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

GILLES PISIER

Remarques sur les classes de Vapnik-Cervonenkis

Annales de l'I. H. P., section B, tome 20, n° 4 (1984), p. 287-298

<http://www.numdam.org/item?id=AIHPB_1984__20_4_287_0>

Ann. Inst. Henri Poincaré,
Vol. 20, n° 4, 1984, p. 287-298. *Probabilités et Statistiques*

Remarques sur les classes de Vapnik-Červonenkis

par

Gilles PISIER

Équipe d'Analyse, Équipe de Recherche associée au C. N. R. S. n° 294,
Université Paris VI, 4, Place Jussieu, 75230 Paris Cedex 05,
Tour 46/0, 4e Étage

RÉSUMÉ. — Soit $M(\Omega, \mathscr{A})$ l'espace de Banach de toutes les mesures bornées, sur un espace mesurable $(\Omega, \mathscr{A})$. Soit $\mathscr{C}$ une sous-classe de $\mathscr{A}$. Nous considérons l'opérateur $J : M(\Omega, \mathscr{A}) \to l^\infty(\mathscr{C})$ défini par $J(\mu)$: $(\mu(C))_{C\in\mathscr{C}}$. Nous montrons que J est de type 2, si et seulement si, il est de type p pour un certain $p > 1$, et que cela arrive, si et seulement si, $\mathscr{C}$ est une classe de Vapnik-Červonenkis. Les démonstrations reposent fortement sur des idées de Dudley. Une application aux espaces de Banach est donnée, qui généralise un résultat précédent de Milman.

SUMMARY. — Let $M(\Omega, \mathscr{A})$ be the Banach space of all bounded signed measures on a measure space $(\Omega, \mathscr{A})$. Let $\mathscr{C}$ be a subclass of $\mathscr{A}$. We consider the operator $J : M(\Omega, \mathscr{A}) \to l^\infty(\mathscr{C})$ defined by $J(\mu) = (\mu(C))_{C\in\mathscr{C}}$. We show that J is of type 2 iff it is of type p for some $p > 1$, and that this happens iff $\mathscr{C}$ is a Vapnik-Červonenkis class. The proofs rely heavily on some ideas of Dudley. We also give an application to Banach spaces which improves a previous result due to Milman.

§ 1. INTRODUCTION

Dans leur étude de la loi des grands nombres pour les distributions empiriques, Vapnik et Červonenkis [*18*] ont considéré les classes $\mathscr{C}$ formées de parties d'un ensemble Ω pour lesquelles il existe un entier k tel que, pour toute partie à k éléments $A \subset \Omega$, l'ensemble $\{ A \cap C \mid C \in \mathscr{C} \}$ ne contienne pas toutes les parties de A.

Dudley [*1*] a étudié ces classes de parties sous le nom de classes de Vapnik-Červonenkis (en abrégé classes V. C.) ; comme lui, nous noterons $v(\mathscr{C})$ le plus petit entier $k \geqq 1$ possédant la propriété précédente. Dudley et ses collaborateurs ont montré, parmi bien d'autres choses, que l'on a pour ces classes un théorème limite central ([*1*]), une loi du logarithme itéré ([*10*]) et un principe d'invariance ([*3*]) pour les distributions empiriques.

Nous renvoyons à [*4*] pour plus de détails.

Le but de cette note est de remarquer que le point de départ des travaux de Dudley (précisément le théorème 7.1 de [*1*]) s'interprète « naturellement » dans la théorie des opérateurs de type 2 (cf. [*8*]). Pour un autre exemple où une telle interprétation s'est révélée utile et « éclairante », voir [*19*].

Nous donnons aussi une application aux espaces de Banach qui améliore un résultat de Milman [*13*].

Soit $(\Omega, \mathscr{A})$ un espace mesuré et soit $\mathscr{C} \subset \mathscr{A}$ une classe d'ensembles. On note $M(\Omega, \mathscr{A})$ l'espace de Banach des mesures μ bornées sur Ω munies de la norme $\| \mu \| = | \mu | (\Omega)$. D'autre part, on note $l^\infty(\mathscr{C})$ l'ensemble des familles bornées de scalaires (réels ou complexes) $x = (x_i)_{i \in \mathscr{C}}$ muni de la norme.

$$\| x \|_\infty = \operatorname*{Sup}_{i \in \mathscr{C}} | x_i | .$$

On peut alors considérer l'opérateur $J : M(\Omega, \mathscr{A}) \to l^\infty(\mathscr{C})$ défini par

$$J(\mu) = (\mu(C))_{C \in \mathscr{C}} ,$$

qui est évidemment de norme $\| J \| \leqq 1$.

