PROBABILISTIC LIMIT THEOREMS IN THE SETTING OF BANACH SPACES

M. Ledoux, J. Zinn

University of Toulouse and Texas A & M University

Abstract. — *We survey the developments of Banach space techniques in the context of classical limit theorems in probability theory. Weak convergence and relation to type and cotype, as well as the exponential bounds parts of the concentration phenomenon for products measures are the highlights of the story. Applications to empirical measures and the bootstrap complete the exposition.*

1. Introduction

Throughout this paper, B will denote a real separable Banach space with norm $\|\cdot\|$, and topological dual space B'. The separability assumption is most convenient to avoid a number of measurability questions. Complex Banach spaces are treated similarly. The basic object of investigation is a Borel random variable X on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in B, and a sequence $(X_n)_{n\in\mathbb{N}}$ of independent copies of X. For each $n \geq 1$, set $S_n = X_1 + \cdots + X_n$.

Classical probability theory on $\mathbb{R}$ or $\mathbb{R}^k$ is mostly concerned with the limiting behaviour of the partial sum sequence $(S_n)_{n\geq 1}$. The most important and famous results are the (strong) law of large numbers (LLN), the central limit theorem (CLT) and the law of the iterated logarithmic (LIL) which, for real-valued random variables, may be summarized in the following way. (We refer to [L-T3] for the history of these results.) Let X be a real-valued random variable.

– The sequence $(S_n/n)_{n\geq 1}$ converges almost surely to $\mathbb{E}(X)$ if and only if $\mathbb{E}(|X|) < \infty$ (we then say that X satisfies the LLN, or strong LLN).

– The sequence $(S_n/\sqrt{n})_{n\geq 1}$ converges in distribution (to a normal random variable G) if and only if $\mathbb{E}(X) = 0$ and $\sigma^2 = \mathbb{E}(X^2) < \infty$ (and in this case, G is centered with variance σ^2) (we then say that X satisfies the CLT).

– Define, on $\mathbb{R}^+$, the function $\mathrm{LL}x = \log\log x$ if $x \geq \mathrm{e}$, and $\mathrm{LL}x = 1$ if $x < \mathrm{e}$, and set $a_n = (2n\mathrm{LL}n)^{1/2}$ for every $n \geq 1$. The sequence $(S_n/a_n)_{n\geq 1}$ is almost surely bounded if and only if $\mathbb{E}(X) = 0$ and $\sigma^2 = \mathbb{E}(X^2) < \infty$, and in this case,

$$\liminf_{n\to\infty} \frac{S_n}{a_n} = -\sigma \qquad \text{and} \qquad \limsup_{n\to\infty} \frac{S_n}{a_n} = +\sigma$$

with probability 1. Moreover, the set of limit points of the sequence $(S_n/a_n)_{n\geq 1}$ is almost surely equal to the interval $[-\sigma, +\sigma]$ (we then say that X satisfies the LIL).

With the exception of the last statement on the LIL these statements may be shown to easily extend to finite dimensional random variables, with the obvious modifications.

The definitions of these basic limit theorems extend to random variables taking values in a infinite dimensional real separable Banach space B. For example, weak convergence in the central limit theorem has to be understood as weak convergence in the space of Borel probability measures on the complete separable metric space B. For the LIL, one has to distinguish between a bounded form (the sequence $(S_n/a_n)_{n\geq 1}$ is almost surely bounded in B), and a compact form (the sequence $(S_n/a_n)_{n\geq 1}$ is almost surely relatively compact in B). In the latter case, it may be shown, completely generally [Ku1], that the set of limit points of the sequence $(S_n/a_n)_{n\geq 1}$ is a compact convex symmetric set in B (the unit ball of the reproducing kernel Hilbert space associated to the covariance structure of the random variable X).

Moment conditions on the law of X fully characterize the preceding limit theorems in finite dimension. However, as it was soon realized, this is no longer true in infinite dimension. At this point emphasis was put on understanding what kind on conditions on the space can ensure an extension of the finite dimensional statements, and what new descriptions are available in this setting. In the first part of this survey, we describe the almost sure limit theorems (LLN and LIL). As a main observation, it was established, as a consequence of deep exponential bounds, which are parts of the concentration of measure phenomenon for products measures, that the almost sure statements actually reduce to the corresponding ones in probability or in distribution under necessary moment conditions. It is a main contribution of the Banach space approach to realize that moment conditions are actually used to handle convergence in distribution. This fact is further illustrated in Section 3 in the investigation of the classical central limit theorem using type and cotype. In the last paragraph, we describe applications of these ideas and techniques to empirical processes and bootstrap in statistics. For convenience, we mostly refer to the monograph [L-T3] for a complete account on the subject of probability in Banach spaces, and for further references and historical developments. We also refer to [L-T3] for the complete proofs that are only outlined here.

2. Almost sure limit theorems

In the early fifties, emphasis was made in trying to understand the strong limit theorems (LLN and LIL) for infinite dimensional random variables, following early indications by A. N. Kolmogorov. In this direction, E. Mourier and R. Fortet (cf. [L-T3]) extended the LLN in a statement completely analogous to the finite dimensional setting.

Theorem 1. *Let X be a random variable with values in a Banach space B. Then the sequence $(S_n/n)_{n\geq 1}$ converges almost surely to $\mathbb{E}(X)$ if and only if $\mathbb{E}(\|X\|) < \infty$.*

Here, when $\mathbb{E}(\|X\|) < \infty$, the expectation $\mathbb{E}(X)$ has to be understood as the element of B such that $\langle \xi, \mathbb{E}(X)\rangle = \mathbb{E}(\langle \xi, X\rangle)$ for every $\xi \in B'$. The modern proof (see [HJ3]) of Theorem 1 is rather straightforward.

Proof. The necessity of the moment condition $\mathbb{E}(\|X\|) < \infty$ is proved as in the real case with the Borel-Cantelli lemma. Assume thus that $\mathbb{E}(\|X\|) < \infty$. Without loss of generality, we can then assume that $\mathbb{E}(X) = 0$. For each $\varepsilon > 0$, let then Y be a centered step random variable (taking finitely many values only) in B such that $\mathbb{E}(\|X - Y\|) \leq \varepsilon$. Consider independent copies $(Y_n)_{n\in\mathbb{N}}$ of Y, and for every $n \geq 1$, set $T_n = Y_1 + \cdots + Y_n$. By the finite dimensional LLN,

$$\lim_{n\to\infty} \frac{1}{n}\,\|T_n\| = 0 \tag{1}$$

almost surely. On the other hand, by the triangle inequality,

$$\frac{1}{n}\,\|S_n - T_n\| \leq \frac{1}{n}\sum_{i=1}^{n} \|X_i - Y_i\|,$$

and by the LLN on the line applied to $\|X - Y\|$, with probability 1,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \|X_i - Y_i\| = \mathbb{E}\big(\|X - Y\|\big) \leq \varepsilon. \tag{2}$$

Summarizing (1) and (2),

$$\limsup_{n\to\infty} \frac{1}{n}\,\|S_n\| \leq \varepsilon$$

almost surely. Since $\varepsilon > 0$ is arbitrary, the conclusion follows. □

Soon after Theorem 1, research concentrated for some time on related forms of the strong LLN in Banach spaces, in particular the so-called Kolmogorov LLN that states that if $(Y_i)_{i\in\mathbb{N}}$ are independent, but not necessarily identically distributed, mean-zero real random variables such that

$$\sum_i \frac{1}{i^2}\,\mathbb{E}(Y_i^2) < \infty,$$

then the sequence $\frac{1}{n}\sum_{i=1}^{n} Y_i$ converges almost surely to 0. A. Beck's fundamental discovery in 1962 [Be] was that the extension of Kolmogorov's LLN could not take place in any infinite dimensional Banach space. More importantly, he characterized, through a geometric condition known as

Beck's convexity, those Banach spaces B for which every sequence $(Y_i)_{i\in\mathbb{N}}$ of independent mean-zero B-valued random variables with $\sup_i \|Y_i\|_\infty < \infty$ satisfies the LLN, i.e.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to 0 \quad \text{almost surely.}$$

In particular, spaces such as $\ell_1, \mathrm{L}_1, c_0, C(K)$ do not satisfy Beck's condition, and therefore Kolmogorov's LLN.

