $q = \beta^{-1}(\varepsilon)$. Then

$$q \leq \frac{\varepsilon n \sigma^2}{\varphi^2(\sigma)}. \tag{4.12}$$

Choosing $M = Q_F(\varepsilon)$, we now apply Theorem 2 to $\mathcal{F}_M = \{f \mathbb{1}_{F \leq M} : f \in \mathcal{F}\}$. In fact:

$$\mathbb{E}((\sup_{f \in \mathcal{F}} |Z_n(f)|)^*) \leq \mathbb{E}((\sup_{f \in \mathcal{F}_M} |Z_n(f)|)^*) + 2\sqrt{n}\|F\mathbb{1}_{F>M}\|_1.$$

Considering (4.12), the first term is majorized via (4.1) by

$$K(\varphi(\sigma) + 2\sqrt{n}\varepsilon Q_F(\varepsilon)).$$

Now, on the one hand, by definition of ε,

$$2\sqrt{n}\varepsilon Q_F(\varepsilon) \leq \frac{2\varepsilon}{\sqrt{B(\varepsilon)}}\delta_F(\varepsilon) \leq 2\delta_F(\varepsilon)\frac{\varphi(\sigma)}{\sigma}$$

and on the other hand, by Lemma 4,

$$\|F\mathbb{1}_{F>M}\|_1 \leq \frac{2\delta_F(\varepsilon)\varepsilon}{\sqrt{B(\varepsilon)}} \leq \frac{2\varphi(\sigma)\delta_F(\varepsilon)}{\sigma\sqrt{n}}.$$

Hence

$$\mathbb{E}(\sup_{f \in \mathcal{F}} |Z_n(f)|) \leq K\varphi(\sigma) + 2(K+2)\frac{\delta_F(\varepsilon)\varphi(\sigma)}{\sigma},$$

therefore completing the proof of Theorem 3.

5. WEAKENING THE BRACKETING ASSUMPTIONS

In the spirit of Andersen *et al.* (1988), it is possible to weaken the bracketing assumptions and the enveloppe condition. This requires slight modifications of the proof of Theorem 1 that we shall indicate below.

THEOREM 4. – *Let $(\xi_i)_{i\in\mathbb{Z}}$ be a strictly stationary and β-mixing sequence of random variables with common marginal distribution P. Assume that the sequence $(\beta_n)_{n\geq 0}$ satisfies (2.4). Let $\mathcal{F}$ be a class of functions with $\mathcal{F} \subset \mathcal{L}_{2,\beta}(P)$. Let ρ be a pseudo-metric such that $\|f-g\|_{2,\beta} \leq \rho(f,g)$ for all $f, g \in \mathcal{F}$. Let $B_\rho(f,\delta)$ denote the closed ball of center f and radius δ with respect to ρ in $\mathcal{F}$. Assume that $\mathcal{F}$ satisfies to the local bracketing condition*

$$\text{(5.1)} \qquad for\ any\ function\ f \in \mathcal{F}, \Lambda_{2,\beta}\Big(\sup_{g\in B_\rho(f,\delta)} |f-g|\Big) \leq c\,\delta,$$

where c is a positive constant. Assume furthermore that the metric entropy function of $\mathcal{F}$ with respect to ρ satisfies to the integrability criterion

$$\text{(5.2)} \qquad \int_0^1 \sqrt{H(u,\mathcal{F},\rho)}du < +\infty.$$

Finally, suppose that some positive function F with $|f| \leq F$ for any $f \in \mathcal{F}$ fulfills the following condition

$$\text{(5.3)} \qquad M^2 B(P(F > M))\ tends\ to\ 0 \text{ as } M\ goes\ to\ infinity.$$

Then Z_n converges weakly to some Gaussian process indexed by $\mathcal{F}$ with a.s. uniformly continuous sample paths.

Comments. – Of course Theorem 4 implies Theorem 1. The reason why we proved Theorem 1 first is that we hope that it makes the proofs more readable.

