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

MICHEL TALAGRAND

Applying a theorem of Fernique

Annales de l'I. H. P., section B, tome 32, nº 6 (1996), p. 779-799
<http://www.numdam.org/item?id=AIHPB_1996__32_6_779_0>

Ann. Inst. Henri Poincaré,

Vol. 32, n° 6, 1996, p. 779-799 *Probabilités et Statistiques*

Applying a theorem of Fernique

by

Michel TALAGRAND (*)

Équipe d'Analyse, Tour 46, E.R.A. au C.N.R.S. No. 754, Université Paris-VI,
4, place Jussieu, 75230 Paris Cedex 05, France and
Department of Mathematics, The Ohio State University,
231 W. 18th Ave., Columbus, OH 43210-1174.

ABSTRACT. – Consider a subset A of Hilbert space H, on which the canonical Gaussian process is bounded. Consider a family $\mathcal{U}$ of linear operators on H. Fernique found natural conditions under which the canonical Gaussian process remains bounded on $\mathcal{U}(A) = \{u(a); u \in \mathcal{U}, a \in A\}$. We provide applications, variations and refinements of this result. For example, we prove the following: if $\mathcal{F}$ is a Donsker class of functions on a probability space, if $\mathcal{C}$ is a Vapnik-Cervonenkis class of sets, the class of functions of type $f1_C$ $(f \in \mathcal{F}, C \in \mathcal{C})$ is Donsker. We also combine new entropy estimates in ergodic theory with Fernique's result to obtain recent results of Weber.

RÉSUMÉ. – **Applications d'un théorème de Fernique.** Soit A un GB-ensemble d'un espace de Hilbert, c'est-à-dire un ensemble où le processus canonique Gaussien reste borné. Soit $\mathcal{U}$ une famille d'opérateurs linéaires sur H. Fernique a proposé des conditions naturelles sous lesquelles $\mathcal{U}(A) = \{u(a); u \in \mathcal{U}, a \in A\}$ demeure un GB-ensemble. Nous proposons des raffinements et des applications de ce résultat. Nous montrons par exemple que si $\mathcal{F}$ est une classe de Donsker de fonctions sur un espace probabilisé, et si $\mathcal{C}$ est une classe de Vapnik-Cervonenkis d'ensembles, la classe des fonctions du type $f1_C$ pour $f \in \mathcal{F}, C \in \mathcal{C}$, demeure une classe de Donsker. Dans une direction tout à fait différente, nous montrons ce qui est sans doute le résultat principal de l'article. Il existe un nombre K tel

(*) Work partially supported by an NSF Grant.

A.M.S. Classification numbers: Primary 60 G 15, 47 A 35; Secondary 62 G 30.

que si T est une contraction d'un espace de Hilbert, si x est un vecteur de norme au plus 1, alors pour $0 < \varepsilon < 1$, on peut recouvrir l'ensemble de tous les vecteurs $y_n = n^{-1} \sum_{0 \leq i \leq n-1} T^i(x)$ pour $n \geq 1$ par au plus $K\varepsilon^{-2}$ boules de rayon ε. Combiné avec le résultat de Fernique, cela permet de retrouver sans peine des résultats récents de Weber.

---

1. INTRODUCTION

On a Hilbert space H is defined a canonical Gaussian process $(X_t)_{t \in H}$, the so called isonormal process of covariance $E(X_s X_t) = \langle s, t \rangle$. It became apparent in the late 60's that the understanding of which general Gaussian processes are sample bounded (resp. sample continuous) was equivalent to the understanding of the subsets A of H on which the restriction on the isonormal process is sample bounded (resp. sample continuous). These sets have been called GB (resp. GC) sets ever since. The study of the GB and GC sets culminated in their characterization (in 1985) through majorizing measures [T1] a result on which most of the present work relies.

Since majorizing measures are not easy to construct (or even to understand) this is however not the end of the story. Throughout this paper, we will measure the size of a GB set A by the natural quantity

$$G(A) = E \sup_{t \in A} X_t \tag{1.1}$$

The reader who worries about this potentially tricky supremum of possibly uncountably many random variables can assume A finite, and will lose nothing of the strength of the results we present.

We consider a family $\mathcal{U}$ of operators on H. For simplicity we assume they are contractions (*i.e.* of norm ≤ 1). It would actually suffice that $\mathcal{U}$ is equicontinuous (*i.e.* the norms of the elements of $\mathcal{U}$ are uniformly bounded) but this reduces to the previous case and requires one extra parameter.

We set

$$\mathcal{U}(A) = \{U(a); U \in \mathcal{U}, a \in A\}. \tag{1.2}$$

What would be really nice to have is the inequality

$$G(\mathcal{U}(A)) \leq K[G(A) + \sup_{a \in A} G(\mathcal{U}(a))] \tag{1.3}$$

where of course $\mathcal{U}(a) = \mathcal{U}(\{a\})$, and where K denotes, as in the entire paper, a universal constant (not necessarily the same at each occurrence).

Unfortunately, as will be shown in Section 3, (1.3) does not hold in general, and one has to replace $G(\mathcal{U}(a))$ by a somewhat larger quantity, that we introduce now.

Given a metric space (T, d), we recall that $N(T, d, \epsilon)$ denotes the smallest number of (open) balls of radius ϵ needed to cover T. Given $a \in A$, we consider the distance d_a on $\mathcal{U}$ given by

$$d_a(U, V) = \|U(a) - V(a)\|. \tag{1.4}$$

THEOREM 1.1. – *(Fernique [F]) For all families $\mathcal{U}$ of contractions, we have*

$$G(\mathcal{U}(A)) \leq K\left[G(A) + \sup_{a \in K} \int_0^\infty \sqrt{\log N(\mathcal{U}, d_a, \epsilon)}d\epsilon\right]. \tag{1.5}$$

Since by the so called "metric entropy bound"

$$G(\mathcal{U}(a)) \leq \int_0^\infty K\sqrt{\log N(\mathcal{U}, d_a, \epsilon)}d\epsilon$$

(1.5) is certainly weaker than (1.3).

It must be said that the principle used in Theorem 1.1 can be formulated in more general situations; but the somewhat restricted setting we use here seems appropriate for the present purposes.

While (1.3) does not hold in general, it turns out that if one controls the quantity $G(\mathcal{U}(a))$, not only when $a \in A$, but when $\|a\| \leq 1$, one can use this quantity rather than the entropy integral of (1.5).