Beck's convexity condition was then recognized in the early seventies through the basic work of G. Pisier [Pi1] as equivalent to a probabilistic type condition. This condition, as well as a related "dual" condition, were introduced independently by B. Maurey [Ma] and Hoffmann-Jørgensen [HJ1]. Let us a recall that a Banach space B is of type p, $1 < p \leq 2$, if there is a constant C such that for every $x_1, \ldots, x_n$ in B,

$$\mathbb{E}\bigg(\bigg\|\sum_{i=1}^{n} \varepsilon_i x_i\bigg\|^p\bigg) \leq C \sum_{i=1}^{n} \|x_i\|^p.$$

Here (and throughout this paper), $(\varepsilon_i)_{i\in\mathbb{N}}$ is a Rademacher sequence, that is a sequence of independent random variables with common distribution $\mathbb{P}\{\varepsilon_i = +1\} = \mathbb{P}\{\varepsilon_i = -1\} = \frac{1}{2}$. Similarly, B is said to be of cotype q, $2 \leq q < \infty$, if there is a constant C such that for every $x_1, \ldots, x_n$ in B,

$$\sum_{i=1}^{n} \|x_i\|^q \leq C\, \mathbb{E}\bigg(\bigg\|\sum_{i=1}^{n} \varepsilon_i x_i\bigg\|^q\bigg).$$

L_p-spaces are of type $\min(p,2)$ and of cotype $\max(p,2)$. By Kwapień's theorem [Kw1], Banach spaces of both type 2 and cotype 2 are isomorphic to Hilbert spaces.

G. Pisier's discovery in 1973 was that Beck's convexity condition for a Banach space B is equivalent to the fact that B is of type p for some $p > 1$. Now, the probabilistic type condition is easily seen to be well-adapted to the investigation of the LLN and resulted in the following

theorem due to J. Hoffmann-Jørgensen and G. Pisier [HJ-P] that extends Beck's result. It should be noted that this last paper brought to the forefront the usefulness of symmetrization by Rademacher variables $(\varepsilon_i)_{i\in\mathbb{N}}$. This still has a great impact.

Theorem 2. *A Banach space B is of type $p > 1$ if and only if for every sequence $(Y_i)_{i\in\mathbb{N}}$ of independent mean-zero B-valued random variables such that*

$$\sum_i \frac{1}{i^p}\, \mathbb{E}\big(\|Y_i\|^p\big) < \infty,$$

one has

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to 0$$

almost surely.

Up to this point, the results developed as natural extensions, in spaces with some type, of the classical theorems in finite dimension. An important step was performed with the contribution of V. V. Yurinskii [Yu1], [Yu2] (whose interests were in exponential inequalities), and applications of his ideas by J. Kuelbs [Ku2] (for the LIL) and J. Kuelbs and J. Zinn [K-Z] (for the LLN). With these results, the Banach space conceptualizations started to have an important impact on the probabilistic analysis. Given $Y_1, \ldots, Y_n$ independent integrable Banach space valued random variables, V. V. Yurinskii's observation was that the norm of the sum $S = \sum_{i=1}^n Y_i$ centered at its mean may be written as a sum of martingale differences

$$\|S\| - \mathbb{E}\big(\|S\|\big) = \sum_{i=1}^{n} d_i$$

with respect to the filtration $\mathcal{F}_i = \sigma(Y_1, \ldots, Y_i)$, $i = 1, \ldots, n$, (i.e. $\mathbb{E}(d_i \,|\, \mathcal{F}_{i-1}) = 0$) with the property that, for every $i = 1, \ldots, n$,

$$|d_i| \le \|Y_i\| + \mathbb{E}\big(\|Y_i\|\big). \tag{3}$$

In a sense, $\|S\| - \mathbb{E}(\|S\|)$ is as good as the sum $\sum_{i=1}^n \|Y_i\|$, so that, provided $\mathbb{E}(\|S\|)$ is under control, the classical one-dimensional results should apply similarly. Together with this representation, J. Kuelbs and J. Zinn proved indeed the following.

Theorem 3. *Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of independent B-valued random variables such that, for some p, $1 \le p \le 2$,*

$$\sum_i \frac{1}{i^p}\, \mathbb{E}\big(\|Y_i\|^p\big) < \infty.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to 0 \quad \textit{almost surely}$$

if and only if

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to 0 \quad \textit{in probability.}$$

Proof. Under the hypothesis, $Y_i/i \to 0$ almost surely, so that we may assume that $\|Y_i\|_\infty \le i$ for every $i \ge 1$. We then symmetrize so that blocking is easily managed.

Here is one way to symmetrize (see immediately after the proof for symmetrization in L_p). Set $S_n = Y_1 + \cdots + Y_n$, $n \geq 1$. Clearly, $S_n/n \to 0$ almost surely if $(S_n - S'_n)/n \to 0$ almost surely, where S'_n/n is formed from an independent copy of the original (Y_i), and $S_n/n \to 0$ in probability. Indeed, if $(S_n - S'_n)/n \to 0$ almost surely, then, by Fubini's theorem, we may find ω' so that $S_n/n - S'_n(\omega')/n \to 0$ almost surely which, in particular, implies this last quantity goes to zero in probability. But, then since $S_n/n \to 0$ in probability, $S'_n(\omega')/n \to 0$. Hence, $S_n/n \to 0$ with probability one. We may thus reduce ourselves to the case of independent symmetric random variables Y_i. Assume thus that $S_n/n \to 0$ in probability. Then, as a consequence of the Hoffmann-Jørgensen inequalities ([HJ3], [L-T3, Chapter 6]),

$$\frac{1}{n}\,\mathbb{E}\big(\|S_n\|\big) \to 0$$

as $n \to \infty$. Moreover, by the maximal inequalities for sums of independent symmetric random variables, it is enough to show that

$$\frac{1}{2^n} \sum_{i=2^{n-1}+1}^{2^n} Y_i \to 0$$

with probability 1. By Yurinskii's result, for every $\varepsilon > 0$, and every n,

$$\mathbb{P}\left\{\Big\|\sum_{i=2^{n-1}+1}^{2^n} Y_i\Big\| - \mathbb{E}\Big(\Big\|\sum_{i=2^{n-1}+1}^{2^n} Y_i\Big\|\Big) \geq \varepsilon n\right\}$$

$$\leq \frac{4}{\varepsilon^2 2^{2n}} \sum_{i=2^{n-1}+1}^{2^n} \mathbb{E}\big(\|Y_i\|^2\big) \leq \frac{4}{\varepsilon^2} \sum_{i=2^{n-1}+1}^{2^n} \frac{1}{i^p}\,\mathbb{E}\big(\|Y_i\|^p\big).$$

The conclusion then immediately follows from the Borel-Cantelli lemma. □

The important feature of Theorem 3 is that, under the convergence in probability of the partial sum sequence, no assumption has to be imposed on the Banach space. With respect to Theorem 2, the type condition is actually only used in order to achieve this convergence in probability. Assume indeed the

Y_i's to be centered and denote by $(Y'_i)_{i\in\mathbb{N}}$ an independent copy of the sequence $(Y_i)_{i\in\mathbb{N}}$. Then, by Jensen's inequality and the triangle inequality, for every n, and $p \geq 1$,

$$\begin{aligned}\mathbb{E}\Big(\Big\|\sum_{i=1}^n Y_i\Big\|^p\Big) &\leq \mathbb{E}\Big(\Big\|\sum_{i=1}^n (Y_i - Y'_i)\Big\|^p\Big)\\ &= \mathbb{E}\Big(\Big\|\sum_{i=1}^n \varepsilon_i (Y_i - Y'_i)\Big\|^p\Big) \leq 2^p\,\mathbb{E}\Big(\Big\|\sum_{i=1}^n \epsilon_i Y_i\Big\|^p\Big)\end{aligned} \tag{4}$$

where the Rademacher sequence $(\varepsilon_i)_{i\in\mathbb{N}}$ is independent from the previous sequences. Using the type inequality conditionally on the Y_i's, for some constant C depending only on B,

$$\mathbb{E}\left(\left\|\sum_{i=1}^n Y_i\right\|^p\right) \le 2^p C \sum_{i=1}^n \mathbb{E}\big(\|Y_i\|^p\big).$$

Therefore, if

$$\sum_i \frac{1}{i^p}\,\mathbb{E}\big(\|Y_i\|^p\big) < \infty,$$

then $\frac{1}{n}\sum_{i=1}^n Y_i \to 0$ in L_p by Kronecker's lemma, and thus in probability.

