EMMANUEL RIO

Approximation forte de processus de sommes partielles indexés par des ensembles, 1990

63834

ORSAY
n° d'ordre :

UNIVERSITE PARIS SUD

CENTRE d'ORSAY

THESE

présentée pour obtenir

Le TITRE de DOCTEUR EN SCIENCES

SPECIALITE : MATHEMATIQUES

par

Emmanuel RIO

Sujet : APPROXIMATION FORTE DE PROCESSUS DE SOMMES PARTIELLES INDEXES PAR DES ENSEMBLES.

Soutenue le 15 Février 1990 devant le Jury composé de :

MM.	BRETAGNOLLE	Jean	Président
	ASSOUAD	Patrice	
	AZENCOTT	Robert	
	DEHEUVELS	Paul	
	MASSART	Pascal	

Code matière AMS (1980) : 60 F 17 - 62 G 99.

ABSTRACT

Given a class $\mathcal{S}$ of subsets of $[0,1]^d$ and an array $(X_i)_{i\in\mathbb{Z}_+^d}$ of independent identically distributed $\mathbb{R}^k$-valued random vectors with mean zero and with finite second moment, the (unsmoothed) partial-sum process X_ν is defined by $X_\nu(S) = \nu^{-d/2}\sum_{i\in\nu S} X_i$, $S \in \mathcal{S}$. If $\mathcal{S}$ is not too large (e.g. either $\mathcal{S}$ is a Vapnik-Chervonenkis class or $\mathcal{S}$ has an integrable metric entopy with inclusion), it is known since Alexander's and Pyke's works that a possibly smoothed version of the partial-sum process X_ν converges uniformly over $\mathcal{S}$ to the Brownian process indexed by $\mathcal{S}$.

In the way opened by Komlós, Major, and Tusnády (1975 and 1976), we developp a method of approximation which allows us to prove strong invariance principles for a (possibly smoothed) version of the partial-sum process X_ν with some rate a_ν depending only on some geometrical characteristics of $\mathcal{S}$, and on the tail distribution of X_1.

For instance, when $\mathcal{S}$ is the Vapnik-Chervonenkis class of Euclidean balls, and when the random vectors X_i have a finite r-th moment for some $r > 2$, a_ν may be taken as $\nu^{-1/2}(\log\nu)^{1/2}+\nu^{-d/2}\nu^{d/r}$ a.s.; when $\mathcal{S}$ is the class of subsets of $[0,1]^d$ with α-differentiable boundaries introduced in Dudley (74), when $\alpha > d-1$, a_ν may be taken as $\nu^{(d-1-\alpha)/2\alpha} + \nu^{-d/2}\nu^{d/r}$. In the particular case when $d = 2$ and $\mathcal{S}$ is the class of lower-left orthants, we show that a_ν may be taken as $\nu^{1-2/r}(\log\nu)^2$ a.s.. Moreover, starting from Beck's works (85 and 87), we prove that any of the above rates of approximation is optimal, up to a possible power of $\log\nu$.

AMS 1980 subject classifications.
Primary 60 F 17; Secondary 62 G 99

Key words and phrases.
Partial-sum process, Invariance principle, Strong approximation, Vapnik-Chervonenkis class, Metric entropy , Quantile transformations, Gaussian processes.

Remerciements.

Je veux tout d'abord remercier Pascal Massart, qui a bien voulu me diriger, et qui n'a ménagé ni son temps ni ses efforts pour me conseiller et pour relire mes notes afin de rendre présentable ce travail.

Je tiens ensuite à remercier Jean Bretagnolle, qui a accepté de rapporter cette thèse, pour avoir suggéré plusieurs simplifications conséquentes et pour avoir relu attentivement un manuscrit émaillé de nombreuses fautes. Qu'il soit aussi remercié pour avoir accepté présider le jury de cette thèse.

Que Paul Deheuvels soit remercié pour avoir accepté de rapporter cette thèse et pour les nombreuses corrections qu'il a suggérées ainsi que pour avoir accepté de faire partie de mon jury.

Je remercie également Patrice Assouad qui, par ses remarques au cours de diverses discussions, m'a aidé dans ce travail. Qu'il soit aussi remercié pour avoir accepté de faire partie de mon jury ainsi que Robert Azencott.

Ma gratitude va également à tous ceux qui, par leurs remarques et leurs conseils m'ont soutenus. Citons en particulier Christian Léonard et Francis Comets.

Enfin, je remercie Mesdames Nicole Parvan et Anne Marie Baillet pour avoir assuré la frappe de la première partie de ce manuscrit.

TABLE DES MATIERES

INTRODUCTION

A quoi nous intéressons-nous?

Soit $(X_i)_{i\in\mathbb{Z}_+^d}$ un champ de variables aléatoires indépendantes et équidistribuées (i.i.d. en abrégé) à valeurs dans $\mathbb{R}^k$, centrées et ayant un moment d'ordre 2 fini. δ_x désignant ici la mesure de Dirac au point x, on définit, pour tout ν de $\mathbb{Z}_+$, la mesure aléatoire normalisée associée X_ν par:

$$X_\nu = \nu^{-d/2} \sum_{i\in]0,\nu]^d} \delta_{i/\nu}$$

On considère une classe $\mathcal{S}$ de parties du cube unité $I^d = [0,1]^d$ et on nomme processus de sommes partielles relatif à la classe $\mathcal{S}$ et au champ X le processus $\{X_\nu(S) : S \in \mathcal{S}\}$. Notre but est de prouver que, si $\mathcal{S}$ satisfait à des propriétés entropiques et géométriques convenables, alors, pourvu que le champ X soit défini sur un espace de probabilité assez grand, il existe un champ $(Y_i)_{i\in\mathbb{Z}_+^d}$ de vecteurs aléatoires Gaussiens i.i.d. tel que, si Y_ν désigne la mesure aléatoire normalisée associée au champ Y,

$$\sup_{S\in\mathcal{S}} \mid X_\nu(S) - Y_\nu(S) \mid = O(a_\nu) \quad p.s. \tag{0.1}$$

a_ν étant une vitesse dépendant des caractéristiques entropiques et géométriques de $\mathcal{S}$ et de la loi des variables X_i, les mesures X_ν et Y_ν ayant été préalablement régularisées pour certaines classes. Un résultat de ce type est appelé principe d'invariance fort. Le cas $a_\nu = o(1)$ permet d'obtenir un théorème de la limite centrale uniforme. Mais le cas $a_\nu = o(\sqrt{\log\log\nu})$ (le résultat ci-dessus est alors appelé principe d'invariance fort au sens de Strassen) est tout aussi intéressant, puisqu'il permet de déduire une loi du logarithme itéré fonctionnelle pour le processus X_ν à partir de celle valant pour le processus Gaussien Y_ν. Qui plus est, nous nous proposons d'établir des vitesses optimales, ce qui permet une étude plus précise du processus X_ν indexé par la classe $\mathcal{S}$: donnons maintenant quelques exemples d'applications possibles de tels résultats.

En dimension 1, si S est la classe $\{]0,t] : 0 < t \leq 1\}$, si $S_\nu = \sum_1^\nu X_i$, 0.1 est équivalent à l'existence d'une suite $(Y_i)_{i>0}$ de Gaussiennes centrées indépendantes et équivariantes telle que, si $T_\nu = \sum_1^\nu Y_i$, alors

$$|S_\nu - T_\nu| = o(a_\nu \sqrt{\nu}) \quad a.s.$$

D'après un résultat de Chung (48) (voir Breiman (67) à ce sujet), on sait que, si les variables Gaussiennes Y_i sont réelles, centrées et réduites, alors,

$$\liminf_{\nu \to +\infty} \nu^{-1/2} (\log \log \nu)^{1/2} \max_{k \leq \nu} T_k = \frac{\pi}{\sqrt{8}} \quad p.s.$$

L'analogue de ce résultat pour les sommes partielles de variables i.i.d. non Gaussiennes est difficile à démontrer directement. Mais, si le principe d'invariance vaut avec la vitesse $a_\nu = o((\log \log \nu)^{-1/2})$, alors ce résultat s'obtient immédiatement à partir du résultat correspondant pour les Gaussiennes. Quand la vitesse est du même ordre de grandeur, certains autres résultats, comme le théorème de Darling-Erdõs peuvent s'obtenir pareillement (voir à ce sujet l'article de U. Einmahl, 89).

Quand le terme d'erreur obtenu est plus petit, on peut, à partir de l'étude des accroissements du mouvement Brownien, en déduire des résultats sur les accroissements de la marche aléatoire non gaussiennes. En effet, tout résultat sur les accroissemnts du mouvement Brownien faisant intervenir des constantes de renormalisation supérieures à $\nu^{-1/2} a_\nu$ (en dimension 1) donne immédiatement le résultat correspondant pour les sommes partielles. Les lois ou les constantes limites sont alors indépendantes de la loi des variables X_i. A ce sujet, on pourra regarder le livre de Csörgó et Révész (81), les articles de Hanson et Russo (83 et 89) et la note de Shao (89). Pour étudier les petits accroissements de la marche aléatoire non gaussienne, il est donc intéressant d'obtenir des bonnes vitesses. La généralisation naturelle de ce problème aux processus multiindicés est alors l'étude du comportement asymptotique des petites oscillations du processus de sommes partielles X_ν indexé par la classe S.

Rappelons les principaux résultats sur les principes d'invariance forts connus à ce jour.

Résultats antérieurs.

Nous citons ici des résultats immédiatement comparables à ceux que nous obtiendrons dans cette thèse. Pour des résultats plus généraux, voir Pyke (84). En dimension 1, le cas $S = \{[0,t] : 0 < t \leq 1\}$ est presque entièrement élucidé. Si les variables X_i sont réelles et ont une transformée de Fourier-Laplace finie dans un voisinage de l'origine, alors Komlós, Major et Tusnády (K.M.T., I et II, 75 et 76) ont montré que 0.1 vaut avec $\nu^{1/2}a_\nu = O(\log \nu)$. Cette vitesse est optimale d'après un travail de Bártfai, repris sous une forme plus générale par Erdös et Rényi en 1970 (voir Csörgó et Révész, 81). Dans leur deuxième article, K.M.T. montrent que, si les variables X_i sont réelles et ont un moment d'ordre r supérieur à 2 au sens strict (i.e. $E(|X_1|^r) < +\infty$), alors 0.1 vaut avec $\nu^{1/2}a_\nu = o(\nu^{1/r})$. Ce résultat est optimal (voir Breiman, 67).

Ces résultats ont été généralisés par A.I. Sakhanenko (84) aux sommes de variables aléatoires réelles indépendantes non équidistribuées. Puis U. Einmahl (89) a étendu les résultats de K.M.T. aux sommes de vecteurs aléatoires indépendants, équivariants et à valeurs dans $\mathbb{R}^k$. Dans un article précédent, U. Einmahl étudiait l'approximation des sommes de vecteurs i.i.d. sous des conditions de moment plus générales: résumons les résultats obtenus par U. Einmahl.

Soit ψ une fonction continue strictement croissante de $\mathbb{R}_+$ dans lui-même. telle que $x \to x^{-2}\psi(x)$ soit croissante et telle qu'il existe un réel positif r tel que $x \to x^{-r}\psi(x)$ soit décroissante. On dira que la loi Q a un ψ-moment si:

$$\int_{\mathbb{R}^k} \psi(|x|)\, dQ(x) < +\infty$$

Quand les vecteurs X_i ont un ψ-moment et sont centrées, alors on peut constuire une suite $(Y_i)_{i>0}$ de vecteurs Gaussiens i.i.d. de loi $N(0, Var\, X_1)$ telle que:

$$\sum_{i \leq \nu} (X_i - \sigma_i Y_i) = o(\psi^{-1}(\nu)) \quad p.s.$$

$(\sigma_i)_{i>0}$ étant une suite de réels positifs telle que $1 - \sigma_i = o(\nu^{-1}(\psi^{-1}(\nu))^2)$. A partir de là, il est facile de montrer que si $(x^2 LLx)^{-1}\psi(x)$ est croissante, 0.1. vaut avec $a_\nu = o(\nu^{-1/2}\psi(\nu))$ et si cette même fonction est décroissante, alors 0.1. vaut avec $a_\nu = o(\nu^2\sqrt{LL\nu}\,/\psi(\nu))$.

En situation multidimensionnelle, l'approximation forte du processus empirique multivarié a été étudiée par P.Massart (87). Quand S est la classe des boules euclidiennes en dimension d, Massart obtient comme corollaire d'un théorème plus général le principe d'invariance suivant pour le processus empirique associé à un n-échantillon de loi uniforme:

si $(x_i)_{i>0}$ est une suite de variables aléatoires indépendantes de loi uniforme sur $]0,1]^d$, si $P_n = n^{-1}\sum_1^n \delta_{x_i}$, si $Z^{(n)} = \sqrt{n}(P_n - \lambda)$, λ étant la mesure uniforme sur le cube unité, alors, il existe une suite $(B_n)_{n>0}$ de ponts Browniens presque sûrement uniformément continus sur la classe $\mathcal{S}$ telle que

$$\sup_{S\in\mathcal{S}} \mid Z^{(n)}(S) - B_n(S) \mid = O(n^{-1/2d}(\log n)^{3/2}) \quad p.s.$$

Ce résultat est optimal au facteur $(\log n)^{3/2}$ près, d'après un travail de J. Beck (85). Massart obtient aussi des approximations fortes pour le processus empiriques indexé par des classes plus "grandes" de continuité du drap Brownien telles que la classe des convexes en dimension 2, ou la classe des parties à bords α-différentiables introduites par Dudley (74) avec $\alpha > d-1$. Pour obtenir ces résultats, Massart utilise essentiellement les constructions proposées par K.M.T. (75), et obtient au passage des principes d'invariance forts pour les processus de sommes partielles de variables aléatoires réelles i.i.d. ayant une transformée de Fourier-Laplace finie dans un voisinage de l'origine: par exemple, si $\mathcal{S}$ est la classe des boules euclidiennes, 0.1 vaut avec $a_\nu = \nu^{-1/2}(\log\nu)^{3/2}$ p.s.. Notre propos est de généraliser les résultats de Massart aux champs de vecteurs à valeurs dans $\mathbb{R}^k$ et à des hypothèses de moment plus faibles. Pour obtenir un résultat tel que 0.1. il est nécessaire de contrôler les fluctuations du processus Gaussien Y_ν et du processus X_ν uniformément sur la classe de parties $\mathcal{S}$ quand ν tend vers $+\infty$. La classe $\mathcal{S}$ doit donc posséder certaines propriétés entropiques et Géométriques.

<u>Quelles sont les "bonnes" classes ?</u>

Pour obtenir un principe d'invariance fort, il semble nécessaire de supposer que le processus $\{X_\nu(S) : S \in \mathcal{S}\}$ satisfait à un théorème de la limite centrale uniforme. Pour celà, la classe considérée doit être telle qu'il existe une version de drap Brownien G_λ de covariance $Cov(A,B) = d_\lambda(A,B)$ à trajectoires bornées et uniformément continues sur $(\mathcal{S}, d_\lambda)$. On supposera donc que la classe $\mathcal{S}$ est une classe de continuité du drap Brownien. D'après un travail de Dudley de 1967 (voir aussi Dudley, 73) une telle condition est satisfaite quand $\mathcal{S}$ a une entropie métrique pour l'écart d_λ intégrable. Rappelons que l'entropie est le nombre $N(\varepsilon, \mathcal{S})$ minimal de boules de rayon ε relativement à d_λ nécessaire pour recouvrir la classe $\mathcal{S}$ et que la classe $\mathcal{S}$ est dite d'entropie intégrable si le logarithme de l'entropie, noté $H(\,.\,)$ vérifie

$$\int_0^1 (\varepsilon^{-1} H(\varepsilon))^{1/2}\, d\varepsilon < +\infty.$$

Rappelons maintenant que la loi des variables aléatoires X_i est quelconque. Aussi, si l'on choisit comme loi pour les X_i la loi de Poisson de paramètre p recentrée et renormalisée, alors pour les petites valeurs de p, le processus obtenu est proche d'un processus de Poisson homogène d'intensité $\nu^d p$ recentré, processus qui est, à son tour proche du processus empirique associé à un échantillon de loi uniforme sur le cube unité. Une "bonne" classe, dans le cadre de notre problème sera donc, suivant la terminologie de Dudley (voir Dudley: 78) une λ-classe de Donsker.

Pour obtenir des vitesses explicites dans les principes d'invariance, nous étudierons deux types précis de classes de Donsker:

- Les classes de Vapnik-chervonenkis (ce sont les classes qui, à partir d'un certain rang $D+1$ ne pulvérisent aucune partie de cardinal $D+1$ du cube unité) dont les éléments vérifient une propriété de régularité des bords
- Les classes ayant une entropie avec inclusion relative à d_λ intégrable, l'entropie avec inclusion étant le nombre minimal d'éléments d'une famille $\mathcal{S}(\varepsilon)$ de Boréliens telle que tout élément S de $\mathcal{S}$ il existe deux éléments S^- et S^+ dans $\mathcal{S}(\varepsilon)$ tels que $S^- \subset S \subset S^+$ et $\lambda(S^+ - S^-) \leq \varepsilon$.

Dans le second cas, pour assurer l'équicontinuité asymptotique du processus de sommes partielles, il est nécessaire de considérer des version régularisées des processus X_ν et Y_ν, c'est à dire de remplacer les mesures X_ν et Y_ν par leur convolée avec l'unité approchée $\nu^d \mathbb{1}_{]-1,0]^d}(\nu x)$. Les nouvelles mesures ainsi définies sont asymptotiquement équicontinues (voir Bass (85)). Notons, que pour les classes de Vapnik-Chervonenkis dont les éléments vérifient une condition de régularité de bords, les processus de sommes partielles "brutes" sont asymptotiquement équicontinus: aussi, pour ces classes, il est inutile de régulariser les mesures X_ν et Y_ν.

Pour les deux types de classes ci-dessus, le théorème de la limite centrale uniforme a été étudié: Alexander (87) pour les classes de Vapnik-Chervonenkis, et Alexander et Pyke (86) pour les classes ayant une entropie avec inclusion intégrable, ont montré un théorème de la limite centrale uniforme dès que les variables X_i ont un moment d'ordre 2 fini. Pous obtenir ces résultats, les auteurs montrent que le processus de sommes partielles qu'ils considère vérifie certaines propriétés de tension, puis, en utilisant le théorème de représentation de Strassen, ils obtiennent l'existence d'un champ Y de vecteurs aléatoires Gaussiens i.i.d. tel que 0.1. vaut avec $a_\nu = o(1)$ en probabilité. De plus, avec des hypothèses identiques à celles de Alexander et Pyke, Bass (85) montre une loi du logarithme itéré fonctionnelle pour les champs de variables aléatoires réelles i.i.d.

Par contre, les vitesses dans les principes d'invariance forts pour des champs de variables indépendantes avec un moment d'ordre $r > 2$ fini, indexés par une classe de parties ayant une entropie relativement à l'écart d_λ intégrable ne sont pas connus. L'un des premiers travaux à ce sujet est du à Bass et Pyke (84). Ces auteurs supposent que la classe $\mathcal{S}$ considérée a un exposant d'entropie ς strictement inférieur à 1. Rappelons que

$$\varsigma = \limsup_{x \to 0} \mid \log(H(x)) / \log x \mid$$

$H(.)$ étant le logarithme de l'entropie. En généralisant les techniques de plongement de Skorohod aux processus de sommes partielles, Bass et Pyke obtiennent un principe d'invariance fort et l'utilisent pour démontrer une loi du logarithme itéré ainsi qu'un théorème de la limite centrale uniforme pour des processus de sommes partielles; mais les auteurs obtiennent ce principe d'invariance seulement sous la condition d'existence d'un moment d'ordre r supérieur à $2/(1-\varsigma)$ et la technique utilisée ne permet pas d'expliciter les vitesses obtenues dans ce cas. Les seuls résultats avec des vitesses dans ce domaine sont dus à Morrow et Philipp (86); sous la contrainte $r > 2(1+\varsigma)(1-\varsigma)^{-1}$, les auteurs montrent que 0.1. vaut avec $a_\nu = O((\log \nu)^{-s_1})$ p.s., s_1 étant une constante non explicite: ces résultats sont, comme nous le verrons plus tard, loin d'être optimaux. Dans les deux articles ci-dessus, les auteurs supposent que les éléments de la classe considérée ont des bords réguliers: cette hypothèse permet de controler l'écart entre les approximations finies-dimensionnelles de X_ν et Y_ν sur la classe $\mathcal{S}$ et est cruciale. Nous reprenons donc à notre compte ce type d'hypothèses. Décrivons à présent nos résultats.

Description de notre travail.

Dans la première partie de le thèse, on étudie l'approximation forte des processus de sommes partielles indexés par des classes de Vapnik-Chervonenkis. Dans la suite, les éléments de $\mathcal{S}$ vérifient l'hypothèse supplémentaire de régularité des bords suivante:

$$\sup_{S\in\mathcal{S}} \lambda((\partial S)^\epsilon) \leq K\epsilon^\delta \text{ pour un } 0 < \delta \leq 1, \text{ pour un } K > 0 \tag{0.2}$$

Sous les hypothèses ci-dessus et si $d \neq \delta$, on montre que, quand les vecteurs X_i ont un moment d'ordre r strictement supérieur à $2d(d-\delta)^{-1}$ fini, le principe d'invariance fort vaut avec $a_\nu = \nu^{-\delta/2}(\log\nu)^{1/2}$. Si $\delta = 1$ et si $r > 2d(d-1)^{-1}$ on montre, en reprenant des calculs antérieurs dus à Beck (85), que, si $\mathcal{S}$ est la classe des boules euclidiennes, le résultat obtenu est optimal au facteur $(\log\nu)^{1/2}$ près. On améliore ainsi le résultat précédent de Massart obtenu pour des variables à valeurs réelles ayant une transformée de Fourier-Laplace finie d'un facteur $\log\nu$.

Pour obtenir un principe d'invariance fort dès que les vecteurs X_i ont une variance finie, on introduit une condition de stabilité qui est, par exemple, vérifiée quand la classe $\mathcal{S}$ est stable par contractions. On montre alors que le principe d'invariance fort au sens de Strassen vaut avec les vitesses suivantes. Enoncons les résultats obtenus pour $d > 1$. Si ψ est choisie telle que $x^{-r}\psi(x)$ soit décroissante pour un r strictement inférieur à $2d(d-\delta)^{-1}$, si les vecteurs X_i ont un ψ-moment, alors le principe d'invariance fort vaut avec la vitesse $a_\nu = \nu^{-d/2}\varphi(\nu^d)$, φ étant la fonction définie à partir de la fonction inverse de ψ, par

$$\varphi(x) = \psi^{-1}(x)(1 + (x^{-1}LLx)^{1/2}\psi^{-1}(x))$$

Notons que si $x \to (x^2LLx)^{-1}\psi(x)$ est croissante alors $\varphi = 0(\psi^{-1})$ en $+\infty$: Le résultat obtenu est alors, d'après la remarque de Breiman, optimal. D'autre part, P. Major (76) a montré que le principe d'invariance fort au sens de Strassen n'est pas améliorable quand les variables ont seulement un moment d'ordre 2.

La méthode que nous utilisons pour démontrer ces résultats est une extension de la méthode de construction par divisions dyadiques proposée par K.M.T.. En introduisant une numérotation du quadrant supérieur droit de $\mathbb{Z}^d$ fondée sur un bon ordre, on linéarise le problème de construction des champs; la numérotation est choisie de telle sorte que la classe de parties de $\mathbb{Z}_+$ correspondant à $\nu\mathcal{S}$ soit à bords réguliers dans un sens à préciser. Puis, on construit les suites $(X_i)_{i>0}$ et $(Y_i)_{i>0}$ associées aux champs à construire par la numérotation à l'aide des techniques développées par K.M.T.. Pour pallier aux problèmes apparaissant dans le cadre de notre problème, on introduit cependant deux modifications essentielles dans cette construction:

• On utilise un argument de récurrence du à Sakhanenko, qui nous permet de travailler avec des vecteurs de loi "régulière"
• On utilise des troncatures variables en fonction du pas de construction.
Nous obtenons ainsi une construction "minimisant" l'écart en probabilité entre $X_\nu(S)$ et $Y_\nu(S)$ pour les parties vérifiant la condition 0.2., puis on utilise les propriétés combinatoires et entropiques des classes de Vapnik-Chervonenkis pour controler le suprémum de l'écart sur la classe S.

Dans la seconde partie de la thèse, on utilise la méthode de construction proposée dans la première partie, pour montrer des principes d'invariance forts pour des classes de Boréliens du cube unité ayant une entropie avec inclusion intégrable. Pour obtenir des principes d'invariance, il est alors necessaire de régulariser les processus de sommes partielles précédemment définis. On montre alors que sous la condition 0.2. de régularité des bords, le principe d'invariance fort vaut avec une vitesse qui s'exprime en fonction de l'intégrale d'entropie, du paramètre δ et du moment des vecteurs X_i. Posons

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

Alors 0.1 vaut avec $a_\nu = \nu^{-d/2} b(\nu^d, \varphi)$. Quand $d > 1$, $b(\,.\,, \varphi)$ est définie par:

$$b(x, \varphi) = \int_1^x \varphi(u^{-1}x) u^{-1/2} I(u)\, du$$

On discute les résultats obtenus dans toute une série de cas particuliers pour lesquels on montre que les résultats obtenus sont optimaux. En particulier, quand les variables ont un moment d'ordre r assez grand par rapport à la fonction d'entropie, on montre que $a_\nu = O(I(\nu^{-\delta}))$. Notons que $I(x)$ est un majorant du module de continuité uniforme du processus Gaussien standard indexé par la classe S (voir Dudley, 73).

