The Annals of Probability
1989, Vol. 17, No. 2, 596–631

COMPARISON THEOREMS, RANDOM GEOMETRY AND SOME LIMIT THEOREMS FOR EMPIRICAL PROCESSES

BY M. LEDOUX AND M. TALAGRAND[1]

Université de Strasbourg and Université de Paris VI and The Ohio State University

In this paper, we obtain several new results and developments in the study of empirical processes. A comparison theorem for Rademacher averages is at the basis of the first part of the results, with applications, in particular, to Kolmogorov's law of the iterated logarithm and Prokhorov's law of large numbers for empirical processes. We then study the behavior of empirical processes along a class of functions through random geometric conditions and complete in this way the characterization of the law of the iterated logarithm. Bracketing and local Lipschitz conditions provide illustrations of some of these ideas to concrete situations.

1. Introduction. Let $(S, \mathscr{S}, P)$ be a probability space. $(\Omega, \Sigma, \Pr)$ will usually denote the product of the infinite product probability space $(S^{\mathbb{N}}, \mathscr{S}^{\mathbb{N}}, P^{\mathbb{N}})$ with a rich enough probability space, supporting in particular a Bernoulli or Rademacher sequence $(\varepsilon_i)_{i \in \mathbb{N}}$ and an orthogaussian sequence $(g_i)_{i \in \mathbb{N}}$ as well as others to be specified when they appear. We consider the coordinate functions X_i, $i \in \mathbb{N}$, which are the projections of Ω onto the ith copy of S. Integration with respect to P or Pr is denoted usually by E, but, when necessary to distinguish, probabilities and expectations with respect to the sequences $(X_i), (\varepsilon_i), (g_i)$, etc. are denoted by $P_X = P$, P_ε, P_g, etc. and E_X, E_ε, E_g, etc. The empirical measures P_n on Ω are the random measures on S,

$$P_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)}, \qquad \omega \in \Omega,\ n \in \mathbb{N}.$$

Set $L_p = L_p(S, \mathscr{S}, P)$, $0 \le p \le \infty$, and denote by $\|\cdot\|_p$ its norm ($1 \le p \le \infty$), by d_p the associated distance and by W_p the unit ball. We also need to consider the *random* spaces $L_1(P_n)$ and $L_2(P_n)$ with their norms

$$\|f\|_{n,1} = \frac{1}{n} \sum_{i=1}^{n} |f(X_i)|, \qquad \|f\|_{n,2} = \left(\frac{1}{n} \sum_{i=1}^{n} |f(X_i)|^2 \right)^{1/2},$$

f a function on S, and corresponding distances $d_{n,1}$, $d_{n,2}$.

By class of functions, we will always mean a family $\mathscr{F}$ of (real) measurable functions f on $(S, \mathscr{S})$ such that $\mathscr{F} \subset L_0$ and $\|f(s)\|_{\mathscr{F}} < \infty$ for all s in S. [For any family $(a_f)_{f \in \mathscr{F}}$ of numbers indexed by a class or set $\mathscr{F}$, we let $\|a_f\|_{\mathscr{F}} =$

Received December 1986; revised March 1988.

[1]Work partially supported by an NSF grant.

AMS 1980 *subject classifications*. 60F17, 60B12, 60F05.

Key words and phrases. Empirical processes, comparison theorems, random geometry, limit theorems.

$\sup_{f \in \mathscr{F}} |a_f|$.] The (centered) empirical process on $\mathscr{F}$ is defined as

$$(P_n - P)(f) = \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Ef), \qquad f \in \mathscr{F},\ n \in \mathbb{N}.$$

The study of empirical processes runs into various measurability questions; since we do not want to be concerned with them here, we shall assume all classes $\mathscr{F}$ to be *countable* [we could instead require the separability of the processes $((P_n - P)(f))_{f \in \mathscr{F}}$].

The recent papers by Giné and Zinn [10, 11] have opened new perspectives in the study of limit theorems for empirical processes indexed by families of functions, in particular the central limit theorem (in short CLT), a subject initiated by Dudley [6, 7]. One of their basic tools is "randomization" which allows us to replace $P_n - P$ by $(1/n)\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}$ and $(1/n)\sum_{i=1}^{n} g_i \delta_{X_i}$ and to make use, conditionally, of the results and arguments of the theory of Gaussian and subgaussian processes. The recent characterization of pregaussian classes obtained by Talagrand [22] of course also plays an important role in this context. Their results and ideas lead to complete random geometric pictures of the CLT for bounded and unbounded classes of functions as illustrated in the papers [10, 11, 21].

Randomization and description of pregaussian classes are the main features of these results. However, the applicability of these powerful techniques to different limit theorems and more general situations seems to require overcoming several new difficulties. For example, the Gaussian randomization is, in general, too heavy in the study of empirical processes based on independent but not necessarily identically distributed random variables. This can easily be seen from the example of the class $\mathscr{F} = \{f_i,\ i \in \mathbb{N}\}$ of functions on $\mathbb{N}$ defined by $f_i(j) = \delta_{ij}$ and the constant random variables $X_i = i$, $i \in \mathbb{N}$. Concerning "pregaussianness," it is, in general also, only necessary for the CLT, and the study of other limit theorems, for example, the law of the iterated logarithm (in short LIL) that will be our reference example in the sequel, cannot be directly deduced from what is known for the CLT. The main objective of the present paper is to develop, with applications, some devices in these directions. We will thus mainly show how the Gaussian randomization can efficiently be replaced to allow the investigation of non-equidistributed situations and how it is possible to make use of a pregaussian structure, even if not explicitly given, to describe empirical processes on classes of functions.

Section 2 is devoted to a comparison theorem for Rademacher averages when coordinates are contracted. It can be used, in these situations, like the Gaussian comparison theorems based on Slepian's lemma but avoids the Gaussian randomization when this one is inefficient, like the non-equidistributed case. The main use of this result is provided in the third section which is devoted to an extension of Kolmogorov's LIL for empirical processes. We present at the end of Section 2 two applications in the form of simplified and more transparent proofs of known results: a weak law of large numbers for squares [11] and a CLT under random entropy conditions for uniformly bounded functions [11, 21].

The LIL is studied in the third part of this work. We first recall, in the context of empirical processes, the recent characterization [16] of classes satisfying the LIL that reduces the almost sure statement of the LIL to a weak statement, namely the convergence in probability to 0 of $\|(P_n - P)(f)\sqrt{n/2LLn}\|_{\mathscr{F}}$. The question of how to describe and control this weak statement will be addressed in the sequel. For the moment, we investigate the LIL in the non-equidistributed case. One of the main tools involved in this study is a powerful new isoperimetric inequality for product measures recently obtained in [20]. Together with the preceding comparison theorem and an observation of Pisier [18] concerning randomization by uniformly distributed random variables, we establish a complete extension of Kolmogorov's LIL for empirical processes. The same idea also leads to a generalization of Prokhorov's strong law of large numbers in this context.