Soit (ε_n) une suite de v. a. indépendantes équidistribuées, à valeurs ± 1, et telles que $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2$.

On dit qu'un opérateur $u : X \to Y$ (entre espaces de Banach) est de type p s'il existe une constante C telle que, pour toute suite finie (x_i) d'éléments de X on a :

$$\mathbb{E} \| \Sigma \varepsilon_i u(x_i) \| \leqq C(\Sigma \| x_i \|^p)^{1/p} .$$

On notera $T_p(u)$ la plus petite constante C vérifiant cette propriété. (Signalons que seul le cas $1 < p \leqq 2$ est intéressant).

Nous pouvons alors énoncer ([1])

THÉORÈME 1. — *Avec les notations précédentes, les propriétés suivantes sont équivalentes :*

i) $\mathscr{C}$ *est une classe V. C.*

ii) J *est de type 2.*

iii) J *est de type p pour un* $p > 1$.

Supposons que $\mathscr{C}$ *soit dénombrable. Alors ces propriétés sont équivalentes à iv), il existe une suite* $a_n > 0$ *tendant vers* 0 *avec les propriétés suivantes : pour toute probabilité* $\mathbb{P}$ *sur* $(\Omega, \mathscr{A})$, *la suite des mesures aléatoires*

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\omega_i} - \mathbb{P}$$

est telle que $\frac{1}{a_n} \operatorname*{Sup}_{C \in \mathscr{C}} |\nu_n(C)|$ *est « stochastiquement borné » sur* $(\Omega^{\mathbb{N}}, \mathbb{P}^{\mathbb{N}})$, *c'est-à-dire que* : $\forall \varepsilon > 0 \ \exists R \forall n$

$$\mathbb{P}^{\mathbb{N}}(\{ (\omega_i)_i | \operatorname*{Sup}_{C \in \mathscr{C}} |\nu_n(C)| > a_n R \}) < \varepsilon .$$

v) Avec les mêmes notations qu'en (iv), il existe une suite $a_n \downarrow 0$ *telle que, pour toute probabilité* $\mathbb{P}$

$$\sup_n \frac{1}{a_n} \operatorname*{Sup}_{C \in \mathscr{C}} |\nu_n(C)| < \infty \quad p.\,s.$$

Remarques. — *a*) L'équivalence entre (*ii*) et (*iii*) est plutôt surprenante. On peut se demander si, dans la situation précédente, l'opérateur J se factorise par un espace de type 2 (c'est-à-dire par un espace dont l'opérateur identité soit de type 2). En effet, James [9] a précisément montré que c'est le cas si l'on considère $\Omega = \mathbb{N}$ et $\mathscr{C}$ formé des intervalles d'entiers ; on en déduit (par exemple par ultraproduit) le même résultat pour $\Omega = [0, 1]$ et $\mathscr{C}$ l'ensemble des intervalles. Cela correspond donc exactement à la situation du théorème classique de Glivenko-Cantelli.

b) Au sujet de l'implication (*iv*) $\Rightarrow$ (*i*), signalons qu'un résultat de même nature est démontré dans [5] qui contient en particulier le cas $a_n = \frac{1}{\sqrt{n}}$. L'équivalence (*i*) $\Leftrightarrow$ (*v*) montre que les classes V. C. sont précisément celles pour lesquelles on a une loi des grands nombres « uniforme » pour

([1]) Je tiens à remercier Joel Zinn d'avoir attiré mon attention sur les questions considérées ici.

les distributions empiriques (i. e. $\operatorname{Sup}_{C \in \mathscr{C}} | v_n(C) |$ tend vers 0 p. s. d'une manière uniforme sur $\mathbb{P}$). Notons pour finir que l'équivalence entre (*iv*) et (*v*) résulte aussi des résultats généraux de [*11*], ou bien aussi de [*17*].

c) Par des méthodes de [*8*], on déduit aisément du fait que J est de type 2 que l'on a, dans la situation du théorème 1 (*iii*)

$$\operatorname{Sup}_n \sqrt{n}\, \mathbb{E} \operatorname{Sup}_{C \in \mathscr{C}} | v_n(C) | < \infty$$

et

$$\operatorname{Sup}_n \sqrt{\frac{n}{2 \log \log n}} \operatorname{Sup}_{C \in \mathscr{C}} | v_n(C) | < \infty \quad \text{p. s.}$$

C'est-à-dire que l'on a la « version bornée » du théorème limite central et de la loi du logarithme itéré. Aux prix de quelques complications, on pourrait aussi dériver la version « compacte » de ces théorèmes (obtenue dans [*1*] et [*10*]) d'une propriété de l'opérateur J, nous ne le ferons pas.

d) Bien entendu, on peut remplacer l'hypothèse « $\mathscr{C}$ dénombrable » par une hypothèse convenable de séparabilité du processus $(v_n(C))_{C \in \mathscr{C}}$.