Another aspect of the preceding proof is the use of the Hoffmann-Jørgensen inequalities. Again, these may be considered as a consequence of the Banach space conceptualization. In its early formulation, the Hoffmann-Jørgensen inequality [HJ1], [HJ2] indicates that whenever $Y_1,\ldots,Y_n$ are independent symmetric random variables with values in B, for every s and $t>0$,

$$\mathbb{P}\big\{\|S\|\ge s+2t\big\} \le \mathbb{P}\Big\{\max_{1\le i\le n}\|Y_i\|\ge s\Big\} + 4\Big(\mathbb{P}\big\{\|S\|\ge t\big\}\Big)^2 \qquad (5)$$

where $S=\sum_{i=1}^n Y_i$. Inequality (5) is one amongst a variety of similar inequalities. Typically, it may be used to show that if $(a_n)_{n\in\mathbb{N}}$ is a sequence of positive numbers increasing to infinity and if $(Y_i)_{i\in\mathbb{N}}$ is a sequence of independent symmetric random variables, then, whenever the sequence

$$\frac{1}{a_n}\sum_{i=1}^n Y_i, \quad n\ge 1,$$

is bounded or converges to 0 in probability in B, the sequence

$$\frac{1}{a_n}\sum_{i=1}^n Y_i\,\mathrm{I}_{\{\|Y_i\|\le Ca_n\}}, \quad n\ge 1,$$

is bounded or converges to 0 in $\mathrm{L}^p(B)$ for any $p>0$ (cf. [HJ3], [L-T3] and the references therein).

The main consequence of Theorem 3 is that the classical probabilistic limit theorems have to be investigated, in a Banach space setting, in two distinct steps. Namely, under the classical moment conditions, prove convergence in probability or in distribution with the help of the type (or cotype) conditions. The resulting statements thus only hold for classes of Banach spaces with the appropriate geometric conditions. One typical and fundamental example of this situation is the central limit theorem which we investigate in the next section. Once the convergence in probability is achieved, or simply assumed, prove, in any Banach space, the corresponding almost sure statement. The lesson learned for limit theorems in Banach spaces is that moment conditions are needed to ensure convergence in probability and that, more or less, convergence in probability then always implies almost sure convergence.

Provided with these fundamental observations, we turn to some more refined almost sure statements, such as the law of the iterated logarithm (LIL). As in Theorem 1, let X be a B-valued random variable, and let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent copies of X. For each $n \geq 1$, $S_n = X_1 + \cdots + X_n$. Recall also $a_n = (2n\mathrm{LL}n)^{1/2}$. As expected from the preceding conclusions, and using exponential bounds on Yurinskii's martingale, J. Kuelbs showed in 1977 [Ku2] that the sequence $(S_n/a_n)_{n\geq 1}$ is relatively compact in B as soon as $\mathbb{E}(\|X\|^2) < \infty$ and $S_n/a_n \to 0$ in probability.

Although this result was a powerful extension of the classical LIL, it was not entirely satisfactory since the moment condition $\mathbb{E}(\|X\|^2) < \infty$ was known not to be necessary for the LIL. The necessary moment condition on the law of X to satisfy the LIL in an infinite dimensional Banach space B actually splits into two parts: first, for every linear functional $\xi \in B'$, the scalar random variable $\langle \xi, X\rangle$ satisfies the LIL, and thus $\mathbb{E}(\langle \xi, X\rangle) = 0$ and $\mathbb{E}(\langle \xi, X\rangle^2) < \infty$. Secondly, if the sequence $(S_n/a_n)_{n\geq 1}$ is almost surely bounded, so is the sequence $(X_n/a_n)_{n\geq 1}$, and thus, by the Borel-Cantelli lemma, $\mathbb{E}(\|X\|^2/\mathrm{L}L\|X\|) < \infty$. The occurence of weak moments with respect to the usual norm conditions made this investigation significantly harder than most of the previous results and showed that it lay at a much deeper level. It was thus open for some time to know whether these necessary moment conditions, together with the control of the sequence $(S_n/a_n)_{n\geq 1}$ in probability, were also sufficient for the LIL to hold. This conjecture was first settled in Hilbert spaces [G-K-Z] using the scalar product structure, and then further extended in smooth spaces (uniformly convex spaces [Le2]).

The final breakthrough was accomplished with the help of the isoperimetric and concentration ideas. The Gaussian isoperimetric inequality may be considered at the origin of this developement (cf. [L-T3, Chapter3]). This inequality in particular implies that if G is a Gaussian random vector with values in B, for every $t \geq 0$,

$$\mathbb{P}\big\{\|G\| - \mathbb{E}\big(\|G\|\big) \geq t\big\} \leq \mathrm{e}^{-t^2/2\sigma^2} \tag{6}$$

where $\sigma^2 = \sup_{\|\xi\|\leq 1} \mathbb{E}(\langle \xi, G\rangle^2)$. In particular,

$$\mathbb{E}\big(\exp\big(\alpha\|G\|^2\big)\big) < \infty$$

if and only if $\alpha < 1/2\sigma^2$. This fundamental Gaussian property led to a first result [L-T1] together with a Gaussian randomization argument put forward in [G-Z2] (and close in spirit to the proof of Lemma 13 below). This was actually the starting point of the deep investigation by M. Talagrand of isoperimetric and concentration inequalities in product spaces, with applications to a number

of various areas in both Banach space and probability theory [Ta2], [Ta3]. In particular, this approach yields optimal extensions of the classical exponential bounds on sums of independent random variables in the spirit of the Gaussian inequality

(6). Let for example $Y_1, \dots, Y_n$ be independent mean-zero B-valued random variables such that $\|Y_i\|_\infty \leq C$ for every $i = 1, \dots, n$. Set $S = Y_1 + \cdots + Y_n$ and define $\sigma^2 = \sup_{\|\xi\| \leq 1} \sum_{i=1}^n \mathbb{E}(\langle \xi, Y_i \rangle^2)$. Then, for every $t \geq 0$,

$$\mathbb{P}\big\{\|S\| - \mathbb{E}\big(\|S\|\big) \geq t\big\} \leq K \exp\left(-\frac{t}{KC} \log\left(1 + \frac{Ct}{\sigma^2 + C\mathbb{E}(\|S\|)}\right)\right) \tag{7}$$

where $K > 0$ is some numerical constant. This type of inequality is, using the centering factor $\mathbb{E}(\|S\|)$, the complete analogue of the classical Bennett inequalities for sums of independent real random variables, as well as the natural extension of (6). It describes the Gaussian behavior of sums of independent random variables for the small values of t, and the Poissonian behavior for the large values. A weaker form of (7) goes back to the early paper [Ta1] by M. Talagrand, which led the author to a complete study of the concentration phenomenon for product measures. The precise form of (7) is taken from the more recent work [Ta4]. An alternate simpler approach based on logarithmic Sobolev inequalities has been recently proposed in [Le3]

Together with such an estimate, it is easy to characterize the LIL in Banach spaces.

Theorem 4. *Let X be a random variable with values in a Banach space B.*

The sequence $(S_n/a_n)_{n\geq 1}$ is almost surely bounded if and only if it is bounded in probability, $\mathbb{E}(\|X\|^2/\mathrm{LL}\|X\|) < \infty$, and for every linear functional $\xi \in B'$, $\mathbb{E}(\langle \xi, X \rangle) = 0$ and $\mathbb{E}(\langle \xi, X \rangle^2) < \infty$.

The sequence $(S_n/a_n)_{n\geq 1}$ is almost surely relatively compact if and only $S_n/a_n \to 0$ in probability, $\mathbb{E}(\|X\|^2/\mathrm{LL}\|X\|) < \infty$, and the family of random variables $\langle \xi, X \rangle^2$, $\|\xi\| \leq 1$, is uniformly integrable.