As an application of Theorem 4 we can provide a CLT in $D[0,1]$. In the i.i.d. case this example was pointed out by Andersen *et al.* as a typical situation where weak bracketing is needed. Let $\{X(t)\}_{t\in[0,1]}$ be a centered uniformly bounded stochastic process with sample paths in $D[0,1]$. Assume that $\sup_{t,\omega} |X(t,\omega)| \leq 1$ a.s. and

$$E|X(s)-X(t)| \leq K|s-t|, \qquad s,\ t \in [0,1],$$

for some positive constant K. Following the lines of Andersen *et al.* (example 4.8) we have, for all fixed t in $[0,1]$

$$E[\sup_{|t-s|\leq\varepsilon} |X(s)-X(t)|] \leq 2K\varepsilon.$$

On the one hand, by Lemma 7, sect. 6, $\|(X(s) - X(t))/2\|_{2,\beta} \leq \sqrt{B(K|s-t|/2)}$. On the other hand it is straightforward to verify that $\Lambda_{2,\beta}(Y) \leq \sqrt{B(E(Y))}$ for any $[0,1]$-valued random variable Y. As a consequence we get that, defining $\rho(\delta_s, \delta_t) = 2\sqrt{B(K|s-t|)}$ for all fixed t in $[0,1]$

$$\Lambda_{2,\beta}(\sup_{\rho(\delta_s,\delta_t)\leq\varepsilon} |X(s) - X(t)|) \leq \varepsilon.$$

Hence (5.1) holds (note that since $\sqrt{B}$ is an increasing and concave function which is null at zero, ρ is a pseudo-metric on $\mathcal{F} = \{\delta_t, t \in [0,1]\}$). As a conclusion we obtain that a strictly stationary sequence $(X_n)_{n\in\mathbb{Z}}$ of stochastic processes with the same distribution as $\{X(t)\}_{t\in[0,1]}$ and with mixing coefficients (β_n) has the CLT property whenever the sequence (β_n) satisfies (2.18) [which ensures that (5.2) holds].

Proof of Theorem 4. – In order to prove Theorem 4 all we have to do is to show that the conclusion of Theorem 2 holds true when substituting (5.1) and (5.2) to (2.10). If so, the inequality of Theorem 3 remains valid with $\delta_F(\varepsilon) \longrightarrow 0$ as $\varepsilon \to 0$ because of assumption (5.3). This implies the tightness of the empirical process. The existence of a uniformly continuous version of the limiting Gaussian process is ensured by (5.2) and the fact that ρ dominates d_Γ. Hence Theorem 4 follows. We set

$$\varphi(\sigma) = \int_0^\sigma \sqrt{H(u, \mathcal{F}_\sigma, \rho)}du.$$

Let $H(\cdot)$ be a function with the properties described in Claim 2 except that $H(\cdot)$ majorizes $H(\cdot, \mathcal{F}_\sigma, \rho)$ instead of $H_\beta(\cdot, \mathcal{F}_\sigma)$.

We want to prove that (2.18) holds. The proof is essentially the same as that of Theorem 2, so we will only sketch it. Let $\delta_k = 2^{-k}\sigma$ and Π_k be a projection of $\mathcal{F}_\sigma$ on a δ_k-net (with respect to the pseudo-metric ρ). Let $\Delta_k f = \sup_{g\in B_\rho(\Pi_k f, \delta_k)} |g - \Pi_k f|$, then $|f - \Pi_k f| \leq \Delta_k f$ and $\Lambda_{2,\beta}(\Delta_k f) \leq c\,\delta_k$. The main difficulty that we have to overcome is that at some stages of the proof of Theorem 2, we used a maximal inequality (Lemma 3) involving a control of the $\mathcal{L}_{2,\beta}$-norm of truncated variables defined from the Δ_k's. In fact what happens is that we can show that for these truncated variables $\|\cdot\|_{2,\beta}$ and $\Lambda_{2,\beta}$ are of the same order. This will be done via the following elementary lemma.

LEMMA 5. – *Let Y be a nonnegative random variable bounded by some constant a. Let Z be a nonnegative random variable and a' be some fixed* *positive constant. Then*

$$\|Y \mathbb{1}_{Z>a'}\|_{2,\beta} \leq \frac{a}{a'} \Lambda_{2,\beta}(Z). \tag{5.4}$$

Proof of Lemma 5. – Noting that $\|\mathbb{1}_{Z>a'}\|_{2,\beta} = \sqrt{B(P(Z > a'))}$ we have by (1.3)

$$\|Y \mathbb{1}_{Z>a'}\|_{2,\beta} \leq a\|\mathbb{1}_{Z>a'}\|_{2,\beta} \leq a\sqrt{B(P(Z > a'))},$$

hence (5.4) follows by definition of $\Lambda_{2,\beta}$.