A class $\mathscr{F}$ in L_2 is said to be (P) pregaussian if the Gaussian process $\{G_P(f); f \in \mathscr{F}\}$ on L_2 with covariance

$$EG_P(f)G_P(g) = E(f - Ef)(g - Eg), \qquad f, g \in L_2,$$

has a version with all the sample functions bounded and uniformly continuous for its L_2 metric (or, equivalently, the L_2 distance d_2 whenever $\|Ef\|_{\mathscr{F}} < \infty$). Pregaussian classes have been characterized in [22] by means of majorizing measures, thus providing an efficient tool in their study. Since classes satisfying the CLT are pregaussian, this necessary condition and its description are of considerable importance in the study of the central limit property. For example, it is shown in [21] how, up to some $L_1(P_n)$ perturbation, the pregaussian character controls the empirical process: More precisely, each pregaussian class $\mathscr{F}$ can be regarded, for each n, as a subset of the direct sum $\mathscr{F}_1 + \mathscr{F}_2 = \{f_1 + f_2; f_1 \in \mathscr{F}_1, f_2 \in \mathscr{F}_2\}$ of two classes $\mathscr{F}_1$ and $\mathscr{F}_2$ (depending on n) such that, with large probability, $\mathscr{F}_2$ equipped with the random distance $d_{n,2}$ is controlled by the pregaussian character of $\mathscr{F}$ in such a way that

$$E_\varepsilon \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n} \right\|_{\mathscr{F}_2} \le kE\|G_P(f)\|_{\mathscr{F}}$$

and concerning $\mathscr{F}_1$, also with large probability,

$$\left\| \sum_{i=1}^{n} |f(X_i)| \right\|_{\mathscr{F}_1} \le kE \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathscr{F}}.$$

Here k depends on the "large" probability. The construction uses the description of pregaussian classes together with Bernstein's inequality to build and control $\mathscr{F}_2$ in the $L_2(P_n)$ norm. The result reduces, in a sense, the study of CLT classes to the behavior of expressions of the type $\|\sum_{i=1}^{n} |f(X_i)|\|_{\mathscr{F}}$ [the class $\mathscr{F}_1$ above, that we called $L_1(P_n)$ perturbation] in which cancellation, one of the main features of sums of independent random variables, does not occur.

As already mentioned, "pregaussianness" does not appear, in general, as a necessary condition for limit theorems other than the CLT, like, for example, the convergence in probability to 0 of $\|(P_n - P)(f)\sqrt{n/2LLn}\|_{\mathscr{F}}$ in the LIL. Our aim will thus be to look for a pregaussian structure in a more general context that could possibly allow the use of the preceding decomposition result in the case of the CLT. The main application we have in mind concerns the LIL which is reduced, as mentioned before, to the behavior in probability of $\|(P_n - P)(f)\sqrt{n/2LLn}\|_{\mathscr{F}}$. In this case, our result reads as follows. Let n be a fixed integer, $\mathscr{F}$ a totally bounded class in L_2. If

$$u \geq E\left\| \sum_{i=1}^{n} g_i f(X_i)/\sqrt{2nLLn} \right\|_{\mathscr{F}},$$

there exist classes $\mathscr{F}_1$, $\mathscr{F}_2$ in L_2 (depending upon n) such that $\mathscr{F} \subset \mathscr{F}_1 + \mathscr{F}_2$ and such that, with large probability,

$$\left\| \sum_{i=1}^{n} |f(X_i)|/\sqrt{2nLLn} \right\|_{\mathscr{F}_1} \leq ku$$

and

$$E_g\left\| \sum_{i=1}^{n} g_i f(X_i)/\sqrt{2nLLn} \right\|_{\mathscr{F}_2} \leq ku.$$

Since the control of $\mathscr{F}_2$ is provided by Gaussian tools, this decomposition theorem reduces, from a theoretical point of view, as before for the CLT, to the study of the LIL for classes under the $L_1(P_n)$ norm (the class $\mathscr{F}_1$ above). The basic idea of the proof tries to find, up to some small $L_1(P_n)$ perturbation, a controlled pregaussian class not too far from the original one (Theorem 14). The argument, which uses, as before, Bernstein's inequality at some crucial point, makes close use of ideas developed in recent papers [22] and [23] on the regularity of Gaussian and stable processes. In the presence of this pregaussian structure, the preceding arguments and conclusions can then be used to yield the expected result.

In the last section, we briefly present, as a natural development of the preceding discussion, the new bracketing or local Lipschitz sufficient conditions for the CLT and the LIL obtained in [2] and [3]. The techniques involved produce a clear way to bound by nonrandom hypotheses the $L_1(P_n)$ and $L_2(P_n)$ portions which appear central in what we described previously. They are, however, combined with a summation by parts argument to ease the control of the $L_1(P_n)$ portion which is the most difficult one in general. We follow [2] in the exposition of the proof, with however a somewhat more general formulation that interpolates between the various known results.

Finally, we would like to refer to the papers by Giné and Zinn [10, 11] for more information on empirical processes, detailed references and notations that are not explained in the text as well as a precise definition and description of the central limit property on which, although implicitly contained, we do not directly concentrate here. Let us also mention that the four sections of this paper (except perhaps Section 3 that uses Section 2) can almost be read independently, as a function of the interest of the reader.

2. A comparison theorem. This section describes a comparison theorem for Rademacher averages; it can be used in many instances like the Gaussian comparison theorems based on Slepian's lemma (cf. [9]), thus avoiding the less efficient Gaussian randomization in the study of empirical processes on nonidentically distributed random variables. The main result is Theorem 5 which is obtained through several steps; it compares averages when coordinates are contracted.

In the following discussion on the comparison theorem, $\mathscr{F}$ will always denote, for simplicity, a class of real functions on $\{1,\dots,n\}$ such that $\|f(i)\|_{\mathscr{F}} < \infty$, $i = 1,\dots,n$, for a fixed integer n. It will also be convenient to use an explicit description of a probability space supporting a Rademacher sequence: On $\mathscr{E} = \{-1,+1\}^n$, consider the uniform probability measure and denote by ε_i: $\mathscr{E} \to \{-1,+1\}$ the ith coordinate function on $\mathscr{E}$: (ε_i) therefore defines a so-called Bernoulli or Rademacher sequence. Accordingly, if ε is any element in $\mathscr{E}$, we also denote by $(\varepsilon_i)_{i\le n}$ its coordinates. Let finally Φ: $\mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing.

Proposition 1. *With the preceding notation,*

$$E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i |f(i)|\right\|_{\mathscr{F}}\right) \le 2E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i f(i)\right\|_{\mathscr{F}}\right).$$

Proof. Let η be a map from $\mathscr{E}\times\{1,\dots,n\}$ into $\{-1,+1\}$. To each ε in $\mathscr{E}$, associate the subset $K(\varepsilon)$ of $\mathscr{E}$ consisting of the elements ξ in $\mathscr{E}$ such that $\xi_i = \eta(\varepsilon, i)$ whenever $\varepsilon_i = 1$. To any $A \subset \mathscr{E}$, set then $B = \bigcup_{\varepsilon\in A} K(\varepsilon)$.

Lemma 2. Card $A \le$ Card B.