§ 2. DÉMONSTRATION DU THÉORÈME 1

Soit $\mathbb{P}$ une probabilité sur $(\Omega, \mathscr{A})$. Suivant Dudley, on associe à $\mathbb{P}$ une distance (ou plutôt un écart $d_{\mathbb{P}}$ sur $\mathscr{A}$, de la manière suivante. On pose

$$\forall t, s \in \mathscr{A} \qquad d_{\mathbb{P}}(t, s) = \mathbb{P}(t \Delta s)^{1/2} = \| 1_t - 1_s \|_{L^2(d\mathbb{P})}$$

Soit T un ensemble muni d'un écart d. Pour tout $\varepsilon > 0$, on notera $N(T, d ; \varepsilon)$ le plus petit nombre de d-boules ouvertes de rayon ε qui forment un recouvrement de T.

L'entropie métrique (i. e. $\operatorname{Log} N(T, d ; \varepsilon)$) a été introduite dans l'étude des processus gaussiens dans [*2*]; plus récemment, Dudley s'en est servi dans son travail [*1*] visant à généraliser le théorème de Kolmogorov-Smirnov (i. e. le théorème limite central pour les distributions empiriques) à des classes d'ensembles plus générales que les intervalles. Nous reprenons ci-dessous l'idée essentielle du théorème 7.1 de Dudley [*1*]. Le théorème suivant n'en est qu'une reformulation plus précise.

THÉORÈME 2. — *Soit* P *une probabilité sur* $(\Omega, \mathscr{A})$. *Posons* $N_P(\varepsilon) = N(\mathscr{C}, d_P ; \varepsilon)$. *On a alors*

$$v(\mathscr{C}) \geq \alpha \operatorname{Sup}_{0 < \varepsilon \leq 1} \frac{\operatorname{Log} N_P(\varepsilon)}{\max \left\{ 1 ; \operatorname{Log} \frac{1}{\varepsilon} \right\}}, \tag{1}$$

où $\alpha > 0$ *est une constante* numérique.

La démonstration du théorème 2 utilise plusieurs lemmes. Le premier est dû indépendamment à Sauer [15], Vapnik-Červonenkis [17] et Shelah. Il joue un rôle fondamental dans l'étude des classes V. C.

LEMME 3. — *Pour toute partie finie* $A \subset \Omega$ *ayant exactement n éléments* $(n \geqq 1)$, *on a* :

$$|\mathscr{C} \cap A| \leqq \sum_{j<v} \binom{n}{j} = \varphi(v, n)$$

où l'on a posé simplement $v = v(\mathscr{C})$.

Notons tout de suite que pour $n \geqq 2$, on a

$$\varphi(v, n) \leqq \sum_{j<v} \frac{n^j}{j!} \leqq \sum_{j<v} \left(\frac{n}{v}\right)^j \frac{v^j}{j!}$$

d'où

(2) $$\varphi(v, n) \leqq \left(\frac{n}{v}\right)^v e^v \qquad \text{pour} \qquad n \geqq v\,.$$

Nous aurons aussi besoin du lemme suivant qui est le point crucial de la démonstration du théorème 7.1 de [1].

LEMME 4. — *Soit* $\varepsilon > 0$. *Soit* $(A_i)_{i \leqq N}$ *une famille de parties de* Ω *telles que* $\mathbb{P}(A_i \Delta A_j) \geqq \varepsilon^2$ *pour tout* $i \neq j$. *Soit* K *un entier tel que*

(3) $$N^2(1 - \varepsilon^2)^K < 1\,.$$

Alors il existe des points $\omega_1, \ldots, \omega_K$ *dans* Ω *tels que*

$$\forall i \neq j \qquad (A_i \Delta A_j) \cap \{\omega_1, \ldots, \omega_K\} \neq \emptyset\,.$$

Démonstration. — (On suit fidèlement Dudley [1], p. 922).

Soient $T_1, \ldots, T_K$ une suite de v. a. indépendantes équidistribuées à valeurs dans Ω et de loi $\mathbb{P}$.