Pour montrer le théorème principal, on utilise un argument de chainage restreint applicable à tout processus indexé par la classe S et vérifiant certaines inégalités exponentielle. Puis, en utilisant cette inégalité à chaque étage de la construction diadique on obtient le résultat demandé. Comme corollaire du résultat général de la seconde partie, mentionnons le théorème suivant. Soit S une classe de Boréliens du cube unité en dimension $d > 1$ vérifiant la condition périmétrique uniforme de Minkowski, c'est à dire la condition 0.2 avec $\delta = 1$, et ayant une entropie avec inclusion (pour d_λ) dont le logarithme

$H(.)$ satisfait $H(u) = O(u^{-\varsigma})$ quand u tend vers 0 pour un réel ς dans $]0,1[$. Si les vecteurs X_i ont un moment d'ordre $r > 2$ distinct de $2d(d-1+\varsigma)^{-1}$, alors 0.1 vaut avec $a_\nu = O(\nu^{(\varsigma-1)/2} + \nu^{d/r-d/2})$ presque sûrement. On étend ainsi un résultat de P. Massart (87) tout en l'améliorant d'un facteur $\log \nu$. Donnons quelques exemples de classes satisfaisant à ces hypothèses: la classe S de parties à bords α-différentiables introduite par Dudley (74) les vérifie avec $\varsigma = (d-1)/\alpha$ si $\alpha > d-1$. La classe des parties convexes du carré en dimension 2 les vérifie avec $\varsigma = 1/2$. Pour ces deux classes, on montre, en utilisant des techniques de théorie des irrégularités de distribution applicables aux classes invariantes par le groupe orthogonal (voir Beck (87)) que les résultats obtenus sont optimaux. De plus, on montre que les mêmes minorations valent pour l'approximation forte du processus multivarié associé à un échantillon de loi uniforme, améliorant ainsi des résultats antérieurs de Borisov. Dans ces deux cas, les vitesses optimales dépendent à la fois de l'entropie de la classe et de la dimension d d'espace.

Enfin, la troisième partie est consacrée à l'étude des processus de sommes partielles indexés par les quadrants en dimension 2. On montre alors que, quand les variables X_i sont réelles et ont un moment fini d'ordre $r > 4$, alors les vitesses obtenues dans le cadre plus général des classes de Vapnik-Chervonenkis peuvent etre améliorées en construisant différemment les champs X et Y. Ainsi, pour tout champ $(X_i)_{i\in\mathbb{Z}^2}$ de variables aléatoires réelles i.e.d. centrées et ayant un moment d'ordre $r > 2$ fini,on peut construire un champ Y de variables aléatoires i.e.d. de loi $N(0, Var\,X_1)$ tel que, si S est la classe des quarts de plan inférieurs gauches, le principe d'invariance fort vaut avec $\nu a_\nu = \nu^{2/r}(\log \nu)^2$ p.s.. D'après la remarque de Breiman, ce résultat est optimal au facteur $(\log \nu)^2$ près, même quand les variables ont un moment d'ordre r supérieur à 4. La technique de construction utilisée est alors la construction analogue dans le cas des processus de sommes partielles de la construction bidimensionnelle jointe du pont empirique et du pont Brownien proposée par G. Túsnady (78). Ce résultat confirme la différence de structure existant entre la classe des boules et celle des quadrants. Cette différence de comportement avait déjà été étudiée en théorie des irrégularités de distribution; à ce sujet, on pourra regarder les travaux de J. Beck (de 83 à 88).

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.

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

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

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

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

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

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

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

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

DUDLEY, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **7** 899-929.

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

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

EINMAHL, U. (1989). The Darling-Erdös Theorem for Sums of I.I.D. Random Variables. *Probab. Th. Rel. Fields.* **82** 241-247.

HANSON, D. L. and RUSSO, R. P. (1983). Some Results on the Wiener Process with Applications to LAG Sums of I.I.D. Random Variables. *Ann. Probab.* **11** 609-623.

HANSON, D.L. and RUSSO, R. P. (1989). Some Liminf Results for Increments of a Wiener Process. *Ann. Probab.* **17** 1063-1082.

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** 111-6131.

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.

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

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

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

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

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* 4-49. *Trudy Inst. Mat.* Vol. 3, "Nauka" Sibirsk.Otdel, Novosibirsk [Russian].

SHAO, Q-M. (1989). On a Problem of Csörgó et Révész. *Ann. of Probab.* **17** 809-812.

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

TUSNÁDY, G. (1978). A remark on the approximation of the sample DF in the multidimensional case. *Period. Math. Hung.* **8.** 53-55.

CHAPITRE I

STRONG APPROXIMATION FOR SET-INDEXED PARTIAL SUM PROCESSES, VIA K.M.T. CONSTRUCTIONS I.

RIO Emmanuel
URA D 0 743 CNRS - Statistique Appliquée, Bât. 425 Mathématique
91405 ORSAY Cedex France

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

The most striking result is that, both of 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 $\mathcal{N}(0, Var\, X_1)$-distributed random vectors may be constructed in such a way that, up to a power of $log\ \nu$:

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

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

From a paper of Beck (85), it is straightforward to prove that such a result cannot be improved, when $\mathcal{S}$ is the class of Euclidean balls.

Résumé : Soit $(X_i)_{i\in\mathbb{Z}_+^d}$ un champ de vecteurs indépendants, équidistribués, centrés et à valeurs dans $\mathbb{R}^k$. Quand $E(|X_1|^r) < +\infty$ pour un $r > 2$, on obtient l'approximation forte du processus de sommes partielles $(\sum_{i\in\nu S} X_i : S \in \mathcal{S})$ par un processus de sommes partielles gaussien $(\sum_{i\in\nu S} Y_i : S \in \mathcal{S})$, uniformément sur tous les éléments d'une classe de Vapnik-Chervonenkis $\mathcal{S}$ de boréliens de $[0,1]^d$.

Le résultat principal est le suivant : on peut construire simultanément un champ $(X_i)_{i\in\mathbb{Z}_+^d}$ de vecteurs indépendants équidistribués et un champ $(Y_i)_{i\in\mathbb{Z}_+^d}$ de vecteurs gaussiens indépendants centrés de variance $Var\, X_1$, tels que, à une puissance de $log\ \nu$ près

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

pour toute classe de Vapnik-Chervonenkis $\mathcal{S}$ vérifiant la condition de Minkowsky uniforme.
En partant d'un article de Beck (85), on démontre ensuite qu'un tel résultat est optimal quand $\mathcal{S}$ est la classe des boules euclidiennes.

Code Matière A.M.S. (1980) - 60 F 17 - 62 G 99

Mots-clés : Central limit theorem, Set-indexed process, Partial-sum process, Invariance principle, Vapnik-Chervonenkis Class, Metric entropy, Random measure.

1. INTRODUCTION

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

$$\sup_{S\in\mathcal{S}} | X_\nu(S) - Y_\nu(S) | = O(\nu^{1/r}) \quad a.s.$$

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

$$\sup_{S\in\mathcal{S}} | X_\nu(S) - Y_\nu(S) | = O(\log \nu) \quad a.s.$$

It is worth noticing that the rates of strong approximation appearing above are optimal (this comes from Breiman's remark (67) when the r-th moment is finite and from Erdős and Rényi, 70: cf. Csr̃gó and Révész, 81) when the moment-generating function is finite). Recently, U. Einmahl (89) 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: to 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 : to Bass (85) for a functional law of iterated logarithm and to Alexander and Pyke (86) 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 : to Alexander (87) for independent arrays indexed by Vapnik-Chervonenkis classes : to Morrow and Philipp (86) for invariance principles and rates of convergence in the independent case for i.i.d. random vectors with a finite r-th-moment. However, these rates are not explicit because of the techniques used by Morrow and Philipp. On the other hand, Massart (89) 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 $\mathcal{S}$ is the class of Euclidean balls. Here, our aim is to strengthen and to unify the results obtained by Massart, Morrow and Philipp, Bass and Pyke, to almost sure invariance principles with optimal rates of convergence.

We are interested by independent arrays of $\mathbb{R}^k$-valued random vectors with a finite r-th moment ($r > 2$) indexed by Vapnik-Chervonenkis (V-C) classes of sets. In the forthcoming paper II we shall study the case of totally bounded with inclusion classes having a convergent entropy integral. Note that, in this case, it is necessary to consider a smoothed version of the partial-sum process (cf. Alexander and Pyke (86)). We mention in advance that we obtain an almost sure invariance principle with an optimal rate of convergence for any $r > 2$ while Bass and Pyke, Morrow and Philipp had to assume very stringent conditions of moments.

Now, we discuss further the scope of results and the related literature. In Sections 3 and 4, using the extension made by Einmahl (89) to the multidimensional case of K.M.T.'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 $\mathcal{S}$, we obtain much better upper bounds on $pr(\mid X_\nu(S) - Y_\nu(S) \mid \geq t)$ than the Bass and Pyke's ones (86 : cf. 5.4). Now, 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 $\mathcal{S}$. Let λ be the Lebesgue measure on $\mathbb{R}^d$. Given a norm $\mid . \mid$ on $\mathbb{R}^d$ and a Borelian subset S of $\mathbb{R}^d$, we set

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

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

$$\sup_{S \in \mathcal{S}} \lambda((\partial S)^\epsilon) \leq K\epsilon^\delta \text{ for any } 0 < \epsilon \leq 1, \text{for some } \delta \in]0,1]. \tag{1.1}$$

If $\delta = 1$ this condition is the uniform Minkowsky condition previously used by Massart (89), Bass and Pyke (84). When $\mathcal{S}$ is a V-C class fulfilling 1.1., our upper bounds and the combinatorial properties of the V-C classes yield:

$$\sup_{S \in \mathcal{S}} \mid X_\nu(S) - Y_\nu(S) \mid = O(\nu^{(d-\delta)/2}(\log \nu)^{1/2} + \nu^{d/r}) \quad a.s.$$

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, we prove that such a result cannot be improved when $\delta = 1$, $r > 2d(d-1)^{-1}$, and $\mathcal{S}$ is the class of Euclidian 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 multidimensional quantile transformation.

2. *DEFINITIONS AND RESULTS* .

Throughout this section, the probability space Ω is assumed to be rich enough in the following sense : there exists a random variable, defined on Ω, with uniform distribution over $[0, 1]$, which is independent of the observations.

For any bounded Borelian subset B of $\mathbb{R}^d$, define :

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

If $\mathcal{S}$ is any family of Borel subsets of the unit cube I^d , we define the class $\nu\mathcal{S}$ of subsets of $\mathbb{Z}^d$ for any integer ν by :

$$\nu\mathcal{S} = \{\nu S \cap \mathbb{Z}^d : S \in \mathcal{S}\}$$

where $\nu S = \{\nu x : x \in S\}$. In order to get nice asymptotic properties for a normalized version of the partial sum process $(X(\nu S) : S \in \mathcal{S})$, we need to have some reasonable growth conditions on $\nu\mathcal{S}$ when ν tends to infinity . So , we shall assume that $\mathcal{S}$ is a V-C class. We recall that this means :

$$D(\mathcal{S}) = \sup\{n \in \mathbb{N} : \sharp(A \cap \mathcal{S}) = 2^n \text{ for some set } A \text{ with } \sharp A = n\} < \infty$$

where $A \cap \mathcal{S} = \{A \cap S : S \in \mathcal{S}\}$ and $\sharp E$ denotes from now on the cardinality of the set E. We call $D(\mathcal{S})$ the density of $\mathcal{S}$ (See Assouad for many examples and properties of such classes).

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

THEOREM 1. *Assume that $d \neq \delta$ and let $\mathcal{S}$ be a family of Borelian subsets of $[0,1]^d$. Assume that $\mathcal{S}$ is a V-C class fulfilling 1.1. for some $0 < \delta \leq 1$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, fulfilling:*

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

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 $N(0, VarQ)$-distributed random vectors such that

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

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 (87). From that point of view, our 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 Komlós, Major, and Tusnády prove that the rate of approximation is of the order of $0(\log \nu)$ w.p. 1.

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 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 (83) for more about 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 distributions functions F and G, and let $\mathcal{S}$ denote the class of intersections of Euclidean balls with the unit cube . Then, there exists some positive constant $c(d)$ depending only on d such that :*

$$E(\nu^{1-d}\sup_{S\in\mathcal{S}}(X(\nu S)-Y(\nu S))^2)\geq(c(d)W(F,G))^2$$

and

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

Comments . From Theorem 2, it follows that the rate of approximation appearing in Theorem 1 is nearly optimal.

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. But, 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 $\mathcal{S}$. Let ψ be a mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$ such that:

(1.2.) (i) $\int\psi(\mid x\mid)dQ(x)<+\infty$, and $x^{-2}\psi(x)$ is a one to one continuous, increasing mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$.

(ii) There exists $r>2$ such that $x^{-r}\psi(x)$ is nonincreasing.

(iii) Furthermore, if there does not exist $r<4$ such that $x^{-r}\psi(x)$ is nonincreasing, then $(x^2LLx)^{-1}\psi(x)$ is nondecreasing.

Note that, when Q has a finite second moment, such a mapping exists (see Major (76)). Throughout, ψ^{-1} denotes the inverse function of ψ.

THEOREM 3. *Let $\mathcal{S}$ be a family of Borel subsets of $[0,1]^d$. Assume that $\mathcal{S}$ is a V-C class fulfilling 1.1. for some $0<\delta\leq 1$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, and let ψ be a mapping fulfilling 1.2 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 $N(0,VarQ)$-distributed random vectors such that:*

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

Moreover, if we assume that

$$\mathcal{V} = \bigcup_{\nu\in\mathbb{N}} \nu \mathcal{S} \text{ is a V-C class,} \tag{1.3}$$

Then, setting

$$\varphi(x) = \psi^{-1}(x)(1 + (x^{-1}LLx)^{1/2}\psi^{-1}(x))$$

we have:

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

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

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

Comments. 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 also Major (76)).

Theorem 2 provides a rate of the order of $\nu^{-d/2}\psi^{-1}(\nu^d)$ in Alexander's uniform central limit Theorem (cf Alexander (87): Corollary 4.4). From Breiman's remark (67), we believe that this rate is optimal.

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

COROLLARY 1. *Let $\mathcal{S}$ be a family of Borel subsets of $[0,1]^d$. Assume that $\mathcal{S}$ is a V-C 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 , fulfilling:*

$$\int_{\mathbb{R}^k} \mid x \mid^r dQ(x) < +\infty \text{ for some some } r > 2 \text{ with } r \neq 2d(d-\delta)^{-1}$$

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 $N(0, Var\, Q)$-distributed random vectors such that:

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

Furthermore, if $\mathcal{S}$ fulfills 1.3, 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.

Consequences of Theorem 3: When $\mathcal{S}$ is contraction closed, *i.e.* $t\mathcal{S} \subset \mathcal{S}$ for any $0 < t < 1$ (note that this condition is fulfilled by many classes of interest), the class $\mathcal{V}$ defined in Theorem 3 is a V-C class with $D(\mathcal{V}) = D(\mathcal{S})$.

Proof. Let $m = D(\mathcal{S}) + 1$. For any subset A of $\mathbb{R}^d$ with cardinality m , there exists an integer p such that $A \cap \mathcal{V} = A \cap p\mathcal{S}$. Hence, $\mid A \cap \mathcal{V} \mid < 2^m$.

Part (b) of Theorem 3 is a weak invariance principle in the sense of Philipp (80) while (c) is a strong invariance principle where the function $x \to x^2 LLx$ plays an important role. In fact, we can derive from Theorem 3 two different results according to the monotonicity of the function $x \to \psi(x)(x^2 LLx)^{-1}$.

COROLLARY 2. *Let $\mathcal{S}$ be a family of Borel subsets of $[0,1]^d$. Assume that $\mathcal{S}$ is a V-C class fulfilling 1.3. and 1.1. for some δ in $]0,1]$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance , and let ψ be a mapping fulfilling the assumptions of Theorem 3 for some $r < 2d(d-\delta)^{-1}$. Furthermore assume that $\psi(x)(x^2 LLx)^{-1}$ 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 $N(0, VarQ)$-distributed random vectors such that:

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

Comments. From Breiman's remark (67), it follows that this result is optimal when $\mathcal{S}$ contains all the singletons of I^d.

On the other hand, when $x \rightarrow \psi(x)(x^2 LLx)^{-1}$ is nonincreasing, we obtain the following Strassen's type invariance principle from Theorem 3:

COROLLARY 3. *Let $\mathcal{S}$ be a family of Borel subsets of $[0,1]^d$. Assume that $\mathcal{S}$ is a V-C class fulfilling 1.3. and 1.1. for some δ in $]0,1]$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, and let ψ be a mapping fulfilling the assumptions of Theorem 3. Furthermore assume that $x \rightarrow \psi(x)(x^2 LLx)^{-1}$ 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 $N(0, Var\, Q)$-distributed random vectors such that

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

Comments. When $\mathcal{S}$ contains the class of lower-left orthants and only the second moment is assumed to be finite, this result is optimal (cf Major (76)). moreover, when $\psi(x) = x^2(LLx)^\alpha$ for some $0 < \alpha < 1$ the power of $LL\nu$ in this result cannot be improved; cf. Einmahl (87, Theorem 4) for more about this.

Now , from Corollary 3 , we can derive a functional law of the iterated logarithm . 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 $VarQ = I_k$, and let $\mathcal{S}$ be a family of Borel subsets of I^d fulfilling the assumptions of Corollary 3 . For any function F from $\mathcal{S}$ into $\mathbb{R}^k$, let :

$$\|F\|_{\mathcal{S}} = \sup_{S \in \mathcal{S}} \mid F(S) \mid$$

where $\mid x \mid$ denotes the euclidean norm of x , and let $\mathcal{K}$ be the subset of functions from $\mathcal{S}$ to $\mathbb{R}^k$ given by :

$$\mathcal{K} = \{F : \text{for some } f : I^d \rightarrow \mathbb{R}^k \text{ with } \int_{I^d} \mid f(t) \mid^2 dt \leq 1, \quad F(S) = \int_S f(t)dt \text{ for all } S \in \mathcal{S}\}$$

When $\mathcal{S}$ contains the class of lower-left orthants, f is uniquely defined. So, we shall assume that $\mathcal{S}$ contains the class of lower-left orthants. Then, the approximating Gaussian process constructed in Corollary 3 fulfills the conditions of Theorem 3.1 of Bass and Pyke (84). So, the following result is available.

Law of the iterated logarithm. $((2\nu^d LL\nu)^{-1/2} X(\nu S) : S \in \mathcal{S})$ *is relatively compact with respect to the metric* $\| \, . \, \|_{\mathcal{S}}$ *, and the set of limits points is exactly* $\mathcal{K}$ a.s.

Comments. This result is new, as far as we know. Note that Bass (85) 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 2. the proof of these theorems is 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 0 and positive definite covariance. We may without loss of generality assume that $VarQ = I_k$. $(X_i)_{i \in \mathbb{Z}_+^d}$ denotes an array of independent random vectors with common law Q , and ψ is any mapping fulfilling 1.2, such that $E(\psi(| X_1 |)) < +\infty$.

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 so defined sequences into arrays. Here, we shall need the following Lemma, due to Skorohod.

Skorohod-Lemma (76). Given two polish spaces R_1, R_2 and a random variable V from Ω to R_2, 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, defined on Ω, which is independent of V and uniformly distributed on $[0, 1]$. 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 (78) 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 (see also Major (78) for the difficulties that arise when one wants to work with non regular laws).

In order to avoid these technical difficulties,we shall use an argument of Sakhanenko (84), which consists of iterating the same construction method, that turns a sequence of independent standard Gaussian vectors into a sequence $(Z_i)_{i\geq 1}$ 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 $(Y_i^0)_{i\geq 1}$ be a sequence of independent standard normal random vectors. Suppose that there also exist a sequence $(\delta_l)_{l\geq 0}$ of independent random variables having uniform distribution on $[0, 1]$ and being independent of the sequence $(Y_i^0)_{i\geq 1}$. By means of a construction method which shall be explained later (cf. *), we define a sequence $(Z_i^0)_{i\geq 1}$ of independent random vectors from $(Y_i^0)_{i\geq 1}$ and δ_0 such that:

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

(This means that $(Z_i^0)_{i\geq 1}$ is a deterministic measurable function of $(Y_i^0)_{i\geq 1}$ and δ_0 , which shall be explained later).

Now, we define the sequence $(Y_i^1)_{i\geq 1}$ of independent Gaussian random vectors 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 $(X_{2i-1})_{i\geq 1}$ are independent with common law Q . It remains to define the random vectors $(X_{2i})_{i\geq 1}$. By Skorohod's lemma ,there exists a sequence $(Z_i^1)_{i\geq 1}$ of independent $\sigma(\delta_1, Y_i^1 : i \geq 1)$-measurable random vectors such that :

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

So ,by induction, for each natural l, there exists a sequence $(Y_i^l, Z_i^l)_{i\geq 1}$ such that:

- For any positive i, Z_i^l is $\sigma(\delta_l, Y_i^l : i \geq 1)$-measurable.
- $(Z_{2^l i}^l, Y_{2^l i}^l)_{i\geq 1}$ has the same law as the sequence $(Z_i^0, Y_i^0)_{i\geq 1}$.
- For any positive i, $Y_{2^{l+1} i}^{l+1} = Z_{2^{l+1} i}^l$

Then, for each natural l, for each odd natural i, we set $X_{2^l i} = Z_{2^l i}^l$. Clearly, the so defined sequence will be a sequence of independent random vectors with common law Q (cf Einmahl (89)).

Now, it remains to explain the method of construction of the sequences $(Z_i^0)_{i\geq 1}$ and $(Y_i^0)_{i\geq 1}$ in a common probability space. By Skorohod's lemma again, it suffices to construct the sequence $(Y_i^0)_{i\geq 1}$ from $(Z_i^0)_{i\geq 1}$. Our construction method uses the dyadic scheme previously introduced in K.M.T. (75). But, if one wants to use the dyadic scheme exactly as in K.M.T., 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 illustraded by the work of Bass and Pyke (84). So, we shall adapt the technique of adaptative truncatures initiated by Bass (85) in his paper on the functional law of the iterated logarithm for set-indexed partial-sum processes to the K.M.T.'s dyadic scheme .

(*) Construction of the two sequences

Notation. Throughout the construction, the intervals $]l,m]$ have to be interpreted as subsets of $\mathbb{Z}_+$. $l^2(\mathbb{Z}_+)$ is given the canonical inner product, which we denote by $(\,.\, | \,.\,)$, and $l^2(]l,m])$ denotes the subspace of $l^2(\mathbb{Z}_+)$ of functions with support included in $]l,m]$. Now, we want to define the finite sequence $(Y_i)_{2^L<i\leq 2^{L+1}}$ as a deterministic function of $(Z_i)_{2^L<i\leq 2^{L+1}}$. Here, it will be convenient to define a dyadic orthogonal basis, as Massart (89) does in his paper.

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 $\tilde{e}_{j,p} = e_{j,p} - 2e_{j-1,2p}$. It is easy to see that the system $\{e_{L,1}, \tilde{e}_{j,p} \mid 0 < j \leq L, 2^{L-j} \leq p < 2^{1+L-j}\}$ is an orthogonal basis of $l^2(]2^L, 2^{L+1}])$. So , in order to construct the sequence $(Y_i)_{2^L<i\leq 2^{L+1}}$, 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 sequence f of reals numbers 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 $(Z_i)_{2^L<i\leq 2^{L+1}}$, we now introduce a nonincreasing filtration $(\mathcal{F}_{j,L})_{0<j\leq L}$ of σ-algebras, related to different levels of truncatures at each stage of the dyadic scheme.

<u>A dyadic filtration</u> Let define the increasing sequence $(M_j)_{j\geq 0}$ by $\psi(M_j) = 2^{j+1}$ for any natural j , let $\bar{Q}_j$ be the distribution of $X\mathbb{1}_{|X|\leq M_j}$ and let Q_j be the conditional distribution of X, given $(\mid X \mid\leq M_j)$, where X is a random vector with law Q.

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 :

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

Now , we define $\mathcal{F}_{j,L}$, for any $0 < j \leq L$ by :

$$\mathcal{F}_{j,L} = \sigma(B_j, \mid X_i : i \in B_j, \mid U_{j,p}^0 : 2^{L-j} \leq p < 2^{1+L-j})$$

Clearly, $(\mathcal{F}_{j,L})_{0<j\leq L}$ is a nonincreasing filtration. So, if we define random vectors $V_{L,1}$ and $\tilde{V}_{j,p}$ such that:

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

then , the random vectors $V_{L,1}$ and $\tilde{V}_{j,p}$ will be independent with a Gaussian law. Now, in order to construct these Gaussian random vectors, we introduce further notation and definition. So, we set :

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

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

$$b_{j,p} =\mid B_j \cap I_{j,p} \mid \text{ and } \tilde{b}_{j,p} = b_{j,p} - 2b_{j-1,2p}$$

Now, we want to define $\tilde{V}_{j,p}$ from $\mathcal{F}_{j,L}$ and $\tilde{U}_{j,p}^0$. Clearly, the r.v.'s $(Z_i)_{i\notin B_j}$ are, given $\{B_j : Z_i \mid i \in B_j\}$, conditionally independent with conditional distribution Q_j when i

is odd, and $N(0, I_k)$ when i is even. Hence, conditionally given $\mathcal{F}_{j,L}$, the random vectors $\{\tilde{U}^0_{j,p} : 2^{L-j} \leq p < 2^{1+L-j}\}$ are independent , and , for each p , the conditional law of $\tilde{U}^0_{j,p}$ 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 $U^0_{j,p}$. So, we may define $\tilde{V}_{j,p}$ as a function of $(b_{j,p}, \tilde{b}_{j,p}, U^0_{j,p}, \tilde{U}^0_{j,p})$.