Proof. Let us concentrate only on the bounded form of the LIL. The conditions have been seen to be necessary. As in the LLN, by a classical blocking argument, it is enough to show that

$$\sup_n \frac{1}{a_{2^n}} \left\| \sum_{i=1}^{2^n} X_i \right\| < \infty \tag{8}$$

almost surely. One can show that, under the integrability condition $\mathbb{E}(\|X\|^2/\mathrm{LL}\|X\|) < \infty$, there exists a sequence of integers $(k_n)_{n\in\mathbb{N}}$ such that $\sum_n 2^{-k_n} < \infty$ and

$$\sum_n \mathbb{P}\left\{\sum_{r=1}^{k_n} \|X_{2^n}^{(r)}\| \geq a_{2^n}\right\} < \infty$$

where $X_{2^n}^{(r)}$ is the r-th largest element of the sample $(\|X_i\|)_{1\leq i\leq 2^n}$. In particular, it is enough to prove (8) with the X_i's replaced by $Y_i = X_i I_{\{\|X_i\| \leq a_{2^n}/k_n\}}$, $i = 1, \dots, 2^n$. Since the sequence $(S_n/a_n)_{n\geq 1}$ is bounded in probability, by the Hoffmann-Jørgensen's inequalities, for some finite constant M,

$$\sup_n \frac{1}{a_{2^n}} \mathbb{E}\bigg(\bigg\|\sum_{i=1}^{2^n} Y_i\bigg\|\bigg) \le M.$$

On the other hand, by the closed graph theorem,

$$\sigma^2 = \sup_{\|\xi\|\le 1} \mathbb{E}\big(\langle \xi, X\rangle^2\big) < \infty.$$

Apply then (6) to $S = Y_1 + \cdots + Y_{2^n}$ to get, for $t = Ta_{2^n}$, $T > 0$,

$$\mathbb{P}\bigg\{\bigg\|\sum_{i=1}^{2^n} Y_i\bigg\| - \mathbb{E}\bigg(\bigg\|\sum_{i=1}^{2^n} Y_i\bigg\|\bigg) \ge Ta_{2^n}\bigg\}$$
$$\le K \exp\bigg(-\frac{Mk_n}{K}\log\bigg(1 + \frac{M}{(\sigma^2 k_n/LL2^n) + 1}\bigg)\bigg).$$

Since $\sum_n 2^{-k_n} < \infty$, the preceding exponential bound is the general term of a summable sequence for any T large enough. The conclusion follows. □

As for the LLN, it is an easy exercise to see that in a type 2 Banach space, whenever X has mean zero and $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$, then $S_n/a_n \to 0$ in probability. As a corollary to Theorem 4, one thus obtains quite a nice characterization of the LIL in type 2 Banach spaces only relying on moment conditions on the law of X.

Corollary 5. *Let X be a mean-zero random variable with values in a type 2 Banach space B. Then the sequence $(S_n/a_n)_{n\ge 1}$ is almost surely relatively compact if and only if $\mathbb{E}(\|X\|^2/\mathrm{LL}\|X\|) < \infty$ and the family of random variables $\langle \xi, X\rangle^2$, $\|\xi\| \le 1$, is uniformly integrable.*

Again, the type property on the Banach space B is only use to ensure the convergence in probability, or weak convergence. As we will now develop it, type and cotype are actually intimately connected with weak

convergence, and in particular the central limit theorem.

3. The central limit theorem and weak convergence

It was soon realized, at the beginning of probability in Banach spaces, that the classical moment conditions $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$ are neither sufficient, nor necessary for a random variable X to satisfy the central limit theorem in an arbitrary Banach space B. They are actually sufficient (only) in type 2 spaces, and necessary (only) in cotype 2 spaces.

However, in these first considerations of the CLT in Banach spaces, the interplay, which we now understand well, between the geometry of the space and the conditions on the random variable was still unclear. Results analogous to the classical results were proved in special spaces such as Hilbert space and for specialized random variables with values in spaces with a basis. Results were also proved in the form of limit theorems for stochastic processes. Some of the early highlights were S. Varadhan's central limit theorem in Hilbert space [Va], the results of R. M. Dudley and V. Strassen [D-S] on processes in $C[0,1]$ and R. M. Dudley's work on empirical processes [Du]. For more details on the historical development, see [L-T3], p. 296.

In a conference in Durham, England 1975, R. M. Dudley posed the question: In which (separable) Banach spaces do the classical conditions of mean zero and finite variance ($\mathbb{E}(\|X\|^2) < \infty$) always imply that the CLT holds? The question was answered at this conference and resulted in the theorem of J. Hoffmann-Jørgensen and G. Pisier [HJ-P] below.

Theorem 6. *A random variable X with values in a type 2 Banach space B satisfies the CLT as soon as $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$. Conversely, if in a Banach space B, all random variables X such that $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$ satisfy the CLT, then B is of type 2.*

The companion theorem for cotype soon followed.

Theorem 7. *A random variable X with values in a cotype 2 Banach space B satisfying the CLT is such that $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$. Conversely, if in a Banach space B, all random variables X satisfying the CLT are such that $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$, then B is of cotype 2.*

To illustrate the idea of these statements, let us outline the proof of Theorem 6.

Proof of Theorem 6. Assume B is of type 2. By the symmetrization argument (4), for every $n \geq 1$,

$$\mathbb{E}\Bigg(\bigg\|\sum_{i=1}^{n} X_i\bigg\|^2\Bigg) \leq 4Cn\,\mathbb{E}\big(\|X\|^2\big). \tag{9}$$

For $\varepsilon > 0$, let Y be a mean-zero step random variable such that $\mathbb{E}(\|X - Y\|^2) \leq \varepsilon/4C$. Therefore, if $(Y_i)_{i\in\mathbb{N}}$ is a sequence of independent copies of Y, (9) applied to $X - Y$ yields

$$\sup_{n\geq 1} \mathbb{E}\Bigg(\bigg\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - Y_i)\bigg\|^2\Bigg) \leq \varepsilon.$$

Since Y is finite dimensional, it satisfies the CLT. The sequence $(S_n/\sqrt{n})_{n\geq 1}$ is thus uniformly close in L^2 to a weakly convergent sequence, and thus is tight. Since, by the finite dimensional CLT, the only possible limit is Gaussian with the same covariance structure as X, X satisfies the CLT. Conversely, assuming that every mean-zero random variable X with values in B such that $\mathbb{E}(\|X\|^2) < \infty$ satisfies the CLT, we deduce from a closed graph argument that there is a finite constant C such that

$$\sup_{n\geq 1} \mathbb{E}\biggl(\biggl\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \biggr\|^2\biggr) \leq C\,\mathbb{E}\bigl(\|X\|^2\bigr).$$

Applying this inequality to a random variable X taking finitely many values $x_1, \ldots, x_n$, then shows that B is of type 2. □

Together with Kwapień's isomorphic characterization of Hilbert spaces, one can deduce from Theorems 6 and 7 the following "probabilistic" characterization of Hilbert spaces to which several authors contributed.

Corollary 8. *A Banach space B is isomorphic to a Hilbert space if and only if the classical moment conditions $\mathbb{E}(X) = 0$ and $\mathbb{E}(\|X\|^2) < \infty$ are necessary and sufficient for a random variable X with values in B to satisfy the CLT.*

As for the LIL, the strong second moment $\mathbb{E}(\|X\|^2) < \infty$ is not always well-adapted, and one has rather to consider weak moments. In the context of the CLT, an additional necessary condition for a random variable X to satisfy the CLT is the existence of a (centered) Gaussian random vector G with values in B (the limiting distribution) with the same covariance structure as X, that is $\mathbb{E}(\langle \xi, X\rangle^2) = \mathbb{E}(\langle \xi, G\rangle^2)$ for every ξ in B'. In an infinite dimensional context, it is not always true that there exists a Gaussian distribution with a given covariance structure. Actually, Theorems 6 and 7 may be rephrased replacing the central limit property by only the existence of a limiting Gaussian distribution. The proofs are rather similar.