The notations being the same as in the proof of Theorem 2, the controls of $\mathbb{E}_1$ and $\mathbb{E}_6$ are not modified. In the same way, since Lemma 4 [inequality (*b*)] provides an upper bound involving $\Lambda_{2,\beta}$, we do not have to modify the majorization of the quantities $\|\Delta_k f \mathbb{1}_{\nu(f)=k, q_k \Delta_k f > b_k}\|_1$ appearing in the controls of $\mathbb{E}_i$, $i = 2, 3, 4$. The corresponding term $\|\Delta_{N-1} f\|_1$ appearing in the control of $\mathbb{E}_5$ is bounded by $2c\,\delta_{N-1}$ via Lemma 4, inequality (*a*), with $\varepsilon = 1$. To finish the control of $\mathbb{E}_2$ we use Lemma 5 with $Y = Z = \Delta_0 f$, $a = 2M$ and $a' = b_0/q_0$, thus

$$\|\Delta_0 f \mathbb{1}_{q_0 \Delta_0 f > b_0} - E[\Delta_0 f \mathbb{1}_{q_0 \Delta_0 f > b_0}]\|_{2,\beta} \leq 2(1+\theta) M q_0 c \sqrt{\frac{\mathbb{H}(\delta_1)}{n}}.$$

It follows from the definition of q_0 that $q_0 \leq q + n\beta_q/\mathbb{H}(\delta_1)$ for any positive integer q. Hence we obtain the same bound for $\mathbb{E}_2$ (with different constants of course) as in the proof of Theorem 2. To modify the control of $\mathbb{E}_3$, we note that, using Lemma 5 with $Y = \Delta_{k-1} f \mathbb{1}_{\nu(f)=k}$, $Z = \Delta_k f$, $a = b_{k-1}/q_{k-1}$ and $a' = b_{k-1}/q_k$

$$\begin{aligned} &\|\Delta_{k-1} f \mathbb{1}_{\nu(f)=k, q_k \Delta_k f > b_{k-1}} - E[\Delta_{k-1} f \mathbb{1}_{\nu(f)=k, q_k \Delta_k f > b_{k-1}}]\|_{2,\beta} \\ &\quad \leq c(1+\theta)\delta_k \frac{q_k}{q_{k-1}}. \end{aligned}$$

Since (q_k) is nonincreasing, we can finish the control of $\mathbb{E}_3$ as before. To modify the control of $\mathbb{E}_4$, we note that, using Lemma 5 with $Y = \Delta_k f \mathbb{1}_{q_k \Delta_k \leq b_{k-1}, \nu(f)=k}$, $Z = \Delta_k f$, $a = b_{k-1}/q_k$ and $a' = b_k/q_k$

$$\begin{aligned} &\|\Delta_k f \mathbb{1}_{\nu(f)=k, q_k \Delta_k f \leq b_{k-1}} - E[\Delta_k f \mathbb{1}_{\nu(f)=k, q_k \Delta_k f \leq b_{k-1}}]\|_{2,\beta} \\ &\quad \leq c(1+\theta)\delta_k \frac{b_{k-1}}{b_k}. \end{aligned}$$

But b_{k-1}/b_k is bounded by $2\sqrt{17}$ so that the control of $\mathbb{E}_4$ may be ended as before. Finally in order to control $\mathbb{E}_5$, we simply notice that

$$\|\Delta_{N-1} f \mathbb{1}_{q_{N-1} \Delta_{N-1} f \leq b_{N-1}}\|_{2,\beta} \leq \theta \frac{b_{N-1}}{q_{N-1}} \leq 2\theta\, \delta_{N-1} \sqrt{\frac{n}{\mathbb{H}(\delta_N)}},$$

which allows to end the majorization of $\mathbb{E}_5$ via Claim 2, using Lemma 3 and the definition of N. Thus an analogue of Theorem 2 is obtained and Theorem 4 follows.