From now on, let $b_{j,p}$ and $\tilde{b}_{j,p}$ be given. First, we define a random vector $\tilde{W}_{j,p}$ from $(U^0_{j,p}, \tilde{U}^0_{j,p})$ such that, given B_j, the random vectors $U^0_{j,p}$ and $\tilde{W}_{j,p}$ are uncorrelated. So, we set :

$$\tilde{W}_{j,p} = \tilde{U}^0_{j,p} + \tilde{b}_{j,p} VarQ_j (Var(U^0_{j,p} \mid B_j))^{-1} U^0_{j,p}$$

We also set :

$$\tilde{W}^0_{j,p} = (Var(\tilde{W}_{j,p} \mid B_j))^{-1/2} \tilde{W}_{j,p}$$

Definition of $\tilde{V}_{j,p}$. Given B_j , $\tilde{V}_{j,p}$ 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 1.

Now , for each natural j , we set $\Gamma_j = \frac{1}{\sqrt{2}}(I_k + Var\bar{Q}_j)^{1/2}$

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 $\mathcal{F}_{L,L}$-measurable with law $N(0, 2^L I_k)$ and, for each $j > 1$, the random vectors $(\tilde{V}_{j,p})_{p>0}$ are $\mathcal{F}_{j-1,L}$-measurable, and, given $\mathcal{F}_{j,L}$, conditionally independent with common law $N(0, 2^j I_k)$. It remains now to define Y_1, Y_2, and the random vectors $(\tilde{V}_{j,p})_{p>0}$. Here, we may assume that the probability space is rich enough in order 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 above defined sequence is a sequence of independent standard normal random vectors. Moreover, the following nice property holds :

Property 3.1. 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 so constructed sequences $(Y_i)_{i>0}$ and $(X_i)_{i>0}$ into arrays. Here, it will be convenient to define a multidimensional dyadic order on $\mathbb{Z}^d_+$.

Notation. $\mathbb{Z}^d$ is provided with the usual sum, product and with the product order. We define now the subset J of $\mathbb{N}^d$ by:

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

It is obvious that the map from J onto $\mathbb{N}$ which maps $(j_1, \ldots, j_d)$ onto $(j_1 + j_2 + \cdots + j_d)$ is a nondecreasing, one to one mapping. So, for each integer j, we shall call $(j_1, j_2, \ldots, j_d)$ the unique element of J such that $j = j_1 + j_2 + \cdots + j_d$. Let R_j be the lattice of multiples of $(2^{j_1}, 2^{j_2}, ..., 2^{j_d})$: we define the box $C'_{j,p}$ for any p of R_j by (here $\mathbb{1} = (1, 1, ..., 1) \in \mathbb{R}^d$):

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

We now define our multidimensional order. Let a and b be two different elements of $\mathbb{Z}^d_+$. We set:

$$j(a, b) = \min\{j \in \mathbb{N} \mid \exists p \in R_j \text{ such that } a \in C'_{j,p} \text{ and } b \in C'_{j,p}\}.$$

Clearly, the mapping $(a, b) \to j(a, b)$ is a distance on $\mathbb{Z}^d_+$. Moreover, this distance has the following nice properties: for any elements a, b and c in $\mathbb{Z}^d_+$,

$$j(a, c) = \max(j(a, b), j(b, c)) \qquad 3.2(a)$$

and

$$\text{either } j(a, b) < j(a, c) \text{ or } j(b, c) < j(a, c) \qquad 3.2(b)$$

Now, we define the order relation $\prec$ as follows. For any $a \in \mathbb{Z}^d_+$, $a \prec a$. if a and b are two different elements of $\mathbb{Z}^d_+$, then $j = j(a, b) > 0$. p being the element of R_j, such that $\{a, b\} \subset C'_{j,p}$, $a \prec b$ iff $a \in C'_{j(a,b)-1,p}$. Clearly, $\prec$ is an antisymmetric relation and, for any a and b in $\mathbb{Z}^d_+$, either $a \prec b$ or $b \prec a$. Now, let a b and c be three different positive integers and assume that $a \prec b$ and $b \prec c$. From 3.2 (a) it follows that either $j(b, c) = j$ or $j(a, b) = j$. In the first situation, b is in $C_{j-1,p}$ and, by (b) of 3.2, $j(a, b) < j$: then $a \prec c \Leftrightarrow a \prec b$: hence $a \prec c$. In the second situation, $j(b, c) < j$, and $a \prec c$. So, $\prec$ is a transitive relation and a total order. Moreover, the following lemma is avaiblable.

LEMMA 0. *$(\mathbb{Z}^d_+, \prec)$ is well-ordered. Hence, there exists an unique nondecreasing one to one mapping σ from $(\mathbb{Z}^d_+, \prec$ onto $(\mathbb{Z}_+, \le)$. Moreover, σ maps the boxes $C'_{N,0}$ onto the intervals $]0, 2^N]$, the boxes $C'_{j,k}$ on the intervals $I_{j,p}$.*

Proof. Let B be any subset of $\mathbb{Z}_+^d$ and let $N = j(\mathbb{1}, B) = \inf_{b\in B} j(\mathbb{1}, b)$. We define the subset $B(N)$ of B by

$$B(N) = \{b \in B \text{ such that } j(\mathbb{1}, b) = N\}$$

Clearly, $B(N)$ is a finite, nonempty set included in the box $C'_{N,0}$. So, there exists an element a in $B(N)$ such that $a \prec b$ for any b in $B(N)$. Now, for any b in $B \setminus B(N)$, $j(\mathbb{1}, b) > N$. Hence, $j(a, b) > N$ and $j(\mathbb{1}, a) = N$. Hence, using again the same arguments, we obtain that $a \prec b$. So, $(\mathbb{Z}_+^d, \prec)$ is well-ordered.

Then, noting that for any (j,p) in $\mathbb{N} \times R_j$, for any m in $C'_{j,p}$, $p + \mathbb{1} \prec m \prec p + (2^{j_1}, \ldots, 2^{j_d})$ it is easily seen that the second part of our lemma holds.

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

4. *UPPER BOUNDS FOR THE CONSTRUCTION.*

Notation. If B is any subset of $\mathbb{Z}_+^d$, let $\sigma B = \{\sigma(i) : i \in B\}$ For any class $\mathcal{V}$ of subsets of $\mathbb{Z}_+^d$, we set $\sigma\mathcal{V} = \{\sigma V : V \in \mathcal{V}\}$ and, for each integer ν, we define the classes $\mathcal{S}_\nu$ and $\mathcal{A}_\nu$ by :

$$\mathcal{S}_\nu = \bigcup_{p\leq\nu} p\mathcal{S} \text{ and } \mathcal{A}_\nu = \sigma\mathcal{S}_\nu$$

By Lemma 0, $\mathcal{A}_\nu$ is a class of subsets of $]0,n]$. Moreover, when $\mathcal{S}$ fulfills the condition 1.3. of theorem 2 , $\mathcal{A}_\nu$ is a V-C class of subsets of $\mathbb{Z}_+$ and $D(\mathcal{A}_\nu) \leq D(\mathcal{V})$. This remark shall be used in the proof of Theorem 3.

Clearly, there is no loss of generality in assuming that $\nu = 2^N$, which we shall do in the sequel. Given a finite subset A of $\mathbb{N}$ and a sequence $(u_i)_{i>0}$ of vectors of $\mathbb{R}^k$, we set $u(A) = \sum_{i\in A} u_i$. It is obvious that:

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

So, henceforward we work with the sequences $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ defined in Section 3.

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

$$\mathcal{B}_o = \{\tilde{e}_j : 0 \leq j < Nd\} \cup \{e_{o,o}\} \quad and \quad \mathcal{B}_j = \{\tilde{e}_{j,p} : 0 < p < 2^{Nd-j}\}$$

Then $\mathcal{B} = \bigcup_{j=0}^{Nd-1} \mathcal{B}_j$ is an orthogonal basis of $l^2(]0, 2^{Nd}])$. Now, let Π_j be the orthogonal projector on the space generated by $\bigcup_{l=1}^{j} \mathcal{B}_l$. 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 $(\Pi_j f \mid \Pi_j f)$. From now on, for convenience, we shall confuse the class $\mathcal{A}_\nu$ with the class of indicator functions of the elements of $\mathcal{A}_\nu$. Then, the uniform control on $\mathcal{A}_\nu$ of the above inner products is ensured via the geometrical assumption 1.1. on the boundaries of elements of $\mathcal{S}$ and the perimetric properties of the mapping σ:

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

$$\sum_{i=1}^{n} \mid \Pi_j f(i) \mid \leq 2K 2^{N(d-\delta)} 2^{j\delta/d}$$

<u>Proof.</u> First, we note that, for any function f taking its values in $[0, 1]$,

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

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 lower-left vertice $i - \mathbb{1}$. For each natural p , we define the closed subset $C_{j,p}$ of $\mathbb{R}^d$ from $I_{j,p} =]p2^j, (p+1)2^j]$ by:

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

For each element f of $\mathcal{A}_\nu$, there exists an integer m smaller than ν and an element S of the family $\mathcal{S}$ such that $f = 1\!\!1_{\sigma(mS)}$. If the boundary of mS does not meet the box $C_{j,p}$, $\Pi_j f(i) = 0$ for any i in $I_{j,p}$. Now, we may assume that $\mathbb{R}^d$ is provided with the norm of the supremum. Then, if the boundary of mS meets the box $C_{j,p}$, $C_{j,p}$ is included in the Borelian set $(m\partial S)^a$, where $a = 2^{jd}$. Now, recall that the interiors of the boxes $C_{j,p}$ are disjoint. Hence,

(4.2) the cardinality of the set of integers p such that $\Pi_j(f) \not\equiv 0$ on $I_{j,p}$ is no more than $2^{-j}\lambda((m\partial S)^a)$.

We complete the proof by combining the above inequality and the condition 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

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

for any natural l, for any odd integer i in $[2^L, 2^{L+1}[$. Now, let $n = \nu^d$. Clearly,

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

So, it will be enough to control each of the terms on right hand. 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 2, being straightforward and using only the integrability assumption $\int \psi(| X_1 |) dQ(x) < +\infty$.

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 shall use the dyadic decomposition previously introduced in K.M.T.(75). If f is any function from $\mathbb{Z}_+$ to $\mathbb{R}$ with support included in $]0, n]$, we set $\gamma_j(f) = 2^{-j}(f \mid \tilde{e}_j)$ and $\gamma_{j,p}(f) = 2^{-j}(f \mid \tilde{e}_{j,p})$. Then, the orthogonal expansion of the function f with respect to the orthogonal basis $\mathcal{B}$ 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 introduce now further notation. Let define the random sequences $(\bar{\xi}_i^j)_{i>0}$ and $(\xi_i^j)_{i>0}$ by :

$$\bar{\xi}_i^j = \mathbb{1}_{M_j<|\bar{X}_i|\le M_{j+1}}\bar{X}_i \text{ and } \xi_i^j = \bar{\xi}_i^j - E(\bar{\xi}_i^j)$$

and let

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

We define now the random vectors $D_j(f)$ by :

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

for any positive j, and $D_0(f)$ by

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

It can easily be seen that

$$Y(f) - \tilde{X}(f) = D_0(f) + \sum_{j=1}^{Nd-1} D_j(\Pi_j f). \tag{4.4}$$

So, in order to control the random vector $Y(f) - \tilde{X}(f)$, it will be enough to control each of the r.v.'s $D_j(\Pi_j f)$, and the random vector $D_0(f)$. First, we note that, for any function f with values in $[-1,1]$, $|\gamma_j(f)| \le 1$ and $|\gamma_{j,p}(f)| \le 1$.

Hence, setting $D_0 = | Y_1 - \tilde{X}_1 | + \sum_{j=0}^{Nd-1} | \tilde{U}_j - \tilde{V}_j |$, $|D_0(f)| \le D_0$ for any $f \in \mathcal{A}_\nu$.

Now, let $D_j = \sup_{A\in\mathcal{A}_\nu} | D_j(\Pi_j f) |$. it is obvious that:

$$\sup_{f\in\mathcal{A}_\nu} | Y(f)-\tilde{X}(f) | \leq \sum_{j=0}^{Nd-1} D_j \tag{4.5}$$

The control of D_0 and $D_j(f)$ 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 1).

Let us define the random vectors $\tilde{T}^0_{j,p}$ and T^0_j by :

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

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

LEMMA 3. *There exists an universal positive constant c_1 and a summable sequence $(\alpha_j)_{j\geq 0}$ of positive numbers each bounded by 1/2 such that:*

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

and

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

Now, Lemma 3 and property 3.1 of the construction allow us to prove exponential bounds on the r.v.'s $D_j(f)$. The proof is carried out at the end of the section.

PROPOSITION 1. *There exist a positive 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,*

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

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

$$pr\big(\mid D_j(g) \mid \geq c_2\sqrt{v}(\psi^{-1}(2^j)t + \varphi(2^j)u)\big) \leq$$
$$4k\exp\big(t^2(1+v^{-1/2}t)^{-1}\log\beta_j\big) + 4k\exp\big(-2u^2\log(1+j)\big) \qquad 4.6.(b)$$

Now, we give some consequences of the above proposition. An immediate consequence of (a) is:

$$D_0 = 0(\varphi(n)) \text{ a.s. and } D_0 = 0_P(\psi^{-1}(n)). \tag{4.7}$$

We now use (b) of Proposition 1 to conclude the proof of theorem 1.

<u>Proof of theorem 1.</u>

In order to prove Theorem 1, one can choose $\psi(x) = x^r$. Now, by proposition 1, with $u = t^2(1 + tv^{-1/2})^{-1}$, there exists some positive constant c_3 such that:

$$pr(\mid D_j(g) \mid \geq c_3 2^{j/r}(1 + u\sqrt{v} + u^2) \leq 2e^{-u^2} \tag{4.8}$$

We now apply 4.8. with $g = \Pi_j f$. Then, by lemma 1, one can choose $v = 2K2^{(Nd-j)(d-\delta)/d}$. Hence, by 4.4., there exists some constant c_4 such that, for any f in $\mathcal{A}_\nu$,

$$\begin{aligned} pr(\sup_{f \in \mathcal{A}_\nu} \mid Y(f) - \tilde{X}(f) \mid \ &\geq c_4(\nu^{d/r}(1 + u^2) + \nu^{(d-\delta)/2)}u)) \\ &\leq Nd \sharp \mathcal{A}_\nu \exp(-u^2) \end{aligned} \tag{4.9}$$

To complete the proof, it suffices to control the cardinality of $\mathcal{A}_\nu$. Now, by definition of $\mathcal{A}_\nu$,

$$\sharp \mathcal{A}_\nu \leq \sum_{p=1}^{\nu} \sharp(p^{-1}\mathbb{Z}^d \cap S)$$

Now, recall that S is a V-C class of subsets of the unit cube I^d. Hence, the control of the cardinality of $\mathcal{A}_\nu$ follows from the following combinatioral lemma, namely Sauer's lemma (cf. Assouad (83: Lemma 1.8., and 1.9.) for a proof).

LEMMA. Let S be a Vapnik-Chervonenkis class of subsets of I^d with density D. For any subset A of I^d with cardinality m, the cardinality of $A \cap S$ is no more than m^D.

So, by Sauer's lemma, for each p,

$$\sharp(p^{-1}\mathbb{Z}^d \cap S) \leq p^{dD(S)}$$

From which it follows that there exists some positive s such that $Nd\sharp\mathcal{A}_\nu \leq n^s$. Now, the end of the proof is straightforward, using 4.9 with $u^2 = (s+2)\log n$ and Borel-Cantelli Lemma.

Proof of Theorem 3.

We shall prove only (b) and (c). The proof of (a) will be omitted, using the same arguments. In order to prove theorem 3, it will be necessary to use the entropy properties of V-C classes of sets. So, we recall some well-known results on V-C classes.

Let P be a probability law on $\mathbb{R}_+^d$ and let $\mathcal{V}$ be the V - C class already defined (cf. 2.3.). Take on $\mathcal{V}$ the usual pseudo-metric d_p associated with P (i.e. for any (S, S') in $\mathcal{V} \times \mathcal{V}$, $d_P(S, S') = P(S \triangle S')$, and let $\mathcal{N}(\epsilon, \mathcal{V}, P)$ denote the minimal cardinality of a collection $\mathcal{V}(\epsilon)$ of elements of $\mathcal{V}$ such that for any S in $\mathcal{V}$ there exists $S(\epsilon)$ in $\mathcal{V}(\epsilon)$ with $d_P(S, S(\epsilon)) \leq \epsilon$. Then *log* $\mathcal{N}(\epsilon, \mathcal{V}, P)$ is called a metric entropy. When $\mathcal{V}$ is a V-C class, the following nice result holds :

LEMMA (Dudley (78)). *There exists a constant C, depending only on $D(\mathcal{V})$ such that for any P, for $0 < \epsilon \leq 1/2$,*

$$\mathcal{N}(\epsilon, \mathcal{V}, P) \leq C\left(\epsilon^{-1} \mid \log \epsilon \mid\right)^{D(\mathcal{V})}$$

(More about this Lemma is given in Assouad (83 sec. 4)).

Now, from the assumption 2.3 and from the definition of $\mathcal{A}_\nu$, is a Vapnik-Chervonenkis class of subsets of $]0, n]$ with furthermore $D(\mathcal{A}_\nu) \leq D(\mathcal{V})$. So, when P is the uniform distribution on $]0, n]$, Dudley's Lemma yields : for each j in $]0, Nd[$, there exists a finite net $\mathcal{F}_j$ of elements of $\mathcal{A}_\nu$ such that:

(4.10) (i) $\sharp\mathcal{F}_j \leq C\left(2^{Nd-j}(Nd-j)\right)^{D(\mathcal{V})}$

(ii) For each $A \in \mathcal{A}_\nu$, there exists $A' \in \mathcal{F}_j$ such that $\sharp(A \triangle A') \leq 2^j$

Now, let define the subset $\mathcal{U}_{j,N}$ of $\mathcal{A}_\nu \times \mathcal{A}_\nu$ by :

$$\mathcal{U}_{j,N} = \left\{(A, A') \in \mathcal{A}_\nu \times \mathcal{A}_\nu : \sharp(A \triangle A') \leq 2^j\right\}.$$

Clearly,

$$D_j \le \sup_{f\in\mathcal{F}_j} |D_j(\Pi_j f)| + \sup_{(f,g)\in\mathcal{U}_{j,N}} |D_j(\Pi_j(f-g))|. \tag{4.11}$$

Now, for any function g, by definition of $D_j(.)$

$$D_j(\Pi_j g) = \sum_{p=1}^{2^{Nd-j}} \left(\gamma_{j,p}(g)(\tilde{V}_{j,p} - \tilde{U}_{j,p}) - 2^{-j}(g \mid e_{j,p})\xi^j(e_{j,p})\right) + \xi^j(g)$$

Now, if (f,g) is an element of $\mathcal{U}_{j,N}$, it is easily seen that :
$\sum_p |\gamma_{j,p}(f-g)| \le 1$ and $\sum_p |2^{-j}(g-f \mid e_{j,p})| \le 1$. From which it follows that :

$$\begin{aligned}\sup_{(f,g)\in\mathcal{U}_{j,N}} D_j(f-g) \;&\le \max_{0<p<2^{Nd-j}} \left|\tilde{U}_{j,p} - \tilde{V}_{j,p}\right| \\ &+ \max_{0<p<2^{Nd-j}} \left|\xi^j(e_{j,p})\right| + \sup_{(f,g)\in\mathcal{U}_{j,N}} \left|\xi^j(f-g)\right|\end{aligned} \tag{4.12}$$

It remains now to control each of the random variables defined in 4.11 and 4.12. Theorem 3 will be a consequence of proposition 1 (b), and of the following oscillation control for the random process $(\xi^j(f) : f \in \mathcal{A}_\nu)$.

PROPOSITION 2. *With probability 1,*

$$\sup_{(f,g)\in\mathcal{U}_{j,N}} \left|\xi^j(f-g)\right| = O\big(\psi^{-1}(2^j)2^{(Nd-j)(d-\delta)/2d}(Nd-j)\big)$$

Before proving proposition 2, we complete the proof of Theorem 3. First we note that:

$$\tilde{V}_{j,p} - \tilde{U}_{j,p} = D_j(\tilde{e}_{j,p}) - \xi^j(\tilde{e}_{j,p}) \text{ and } \xi^j(e_{j,p}) = D_j(e_{j,p}).$$

So, by Lemma 2 and proposition 1 (b) applied with $v = v_j = 2K2^{(Nd-j)(d-\delta)/d}$ (here $K \ge 1$), and by 4.11 and 4.12, for any positive u and t

$$\begin{aligned} pr\big(D_j - \sup_{(f,g)\in\mathcal{U}_{j,N}} \left|\xi^j(f-g)\right| \;&\ge 4c_2 v_j^{1/2}(\psi^{-1}(2^j)t + \varphi(2^j)u)\big) \\ &\le 4k\left(\sharp\mathcal{F}_j + 3.2^{Nd-j}\right)\left(\beta_j^{t^2(1+v_j^{-1/2}t)^{-1}} + (1+j)^{-2u^2}\right)\end{aligned} \tag{4.13}$$

Now, from 4.10 (c), there exist c_5 and s such that :

$$\sharp\mathcal{F}_j + 2^{1+Nd-j} \leq c_5 2^{(Nd-j)s}$$

So, there exists $c_6 > 0$ such that, when $t = u = (Nd-j)^{1/2}c_6$

$$\left(\sharp\mathcal{F}_j + 2^{1+Nd-j}\right)\left(\beta_j^{t^2(1+v_j^{-1/2}t)^{-1}} + (1+j)^{-2u^2}\right) \leq c_6\left(\beta_j + (1+j)^{-2}\right)\Big)^{Nd-j}$$

Now, the bounds given in the above inequality are summable on (j, N) : hence, by inequality 4.13 and by Borel-Cantelli Lemma,

$$D_j - \sup_{(f,g)\in\mathcal{U}_{j,N}} \left|\xi^j(f-g)\right| = 0\left(\varphi(2^j)(2^{(Nd-j)(d-\delta)/d}(Nd-j))^{1/2}\right) \ w.p.1 \qquad (4.14)$$

Moreover, there exists $c_7 > 0$ such that, if $t = u\sqrt{log(1+j)} = c_7\sqrt{Nd-j+x}$,

$$\left(\sharp\mathcal{F}_j + 2.2^{Nd-j}\right)\left(\beta_j^{t^2(1+v_j^{-1/2}t)^{-1}} + (1+j)^{-2u^2}\right) \leq 2e^{j-Nd-x.} \qquad (4.15)$$

Hence, by inequality 4.13 and 4.15, we have :

$$\sum_{0<j<Nd}\left(D_j - \sup_{(f,g)\in\mathcal{U}_{j,N}} \mid \xi^j(f-g) \mid\right) \qquad (4.16)$$
$$= O_P\left(\sum_{0<j<Nd}\left(2^{(Nd-j)(d-\delta)/d}(Nd-j)\right)^{1/2}\psi^{-1}(2^j)\right).$$

Now, recall that there exists $\gamma > (d-\delta)/2d$ such that $x^{-\gamma}\psi^{-1}(x)$ is nondecreasing. Hence, there exists some $\epsilon > 0$ such that, for any j in $]0, Nd[$,

$$2^{(Nd-j)(d-\delta)/2d}(Nd-j)\psi^{-1}(2^j) \leq 2^{(j-Nd)\epsilon}\psi^{-1}(2^{Nd}).$$

So, by lemma 3, 4.7, 4.13, 4.16 and the above inequality, (b) of theorem 2 holds. Moreover, when $x^{-1/r}\varphi(x)$ is nonincreasing, by lemma 3, 4.7, 4.14, 4.16, and proposition 2, (c) of theorem 2 holds. It remains now to prove proposition 2.

Proof of proposition 2.

Notation. Let $p_j = pr(|X_1| > M_j)$ and let $a_j = 2^j p_j$. From the integrability assumptions on Q, it follows that $(a_j)_{j>0}$ is a summable sequence of positive numbers each bounded by $1/2$. Moreover, by definition of $\bar{\xi}_i^j$,

$$| \bar{\xi}_i^j | \leq M_{j+1} \text{ a.s. and } pr\left(\bar{\xi}_i^j \neq 0\right) \leq 2^{-j} a_j. \tag{4.18}$$

Hence, for each (f,g) in $\mathcal{U}_{j,N}$, $\left|E(\bar{\xi}^j(f-g))\right| \leq a_j \psi^{-1}(2^{j+2})$.