Proof. Let η': $\mathscr{E}\times\{1,\dots,n\} \to \{-1,+1\}$ be defined by $\eta'(\varepsilon,1) = 1$ and $\eta'(\varepsilon,i) = \eta(\varepsilon,i)$ for $i \ge 2$. K' is associated to η' as K to η before and $B' = \bigcup_{\varepsilon\in A} K'(\varepsilon)$. To establish the lemma, it will be enough to show that Card $B' \le$ Card B since if by this procedure we replace each coordinate by 1, we reduce to the case $\eta \equiv 1$ for which trivially $A \subset B$ and the lemma holds. For ξ in $\mathscr{E}$, let $\bar{\xi}$ denote the element of $\mathscr{E}$ obtained by changing the sign of ξ_1. To prove that Card $B' \le$ Card B, it is enough to show that $\mathrm{Card}(B' \cap \{\xi,\bar{\xi}\}) \le \mathrm{Card}(B \cap \{\xi,\bar{\xi}\})$ for each ξ in $\mathscr{E}$ such that $\xi_1 = 1$. If $\bar{\xi} \in B'$, then $\bar{\xi} \in K'(\varepsilon)$ for some $\varepsilon \in a$, so $\varepsilon_1 = -1$, and hence $\xi, \bar{\xi} \in K(\varepsilon) \subset B$. If $\bar{\xi} \notin B'$, $\xi \in B'$, then $\xi \in K'(\varepsilon)$ for some $\varepsilon \in A$, and $\varepsilon_1 = 1$ since $\bar{\xi} \notin B'$. If $\eta(\varepsilon,1) = 1$ [resp. $\eta(\varepsilon,1) = -1$], then $\xi \in K(\varepsilon)$ [resp. $\bar{\xi} \in K(\varepsilon)$]. This completes the proof of Lemma 2. □

The next lemma is an immediate consequence of the preceding one and the marriage lemma (see, e.g., [5]).

Lemma 3. *Let η be a map from $\mathscr{E}\times\{1,\dots,n\}$ into $\{-1,+1\}$. There is a one-to-one map θ: $\mathscr{E}\to\mathscr{E}$ such that for each ε in $\mathscr{E}$ and $i = 1,\dots,n$, $\theta(\varepsilon)_i \eta(\varepsilon,i) \ge \varepsilon_i$, that is, $\theta(\varepsilon)_i = \eta(\varepsilon,i)$ whenever $\varepsilon_i = 1$.*

We are now in a position to prove Proposition 1. To each $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ in $\mathscr{E}$, let f_ε in $\mathscr{F}$ be such that

$$\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i |f(i)|\right\|_{\mathscr{F}}\right) = \Phi\left(\left|\sum_{i=1}^{n} \varepsilon_i |f_\varepsilon(i)|\right|\right),$$

where we write for simplicity an equality for what is only, in general, an arbitrarily close approximation. Note that $f_{-\varepsilon}$ can be chosen to be f_ε, where $-\varepsilon = (-\varepsilon_1, \dots, -\varepsilon_n)$. Let η be defined as $\eta(\varepsilon, i) = \operatorname{sign}(f_\varepsilon(i))$ so that $f_\varepsilon(i) = \eta(\varepsilon, i)|f_\varepsilon(i)|$ and let θ be the bijection of Lemma 3 corresponding to this map η. If $\sum_{i=1}^{n} \varepsilon_i |f_\varepsilon(i)| \geq 0$,

$$\Phi\left(\left|\sum_{i=1}^{n} \varepsilon_i |f_\varepsilon(i)|\right|\right) \leq \Phi\left(\sum_{i=1}^{n} \theta(\varepsilon)_i \eta(\varepsilon, i) |f_\varepsilon(i)|\right) = \Phi\left(\left|\sum_{i=1}^{n} \theta(\varepsilon)_i f_\varepsilon(i)\right|\right)$$

and if $\sum_{i=1}^{n} \varepsilon_i |f_\varepsilon(i)| \leq 0$, since $f_{-\varepsilon} = f_\varepsilon$,

$$\Phi\left(\left|\sum_{i=1}^{n} \varepsilon_i |f_\varepsilon(i)|\right|\right) = \Phi\left(\sum_{i=1}^{n} (-\varepsilon)_i |f_{-\varepsilon}(i)|\right) \leq \Phi\left(\left|\sum_{i=1}^{n} \theta(-\varepsilon)_i f_{-\varepsilon}(i)\right|\right).$$

Thus

$$\begin{aligned} E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i |f(i)|\right\|_{\mathscr{F}}\right) &= 2^{-n} \sum_{\varepsilon \in \mathscr{E}} \Phi\left(\left|\sum_{i=1}^{n} \varepsilon_i |f_\varepsilon(i)|\right|\right) \\ &\leq 2^{-n}\left(\sum_{\varepsilon \in \mathscr{E}} \Phi\left(\left|\sum_{i=1}^{n} \theta(\varepsilon)_i f_\varepsilon(i)\right|\right) + \sum_{\varepsilon \in \mathscr{E}} \Phi\left(\left|\sum_{i=1}^{n} \theta(-\varepsilon)_i f_{-\varepsilon}(i)\right|\right)\right) \\ &\leq 2E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i f(i)\right\|_{\mathscr{F}}\right), \end{aligned}$$

which is the result. □

The next proposition studies contractions on $\mathbb{R}_+$. It will then lead to the general result together with the preceding proposition. Let Φ be as before.

Proposition 4. *Let φ_i: $\mathbb{R}_+ \to \mathbb{R}$, $i = 1, \dots, n$, be contractions such that $\varphi_i(0) = 0$. Let also $\mathscr{F}$ be a class of positive functions on $\{1, \dots, n\}$. Then*

$$E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i \varphi_i(f(i))\right\|_{\mathscr{F}}\right) \leq E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i f(i)\right\|_{\mathscr{F}}\right).$$

Proof. Let ε be a Rademacher random variable independent of $\varepsilon_1, \dots, \varepsilon_n$. Since

$$E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i \varphi_i(f(i))\right\|_{\mathscr{F}}\right) = E\Phi\left(\left\|\varepsilon \sum_{i=1}^{n-1} \varepsilon_i \varphi_i(f(i)) + \varepsilon_n \varphi_n(f(n))\right\|_{\mathscr{F}}\right),$$

by conditioning on $\varepsilon_1, \dots, \varepsilon_{n-1}$ and iterating, it is enough to show that if φ is a contraction from $\mathbb{R}_+$ into $\mathbb{R}$ such that $\varphi(0) = 0$ and $\mathscr{F}$ is a class of functions f on $\{1,2\}$ such that $f(2) \geq 0$, one has

$$E\Phi\big(\|\varepsilon_1 f(1) + \varepsilon_2\varphi(f(2))\|_{\mathscr{F}}\big) \leq E\Phi\big(\|\varepsilon_1 f(1) + \varepsilon_2 f(2)\|_{\mathscr{F}}\big).$$

By a simple perturbation, we may assume Φ to be strictly increasing. Let f and g in $\mathscr{F}$ be such that

$$\begin{aligned} I &= E\Phi\big(\|\varepsilon_1 f(1) + \varepsilon_2\varphi(f(2))\|_{\mathscr{F}}\big) \\ &= \tfrac{1}{2}\Phi\big(|f(1) + \varphi(f(2))|\big) + \tfrac{1}{2}\Phi\big(|g(1) - \varphi(g(2))|\big), \end{aligned}$$

where we write again an equality for simplicity and may and do assume that

$$\Phi\big(|f(1) + \varphi(f(2))|\big) \geq \Phi\big(|g(1) + \varphi(g(2))|\big) \tag{1}$$

and

$$\Phi\big(|g(1) - \varphi(g(2))|\big) \geq \Phi\big(|f(1) - \varphi(f(2))|\big). \tag{2}$$

We distinguish between the following cases.