6. APPENDIX

A. Properties of the space $\mathcal{L}_{2,\beta}(P)$

1. *Proof of Lemma 1.* – In order to prove Lemma 1, we show that $\|f\|_{2,\beta}$ can be viewed as the supremum of ℓ^2-type norms, via a classical result of Fréchet (1951, 1957).

Fréchet's result. – Let F and G be the d.f.'s of positive real-valued random variables and $\mathcal{L}(F,G)$ denote the class of bivariate r.v's (X,Y) with given marginal d.f.'s F and G. Fréchet (1951, 1957) proved that, the maximum of $\mathbb{E}(XY)$ over $\mathcal{L}(F,G)$ is obtained when $(X,Y) = (F^{-1}(U), G^{-1}(U))$, where U is uniformly distributed over $[0,1]$ [*see* Bártfai (1970) for a detailed proof]. In other words, we have:

$$\sup_{(X,Y)\,\in\mathcal{L}(F,G)} \mathbb{E}(XY) = \int_0^1 F^{-1}(u)G^{-1}(u)du \tag{6.1}$$

Let G_β be the distribution function on $\mathbb{N}$ defined by: $G_\beta(n) = 1 - \beta_n$, for any $n \in \mathbb{N}$ [note that $G_\beta(0) = 0$]. Clearly, $G_\beta^{-1}(u) = \beta^{-1}(1-u)$. So, noticing that, for any $f \in \mathcal{L}_{2,\beta}(P)$, the inverse of the d.f. of $f^2(\xi_0)$ is exactly $u \to Q_f^2(1-u)$ and applying (6.1) and Skorohod's Lemma (1976), we obtain:

$$\|f\|_{2,\beta} = \sup_{b\in\mathcal{L}(\beta)} \sqrt{\mathbb{E}(bf^2(\xi_0))}. \tag{6.2}$$

where $\mathcal{L}(\beta)$ denotes the class of integer valued random variables b on $(\Omega, \mathcal{T}, \mathbb{P})$ with d.f. G_β. Now, let

$$\|f\|_{2,b} = \sqrt{\mathbb{E}(bf^2(\xi_0))} \tag{6.3}$$

be the Hilbertian norm associated with the r.v. b. Since $\|f\|_{2,\beta}$ is the maximum of the norms $\|f\|_{2,b}$ over the class $\mathcal{L}(\beta)$, $f \to \|f\|_{2,\beta}$ is a norm.

Since $\|.\|_{2,b}$ is a ℓ^2-type norm, $|f| \le |g|$ implies $\|f\|_{2,b} \le \|g\|_{2,b}$. Hence, by (6.2), $\|.\|_{2,\beta}$ verifies (2.1). Moreover, since $b \ge 1$ a.s., for any f in $\mathcal{L}_{2,\beta}(P)$, $\|f\|_2 \le \|f\|_{2,\beta}$, therefore completing the proof.

2. *Comparisons with the Orlicz norms.* – In this part, we prove Lemma 2. Recall that, if U has the uniform distribution over $[0,1]$, $Q_f(U)$ has the same distribution as $|f(\xi_0)|$. Hence, for any f in $\mathcal{L}_{\phi,2}(P)$,

$$\|f\|_{\phi,2} = \inf\left\{c > 0 : \int_0^1 \phi((Q_f(u)/c)^2)du \le 1\right\}.$$

Hence, for any $c > \|f\|_{\phi,2}$, Young's inequality $xy/c^2 \leq \phi^*(x) + \phi(y/c^2)$ ensures that

$$\frac{\|f\|_{2,\beta}^2}{c^2} = \int_0^1 \beta^{-1}(u)[Q_f(u)/c]^2 du \leq 1 + \int_0^1 \phi^*(\beta^{-1}(u))\, du\,,$$

therefore establishing (*a*) of Lemma 2.

Now $[Q_f(u)]^2 \leq \phi^{-1}(1/u)[\Lambda_\phi(f)]^2$, which implies (*b*) of Lemma 2.

B. A monotonic saturation Lemma

In this section, we establish the following lemma, which implies Claim 2 of section 4.