Now let define the sequence $(b_i^j)_{i>2^j}$ of independent r.v's by $b_i^j = \mathbb{1}_{(\bar{\xi}_i^j \neq 0)}$ and let define the class $\mathcal{A}_{N,j}$ by

$$\mathcal{A}_{N,j} = \{A \Delta A' : (A, A') \in \mathcal{U}_{j,N}\}.$$

From the above remarks, it follows that

$$\sup_{(f,g)\in \mathcal{U}_{j,N}} \left|\xi^j(f-g)\right| \leq \psi^{-1}(2^{j+2})\Big(1 + \sup_{A\in\mathcal{A}_{N,j}} \sum_{i\in A} b_i^j\Big). \tag{4.19}$$

So, proposition 2 is a consequence of the following lemma :

LEMMA 5. *There exists a positive constant* c_8 such that :

$$pr\Big(\sup_{A\in\mathcal{A}_{N,j}} \sum_{i\in A} b_i^j \geq c_8(Nd-j)2^{(Nd-j)(d-\delta)/2d}\Big) \leq c_8 a_j^{Nd-j}.$$

Proof : by 4.18 and by definition of $\mathcal{A}_{N,j}$, for each A in $\mathcal{A}_{N,j}$,

$$pr(b_i^j \neq 0 \textit{ for some } i \in A) \leq 1/2.$$

So, by Levy's symmetrization inequality (cf. Pollard p. 14 for a proof in our context) if $(B_i)_{i>2^j}$ denotes a sequence of independent r.v's with Bernouilli law $B(2^{-j})$, if $(\epsilon_i)_{i>2^j}$ denotes a sequence of independent symmetric r.v's, with values in $\{-1,0,1\}$ such that $pr(\epsilon_i \neq 0) = 2^{1+j} pr(b_i^j = 0)$, and if the sequences $(B_i)_i$ and $(\epsilon_i)_i$ are independent, then,

$$pr\left(\sup_{A\in\mathcal{A}_{N,j}} \sum_{i\in A} b_i^j \geq t\right) \leq 2\, pr\left(\sup_{A\in\mathcal{A}_{N,j}} \sum_{i\in A} B_i\epsilon_i \geq 1+t\right) \tag{4.20}$$

Now, let B denote the random set of integers i such that $B_i = 1$; for convenience, we also set :

$$R_j = \sup_{A\in\mathcal{A}_{N,j}} \sum_{i\in A} \epsilon_i B_i.$$

By definition of the r.v's ϵ_i, for each A in $\mathcal{A}_{N,j}$

$$\log E\Big(\exp(t \sum_{i \in A \cap B} \epsilon_i) \mid B\Big) \leq \sharp(A \cap B) 2a_j(ch\ t - 1). \tag{4.21}$$

In order to control the conditional Laplace of R_j given B, it will be necessary to give an upper bound for the r.v. $\sup_{A \in \mathcal{A}_{N,j}} \sharp(A \cap B)$. Now, recall that each $A \in \mathcal{A}_{N,j}$ is the symmetric difference of two elements of $\mathcal{A}_N$, with furthermore $\sharp A \leq 2^j$. Hence, from Lemma 2 it follows that for each A in $\mathcal{A}_{N,j}$,

$$\sharp \left\{p \in]0, 2^{Nd-j}[: I_{j,p} \cap A \neq \emptyset\right\} \leq v_j \tag{4.22}$$

(recall $v_j = 2K2^{(Nd-j)(d-\delta)/d}$).

Now, let $n(j) = \max_{0<p<2^{Nd-j}} \sharp(I_{j,p} \cap B)$; by inequalities 4.21, and 4.22, conditionally given B,

$$\log E(\exp(t \sum_{i \in A \cap B} \epsilon_i) \mid B) \leq 2.v_j n(j) a_j (ch\ t - 1)$$

Now, recall that $\mathcal{A}_\nu$ is a V-C class of sets with $D(\mathcal{A}_\nu) \leq D(\mathcal{V})$. Let $D = D(\mathcal{V})$: by definition of $\mathcal{A}_{N,j}$, for any subset B of $]0, 2^{Nd}]$,

$$\sharp(\mathcal{A}_{N,j} \cap B) \leq (\sharp(\mathcal{A}_\nu \cap B))^2 \leq (\sharp B)^{2D}.$$

For convenience, we set $x(t) = \exp\big(a_j v_j (ch\ t - 1)\big)$. Using the above remarks, it is straightforward to prove that, for each positive t,

$$E(\exp(tR_j) \mid B) \leq (2^{Nd-j} n(j))^{2D} x(t)^{n(j)}. \tag{4.23}$$

It remains then to control $E(n(j)^{2D} x(t)^{n(j)})$. Let $n_{j,p} = \sharp(B \cap I_{j,p})$. By definition of $n(j)$, for each $x \geq 1$,

$$n(j)^{2D} x^{n(j)} \leq \sum_{0<p<2^{Nd-j}} \frac{\partial^{2D}}{\partial x^{2D}} \left(x^{2D+n_{j,p}}\right).$$

Since the r.v.'s $n_{j,p}$ have the binomial distribution $B(2^j, 2^{-j})$, we infer :

$$E(\exp tR_j) \leq 2^{C(D)+(1+2D)(ND-j)+2(\log\ 2)^{-1} x(t)}.$$

Hence, setting $t = t_j = Argch(1 + (2a_j v_j)^{-1})$, we get :

$$E(\exp t_j R_j) \leq 2^{C(D)+(1+2D)(ND-j)} e^{2e}. \tag{4.24}$$

Then by Markov's inequality applied to $exp(t_j R_j)$,

$$pr(R_j \geq t_j^{-1}(Nd - j)(1 + 2D)log(2a_j^{-1})) \leq \ e^{2e} 2^{C(D)} a_j^{Nd-j}. \tag{4.25}$$

Noting that $\log(2a_j^{-1}) = 0(Argch(2a_j^{-1}))$ and that, for any $0 < a \leq b$, $(argch\ b)^{-1} Argch\ a \leq \sqrt{ab^{-1}}$, it is easy to see that $t_j^{-1} \log(2a_j^{-1}) = O(v_j^{1/2})$.

So, by 4.25, Lemma 5 holds: hence, proposition 2 is proved, and the proof of Theorem 3 is complete. Now, we prove proposition 1.

Proof of proposition 1.

We prove only (b). The proof of (a) will be omitted, using the same arguments. In order to prove proposition 1, we need the following large deviation lemma, which is due to Massart (89). Let a and r be positive reals and let $\bar{H}(a, r)$ be the class of random variables Z such that

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

We denote by $H(a, r)$ the class of the random variables $Z - EZ$, with Z in $\bar{H}(a, r)$.

LEMMA. 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 $0 < t < a$:

$$\log\big(E(\exp(tT(w))\big) \leq v(t/a)^2, \tag{a}$$

and, for any positive u,

$$pr(a \mid T(w) \mid \geq u) \leq 2exp(-u^2(4v + u)^{-1}). \tag{b}$$

(cf. Massart (89) for a proof).

Remark. (b) follows from (a) via the classical Cramer-Chernoff calculation.

Let us now introduce the following notations and definitions. We define the random sequence $\bar{X}^l$ by $\bar{X}^l = \mathbb{1}_{2^l(2\mathbb{N}+1)}\bar{X}$ and the random sequence $\tilde{X}^l$ from the already defined sequence $\bar{X}^l$ by: $\tilde{X}_i^l = \bar{X}_i^l - E(\bar{X}_i^l)$ for any natural i. We also set $\tilde{V}_{j,p}^l = Y^l(\tilde{e}_{j,p})$, and for any $l \in]0, j-1]$,

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

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

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

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

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

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

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

Now, we set

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

and

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

By definition of $D_j(g)$,

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

(b) of proposition 1 follows from classical Cramer-Chernoff calculations. So, we have to bound the Laplace transforms of the above random vectors. First, we control the

Laplace transform of $D_{j,2}(g)$. By convexity of $x \to \exp(tx)$ and by Massart's lemma, for any sequence $(u_l)_l$ of positive numbers such that $u_0 + \cdots + u_{j-1} = 1$, we have:

$$E\Big(\exp(t \mid D_{j,2}(g))\Big) \leq \sup_{0 \leq l < j} E\Big(\exp(tu_l^{-1} \mid D_{j,2}^l(g))\Big)$$

Now, recall that the sequences $(\tilde{T}_{j,p}^l)_{p>0}$ and $(\tilde{T}_{j-l,p}^0)_{p>0}$ are identically distributed, and the random vectors $\tilde{T}_{j,p}^l$ are independent elements of $H(c_1 \log \alpha_j^{-1} / \psi^{-1}(2^j)\ , 3)$. Furthermore, since $\mathcal{B}$ is a dyadic orthogonal basis of $l^2(]0,n])$, the reals numbers $\gamma_{j,p}(g)$ are each bounded by 1, and $\sum_p \gamma_{j,p}^2(g) \leq v$. So, by Massart's lemma for each l in $[0, j[$, we have:

$$\log E\Big(\exp(tu_l^{-1} \mid D_{j,2}^l(g))\Big) \leq v(b_{j-l}u_l)^{-2} \mid t \mid^2$$

for any $\mid t \mid < b_{j-l}u_l$, where $b_m = c_1 \log \alpha_m^{-1}/\psi^{-1}(2^m)$, for any natural m. Let $s = r^{-1}$. By 1.2, we have:

$$b_{j-l}u_l \geq c_1(\psi^{-1}(2^j))^{-1}2^{ls}u_l|\log \alpha_{j-l}|$$

Now, one can choose $u_l = u_0(1+l)2^{-ls}$. Then, setting $\beta_j = \sup_{l \leq j} \alpha_{j-l}^{l+1}$, we obtain:

$$\log E\Big(\exp(t \mid D_{j,2}(g))\Big) \leq cv \mid t \mid^2 (\psi^{-1}(2^j)/|\log \beta_j|)^2$$

for any t fulfilling $\mid t \mid \psi^{-1}(2^j) \leq \beta_j$, where c is a positive constant. Clearly, the so defined sequence $(\beta_j)_{j>0}$ fulfills the conditions of proposition 2. Moreover, from (b) of Massart's lemma, it follows that there exists a constant c_5 such that, for any positive real t,

$$pr(\mid D_{j,2}(g) \mid \geq c_5\sqrt{v}\psi^{-1}(2^j t)) \leq 2k \exp(t^2(1 + v^{-1/2}t)^{-1} \log \beta_j) \qquad (4.26)$$

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,

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

So, using again the same arguments, it will be enough to prove that, for each $j > 0$, the random vectors $(\bar{\xi}_i^j)_{i>0}$ are independent and such that:

$$E\big(\exp(c_6 |\log a_j|\, |\bar{\xi}_i^j|/\psi^{-1}(2^j)\big) \le 1 + 2^{-j} \tag{4.27}$$

Clearly, the random vectors $\bar{\xi}_i^j$ are independent. Moreover, by 4.18,

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

and 4.27. follows immediately from the above inequality. So, we have proved the following result: there exists some positive constant c_7 and a summable sequence $(\beta_j)_{j>0}$ fulfilling the conditions of proposition 1 such that, for any positive t,

$$pr(|\, D_j(g) - D_{j,1}(g)\, | \ge c_7\sqrt{v}\psi^{-1}(2^j)t) \le 4k \exp\big(t^2(1 + v^{-1/2})^{-1} \log \beta_j\big) \tag{4.28}$$

It remains to bound the Laplace transform of $D_{j,1}(g)$. First, we note that, for each l, $D_{j,1}^l(g)$ is a centered gaussian random vector fulfilling:

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

Now, it can easily be seen (cf. Einmahl (87)) that:

$$\|\Delta_j\| = O\Big(2^{-j}(\psi^{-1}(2^j))^2\Big) \text{ for any } i > 0$$

So, by convexity of $x \to \exp(tx)$, for any sequence $(u_l)_l$ of positive numbers such that $u_0 + \cdots + u_{l-1} = 1$, we have:

$$\log\big(E(\exp(t \mid D_{j,1}(g)))\big) \le c_8 v \mid t \mid^2 \max_{0<l\le j}\big(u_{j-l}^{-1} 2^{-l}(\psi^{-1}(2^l))^4\big)$$

Now, by definition of φ, $(\psi^{-1}(x))^4/x \le (\varphi(x))^2/LLx$. Moreover, from (iii) of 1.2 it follows that there exists some positive s such that, for any $0 < t < 1$, $\varphi(xt) \le t^s\varphi(x)$. So, if one choose $u_l = 2^{-ls}u_0$, then there exists a positive constant c_9 such that:

$$\log\big(E(\exp(t \mid D_{j,1}(g)))\big) \le c_9 v \mid t \mid^2 \big((\varphi(2^j))^2/\log(1+j)\big) \tag{4.29}$$

So, using the classical Cramer-Chernoff calculation, we get:

$$pr\big(D_{j,1}(g) \ge c_{10}\sqrt{v}\ \varphi(2^j)u\big) \le 2k \exp(-2u^2 \log(1+j)) \tag{4.30}$$

We complete the proof of proposition 1 by collecting the inequalities 4.28 and 4.30

5. *LOWER BOUNDS FOR THE APPROXIMATION OF PARTIAL-SUM PROCESSES.*

In this section, starting from a paper of Beck (85) on lowers 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, $\mathcal{S}$ shall be the class

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

which was previously used by Beck (85).

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

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

From a result of Bartfai (cf. Major (78) for a proof in a more general context), 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 define M_ν by:

$$M_\nu = \sup_{S\in\mathcal{S}} \left| \sum_{i\in\nu S} (X_i - Y_i) \right|$$

Throughout this section, $| \ . \ |$ denotes the Euclidian norm on $\mathbb{R}^d$. We define an empirical measure μ associated with the arrays $(X_i)_i$and $(Y_i)_i$ by :

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

As Beck does, we set :

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

for any ρ in $[0,1]$ (note that $h(\rho,t)$ depends only on ρ and $| \, t \, |$). Using 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 (85: inequality (29)) $h(1,t) \geq c^2\nu^{d-1}h\left(\frac{1}{2\nu},t\right)$, for any t in $\mathbb{R}^d$, for some positive universal constant c. Hence, we have:

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

From which it follows that there exists a ball $B(x,r)$ with radius $r \in [\frac{1}{2},1]$ such that :

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

Now, by definition of W(F,G), $E\left((X_i - Y_i)^2\right) \geq W^2(F,G)$ for any i. So, the first part of Theorem 2 holds.

Let F_n (*resp.* G_n) be the empirical distribution function of $(X_i)_{i\in]0,\nu]^d}$ (*resp.* $(Y_i)_{i\in]0,\nu]^d}$). By Bartfai's result,

$$M_\nu^2 \geq c^2\nu^{d-1}\int_0^1 (F_n^{-1}(u) - G_n^{-1}(u))^2 du \ \ a.s.$$

We complete the proof by collecting Glivenko-Cantelli theorem, the above inequalities and standard arguments of measure theory.

Remark. We are unable to prove such a result when S is the class of lower-left orthants. This is not surprising in view of Beck's work.

APPENDIX 1 : 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:

- Y is a $\mathcal{N}(0, I_s)$-distributed random vector
- 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 Q_1 denote the distribution function of X_1, and for each positive integer l, let $Q_l(\,.\, \mid X_1, \ldots, X_{l-1})$ denote the conditional distribution function of X_l, for given $(X_1, \ldots, X_{l-1})$. We define the random vector $U = (U_1, \ldots, U_s)$ from $(X_1, \ldots, X_l)$ by:

$$U_1 = Q_1(X_1) \text{ and, for each } l > 1, \qquad U_l = Q_l(X_l \mid X_1, \ldots, X_{l-1})$$

Clearly, U_l is a measurable function of $(X_1, \ldots, X_l)$. Moreover, if we assume that Q is absolutly 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 an 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 fulfills the conditions (i) and (ii). Henceforth, we call the so defined transformation multivariate quantile transformation. The conditional transformation of $(X_l, \ldots, X_s)$ given $(X_1, \ldots, X_{l-1})$ is called multivariate conditional quantile transformation of $(X_l, \ldots, X_s)$ given $(X_1, \ldots, X_{l-1})$. Now, we recall some recents 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 variables, via Rosenblatt's transformation. However, we need to modify slightly his main result.

We consider a sequence $\xi_1, \ldots, \xi_m$ of independent mean zero random vectors with values in $\mathbb{R}^s$ such that:

$$Var(\xi_1 + \cdots + \xi_m) = V_m I_s$$

and, for any $p \leq m$,

$$E\big(\exp |t\xi_p|\big) < +\infty \text{ for any } t \in \mathbb{R}.$$

One can choose $\alpha_0 > 0$ in a way such that:

$$\sum_{p=1}^{m} \alpha_0 E(|\ \xi_p\ |^3 \exp(\alpha_0\ |\ \xi_p\ |) = V_m \tag{a.1}$$

Then, we set $\alpha = \alpha_0 \wedge (1/2)$.

THEOREM 7 (Einmahl (89) sect. 3) : Let $\xi_1, \ldots, \xi_m)$ be mean zero 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 $N(0, vB_m I_s)$ 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 such that, if $V_m \geq C(v)\alpha^{-2}$, the following holds true:

$$|\ S_m - \sqrt{V_m} Y\ | \leq C(v)\alpha^{-1}\Big(\frac{|\ S_m\ |^2}{V_m} + 1\Big). \tag{a.2}$$

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

Comments. Our version of theorem 7 is a corollary of Einmahl's theorem. Note that Einmahl had to assume $V_m \geq c\alpha^{-2}\log\alpha^{-1}$ (cf. Einmahl 89, p.43) but this condition comes from $\exp(-\frac{3}{8}c^2\alpha^2 V_m) \leq \beta_m$: here, $\beta_m = 1/(2\alpha\sqrt{V_m})$. Hence, a.1. still holds when $m \geq c\alpha^{-2}$. Moreover, the condition $v^{(s+1)/2} \geq Q_m(2c\alpha)$ in Einmahl's result is ensured by $Q_m(2c\alpha) \leq -V_m\alpha^2 c$.

Now, from theorem 7, 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 $V_m \geq C(v)\alpha^{-2}$,

$$|\ T_m\ | \leq C(v)\alpha^{-1}\Big(1 + \frac{|\ S_m\ |^2}{V_m}\Big)\mathbb{1}_{(|S_m| \leq C(v)\alpha V_m)} + |\ T_m\ |\ \mathbb{1}_{(|S_m| > C(v)\alpha V_m)}$$

Integrating by parts (cf. Massart 89 : sect. 3, Lemma 4 (ii)) we obtain that $m^{-1}|S_m|^2 \mathbb{1}_{|S_m| \leq C(v)\alpha V_m}$ belongs to $H(c_1, 3)$. Now, from the definition of α, it follows that

$$E\big(\exp(t\ |\ S_m)\big) \leq exp(2\ |\ t\ |^2)\ for\ any\ \ |\ t\ | \leq \alpha.$$

So, the classical Cramer-Chernoff calculation yields :

$$pr(\mid S_m \mid > C(v)\alpha V_m) \leq 2s\ exp(-c_2 V_m \alpha^2)$$

and, by Cauchy-Schwarz inequality, for any $t \leq \alpha/2$,

$$\begin{aligned} E(exp(t \mid S_m \mid \mathbb{1}_{\mid S_m \mid > C(v)\alpha V_m}) &\leq 1 + exp(-c_3 V_m \frac{\alpha^2}{2}) E(exp(2t \mid T_m \mid))^{1/2} \\ &\leq 1 + exp(V_m(8t^2 - c_3 \frac{\alpha^2}{2})) \end{aligned}$$

Hence, $\mid S_m \mid \mathbb{1}_{\mid S_m \mid > C(v)\alpha V_m}$ belongs to $\bar{H}(\frac{\alpha}{4}\sqrt{c_3}, 3)$, and using the convexity of $x \rightarrow e^{tx}$,

$$\mid T_m \mid \in \bar{H}(c\alpha, 3)\ for\ some\ c > 0 \tag{a.3}$$

Now, we prove (b) of lemma 3. The proof of (a) will be omitted, using the same arguments. From now on we work conditionally given B_j. For given B_j, one can write $(U^0_{j,p}, \tilde{U}^0_{j,p})$ as a sum of 2^{j-1} independent random vectors with values in $\mathbb{R}^{2k}$. now, we set:

$$W^0_{j,p} = \left(Var(U^0_{j,p} \mid B_j)\right)^{-1/2} U^0_{j,p} \text{ and } S_{j,p} = 2^{j/2}(W^0_{j,p}, \tilde{W}^0_{j,p})$$

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 fulfills the condition a.1 of Einmahl's theorem, we have:

$$\mid 2^{j/2}\tilde{W}^0_{j,p} - \tilde{V}_{j,p} \mid \in \bar{H}(c\alpha, 3) \tag{a.4}$$

Moreover, it is easily seen that for any B_j, there exists some symmetric matrix Γ depending on p and B_j, satisfying $I_{2k} \leq 4\Gamma \leq 16 I_{2k}$, and such that:

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

Hence, if ς_j is the positive number such that

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

then, the random variable $|2^{j/2}\tilde{W}^0_{j,p} - \tilde{V}_{j,p}|$ is an element of $\bar{H}(c_4\varsigma_j, 3)$, for some universal constant c_4 (here the constant v of theorem 7 satisfies $v \geq 1/4$). Now, the followimg lower bound on ς_j is the main technical tool for the proof of lemma 3.

LEMMA A.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:

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

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

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

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

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

By convexity of $x \to e^x$, it suffices to control each of the terms on right hand. First we note that the above defined matrix are elements of the commutative ring generated by $VarQ_j$. Hence, we have:

$$Var\tilde{W} = Var\tilde{U} - A\ CoV(U,\tilde{U})$$

Moreover, by definition of $(U,\tilde{U})$, $VarU \geq 2^{j-1}I_k$, and $CoV(U,\tilde{U}) = -\tilde{b}_{j,p}VarQ_j$. 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, $\|Var\tilde{W}\| \leq 3.2^j$. Hence, by a.5, we get:

$$|\,\tilde{W} - (Var\tilde{W})^{1/2}Y\,| \in \bar{H}(c_4\varsigma_j/3,\ 3) \tag{a.7}$$

On the other hand, the r.v. $T_2 = \big((2^{j/2}\Gamma_j - (Var\tilde{W})^{1/2}\big)Y$ 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 calculation proves that:

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

From which 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 $\bar{H}(c_6, 3)$. It remains to control the random vector AU. Let $V^1 = \sum_{i \in I_{j,p} \cap 2\mathbb{N}} Z_i^0$: 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, recalling that $b_{j,p}$ has a binomial law $B(2^{j-1}, p_j)$, it is easy to see that $|AU|$ belongs to the class $\bar{H}(-c_7 \log a_j / M_j, 3)$ for some universal positive constant c_7. Then, we complete the proof of lemma 3 by collecting the above inequalities and a.7. Now, we pass to the proof of a.5.

Proof of Lemma A.5.

First, we note that, for any positive t,

$$\alpha t^3 exp(\alpha t) \leq t^2(\exp(2\alpha t) - 1).$$

From which it follows that there exists some constant C_0 depending only on ψ such that:

$$\int_{\mathbb{R}^k} \alpha |x|^3 e^{|\alpha x|} d\bar{Q}_j(x) \leq$$

$$C_0 \Big(1 + \sum_1^j a_{l-1}(\exp(2\alpha M_l) - 1)\Big) \tag{4.8}$$

Hence, to prove lemma A.5, it suffices to prove that there exists some positive constants C_1 and C_2 such that:

$$\sum_1^j a_{l-1}(\alpha_j^{-C_1 M_l / M_j} - 1) \leq C_2 \tag{4.9}$$

Now, recall that there exists some $\epsilon > 0$ satisfying $M_l \leq M_j 2^{(l-j)\epsilon}$. Hence, there exists some positive constant C_3 such that:

$$\sum_1^j \sqrt{M_l M_j^{-1}} \leq C_3 \text{ and } \sqrt{M_l M_j^{-1}}(1+j-l) \leq C_3$$

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

$$\sum_1^j a_{l-1}(\alpha_j^{M_l/(C_3 M_j)} - 1) \leq \sum_1^j (1-a_l)\sqrt{M_l M_j^{-1}} \leq C_3$$

Hence, 4.9 holds true and the proof of a.5 is complete. ■

APPENDIX 2 : Proof of lemma 2.

By Kronecker's lemma, lemma 2 follows from :

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

So, by three-series lemma, it suffices to prove that :

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

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

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

Hence, we have:

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

By 1.2, there exists some positive constant c such that $\sum_{l\geq 0} 1/\psi^{-1}(i2^l) \leq c/\psi^{-1}(i)$. Moreover, it is well-known that the serie

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

is convergent if and only if $\int_{\mathbb{R}^k} \psi(|x|)dQ(x) < +\infty$. So, 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** (3) 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.

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

DUDLEY, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **7** 899-929.

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

EINMAHL, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Analysis* **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** 111-6131.

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 Analysis* **8** 487-517.

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

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

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

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 Series in Statistics.

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* 4-49. *Trudy Inst. Mat.* Vol. 3, "Nauka" Sibirsk.Otdel, Novosibirsk [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. Geb.* **3** 211-226.

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

CHAPITRE II

STRONG APPROXIMATION FOR SET-INDEXED PARTIAL SUM PROCESSES, VIA K.M.T. CONSTRUCTIONS II.

RIO Emmanuel
URA D 0 743 CNRS - Statistique Appliquée, Bât. 425 Mathématique
91405 ORSAY Cedex France

Summary : Let $(X_i)_{i\in\mathbb{Z}_+^d}$ be an array of independent identically distributed random vectors with values in $\mathbb{R}^k$, with mean zero and finite variance, and let $\mathcal{S}$ be a class of Borelians subsets of $[0,1]^d$. If, for the usual metric, $\mathcal{S}$ is totally bounded and has a convergent entropy integral, we obtain a strong invariance principle for an appropriately smoothed version of the partial sum process $\{\sum_{i\in\nu S} X_i : S \in \mathcal{S}\}$ with an error term depending only on $\mathcal{S}$ and on the queue of distribution of X_1. In particular, when $\mathcal{S}$ is the class of subsets of $[0,1]^d$ with α-differentiable boundaries introduced in Dudley (74), we prove that our result is optimal.

Résumé : Soit $(X_i)_{i\in\mathbb{Z}_+^d}$ un champ de vecteurs indépendants, équidistribués, centrés et à valeurs dans $\mathbb{R}^k$, et de variance finie et soit $\mathcal{S}$ une classe de Boreliens de $[0,1]^d$. Quand, pour la pseudométrique usuelle, $\mathcal{S}$ est totalement bornée pour l'inclusion et a une intégrale d'entropie convergente, on obtient un principe d'invariance fort pour une version régularisée du processus de sommes partielles $\{\sum_{i\in\nu S} X_i : S \in \mathcal{S}\}$ dans lequel le terme d'erreur dépend explicitement de la queue de distribution de X_1 et des caractéristiques entropiques et géométriques de la classe $\mathcal{S}$. En particulier, si $\mathcal{S}$ est la classe des ouverts de $\mathbb{R}^d$ dont les bords sont l'image par une application α-Hölderienne de la sphère de dimension $d-1$ introduite par R.M. Dudley en 1974, on montre que les résultats obtenus ne sont pas améliorables.