In the presence of a limiting Gaussian distribution, the natural necessary moment condition on the norm, in order for a random variable X with values in a Banach space B to satisfy the CLT is that

$$\lim_{t\to\infty} t^2 \mathbb{P}\bigl\{\|X\| \geq t\bigr\} = 0. \tag{10}$$

In particular, $\mathbb{E}(\|X\|^p) < \infty$ only for $0 < p < 2$. It is then a challenging question to characterize those Banach spaces B in which the preceding condition (10) together with the existence of a limiting Gaussian

distribution are (necessary and) sufficient for a B-valued mean-zero random variable X to satisfy the CLT. As shown by G. Pisier and J. Zinn [P-Z], L_p-spaces with $2 \leq p < \infty$ share this property. If, in a Banach space a (certain) version of an inequality of H. P. Rosenthal holds, then [G-K-Z] the CLT holds if and only if (10) holds and the limiting Gaussian distribution exists (in the space). An attempt is made in [Le1] to characterize such spaces. These ideas lead to necessary and sufficient conditions for the CLT in spaces of the form $\mathrm{L}_p(\mathrm{L}_q)$ [G-Z1].

In a general Banach space, one cannot hope for any reasonable description of random variables satisfying the CLT. However, one can give sufficient conditions for some classes of random variables. Let for example $C(K)$ be

the Banach space of continuous functions on a compact metric space (K, d)

equipped with the uniform norm $\|\cdot\|_\infty$. Consider the class of Lipschitz random variables X on $C(K)$, that is such that for some non-negative random variable M,

$$\big|X(s,\omega) - X(t,\omega)\big| \le M(\omega)d(s,t), \quad s,t \in K,$$

for all (or almost all) ω. There were many partial steps in the direction of the "final" theorems in this direction, with contributions in [D-S], [Du] and [G]. A big step was taken in [J-M] with the following result.

Theorem 9. *Let X be a Lipschitz mean-zero random variable such that $\mathbb{E}(M^2) < \infty$. Then, whenever $d \le d_G$ where $d_G(s,t) = \|G(s) - G(t)\|_2$, $s,t \in K$, is the L_2-metric of a Gaussian random vector G in $C(K)$, then X satisfies the CLT.*

Proof. We follow the approach of [Zi] which is a nice illustration of the type 2 ideas. Consider indeed the canonical injection map $j : \mathrm{Lip}(K) \to C(K)$. The space $\mathrm{Lip}(T)$ is equipped with the norm

$$\|x\|_{\mathrm{Lip}} = \|x\|_\infty + \sup_{s\neq t} \frac{|x(s) - x(t)|}{d(s,t)}\,.$$

Although $C(K)$ is of no type $p > 1$, under the hypothesis of the theorem, the linear operator j is actually of type 2. Let indeed $x_1, \ldots, x_n$ be elements in $\mathrm{Lip}(K)$ such that, by homogeneity, $\sum_{i=1}^n \|x_i\|_{\mathrm{Lip}}^2 = 1$. By an elementary comparison (cf. [L-T3]),

$$\mathbb{E}\bigg(\bigg\|\sum_{i=1}^n \varepsilon_i j(x_i)\bigg\|_\infty^2\bigg) \le 2\,\mathbb{E}\bigg(\bigg\|\sum_{i=1}^n g_i j(x_i)\bigg\|_\infty^2\bigg)$$

where g_i are independent standard normal variables. Since the x_i's are element of $\mathrm{Lip}(K)$, the Gaussian process

$$\widetilde{G}(t) = \sum_{i=1}^n g_i x_i(t), \quad t \in K,$$

is such that, for every $s,t \in K$,

$$\mathbb{E}\big(\big|\widetilde{G}(s) - \widetilde{G}(t)\big|^2\big) = \sum_{i=1}^n \big|x_i(s) - x_i(t)\big|^2 \le d_G(s,t)^2.$$

Classical Gaussian comparison theorems (cf. [L-T3, Chapter 3]) then shows that

$$\mathbb{E}\bigg(\bigg\|\sum_{i=1}^n \varepsilon_i j(x_i)\bigg\|_\infty^2\bigg) \le C$$

where $C > 0$ only depends on G. Therefore, by homogeneity, j is an operator of type 2 in the sense that, whenever $x_1, \ldots, x_n$ are elements in $\mathrm{Lip}(K)$, then

$$\mathbb{E}\bigg(\bigg\|\sum_{i=1}^{n}\varepsilon_i j(x_i)\bigg\|_\infty^2\bigg) \leq C\sum_{i=1}^{n}\|x_i\|_{\mathrm{Lip}}^2.$$

Mimicking the proof of Theorem 6, we get that

$$\frac{1}{\sqrt{n}}\,\mathbb{E}\bigg(\bigg\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} j(X_i)\bigg\|_\infty^2\bigg) \leq C\mathbb{E}\big(\|X\|_{\mathrm{Lip}}^2\big). \tag{11}$$

One then concludes as in Theorem 6. There is however a little difficulty due to the fact that $\mathrm{Lip}(K)$ need not be separable. To handle this problem, show that, for every $\varepsilon > 0$, there exists a finite dimensional subspace F of $C(K)$ such that if Q_F is the quotient map $C(K) \to C(K)/F$, the type 2 constant of the operator $j \circ Q_F$ is less than ε. Applying (11) to $j \circ Q_F$ then easily yields the result. □

With some further work, the integrability condition on M may be weakened into $\lim_{t\to\infty} t^2\mathbb{P}\{M \geq t\} = 0$, provided it

is assumed in addition that there is a continuous Gaussian process with the same covariance structure as X.

To some extent, the preceding analysis of the classical central limit theorem with Gaussian limits may be developed similarly in case of the general central limit theorem with infinitely divisible limits

[A-A-G]. Under the assumptions of tightness of the partial sum sequence, limits are identified more or less as in the scalar case. Type or cotype assumptions are sufficient or necessary for tightness. In particular, stable limits may be characterized through stable type. We refer to the book [A-G] for an account on the general central limit theorem.

There are several other concepts and results from Banach space theory which were crucial to the investigation of stable limit theorems and the CLT for triangular arrays. In particular, Banach spaces in which ℓ_∞ is not finitely representable [Ma-P] (of finite cotype) may be used to investigate Poissonnization and the accompanying laws in infinite dimension [A-G-M-Z]. Banach spaces B not containing c_0 were characterized by J. Hoffmann-Jørgensen [HJ2] and S. Kwapień [Kw2] as those in which almost sure bounded partial sum sequences (S_n) of independent symmetric B-valued random variables are almost surely convergent. There was a nice interplay between the Banach space and probability communities around stable type. In [M-P] a representation theorem for stable random variables and vectors (obtained in [L-W-Z]) was extended and used to analyze stable random Fourier series. Let X be a p-stable $(0 < p < 2)$ random vector with values in B with spectral measure m symmetrically distributed on the unit sphere of B (cf. [L-T3, Chapter 5]). Denote by $(Y_j)_{j\in\mathbb{N}}$ independent random variables distributed as $m/|m|$. Let furthermore $,\ _j = \lambda_1 + \cdots + \lambda_j$, $j \geq 1$, where the λ_i's are independent standard exponential random variables and assume the sequences $(Y_j)_{j\in\mathbb{N}}$ and $(,\ _j)_{j\in\mathbb{N}}$ to be independent. Then, the series

$$\sum_{j=1}^{\infty} ,\ _j^{-1/p} Y_j$$

converges almost surely and is distributed as $c_p|m|^{-1/p}X$ for some $c_p > 0$ only depending on p. This representation later allowed G. Pisier in [Pi3] to obtain an extension

of results of W. B. Johnson and G. Schechtman [J-S] on spaces in which ℓ_p, $1 \leq p < 2$, is not finitely representable. Since $,\ _j \sim j$ by the LLN, Pisier's idea was to use this representation together with Yurinskii's inequality to produce a proof similar to the Gaussian proof of Dvoretzky's theorem based on (5) (cf. [Pi4]).

4. Bootstrap and empirical processes

Exponential bounds on empirical processes have been proved extremely useful very recently in the selection of models in statistics (cf. [B-M]). To describe one inequality more precisely, let, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $(X_i)_{i\in\mathbb{N}}$ be independent indentically distributed random variables with values in some measurable space $(S, \mathcal{S})$ and with common law P. For every $n \geq 1$, consider the empirical measures

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

Now, let $\mathcal{F}$ be a class of measurable functions on S with real values. The theory of empirical processes runs into various measurability questions in which we do not want to enter here. So let us assume for simplicity the class $\mathcal{F}$ to be countable.