LEMMA 6. – *Let* $G :]0,1]\times]0,+\infty[\to \mathbb{R}^+$ *be a function fulfilling the following conditions: for any positive* y, $x \to G(x,y)$ *is nonincreasing, and, for any* x *in* $]0,1]$, $y \to (G(x,y)/y)$ *is nondecreasing. Let* ψ: $]0,1] \to \mathbb{R}^+$ *be a nonincreasing function. Suppose that*

$$\text{(i)} \qquad \int_0^1 G(x,\psi(x))dx < +\infty.$$

Then, for any $a > 1$, *there exists a nonincreasing mapping* $\psi_a :]0,1] \to \mathbb{R}^+$ *such that* $x \to x^a\psi_a(x)$ *is nondecreasing,* $\psi_a \geq \psi$ *and, for any* ε *in* $]0,1]$,

$$\text{(ii)} \qquad \int_0^\varepsilon G(x,\psi_a(x))dx \leq 2(1-2^{1-a})^{-1}\int_0^\varepsilon G(x,\psi(x))dx.$$

Proof of Lemma 6. – We proceed exactly as in the proof of Claim 1 in Doukhan, Massart and Rio (1994). Let $\psi_a(x) = \sup_{u\leq x} (u/x)^a\psi(u)$. Clearly, ψ_a is a nonincreasing function, $\psi_a \geq \psi$ and $x \to x^a\psi(x)$ is nondecreasing. The monotonicity properties of the above functions imply that

$$\text{(6.4)} \qquad \psi_a(x) \leq 2^{-a}\psi_a(x/2) \vee \psi(x/2).$$

Assume now that ψ is uniformly bounded over $]0,1]$. Then, both (6.4) and the monotonicity properties of F and ψ ensure that

$$\text{(6.5)} \qquad \int_0^\varepsilon G(x,\psi_a(x))dx$$
$$\leq \int_0^\varepsilon G(x/2, 2^{-a}\psi_a(x/2))dx + \int_0^\varepsilon G(x/2,\psi(x/2))dx$$
$$\leq 2^{1-a}\int_0^\varepsilon G(x,\psi_a(x))dx + 2\int_0^\varepsilon G(x,\psi(x))dx,$$

therefore establishing Lemma 6 for uniformly bounded functions ψ. The corresponding result for unbounded functions follows from (6.5) applied to $\psi^A(t) = \psi(t) \wedge A$ and from the fact that $\psi_a^A = \lim_{A \nearrow \infty} \uparrow \psi_a^A$ combined with Beppo-Levi lemma.

C. Inequalities involving $\mathcal{L}_\infty$-type norms

Let F be some positive element of $\mathcal{L}_{2,\beta}(P)$, and g be any function in $\mathcal{L}_\infty(P)$. In this subsection, we compare the $\mathcal{L}_{2,\beta}$-norm of gF with the $\mathcal{L}_{2,\beta}$-norm of F. Here the following lemma yields an upper bound on $\|gF\|_{2,\beta}$, which is used to derive application 4 of Theorem 1.

LEMMA 7. – *Let F be some element of $\mathcal{L}_{2,\beta}(P)$ satisfying $F \geq 1$, and let*

$$R(t) = \sqrt{\int_0^t \beta^{-1}(u)[Q_F(u)]^2 du}.$$

Then, for any g in $\mathcal{L}_\infty(P)$, $g \neq 0$,

$$\|gF\|_{2,\beta} \leq \|g\|_\infty R(\varepsilon) \leq \|g\|_\infty \|F\|_{2,\beta},$$

where ε is the real number in $[0, 1]$ defined by the equation

$$\int_0^\varepsilon [Q_F(u)]^2 du = \left[\frac{\|gF\|_2}{\|g\|_\infty}\right]^2.$$

Proof of Lemma 7. – Clearly, there is no loss of generality in assuming that $\|g\|_\infty = 1$. Let $f = gF$. We set $\delta = \|f\|_2$. Since $|f| \leq F$, $Q_f \leq Q_F$. Hence the elementary equality

$$\|f\|_{2,\beta}^2 = \sum_{n \geq 0} \int_0^{\beta_n} [Q_f(u)]^2 du$$

implies that

$$\|f\|_{2,\beta}^2 \leq \sum_{n \geq 0} \delta^2 \wedge \int_0^{\beta_n} [Q_F(u)]^2 du = \sum_{n \geq 0} \int_0^{\beta_n \wedge \varepsilon} [Q_F(u)]^2 du = R^2(\varepsilon),$$

which establishes Lemma 7.