Code Matière A.M.S. (1980) - 60 F 17 - 62 G 99
Mots-clés : Central limit theorem, Set-indexed process, Partial-sum process, Invariance principle, Metric entropy with inclusion, Multivariate empirical processes.

1. INTRODUCTION

In this paper, we continue the research started in our previous paper I (cf. Rio (90)). So, the purpose of this paper is to etablish strong invariance principles for partial-sum processes indexed by a family of subsets of the unit cube $I^d = [0,1]^d$. As motivation and potential application for our results, we refer the reader to Pyke's work (84). The context of the problem is as follows. Let $(X_i)_{i \in \mathbb{Z}_+^d}$ be an array of independent identically distributed $\mathbb{R}^k$-valued random vectors with mean 0 and finite variance. For any subset B of $\mathbb{R}^d$, define:

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

If $\mathcal{S}$ is any collection of subsets of $[0,1]^d$ we define the normalized partial sum process $\{S_\nu(S)) : S \in \mathcal{S}\}$ by :

$$S_\nu(S) = \nu^{-d/2} \sum_{i \in \nu S} X_i.$$

Our aim is to study the rate of convergence in the strong invariance principle. The general approach is then analogous to that introduced by Komlos, Major and Tusnady (K.M.T.) for their results on the strong approximations of sums of i.i.d. random variables; cf. K.M.T. (75) I and II. In the spirit of Massart (89), the methods are then extended to obtain strong invariance principles for multidimensionally-indexed processes.

Now, let us recall some recents results on stong approximation. When $d = 1$, $\mathcal{S} = \{[0,t], 0 \le t \le 1\}$, Einmahl (87, 89) has recently obtained the following results via K.M.T. type constructions. Let ψ be some mapping satisfying:

(i) $E(\psi(|X_1|)) \le +\infty$

(ii) $x \to x^{-2}\psi(x)$ is non decreasing, and $x \to x^{-r}\psi(x)$ is nonincreasing for some $r > 2$.

and define $Lx = \log(x \vee e)$ and $LLx = L(Lx)$. If furthermore $x \to (x^2 LLx)^{-1}\psi(x)$ is non-decreasing, then a sequence $(Y_i)_{i \ge 1}$ of i.i.d. Gaussian random vectors may be constructed in such a way that denoting by T_ν the partial sum process associated with $(Y_i)_{i \ge 1}$,

$$\sup_{S \in \mathcal{S}} | T_\nu(S) - S_\nu(S) | = O(\psi^{-1}(\nu^d)\nu^{-d/2}) \quad a.s.$$

Morever, if $k = 1$ and if the moment-generating function of X_1 is finite in a neighborhood of 0, K.M.T (75) proved that

$$\sup_{S\in\mathcal{S}} | X_\nu(S) - Y_\nu(S) |= O(\nu^{-1/2} log\ \nu) \quad a.s.$$

It is worth noticing that the rates of strong approximation appearing above are optimal (this comes from Breiman's remark (67) when the r-th moment is finite and from Bartfai (66) when the moment-generating function is finite).

On the other hand, Massart (87) has obtained optimal rates in the strong invariance principle for unidimensional ($k = 1$) set-indexed partial-sum processes, when the class $\mathcal{S}$ is not too large (i.e. $\mathcal{S}$ is a Vapnik-Chervonenkis (V-C) class or $\mathcal{S}$ fulfills a suitable condition of entropy with inclusion). However, he had to assume the existence of the moment generating function of the r.v.'s X_i. Recently, we generalized Massart's results for V-C classes of sets to $\mathbb{R}^k$-valued random vectors under weaker moment conditions. Here, using the rather complicated K.M.T. type embedding previously introduced in our previous paper I, we study the rates of convergence in the strong invariance principle for partial-sum processes indexed by classes $\mathcal{S}$ of sets with an *integrable entropy with inclusion* $N_I(\varepsilon, \mathcal{S})$. However, it is necessary to consider a smoothed version of the partial sum process (cf. Alexander and Pyke (86)). Then, due to the optimality of our embedding, we obtain a uniform central limit theorem (CLT) and a strong Strassen invariance principle when only the second moment is assumed to be finite. In the same context Bass and Pyke (84) and Morrow and Philipp (86) had to assume very stringent conditions of moments. Considering the papers of Alexander and Pyke (86) on the uniform CLT, and Bass's work (85) on the law of the iterated logarithm, this is not surprising.

Now, we discuss further the scope of results: we need an extra condition on boundaries of elements of the class $\mathcal{S}$. Let λ be the Lebesgue measure on $\mathbb{R}^d$. Given a norm $| \, . \, |$ on $\mathbb{R}^d$ and a borelian subset S of $\mathbb{R}^d$, we set

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

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

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

If $\delta = 1$ this condition is the uniform Minkowsky condition previously used by Massart (89), Bass and Pyke (84). Let $H(\varepsilon)$ be the log-entropy (with inclusion) of $\mathcal{S}$ and define

$I(x) = \int_0^x (H(u)/u)^{1/2} du$. It is worth noticing that $I(x)$ is an upper bound on the modulus of continuity of the standard Brownian process indexed by $\mathcal{S}$ with covariance function $Cov(A,B) = \lambda(A\Delta B)$. When $\mathcal{S}$ fufills 1.1 and has an integrable entropy with inclusion and $E(\mid X_1 \mid^r) < +\infty$ for some large enough r, we obtain a rate of convergence in the strong invariance principle of the order of $I(\nu^{-\delta})$. Moreover, when $\mathcal{S}$ is the class of sets with α-differentiable boundaries introduced in Dudley (74), we even prove that this result is optimal (cf. sect. 4).

On the other hand, when $\mathcal{S}$ is not too large and the moment is closer to the second moment only, we prove that the rate of approximation is of the order of $o(\nu^{-d/2}\psi^{-1}(\nu^d))$ in probability. From Breiman's remark this result is optimal too. Now, we introduce notation and we state our results.

2. *DEFINITIONS AND RESULTS* .

Throughout this section, the probability space Ω is assumed to be so that there exists a random variable, defined on Ω, with uniform distribution over $[0,1]$, which is independent of the observations.

For any bounded Borelian subset B of $\mathbb{R}^d$, we define the smoothed partial sum $X(B)$ by:

$$X(B) = \sum_{i \in B} \lambda([i-1,i] \cap B) X_i$$

Let $\mathcal{S}$ be a family of Borel subsets of the unit cube I^d satisfying the smoothness condition 1.1 on the boundaries of elements of $\mathcal{S}$ for some $\delta > 0$. We define the class $\nu\mathcal{S}$ of subsets of $\mathbb{R}^d$ for any integer ν by :

$$\nu\mathcal{S} = \{\nu S : S \in \mathcal{S}\}$$

where $\nu S = \{\nu x : x \in S\}$. In order to get nice asymptotic properties for a normalized version of the smoothed partial-sum process $(X(\nu S) : S \in \mathcal{S})$, we need some reasonable growth conditions on $\nu\mathcal{S}$ when ν tends to infinity . Define the pseudometric d_λ by $d_\lambda(A,B) = \lambda(A\Delta B)$. We assume that $\mathcal{S}$ is *totally bounded with inclusion* and has a *convergent entropy integral* with respect to d_λ. It means that, first, for every positive ε there exists a finite collection (called an ε-net) $\mathcal{S}(\varepsilon)$ such that for any S in $\mathcal{S}$, there exists S^+ and S^- in $\mathcal{S}(\varepsilon)$ with $S^- \subset S \subset S^+$ and $d_\lambda(S^-, S^+) \leq \varepsilon$. And secondly, that the minimal cardinality of such a collection $\mathcal{S}(\varepsilon)$ which we denote by $N_I(\varepsilon, \mathcal{S})$ satisfies:

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

Define

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

We may, by enlarging the class $\mathcal{S}$ a little if necessary, assume that $H(\varepsilon) > |\log \varepsilon|$ for any $0 < \varepsilon < 1$. Yet, in order to get strong invariance principle when only the second moment of Q is assumed to be finite, we have to put the following condition on $\mathcal{S}$. So, $\mathcal{S}$ has to be supposed *contraction closed.* We recall that this means:

$$\text{for all } t \in [0,1[\text{ for all } S \in \mathcal{S},\ tS \in \mathcal{S} \tag{2.2}$$

In many cases of interest, there is no loss of generality in making this assumption, because if the approximating collections $\mathcal{S}(\varepsilon)$ fulfill the uniform condition 1.1. of smoothness of the boundaries with some constants δ and K, and if $\mathcal{S}^* = \{tS : S \in \mathcal{S},\ 0 \leq t \leq 1\}$ has log-entropy $H^*(\varepsilon)$, then it is obvious that $\mathcal{S}^*$ is totally bounded with inclusion and satisfies $H^*(\varepsilon) = O(H(\varepsilon))$.

Now, we state our main result, which provides an invariance principle with an error term depending only on the queue of distribution of the r.v.'s X_i and on the log-entropy of $\mathcal{S}$ in general. So, let ψ be a mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$ such that:

(2.3.) (i) $\int \psi(|x|)dQ(x) < +\infty$, and $x^{-2}\psi(x)$ is a one to one continuous, increasing mapping from $\mathbb{R}^+$ onto $\mathbb{R}^+$

(ii) There exists $r > 2$ such that $x^{-r}\psi(x)$ is nonincreasing.

(iii) Furthermore, if there does not exist $r < 4$ such that $x^{-r}\psi(x)$ is nonincreasing, then $(x^2 LLx)^{-1}\psi(x)$ is nondecreasing.

Throughout, ψ^{-1} denotes the inverse function of ψ. Note that, when Q has a finite second moment, such a mapping ψ exists (see Major (76)). We also need to introduce the following notations. Let $H_1(\varepsilon) = \varepsilon^{-1}H(\varepsilon)$ and let H_1^{-1} be the inverse of H_1 (i.e. $H_1^{-1}(x) = y \Leftrightarrow H(y^+) \leq xy \leq H(y^-)$ for any positive y). For every positive nondecreasing function f, let define:

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

Then, the following result is available:

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

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

and, setting

$$\varphi(x) = \psi^{-1}(x)(1 + (x^{-1}LLx)^{1/2}\psi^{-1}(x))$$

we have:

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

Comments. When $E(|X_1|^2 LL(|X_1|)) < +\infty$, there exists a mapping ψ in such a way that $(x^2LLx)^{-1}\psi(x)$ is nondecreasing. Then, it is obvious that $\varphi(x) = 0(\psi^{-1}(x)$ as $x \to +\infty$ and (b) of Theorem 1 holds true with $o(b(\nu^d, \psi^{-1}))$ a.s.

Let $y = H_1^{-1}(u)$. Clearly, $\sqrt{u}H_1^{-1}(u) = \sqrt{yH(y)} \leq I(y)$ Now, note that, necessarily $H(\varepsilon) = O(\varepsilon^{-1})$ as $\varepsilon \to 0$. Hence, $H_1^{-1}(u) = O(u^{-1/2})$ as $u \to +\infty$. So, when $d > 1$ or $\delta \leq 1/2$ it is obvious that

$$H_1^{-1}(u) + u^{-1/2}I(u^{-\delta/d}) = O(I(u)) \text{ as } u \to +\infty$$

From (b) of Theorem 1, it follows that the uniform central limit theorem holds as soon as

$$\int_1^{+\infty} \psi^{-1}(t)t^{-3/2}dt < +\infty.$$

For example, Theorem 1 provides an uniform central limit theorem and a strong invariance principle when $E(|X_1|^2(L|X_1|)^{2+\epsilon}) < +\infty$ for some positive ϵ. So, our result is not optimal in the general case: in order to get optimal results, we need to put additional conditions on S and on the law Q. Throughout, we assume that either Q has a finite r-th moment for some $r > 2$ or S has an entropy exponent ς in $[0,1[$. We recall that this means:

$$\limsup_{\varepsilon\to 0} \mid \log(H(\varepsilon)/\log\varepsilon \mid = \varsigma$$

Let us give an example of such a class. If $d = 1$ and if $\mathcal{A}(\delta)$ is the class of Borelian subsets of the unit interval satisfying 1.1 for some $1/2 < \delta < 1$ and for some positive K, $\mathcal{A}(\delta)$ is contraction closed and it is easily seen that $\mathcal{A}(\delta)$ is totally bounded with inclusion and that its log-entropy $H_\delta(\,.\,)$ satisfies:

$$H_\delta(\varepsilon) = O(\varepsilon^{1-1/\delta}|\log\varepsilon|) \tag{2.4}$$

Hence, $\mathcal{A}(\delta)$ has the entropy exponent $\varsigma = \delta^{-1} - 1$. From this remark, it follows that, for any positive integer d,for any class S of Borelian subsets of the unit cube I^d satisfying 2.1, 2.2, 1.1 for some $0 < \delta \le 1$ and having an exponent of entropy ς in $[0,1[$, the following result holds true: for any $s > \varsigma$,

$$H_1^{-1}(u) + u^{-1/2}I(u^{-\delta/d}) = O(u^{(d-\delta(1-s))/2d}) \text{ as } u \to +\infty$$

So, if S has an exponent of entropy ς in $[0,1[$ and if the moment of Q is between 2 and $2d(d-\delta(1-\varsigma))^{-1}$, the rate of approximation does not depend on the class S anymore. Then, the following corollary generalizes Einmahl's results (87 and 89) to the multidimensional case.

COROLLARY 1. *Let S be a family of Borel subsets of $[0,1]^d$ fulfilling 2.1 and 1.1 for some $0 < \delta \le 1$. Assume S to be totally bounded with inclusion and to have an exponent of entropy ς in $[0,1[$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance, and let ψ be a mapping fulfilling 2.3 for some $r < 2d(d-\delta(1-\varsigma))^{-1}$. 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* $N(0, VarQ)$*-distributed random vectors such that:*

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

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

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

Comments. Clearly, this is sufficient to obtain a construction of the arrays such that (a) and (b) hold with respective rates $O_P(\psi^{-1}(\nu^d))$ and $O(\varphi(\nu^d))$ *a.s.* (cf. Major (76)).

Corollary 1 provides a rate of the order of $\nu^{-d/2}\psi^{-1}(\nu^d)$ in the uniform central limit theorem of Alexander and Pyke (86). Taking in account a remark of Breiman (67), we believe that this rate is optimal.

Part (a) of Corollary 1 is a weak invariance principle in the sense of Philipp (80) while (b) is a strong invariance principle where the function $x \to x^2 LLx$ plays an important role. In fact, Corollary 1 yields two different results according to the monotonicity of the function $x \to \psi(x)(x^2LLx)^{-1}$. When $x \to \psi(x)(x^2LLx)^{-1}$ is nondecreasing, $\varphi = O(\psi^{-1})$ and (b) of Corollary 1 holds with the error term $o(\psi^{-1}(\nu^d))$. From Breiman's remark, this result is optimal when $\mathcal{S}$ contains the class of boxes.

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

COROLLARY 2. *Let* $\mathcal{S}$ *be a family of Borel subsets of* $[0,1]^d$ *satisfying the assumptions of Corollary 1. Let Q be a law on* $\mathbb{R}^k$ *with mean zero and positive definite covariance matrix, and let* ψ *be a mapping fulfilling 2.3. Furthermore assume that* $x \to \psi(x)(x^2LLx)^{-1}$ *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* $N(0, VarQ)$*-distributed random vectors such that:*

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

Comments. When S contains the class of lower-left orthants and when only the second moment is assumed to be finite, this result is optimal (cf Major (76)). moreover, When $\psi(x) = x^2(LLx)^\alpha$ for some $0 < \alpha < 1$, the power of $LL\nu$ cannot be improved; Cf. Einmahl (87), Theorem 4.

Note that Corollary 3 and Theorem 3.1 of Bass and Pyke (84) yield a functional law of the iterated logarithm (cf. Bass and Pyke (84) for more about this). However, the law of iterated logarithm has been proved by Bass (85), by means of enterely different methods

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

$$\mathcal{H}(\varsigma) \qquad \liminf_{\varepsilon \to 0} \mid \log(H(\varepsilon))/\log \epsilon \mid \geq \varsigma$$

Then, the following result is available.

COROLLARY 3. *Let $d > 1$ and let S be a family of Borel subsets of the unit cube fulfilling $\mathcal{H}(\varsigma)$ for some ς in $[0,1]$ and 1.1 for some δ in $]0,1]$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite variance matrix, satisfying*

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

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 $N(0, VarQ)$-distributed random vectors such that:

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

Comments. *When S fulfills the uniform Minkowsky condition (i.e. $\delta = 1$), Corollary 4 means that in some sense, the rate of convergence in the uniform central limit of Alexander and Pyke (86) is of the order of $I(\nu^{-1})$. This result generalizes a previous result by Massart (87).*

Before discussing further our results, we give a consequence of corollaries 1 and 3.

COROLLARY 4. *Let $d > 1$ and let $\mathcal{S}$ be a family of Borel subsets of $[0,1]^d$. Assume that $\mathcal{S}$ is totally bounded with inclusion and has a log-entropy $H(.)$ satisfying $H(\varepsilon) = O(\varepsilon^{-\varsigma})$ as $\varepsilon \to 0$ for some $0 < \varsigma < 1$. Furthermore assume that $\mathcal{S}$ fulfills 1.1 with $\delta = 1$. Let Q be a law on $\mathbb{R}^k$ with mean zero and positive definite covariance matrix, satisfying*

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

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

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

Comments. Among examples of index families which satisfy the conditions of Corollary 4 for some positive ς, let consider the following. If $\mathcal{C}^2$ denotes the class of convex subsets of I^2. Dudley (74) has shown that $\varsigma = 1/2$. Now, if $d > 1$ and if $J(\alpha, d, M)$ denote the class of sets introduced in Dudley (74), whose boundaries are images of α-differentiables mappings of the $(d-1)$-sphere into $\mathbb{R}^d$, with all derivatives of order up to α uniformly bounded by M, $\varsigma = (d-1)/\alpha$ (cf. Dudley (78)). This class shall be precisely defined in section 4. In these cases, we prove that our result cannot be improved. Let us now state the related result.

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

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

Comments . From Theorem 2, it follows that the rate of approximation appearing in Corollary 4 is optimal in many cases of interest.

Before proving Theorem 1, we give another corollary in the unidimensional case.

COROLLARY 6. *Let $d = 1$, let $\delta > 1/2$, and let Q be a law on $\mathbb{R}^k$ with mean 0, with positive definite covariance, satisfying:*

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

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 $N(0, VarQ)$-distributed random vectors such that:

$$\sup_{S \in \mathcal{A}(\delta} \mid X(\nu S) - Y(\nu S) \mid = O(\nu^{1-\delta}(\log \nu)^{\delta}) \ a.s.$$

Comment. This result generalizes the results of K.M.T.. Moreover, we believe that this result is optimal, too.

Now, we prove theorem 1 . The proof of this theorem is based on the methods of a common probability space previously introduced by Komlos, Major, and Tusnady. Here, our method of construction of the two arrays of independent random vectors is exactly the same as in our previous paper (cf. Rio (90)). Yet, we need to recall some basic properties of our construction.

3. STRONG APPROXIMATION.

Throughout this section, Q is a law on $\mathbb{R}^k$ with mean 0 and finite variance . We may without loss of generality assume that $VarQ = I_k$. $(X_i)_{i \in \mathbb{Z}_+^d}$ denotes an array of independent random vectors with common law Q , and ψ is any mapping fulfilling 2.3, such that $E(\psi(\mid X_1 \mid)) < +\infty$.

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 $N(0, I_k)$. This construction is exactly the same as in our previous paper; so, we refer the reader to Rio (90) for this construction Then , by means of the mapping σ from $\mathbb{Z}_+$ onto $\mathbb{Z}_+^d$ which was defined in our previous paper, we will turn the so defined sequences into arrays. Now, we recall the definition and the geometrical properties of this mapping.

First, we define a multidimensional dyadic order on $\mathbb{Z}_+^d$:

Notation. $\mathbb{Z}^d$ is provided with the usual sum, product and order. We define now the subset J of $\mathbb{N}^d$ by :

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

It is obvious that the map from J onto $\mathbb{N}$ which maps $(j_1, \dots, j_d)$ onto $(j_1 + j_2 + \dots + j_d)$ is a non-decreasing, one to one correspondence . So , for each integer j , we shall call $(j_1, j_2, \dots, j_d)$ the unique element of J such that $j = j_1 + j_2 + \dots + j_d$. Let R_j be the lattice of multiples of $(2^{j_1}, 2^{j_2}, ..., 2^{j_d})$: we define the box $C'_{j,p}$ for any p of R_j by (here $\mathbb{1} = (1, 1, ..., 1) \in \mathbb{R}^d$) :

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

Then, the following lemma is avaiblable (cf. I for a proof.)

LEMMA. There exists a one to one mapping σ from $\mathbb{Z}_+^d$ to $\mathbb{Z}_+$ such that σ^{-1} maps the intervals $[1, 2^N]$ onto the boxes $[\mathbb{1}, (2^{N_1}, ..., 2^{N_d})]$, the intervals $I_{j,p}$ onto the boxes $C'_{j,p}$.

Then, we set $Y_i = Y_{\sigma(i)}$ and $X_i = X_{\sigma(i)}$ for any $i \in \mathbb{Z}_+^d$. Now, we define the class $\mathcal{A}_\nu$ of subsets of $]0, \nu^d]$ from the class νS and we recall some fundamental properties of the class $\mathcal{A}_\nu$.

Notations. If B is any subset of $\mathbb{Z}_+^d$, let $\sigma^* B$ be the mapping from $\mathbb{Z}_+$ to $[0, 1]$ defined by:

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

and, for each integer ν, we define the class $\mathcal{A}_\nu$ by :

$$\mathcal{A}_\nu = \{\sigma^* B : B \in \nu S\}$$

When S is contraction closed, $(\mathcal{A}_\nu)_{\nu > 0}$ is a nondecreasing sequence of families of elements of $l^2(\mathbb{Z}_+)$. Hence, there is no loss of generality in assuming that $\nu = 2^N$, which we shall do in the sequel. Given a function f and a sequence $(u_i)_{i>0}$ of vectors of $\mathbb{R}^k$, we set $u(f) = \sum f(i) u_i$. It is obvious that:

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

So, henceforward we work with the sequences $(X_i)_{i>0}$ and $(Y_i)_{i>0}$ defined in our previous paper. Throughout the sequel, the intervals $]l,m]$ have to be interpreted as subsets of $\mathbb{Z}_+$. $l^2(\mathbb{Z}_+)$ is given the canonical inner product, which we denote by $(\, . \mid . \,)$, and $l^2(]l,m])$ denotes the subspace of $l^2(\mathbb{Z}_+)$ of functions with support included in $]l,m]$.

In order to control the random vector $Y(f) - X(f)$, uniformly on the class $\mathcal{A}_\nu$, It will be convenient to define a dyadic orthogonal basis, as Massart (89) does in his paper.

Let $I_{j,p} =]p2^j, (p+1)2^j]$, and let $e_{j,p}$ be the characteristic function of $I_{j,p}$. For any positive integers p and j, we set $\tilde{e}_{j,p} = e_{j,p} - 2e_{j-1,2p}$. Let $\tilde{e}_j = e_{j,1}$. We define the orthogonal systems $\mathcal{B}_o$ and $\mathcal{B}_j$ by :

$$\mathcal{B}_o = \{\tilde{e}_j : 0 \leq j < Nd\} \cup \{e_{o,o}\} \quad \textit{and} \quad \mathcal{B}_j = \{\tilde{e}_{j,p} : 0 < p < 2^{Nd-j}\}$$

Then $\mathcal{B} = \bigcup_{j=0}^{Nd-1} \mathcal{B}_j$ is an orthogonal basis of $l^2(]0,n])$. Now, let Π_j be the orthogonal projector on the space generated by $\bigcup_{l=1}^{j} \mathcal{B}_l$. For any function bounded by 1, the control of $X(f) - Y(f)$ depends mainly on the quantities $(\Pi_j f \mid \Pi_j f)$. Moreover, the uniform control on $\mathcal{A}_\nu$ of the above inner products is insured via the geometrical assumption 1.1. on the boundaries of elements of S and the perimetric properties of the mapping σ: let recall the related lemma (cf. Rio (90) for a proof).

LEMMA 1. *Assume that S is a class of Borelian subsets of the unit cube fulfilling the condition 1.1. for some constants $\delta > 0$ and $K \geq 1$. Let $n = \nu^d$. Then, for any element $f \in \mathcal{A}_\nu$, $\Pi_j f$ takes its values in $[-1,1]$:*

$$\sum_{i=1}^{n} \mid \Pi_j f(i) \mid \leq 2Kn^{(d-\delta)/d} 2^{j\delta/d}$$

Moreover, if $a = 2^{jd}$

(4.2) the cardinality of the set of integers p such that $\Pi_j(f) \not\equiv 0$ on $I_{j,p}$ is no more than $2^{-j}\lambda((\nu\partial S)^a)$.

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

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

for any nonnegative integer l, for any odd integer i in $[2^L, 2^{L+1}[$. Clearly, we have:

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

So, it will be enough to control each of the terms on right hand. First, the control of the sequence $\sum_{i=1}^{n} | X_i - \tilde{X}_i |$ is insured via the following lemma (cf. Rio (90) for a proof).