The statistical treatment of empirical measures shows that the unknown law P can be recovered from the observations P_n on some class $\mathcal{F}$, the larger the class, the more accurate the result. In particular, a class $\mathcal{F}$ is said to be a Glivenko-Cantelli class if

$$\sup_{\mathcal{F}} \big| P_n(f) - P(f) \big| \to 0$$

with probability 1. $\mathcal{F}$ is said to be a Donsker class if the sequence

$$\sqrt{n}\big(P_n(f) - P(f)\big), \quad f \in \mathcal{F},$$

converges in distribution (in a sense to be made precise) to a centered Gaussian process G_P indexed by $\mathcal{F}$ with covariance $P(fg) - P(f)P(g)$, $f, g \in \mathcal{F}$. These definitions extend the classical results of Glivenko-Cantelli and Donsker for the class of the characteristic functions of the intervals $(-\infty, x]$, $x \in \mathbb{R}$. In

statistical applications however, one is interested in estimates at finite

range, that is on P_n for fixed n. To this end, the exponential bound (7) is of fundamental importance. Assume that $|f| \leq C$ for every $f \in \mathcal{F}$ and set

$$Z_n = \sup_{f \in \mathcal{F}} \big| P_n(f) - P(f) \big|.$$

Then, for every $t \geq 0$,

$$\mathbb{P}\big\{ Z_n - \mathbb{E}(Z_n) \geq t \big\} \leq K \exp\bigg(-\frac{nt}{KC} \log\bigg(1 + \frac{Ct}{\sigma^2 + C\mathbb{E}(Z_n)} \bigg) \bigg) \tag{12}$$

where $\sigma^2 = \sup_{f \in \mathcal{F}} (P(f^2) - P(f)^2)$ (cf. [Ta4], [Le3]). A similar inequality also holds for $\mathbb{E}(Z_n) - Z_n$ which thus yields a concentration property. It is very important in statistical applications that (12) holds with $\mathbb{E}(Z_n)$ and not a multiple of $\mathbb{E}(Z_n)$ as was the case in the earlier bounds [L-T3].

A special class $\mathcal{F}$ is given by the example of the family of characteristic functions of Vapnik-Cervonenkis classes of sets. Let S be a set and $\mathcal{C}$ be a class of subsets of S. Let A be a subset of S of cardinality k. Say that $\mathcal{C}$ shatters A if each subset of A is the trace of an element of $\mathcal{C}$. $\mathcal{C}$ is said to be a Vapnik-Cervonenkis class (VC class in short) if there is an integer $k \geq 1$ such that no subset A of S of cardinality k is shattered by $\mathcal{C}$. Denote by $v(\mathcal{C})$ the smallest k with this property. The class of all interval $(-\infty, x]$, $x \in \mathbb{R}$, is a VC class with $v(\mathcal{C}) = 2$. The most striking fact about VC classes is that whenever $\mathcal{C}$ is a VC class in S and $v = v(\mathcal{C})$, any subset A of S with $\mathrm{Card}(A) = n \geq v$ satisfies

$$\mathrm{Card}(\mathcal{C} \cap A) \leq \Big(\frac{\mathrm{e}n}{v} \Big)^v. \tag{13}$$

Let now Q be a probability measure on a measurable space $(S, \mathcal{S})$ and consider a VC class $\mathcal{C}$ of subsets of S. For any measurable subsets A, B in $\mathcal{S}$, set

$$d_Q(A, B) = \| I_A - I_B \|_2$$

(where the norm is understood with respect to Q). For any $\varepsilon > 0$, let $N(\varepsilon) = N(\mathcal{C}, d_Q; \varepsilon)$ be the minimal number of balls of radius ε in the metric d_Q which are necessary to cover $\mathcal{C}$. As a consequence of (13), it was shown by R. M. Dudley [Du] that the growth of $N(\varepsilon)$ as ε goes to zero is controlled by $v(\mathcal{C})$. Indeed, for any $\varepsilon > 0$,

$$\log N(\varepsilon) \leq K v(\mathcal{C}) \Big(1 + \log \frac{1}{\varepsilon} \Big). \tag{14}$$

¿From this result, it is not so difficult to deduce that $\mathcal{C}$ is a Donsker class for every probability measure P on $(S, \mathcal{S})$. Rather than to directly prove such a property, let us relate, as in Section

3, the nice limit properties of VC classes to the type 2 property of a certain operator between Banach spaces. Denote by $M = M(S, \mathcal{S})$ the Banach space of all bounded measures μ on $(S, \mathcal{S})$ equipped with the norm $\|\mu\| = |\mu|(S)$. Consider the operator $j : M \to \ell_\infty(\mathcal{C})$ defined by $j(\mu) = (\mu(C))_{C \in \mathcal{C}}$. Denote by $T_2(j)$ the type 2 constant of j, that is the smallest constant C such that for all $\mu_1, \ldots, \mu_n$ in M,

$$\mathbb{E}\biggl(\biggl\|\sum_{i=1}^n \varepsilon_i j(\mu_i)\biggr\|_{\mathcal{C}}^2\biggr) \le C \sum_{i=1}^n \|\mu_i\|^2.$$

The next theorem has been observed by G. Pisier [Pi2].

Theorem 10. *For some numerical constant $K > 0$,*

$$K^{-1}\sqrt{v(\mathcal{C})} \le T_2(j) \le K\sqrt{v(\mathcal{C})}\,.$$

Proof. We only show the right-hand side inequality. Let $\mu_1, \ldots, \mu_n$ in M. To prove the type 2 inequality, we may assume that the measures μ_i are positive and, by homogeneity, that $\sum_{i=1}^n \|\mu_i\|^2 = 1$. Set $Q = \sum_{i=1}^n \|\mu_i\|\mu_i$. Then Q is a probability measure on $(S, \mathcal{S})$, and we clearly have that

$$\biggl(\sum_{i=1}^n \bigl|\mu_i(A) - \mu_i(B)\bigr|^2\biggr)^{1/2} \le d_Q(A, B)$$

for all subsets A and B. A classical entropic bound on the Rademacher process $\sum_{i=1}^n \varepsilon_i \mu_i(C)$, $C \in \mathcal{C}$ (cf. [L-T3, Chapter 11]), then shows that

$$\begin{aligned}\mathbb{E}\biggl(\biggl\|\sum_{i=1}^n \varepsilon_i \mu_i\biggr\|_{\mathcal{C}}^2\biggr) &= \mathbb{E}\biggl(\sup_{C \in \mathcal{C}}\biggl|\sum_{i=1}^n \varepsilon_i \mu_i(C)\biggr|^2\biggr) \\ &\le 1 + K\int_0^\infty \bigl(\log N(\varepsilon)\bigr)^{1/2} d\varepsilon \le 1 + K'\sqrt{v(\mathcal{C})}\end{aligned}$$

where we used (14) in the last step. Since $v(\mathcal{C}) \ge 1$, the claim is established. □

One can then deduce that VC classes are uniformly Donsker by the type theory of Section 3. Further and refined limit theorems for empirical processes have been developed, especially in [Du] and [G-Z2], under various conditions (entropy with bracketing, random entropy etc).

The following is the Banach space formulation of Theorem 10.

Theorem 11. *Let $x_1, \ldots, x_n$ be functions on some set T taking the values ± 1. Let*

$$r(T) = \mathbb{E}\biggl(\sup_{t \in T}\biggl|\sum_{i=1}^n \varepsilon_i x_i(t)\biggr|\biggr).$$

There exists a numerical constant $K > 0$ such that for every $k \le r(T)^2/Kn$, one can find $m_1 < m_2 < \cdots < m_k$ in $\{1, \ldots, n\}$ such that the set of values $\{x_{m_1}(t), \ldots, x_{m_k}(t)\}$, $t \in T$, is exactly $\{-1, +1\}^k$. In other words, the subsequence $\{x_{m_1}, \ldots, x_{m_k}\}$ generates a subspace isometric to ℓ_1^k in $\ell_\infty(T)$.

For the proof, apply Theorem 10 to the class $\mathcal{C}$ of subsets of $\{1, 2, \ldots, n\}$ of the form

$$\big\{i \in \{1, \ldots, n\}; x_i(t) = 1\big\}, \quad t \in T.$$

The functional calculus of probability has also been proved useful in other statistical problems, and, in the final part of this exposition, we present applications to the bootstrap in statistics.