Assume now that the envelope function of $\mathcal{F} \subset \mathcal{L}_2(P)$ is in $\mathcal{L}_{2,\beta}(P)$. Then there exists some measurable function F such that

$$F \geq 1, \quad F \in \mathcal{L}_{2,\beta}(P), \quad \text{and, for any } f \in \mathcal{F}, \quad |f| \leq F.$$

By Lemma 7, condition (2.10) in Theorem 1 holds as soon as

$$\int_0^1 H_2^{1/2}\left(\sqrt{\int_0^u [Q_F(t)]^2 dt}\right) dR(u) < +\infty. \tag{6.6}$$

In the general case, this integral condition leads to a very intricated summability condition on the mixing coefficients. However this condition is much more tractable if the envelope function F is in $\Lambda_{2r}(P)$. In this case, we may, by increasing Q_F if necessary, assume that $Q_F(u) = cu^{-1/(2r)}$, and (6.6) implies the criterions (2.15), (2.16) and (2.17).

REFERENCES

N. T. Andersen, E. Giné, M. Ossiander and J. Zinn, The central limit theorem and the law of iterated logarithm for empirical processes under local conditions, *Probab. Th. Rel. Fields*, Vol. **77**, 1988, pp. 271-305.

W. K. Andrews and D. Pollard, An introduction to functional central limit theorems for dependent stochastic processes, *Inst. Stat. review*, Vol. **62**, 1994, pp. 119-132.

M. A. Arcones and B. Yu, Central limit theorems for empirical and U-processes of stationary mixing sequences, J. *Theoret. Prob.*, Vol. **7**, 1994, pp. 47-71.

P. Bártfai, Über die Entfernung de Irrfahrtswege, *Studia Sci. Math. Hungar*, Vol. **1**, 1970, pp. 161-168.

R. F. Bass, Law of the iterated logarithm for set-indexed partial sum processes with finite variance, *Z. Warsch. verw. Gebiete*, Vol. **70**, 1985, pp. 591-608.

H. C. P. Berbee, *Random walks with stationary increments and renewal theory*, Math. Cent. Tracts, Amsterdam, 1979.

I. Berkes and W. Phillip, An almost sure invariance principle for the empirical distribution of mixing random variables, *Z. Wahrsch. Verw. Gebiete*, Vol. **41**, 1977, pp. 115-137.

E. Bolthausen, Weak convergence of an empirical process indexed by the closed convex subsets of I^2, *Z. Wahrsch. Verw. Gebiete*, Vol. **43**, 1978, pp. 173-181.

R. C. Bradley, Basic properties of strong mixing conditions, in *Dependence in probability and statistics a survey of recent results*, Oberwolfach 1985, Birkhäuser, 1986.

Yu. A. Davydov, Mixing conditions for Markov chains, *Theory Probab. Appl.*, Vol. **28**, 1973, pp. 313-328.

S. Dhompongsa, A note on the almost sure approximation of empirical process of weakly dependent random vectors, *Yokohama math. J.*, Vol. **32**, 1984, pp. 113-121.

M. Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov's theorems, *Ann. Math. Stat.*, Vol. **23**, 1952, pp. 277-281.

J. L. Doob, *Stochastic Processes*, Wiley, New-York, 1953.

P. Doukhan, Mixing: properties and examples, *Lecture notes in Statistics* 85, Springer, 1994.

P. Doukhan, J. León and F. Portal, Principe d'invariance faible pour la mesure empirique d'une suite de variables aléatoires dépendantes, *Probab. Th. rel. fields*, Vol. **76**, 1987, pp. 51-70.

P. Doukhan, P. Massart and E. Rio, The functional central limit theorem for strongly mixing processes, *Ann. Inst. H. Poincaré, Probab. Stat.*, Vol. **30**, 1994, pp. 63-82.

R. M. Dudley, Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces, *Illinois J. Math.*, Vol. **10**, 1966, pp. 109-126.