Alors, pour $i \neq j$ et $m \leqq K$ fixés, la probabilité que T_m ne soit pas dans $A_i \Delta A_j$ vaut $1 - P(A_i \Delta A_j) \leqq 1 - \varepsilon^2$. Donc

$$\text{Proba}\,\{(A_i \Delta A_j) \cap \{T_1, \ldots, T_K\} = \emptyset\} \leqq (1 - \varepsilon^2)^K\,.$$

Par conséquent la probabilité qu'il existe $i, j \leqq N$ pour lesquels ce dernier événement se produit est majorée par $N^2(1 - \varepsilon^2)^K$. La conclusion du lemme est donc claire.

Démonstration du théorème 2. — Soit $(A_i)_{i \leqq N}$ une famille maximale d'éléments de $\mathscr{C}$ tels que $d_P(A_i, A_j) \geqq \varepsilon$ pour tout $i \neq j$. On a nécessairement $N \geqq N_P(\varepsilon)$ par la maximalité de N. Soit alors K comme au lemme 4. D'après les lemmes 3 et 4 on a nécessairement (d'après (2))

$$(4) \qquad N \leqq \varphi(v, K) \leqq \left(\frac{eK}{v}\right)^v \quad \text{si} \quad K \geqq v.$$

Pour que l'on ait (3), il suffit que $2 \operatorname{Log} N + K \log(1-\varepsilon^2) < 0$, on peut donc prendre $K = \left[\frac{2}{\varepsilon^2} \operatorname{Log} N\right] + 1$. Supposons que $\frac{2}{\varepsilon^2} \operatorname{Log} N \geqq v \geqq 1$, de sorte que $K \geqq v$ et $K \leqq \frac{4}{\varepsilon^2} \log N$. On tire de (4)

$$\frac{\operatorname{Log} N}{v} \leqq \operatorname{Log}\left(\frac{4e}{\varepsilon^2}\right) + \operatorname{Log}\left(\frac{\operatorname{Log} N}{v}\right).$$

Comme $\operatorname{Log} x \leqq \frac{x}{e}$ pour $x > 0$, on obtient

$$(5) \qquad \left(1 - \frac{1}{e}\right) \frac{\operatorname{Log} N}{v} \leqq \operatorname{Log} \frac{4e}{\varepsilon^2}$$

d'où on tire l'estimation (1). [On notera que si $\frac{2}{\varepsilon^2} \operatorname{Log} N < v$, alors (5) est encore vérifiée car $\varepsilon \leqq 1$.]

Pour démontrer le théorème 1, nous utiliserons une majoration (maintenant bien connue) pour les processus gaussiens ou sous-gaussiens, qui est dûe à Dudley. Pour la démonstration, cf. par exemple [*12*], p. 25.

LEMME 5. — *Soit $x_1, \dots, x_n$ des fonctions bornées sur un ensemble* T. *On pose $d(s, t) = (\Sigma \mid x_i(t) - x_i(s) \mid^2)^{1/2}$ pour tous s, t dans* T.

$$(6) \qquad \mathbb{E} \| \Sigma \varepsilon_i x_i \|_{l^\infty(T)} \leqq D \left\{ \sup_{t \in T} (\Sigma \mid x_i(t) \mid^2)^{1/2} + \int_0^\infty (\operatorname{Log} N(T, d; \varepsilon))^{1/2} d\varepsilon \right\},$$

où D *est une constante numérique.*

Démonstration du théorème 1. — Montrons tout d'abord $(i) \Rightarrow (ii)$. Soit $\mu_1, \dots, \mu_n$ dans $M(\Omega, \mathscr{A})$. Soit $P = \frac{1}{\Sigma \mid \mu_i \mid (\Omega)^2} \sum_1^n \mid \mu_i \mid (\Omega) \mid \mu_i \mid$.

Alors P est une probabilité sur $(\Omega, \mathscr{A})$ et on a

$$\Sigma_1^n \mid \mu_i(t)-\mu_i(s) \mid^2 \leqq \Sigma \mid \mu_i \mid (t \Delta s)^2 \leqq \Sigma \| \mu_i \|^2 \mathrm{P}(t \Delta s).$$

On déduit donc de (6)

$$\mathbb{E} \| \Sigma \varepsilon_i \mathrm{J}(\mu_i) \|_{l^\infty(\mathscr{C})} \leqq \mathrm{D} \left\{ (\Sigma \| \mu_i \|^2)^{1/2} \right\} \left(1 + \int_0^\infty (\operatorname{Log} \mathrm{N}_p(\varepsilon))^{1/2} d\varepsilon \right).$$

D'après (1), on conclut que J est de type 2 et

$$\mathrm{T}_2(\mathrm{J}) \leqq \mathrm{D}(1 + \alpha' \sqrt{v(\mathscr{C})}) \tag{7}$$

où D et α' sont des constantes numériques.