The "bootstrap", introduced by B. Efron in 1979, is a resampling method for approximating the distribution of various statistics. For every $\omega \in \Omega$, let $\hat{X}^\omega_{ni}$, $i = 1, \ldots, n$, be independent and identically distributed random variables with common distribution $P_n(\omega)$. Denote, for each $n \ge 1$, $\hat{P}_n(\omega)$ the empirical distribution based on $\hat{X}^\omega_{ni}$, $i = 1, \ldots, n$, that is

$$\hat{P}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{\hat{X}^\omega_{ni}}.$$

It is then expected that the distribution of statistics $\hat{H}_n(\omega) = H_n\big(\hat{X}^\omega_{n1}, \ldots, \hat{X}^\omega_{nn}; P_n(\omega)\big)$ is ω-almost surely asymptotically close to that of $H_n(X_1, \ldots, X_n; P)$. This suggestive method has been validated with limit theorems for many particular statistics H_n, improving rates with respect to the classical central limit theorem approximations. Let $F = \sup_{f \in \mathcal{F}} |f|$. The following has been obtained in [G-Z3].

Theorem 12. *The conditions $\int F^2 dP < \infty$ and P is a Donsker class are necessary and sufficient for*

$$\sqrt{n}\,\big(\hat{P}_n(\omega) - P_n(\omega)\big) \to G \quad \textit{weakly}$$

ω-almost surely for a centered Gaussian process G independent of ω. Furthermore $G = G_P$.

This theorem completely settles (modulo measurability) the question of the validity of the bootstrap for the CLT for empirical processes. The proof of Theorem 12 is based on an almost sure version of the CLT which we state in the setting of Banach space valued random variables for simplicity. In the statement, X is a Banach space valued random variable and g a standard normal random variable independent of X. Accordingly, if $(X_i)_{i \in \mathbb{N}}$, resp. $(g_i)_{i \in \mathbb{N}}$, is a sequence of independent copies of X, resp. g, the two sequences are understood to be independent (i.e. constructed on different probability spaces). The following result is due to M. Talagrand and the authors [L-T2].

Theorem 13. *Let X be a random variable with values in a Banach space B*

and let g be a standard normal random variable independent of X. The following are equivalent.

(i) $\mathbb{E}(\|X\|^2) < \infty$ *and X satisifies the CLT with limiting Gaussian distribution G.*

(ii) For almost every ω of the probability space supporting the X_i's, the sequence $\left(\sum_{i=1}^n g_i X_i(\omega)/\sqrt{n}\right)_{n\geq 1}$ *converges in distribution.*

In either case, the limit of $\left(\sum_{i=1}^n g_i X_i(\omega)/\sqrt{n}\right)_{n\geq 1}$ *does not depend on ω and is distributed as G.*

Proof. We concentrate on the implication $(i) \Rightarrow (ii)$. It is not difficult to see that it is enough to establish the following lemma.

Lemma 14. *If* $\mathbb{E}(\|X\|^2) < \infty$, *almost surely*

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}_g\left(\left\| \sum_{i=1}^n g_i X_i \right\|\right) \leq 4 \limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}\left(\left\| \sum_{i=1}^n g_i X_i \right\|\right)$$

where we denote by $\mathbb{E}_g$ partial integration with respect to the sequence $(g_i)_{i\in\mathbb{N}}$.

Proof. We only outline one argument. Let

$$M = \limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}\left(\left\| \sum_{i=1}^n g_i X_i \right\|\right)$$

assumed to be finite. By the Borel-Cantelli lemma, it suffices to show that for every $\varepsilon > 0$,

$$\sum_n \mathbb{P}\left\{ \sup_{2^{n-1}<k\leq 2^n} \frac{1}{\sqrt{k}} \mathbb{E}_g\left(\left\| \sum_{i=1}^k g_i X_i \right\|\right) \geq 2(2M+5\varepsilon) \right\} < \infty.$$

By Lévy's inequalities for symmetric random vectors, and a simple truncation argument under the hypothesis $\mathbb{E}(\|X\|^2) < \infty$, it is enough to show that

$$\sum_n \mathbb{P}\left\{ \mathbb{E}_g\left(\left\| \sum_{i=1}^{2^n} g_i U_i \right\|\right) \geq 2M+5\varepsilon \right\} < \infty$$

where, for every $i = 1, \ldots, 2^n$,

$$U_i = U_i(n) = 2^{-n/2} X_i I_{\{\|X_i\|\leq \varepsilon\sqrt{2^n}\}}.$$

By Hoffmann-Jørgensen's inequality (5), for every n,

$$\mathbb{P}\left\{ \mathbb{E}_g\left(\left\| \sum_{i=1}^{2^n} g_i U_i \right\|\right) \geq 2M+5\varepsilon \right\} \leq \left(\mathbb{P}\left\{ \mathbb{E}_g\left(\left\| \sum_{i=1}^{2^n} g_i U_i \right\|\right) \geq M+2\varepsilon \right\} \right)^2,$$

so that, by definition of M, it finally suffices to show that

$$\sum_n \left(\mathbb{P}\left\{\mathbb{E}_g\left(\left\|\sum_{i=1}^{2^n} g_i U_i\right\|\right) - \mathbb{E}\left(\left\|\sum_{i=1}^{2^n} g_i U_i\right\|\right) \geq \varepsilon\right\}\right)^2 < \infty.$$

By Yurinskii's martingale representation and Chebyshev's inequality,

$$\mathbb{P}\left\{\mathbb{E}_g\left(\left\|\sum_{i=1}^{2^n} g_i U_i\right\|\right) - \mathbb{E}\left(\left\|\sum_{i=1}^{2^n} g_i U_i\right\|\right) \geq \varepsilon\right\} \leq \frac{4}{\varepsilon^2}\sum_{i=1}^{2^n} \mathbb{E}(d_i^2).$$

Now,

$$d_i = \mathbb{E}(f_i \,|\, \mathcal{F}_i) - \mathbb{E}(f_i \,|\, \mathcal{F}_{i-1})$$

where

$$f_i = \mathbb{E}_g\left(\left\|\sum_{j=1}^{2^n} g_j U_j\right\|\right) - \mathbb{E}_g\left(\left\|\sum_{j=1, j\neq i}^{2^n} g_j U_j\right\|\right),$$

$i = 1, \ldots, 2^n$. For every i, $0 \leq f_i \leq \|g_i U_i\|$. Moreover, by identical distribution,

$$\mathbb{E}(f_i) \leq \frac{1}{2^n}\,\mathbb{E}\left(\left\|\sum_{j=1}^{2^n} g_j U_j\right\|\right) \leq \frac{M+\varepsilon}{2^n}$$

for every n large enough. Therefore,

$$\begin{aligned}
\mathbb{E}(d_i^2) &\leq \mathbb{E}(f_i^2) \\
&\leq \mathbb{E}\big(\|g_i U_i\|^{3/2} f_i^{1/2}\big) \\
&\leq \big(\mathbb{E}(\|g_i U_i\|^3)\big)^{1/2}\big(\mathbb{E}(f_i)\big)^{1/2} \leq \left(\frac{M+\varepsilon}{2^n}\,\mathbb{E}\big(\|g_i U_i\|^3\big)\right)^{1/2}.
\end{aligned}$$

One is thus left to show that the series

$$\sum_n 2^{-n/2}\mathbb{E}\big(\|X\|^3 I_{\{\|X\| \leq \varepsilon\sqrt{2^n}\}}\big)$$

is summable which is an easy consequence of the second moment assumption. □

We briefly conclude the proof of Theorem 13. Since X satisfies the CLT, gX also satisfies the CLT. Choose then, for every $k \geq 1$, a finite dimensional subspace F_k of B such that if $T_k = T_{F_k} : B \to B/F_k$ is the quotient map,

$$\sup_{n\geq 1} \frac{1}{\sqrt{n}}\,\mathbb{E}\left(\left\|\sum_{i=1}^{n} g_i T_k(X_i)\right\|\right) \leq \frac{1}{k}\,.$$

Apply Lemma 14 to $T_k(X)$ for every k. There exists Ω_k with $\mathbb{P}(\Omega_k) = 1$ such that for every $\omega \in \Omega_k$,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}_g \left(\left\| \sum_{i=1}^{n} g_i T_k\big(X_i(\omega)\big) \right\| \right) \leq \frac{4}{k} \,.$$

Let also Ω_0 be the set of full probability obtained when Lemma 14 is applied to X itself. Let $\Omega^0 = \bigcap_{k\geq 0} \Omega_k$. If $\omega \in \Omega^0$, for each $\varepsilon > 0$, there exists a finite dimensional subspace F of B such that if $T = T_F$ is the quotient map,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}_g \left(\left\| \sum_{i=1}^{n} g_i T\big(X_i(\omega)\big) \right\| \right) \leq \varepsilon^2 \,.$$

Hence, if $n \geq n(\varepsilon)$,

$$\mathbb{P}_g \left\{ \left\| T \left(\sum_{i=1}^{n} g_i X_i(\omega)/\sqrt{n} \right) \right\| \geq \varepsilon \right\} \leq \varepsilon.$$

It follows that the sequence $\sum_{i=1}^{n} g_i X_i(\omega)/\sqrt{n}$, $n \geq 1$, is tight. The proof is easily completed by identifying the limit. □

References

[A-A-G] A. de Acosta, A. Araujo, E. Giné. On Poisson measures, Gaussian measures and the central limit theorem in Banach spaces. Advances in Probability 4, 1–68 (1978). Dekker.