R. M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, *J. Functional Analysis*, Vol. **1**, 1967, pp. 290-330.

R. M. Dudley, Central limit theorems for empirical measures, *Ann. Probab.*, Vol. **6**, 1978, pp. 899-929.

R. M. Dudley, A course on empirical processes. Ecole d'été de probabilités de Saint-Flour XII-1982. *Lectures Notes in Math.*, Vol. **1097**, Springer, Berlin, 1984, pp. 1-142.

M. Fréchet, Sur les tableaux de corrélation dont les marges sont données, *Annales de l'université de Lyon, Sciences, section A*, Vol. **14**, 1951, pp. 53-77.

M. Fréchet, Sur la distance de deux lois de probabilité, *C. R. Acad. Sci. Paris*, Vol. **244**, No. 6, 1957, pp. 689-692.

E. Giné and J. Zinn, Some limit theorems for empirical processes, *Ann. Probab.*, Vol. **12**, 1984, pp. 929-989.

N. Herrndorf, A functional central limit theorem for strongly mixing sequences of random variables, *Z. Wahr. Verv. Gebiete*, Vol. **69**, 1985, pp. 541-550.

I. A. Ibragimov, Some limit theorems for stationary processes, *Theory Probab. Appl.*, Vol. **7**, 1962, pp. 349-382.

V. I. Kolchinskii, On the central limit theorem for empirical measures (In Russian), *Teor. vero. i. mat. stat. (Kiev)*, Vol. **24**, 1981, pp. 63-75.

A. N. Kolmogorov and Y. A. Rozanov, On the strong mixing conditions for stationary gaussian sequences, *Theory Probab. Appl.*, Vol. **5**, 1960, pp. 204-207.

P. Massart, *Invariance principles for empirical processes: the weakly dependent case*, Quelques problèmes de vitesse de convergence pour des mesures empiriques. Thèse d'Etat, Université de Paris-Sud, 1987.

A. Mokkadem, Propriétés de mélange des modèles autorégressifs polynomiaux, *Ann. Inst. Henri Poincaré, Probab. Stat.*, Vol. **26**, 1990, pp. 219-260.

M. Ossiander, A central limit theorem under metric entropy with L_2-bracketing, *Ann. Probab.*, Vol. **15**, 1987, pp. 897-919.

W. Philipp and L. Pinzur, Almost sure approximation theorems for the multivariate empirical processes, *Z. Wahr. Verv. Gebiete*, Ser. A, Vol. **54**, 1980, pp. 1-13.

D. Pollard, A central limit theorems for empirical processes, *J. Aust. Math. Soc.*, Vol. **33**, 1982, pp. 235-248.

D. Pollard, *Convergence of stochastic processes*, Springer, Berlin, 1984.

D. Pollard, *Empirical processes: theory and applications*, NSF-CBMS Regional Conference Series in Probability and Statistics IMS-ASA, Hayward-Alexandria, 1990.

P. Révész, Three theorems of multivariate empirical process. *Lectures Notes in Math.*, Vol. **566**, Springer, Berlin, 1976, pp. 106-126.

E. Rio, Covariance inequalities for strongly mixing processes, *Ann. Int. H. Poincaré, Prob. Stat.*, Vol. **29**, 1993, pp. 587-597.

M. Rosenblatt, A central limit theorem and a strong mixing condition, *Proc. Nat. Ac. Sc. U.S.A.*, Vol. **42**, 1956, pp. 43-47.

A. V. Skorohod, On a representation of random variables, *Theory Probab. Appl.*, Vol. **21**, 1976, pp. 628-632.

T. G. Sun, *Ph. D. dissertation*, Dept of Mathematics, Univ. of Washington, Seattle, 1976.

M. Talagrand, Regularity of Gaussian processes, *Acta Math.*, Vol. **159**, 1987, pp. 99-149.

V. A. Volkonskii and Y. A. Rozanov, Some limit theorems for random functions, Part I, *Theory Probab. Appl.*, Vol. **4**, 1959, pp. 178-197.

K. Yoshihara, Note on an almost sure invariance principle for some empirical processes, *Yokohama math. J.*, Vol. **27**, 1979, pp. 105-110.

(Manuscript received November 22, 1993.)