$(ii) \Rightarrow (iii)$ est trivial.

$(iii) \Rightarrow (iv)$ est standard, cf. e. g. [*8*]. En effet, par des arguments de symétrisation classiques, on montre que l'on a (en supposant (ε_i) indépendant de (ω_i))

$$\left\{ \begin{aligned} & \mathbb{E} \| \Sigma_1^n \mathrm{J}(\delta_{\omega_i} - \mathbb{P}) \| \leqq \mathbb{E} \| \Sigma_1^n \mathrm{J}(\delta_{\omega_i} - \delta_{\omega_{i+n}}) \| \\ & = \mathbb{E} \| \Sigma_1^n \varepsilon_i \mathrm{J}(\delta_{\omega_i} - \delta_{\omega_{i+n}}) \| \leqq 2\, \mathbb{E} \| \Sigma_1^n \varepsilon_i \mathrm{J}(\delta_{\omega_i}) \| . \end{aligned} \right. \tag{8}$$

[La première inégalité s'obtient en intégrant par rapport aux ω_{i+n} à l'intérieur de la norme et la dernière résulte de l'inégalité triangulaire ; l'égalité résulte de la symétrie de $\delta_{\omega_i} - \delta_{\omega_{n+i}}$.]

On en déduit donc

$$\mathbb{E} \| \Sigma_1^n \mathrm{J}(\delta_{\omega_i} - \mathbb{P}) \| \leqq 2\mathrm{T}_p(\mathrm{J}) n^{1/p}$$

d'où $\mathbb{E} \operatorname{Sup}_{\mathrm{C} \in \mathscr{C}} \mid v_n(\mathrm{C}) \mid \leqq 2\mathrm{T}_p(\mathrm{J}) n^{-1/p'}$, avec $\dfrac{1}{p} + \dfrac{1}{p'} = 1$, ce qui entraîne évidemment (iv) avec $a_n \approx n^{-1/p'}$.

Suivant les mêmes idées, on montre aisément que $(iii) \Rightarrow (v)$. En effet (iii) entraîne que la série $\Sigma_1^\infty n^{-\alpha} \mathrm{J}(\delta_{\omega_n} - \mathbb{P})$ est convergente, si $\alpha > 1/p$, p. s. et dans $\mathrm{L}^1(\mathbb{P}^{\mathbb{N}} ; l^\infty(\mathscr{C}))$; d'où (d'après le lemme de Kronecker) la convergence p. s. vers 0 de $\left\| n^{-\alpha} \sum_{i=1}^n \mathrm{J}(\delta_{\omega_i} - \mathbb{P}) \right\|$, ce qui donne (v) avec $a_n = n^{\alpha - 1}$ et $1/p < \alpha < 1$. On a évidemment $(v) \Rightarrow (iv)$. Il ne nous reste donc plus qu'à montrer que $(iv) \Rightarrow (i)$.

Supposons donc (iv). Il résulte alors des inégalités d'Hoffmann-Jørgensen (cf. [*7*], p. 163-165) que

$$\operatorname{Sup}_n \frac{1}{a_n} \mathbb{E} \operatorname{Sup}_{\mathrm{C} \in \mathscr{C}} \mid v_n(\mathrm{C}) \mid < \infty$$

pour toute probabilité $\mathbb{P}$ sur $(\Omega, \mathscr{A})$.

Soit (ε_i) une suite de v. a. de Bernoulli indépendante de la suite (X_i).

On peut évidemment supposer que $\sup_n \frac{1}{a_n\sqrt{n}} \leqslant \beta < \infty$.

Posons $\Phi(\mathbb{P}) = \sup_n \frac{1}{na_n} \mathbb{E} \sup_{C\in\mathscr{C}} | \langle \Sigma_1^n \varepsilon_i \delta_{\omega_i}, C \rangle |$ (où l'espérance porte sur tous les (ε_i) et (ω_i).).

On a alors $\Phi(\mathbb{P}) < \infty$ pour toute probabilité $\mathbb{P}$.

En effet, on a (d'après (8))

$$\begin{aligned}\Phi(\mathbb{P}) &\leqq \beta + \sup_n \left(\frac{1}{na_n}\right) \mathbb{E} \sup_{C\in\mathscr{C}} | \langle \Sigma_1^n \varepsilon_i(\delta_{\omega_i} - \mathbb{P}), C \rangle | \\ &\leqq \beta + 2 \sup_n \frac{1}{a_n} \mathbb{E} \sup_{\mathscr{C}} | \nu_n(C) | .\end{aligned}$$

Donc $\Phi(\mathbb{P}) < \infty$.