THEOREM 1.2. – *For all $A \subset H$, all families $\mathcal{U}$ of operators, we have*

$$G(\mathcal{U}(A)) \leq K[G(A) + \sup_{\|x\| \leq 1} G(\mathcal{U}(x)) \sup_{x \in A} \|x\|] \tag{1.6}$$

The simple proof of Theorem 1.2, that will be given in Section 2, leaves considerable room. For example, setting $\alpha = \sup_{x \in A} \|x\|$, the same proof shows that

$$\begin{aligned} G(\mathcal{U}(A)) \leq & K(G(A) \\ & + \sum_{n \geq 1} \sup\{G(\mathcal{U}(x)); x \in A - A, \|x\| \leq 2^{-n}\alpha\}) \end{aligned} \tag{1.7}$$

However, an elegant common extension to Theorem 1.1 and Theorem 1.2 remains to be found.

In Section 4, we will show how either Theorem 1.1 or Theorem 1.2 imply recent results of Weber related to Ergodic theory. This application will be based on the following estimate, that is conceivably the main result of the present paper. Consider an isometry T of H, and set

$$U_n = \frac{1}{n}(I + T + \cdots + T^{n-1}) \tag{1.8}$$

THEOREM 1.3. – *For some universal constant K, any isometry U and any $x \in H$, we have, for all $0 < \epsilon \leq \|x\|$,*

$$N(\{U_n(x), n \geq 1\}, \epsilon) \leq \frac{K\|x\|^2}{\epsilon^2}.$$

In Section 5, we will apply either Theorem 1.1 or Theorem 1.2 to the theory of Donsker classes. (A Donsker class $\mathcal{F}$ of functions is essentially a class where the central limit theorem holds uniformly). The most noticeable feature of the next result is that it holds without extra conditions on $\mathcal{F}$.

THEOREM 1.4. – *If $\mathcal{F}$ is a Donsker class of functions on a probability space, and $\mathcal{C}$ is a Vapnik-Cervonenkis class of sets, the class of functions of the type $f1_C(f \in \mathcal{F}, C \in \mathcal{C})$ is a Donsker class of functions (under measurability).*

It turns out that the natural proof of Theorem 1.1 yields a somewhat stronger result. While this strengthening does not seem to be relevant in any conceivable application, it does have some theoretical interest. For a metric space (T, d), we introduce the quantity

$$S(T, d) = \sup_{t \in T} \int_0^\infty \sqrt{\log N(B(t, 2^7\epsilon), d, \epsilon)}\, d\epsilon$$

where $B(t, 2^7\epsilon)$ is the d-ball centered at t, of radius $2^7\epsilon$. It is obvious that $S(T, d)$ is dominated by the metric entropy integral

$$I(T, d) = \int_0^\infty \sqrt{\log N(T, \epsilon, d)}\, d\epsilon.$$

It is however simple to construct situations where $I(T, d) = \infty$ but $S(T, d) < \infty$. The reader who wishes to truly understand majorizing measures and abstract Gaussian processes should complete the mostly routine proof of the fact that for $A \subset H$, we have

$$G(A) \leq KS(A, d) \tag{1.9}$$

where d denotes the distance induced by the norm, and should construct examples where $G(A) < \infty, S(A,d) = \infty$.

The quantity $S(A,d)$ can be seen as a kind of intermediate measure of the size of A between $G(A)$ and $I(A,d)$, and it is actually rather different from either of these. In Section 6, we will prove the following refinement of Theorem 1.1.

THEOREM 1.5. – *For all $A \subset H$, all families $\mathcal{U}$ of operators, we have*

$$G(\mathcal{U}(A)) \leq K[G(A) + \sup_{a \in A} S(\mathcal{U}, d_a)]. \tag{1.10}$$

2. PROOF OF THEOREM 1.2

It is not known at the present time how to prove any result in the line of Theorem 1.1 without relying upon the majorizing measure theorem of [T1]. The form of the theorem we use (*see* [T3 section 2]) is as follows. Consider the largest integer n_0 with $2^{-n_0} \geq diamA$ (when working with subsets of H the distance we use is always induced by the norm.) Then we can find an increasing sequence $(\mathcal{B}_n)_{n \geq n_0}$ of finite partitions of A such that each element of $\mathcal{B}_n$ has diameter $\leq 2^{-n}$, that $\mathcal{B}_{n_0} = \{\mathcal{A}\}$, and a probability measure μ on A such that, if for $a \in A$, we denote by $\mathcal{B}_n(a)$ the unique element of $\mathcal{B}_n$ that contains a, we have

$$\forall a \in A, \sum_{n \geq n_0} 2^{-n} \sqrt{\log \frac{1}{\mu(B_n(a))}} \leq KG(A). \tag{2.1}$$

Conversely, for any such probability measure μ on A, and any increasing sequence of partitions such that each element of $\mathcal{B}_n$ has diameter at most 2^{-n} we have

$$G(A) \leq K \sup_{a \in A} \sum_{n \geq n_0} 2^{-n} \sqrt{\log \frac{1}{\mu(B_n(a))}} \tag{2.2}$$

A rather minor modification of the argument that gives (2.2) will prove Theorem 1.2. The basic argument is as follows.

LEMMA 2.1. – *Consider subsets $(C_i)_{i \leq p}$ of H, and assume that C_i is contained in the ball centered at the origin of radius R_i. Then*

$$G(\bigcup_{i \leq p} C_i) \leq \max_{i \leq p} \left(G(C_i) + KR_i\sqrt{\log i}\right) + K \max_{i \leq p} R_i \tag{2.3}$$

The following special cases will be used. First, the case where $R_i = R$ for all $i \leq p$. Then (separating the case $p = 1$), (2.3) becomes

$$G(\bigcup_{i\leq p} C_i) \leq \max_{i\leq p} \big(G(C_i) + KR\sqrt{\log i}\big) \tag{2.4}$$