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

Now, in order to obtain an exponential bound on the random vector $Y(f) - \tilde{X}(f)$, we shall use the dyadic decomposition previously introduced in K.M.T. (75), exactly as in our precedent paper.

If f is any function from $\mathbb{Z}_+$ to $\mathbb{R}$ with support included in $]0, n]$, we set $\gamma_j(f) = 2^{-j}(f \mid \tilde{e}_j)$ and $\gamma_{j,p}(f) = 2^{-j}(f \mid \tilde{e}_{j,p})$. Then, the orthogonal expansion of the function f with respect to the orthogonal basis $\mathcal{B}$ has the following form :

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

We introduce now further notations. Let define the random sequences $(\bar{\xi}_i^j)_{i>0}$ and $(\xi_i^j)_{i>0}$ by :

$$\bar{\xi}_i^j = 1\!\!1_{M_j < |\bar{X}_i| \leq M_{j+1}} \bar{X}_i \text{ and } \xi_i^j = \bar{\xi}_i^j - E(\bar{\xi}_i^j)$$

and let

$$\tilde{U}_{j,p} = \tilde{X}(\tilde{e}_{j,p}) - \sum_{l \geq j} \xi^l(\tilde{e}_{j,p}), \quad \tilde{U}_j = \tilde{X}(\tilde{e}_j) \text{ and } \tilde{V}_j = Y(\tilde{e}_j)\ \tilde{V}_{j,p} = Y(\tilde{e}_{j,p}).$$

We define now the random vectors $D_j(f)$ by :

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

and

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

It can easily be seen that

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

So, in order to control the random vector $Y(f) - \tilde{X}(f)$, it suffices to control each of the r.v.'s $D_j(\Pi_j f)$, and the random vector $D_0(f)$. First, we note that, for any function f with values in $[-1,1]$, $|\gamma_j(f)| \leq 1$ and $|\gamma_{j,p}(f)| \leq 1$.

Hence, setting $D_0 = \mid Y_1 - \tilde{X}_1 \mid + \sum_{j=0}^{Nd-1} \mid \tilde{U}_j - \tilde{V}_j \mid$, $D_0(f) \leq D_0$ for any $f \in \mathcal{A}_\nu$.

Now, let $D_j = \sup_{A \in \mathcal{A}_\nu} \mid D_j(\Pi_j f) \mid$. It is obvious that:

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

The control of D_0 and $D_j(f)$ is based on the following inequalities which are consequences of Einmahl's results on the normal approximation via multivariate transformations of smoothed partial sums of smoothed random vectors. we refer to our previous paper for more about this.

PROPOSITION 1. *There exist a positive 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,*

$$pr\big(D_0 \geq c_2(\psi^{-1}(n)t + \varphi(n)u)\big) \leq 4k\big((\beta_{Nd})^t + \exp(-2u^2 LLn)\big) \qquad 3.6.(a)$$

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

$$pr(| D_j(g) | \geq c_2\sqrt{v}(\psi^{-1}(2^j)t + \varphi(2^j)u)) \leq$$
$$4k\exp\left(t^2(1+v^{-1/2}t)^{-1}\log\beta_j\right) + 4k\exp\left(-2u^2\log(1+j)\right) \qquad 3.6.(b)$$

Now, we give some consequences of the above proposition. An immediate consequence of (a) is:

$$D_0 = O(\varphi(n)) \text{ a.s. and } D_0 = O_P(\psi^{-1}(n)) \qquad (3.7)$$

We now use (b) of proposition 1 to conclude the proof of theorem 1. Clearly, it will be enough to control each of the random variables D_j.

Proof of theorem 1.

In order to prove Theorem 1, it is necessary to use the entropy properties of the already defined class $\mathcal{A}_\nu$. First, we note that, for any Borelian subset B of $\mathbb{R}_+^d$,

$$\lambda(B) = \nu^{-d} \sum_{i \in \mathbb{Z}_+} \sigma^\star B(i)$$

Moreover, it is obvious that, for any borelians subsets A and B, $A \subset B$ implies $\sigma^\star A \leq \sigma^\star B$. Hence, the entropy with bracketing $N(\varepsilon, \mathcal{A}_\nu)$ with respect to the uniform measure P on $\mathbb{N}\cap]0, \nu^d]$ of the family $\mathcal{A}_\nu$ satisfies $N(\varepsilon, \mathcal{A}_\nu) \leq N_I(\varepsilon, \mathcal{S})$. Recall that the entropy with bracketing is the minimal cardinality of a collection $\mathcal{A}(\varepsilon)$ of functions such that, for any f in $\mathcal{A}_\nu$, there exists f^+ and f^- in $\mathcal{A}(\varepsilon)$ such that $f^- \leq f \leq f^+$ and $P(f^+ - f^-) \leq \varepsilon$.

So, henceforth, we work with the family $\mathcal{A}_\nu$. In order to prove theorem 1, it will be necessary to use theorem 1, and a *chaining argument* (cf. Pollard (84) p.142 for more about this). If furthermore the entropy with bracketing is integrable, one can use a restricted chaining argument (see also Pollard (84, p.159) for an example). By means of a restricted chaining, in the spirit of Bass (85), we now control each of the random variables D_j. Here, we need to state a general oscillation lemma

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

(3.8) (i) $\mathcal{G}$ contains the null function.

(ii) $\mathcal{F}$ has an integrable entropy with bracketing, which we denote by $N(\varepsilon, \mathcal{F})$.

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

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

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

(3.9) (i) for any $g \in \mathcal{G}$, $|A(g)| \leq A^*(|g|)$.

(ii) for any positive ϵ, for any positive t,

$$pr(A(f-g) \geq (a \wedge \epsilon)t) \leq \Lambda \exp(-2\theta(\epsilon \wedge a)t^2(1+t)^{-1})$$

(iii) for any positive ϵ, for any positive g in $\mathcal{G}$ satisfying $\|g\|_{1,\mu} \leq \epsilon$,

$$pr(A^*(g) \geq \epsilon t) \leq \Lambda \exp(-\theta \epsilon t)$$

When the random process A fulfills the above conditions, the following lemma is available:

LEMMA 3. *Let $\mathcal{F}$ be a class of measurable functions from $\mathcal{X}$ to $[-1, 1]$ satisfying 3.8 and let A be a linear process fulfilling 3.9 . Then, for any positive x,*

$$pr\Big(\sup_{f \in \mathcal{F}} A(f) > 7\theta^{-1/2}I(a) + 2H_1^{-1}(\theta/3) + 2x + 7\sqrt{ax}\Big) \leq \Lambda\Big(1 + \frac{\pi^2}{6}\Big)\exp(-\theta x)$$

Comments. From lemma 3, it follows that the random process A has continuous paths; moreover, lemma 3 yields an upper bound on the uniform modulus of continuity of A (it suffices to apply lemma 3 with $\mathcal{F}_1 = \mathcal{U}_b$ and $\{A(f-g) : (f, g) \in \mathcal{U}_b\}$).

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

- for any $f^* \in \mathcal{F}_\epsilon^*$, $\|f^*\|_{1,\mu} \le \epsilon$.
- for any f in $\mathcal{F}$ there exists some $(f_\epsilon, f_\epsilon^*)$ in $\mathcal{F}_\epsilon \times \mathcal{F}_\epsilon^*$ such that $|f - f_\epsilon| \le f_\epsilon^*$.
- The minimal cardinality of such pairs is no more than $\exp(H(\epsilon))$.

By definition of both processes A and A^*, for any positive ϵ,

$$\sup_{f\in\mathcal{F}} A(f) \le \sup_{f\in\mathcal{F}_\epsilon} A(f) + \sup_{f^*\in\mathcal{F}_\epsilon^*} A^*(f^*) \tag{3.10}$$

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

$$\varepsilon(x) - 3\theta^{-1}H^+(\varepsilon(x)) \le x \le \varepsilon(x) - 3\theta^{-1}H(\varepsilon(x)) \tag{3.11}$$

1. If $x > 1$. It is obvious that:

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

So, by (iii) of 3.9, inequality (a) of lemma 3 holds true.

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

$$pr\big(\sup_{f^*\in\mathcal{F}_{\varepsilon(x)}^*} A^*(f^*) \ge \varepsilon(x)\big) \le \exp(-\theta x) \tag{3.12}$$

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

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

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

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

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

$$pr(\sup_{f^*\in\mathcal{F}^*_{\varepsilon(x)}} A^*(f^*) \geq x + H_1^{-1}(\theta/3)) \leq \Lambda \exp(-\theta x) \tag{3.14}$$

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

$$pr(\sup_{f\in\mathcal{F}_{\varepsilon(x)}} A(f) \geq \varepsilon(x)) \leq \Lambda \exp(-\theta x)$$

Then, we complete the proof of (a) of lemma 3 by collecting inequalities 3.13, 3.14 and the above inequality. On the other hand, when $x \leq a$, we need to use a restricted chaining argument.

Restricted chaining.

Let us now introduce the following notations. Let $\varepsilon_0 = \varepsilon(x)$. For any integer j, we set $\varepsilon_j = 2^j\varepsilon_0$, and we define the sequence $\mathcal{F}_j$ of approximating nets associated with the sequence $(\varepsilon_j)_j$ as follows: $\mathcal{F}_0 = \mathcal{F}_{\varepsilon_0}$ and, for any natural j, $\mathcal{F}_{j+1} = \mathcal{F}_j$ iff $H(\varepsilon_{j+1}) = H(\varepsilon_j)$, and $\mathcal{F}_{j+1} = \mathcal{F}_{\varepsilon_{j+1}}$ otherwise. Clearly, the so defined collection $\mathcal{F}_j$ is an ε_j-net. Hence, there exists mappings ϕ_j from $\mathcal{F}_j$ to $\mathcal{F}_{j+1}$ such that:

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

Now, let $l = \sup\{j \in \mathbb{N} : \varepsilon_j \leq a\}$. For any f_0 in $\mathcal{F}_0$, we define the mappings $(f_j)_{j>0}$ by $f_{j+1} = \phi_j(f_j)$ for any natural j. Let J_0 be the subset of $\mathbb{N}$ defined by

$J_0 = \{j < l \text{ such that } \mathcal{F}_j \neq \mathcal{F}_{j+1}\} \cup \{l\}$. For any f_0 in $\mathcal{F}_0$, one can write:

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

Starting from the above equality, it is easily seen (cf. Pollard (84) p.142) that, for any sequence $(t_j)_{j\geq 0}$ of positive numbers,

$$pr(\sup_{f_0\in\mathcal{F}_0} A(f_0) \geq \sum_0^l t_j) \leq \sum_{j\in J_0} |\mathcal{F}_j| \sup_{f\in\mathcal{F}_j} pr(A(f - \phi_j(f)) \geq t_j) \tag{3.15}$$

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

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

for any positive number u_j. So, setting $t_j = u_j \varepsilon_{j+1}$, and choosing u_j such that

$$3H(\varepsilon_j) + \theta x = \theta \varepsilon_j h(u_j),$$

we get:

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

From which it follows that

$$pr(\sup_{f_0 \in \mathcal{F}_0} A(f_0) \geq \sum_0^l t_j) \leq \Lambda \frac{\pi^2}{6} \exp(-\theta x) \tag{3.16}$$

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

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

Then, collecting the inequalities 3.16 and 3.17 we get that

$$pr(\sup_{f \in \mathcal{F}_0} A(f) \geq 7\theta^{-1/2} I(a) + H_1^{-1}(\theta/3) + +x + 7\sqrt{ax}) \leq \Lambda(1 + \frac{\pi^2}{6}) \exp(-\theta x) \tag{3.18}$$

and then, the end of the proof is straightforward.

<u>End of the proof of theorem 1.</u>

To complete the proof of theorem 1, we control each of the components of the random vector $D_j(\Pi_j f)$. So, in order to control the random variables D_j, there is no loss of generality in assuming that $k = 1$, which we shall do throughout the sequel. Now, we define a positive linear random process $D_j^*(\ .\ )$ associated with $D_j(\ .\ )$. Let

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

Clearly, for any g in $l^1(]0,n])$, $D_j(g) \leq D_j^*(|g|)$. In order to apply lemma 3, it then suffices to prove that, for any mapping g from $\mathbb{Z}_+$ into $[-1,1]$, For any positive $v \geq 2^{-j}(g \mid g)$,

$$pr\big(D_j^*(g) - E(D_j^*(g)) \geq c_2\sqrt{v}(\psi^{-1}(2^j)t + \varphi(2^j)u)\big) \leq$$
$$4k\exp\big(t^2(1+v^{-1/2}t)^{-1}\log\beta_j\big) + 4k\exp\big(-2u^2\log(1+j)\big) \qquad (3.19)$$

The proof of 3.19 will be omitted, since uses exactly the same arguments as the proof of (b) of proposition 1.

Now, let P be the uniform law on $]0,n]$. It is obvious that

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

First, we prove (b) of theorem 3. We may w.l.o.g. assume, by increasing β_j a little, that $\beta_j \geq (1+j)^{-2}$. Then, setting $a = 1 \wedge 2K2^{(j-Nd)\delta/d}$, $A_j(g) = (4c_2\varphi(2^j))^{-1}2^{j-Nd}D_j(g)$ and $A_j^*(g) = (4c_2\varphi(2^j))^{-1}2^{j-Nd}D_j^*(g)$, and collecting lemma 1, proposition 1 and the inequalities 3.19 and 3.20, we obtain the existence of some positive constant c_3 such that, for any positive t,

$$pr(A_j(g) \geq (\|g\|_{1,P} \wedge a)t) \leq 8k\exp(c_3h(t)|\log\beta_j|) \qquad 3.21.(a)$$

and

$$pr(A_j^*(g) \geq t\|g\|_{1,P}) \leq 8k\exp(c_3t|\log\beta_j|) \qquad 3.21.(b)$$

Hence, by lemma 3,setting $u_j = 2^{Nd-j}$, there exists some positive constant c_4, such that for any positive t,

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

Now, the end of the proof is straightforward, setting $t_j = Nd - j$ in the above inequalities and using Borel-Cantelli lemma. In order to prove (a), it suffices to notice that

$$pr(A_j(g) \geq (\|g\|_{1,P} \wedge a)t) \leq 8k\exp(-c_3h(t))$$

and

$$pr(A_j^*(g) \geq \|g\|_{1,P}t) \leq 8k\exp(-c_3h(t))$$

Then, starting from lemma 3, one can complete the proof of (a) exactly in the same way

4. LOWER BOUNDS ON THE APPROXIMATION OF PARTIAL-SUM AND MULTIVARIATE EMPIRICAL PROCESSES.

In this section, starting from techniques previously initiated by Beck (85, 87) for lower bounds in theory of irregularities of distribution, we prove that our Corollary 4 is optimal when S is either the class $J(\alpha, d, M)$ introduced in Dudley (74), either the class C^2 of convex subsets of I^2. We shall also apply these techniques to provide lower bounds on the normal approximation of the multivariate homogenous empirical process. Before proving Theorem 2, let us define precisely the class $J(\alpha, d, M)$.

Throughout this section, $|\,.\,|$ denotes the Euclidian norm on $\mathbb{R}^d$. Let β be the greatest integer lower than α (i.e. $\alpha < \beta$) and let $\gamma = \alpha - \beta$. For any open set $U \subset \mathbb{R}^d$, let $F(U, \alpha)$ be the set of all real functions f on U such that:

(a) f is $C^{(\beta)}$ i.e. the partial derivatives $D^p f$exist and are continuous for any multiindex $p = (p_1, \ldots, p_d)$ with $|p| = p_1 + \cdots + p_d \leq \beta$

(b) $\|f\|_\alpha < +\infty$ where

$$\|f\|_\alpha = \sup_{|p|=\beta} \sup_{\substack{x \neq y \\ (x,y) \in U^2}} \frac{|D^p f(x) - D^p f(y)|}{|x-y|^\gamma} + \sup_{|p| \leq \beta} \sup_{x \in U} |D^p f(x)|$$

Let S^{d-1} be the unit sphere of $\mathbb{R}^d$. We can cover S^{d-1} by finitely many coordinate patches V_j, so that there are C^∞ isomorphisms $\varphi_j : U \to V_j$ where U is the open ball $\{y : |y| < 1\} \subset \mathbb{R}^{d-1}$. Furthermore, we may assume that $\varphi_j = \psi_{j|U}$, for some mapping $\psi_j \in F(W, \alpha)$, for some open set $W \supset \overline{U}$. Then, let $I(\alpha, d, M)$ be the set of $\mathbb{R}^d$-valued functions on S^{d-1} such that $\sup_j \|f \circ \varphi_j\|_\alpha \leq M$, and let:

$$J(\alpha, d, M) = \{I(f) : f \in I(\alpha, d, M)\}$$

where I(f) is the subset of vectors $x \in \mathbb{R}^d \setminus range(f)$ such that, in $\mathbb{R}^d \setminus \{x\}$, f is not homotopic to any constant map. Then, one can prove that $\partial(I(f)) = range(f)$ (cf. Dudley (74)). Moreover, the so defined class $J(\alpha, d, M)$ is invariant by any orthogonal transformation. Now, recall that $\mathcal{S}$ is a class of subsets of the unit cube. So, in order to work with a family of subsets of I^d, we define $\mathcal{S}$ by $\mathcal{S} = \{I^d \cap S : S \in J(\alpha, d, M)\}$. In fact, we derive theorem 2 from the more general result which is stated below.

THEOREM 3. *Let $2 < r \leq 4$ and let F and G be different probability laws on $\mathbb{R}$ with mean zero and finite rth-moment. 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 distributions functions F and G , and let $\mathcal{S}$ denote the class $J(\alpha, d, M)$. Then, there exists some positive constant $C(F, G)$ such that:*

$$pr\big(\nu^{(1-d)(1+1/\alpha)/2} \sup_{S \in \mathcal{S}} \mid X(\nu S) - Y\nu S) \mid \leq C(F, G)\big) \leq C(F, G)\nu^{(1-r/2)(\alpha+d-1)/\alpha}$$

Comments. When $\mathcal{S}$ is the class $\mathcal{C}_2$ of convex subsets of $\mathbb{R}^2$ Theorem 3 still holds with $\alpha = 2$.

Before proving Theorem 1, we still give another theorem whose proof also uses the Fourier analysis techniques initiated by Beck. Let λ_0 be the Lebesgue measure on the unit cube and let $(x_i)_{i>0}$ be a sequence of i.i.d. random vectors with distribution λ_0. The following result is a partial converse of Massart's results on the strong approximation of the multivariate empirical bridge; cf. Massart (89). Let P_n denote the empirical measure associated with $(x_1, \cdots, x_n)$: $nP_n = \sum_1^n \delta_{x_i}$. We call the empirical bridge the centered and normalized process $Z_n = \sqrt{n}\,(P_n - \lambda_0)$.

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

$$E\big(\sup_{S \in \mathcal{S}} \mid Z_n(S) - Z^{(n)}(S) \mid\big) \geq C_0(d)\, n^{(d-1-\alpha)/(2d\alpha)}$$

comments. When S is the class of convex subsets of $\mathbb{R}^2$, Theorem 4 still holds true with $\alpha = 2$. It is worth noticing that this result improves previous lower bounds obtained by Borisov (85).

Proof of Theorem 3. Let μ be a signed Borel measure on I^d absolutely continuous with respect to λ. We shall give a lower bound on the variable $\sup_{S \in \mathcal{S}} |\mu(S)|$. We need first to introduce notations and definitions.

Notations. Let define the test function ϱ by $\varrho(x) = 0$ iff $|x| \leq 1$ and

$$\varrho(x) = \frac{1}{2\sqrt{d}} \exp(-(1 - |x|)^{-1}) \text{ iff } |x| \geq 1$$

and let define the open set C from ϱ by:

$$C = \{(x_1, \dots, x_d) \in \mathbb{R}^d : |x| < \varrho(x_1, \dots, x_d)\}$$

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

Now, we define a subadditive functional $\Delta(\mu, A)$ from the signed measure μ. So, let

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

where $v.\chi_A(x) = \chi_A(v^{-1}x) = \mathbb{1}_{vC_A}$.

As the first step of the proof, we give the following lower bound on the variable $\sup_{S \in \mathcal{S}} | \mu(S) |$.

LEMMA 4. *Let $A = (M\|\varrho\|_\alpha)^{1/\alpha}\nu^{1-1/\alpha}$ and let μ be a signed measure with support included in I^d. Then, there exists a positive constant C_0 depending only on d, such that*

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

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

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

Clearly $\Psi_\varepsilon \in F(\mathbb{R}^{d-1}, \alpha)$ and $\|\Psi_\varepsilon\|_\alpha \le M$. Now, we derive from the functions Ψ_ε a family of subsets of I^d with α-smooth boundaries. So, we set:

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

It is easily seen that, for all $v \in O(d)$, there exists some $S(\varepsilon, y, v)$ in $J(\alpha, d, M)$ such that:

$$S(\varepsilon, y, v) \cap I^d = v.\mathcal{C}(\varepsilon, y) \cap I^d$$

So, recalling that the support of μ is included in I^d, we get:

$$\begin{aligned} \sup_{S \in \mathcal{S}} \mid \mu(S) \mid &\ge \frac{1}{2} \sup_\varepsilon (\mu(v.\mathcal{C}(\varepsilon, y)) - \mu(v.\mathcal{C}(-\varepsilon, y))) \\ &\ge \frac{1}{2} \sum_{p \in \mathbb{Z}^{d-1}} |\mu(v.(\mathcal{C}_A + (\nu^{-1}A, \ldots, \nu^{-1}A, 1)y))| \end{aligned} \tag{4.1}$$

Let $y = (y_1, \ldots, y_d)$. If $|y_d| > 1 + d$, the terms on right hand in the above inequality are null. Hence, integrating, we get:

$$2^d(1+d) \sup_{S \in \mathcal{S}} |\mu(S)| \ge \iint_{O(d) \times [-1,1]^{d-1} \times \mathbb{R}} dv dy \sum_{p \in \mathbb{Z}^{d-1}} |\mu(v.(\mathcal{C}_A + (\nu^{-1}A, \ldots, \nu^{-1}A, 1)y)|$$

and, from now on, the end of the proof is straightforward

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

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

The signed measure μ is said to fulfill the condition $\mathcal{K}(A, C)$ iff:

$\mathcal{K}(A, C)$ $$\Delta(\mu, A) \ge A^{(1-d)/2} C\, D(\mu, A)$$

If μ is a random measure,we set $\Gamma(C) = \{\omega \in \Omega : \mu(\omega) \text{ fulfills } \mathcal{K}(A, C)\}$. We shall see later that, when μ is the empirical measure associated with the difference of the two partial-sum processes X and Y, there exists some positive constant C not depending on ν such that,

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

Then, theorem 3 is a straight consequence of the following proposition:

PROPOSITION 2. *There exists some positive constant $c(d)$ such that, for all signed measure μ*

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

Proof. By Parseval-Plancherel identity,

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

From which it follows that:

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

Now, let:

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

Clearly, proposition 2 follows from the next lemma, which is the main technical tool for the proof.

LEMMA 5. *There exists some positive constant $c(d)$ such that, for any $\rho \leq \pi\sqrt{d}$, $g(A, \rho) \geq c(d) A^{d-1}$.*

Proof. Let $t = (0, \ldots, 0, \rho)$. From the revolution symmetry of C_A, it follows that, if ds denotes the invariant measure on S^{d-1} normalized such that $\int_{S^{d-1}} ds = 1$,

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

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

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

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

$$\nu^d \Re e\, \widehat{\chi}_A(\rho\nu s) = \int_{\nu C_A} \cos(\rho s \mid x)\, dx$$

Now, it is easily seen that, for any s in V_ϵ, for any x in νC_A, for any $\rho \leq \pi\sqrt{d}$, $(\rho s \mid x) \leq \frac{\pi}{6} + \epsilon A \pi \sqrt{d}$. Hence, if one choose ϵ such that $\epsilon A \sqrt{d} = 1/6$,

$$\nu^d \Re e\, \widehat{\chi}_A(\rho\nu s) \geq \frac{1}{2}\lambda(\nu C_A)$$

Now, recall that $A \geq 1$. So, the measure of V_ϵ is greater than $c_2 A^{1-d}$. Then, the proof of lemma 5 is achieved by noting that $\lambda(C_A) \geq c_3 A^{d-1}$ and by collecting the above inequalities.

Now, we define an empirical measure associated with the difference of the two partial-sum processes. So, we set:

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

and we define μ_Y from the array $(Y_p)_p$ exactly in the same way. Then, let $\mu = \mu_X - \mu_Y$. A few calculation proves that:

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

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

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

So, by proposition 2 and by Parseval-Plancherel identity again,

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

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

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

From a result of Bartfai (cf. Major (78) for a proof in a more general context), it follows that:

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

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

$$D(\mu, A) \geq (2/\pi)^d c(d)\, A^{d-1} W^2(F_n, G_n) \tag{4.3}$$

Now, let $d(F,G)$ denote the levy distance of the laws F and G. it is easily seen that for any laws F and G, $W^2(F,G) \geq d^3(F,G)$. So, it suffices to give a lower bound on $d(F_n, G_n)$. First, we note that:

$$d(F_n, G_n) \geq d(F,G) - d(F,F_n) - d(G,G_n)$$

Let $b = d(F,G)/4$. Starting from Dvoretsky-Kiefer-Wolfowitz inequality (cf. Massart, 90) it is straightforward to prove that:

$$pr(d(F,F_n) > b) \leq 2\exp(-2nb^2) \qquad (resp. G)$$

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

$$pr(W^2(F_n, G_n) \geq c_4) \leq 4\exp(-nc_4) \tag{4.4}$$

So, by 4.4, and 4.3, there exists some positive constant c_5 such that,

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

Now, in order to conclude the proof of Theorem 3, it only remains to prove that μ satisfies 4.2 for some positive constant C not depending on ν.