Nous allons maintenant montrer qu'il existe une constante C telle que

$$\Phi(\mathbb{P}) \leqq C \qquad \text{pour toute probabilité } \mathbb{P} . \tag{9}$$

Pour cela, montrons tout d'abord que si Q_0, Q_1 sont deux probabilités sur $(\Omega, \mathscr{A})$, si $0 \leqq \alpha \leqq 1$ et si $\mathbb{P} = \alpha Q_0 + (1-\alpha)Q_1$, on a

$$\Phi(\mathbb{P}) \geqq \alpha\Phi(Q_0) .$$

En effet, on peut trouver sur un même espace de probabilité, (Ω', P') des v. a. Y^0 et Y^1 à valeurs dans $(\Omega, \mathscr{A})$ et de lois respectives Q_0 et Q_1 ; soit alors X_n la v. a. définie sur $\Omega'' = (\{0, 1\} \times \Omega')^{\mathbb{N}}$ de la manière suivante

$$X_n((t_k, \omega'_k)) = Y^{t_n}(\omega'_n) .$$

On munit Ω'' de la probabilité produit infini de $(\alpha\delta_0 + (1-\alpha)\delta_1) \times \mathbb{P}'$, de sorte que les v. a. (X_n) sont indépendantes équidistribuées et de loi $\mathbb{P}$. Posons $| \mu |_{\mathscr{C}} = \sup \{ | \mu(C) |, C \in \mathscr{C} \}$ pour tout μ dans $M(\Omega, \mathscr{A})$. On peut écrire, pour $(t_k, \omega'_k)_k$ fixé dans Ω''

$$\begin{aligned}\mathbb{E}_\varepsilon | \Sigma_1^n \varepsilon_i \delta_{X_i} |_{\mathscr{C}} &\geqq \mathbb{E}_\varepsilon | \Sigma_1^n \varepsilon_i 1_{\{t_i=0\}} \delta_{X_i} |_{\mathscr{C}} \\ &\geqq \mathbb{E}_\varepsilon | \Sigma_1^n \varepsilon_i 1_{\{t_i=0\}} \delta_{Y^0(\omega'_i)} |_{\mathscr{C}} .\end{aligned}$$

D'où en intégrant en (t_i, ω'_i)

$$\begin{aligned}(na_n)\Phi(\mathbb{P}) &\geqq \mathbb{E}_{(t_i)} \mathbb{E}_{(\omega'_i)} \mathbb{E}_\varepsilon | \Sigma_1^n \varepsilon_i 1_{\{t_i=0\}} \delta_{Y^0(\omega'_i)} |_{\mathscr{C}} \\ &\geqq \mathbb{E}_{(\omega'_i)} \mathbb{E}_\varepsilon | \Sigma_1^n \varepsilon_i \alpha \delta_{Y^0(\omega'_i)} |_{\mathscr{C}}\end{aligned}$$

(On a utilisé l'inégalité de Jensen par rapport aux variables (t_i)). On a donc finalement l'inégalité annoncée $\Phi(\mathbb{P}) \geqq \alpha\Phi(Q_0)$.

L'inégalité (9) en résulte, car sinon il existerait une suite de probabilités P_k sur $(\Omega, \mathscr{A})$ telle que $\Phi(P_k) \geqq 4^k$ pour tout k ; on aurait alors, pour $P = \Sigma_1^\infty 2^{-k} P_k$ $\Phi(P) \geqq 2^{-k}\Phi(P_k) = 2^k$, ce qui contredirait le fait que $\Phi(P) < \infty$.

Enfin, pour conclure, montrons que (9) entraîne (*i*). Supposons que $\mathscr{C}$ n'est pas une classe V. C. Alors pour tout k, il existe une partie $A = \{ \omega_1, \ldots, \omega_k \}$ de Ω telle que toute partie de A soit de la forme $C \cap A$ pour un C dans $\mathscr{C}$.

Notons que l'on a alors trivialement :

$$(10) \qquad | \Sigma_1^k \alpha_i \delta_{\omega_i} |_{\mathscr{C}} \geqq \frac{1}{2} \Sigma | \alpha_i | \qquad \forall (\alpha_i) \in \mathbb{R}^k .$$

On va aboutir à une contradiction.

En effet, soit (Z_n) une suite de v. a. indépendantes à valeurs dans Ω, de loi $Q = \frac{1}{k} \Sigma_1^k \delta_{\omega_i}$. On a $\Phi(Q) \leqq C$, donc, pour tout n,

$$\mathbb{E} | \Sigma_1^n \varepsilon_i \delta_{Z_i} |_{\mathscr{C}} \leqq Cna_n .$$

Soit $\Omega_0 = \{ Z_i \neq Z_j, \forall i \neq j \leqq n \}$.