Second, when $R_i = R2^{-i}$, (2.3) becomes

$$G(\bigcup_{i\leq p} C_i) \leq \max_{i\leq p} G(C_i) + KR. \tag{2.5}$$

Proof. – We set $Z_i = \sup_{t\in C_i} X_t$. The key point is the deviation inequality of [I-S-T]

$$P\big(Z_i \geq G(C_i) + t\big) \leq \exp\big(-\frac{t^2}{2R_i^2}\big)$$

so that, setting $A = \max_{i\leq p} \big(G(C_i) + 2R_i\sqrt{\log i}\big)$ we have, for $t \geq 0$,

$$\begin{aligned} P(Z_i \geq A + t) &\leq \exp\big(-\frac{1}{2R_i^2}(t + 2R_i\sqrt{\log i})^2\big) \\ &\leq \frac{1}{i^2}\exp\big(-\frac{t^2}{2R}\big) \end{aligned}$$

where $R = \max_{i\leq p} R_i$. Thus, if $Z = \sup_{i\leq p} Z_i$, we have

$$P(Z \geq A + t) \leq \sum_{i\geq 1} P(Z_i \geq A + t) \leq 2\exp\big(-\frac{t^2}{2R^2}\big)$$

and by a routine argument,

$$EZ = G\big(\bigcup_{i\leq p} C_i\big) \leq A + KR$$

from which (2.3) follows. □

We start the proof of Theorem 1.2. There is no loss of generality to assume that A is finite. We find m large enough that

$$\forall B \in \mathcal{B}_m, \quad \mathrm{card}\mathcal{B} \cap \mathcal{A} = 1$$

and we set

$$R = \sup\{G(\mathcal{U}(x)); \|x\| \leq 1\}.$$

We show by decreasing induction over $n \leq m$ that

$$(2.6)\qquad \begin{cases} \forall B \in \mathcal{B}_n, \quad \forall x \in B, \\ G(\mathcal{U}(B - x)) \leq K\left(2^{-n}R + \sup\limits_{t\in B} \sum\limits_{k\geq n+1} 2^{-k}\sqrt{\log \dfrac{1}{\mu(B_k(t))}}\right) \end{cases}$$

This is certainly true for $n = m$ since the left-hand side is zero; and for $n = n_0$, this implies Theorem 2.1 since $G(\mathcal{U}(B)) \leq G(\mathcal{U}(B - x)) + G(\mathcal{U}(x))$, since $G(\mathcal{U}(x)) \leq R \sup\limits_{x\in B} \|x\|$ for $x \in B$, and since $B = A$ for $B \in \mathcal{B}_{n_0}$.

Assuming that (2.6) holds for n, we start the proof that it holds for $n - 1$. Consider B' in $\mathcal{B}_{n-1}$, and enumerate the elements of $\mathcal{B}_n$ it contains as $(B_i)_{i\leq p}$. We can pick this enumeration in a way that the sequence $(\mu(B_i))_{i\leq p}$ decreases. We then have $\mu(B_i) \leq 1/i$, so that

$$(2.7)\qquad \sqrt{\log \frac{1}{\mu(B_i)}} \geq \sqrt{\log i}$$

The key of the argument is that, for some universal constant K_1, and any $x \in B'$,

$$G(\mathcal{U}(B' - x)) \leq \max_{i\leq p}(G(\mathcal{U}(B_i - x)) + K_1 2^{-n}\sqrt{\log i})$$

This follows from (2.4) and the fact that

$$\mathcal{U}(B' - x) = \bigcup_{i\leq p} \mathcal{U}(B_i - x)$$

Consider now, for $i \leq p$, a point x_i in B_i. Thus

$$\begin{aligned} G(\mathcal{U}(B_i - x)) &= G(\mathcal{U}(B_i - x_i)) + G(\mathcal{U}(x_i - x)) \\ &\leq G(\mathcal{U}(B_i - x_i)) + 2^{-n-1}R \end{aligned}$$

We also observe that, by (2.7), and since $B_n(t) = B_i$ for $t \in B_i$, we have

$$\begin{aligned} &\sup_{t\in B} \sum_{k\geq n} 2^{-k}\sqrt{\log \frac{1}{\mu(B_k(t))}} \\ &\quad \geq \max_{i\leq p}\left(2^{-n}\sqrt{\log i} + \sup_{t\in B_i} \sum_{k\geq n+1} 2^{-k}\sqrt{\log \frac{1}{\mu(B_k(t))}}\right) \end{aligned}$$

Thus, (2.6) is proved by induction, provided $K \geq \max(2, K_1)$. The proof of Theorem 2.1 is complete. □

3. AN EXAMPLE

The purpose of this section is to provide an example showing that (1.3) does not hold in general.

We consider an integer p, and, for $k \leq p$, we consider an integer $n_k \geq 2$. (The values of these will be specified later.) For $1 \leq k \leq p$, we set

$$R_k = \{1, \cdots, n_1\} \times \cdots \times \{1, \cdots, n_k\}.$$

Thus, an element τ of R_p is a sequence $(\tau_1, \cdots, \tau_p)$ of integers, with $\tau_k \leq n_k$. We will denote by $\tau|k$ the truncated sequence $(\tau_1, \cdots, \tau_k) \in R_k$.

In a fixed Hilbert space, we consider an orthonormal family of vectors $e_\rho, \rho \in \bigcup_{k \leq p} R_k$. Consider a number $r \geq 2$. For $\tau \in R_p$, we set

$$x_\tau = \sum_{1 \leq k \leq p} r^{-k} e_{\tau|k}.$$

We set $A = \{x_\tau; \tau \in R_p\}$. It is a simple matter to check that

$$G(A) \leq K \sum_{1 \leq k \leq p} r^{-k} \sqrt{\log m_k} \tag{3.1}$$

where $m_k = n_1 \ldots n_k$. For $\tau \in R_p$, let us consider an isometry U_τ of the Hilbert space, with the following property. For $\rho \in R_k$, we have $U_\tau(e_\rho) = 0$ if either $k = p$ or if $\rho \neq \tau|k$. If $\rho = \tau|k$, and $k \leq p-1$, we have $U_\tau(e_\rho) = e_{\rho,\tau_{k+1}}$, where $e_{\rho,\tau_{k+1}}$ denotes of course the vector $e_\sigma, \sigma \in R_{k+1}$ being the sequence $\rho_1, \cdots, \rho_k, \tau_{k+1}$.

We consider the family $\mathcal{U}$ of the isometries $U_\tau, \tau \in R_p$. The set $\mathcal{U}(A)$ contains in particular the vectors $y_\tau = U_\tau(x_\tau)$, and

$$y_\tau = \sum_{1 \leq k \leq p-1} r^{-k} e_{\tau|k+1} = \sum_{2 \leq k \leq p} r^{-k+1} e_{\tau|k}$$

(the term $+1$ in the exponent is crucial).