Proof of 4.2. The control of the tail of $\Gamma(C)$ is performed via the following classical inequality (cf. Meyer, 72 sect.48).

LEMMA. (Marcinkiewicz-Zygmund's inequality) Let $\xi_1, \dots, \xi_n$ be independent random variables with mean 0. Moreover, assume that $E(|\xi_p|^r) \le 1$, for any p, for some $r > 1$. Let $S_n = \sum_1^n \xi_p$. Then,

$$E(| S_n |^r) \le C(r) n^{\max(1, r/2)}$$

Now, we introduce further notations. for any signed measure β let define $k(\beta, v, x) = |\beta * v.\chi_A(x)|$. for convenience, we write $k^2(\beta, v, x) = (k(\beta, v, x))^2$. Then, for any positive a, we have:

$$2a\Delta(\mu, A) \ge D(\mu, A) - \iint_{O(d)\times\mathbb{R}^d} dvdx\, k^2(\mu, v, x) \mathbb{1}_{k(\mu,v,x)>2a} \ge$$
$$D(\mu, A) - 4 \iint_{O(d)\times\mathbb{R}^d} dvdx (k^2(\mu_X, v, x) \mathbb{1}_{k(\mu_X,v,x)>a} + k^2(\mu_Y, v, x) \mathbb{1}_{k(\mu_Y,v,x)>a})$$

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

$$pr\left(16 \iint_{O(d)\times\mathbb{R}^d} dvdx\, k^2(\mu_X, v, x) \mathbb{1}_{k(\mu_X,v,x)>a} \ge c_5 A^{d-1}\right)$$
$$= O(\nu^{(1-r/2)(\alpha+d-1)/\alpha}) \tag{4.6}$$

Now, we prove 4.6. For all $(x, v) \in \mathbb{R}^d \times O(d)$, there exists a family of positive numbers $(\alpha_p)_p$ each being bounded by 1 such that $k(\mu_X, v, x) = |\sum_p \alpha_p X_p|$. Moreover, the cardinality of the set $\{p : \alpha_p \ne 0\}$ is smaller than $2.3^d(1 + d)A^{d-1}$. Hence, by Marcinkiewicz-Zygmund inequality,

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

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

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

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

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

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

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

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

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

Hence, by Jensen's inequality,

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

Finally, 4.6 follows from 4.7 and Markov's inequality applied to the above random variable. So, the proof of Theorem 3 is achieved. Now, we prove Theorem 4

Proof of Theorem 4.

Let introduce the following notations and definitions:

$$B_n = \sqrt{n} Z^{(n)} \text{ and } U_n = \sqrt{n} Z_n = \sum_1^n \delta_{x_p} - n\lambda_0$$

We define the real number ν by $\nu^d = n$ and we define A from ν exactly as in proposition 2. Let $\mu = U_n - B_n$. Using the same arguments as in the proof of lemma 4, it can easily be proven that:

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

Now, let us define the functional $D_4(\mu, A)$ by:

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

From Hölder's inequality, it follows that:

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

Moreover, it is straightforward to prove that there exists a positive universal constant c_7 such that:

$$E(D_4(\mu, A)) \leq c_7 A^{2(d-1)}$$

Hence, Theorem 4 is a consequence of the following proposition:

PROPOSITION 3. *There exists some positive constant C_1 such that:*

$$E(D(\mu, A)) \geq C_1 A^{d-1}$$

<u>Proof.</u> First, we note that, necessarily, the Gaussian process $B_n(v.(C_A + x))$ is almost surely uniformly continuous on $O(d) \times \mathbb{R}^d$ and with compact support. So, we may employ the theory of Fourier transformation. By Parseval-Plancherel identity,

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

Hence, by lemma 5,

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

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

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

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

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

Hence, proposition 3 follows from the above lemma:

LEMMA 6. *There exists some positive constant C_2 such that*

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

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

$$\bar{X}(n,x)=\sum_1^n \kappa(\nu(x-x_p)) \text{ and } X(n,x)=\bar{X}(n,x)-E(\bar{X}(n,x))$$

By definition, $X(n,x)$ has distribution $F_{n,x}$. Moreover, an elementary computation gives:

$$E(\bar{X}(n,x))=\int_{[(x-1)\nu,x\nu]}\kappa(t)\,dt$$

Now, recall that $\kappa=(\sin\pi t/\pi t)^{\otimes d}$. Hence the above integral is semi-convergent. So, we have:

$$E(\bar{X}(n,x)=(2\pi)^d+o(1)$$

It remains now to study the convergence in distribution of $\bar{X}(n,x)$. Let f be any $C^{(2)}$ mapping from $\mathbb{R}$ into $\mathbf{C}$ satisfying $f(0)=0$. First we prove that:

$$nE(f(\kappa(\nu(x-x_p)))=\int_{\mathbb{R}^d} f(\kappa(t))\,dt\ +o(1) \tag{4.9}$$

Since f is twice differentiable,

$$f(t)=tf'(0)+g(t) \text{ with } g(t)=O(t^2) \text{ as } t\to 0$$

So, by integrating,

$$nE(f(\kappa(x-x_p)))=\int_{[(x-1)\nu,x\nu]}\big(f'(0)\kappa(t)+g(\kappa(t))\big)dt$$

Now, it is obvious that $g(\kappa(t))=O((\kappa(t))^2)$ as $|t|\to+\infty$. Hence, recalling that $\kappa\in L^2(\mathbb{R}^2)$ we obtain 4.9. Now, by Levy's theorem, $\bar{X}(n,x)$ converges in distribution iff $E(\exp(i\xi\bar{X}(n,x))$ converges to a continuous function. Morevover, by 4.9,

$$E(\exp(i\xi\bar{X}(n,x))) = \left(1 + \frac{1}{n}\int_{\mathbb{R}^d} \exp(i\kappa(t))dt + o(1/n)\right)^n$$

From which it follows that $\bar{X}(n,x)$ converges weakly to a non-Gaussian infinitely divisible distribution. Hence, theorem 4 holds true.

When $\mathcal{S}$ is the class $\mathcal{C}_2$ of convex subsets of $\mathbb{R}^2$, we may apply the same techniques to provide lower bounds on the approximation. For example, let $\mathcal{P}_m$ denote the m-gone with vertices $\exp(i2k\pi/m)$ and let $\mathcal{D}_m = \mathcal{P}_m \setminus \mathcal{P}_{2m}$, and let $C(\mathcal{D}_m)$ denote the set of connex components of the interior of $\mathcal{D}_m$. Then, it is easily seen that, for any signed measure μ, for any v in $O(2)$, and for any (τ, x) in $I \times \mathbb{R}$,

$$2 \sup_{S \in \mathcal{C}_2} \mid \mu(S) \mid \geq \sum_{T \in C(\mathcal{D}_m)} |\mu(\tau v T + x)|$$

Then, choosing m of the order of $\sqrt{\nu}$, and using the same techniques of *rotation discrepancy* previously introduced by Beck (87), one can prove the corresponding lower bounds for the approximation of partial-sum or multivariate empirical processes indexed by $\mathcal{C}_2$

REFERENCES

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

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

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

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

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

BORISOV, I.S. (1985). Upper and lower estimates of the convergence rate in the invariance principle. *Proceedings of the fourth Vilnius conference I, VNU Science Press.* 251-259.

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

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

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

DUDLEY, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **7** 899-929.

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

EINMAHL, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Analysis* **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** 111-6131.

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 Analysis* **8** 487-517.

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

MASSART, P. (1989). About the constant in the DKW inequality. To appear in *Ann. of Probab.*

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

MEYER, P.A. (1972). Martingales and stochastics integrals I. Lecture Notes in Mathematics **284.**

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

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 Series in Statistics.

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* 4-49. *Trudy Inst. Mat.* Vol. 3, "Nauka" Sibirsk.Otdel, Novosibirsk [Russian].

CHAPITRE III

III. APPROXIMATION FORTE POUR DES PROCESSUS DE SOMMES PARTIELLES INDEXES PAR LES QUADRANTS.

Résumé : Soit $(X_i)_{i\in\mathbb{Z}_+^2}$ un champ de variables aléatoires réelles de variance finie égale à 1, indépendantes, équidistribuées et centrées. Dans la première partie de la thèse, nous avons vu, dans un cadre plus général que, si les variables X_i ont un moment d'ordre r inférieur à 4, le principe d'invariance fort pour le processus de sommes partielles $\{\sum_{i\leq p} X_i : p \leq (\nu,\nu)\}$ vaut avec un terme d'erreur de l'ordre de $o(\nu^{2/r})$. Cependant, quand les variables X_i ont un moment fini d'ordre r supérieur à 4, la méthode générale utilisée donne un terme d'erreur de l'ordre de $(\nu \log \nu)^{1/2}$. Nous montrons ici, à l'aide d'un procédé de construction adapté à la classe des quadrants que, dans ce cas, on peut obtenir un principe d'invariance fort avec un terme d'erreur de l'ordre de $o(\nu^{2/r}(\log \nu)^2)$.

D'après une remarque de Breiman (67), ce résultat est optimal au facteur $(\log \nu)^2$ près et confirme le fait que les vitesses optimales dans le principe d'invariance fort dépendent des propriétés géométriques de la classe de parties considérée.

Code Matière A.M.S. (1980) - 60 F 17 - 62 G 99

Mots-clés : Central limit theorem, Set-indexed process, Partial-sum process, Invariance principle, Multivariate empirical processes.

1. INTRODUCTION ET RESULTATS.

Dans cette partie de la thèse on s'interesse à l'approximation forte des processus de sommes partielles indexés par les quadrants en dimension 2. Dans la suite $(X_i)_{C \in \mathbb{Z}_+^2}$ désignera donc un champ de variables aléatoires réelles, indépendantes, équidistribuées, et telles que $E(X_1) = 0$ et $E(X_1^2) = 1$. On définit alors le processus de sommes partielles attaché au champ $(X_i)_i$ de la façon suivante. $\mathbb{R}^2$ et $\mathbb{Z}^2$ étant munis de l'ordre produit usuels, on pose $S_p = \sum_{i \le p} X_i$ pour tout $p \in \mathbb{Z}_+^2$.

Les propriétés asymptotiques du processus $\{S_p : p \le (\nu, \nu)\}$ ont été étudiés pour la première fois par Wichura (73), qui a montré que le processus $\{(2LL\nu)^{-1/2}\nu^{-1}S_{\lfloor nx \rfloor} : x \in [0,1]^2\}$ suit une loi du logarithme itéré fonctionnelle au sens de V. Strassen. Puis, P. Major (76), à la suite des travaux de Komlós, Major, et Tusnády sur l'approximation forte des sommes partielles de variables aléatoires indépendantes, a montré que le principe d'invariance fort obtenu par Strassen s'étend aux processus de sommes partielles indexés par les quadrants: autrement dit, on peut construire un champ $(Y_i)_{i \in \mathbb{Z}^2}$ de variables aléatoires gaussiennes centrées réduites tel que, si $T_p = \sum_{i \le p} Y_i$,

$$\sup_{p \le (\nu,\nu)} | S_p - T_p | = o(\nu (LL\nu)^{1/2}) \ \ p.s.$$

Ainsi, toute loi du logarithme itéré qui vaut pour le processus Gaussien vaut aussi pour le processus de sommes partielles non Gaussien. Cependant, pour de nombreux problèmes, il est interessant d'avoir un terme d'erreur plus petit dans le principe d'invariance ci-dessus. Pour réduire le terme d'erreur, il est alors nécessaire de supposer que la loi de X_1 satisfait des hypothèses supplémentaires d'intégrabilité (voir aussi Major :76).

Dans la première partie de la thèse, nous avons démontré, dans le cadre plus général des processus de sommes partielles indexés par des classes de Vapnik-Chervonenkis, le résultat suivant. Soit ψ une fonction monotone continue de $\mathbb{R}_+$ dans lui-même telle que:

(i) $E(\psi(| X_1 |)) < +\infty$ et $x \to x^{-2}\psi(x)$ est continue, croissante sur $\mathbb{R}_+$.

(ii) il existe un réel r strictement inférieur à 4 tel que $x \to x^{-r}\psi(x)$ soit décroissante pour x assez grand.

(Notons qu'une telle fonction dépend, bien sur, de la loi de X_1). Alors si on définit la fonction φ par:

$$\varphi(x) = \psi^{-1}(x)(1 + (LLx/x)^{1/2}\psi^{-1}(x))$$

on peut construire un champ $(Y_i)_{i \in bkZ^+}$ de variables aléatoires gaussiennes centrées réduites indépendantes tel que:

$$\sup_{p \leq (\nu,\nu)} | S_p - T_p | = o(\varphi(\nu^2)) \quad p.s.$$

Ce résultat généralise un résultat précédent de U. Einmahl (87) sur l'approximation des sommes partielles en dimension 1 et est de plus optimal. Par contre, quand les variables X_i ont un moment d'ordre supérieur à 4 (i.e. il existe $r > 4$ tel que $E(| X |^r) < +\infty$), la construction des champs $(X_i)_i$ et $(Y_i)_i$ proposée dans la première partie donne le résultat suivant:

$$\sup_{p \leq (\nu,\nu)} | S_p - T_p | = o((\nu \log \nu)^{1/2}) \quad p.s.$$

Ce résultat, dont l'analogue pour la classe des boules euclidiennes est optimal au facteur $(\log n)^{1/2}$ près, n'est pas satisfaisant pour la classe des quadrants. De fait, nous nous proposons de montrer ici de montrer que le principe d'invariance fort vaut avec une erreur de l'ordre de $\nu^{2/r}$, à une puissance de $\log n$ près et donc que l'ordre de grandeur de l'écart est entièrement différent pour les deux classes ci-dessus. On retrouve ici une différence de comportement déjà connue en dimension 2 pour le processus empirique associé à un n-échantillon de loi uniforme: on pourra, par exemple, comparer le résultat de Massart (87) pour les classes de Vapnik-Chervonenkis et celui de G. Tusnady (78) pour la classe des quadrants, qui prouve que l'écart uniforme entre le processus empirique et le pont Brownien est au plus de l'ordre de $(\log n)^2$. Détaillons maintenant les résultats que nous obtenons.

Définitions et résultats.

Pour démontrer le résultat annoncé dans l'introduction, nous allons procéder ainsi: on démontre d'abord un principe d'invariance fort avec une inégalité exponentielle accompagnante pour un champ de variables indépendantes, équidistribuées et bornées par une constante K, dans laquelle les constantes sont calculées en fonction de K. Puis on en déduit les vitesses d'approximation dans le cas des moments à l'aide de troncatures. Cette démarche est identique à celle utilisée par U. Einmahl (89) pour généraliser les théorèmes de Komlos, Major et Tusnady aux sommes de vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^k$. Enonçons donc le résultat principal.

THEOREME 1. *Soit $(X_i)_{i\in\mathbb{Z}_+^2}$ un champ de variables indépendantes équidistribuées, centrées, réduites et bornées par K. Alors, pour tout couple (N_1, N_2) d'entiers positifs, on peut construire un champ $(Y_i)_{i\in\mathbb{Z}_+^2}$ de variables aléatoires indépendantes, Gaussiennes, centrées et réduites, tel que, si $N = N_1 + N_2$, pour tout t positif,*

$$pr\Big(\sup_{p\leq(2^{N_1},2^{N_2})} \mid S_p - T_p \mid \geq cKN(N+t)\Big) \leq \Lambda e^{-t}$$

Λ et c étant des constantes universelles positives.

A partir du théorème 1, on peut alors obtenir des principes d'invariance avec des vitesses optimales à une puissance de $\log\nu$ près. Soit $\mathcal{H}$ l'ensemble des fonstions ψ de $\mathbb{R}_+$ dans lui-même telles que $x^{-4}\psi(x)$ soit continue et croissante, et telles qu'il existe un réel positif r tel que $x^{-r}\psi(x)$ soit décroissante.

COROLLAIRE 1. *Soit $(X_i)_{i\in\mathbb{Z}_+^2}$ un champ de variables i.e.d. centrées et réduites. Supposons de plus qu'il existe une fonction ψ dans $\mathcal{H}$ telle que $E(\psi(\mid X_1 \mid) < +\infty$. Alors, on peut construire un champ $(Y_i)_{i\in\mathbb{Z}_+^2}$ de variables aléatoires Gaussiennes centrées et réduites tel que:*

$$\sup_{p\leq(\nu,\nu)} \mid S_p - T_p \mid = o(\psi^{-1}(\nu^2)(\log\nu)^2) \quad p.s.$$

Remarque. D'après une remarque de Breiman (1967), ce résultat est optimal, au facteur $(\log\nu)^2$ près.

Quand X_1 a une transformée de Laplace finie sur un voisinage de l'origine, on peut facilement déduire du théorème 1 le corollaire suivant:

COROLLAIRE 2. *Soit $(X_i)_{i\in\mathbb{Z}_+^2}$ un champ de variables aléatoires i.e.d. centrées réduites. Supposons de plus qu'il existe $t_0 > 0$ tel que $E(\exp|t_0X_1|) < +\infty$. Alors, il existe un champ $(Y_i)_{i\in\mathbb{Z}_+^2}$ de variables indépendantes Gaussiennes centrées réduites et une constante C_0 dépendant uniquement de la loi de X_1 tels que, pour tout t positif*

$$pr\Big(\sup_{p\leq(\nu,\nu)} \mid S_p - T_p \mid \geq C_0(\log\nu)^3(t+\log\nu)\Big) \leq e^{-t}$$

Remarques. Il résulte du corollaire 2 que l'écart presque sûr entre le processus de sommes partielles non Gaussien et le processus approximant Gaussien pour la norme uniforme est de l'ordre de $(\log n)^4$ au plus. Cependant, on peut penser que cette vitesse n'est pas encore optimale: en effet, pour le processus empirique, l'écart est de l'ordre de $(\log n)^2$ (voir G. Tusnady, 1978).

Nous allons maintenant démontrer le théorème 1. Cette preuve se fonde sur un une construction bidimensionnelle des champs X et Y à la hongroise.

2. UNE CONSTRUCTION BIDIMENSIONNELLE.

Dans la suite, on suppose que l'espace probabilisé Ω est assez grand pour contenir une variable aléatoire uniforme indépendante du champ de variables aléatoires considéré. Cette hypothèse permet d'utiliser le lemme suivant:

Lemme de Skorohod (76). Soient R_1 et R_2 deux espaces polonais, et soit Q une mesure de probabilité sur $R_1 \times R_2$ de loi marginale q sur R_2. Alors, il existe une application mesurable Ψ telle que, si U et V sont deux variables aléatoires indépendantes de lois respectives q et la loi uniforme sur $[0,1]$ sur un espace probabilisé Ω, $(\Psi(U,V),V)$ a pour loi Q.

On se propose maintenant de construire les deux champs finis du théorème 1. On se donne donc un champ $(Y_i)_{i\leq(2^{N_1},2^{N_2})}$ de variables aléatoires i.e.d. de loi normale centrée réduite et on se propose de construire un champ $(X_i)_{i\leq(2^{N_1},2^{N_2})}$ de variables aléatoires i.e.d. centrées réduites et bornées par K. Pour celà, nous utiliserons fréquemment le résultat suivant, du à A.I. Sakhanenko (1984).

THEOREME. -Sakhanenko- *Soit $(\xi_i)_{i\in\mathbb{Z}_+}$ une suite de variables aléatoires centrées indépendantes telle qu'il existe une constante α positive satisfaisant: pour tout entier i positif,*

$$\alpha E(|\xi_i| \exp |\alpha \xi_i|) \leq E(\xi_i^2)$$

Alors, on peut construire une suite de variables aléatoires η_i de loi normale $N(0, E(X_i^2))$ indépendantes telle que:

$$E\left(\exp \left|c\alpha \sup_{p \leq n} \sum_1^p (\xi_i - \eta_i)\right|\right) \leq 1 + \alpha B$$

c étant une constante universelle positive et B désignant l'écart type de $\sum_1^n \xi_i$.

Quand les variables ξ_i sont presque surement bornées par une même constante K_0, on peut choisir $\alpha = (2K_0)^{-1}$ dans le théorème ci-dessus. En appliquant l'inégalité de Markov, on obtient alors:

$$pr\left(\sup_{p \leq n} \left|\sum_1^p (\xi_i - \eta_i)\right| \geq 2K_0 c^{-1}(t + \log(1+n))\right) \leq e^{-t} \tag{3.1}$$

Ce résultat remarquable est à la base des majorations et de la construction que nous allons maintenant détailler.

Le champ de variables aléatoires Gaussiennes étant donné, on construit dans un premier temps un champ auxiliaire $(Z_i^0)_{i \leq (2^{N_1}, 2^{N_2})}$ de variables aléatoires indépendantes ayant la loi des variables X_i. Puis on construit une permutation aléatoire Π indépendante du champ $(Z_i)_i$ par étapes successives et on pose $X_i = Z_{\Pi(i)}$. Comme les variables Z_i sont indépendantes et équidistribuées, il en va alors de même pour le champ $(X_i)_i$ ainsi défini. Mais, avant de définir le champ $(Z_i)_i$, nous allons introduire une filtration croissante associée au champ de variables aléatoires Gaussiennes.

Une filtration croissante. Dans cette partie, $]l, m]$ sera considéré comme un sous-intervalle de $\mathbb{Z}_+$. $l^2(\mathbb{Z}_+^2)$ est muni du produit scalaire usuel, et l'on confondra les parties de $\mathbb{Z}_+^2$ et leurs fonctions indicatrices. Pour définir le champ X, il sera commode d'introduire le système orthogonal dyadique suivant: soit $I_{j,p} =]p2^j, (p+1)2^j]$, et soit $e_{j,p}$ la fonction indicatrice de $I_{j,p}$. pour tout couple (p, j) d'entiers positifs, on pose $\tilde{e}_{j,p} = e_{j,p} - 2e_{j-1,2p}$.

Il est facile de montrer que le système $\{e_{N_2,0}, \tilde{e}_{j,p} : 0 < j \leq N_2, 0 \leq p < 2^{N_2 - j}\}$ est une base orthogonale de $l^2(]0, 2^{N_2}])$. Pour tout champ $(u_i)_{i \in \mathbb{Z}^2}$ de vecteurs, pour toute fonction f de $l^2(\mathbb{Z}^2)$ à support fini, on pose $u(f) = \sum_{i \in \mathbb{Z}^2} f(i) u_i$. Pour construire le champ X, il suffit de construire les vecteurs aléatoires $U_{l,N_2,0} = X(e_{0,l} \otimes e_{N_2,0})$ et $\tilde{U}_{l,j,p} = X(e_{0,l} \otimes \tilde{e}_{j,p})$ à partir des vecteurs correspondants attachés au champ Gaussien Y. Aussi, on pose:

$$V_{l,j,p} = Y(e_{0,l} \otimes e_{j,p}) \text{ et } \tilde{V}_{l,j,p} = Y(e_{0,l} \otimes \tilde{e}_{j,p})$$

On suppose qu'il existe une suite $(\delta_i)_{i>0}$ de variables aléatoires indépendantes de loi uniforme sur un intervalle, indépendante du champ de Gaussiennes Y, et on définit les tribus $\mathcal{F}_j$ et $\mathcal{G}_j$ ainsi : $\mathcal{F}_{N_2}$ est la tribu engendrée par δ_{N_2} et les variables Gaussiennes $V_{l,0,N_2}$, l décrivant l'intervalle $]0, 2^{N_2}]$; $\mathcal{G}_j$ est la tribu engendrée par δ_j et par le système de variables Gaussiennes $\{\tilde{V}_{l,j,p} : l \leq 2^{N_1} : p \leq 2^{N_2-j}\}$ et les tribus $\mathcal{F}_j$ sont définis par la relation de récurrence descendante $\mathcal{F}_{j-1} = \mathcal{F}_j \vee \mathcal{G}_j$.

La suite de tribus $(\mathcal{F}_j)_{j \leq N_2}$ ainsi définis est évidemment décroissante; de plus $\mathcal{G}_j$ est indépendante de $\mathcal{F}_j$ et représente donc *l'innovation dyadique* nécessaire pour passer de $\mathcal{F}_j$ à $\mathcal{F}_{j-1}$. Nous allons maintenant construire le champ X par récurrence de telle sorte que les variables $U_{l,0,N_2}$ soient dans la tribu $\mathcal{F}_{N_2}$, et les variables $\tilde{U}_{l,j,p}$ soient dans la tribu $\mathcal{F}_{j-1}$. Le premier pas est la construction d'un champ Z de variables équidistribuées, indépendantes ayant la loi F des variables X_i.

<u>Définition du champ Z.</u>

D'après le théorème de Sakhanenko et le lemme de Skorohod, il existe un champ Z de variables aléatoires i.i.d. de loi F tel que, si l'on pose $U_{l,N_2,0} = \sum_{i \leq 2^{N_2}} Z_{l,i}$, alors, pour tout t positif,

$$pr\left(\sup_{m \leq 2^{N_1}} \Big|\sum_{l \leq m} (U_{l,N_2,0} - V_{l,N_2,0})\Big| \geq cK(N+t)\right) \leq e^{-t} \tag{3.2}$$

Le champ que nous venons de définir réalise donc une bonne approximation du champ Gaussien sur les rectangles de la forme $]0, l] \times]0, 2^{N_2}]$. Pour définir le champ X, il est donc naturel de procéder à l'aide de permutations portant sur les variables $Z_{l,i}$ laissant globalement invariantes chaque colonne.