Supposons $n^2 < k/2$ de sorte que $\text{Proba}(\Omega_0) > 1 - \frac{n^2}{k} \geqq \frac{1}{2}$. On peut écrire

$$\mathbb{E} 1_{\Omega_0} | \Sigma_1^n \varepsilon_i \delta_{Z_i} |_{\mathscr{C}} \leqq Cna_n$$

d'où d'après (10)
$$\frac{n}{4} \leqq Cna_n ,$$

ce qui contredit l'hypothèse $a_n \to 0$ quand $n \to \infty$. C. Q. F. D.

REMARQUE 6. — On déduit de (7) que l'on a :

$$(10) \qquad T_2(J) \leq C' \sqrt{v(\mathscr{C})}$$

pour une constante numérique C'.

On peut remarquer que l'inégalité inverse est aussi vraie :

$$(11) \qquad T_2(J) \geqq C'' \sqrt{v(\mathscr{C})}$$

pour une constante numérique $C'' > 0$.

En effet, soit $v = v(\mathscr{C})$, soit $A \subset \Omega$ avec $| A | = v - 1$ $(v \geqq 1)$ et tel que toute partie de A soit de la forme $A \cap C$ pour un C dans $\mathscr{C}$. Alors, on a, si $A = \{ \omega_1, \ldots, \omega_{v-1} \}$

$$\mathbb{E} \left\| \sum_{i \leqq v-1} \varepsilon_i J(\delta_{\omega_i}) \right\|_{l^\infty(\mathscr{C})} \leqq T_2(J) \sqrt{v-1} .$$

D'où d'après (10) : $\frac{1}{2}(v-1) \leqq T_2(J)\sqrt{v-1}$.

Soit $T_2(J) \geqq \frac{1}{2}\sqrt{v-1}$, ce qui donne (11). On montrerait de même que $T_p(J)$ est équivalent à $v(\mathscr{C})^{1/p'}$, $\frac{1}{p}+\frac{1}{p'}=1$.

Remarque. — Soit $\mathscr{C} \subset \mathscr{A}$ une classe (supposée dénombrable) d'ensembles mesurables. Supposons que pour toute suite de v. a. indépendantes (Y_n) [non nécessairement équidistribuées] à valeurs dans $(\Omega, \mathscr{A})$ on a :

$$(12) \qquad \sup_{C\in\mathscr{C}} \left| \frac{1}{n} \Sigma_1^n (\delta_{Y_i}(C) - \mathbb{P}\{ Y_i \in C \}) \right| \to 0$$

en probabilité quand $n \to \infty$.

Alors $\mathscr{C}$ est nécessairement une classe V. C. Il suffit pour le voir de reprendre la démonstration précédente.

En effet posons

$$\lambda_n = \sup \mathbb{E} \left\| \frac{1}{n} \Sigma_1^n \varepsilon_i \delta_{X_i}(C) \right\|_{l^\infty(\mathscr{C})}$$

où le sup porte sur toutes les suites de v. a. indépendantes *équidistribuées* (X_i), à valeurs dans $(\Omega, \mathscr{A})$. On voit aisément que si $\lambda_n \not\to 0$ alors on peut construire une suite (Y_n) comme ci-dessus qui ne vérifie pas (12).

Donc (12) entraîne que $\lambda_n \to 0$, c'est-à-dire qu'il existe une suite $a_n \to 0$ pour laquelle on a (9); or on vient de voir que (9) a lieu ssi $\mathscr{C}$ est une classe de V. C.

§ 3. APPLICATIONS AUX ESPACES DE BANACH

Le théorème 2 nous permet d'améliorer un résultat de Milman (cf. [*13*]). lui-même inspiré d'un résultat de [*15*] (§ 7).

THÉORÈME 7. — *Soit $x_1, \ldots, x_n$ des fonctions à valeurs dans $\{-1, +1\}$ sur un ensemble* T.

Posons $M = \mathbb{E} \sup_{t\in T} | \Sigma_1^n \varepsilon_i x_i(t) |$.

Alors, pour tout k tel que $k \leqq \gamma \frac{M^2}{n}$ (où γ est une constante numérique), il existe une suite extraite $m_1 < m_2 < \ldots < m_k$ telle que l'ensemble des valeurs $\{ x_{m_1}(t), \ldots, x_{m_k}(t) \}$ soit égal à $\{-1, +1\}^k$. La suite extraite engendre donc dans $l^\infty(T)$ un sous-espace isométrique à l_k^1.