It is a simple matter (related to (3.1)) to see that

$$G(\{y_\tau; \tau \in R_p\}) \geq \frac{1}{K} \sum_{2 \leq k \leq p} r^{-k+1} \sqrt{\log n_k}. \tag{3.2}$$

Next, given $\tau \in R_p$, we set

$$X_\tau = \{U_\sigma(x_\tau); \sigma \in R_p\}$$

and we aim to estimate $G(X_\tau)$ from above. We set, for $1 \leq q < p$,

$$Z_q = \{\sigma \in R_p; \sigma|q = \tau|q, \sigma|q+1 \neq \tau|q+1\}.$$

For $\sigma \in Z_q$, the vector $U_\sigma(x_\tau)$ depends only upon $\sigma|q+1$. Thus the set $\{U_\sigma(x_\tau); \sigma \in Z_q\}$ consists of $n_{q+1} - 1$ vectors, each of which within distance $3 \cdot r^{-q}$ of $U_\tau(x_\tau)$. It is then a simple matter to *see* (using (2.5))

$$G(X_\tau) \leq K \sup_{1 \leq q \leq p-1} r^{-q} \sqrt{\log n_{q+1}} \tag{3.3}$$

It remains to choose the parameters. We take n_k such that $\sqrt{\log n_k} \simeq r^k$, and we take $p = r$. Thus,

$$G(A) + \sup_\tau G(X_\tau) \leq Kr$$

while $G(\mathcal{U}(A)) \geq r^2/K$. As r is arbitrarily large, this completes the construction.

The reader might like to note that in this case the entropy integral of (1.5) is of order r.

4. A THEOREM OF WEBER

If U is an isometry of H, if U_n is given by (1.8) and if $\mathcal{U} = \{U_n; n \geq 1\}$, it follows from Theorem 1.3 and either Theorem 1.1 or Theorem 1.2 that

$$G(\mathcal{U}(A)) \leq K(G(A) + \sup\{\|a\|; a \in A\})$$

This had been proved earlier by Weber [W] in the special case where $H = L^2(\Omega, P)$ and T is induced by an ergodic transformation of Ω. (Weber's result partially motivated the present paper.)

Proof of Theorem 1.3. – There is no loss of generality to assume $\|x\| = 1$. By the spectral theorem (e.g. as in [K] p. 94) there is a probability measure μ on $[-\pi, \pi]$ such that

$$\langle T^n(x), T^m(x) \rangle = \int_{-\pi}^{\pi} e^{i(m-n)\theta} d\mu(\theta) \tag{4.1}$$

for all n, m in $\mathbb{Z}$, where $\langle \cdot, \cdot \rangle$ denotes of course the scalar product in H. Thus, if we set

$$V_m(\theta) = \frac{1}{m}(1 + \cdots + e^{i(m-1)\theta}) = \frac{1}{m} \frac{e^{im\theta} - 1}{e^{i\theta} - 1}$$

it follows from (5.1) that whenever $n, m \geq 1$, we have

$$(4.2) \qquad \|U_n(x) - U_m(x)\|^2 = \int_{-\pi}^{\pi} |V_m(\theta) - V_n(\theta)|^2 d\mu(\theta)$$

For $\ell \geq 0$, consider the set

$$J_\ell = \{\theta; 2^{-\ell-1}\pi < |\theta| \leq 2^{-\ell}\pi\}$$

and set $a_\ell = \mu(J_\ell)$, so that $\sum_{\ell \geq 0} a_\ell \leq 1$.

Consider the sequence

$$b_\ell = \sum_{k \geq 0} 2^{-|k-\ell|} a_k$$

so that $\sum_{\ell \geq 0} b_\ell \leq 3$, and

$$(4.3) \qquad 1/2 \leq b_{\ell+1}/b_\ell \leq 2,$$

unless $a_k = 0$ for all k, so that μ is a point mass at zero and $Tx = x$. We now fix $1 \geq \epsilon > 0$ once and for all.

By (5.3), the sequence $b_m 2^{2m}$ increases to infinity unless $Tx = x$, an uninteresting case. Thus, for $k \geq 1$, there exists a smallest integer $m(k) \geq 1$ such that

$$(4.4) \qquad b_{m(k)} 2^{2m(k)} \geq 2^{2k} \epsilon^2$$

It is a simple matter to deduce from (4.3) that

$$(4.5) \qquad m(k) \leq m(k+1) \leq m(k) + 2$$

We now construct a subset F of $\mathbb{N}$ as follows. For $n \in \mathbb{N}$, $n \geq 1$, consider k such that $2^k \leq n < 2^{k+1}$. We define

$$f(n) = \sum_{\ell < k} b_{m(\ell)} + (2^{-k} n - 1) b_{m(k)}.$$

We put n in F whenever, for some $p \geq 1$, we have $f(n-1) < p\epsilon^2$, $f(n) \geq p\epsilon^2$. Observe in particular that when $b_{m(k)} > \epsilon^2$ the interval $[2^k, 2^{k+1}]$ contains about $b_{m(k)}/\epsilon^2$ points of F that are about evenly distributed. We will show two things:

$$(4.6) \qquad \text{card} F \leq \frac{K}{\epsilon^2}$$

(4.7) For each $n \geq 1$, there exists m in F such that $\|U_n(x) - U_m(x)\| \leq K\epsilon$.

Together these two facts prove Theorem 1.3.

First, we note that if $m(k) = 1$, we have $2^{2k}\epsilon^2 \leq 4b_1 \leq 12$. Thus, the number of values of k for which this holds is certainly $\leq K/\epsilon^2$. When $m(k) > 1$, the definition of $m(k)$ shows that

$$b_{m(k)-1}2^{2m(k)-2} < 2^{2k}\epsilon^2.$$

It then follows that, since $b_{m(k)} \leq 2b_{m(k)-1}$,

$$b_{m(k)}2^{2m(k)} < 2^{2k+3}\epsilon^2 \leq b_{m(k+2)}2^{2m(k+2)} \tag{4.8}$$

and thus $m(k+2) > m(k)$. Thus we have

$$\sum_{k\geq 1, m(k)>1} b_{m(k)} \leq 2\sum_{m\geq 1} b_m \leq 6$$

and this proves that $f(n) \leq 6$ and hence (4.6).

We turn to the proof of (4.7).