CASE 1. $a = f(1) + \varphi(f(2)) \geq 0$, $b = g(1) - \varphi(g(2)) \geq 0$. Assume to begin with that $g(2) \leq f(2)$, so that, since φ is a contraction,

$$c = f(2) - \varphi(f(2)) \geq g(2) - \varphi(g(2)) = d \geq 0.$$

Note that, by (1),

$$f(1) + \varphi(f(2)) \geq g(1) + \varphi(g(2)) \geq g(1) - g(2)$$

so that $a \geq b - d$. Using then that $\Phi(|\cdot + t|) - \phi(|\cdot|)$ is increasing for $t \geq 0$ since Φ is convex and increasing, it is easily seen from $a \geq b - d$ and $c \geq d \geq 0$ that

$$\begin{aligned} \Phi(|a + c|) - \Phi(|a|) &\geq \Phi(|(b - d) + c|) - \Phi(|b - d|) \\ &\geq \Phi(b) - \Phi(|b - d|), \end{aligned}$$

that is,

$$2I \leq \Phi\big(|f(1) + f(2)|\big) + \Phi\big(|g(1) - g(2)|\big),$$

which is the result in this situation. If $f(2) \leq g(2)$, let

$$p = g(1) + g(2) - a, \qquad q = f(1) - f(2) - b,$$

so that $p + q \geq 0$ by contraction. Note that $a \geq b + q$. Observe also that, by (2),

$$g(1) - \varphi(g(2)) - f(1) + \varphi(f(2)) \geq 0$$

and, by contraction,

$$-\varphi(g(2)) + \varphi(f(2)) \leq g(2) - f(2) \leq g(2) - \varphi(f(2))$$

so that $p \geq 0$. As previously, we then have

$$\begin{aligned} \Phi(|a + p|) - \Phi(|a|) &\geq \Phi\big(|(b + q) + p|\big) - \Phi(|b + q|) \\ &\geq \Phi(b) - \Phi(|b + q|), \end{aligned}$$

that is,

$$2I \leq \Phi(|g(1) - g(2)|) + \Phi(|f(1) - f(2)|),$$

which again gives the result. The expected inequality is thus proved in this first case.

CASE 2. $f(1) + \varphi(f(2)) \leq 0$, $g(1) - \varphi(g(2)) \leq 0$. It is completely similar to the preceding case.

CASE 3. $f(1) + \varphi(f(2)) \geq 0$, $g(1) - \varphi(g(2)) \leq 0$. We have

$$2I = \Phi(f(1) + \varphi(f(2))) + \Phi(-g(1) + \varphi(g(2))).$$

Since Φ is increasing and $\varphi(f(2)) \leq f(2)$, $\varphi(g(2)) \leq g(2)$,

$$2I \leq \Phi(f(1) + f(2)) + \Phi(-g(1) + g(2))$$

and the result follows.

CASE 4. $f(1) + \varphi(f(2)) \leq 0$, $g(1) - \varphi(g(2)) \geq 0$. Similar to Case 3.

The proof is complete. □

Propositions 1 and 4 can now be combined to yield the general comparison theorem. For Gaussian averages, a similar conclusion easily follows from the Gaussian comparison theorems.

THEOREM 5. *Let Φ: $\mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing. Let φ_i: $\mathbb{R} \to \mathbb{R}$ be contractions such that $\varphi_i(0) = 0$, $i = 1, \dots, n$, and let $\mathscr{F}$ be a class of functions on $\{1, \dots, n\}$. Then*

$$E\Phi\left(\frac{1}{2}\left\|\sum_{i=1}^{n} \varepsilon_i \varphi_i(f(i))\right\|_{\mathscr{F}}\right) \leq \frac{3}{2} E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i f(i)\right\|_{\mathscr{F}}\right).$$

PROOF. Since $f^+ = \frac{1}{2}(f + |f|)$, by Proposition 1 and convexity,

$$E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i f^+(i)\right\|_{\mathscr{F}}\right) \leq \frac{1}{2}\left(E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i f(i)\right\|_{\mathscr{F}}\right) + E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i |f(i)|\right\|_{\mathscr{F}}\right)\right)$$

$$\leq \frac{3}{2} E\Phi\left(\left\|\sum_{i=1}^{n} \varepsilon_i f(i)\right\|_{\mathscr{F}}\right)$$

and similarly for f^-. Since $\varphi_i(0) = 0$, note that $\varphi_i(f) = \varphi_i(f^+) + \varphi_i(-f^-)$ for each f. Since $\varphi_i(\cdot)$ and $\varphi_i(-\,\cdot\,)$ are contractions on $\mathbb{R}_+$, it follows from

Proposition 4 that

$$\begin{aligned} E\Phi\left(\tfrac{1}{2}\left\|\sum_{i=1}^n \varepsilon_i\varphi_i(f(i))\right\|_{\mathscr{F}}\right) &\le \tfrac{1}{2}\left(E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i\varphi_i(f^+(i))\right\|_{\mathscr{F}}\right) + E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i\varphi_i(f^-(i))\right\|_{\mathscr{F}}\right)\right) \\ &\le \tfrac{1}{2}\left(E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i f^+(i)\right\|_{\mathscr{F}}\right) + E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i f^-(i)\right\|_{\mathscr{F}}\right)\right) \\ &\le \tfrac{3}{2}E\Phi\left(\left\|\sum_{i=1}^n \varepsilon_i f(i)\right\|_{\mathscr{F}}\right) \end{aligned}$$

and the theorem is established. □

In the last part of this section, we present two applications of the preceding comparison theorem in the form of simplified proofs of results of Giné and Zinn. The next section will contain further and more important applications to strong limit theorems. We should point out that in all these applications Theorem 5 is mainly used in the following simple form with, in particular, $\Phi(t) = t$:

$$\begin{aligned} E\left\|\sum_{i=1}^n \varepsilon_i f^2(i)\right\|_{\mathscr{F}} &= 2E\left\|\sum_{i=1}^n \varepsilon_i\varphi_i(\|f(i)\|_{\mathscr{F}} f(i))\right\|_{\mathscr{F}} \\ &\le 6E\left\|\sum_{i=1}^n \varepsilon_i\|f(i)\|_{\mathscr{F}} f(i)\right\|_{\mathscr{F}} \\ &\le 6\max_{i\le n}\|f(i)\|_{\mathscr{F}} E\left\|\sum_{i=1}^n \varepsilon_i f(i)\right\|_{\mathscr{F}}, \end{aligned}$$

where $\varphi_i(t) = \min(t^2/2\|f(i)\|_{\mathscr{F}}^2, \|f(i)\|_{\mathscr{F}}^2/2)$, $i = 1,\dots,n$, and where the last inequality follows by the contraction principle [13]. The above may be considered as a version of the "square root trick" in terms of expectations (e.g., [11], proof of Lemma 1.3.3). The first application concerns a weak law of large numbers for squares. Before stating it, we need to give a symmetrization inequality close to [11], Lemma 1.2.3, and proved similarly. We take again the notation described in the introduction.

LEMMA 6. *Let $\mathscr{F}$ be a class in L_1. For any $t > 0$ and integer n,*

$$\begin{aligned} \Pr\left\{\left\|\sum_{i=1}^n (f(X_i) - Ef)\right\|_{\mathscr{F}} > 3t\right\} &\le 2\Pr\left\{\left\|\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}} > t\right\} + \sup_{f\in\mathscr{F}}\Pr\left\{\left|\sum_{i=1}^n (f(X_i) - Ef)\right| > t\right\}. \end{aligned}$$

The following theorem has been essentially obtained in [11], Theorem 1.4.6.