[A-G] A. Araujo, E. Giné. The central limit theorem for real and Banach space valued random variables. Wiley (1980).

[A-G-M-Z] A. Araujo, E. Giné, V. Mandrekar, J. Zinn. On the accompanying laws in Banach spaces. Ann. Probability 9, 202–210 (1981).

[Be] A. Beck. A convexity condition in Banach spaces and the strong law of large numbers. Proc. Amer. Math. Soc. 13, 329–334 (1962).

[B-M] L. Birgé, P. Massart. Book to appear.

[Du] R. M. Dudley. Central limit theorems for empirical measures. Ann. Probability 6, 899–929 (1978).

[D-S] R. M. Dudley, V. Strassen. The central limit theorem and ε-entropy. Probability and Information Theory. Lecture Notes in Mathematics, vol. 89, 224–231. Springer 1969.

[G-Z1] E. Giné, J. Zinn. Central limit theorems and weak laws of large numbers in certain Banach spaces. Z. Wahrscheinlichkeitstheorie verw. Gebiete 62, 323–354 (1983).

[G-Z2] E. Giné, J. Zinn. Some limit theorems for empirical processes. Ann. Probability 12, 929–989 (1984).

[G-Z3] E. Giné, J. Zinn. Bootstrapping general empirical measures. Ann. Probability 18, 851–869 (1990).

[G-K-Z] V. Goodman, J. Kuelbs, J. Zinn. Some results on the LIL in Banach space with applications to weighted empirical processes. Ann. Probability 9, 713–752 (1981).

[HJ1] J. Hoffmann-Jørgensen. Sums of independent Banach space valued random variables. Aarhus Univ. Preprint Series 1972/73, no. 15 (1973).

[HJ2] J. Hoffmann-Jørgensen. Sums of independent Banach space valued random variables. Studia Math. 52, 159–186 (1974).

[HJ3] J. Hoffmann-Jørgensen. Probability in Banach spaces. Ecole d'Eté de Probabilités de St-Flour 1976. Lecture Notes in Mathematics, vol. 598, 1–186. Springer 1976.

[HJ-P] J. Hoffmann-Jørgensen, G. Pisier. The law of large numbers and the central limit theorem in Banach spaces. Ann. Probability 4, 587–599 (1976).

[J-M] N. C. Jain and M. B. Marcus. Central limit theorem for $C(S)$-valued random variables. J. Funct. Anal. 19, 216–231 (1975).

[J-S] W. B. Johnson and G. Schechtman. Embedding ℓ_p^m into ℓ_p^1 Acta Math. 149, 71–85 (1982).

[Ku1] J. Kuelbs. A strong convergence theorem for Banach space valued random variables. Ann. Probability 4, 744–771 (1976).

[Ku2] J. Kuelbs. Kolmogorov's law of the iterated logarithm for Banach space valued random variables. Illinois J. Math. 21, 784–800 (1977).

[K-Z] J. Kuelbs, J. Zinn. Some stability results for vector valued random variables. Ann. Probability 7, 75–84 (1979).

[Kw1] S. Kwapień. Isomorphic characterization of inner product spaces by orthogonal series with vector valued coefficients. Studia Math. 44, 583–595 (1972).

[Kw2] S. Kwapień. On Banach spaces containing c_0. Studia Math. 52, 187–190 (1974).

[Le1] M. Ledoux. Sur une inégalité de H. P. Rosenthal et le théorème limite central dans les espaces de Banach. Israel J. Math. 50, 290–318 (1985).

[Le2] M. Ledoux. The law of the iterated logarithm in uniformly convex Banach spaces. Trans. Amer. Math. Soc. 294, 351–365 (1986).

[Le3] M. Ledoux. On Talagrand's deviation inequalities for product measures. ESAIM Prob. & Stat. 1, 63–87 (1996).

[L-T1] M. Ledoux, M. Talagrand. Characterization of the law of the iterated logarithm in Banach spaces. Ann. Probability 16, 1242–1264 (1988).

[L-T2] M. Ledoux, M. Talagrand. Un critère sur les petites boules dans le théorème limite central. Probab. Theor. Rel. Fields 77, 29–47 (1988).

[L-T3] M. Ledoux, M. Talagrand. Probability in Banach spaces. Springer 1991.

[L-W-Z] R. LePage, M. Woodroofe, J. Zinn. Convergence to a stable distribution via order statistics. Ann. Probability 9, 624–632 (1981).

[M-P] M. B. Marcus and G. Pisier. Characterizations of almost sureely continuous p-stable random Fourier series and strongly stationary processes. Acta. Math. 152, 245–301 (1984).

[Ma] B. Maurey. Espaces de cotype p. Séminaire Maurey-Schwartz 1972-73. Ecole Polytechnique, Paris (1973).

[Ma-P] B. Maurey and G. Pisier. Séries de variables aléatoires vectorielles indépendantes et géométrie des espaces de Banach. Studia Math. 58, 45–90 (1976).

[Pi1] G. Pisier. Sur les espaces qui ne contiennent pas de ℓ_n^1 uniformément. Séminaire Maurey-Schartz 1973-74. Ecole Polytechnique, Paris (1974).

[Pi2] G. Pisier. Remarques sur les classes de Vapnik-Cervonenkis. Ann. Inst. H. Poincaré 20, 287–298 (1984).

[Pi3] G. Pisier. On the dimension of the ℓ_p^n-subspaces of Banach spaces, for $1 \leq p < 2$. Trans. Amer. Math. Soc. 276, 201–211 (1983).

[Pi4] G. Pisier. Probabilistic methods in the geometry of Banach spaces. Probability and Analysis, Varenna (Italy) 1985. Lecture Notes in Mathematics, vol. 1206, 167–241. Springer 1986.

[P-Z] G. Pisier, J. Zinn. On the limit theorems for random variables with values in the spaces L_p $(2 \leq p < \infty)$. Z. Wahrscheinlichkeitstheorie verw. Gebiete 41, 289–304 (1978).

[Ta1] M. Talagrand. Isoperimetry and integrability of the sumof independent Banach space valued random variables. Ann. Probability 17, 1546–1570 (1989).

[Ta2] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l'I.H.E.S., 81, 73–205 (1995).

[Ta3] M. Talagrand. A new look at independence (1994). Ann. Probability, 24, 1–34 (1996).

[Ta4] M. Talagrand. New concentration inequalities in product spaces (1995). Invent. math. 126, 505–563 (1996).

[Va] S. R. S. Varadhan. Limit theorems for sums of independent random variables with values in a Hilbert space. Sankhÿa A24, 213–238 (1962).

[Yu1] V. V. Yurinskii. Exponential bounds for large deviations. Theor. Probability Appl. 19, 154–155 (1974).

[Yu2] V. V. Yurinskii. Exponential integrability for sums of random vectors. J. Multivariate Anal.

6, 476–499 (1976).

[Zi] J. Zinn. A note on the central limit theorem in Banach spaces. Ann. Probab. 5, 283–286 (1977).

M. Ledoux

Département de Mathématiques

Université Paul-Sabatier

31062, Toulouse France

ledoux@cict.fr

J. Zinn

Department of Mathematics

Texas A & M University

77843 Texas U.S.A.

Joel.Zinn@math.tamu.edu