LEMMA 4.1. – *We have the following:*

$$|V_n(\theta)| \leq \frac{4}{n|\theta|} \tag{4.9}$$

$$|V_n(\theta) - V_m(\theta)| \leq K|\theta||n-m|. \tag{4.10}$$

Proof. – Since (4.9) is obvious we prove only (4.10). Consider the function

$$\varphi(u) = \frac{e^{i\theta u} - 1}{u}$$

Thus

$$\varphi'(u) = \frac{i\theta e^{i\theta u}}{u} - \frac{e^{i\theta u} - 1}{u^2} = -\frac{(1 - ui\theta)e^{i\theta u} - 1}{u^2}$$

Now,

$$|(1-z)e^z - 1| \leq K|z|^2 \text{ for } |z| \leq 1,$$

so that $|\varphi'(u)| \leq K\theta^2$ if $|\theta u| \leq 1$. On the other hand, if $|\theta u| \geq 1$, then

$$|\varphi'(u)| \leq \frac{|\theta|}{u} + \frac{2}{u^2} \leq 3\theta^2$$

so that in any case $|\varphi'(u)| \leq K\theta^2$. Thus

$$|\varphi(n) - \varphi(m)| \leq |n-m| \sup|\varphi'(u)| \leq K|n-m||\theta^2|,$$

and the result follows since $|e^{i\theta} - 1| \geq |\theta|/K$ for $|\theta| \leq \pi$. □

Consider $n \in \mathbb{N}$, $n \geq 1$, and k such that $2^k \leq n < 2^{k+1}$. Consider the largest integer p such that $p\epsilon^2 \leq f(n)$, (so that $f(n) < (p+1)\varepsilon^2$) and the smallest integer n' such that $f(n') \geq p\epsilon^2$. By definition, $n' \in F$, and we will show that $\|U_n(x) - U_{n'}(x)\| \leq K\epsilon$. Consider k' with $2^{k'} \leq n' < 2^{k'+1}$. Thus $k' \leq k$. Then

$$\text{(4.11)} \quad \epsilon^2 \geq f(n) - f(n') = (2^{-k}n - 1)b_{m(k)} + \sum_{k'<\ell<k} b_{m(\ell)} + (2 - 2^{-k'}n')b_{m(k')}.$$

We now observe that

$$\text{(4.12)} \qquad 2^{-k}(n - n')b_{m(k)} \leq K\epsilon^2$$

$$\text{(4.13)} \qquad \sum_{m(k')<\ell<m(k)} b_\ell \leq K\epsilon^2.$$

To prove (4.12), we first observe that, combining (4.3), (4.5), we have $1/K \leq b_{m(k+1)}/b_{m(k)} \leq K$. Thus (4.12) is obvious from (4.11) if either $k' = k$ or $k' = k-1$. But if $k' < k-1$, then (4.11) implies $b_{m(k-1)} < \epsilon^2$, so that $b_{m(k)} < K\epsilon^2$ and (4.12) follows.

To prove (4.13), we can assume $k' < k-1$. It is then a consequence of (4.3), (4.5) and $\sum_{k'<\ell<k} b_{m(\ell)} \leq \epsilon^2$.

We set

$$c_\ell = \int_{J_\ell} |V_n(\theta) - V_{n'}(\theta)|^2 d\mu(\theta)$$

so that we try to control $\sum_{\ell \geq 0} c_\ell$. If $\ell \leq m(k')$, we use (4.9) and the inequality $|u+v|^2 \leq 2|u|^2 + 2|v|^2$ to get

$$c_\ell \leq \frac{K}{n'^2} a_\ell 2^{2\ell}$$

so that

$$\text{(4.14)} \qquad \sum_{\ell \leq m(k')} c_\ell \leq \frac{K}{n'^2} \sum_{\ell \leq m(k')} b_\ell 2^{2\ell} \leq \frac{K}{n'^2} b_{m(k')} 2^{2m(k')} \leq K\epsilon^2$$

where we have used (4.3) and (4.8).

If $m(k') < \ell < m(k)$, we use the trivial bound $|V_n(\theta)| \leq 1$ to get

$$\sum_{m(k')<\ell<m(k)} c_\ell \leq \sum_{m(k')<\ell<m(k)} a_\ell \leq K\epsilon^2 \tag{4.15}$$

using $a_\ell \leq b_\ell$ and (4.13).

If $\ell \geq m(k)$, we use (4.10) to get

$$c_\ell \leq K a_\ell 2^{-2\ell} |n - n'|^2$$

so that

$$\begin{aligned}\sum_{\ell \geq m(k)} c_\ell &\leq K|n - n'|^2 \sum_{\ell \geq m(k)} a_\ell 2^{-2\ell} \\ &\leq K|n - n'|^2 \sum_{\ell \geq m(k)} b_\ell 2^{-2\ell} \\ &\leq K|n - n'|^2 b_{m(k)} 2^{-2m(k)}\end{aligned}$$

using (4.3). But, using (4.12)

$$(n - n')^2 b_{m(k)} 2^{-2m(k)} \leq K\epsilon^4 2^{2k-2m(k)}/b_{m(k)} \leq K\epsilon^2$$

using (4.8). Thus

$$\sum_{\ell \geq m(k)} c_\ell \leq K\epsilon^2.$$

Combining with (4.14), (4.15), we have shown that $\|U_n(x) - U_{n'}(x)\| \leq K\epsilon$. □

Remark. – Both M. Weber and an anonymous referee pointed out to me that Theorem 1.3 holds as well for all contractions of Hilbert space rather than just isometries. Indeed, in that case, by a dilation theorem of Sz. Nagy, (4.2) still holds (with inequality rather than equality).

5. DONSKER CLASSES

Donsker classes are characterized by numerous equivalent properties. We will recall only the technically useful (but uninspiring) characterization we need, and refer the reader to [G-Z] for more material.

Consider a probability space (Ω, P), and an i.i.d. sequence of Ω valued r.v. (X_i) of law P. Consider a sequence (g_i) of standard Gaussian r.v. A class $\mathcal{F}$ of functions on Ω (that we assume to be countable to avoid well understood measurability problems) is called a Donsker class if $\mathcal{F} \subset \mathcal{L}_2(\Omega)$ and if, for any $\delta > 0$, we can find a *finite* covering $(\mathcal{F}_k)$ of $\mathcal{F}$ such that

$$\forall k, \quad \limsup_{n\to\infty} E \sup_{f,h\in\mathcal{F}_k} \left| \frac{1}{\sqrt{n}} \sum_{i\le n} g_i(f(X_i) - h(X_i)) \right| \le \delta. \tag{5.1}$$

We recall that a class $\mathcal{C}$ of subsets of Ω is called a VC (=Vapnik Cervonenkis) class (of dimension $\le d$), if it does not shatter any subset F of cardinality d, that is, for any subset F of $\Omega, \mathrm{card} F = d$, at least one subset of F is not of the type $C \cap F, C \in \mathcal{C}$.

The crucial property of VC classes is as follows.