THEOREM 7. *Let $\mathscr{F}$ be a class of functions such that $\|Ef\|_{\mathscr{F}} < \infty$. If the sequence $(\|\sum_{i=1}^n (f(X_i) - Ef)/\sqrt{n}\|_{\mathscr{F}})_{n \in \mathbb{N}}$ is bounded in probability and if $\lim_{t\to\infty} t^2 P(\|f\|_{\mathscr{F}} > t) = 0$ (in particular if $\mathscr{F}$ satisfies the* CLT), *then*

$$\frac{1}{n}\left\|\sum_{i=1}^n \left(f^2(X_i) - Ef^2\right)\right\|_{\mathscr{F}} \to 0 \quad \text{in probability,}$$

that is, the class $\mathscr{F}^2 = \{f^2;\ f \in \mathscr{F}\}$ satisfies the weak law of large numbers.

PROOF. We first show that

$$\frac{1}{n}\left\|\sum_{i=1}^n \varepsilon_i f^2(X_i)\right\|_{\mathscr{F}} \to 0 \quad \text{in probability.}$$

Let $\varepsilon > 0$ be fixed. For each n,

$$\Pr\left\{\left\|\sum_{i=1}^n \varepsilon_i f^2(X_i)\right\|_{\mathscr{F}} > \varepsilon n\right\}$$

$$\leq n\Pr\{\|f(X_1)\|_{\mathscr{F}} > \varepsilon^2\sqrt{n}\} + \frac{1}{\varepsilon n}E_X E_\varepsilon\left\|\sum_{i=1}^n \varepsilon_i f^2(X_i) I_{\{\|f(X_i)\|_{\mathscr{F}} \leq \varepsilon^2\sqrt{n}\}}\right\|_{\mathscr{F}}.$$

We use Theorem 5 conditionally on the X_i's to bound the second term on the right-hand side of the preceding inequality by

$$\frac{6\varepsilon}{\sqrt{n}}E\left\|\sum_{i=1}^n \varepsilon_i f(X_i) I_{\{\|f(X_i)\|_{\mathscr{F}} \leq \varepsilon^2\sqrt{n}\}}\right\|_{\mathscr{F}}.$$

By Hoffmann-Jørgensen's inequality ([13], pages 164–165 or [11], Lemma 1.2.6), the stochastic boundedness of the sequence $(\|\sum_{i=1}^n (f(X_i) - Ef)/\sqrt{n}\|_{\mathscr{F}})_{n \in \mathbb{N}}$ implies that

$$\sup_n \frac{1}{\sqrt{n}}E\left\|\sum_{i=1}^n \varepsilon_i f(X_i) I_{\{\|f(X_i)\|_{\mathscr{F}} \leq \varepsilon^2\sqrt{n}\}}\right\|_{\mathscr{F}} < \infty$$

uniformly in $\varepsilon \leq 1$. Thus, letting n go to infinity, our claim is established. To get the full conclusion, first note that the stochastic boundedness of the sequence $(\|\sum_{i=1}^n (f(X_i) - Ef)/\sqrt{n}\|_{\mathscr{F}})_{n \in \mathbb{N}}$ implies that $\mathscr{F}$ is totally bounded in L_2. Indeed, by considering finite subclasses of $\mathscr{F}$ and letting n tend to infinity, it follows from the finite dimensional CLT that the set of all $f - Ef$ in L_2 defines a GB set, which is thus totally bounded. Since $\|Ef\|_{\mathscr{F}} < \infty$, $\mathscr{F}$ is also totally bounded in L_2. In particular, we see that

$$\lim_{n\to\infty} \frac{1}{n}\sup_{f \in \mathscr{F}} E\left(\max_{i \leq n} f^2(X_i)\right) = 0.$$

Therefore, by Hoffmann-Jørgensen's inequality and the preceding claim,

$$\sup_{f\in\mathscr{F}}\Pr\left\{\left|\sum_{i=1}^{n}\left(f^2(X_i)-Ef^2\right)\right|>\varepsilon n\right\}\le\frac{2}{\varepsilon n}\sup_{f\in\mathscr{F}}E\left|\sum_{i=1}^{n}\varepsilon_i f^2(X_i)\right|\to 0\quad\text{as } n\to\infty$$

for each $\varepsilon > 0$. The conclusion then follows from Lemma 6. □

The preceding proof can actually be easily adapted to yield a general phenomenon involving the comparison theorem that can, moreover, be applied to independent but not necessarily identically distributed random variables X_i, $i \in \mathbb{N}$. Let $\varphi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be such that $\varphi(0) = 0$ and φ increases to infinity. Assume, moreover, that

$$\limsup_{\varepsilon\to 0}\limsup_{t\to\infty}\frac{t\varphi'(\varepsilon t)}{\varphi(t)}=0.$$

Then, if $\mathscr{F}$ is a class of functions such that the family $\{\varphi(|f(X_i)|);\ f\in\mathscr{F},\ i\in\mathbb{N}\}$ is uniformly integrable, $\lim_{n\to\infty}\sum_{i=1}^{n}\Pr\{\|f(X_i)\|_{\mathscr{F}}>\varepsilon\varphi^{-1}(n)\}=0$ for each $\varepsilon>0$ and the sequence $(\|\sum_{i=1}^{n}(f(X_i)-Ef(X_i))/\varphi^{-1}(n)\|_{\mathscr{F}})_{n\in\mathbb{N}}$ is bounded in probability, we have

$$\frac{1}{n}\left\|\sum_{i=1}^{n}\left(\varphi(|f(X_i)|)-E\varphi(|f(X_i)|)\right)\right\|_{\mathscr{F}}\to 0\quad\text{in probability.}$$

The second application deals with random entropy conditions in the CLT for classes of uniformly bounded functions ([11], Theorem 2.2.1 and [21], Theorem 3). Our elementary proof avoids and actually enlightens through the comparison theorem the so-called "square root trick." Note also the simplification of constants in the statement.

Given a metric space (T, d), $N(T, d; \varepsilon)$ denotes the minimal number of d-balls of radius $\varepsilon > 0$ that cover T. Recall also from [11] the notation $\mathscr{F}'_{\varepsilon,n}$ as the class of elements of the form $f-g$ for f and g in a class $\mathscr{F}$ satisfying $E|f-g|^2\le\varepsilon_n^{-1/2}$. Recall we write $\mathscr{F}\subset\mathscr{F}_1+\mathscr{F}_2$ to mean that any element f of a class $\mathscr{F}$ can be written as f_1+f_2 with $f_1\in\mathscr{F}_1$ and $f_2\in\mathscr{F}_2$. Let finally $W_{n,1}$ denote the unit ball of $L_1(P_n)$.