<u>Définition du champ X et majorations.</u>

En reprenant la démarche utilisée par Komlos, Major et Tusnady (75) pour démontrer un résultat d'approximation forte de type Kiefer pour le processus empirique, on peut remarquer que, si il existe un champ X de variables aléatoires i.i.d. vérifiant l'inégalité 3.2 et satisfaisant de plus les inégalités suivantes:

$$pr\left(\sup_{m\leq 2^{N_1}} \left|\sum_{l\leq m}(\tilde{U}_{l,j,p} - \tilde{V}_{l,j,p})\right| \geq cK(N+t)\right) \leq e^{-t} \tag{3.3}$$

pour tout t positif, pour tout $j \in]0, N_2]$ pour tout $p < 2^{N_2-j}$, alors les champs X et Y ainsi construits vérifient le théorème 1.

Rappelons maintenant que nous allons obtenir le champ X au moyen de permutations du champ Z laissant invariantes les colonnes. Le premier pas est alors de définir un champ Z^1 de variables aléatoires indépendantes de loi F à partir du champ Z et de la tribu $\mathcal{G}_{N_2}$ de la forme $Z_i^1 = Z_{\Pi_1(i)}$, Π_1 étant une permutation aléatoire du rectangle $]0, 2^{N_1}]\times]0, 2^{N_2}]$ laissant globalement invariantes les colonnes, indépendante du champ Z, $\mathcal{F}_{N_2-1}$-mesurable et telle que, si l'on pose $\tilde{U}_{l,N_2,0} = Z^1(e_{0,l} \otimes \tilde{e}_{N_2,0})$, alors les vecteurs ainsi définis vérifient l'inégalité 3.3 . Supposons pour l'instant que l'on puisse construire un tel champ: nous détaillerons ultérieurement la méthode de construction utilisée (voir (*)).

Pour définir le champ X on procède alors par récurrence : si l'on a défini un champ Z^k $\mathcal{F}_{N_2-k}$-mesurables à l'aide de permutations laissant globalement invariantes les colonnes tel que, pour tout $j > N_2 - k$, pour tout $p < 2^{N_2-j}$, les suites de variables $(\tilde{U}^k_{l,j,p})_{l>0}$ associées au champ Z^k et les variables $(\tilde{V}_{l,j,p})_{l>0}$ vérifient l'inégalité 3.3, alors, au pas suivant, on définit un champ Z^{k+1} $\mathcal{F}_{N_2-k-1}$-mesurable à partir du champ Z^k possédant la propriété suivante:

- Pour tout $p < 2^k$, $\sum_{i\in I_{N_2-k,p}} Z_i^{k+1} = \sum_{i\in I_{N_2-k,p}} Z_i^k$

Il est clair que les accroissements dyadiques $\tilde{U}^{k+1}_{l,j,p}$ associés au champ Z^{k+1} sont, pour tout $j > N_2 - k$, identiques à ceux associés au champ Z^k. On peut donc remplacer Z^k par Z^{k+1} dans les inégalités précédentes. Mais on demande aussi au champ Z^{k+1} de réaliser une bonne approximation à l'échelle 2^{N_2-k-1}. Pour cela, on construit Z^{k+1} à partir du champ Z^k au moyen d'une permutation aléatoire indépendante laissant globalement invariantes les parties de la forme $\{l\} \times I_{N_2-k,p}$. De plus, pour chaque $p \in [0, 2^k[$, la restriction de cette permutation au sous rectangle $]0, 2^{N_1}] \times I_{N_2-k,p}$ est construite à partir des variables $\tilde{V}_{l,N_2-k,p}$ et de la restriction du champ Z^k à ce même sous-rectangle par un procédé identique à celui utilisé pour construire Z^1 à partir de Z. Alors les variables $\tilde{U}_{l,N_2-k,p} = Z^{k+1}(e_{0,l} \otimes \tilde{e}_{N_2-k,p})$ vérifient l'inégalité suivante:

$$pr\big(\sup_{l\leq 2^{N_1}} \mid \sum_{m=1}^{l} |\tilde{U}_{l,N_2-k,p} - \tilde{V}_{l,N_2-k,p}| \geq cK(N-k+t)\big) \leq e^{-t}$$

Au dernier pas de la construction, on pose $Z_{l,i}^{N_2} = X_{l,i}$. Le champ X est alors un champ de variables aléatoires indépendantes de loi F et les champs X et Y vérifient le théorème 1. Il reste maintenant à expliquer comment construire Z^1 à partir de Z et de la tribu $\mathcal{G}_{N_2}$.

(*) Construction du champ Z^1.

Rappelons que l'on se propose de construire un champ Z^1 tel qu'il existe une permutation Π aléatoire, indépendente de Z satisfaisant: pour tout couple (l,i), $Z^1_{(l,i)} = Z_{\Pi((l,i))}$. Définissons tout d'abord la forme et la loi de la permutation Π: Π sera un produit de transpositions entre (l,i) et $(l,i+2^{N_2-1})$, i Θtant un entier plus petit que 2^{N_2-1}, chacune de ces transpositions ayant une probabilité d'occurence dans le produit égale à 1/2, les occurences étant indépendantes entre elles.

Se donner une telle permutation équivaut à se donner un champ $(\varepsilon_{l,i})_{(l\leq 2^{N_1}, i\leq 2^{N_2-1})}$ de signes symétriques indépendants, indépendant du champ Z. Posons $z_{l,i} = Z_{l,i} - Z_{l,i+2^{N_2-1}}$ D'après ce qui précède, pour construire le champ Z^1, il suffit de construire un champ $(\varepsilon_{l,i})_{l,i}$ de signes symétriques indépendants de telle sorte que

$$pr\big(\sup_{l\leq 2^{N_1}} \mid \sum_{m\leq l}(\tilde{V}_{l,N_2,0} - \sum_{i} \varepsilon_{l,i} z_{l,i})| \geq cK(N+t)\big) \leq e^{-t}$$

D'après le lemme de Skorohod, un tel champ existe si il existe, dans un autre espace probabilisé, trois suites $(\varepsilon_i)_{i\leq 2^{N-1}}$, $(z_i)_{i\leq 2^{N-1}}$, et $(y_i)_{i\leq 2^{N-1}}$ de variables aléatoires telles que:

- Chacune des trois suites est une suite de variables aléatoires indépendantes équidistribuées.
- La suite $(z_i)_{i>0}$ est indépendante de la suite $(\varepsilon_i)_{i>0}$ ainsi que de la suite (y_i).
- Les v.a. ε_i sont des variables de signe symétriques, les variables z_i ont pour loi $F * \check{F}$, et les variables y_i sont des Gaussiennes centrées de variance 2.

Ces trois suites vérifiant l'inégalité suivante : pour tout t positif,

$$pr(\sup_{l\leq 2^{N-1}} |\sum_1^l y_i - z_i\varepsilon_i| \geq cK(N+t)) \leq e^{-t} \tag{3.4}$$

c étant une constante universelle positive.

On aura donc achevé la preuve du théorème 1 quand on aura démontré 3.4 . Or, d'après le théorème de Sakhanenko, si $(z_i)_{i>0}$ est une suite de variables indépendantes de loi $F * \check{F}$ (Rappelons que ces variables sont bornées par $2K$ en valeur absolue) il existe des suites $(\varepsilon_i)_{i>0}$ de signes symétriques indépendants, et $(\eta_i)_{i>0}$ de Gaussiennes centrées réduites indépendantes chacune indépendante de la suite $(z_i)_{i>0}$ et telles que

$$pr(\sup_{l\leq 2^{N-1}} |\sum_1^l z_i(\varepsilon_i - \eta_i)| \geq cK(N+t)) \leq e^{-t} \tag{3.5}$$

Donc, l'inégalité 3.4 est une conséquence du lemme suivant:

LEMME 1. *Soit $(z_i)_{i>0}$ une suite de variables équidistribuées de loi G à support dans $[-K, K]$ satisfaisant*

$$\int_{\mathbb{R}} x^2\, dG(x) = 1$$

et $(\eta_i)_{i>0}$ une suite de variables aléatoires Gaussiennes indépendantes centrées réduites, indépendante de la suite $(z_i)_{i>0}$. Alors, il existe une suite $(y_i)_{i>0}$ de Gaussiennes centrées réduites indépendantes et indépendante de la suite $(z_i)_{i>0}$ telle que, pour tout t positif,

$$pr(\sup_{l\leq 2^{N}} |\sum_1^l (y_i - z_i\eta_i)| \geq cK(N+t)) \leq e^{-t} \tag{3.6}$$

Preuve. Pour construire la suite $(y_i)_{i>0}$, nous allons introduire les notations suivantes: on pose $w_i = z_i\eta_i$ et

$$W_N = \sum_1^{2^N} z_i\eta_i\,,\ W_{j,p} = w(e_{j,p}) \text{ et } \tilde{W}_{j,p} = w(\tilde{e}_{j,p})$$

et on appelle encore V_N et $\tilde{V}_{j,p}$ les variables Gaussiennes correspondantes à définir (i.e. $V_N = y(e_{N,0})$ et $\tilde{V}_{j,p} = y(\tilde{e}_{j,p})$). On définit aussi les variances conditionnelles à la suite $(z_i)_{i>0}$ associées à la suite $(w_i)_{i>0}$ que l'on note D_N et $D_{j,p}$ par:

$$D_N = \sum_1^{2^N} z_i^2 \quad \text{et} \quad D_{j,p} = \sum_{i \in I_{j,p}} z_i^2$$

Puis on pose $\tilde{D}_{j,p} = D_{j,p} - 2D_{j-1,2p}$ et on définit les variables V_N et les variables $\tilde{V}_{j,p}$ ainsi: On pose $V_N = 2^{N/2} D_N^{-1/2} W_N$, et on définit $\tilde{V}_{j,p}$ pour tout j inférieur à N par:

$$\tilde{V}_{j,p} = 2^{j/2} Var(\tilde{V}^0_{j,p})^{-1/2} \tilde{V}^0_{j,p}$$

$\tilde{V}^0_{j,p}$ étant la variable Gaussienne obtenue par projection orthogonale de $\tilde{W}_{j,p}$ sur l'orthogonal de $W_{j,p}$. Ainsi:

$$\tilde{V}^0_{j,p} = \tilde{W}_{j,p} - D_{j,p}^{-1} \tilde{D}_{j,p} W_{j,p}$$

Conditionnellement aux v.a. $(z_i)_{i>0}$, les variables ainsi définies sont indépendantes, de loi Gaussienne et de variance ad hoc: on a donc défini une suite $(y_i)_{i \leq 2^N}$ de Gaussiennes indépendantes centrées et réduites (la preuve de cette assertion utilise des arguments analogues à ceux utilisés dans la première partie de la thèse).

Posons $S_p = \sum_1^p \eta_i z_i$ et $T_p = \sum_1^p y_i$. Pour démontrer que les deux suites $(w_i)_i$ et $(y_i)_i$ vérifient l'inégalité 3.6, on approxime p par l'entier $p(M)$ multiple de 2^M le plus proche de p, M étant un entier naturel à choisir. Alors,

$$| S_p - T_p | \leq |S_{p(M)} - T_{p(M)}| + |T_p - T_{p(M)}| + |S_p - S_{p(M)}|$$

Soit t un réel supérieur à 1. Si on choisit M tel que $2^{M-1} \leq K^2 t < 2^M$, alors, en appliquant l'inégalité de Bernstein (voir Bennett (62)), on obtient:

$$pr(| S_p - S_{p(M)} | \geq 4Kt) \leq 2e^{-t/4}$$

Comme les variables y_i sont indépendantes, de loi $N(0,1)$, il est clair que $| T_p - T_{p(M)} |$ vérifie la même inégalité; si $M > N$, on peut prendre $p(M) = 0$; alors $S_{p(M)} = T_{p(M)} = 0$, et d'après ce qui précède, pour tout $p \leq 2^N$,

$$pr(| T_p - S_p | \geq 8Kt) \leq 4e^{-t/4} \tag{3.7}$$

Quand $M \leq N$, on décompose $S_{p(M)} - T_{p(M)}$ à l'aide des accroissements diadiques. Ainsi, on peut montrer que

$$2 \mid T_{p(M)} - S_{p(M)} \mid \leq \sum_{j=M+1}^{N} |\tilde{V}_{j,p(j)} - \tilde{W}_{j,p(j)}| + 2|V_N - W_N|$$

La suite d'entiers $p(j)$ étant telle que les intervalles $I_{j,p(j)}$ forment une suite croissante pour l'inclusion. Par construction, la loi de la variable aléatoire de droite dans l'inégalité ci-dessus ne dépend pas de la suite d'entiers $p(j)$. Donc, si l'on pose:

$$\Delta = \frac{1}{2} \sum_{j=M+1}^{N} \mid \tilde{V}_{j,0} - \tilde{W}_{j,0} \mid + \mid V_N - W_N \mid$$

alors, pour conclure la preuve du lemme 1, il suffit de montrer que Δ satisfait une inégalité analogue à l'inégalité 3.6 (voir K.M.T. pages 119-121) c'est à dire qu'il existe une constante universelle c_0 telle que, pour tout t positif,

$$pr(\Delta \geq c_0 K(N+t)) \leq e^{-t} + 2^N e^{-t/7} \tag{3.8}$$

Preuve de 3.8. Clairement,

$$\mid \tilde{W}_{j,0} - \tilde{V}_{j,0} \mid \leq \mid \tilde{W}_{j,0} - \tilde{V}^0_{j,0} \mid + \mid \tilde{V}^0_{j,0} - \tilde{V}_{j,0} \mid$$

Majorons la première des deux variables de droite:

$$\frac{1}{2} \mid \tilde{V}_{j,0} - \tilde{W}_{j,0} \mid \leq D^{-1}_{j,0}(KW^2_{j,0} + K^{-1}\tilde{D}^2_{j,0})$$

En utilisant des arguments analogues, il est facile de montrer que:

$$\frac{1}{2} \mid \tilde{V}^0_{j,0} - \tilde{V}_{j,0} \mid \leq 2^{-j} K \tilde{V}^2_{j,0} + 2^{-j} K^{-1} \tilde{D}^2_{j,0} + 2^{-j} K^{-1} (D_{j,0} - 2^j)^2$$

On note alors que pour tout $p \leq 2^{N-M-1}$, d'après l'inégalité de Bernstein,

$$pr(\mid D_{M+1,p} - 2^{M+1} \mid \geq 2^M) \leq 2.e^{-t/7}$$

Si on définit l'évènement Θ par

$$\Theta = \bigcap_{p=0}^{2^{N-M-1}} (|D_{M+1,p} - 2^{M+1}| \leq 2^M)$$

alors Θ a une contre-probabilité inférieure à $2^N e^{-t/7}$, et pour tout ω dans Θ,

$$\Delta \leq | V_N - W_N | + 2K \sum_{j>M} 2^{-j}(\tilde{V}_{j,0}^2 + W_{j,0}^2) + \frac{1}{K} \sum_{j>M} 2^{-j}(3\tilde{D}_{j,0}^2 + (D_{j,0} - 2^j)^2)$$

Nous laissons au lecteur le soin de majorer $| V_N - W_N |$ et nous allons maintenant majorer les autres termes de la somme de droite dans l'inégalité ci-dessus: le second terme est une somme de carrés de $N - M$ variables aléatoires Gaussiennes indépendantes centrées et réduites; par conséquent, en utilisant le calcul à la Cramer-Chernoff usuel, on obtient:

$$pr(\sum_{j>M} 2^{-j}\tilde{V}_{j,0}^2 \geq c_1 N + t) \leq \exp(-c_2 t)$$

c_1 et c_2 étant des constantes universelles positives.

Le controle du second utilise l'argument de sommes *glissantes* donné par Komlos, Major, et Tusnady (75, I p.120),

$$\mathbb{1}_\Theta \sum_{j>M} 2^{-j} W_{j,0}^2 \leq \mathbb{1}_\Theta (2^{-M} W_{M,0}^2 + 4 \sum_{M}^{N-1} 2^{-j} W_{j,1}^2) \leq$$

$$6 D_{M,0}^{-1} W_{M,0}^2 + 6 \sum_{M}^{N-1} D_{j,1}^{-1} W_{M,1}^2 \qquad (3.9)$$

Or, le majorant ci-dessus est à un coefficient près la somme de $N+1-M$ carrés de variables Gaussiennes indépendantes centrées réduites, et donc se controle comme précédemment.

Enfin, pour majorer le dernier terme, on note que

$$3\tilde{D}_{j,0}^2 + D_{j,0}^2 \leq 6\big((D_{j-1,0} - 2^{j-1})^2 + (D_{j-1,1} - 2^{j-1})^2\big)$$

et, en utilisant à nouveau l'argument de sommes glissantes, on obtient:

$$\sum_{j>M} 2^{-j}(3\tilde{D}_{j,0}^2 + D_{j,0}^2) \leq 24 \sum_{M}^{N-1} 2^{-j}(D_{j,1} - 2^j)^2 + 24.2^{-M}(D_{M,0} - 2^M)^2$$

Définissons les variables aléatoires T_j pour j supérieur à M par

$$T_M = 2^{-M}(D_{M,0} - 2^M)^2 \mathbb{1}_{|D_{M,0} - 2^m| \leq 2^{M-1}}$$

et

$$T_j = 2^{1-j}(D_{j-1,1} - 2^{j-1})^2 \mathbb{1}_{|D_{j-1,1} - 2^{j-1}| \leq 2^{j-2}}$$

pour $j > M$. Les variables T_j ainsi définies sont indépendantes. De plus, en intégrant par parties, on obtient:

$$E(\exp(tT_j)) = \int_0^{2^{j-3}} te^{tu} pr(||D_{j-1,1} - 2^{j-1}| \geq \sqrt{2^{j-1}t})\, du$$

Remarquons maintenant que $D_{j-1} - 2^{j-1}$ est une somme de variables indépendantes centrées bornées par K^2 et de variance inférieure à K^2; donc en appliquant l'inégalité de Bernstein, on obtient:

$$pr(|D_{j-1,1} - 2^{j-1}| \geq \sqrt{2^{j-1}u}) \leq 2\exp(-\frac{u}{4K^2})$$

On en déduit immédiatement que les variables T_M et T_j vérifient:

$$E(\exp(T_j(8K^2)^{-1})) \leq 3$$

Puis en utilisant le lemme 4 de l'article de P. Massart (89), il vient

$$pr(\sum_M^N T_j \geq c_3 K^2(N+t)) \leq 2e^{-t}$$

pour tout t positif, c_3 étant une constante universelle positive. En récapitulant ce qui précède, on a donc montré l'inégalité 3.8. Donc, en combinant 3.7 et 3.8, on a:

$$pr(|\ S_p - T_p\ | \geq c_4 K(N+t)) \leq e^{-t}$$

pour tout t positif, c_4 étant une c.u. positive et Le lemme 1 découle immédiatement de cette inégalité. Le théorème 1 est donc démontré

3. APPROXIMATION FORTE.

Dans cette partie, nous démontrons à partir du résultat principal les théorèmes 2 et 3. Pour cela, on construit les champs $(X_i)_{i\in\mathbb{Z}_+^2}$ et $(Y_i)_{i\in\mathbb{Z}_+^2}$ à l'aide de la construction donnée dans la deuxième partie comme suit.

Pour tout entier N, $\mathcal{D}_N =]0, 2^{N+1}]^2 \setminus]0, 2^N]^2$ est réunion disjointe de trois carrés de cotés de longueur 2^N. Supposons avoir construit les champs X et Y sur le carré $]0, 2^N]^2$. Au pas suivant, on construit alors de façon indépendante et indépendemment des champs déjà construits les champs X et Y sur chacun des trois carrés de coté 2^N ci-dessus mentionnés. Détaillons cette construction dans le cadre de la preuve du théorème 2.

Supposons donné le champ $(X_i)_{i \in \mathbb{Z}_+^2}$ de variables indépendantes de loi F centrée réduite et telle que $\int_{\mathbb{R}} \psi(|x|)\, dx < +\infty$. On définit alors à partir de X le champ $\bar{X}$ de variables tronquées par:

$$\forall i \in \mathcal{D}_N, \bar{X}_i = X_i \mathbb{1}_{|X_i| \leq \psi^{-1}(2^{2N})}$$

Puis, à partir de $\bar{X}$, le champ renormalisé centré $\tilde{X}$ par:

$$\tilde{X}_i = (Var\, X_i)^{-1/2}(\bar{X}_i - E(\bar{X}_i))$$

Il est clair que les variables $\tilde{X}_i$ sont centrées, réduites et bornées par $2\psi^{-1}(2^{2N})$ quand i est dans $\mathcal{D}_N$. De plus, U. Einmahl (89) a démontré que les champs X et $\tilde{X}$ vérifient:

$$\sum_{i \leq (\nu,\nu)} |\tilde{X}_i - X_i| = o(\psi^{-1}(\nu^2)) \quad p.s.$$

Pour démontrer le théorème 2, il suffit donc de montrer que l'on peut construire un champ Y de Gaussiennes centrées réduites satisfaisant à l'inégalité suivante: il existe une constante positibe c_5 telle que, pour tout t positif,

$$pr(\sup_{p \leq (\nu,\nu)} |\sum_{i \leq p} (\tilde{X}_i - Y_i)| \geq c_5 \psi^{-1}(\nu^2) \log \nu (t + \log \nu)) \leq e^{-t} \tag{3.9}$$

Or, d'après le théorème 1 et le lemme de Skorohod, on peut construire un champ de variables aléatoires indépendantes de loi $N(0,1)$ tel que pour tout entier N positif, on ait:

$$pr(\sup_{p \in \mathbb{Z}^2} | \sum_{i \in \mathcal{D}_N \cap]0,p]} (Y_i - \tilde{X}_i)| \geq c_6 \psi^{-1}(2^{2N+2}) N(N+t)) \leq e^{-t}$$

Rappelons maintenant que pour $s = 1/r$, $x^{-s}\psi^{-1}(x)$ est une fonction croissante. Donc, en sommant les inégalités ci-dessus en N, on obtient immédiatement l'inégalité 3.9 pour les entiers puissances de 2. On a donc démontré le théorème 2.

Preuve du théorème 3.

On considère ici un champ $(X_i)_{i\in\mathbb{Z}_+^2}$ de variables équidistribuées centrées réduites ayant une transformée de Laplace finie sur $[-t_0, t_0]$. Posons $b_0 = E(\exp|t_0 X_1|)$. Les variables X_i sont alors dans la classe exponentielle $H(t_0, b_0)$ définie dans la première partie. En appliquant l'inégalité de Markov à $\exp|t_0 X_1|$, on peut facilement montrer qu'il existe une constante C positive telle que:

$$pr(|X_1| \geq C\,Lx) \leq x^{-4}\ ,\ E(X_1^2 \mathbb{1}_{|X_1|\geq C\,Lx}) \leq x^{-4}$$

On définit alors le champ de variables tronquées $\bar{X}$ à partir du champ X par $\bar{X}_i = X_i \mathbb{1}_{|X_i|\geq CN}$ pour tout i dans $\mathcal{D}_N$, puis le champ $\tilde{X}$ en renormalisant et en recentrant les variables $\bar{X}_i$. Puis, à partir du champ $\tilde{X}$ ainsi défini, on construit comme précédemment un champ Y de Gaussiennes centrées réduites et indépendantes tel que pour tout entier postif N,

$$pr(\sup_{p\in\mathbb{Z}_+^2} |\sum_{i\in\mathcal{D}_N\cap]0,p]} |\tilde{X}_i - Y_i| \geq c_7 N^2(N+t)) \leq e^{-t}$$

Pour démontrer le théorème 3, il suffit de vérifier que la meme inégalité est encore valable si l'on remplace Y_i par $\tilde{X}_i$. En fait, l'écart uniforme entre les champs $\bar{X}$ et $\tilde{X}$ se controle facilement (nous laissons ce point au lecteur) et le théorème 3 découle donc de ce qui précède et de la proposition suivante:

$$\sum_{i\in\mathcal{D}_N} |X_i - \bar{X}_i| \in \bar{H}(t_0/2, c_8)$$

c_8 étant une constante positive dépendant de la loi des variables X_i

<u>Preuve.</u> Puisque les variables considérées sont indépendantes et équidistribuées,

$$\log E\big(\exp(\frac{t_0}{2}\sum_{i\in\mathcal{D}_N} |X_i - \bar{X}_i|)\big) \leq 3.2^{2N} \log E(\exp(|tX_1|\mathbb{1}_{|X_1|\geq CN}))$$

Or, par définition de C, $pr(|X_1| \geq CN) \leq 2^{-4N}$. En appliquant l'inégalité de Cauchy-Schwarz, on en déduit que:

$$E\big(\exp(\frac{1}{2}|t_0 X_1|\mathbb{1}_{|X_1|\geq CN})\big) \leq 1 + 2^{-2N} b_0^{1/2}$$

Et la proposition découle immédiatement des deux inégalités ci-dessus. Le théorème 3 est donc démontré.

Quelques remarques sur les dimensions supérieures.

En utilisant la construction bidimensionnelle proposée dans cette partie et les arguments développés par Massart et par moi dans la première partie de cette thèse, il est possible de montrer que, à une puissance de $\log \nu$ près, l'approximation uniforme sur la classe des quadrants d'un champ $(X_i)_{i\in\mathbb{Z}_+^d}$ de variables équidistribuées centrées et ayant un moment d'ordre $r > 2$ fini vaut avec un terme d'erreur de l'ordre de $\nu^{(d-2)/2} \vee \nu^{d/r}$. Je ne pense pas, cependant, qu'un tel résultat soit optimal

REFERENCES

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

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

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

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

EINMAHL, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivariate Analysis* **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** 111-6131.

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.

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

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

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* 4-49. *Trudy Inst. Mat.* Vol. 3, "Nauka" Sibirsk.Otdel, Novosibirsk [Russian].

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

TUSNÁDY, G. (1978). A remark on the approximation of the sample DF in the multidimensional case. *Period. Math. Hung.* **8.** 53-55.

WICHURA, M.J. (1973). Some Strassen type laws of the iterated logarithm for multiparameter stochastic processes. *Ann. Probab.* **1.** 272-296.