LEMMA 5.1. – *Consider a VC class $\mathcal{D}$ of subsets of Ω of dimension at most d, and points $(x_i)_{i\le n}$ of Ω. Consider numbers $(a_i)_{i\le n}$. We define*

$$a^2 = \sum_{i\le n} a_i^2; b^2 = \sup_{D,D'\in\mathcal{D}} \sum_{i\le n} a_i^2 1_{D\triangle D'}(x_i).$$

Then, given $D_0 \in \mathcal{D}$, we have

$$E \sup_{D\in\mathcal{D}} |\sum_{i\le n} a_i g_i(1_D(x_i) - 1_{D_0}(x_i))| \le K\sqrt{d} b \log \frac{Ka}{b} \le K\sqrt{d} a. \tag{5.2}$$

Proof. – We have to evaluate $E \sup_{D\in\mathcal{D}} |X_D|$, where the Gaussian process X_D indexed by $\mathcal{D}$ is given by

$$X_D = \sum_{i\le n} a_i g_i(1_D(x_i) - 1_{D_0}(x_i)).$$

The canonical distance δ associated to the process is given by

$$\begin{aligned} \delta^2(D,F) &= \sum_{i\le n} a_i^2(1_D(x_i) - 1_F(x_i))^2 \\ &= a^2\nu(D\triangle F) \end{aligned}$$

where the probability ν is given by $\nu(\{x_i\}) = a_i^2/a^2$. The diameter of $(\mathcal{D}, \delta)$ is at most $2b$; moreover Dudley [D] proved that the maximal cardinality of a subset Z of $\mathcal{D}$ such that

$$D, F \in Z \Rightarrow \nu(D\triangle F) \ge \epsilon$$

is at most $(K\epsilon)^{-2d}$ (this is a weak form of the result, that is sufficient here). Thus

$$N(\mathcal{D},\delta,\epsilon) \leq (K\frac{a^2}{\epsilon^2})^{2d}$$

and bounding of the left-hand side of (5.2) by the entropy integral $\int_0^{2b}\sqrt{\log N(\mathcal{D},\delta,\epsilon)d\epsilon}$ yields the result.

LEMMA 5.2. – *Consider a VC class $\mathcal{C}$ of subsets of Ω, of dimension $\leq d$. Consider a function f in $L^2(\Omega)$ and consider $\gamma > 0$. Then there is a finite partition $(\mathcal{C}_k)$ of $\mathcal{C}$ with the following property:*

$$\text{(5.3)} \qquad \limsup_{n\to\infty} E \sup_{C,D\in\mathcal{C}_k}\left|\frac{1}{\sqrt{n}}\sum_{i\leq n} g_i f(X_i)(1_C(X_i)-1_D(X_i))\right| \leq \gamma.$$

Proof. – We first observe that, by Lemma 5.1, if we denote by E_g conditional expectation given X_i, we have

$$E_g \sup_{C,D\in\mathcal{C}_k}\left|\frac{1}{\sqrt{n}}\sum_{i\leq n} g_i f(X_i)(1_C(X_i)-1_D(X_i))\right| \leq K\sqrt{d}\left(\sum_{i\leq n}\frac{f^2(X_i)}{n}\right)^{1/2}.$$

Thus, by the law of large numbers, (5.3) holds for any partition $(\mathcal{C}_k)$ whenever $\|f\|_2 < \gamma/K\sqrt{d}$. Thus, by truncation we can reduce to the case where f is bounded.

We now show that it suffices to decompose $\mathcal{C}$ is such a way that

$$\xi = \sup_k \sup_{C,D\in\mathcal{C}_k} P(C\triangle D)$$

is sufficiently small. (That such a finite partition exists is a consequence of a result of Dudley used before.)

Let us consider the r.v.

$$b_{n,k}(X) = \sup_{C,D\in\mathcal{C}_k}\sum_{i\leq n}\frac{f^2(X_i)}{n}1_{C\triangle D}(X_i)$$

Thus

$$b_{n,k}(X) - \xi \leq \sup_{C,D\in\mathcal{C}_k}\sum_{i\leq n}\frac{f^2(X_i)}{n}(1_{C\triangle D}(X_i) - P(C\triangle D))$$

It follows from a standard argument (brought to light in [G-Z]) that

$$Eb_{n,k}(X) - \xi \le KE \sup_{C,D\in\mathcal{C}_k} \left|\sum_{i\le n} g_i \frac{f^2(X_i)}{n} 1_{C\triangle D}(X_i)\right|.$$

The collection of all sets $C\triangle D (C, D \in \mathcal{C})$ is still a VC class of dimension $\le Kd$. Thus, by Lemma 5.1 and the bound $\sum_{i\le n} f(X_i)^4/n^2 \le \|f\|_\infty^4/n$, we have

$$Eb_{n,k}(X) \le \xi + K\sqrt{d}\frac{\|f\|_\infty^2}{\sqrt{n}}$$

Setting $a_n(X) = \sum_{i\le n} \frac{f^2(X_i)}{n}$, it follows from Lemma 5.1 again that

$$\begin{aligned} E_g \sup_{C,D\in\mathcal{C}_k} &\left|\frac{1}{\sqrt{n}}\sum_{i\le n} g_i f(X_i)(1_C(X_i) - 1_D(X_i))\right| \\ &\le K\sqrt{d} b_{n,k}(X)\sqrt{\log \frac{a_n(X)}{b_{n,k}(X)}} \\ &\le K\sqrt{d b_{n,k}(X) a_n(X)} \end{aligned}$$

The purpose of this brutal last bound is that by Cauchy-Schwarz,

$$E\sqrt{b_{n,k}(X)a_n(X)} \le \sqrt{Eb_{n,k}(X)Ea_n(X)} \tag{5.4}$$

has a lim sup, as $n \to \infty$, at most $K\|f\|_2\sqrt{\xi}$, and this concludes the proof. □

We now prove Theorem 1.4. Consider $\delta > 0$. We have to produce an appropriate finite partition of the class of functions $f1_C, f \in \mathcal{F}, C \in \mathcal{C}$ that witnesses (5.1). Since $\mathcal{F}$ is a Donsker class, there is a finite partition $(\mathcal{F}_k)_{k\le q}$ of $\mathcal{F}$ such that

$$\forall k, \quad \limsup_{n\to\infty} E \sup_{f,h\in\mathcal{F}_k} \left|\frac{1}{\sqrt{n}}\sum_{i\le n} g_i(f(X_i) - h(X_i))\right| \le \delta. \tag{5.5}$$