THEOREM 8. *Let $\mathscr{F}$ be a uniformly bounded pregaussian class. Define, for $\varepsilon, \gamma > 0$ and n integer, the event*

$$A(\varepsilon,n;\gamma)=\Big\{\text{there exists } \mathscr{G}, \text{ finite class of functions on } S, \text{ such that:}$$

$$\int_0^{n^{-1/4}}(\log N(\mathscr{G},d_{n,2};t))^{1/2}\,dt<\gamma \text{ and } \mathscr{F}'_{\varepsilon,n}\subset\gamma n^{-1/2}W_{n,1}+\mathscr{G}\Big\}.$$

Then, if $\liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\Pr\{A(\varepsilon,n;\gamma)\}=1$ *for all* $\gamma>0$, *$\mathscr{F}$ satisfies the CLT.*

PROOF. Since the class $\mathscr{F}$ is pregaussian and uniformly bounded, it is enough to show, by Theorem 2.1.1 of [11], that

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} E\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} = 0.$$

We may and do assume that $\sup_{f\in\mathscr{F}}\|f\|_\infty \le \frac{1}{2}$ (for example). It is also easy to see (cf. [11]) that, changing $\mathscr{G}$ if necessary, we may reduce ourselves to the case where each f in $\mathscr{F}'_{\varepsilon,n}$ can be written on $A(\varepsilon, n; \gamma)$ as a sum $h + g$ where $g \in \mathscr{G}$, $h \in \gamma n^{-1/2}W_{n,1}$ and $\|h\|_\infty \le 2$. Thus we can suppose that

$$D = \sup_{f\in\mathscr{G}}\|f\|_{n,2} \le \sup_{f\in\mathscr{F}'_{\varepsilon,n}}\|f\|_{n,2} + \left(2\gamma n^{-1/2}\right)^{1/2}.$$

On $A(\varepsilon, n; \gamma)$,

$$\int_0^\infty \left(\log N(\mathscr{G}, d_{n,2}; t)\right)^{1/2} dt \le \gamma + \int_{n^{-1/4}}^{2D}\left(\log N\left(\mathscr{G}, d_{n,2}; n^{-1/4}\right)\right)^{1/2} dt$$
$$\le \gamma(1 + 2Dn^{1/4}),$$

and, therefore, by Dudley's majorization theorem (cf. for example [11], Section 1.4),

$$E_\varepsilon\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{G}} \le K\left[D + \gamma(1 + 2Dn^{1/4})\right],$$

where K denotes a numerical constant, possibly changing from line to line below. Now, since on $A(\varepsilon, n; \gamma)$ we have $\mathscr{F}'_{\varepsilon,n} \subset \gamma n^{-1/2}W_{n,1} + \mathscr{G}$, it follows that on this set,

$$\begin{aligned}E_\varepsilon\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} &\le \gamma + K\left[D + \gamma(1 + 2Dn^{1/4})\right]\\ &\le (K+1)\gamma\\ &\quad + K(1 + 2\gamma n^{1/4})\left[\sup_{f\in\mathscr{F}'_{\varepsilon,n}}\|f\|_{n,2} + \left(2\gamma n^{-1/2}\right)^{1/2}\right].\end{aligned}$$

Thus, for $0 < \gamma \le 1$ and $n \ge n_0$ large enough depending on γ,

$$E_\varepsilon\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} \le K\gamma\left(1 + n^{1/4}\left\|\sum_{i=1}^{n} f^2(X_i)/n\right\|_{\mathscr{F}'_{\varepsilon,n}}^{1/2}\right).$$

Using that $Ef^2 \le \varepsilon n^{-1/2}$ for all f in $\mathscr{F}'_{\varepsilon,n}$, we have moreover that for some K, $0 < \varepsilon \le 1$ and $n \ge n_0$,

$$E_\varepsilon\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} \le K\gamma\left(1 + \left\|\sum_{i=1}^{n}\left(f^2(X_i) - Ef^2\right)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}}^{1/2}\right).$$

Let us now choose $\varepsilon \le 1$ and $n_1 \ge n_0$ such that for $n \ge n_1$, $A = A(\varepsilon, n; \gamma)$ satisfies $\Pr(A^c) \le 10^{-2}$ (i.e., small enough). Let $t > 0$ be specified later. For these choices, by what we obtained previously,

$$\begin{aligned}
&\Pr\left\{\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} > t\right\} \\
&\le \Pr(A^c) + \frac{1}{t}\int_A E_\varepsilon\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} dP_X \\
&\le \Pr(A^c) + \frac{K\gamma}{t}\left[1 + E\left(\left\|\sum_{i=1}^{n} \left(f^2(X_i) - Ef^2\right)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}}^{1/2}\right)\right] \\
&\le \Pr(A^c) + \frac{K\gamma}{t}\left[1 + \left(E\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}}\right)^{1/2}\right],
\end{aligned}$$

where in the last step we have used Jensen's inequality and comparison Theorem 5. Choose now

$$t = 10^2 K\gamma\left[1 + \left(E\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}}\right)^{1/2}\right]$$

so that

$$\Pr\left\{\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} > t\right\} \le 2.10^{-2}.$$

We are thus in a position to apply Hoffmann-Jørgensen's inequality ([11], Lemma 1.2.6) to get that for ε as before and $n \ge n_2 \ge n_1$,

$$E\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} \le Kt.$$

By the choice of t, this implies that for some numerical K,

$$E\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathscr{F}'_{\varepsilon,n}} \le K\gamma,$$

from which the conclusion follows. □

3. The law of the iterated logarithm for empirical processes. This section is devoted to the LIL for empirical processes, in particular in non-equidistributed situations. As an introduction, we first translate into the language of empirical processes the recent characterization of the i.i.d. LIL in Banach spaces [16]. This result and its proof will actually be the conducting rod to the further developments.

Recall the empirical measures P_n and let

$$I_n = (P_n - P)\sqrt{n/2LLn}, \qquad n \in \mathbb{N},$$

where $Lt = \max(1, \log t)$, $t \in \mathbb{R}_+$, and $LLt = L(Lt)$.

For a class $\mathscr{F}$ in L_2, define the set $\mathscr{K}_{\mathscr{F}}$ of functions on $\mathscr{F}$ as

$$\mathscr{K}_{\mathscr{F}} = \{ f \to E(fg);\ Eg = 0,\ Eg^2 \le 1 \}.$$

The class $\mathscr{F}$ is said to satisfy the LIL for P (or $\mathscr{F}$ is a P-Strassen class) if, with probability 1, the sequence $(I_n)_{n \in \mathbb{N}}$ is relatively compact in the space $l_\infty(\mathscr{F})$ of all bounded functions on $\mathscr{F}$ with $\mathscr{K}_{\mathscr{F}}$ as the set of limit points.

The main result of [16] can be expressed as follows.

Theorem 9. *Let $\mathscr{F}$ be a class of functions in L_2 such that $\|Ef\|_{\mathscr{F}} < \infty$. The class $\mathscr{F}$ satisfies the* LIL *if and only if the following three conditions are fulfilled*:

(i) *$\mathscr{F}$ is totally bounded in L_2.*
(ii) $E(\|f\|_{\mathscr{F}}^2 / LL\|f\|_{\mathscr{F}}) < \infty$.
(iii) $\|(I_n(f))\|_{\mathscr{F}} \to 0$ *in probability.*

This result is proved as in [16]: Necessity of (ii) and (iii) follows similarly and (i) can be deduced from the compactness of $\mathscr{K}_{\mathscr{F}}$ in $l_\infty(\mathscr{F})$. Conversely, the main step consists in proving, following [16] closely, that for any class $\mathscr{F}$ satisfying (ii) and (iii), with probability 1,

$$\limsup_{n \to \infty} \|I_n(f)\|_{\mathscr{F}} \le K \|Ef^2\|_{\mathscr{F}}^{1/2},$$

where K is a numerical constant. It is then easy to deduce from the fact that $\mathscr{F}$ is totally bounded in L_2 that the sequence $(I_n)_{n \in \mathbb{N}}$ is relatively compact. Identification of the limit set is established as in [8].