Démonstration. — Posons $M' = \mathbb{E} \operatorname{Sup}_t \left| \Sigma_1^n \varepsilon_1 \left(\frac{1 + x_i(t)}{2} \right) \right|$.

Soit $\mathscr{C}$ la classe de parties de $\{ 1, 2, \ldots, n \}$ de la forme

$$A_t = \{ i \in \{ 1, \ldots, n \} \mid x_i(t) = 1 \}.$$

En appliquant le théorème 1 à cette classe, on trouve $M' \leqq T_2(J)\sqrt{n}$ d'où d'après (10)

$$v \geqq \frac{1}{C'^2} \frac{M'^2}{n}.$$

Par conséquent, pour tout $k < C'^{-2}M'^2/n$ on a bien la conclusion du théorème 7. On obtient le résultat annoncé en observant que

$$\frac{1}{2}(M - \sqrt{n}) \leqq M' \leqq \frac{1}{2}(M + \sqrt{n}) \qquad \text{et} \qquad M \geqq \sqrt{\frac{n}{2}}.$$

Remarques. — *i*) Le cas de fonctions arbitraires x_i à valeurs dans l'intervalle $[-1, 1]$ est étudié dans [*6*], pour M de l'ordre de n.

ii) On sait (cf. e. g. [*13*]) que l'estimation du théorème 7 est la meilleure possible.

iii) Le lecteur trouvera dans [*14*] une généralisation à des fonctions (x_i) prenant un nombre fini fixé de valeurs.

BIBLIOGRAPHIE

[*1*] R. DUDLEY, Central limit theorem for empirical measures. *Ann. Probability*, t. **6**, 1978, p. 899-929.

[*2*] R. DUDLEY, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.*, t. **1**, 1967, p. 290-330.

[*3*] R. DUDLEY, W. PHILIPP, *Invariance principles for sums of Banach space valued random elements and empirical processes*. A paraître.

[*4*] R. DUDLEY, A course on empirical processes. École d'Été de St Flour, 1982. *Springer Lecture Notes*, à paraître.

[*5*] M. DURST, R. DUDLEY, Empirical processes, Vapnik-Červonenkis classes and Poisson processes. *Probability and Math. Statistics (Wroclaw)*. t. **1**, 1980, p. 109-115.

[*6*] J. ELTON, Sign embeddings of l_1^n. *Trans. A. M. S.*, t. **279**, 1983, p. 113-124.

[*7*] J. HOFFMANN-JORGENSEN, Sums of independent Banach space valued random variables. *Studia Math.*, t. **52**, 1974, p. 159-186.

[*8*] J. HOFFMAN-JØRGENSEN et G. PISIER, The law of large numbers and the central limit theorem in Banach spaces. *Annals of Proba.*, t. **4**, 1976, p. 587-599.

[*9*] R. C. JAMES, Non reflexive spaces of type 2. *Israël J. Math.*, t. **30**, 1978, p. 1-13.

[*10*] J. KUELBS, R. DUDLEY, Log Log Laws for empirical measures. *Ann. Probability*, t. **8**, 1980, p. 405-418.

[*11*] J. KUELBS, J. ZINN, Some stability results for vector valued random variables. *Annals of Probability*, t. **7**, 1979, p. 75-84.

[12] M. B. MARCUS, G. PISIER, Random Fourier series with applications to Harmonic Analysis. *Annals of Math. Studies*, n° 101, 1981. Princeton Univ. Press.

[13] V. MILMAN, Some remarks about embedding of l_1^k in finite dimensional spaces. *Israël J. Math.*, t. **43**, 1982, p. 129-138.

[14] A. PAJOR, Thèse de 3e cycle en préparation et Plongement de l_n^1 dans les espaces de Banach complexes. *C. R. Acad. Sci.*, t. **296**, 1983, p. 741-743.

[15] G. PISIER, De nouvelles caractérisations des ensembles de Sidon. *Advances in Math. Supplementary Studies*. t. **7 B**, 1981, p. 685-726.

[16] N. SAUER, On the density of families of sets. *J. Combinatiorial theory*, t. **A 13**, 1972, p. 145-147.

[17] M. STEELE, Empirical discrepancies and subadditive processes. *Annals of Probability*, t. **6**, 1978, p. 118.

[18] VAPNIK-ČERVONENKIS, On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Prob. Appl.*, t. **16**, 1971, p. 264-280.

[19] J. ZINN, A note on the central limit theorem in Banach spaces. *Annals of Proba.*, t. **5**, 1977, p. 283-286.

(Manuscrit reçu le 16 janvier 1984)