For each k, we fix $f_k \in \mathcal{F}_k$. Since $\mathcal{F} \subset L^2$, by Lemma 5.2, there is a finite partition $(\mathcal{C}_\ell)_{\ell\le m}$ of $\mathcal{C}$ with the property

$$\begin{cases} \forall k \le q, \quad \forall \ell \le m, \\ \limsup_{n\to\infty} E \sup_{C,D\in\mathcal{C}_\ell} \left|\frac{1}{\sqrt{n}}\sum_{i\le n} g_i f_k(X_i)(1_C(X_i) - 1_D(X_i))\right| \le \delta. \end{cases} \tag{5.6}$$

We fix $C_\ell \in \mathcal{C}_\ell$. We are done if we can prove that, for each k, ℓ

(5.7)
$$\limsup_{n\to\infty} E \sup_{C\in\mathcal{C}_\ell, f\in\mathcal{F}_k} \left| \frac{1}{\sqrt{n}} \sum_{i\le n} g_i(f(X_i)1_C(X_i) - f_k(X_i)1_{C_\ell}(X_i)) \right| \le K\delta.$$

We write

$$\begin{aligned} &f(X_i)1_C(X_i) - f_k(X_i)1_{C_\ell}(X_i) \\ &\quad = (f(X_i) - f_k(X_i))1_C(X_i) + f_k(X_i)(1_C(X_i) - 1_{C_\ell}(X_i)) \end{aligned}$$

We set

$$W_n(X) = E_g \sup_{C\in\mathcal{C}_\ell, f\in\mathcal{F}_k} \left| \frac{1}{\sqrt{n}} \sum_{i\le n} g_i(f(X_i) - f_k(X_i))1_C(X_i) \right|.$$

It follows from (5.6) that we are reduced to prove that $\limsup_{n\to\infty} EW_n(X) \le K\sqrt{d}\delta$. The key is to apply Theorem 1.2 to estimate the expectation E_g at $X_1, \cdots, X_n$ given. The basic Hilbert space is ℓ_n^2; the set $A = A_{k,n}(X)$ is

$$A_{k,n}(X) = \left\{ \pm \frac{1}{\sqrt{n}} (f(X_i) - f_k(X_i))_{i\le n}; f \in \mathcal{F}_k \right\}$$

To each $C \in \mathcal{C}_\ell$, we associate the contraction U_C of ℓ_n^2 given by $U_C((a_i)) = (a_i 1_C(X_i))_{i\le n}$. We set $\mathcal{U} = \{U_C; C \in \mathcal{C}_\ell\}$. It follows from Lemma 5.1 that, for $a \in \ell_n^2$

$$G(\mathcal{U}(a)) \le K\sqrt{d}\|a\|.$$

We also observe that

$$\begin{aligned} \sup_{a\in A} \|a\| &= \sup_{f\in\mathcal{F}_k} \left(\frac{1}{n} \sum_{i\le n} (f(X_i) - f_k(X_i))^2 \right)^{1/2} \\ &\le K E_g \sup_{f\in\mathcal{F}_k} \left| \frac{1}{n} \sum_{i\le n} g_i(f(X_i) - f_k(X_i)) \right|. \end{aligned}$$

Thus, by Theorem 1.2, we have

$$W_n(X) \le K\sqrt{d}G(A_{k,n}(X))$$

That $\limsup_{n\to\infty} EW_n(X) \le K\sqrt{d}\delta$ is then a consequence of (5.5). □

6. PROOF OF THEOREM 1.5

The method is to construct a probability measure ν on $\mathcal{U} \times A$, the image of which under the natural map from $\mathcal{U} \times A$ into $\mathcal{U}(A)$ will witness that $G(\mathcal{U}(A))$ is bounded by the right-hand side of (1.10) through the majorizing measure bound. It makes the proof somewhat clearer to construct, as well as ν, an increasing family $(\mathcal{C}_n)$ of finite partitions of $\mathcal{U} \times A$. We recall the increasing sequence $(\mathcal{B}_n)_{n \geq n_0}$ of partitions of A used in the proof of Theorem 1.2 (and the measure μ on A). We denote by m_0 the largest integer $\leq n_0$ such that

$$\forall a \in A, 2^{-m_0} \geq \operatorname{diam}(\mathcal{U}, d_a).$$

and we set $\mathcal{C}_{m_0} = \{\mathcal{U} \times A\}$. For $m_0 \leq n \leq n_0$, we set $\mathcal{B}_n = \mathcal{B}_{n_0} = \{A\}$. For each $n \geq m_0$, each set of $\mathcal{C}_n$ will be of the form $W \times B$, where $B \in \mathcal{B}_n$, and we now describe the process by which the sequence $\mathcal{C}_n$ is inductively constructed. Assuming that $\mathcal{C}_{n-1}$ has been constructed, an element of $\mathcal{C}_{n-1}$ is of the type $W' \times B'$ $(B' \in \mathcal{B}_{n-1})$, and we have to partition this element into elements of $\mathcal{C}_n$. First, we partition B' into elements of $\mathcal{B}_n$, and then we have to show how to partition a set of the type $W' \times B$ $(B \in \mathcal{B}_n)$. For this purpose we choose one arbitrary element a of B, and a maximum subset Z of W' that is $3 \cdot 2^{-n}$ separated for d_a (*i.e.* any two points of Z are within distance $\geq 3 \cdot 2^{-n}$). The balls of radius $3 \cdot 2^{-n}$ (for d_a) centered on Z cover W'. Thus we can find a partition of W' into sets of diameter $\leq 6 \cdot 2^{-n}$ for d_a that refine this covering. We fix such a partition, and this completes the construction. To discuss the properties of the construction, we denote by $W \times B$ an element of $\mathcal{C}_n$ contained in $W' \times B$.

We observe that whenever $b \in B$, then Z is 2^{-n} separated for d_b. This is a consequence of the fact that B has diameter $\leq 2^{-n}$, and that

$$\begin{aligned}(6.1)\quad |d_a(U,V) - d_b(U,V)| &= |\|U(a) - V(a)\| - \|U(b) - V(b)\|| \\ &\leq \|U(a) - U(b)\| + \|V(a) - V(b)\| \\ &\leq 2\|a - b\|\end{aligned}$$

since U, V are contractions.