Theorem 9 thus reduces the investigation of classes satisfying the LIL to classes for which $\|I_n(f)\|_{\mathscr{F}} \to 0$ in probability, a weak convergence close in some sense to the CLT. It will be one of the purposes of the next paragraphs to find descriptions as well as sufficient conditions for such a property to hold.

The proof of Theorem 9 [16] is based on Gaussian randomization and use of the isoperimetric concentration inequality of the norm of a Gaussian random vector around its median or mean and the Gaussian comparison theorems based on Slepian's lemma. Attempting to investigate the non-equidistributed case in the form of Kolmogorov's LIL (see, e.g., [19], page 269) for empirical processes, we will make use of these ideas. However, we replace the Gaussian concentration inequality by a similar inequality (that follows) for averages by random variables uniformly distributed on $[-1, +1]$, which was pointed out in [18], (2.14), and the Gaussian comparison theorems by the results of Section 2, thus avoiding the Gaussian randomization. The main argument of the proof lies then in an application of the new isoperimetric inequality for product measures recently established in [20]. We obtain in this way an extension of Kolmogorov's LIL for empirical processes. A version of Kolmogorov's LIL for Banach space valued random variables has already been obtained by Kuelbs [14] for normalizers that do not allow two-sided bounds in general. We use weak moments to get this better description.

In what follows, the measurable space $(S, \mathscr{S})$ is equipped with a family of probabilities $(Q_i)_{i \in \mathbb{N}}$ and $(\Omega, \Sigma, \Pr)$ is the product of $(S^{\mathbb{N}}, \Sigma^{\mathbb{N}}, \prod_i Q_i)$ with a rich enough probability space supporting a Bernoulli sequence $(\varepsilon_i)_{i\in\mathbb{N}}$ as well as a sequence $(u_i)_{i\in\mathbb{N}}$ of independent random variables uniformly distributed on $[-1,+1]$. We use the same conventions as those described in the introduction about probabilities and expectations [in particular, P_u and E_u with respect to (u_i)] and assume moreover the sequences (ε_i) and (u_i) to be independent. X_i, $i\in\mathbb{N}$, are the independent coordinate functions, each of them of law Q_i.

Theorem 10. *Let $\mathscr{F}$ be a class of functions. We assume:*

(i) *For each $i\in\mathbb{N}$, $\|Ef^2(X_i)\|_{\mathscr{F}}<\infty$ and $Ef(X_i)=0$, $f\in\mathscr{F}$.*

(ii) *Whenever $s_n^2=\|\sum_{i=1}^n Ef^2(X_i)\|_{\mathscr{F}}$, the sequence (s_n) increases to infinity and for some sequence (K_i) of positive numbers such that $K_i\to 0$, for each i, with probability* 1

$$\|f(X_i)\|_{\mathscr{F}}\le K_i s_i/\left(2LLs_i^2\right)^{1/2}.$$

(iii) *The sequence $(\|\sum_{i=1}^n f(X_i)\|_{\mathscr{F}}/(2s_n^2LLs_n^2)^{1/2})_{n\in\mathbb{N}}$ is bounded in probability.*

Then with probability 1,

$$0<\limsup_{n\to\infty}\frac{\|\sum_{i=1}^n f(X_i)\|_{\mathscr{F}}}{\left(2s_n^2LLs_n^2\right)^{1/2}}<\infty.$$

Before turning to the proof of this result, let us make the following comments. The sequence (s_n) defined in (ii) is constructed as a supremum of weak variances in order to allow the lower bound in the conclusion. Condition (i) is natural and the bound (ii) on $\|f(X_i)\|_{\mathscr{F}}$ is the best possible on the line ($\mathscr{F}$ finite). As usual, $\sup_i K_i<\infty$ is sufficient for the upper bound, $\lim_{i\to\infty}K_i=0$ being used to show that the lim sup is strictly positive, actually greater than 1. Hypothesis (iii) parallels the condition in Theorem 9 and is typical in the context of almost sure limit theorems for infinite dimensional random variables. It is, of course, necessary for the lim sup to be finite. Finally, note that the proof and the zero-one law will actually show that, with probability 1,

$$1\le\limsup_{n\to\infty}\frac{\|\sum_{i=1}^n f(X_i)\|_{\mathscr{F}}}{\left(2s_n^2LLs_n^2\right)^{1/2}}=M<\infty.$$

We do not know whether $M=1$ can be obtained as it is the case on the line, even if (iii) is strengthened into a convergence in probability to 0.

Proof of Theorem 10. Throughout the proof, we set $t_n=(2LLs_n^2)^{1/2}$ in order to ease the notation and follow in this [19]. Let $\rho>1$ and, for each k, let n_k be the smallest integer such that $s_n>\rho^k$. It is easy to see, using (ii), that

$$\frac{s_{n+1}}{s_n}\sim 1,\qquad s_{n_k}\sim\rho^k\quad\text{and}\quad\frac{s_{n_{k+1}}}{s_{n_k}}\sim\rho.$$

In the first part of the proof, we show that the lim sup is finite. It will be enough to take $\rho = 2$ there and we do that for simplicity. We first claim that it is enough to prove that for some M,

$$\sum_k \Pr\left\{\left\|\sum_{i=1}^{n_k} u_i f(X_i)\right\|_{\mathscr{F}} \geq M s_{n_k} t_{n_k}\right\} < \infty. \tag{3}$$

Indeed, by Lévy's inequality and Hoffmann-Jørgensen's integrability theorems [13], we then have that

$$E\left(\sup_n \frac{1}{s_n t_n}\left\|\sum_{i=1}^{n} u_i f(X_i)\right\|_{\mathscr{F}}\right) < \infty.$$

By centering, symmetry and Jensen's inequality,

$$\begin{aligned} &E\left(\sup_n \frac{1}{s_n t_n}\left\|\sum_{i=1}^{n} f(X_i)\right\|_{\mathscr{F}}\right) \\ &\leq 2E\left(\sup_n \frac{1}{s_n t_n}\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right) \\ &\leq 2(E|u_1|)^{-1} E\left(\sup_n \frac{1}{s_n t_n}\left\|\sum_{i=1}^{n} \varepsilon_i |u_i| f(X_i)\right\|_{\mathscr{F}}\right) \\ &= 4E\left(\sup_n \frac{1}{s_n t_n}\left\|\sum_{i=1}^{n} u_i f(X_i)\right\|_{\mathscr{F}}\right), \end{aligned}$$

which establishes our claim.

One of the main tools in the proof of (3) is the recent isoperimetric inequality for independent random variables proved in [20]. In our context, the result reads as follows.