We also observe that, by the same argument, for each b in B, W is of d_b-diameter at most $8 \cdot 2^{-n}$. Thus we see by induction that for a set $W' \times B'$ of $\mathcal{C}_{n-1}$, and b in B', the set W' is of d_b-diameter at most $8 \cdot 2^{-n+1} = 2^{-n+4}$. Thus W' is contained in the ball of center t and d_b-radius 2^{-n+4} whenever $t \in W'$. In particular the points of Z are contained in this ball; since they are 2^{-n} separated we have

$$(6.2)\qquad \forall b \in B,\ \forall t \in W',\quad \operatorname{card} Z \leq N(B(t, 2^{-n+4}), 2^{-n-1}, d_b).$$

We observe that by our construction, $\text{card}Z$ depends only on $W' \times B$, so only on $W \times B$. We set

$$v_n(W \times B) = \frac{1}{\text{card}Z}.$$

so that

$$\sum_{W \times B \subset W' \times B} v_n(W \times B) \leq 1. \tag{6.3}$$

For sets of $\mathcal{C}_n$, we now define the "weights" w_n as follows, by induction over $n \geq m_0$. We set $w_{m_0}(\mathcal{U} \times A) = 1$, and, if $W \times B \in \mathcal{C}_n, W' \times B' \in \mathcal{C}_{n-1}, W \times B \subset W' \times B'$, we set

$$w_n(W \times B) = v_n(W \times B)\mu(B)w_{n-1}(W' \times B') \tag{6.4}$$

Since $\sum \mu(B) \leq \mu(B') \leq 1$, where the sum is over all $B \in \mathcal{A}_n, B \subset B' \in \mathcal{A}_{n-1}$, and by (2.4), we see by induction that for each n,

$$\sum \{w_n(W \times B); W \times B \in \mathcal{C}_n\} \leq 1.$$

It follows that there is a probability measure ν on $U \times A$ such that

$$\forall n \geq m_0, \quad \forall W \times B \in \mathcal{C}_n, \quad \nu(W \times B) \geq 2^{-n+m_0-1} w_n(W \times B).$$

Consider the image η of ν under the map $(U, a) \to U(a)$, and consider $s \in U(A)$. By the Preston-Fernique majorizing measure bound (*see* [L-T] ch. 12) it suffices to prove

$$\int_0^{\infty} \sqrt{\log \frac{1}{\eta(B(s,\epsilon))}} d\epsilon \leq K[G(A) + \sup_{x \in A} S(\mathcal{U}, d_x)]$$

We can write $s = t(a), t \in \mathcal{U}, a \in A$, and it suffices to prove that, if $W_n \times B_n$ is the element of $\mathcal{C}_n$ that contains (U, a), we have

$$I = \sum_{n \geq m_0} 2^{-n} \sqrt{\log \frac{1}{\nu(W_n \times B_n)}} \leq K[G(A) + \sup_{x \in A} S(\mathcal{U}, d_x)] \tag{6.5}$$

Obviously, we have

$$I \leq J = \sum_{n \geq m_0} 2^{-n} \sqrt{\log \frac{2^{n-m_0+1}}{w_n(W_n \times B_n)}}$$

since $\nu(W_n \times B_n) \geq 2^{-n+m_0-1} w_n(W_n \times B_n)$. Using the formulae $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$ for $u, v \geq 0$, and (6.4), we see that

$$J \leq \sum_{n \geq m_0} 2^{-n} \left(\sqrt{\log 2} + \sqrt{\log \frac{1}{\mu(B_n)}} + \sqrt{\log \frac{1}{v_n(W_n \times B_n)}} + \sqrt{\log \frac{2^{n-m_0}}{w_{n-1}(W_{n-1} \times B_{n-1})}} \right) \tag{6.6}$$

Since $B_n = B_n(a)$, it follows from (2.1) that

$$\sum_{n \geq m_0} 2^{-n} \left(\sqrt{\log 2} + \sqrt{\log \frac{1}{\mu(B_n)}} \right) \leq K(2^{-m_0} + G(A)).$$

Also,

$$\sum_{n \geq m_0} 2^{-n} \sqrt{\log \frac{2^{n-m_0}}{w_{n-1}(W_{n-1} \times B_{n-1})}} \leq K 2^{-m_0} + \frac{J}{2}$$

so that (6.6) implies

$$J \leq K \left(2^{-m_0} + G(A) + \sum_{n \geq m_0} 2^{-n} \sqrt{\log \frac{1}{v_n(W_n \times B_n)}} \right)$$

To control this last term, we recall that since $a \in B_n$, by (2.3) and construction of W, we have

$$\frac{1}{v_n(W_n \times B_n)} \leq N(B(t, 2^{-n+4}), d_a, 2^{-n-1}).$$

Also, since

$$2^{-n-2} \leq \epsilon < 2^{-n-1} \Rightarrow N(B(t, 2^{-n+4}), d_a, 2^{-n-1}) \leq N(B(t, 2^7 \epsilon), d_a, \epsilon)$$

we have

$$\sum_{n \geq m_0} 2^{-n} \sqrt{\log \frac{1}{v_n(W_n \times B_n)}} \leq K S(\mathcal{U}, d_a).$$

Since it is easily seen that $2^{-m_0} \leq \sup_{a \in A} S(\mathcal{U}, d_a)$, the proof is complete. □

REFERENCES

[D] R. Dudley, A course on empirical processes, École d'été de probabilités de Saint-Flour, 1982, *Lecture Notes in Mathematics*, Vol. 1097, 1984, pp. 2-142.

[F] X. Fernique, Régularité de fonctions aléatoires gaussiennes à valeurs vectorielles, *Ann. Probab.* , Vol. **18**, 1990, pp. 1739-1745.

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

[I-S-T] I. A. Ibragimov, V. N. Sudakov and B. S. Tsirel'son, Norms of Gaussian sample functions, Proceedings of the third Japan-USSR Symposium on Probability Theory, *Lecture Notes in Math.*, 550, 1976, Springer-Verlag, pp. 20-41.

[K] U. Krengel, *Ergodic theorems* , W. de Gruyter, 1989.

[L-T] M. Ledoux and M. Talagrand, *Probability in a Banach space*, Springer Verlag, 1990.

[T1] M. Talagrand, Regularity of Gaussian processes , *Acta. Math.*, Vol. 159, 1987, pp. 49-99.

[T2] M. Talagrand, A simple proof of the majorizing measure theorem, *Geometric and Functional analysis*, Vol. **2**, 1992, pp. 118-127.

[T3] M. Talagrand, Construction of majorizing measures, Bernoulli processes and cotype, *Geometric and Functional analysis*, Vol. **4**, 1994, pp. 660-717.

[W] M. Weber, Coupling of the GB set property, *J. Theor. Probab.*, Vol. **9**, 1996, pp. 105-112.

(Manuscript accepted April 26, 1995.)