LEMMA 11. *Let n, m, q be fixed integers such that $m \geq q \geq 2$. There is a universal constant K such that if A is a (measurable) subset of S^n satisfying $\Pr\{(X_i)_{i \leq n} \in A\} \geq \frac{1}{2}$, then*

$$\Pr_*(H(A, m, q)) \geq 1 - \left(\frac{K}{q}\right)^m,$$

where

$$H(A, m, q) = \left\{\exists\, x^1, \dots, x^q \in A \text{ such that: } \operatorname{Card}\left\{i \leq n;\ X_i \notin \left\{x_i^1, \dots, x_i^q\right\}\right\} \leq m\right\}.$$

This result is used together with the comparison properties described in Section 2 as well as an exponential concentration inequality of the Gaussian type for averages by uniformly distributed random variables on $[-1, +1]$. This inequality was proved in [18], (2.14), actually as a consequence of the Gaussian inequalities. It implies the following lemma.

LEMMA 12. *Let* $Z = \|\sum_{i=1}^n u_i f(x_i)\|_{\mathscr{F}}$ *where* $x_1, \dots, x_n \in S$ *and set* $\sigma^2 = \|\sum_{i=1}^n f^2(x_i)\|_{\mathscr{F}}$. *Then, for each* $t > 0$,

$$\Pr\{Z > EZ + t\} \le \exp(-t^2/\pi\sigma^2).$$

We can now perform the main step in the proof of (3). First note that under assumptions (iii) and (ii), classical arguments involving symmetry and Hoffmann-Jørgensen's inequality as detailed in [14], show that for some constant C (≥ 1),

$$E\left\|\sum_{i=1}^{n_k} \varepsilon_i f(X_i)\right\|_{\mathscr{F}} \le C s_{n_k} t_{n_k} \tag{4}$$

for all k. Let now k be fixed but arbitrary and assume for simplicity that $\sup_i K_i \le 1$. Define A_1 and A_2 in S^{n_k} by

$$A_1 = \left\{(x_1, \dots, x_{n_k});\ E_u\left\|\sum_{i=1}^{n_k} u_i f(x_i)\right\|_{\mathscr{F}} \le 4C s_{n_k} t_{n_k}\right\},$$

$$A_2 = \left\{(x_1, \dots, x_{n_k});\ \left\|\sum_{i=1}^{n_k} f^2(x_i)\right\|_{\mathscr{F}} \le 52 C s_{n_k}^2\right\}.$$

Using (4), clearly $\Pr\{(X_i)_{i \le n_k} \in A_1\} \ge \frac{3}{4}$. Concerning A_2, first note that by the comparison properties of Section 2 and (ii),

$$\begin{aligned} E\left\|\sum_{i=1}^{n_k} f^2(X_i)\right\|_{\mathscr{F}} &\le s_{n_k}^2 + E\left\|\sum_{i=1}^{n_k} \left(f^2(X_i) - Ef^2(X_i)\right)\right\|_{\mathscr{F}} \\ &\le s_{n_k}^2 + 2E\left\|\sum_{i=1}^{n_k} \varepsilon_i f^2(X_i)\right\|_{\mathscr{F}} \\ &\le s_{n_k}^2 + 12\frac{s_{n_k}}{t_{n_k}} E\left\|\sum_{i=1}^{n_k} \varepsilon_i f(X_i)\right\|_{\mathscr{F}}. \end{aligned}$$

Hence, by (4) again,

$$E\left\|\sum_{i=1}^{n_k} f^2(X_i)\right\|_{\mathscr{F}} \le 13 C s_{n_k}^2$$

and $\Pr\{(X_i)_{i \le n_k} \in A_2\} \ge \frac{3}{4}$. Setting $A = A_1 \cap A_2$, we thus have that $\Pr\{(X_i)_{i \le n_k} \in A\} \ge \frac{1}{2}$. Letting $M = 1 + 14C(2K + 1)$ where K is the universal constant of Lemma 11, we get from this lemma that [with some abuse in the (non-) measurability of $H(A, m, q)$]

$$\begin{aligned} &\Pr\left\{\left\|\sum_{i=1}^{n_k} u_i f(X_i)\right\|_{\mathscr{F}} > M s_{n_k} t_{n_k}\right\} \\ &\quad \le \left(\frac{K}{q}\right)^m + \int_{H(A, m, q)} P_u\left\{\left\|\sum_{i=1}^{n_k} u_i f(X_i)\right\|_{\mathscr{F}} \ge M s_{n_k} t_{n_k}\right\} dP_X. \end{aligned} \tag{5}$$

On $H(a, m, q)$, there exist $x^1, \dots, x^q$ in A and $i_1, \dots, i_j$, $j \le m$, in $\{1, \dots, n_k\}$

such that

$$\{1,\dots,n_k\} = J \cup \{i_1,\dots,i_j\},$$

where $J = \bigcup_{l=1}^{q} I_l$ and $I_l = \{i \le n_k;\ X_i = x_i^l\}$, $l = 1,\dots,q$. Hence, on $H(A, m, q)$, by (ii),

$$\left\| \sum_{i=1}^{n_k} u_i f(X_i) \right\|_{\mathscr{F}} \le m \frac{s_{n_k}}{t_{n_k}} + \left\| \sum_{i\in J} u_i f(X_i) \right\|_{\mathscr{F}}.$$

We note that, by monotonicity of averages and since $x^1,\dots,x^q \in A \subset A_1$,

$$E_u \left\| \sum_{i\in J} u_i f(X_i) \right\| \le \sum_{l=1}^{n_k} E_u \left\| \sum_{i=1}^{n_k} u_i f(x_i^l) \right\| \le 4qCs_{n_k} t_{n_k}.$$

If we let $m = [t_{n_k}^2]$ and $q = [2K] + 1$ (≥ 2), we thus get that on $H(A, m, q)$,

$$P_u\left\{ \left\| \sum_{i=1}^{n_k} u_i f(X_i) \right\|_{\mathscr{F}} > Ms_{n_k} t_{n_k} \right\}$$

$$\le P_u\left\{ \left\| \sum_{i\in J} u_i f(X_i) \right\|_{\mathscr{F}} > E_u \left\| \sum_{i\in J} u_i f(X_i) \right\| + 10C(2K+1) s_{n_k} t_{n_k} \right\}.$$

Applying now Lemma 12, the preceding probability is estimated by

$$\exp\left(-10^2 C^2 (2K+1)^2 s_{n_k}^2 t_{n_k}^2 / \pi \left\| \sum_{i\in J} f^2(X_i) \right\|_{\mathscr{F}} \right).$$

Since $x^1,\dots,x^q \in A \subset A_2$, we see that

$$\left\| \sum_{i\in J} f^2(X_i) \right\|_{\mathscr{F}} \le \sum_{l=1}^{q} \left\| \sum_{i=1}^{n_k} f^2(x_i^l) \right\|_{\mathscr{F}} \le 52C(2K+1) s_{n_k}^2.$$

Therefore, we have finally obtained that, on $H(A, m, q)$, for the choices of q and m described before,

$$P_u\left\{ \left\| \sum_{i=1}^{n_k} u_i f(X_i) \right\|_{\mathscr{F}} > Ms_{n_k} t_{n_k} \right\} \le \exp\left(-t_{n_k}^2 \right).$$

Therefore, by (5),

$$\Pr\left\{ \left\| \sum_{i=1}^{n_k} u_i f(X_i) \right\|_{\mathscr{F}} > Ms_{n_k} t_{n_k} \right\} \le 2^{-[t_{n_k}^2]} + \exp\left(-t_{n_k}^2 \right),$$

which holds for each k (large enough) since k was arbitrary in the preceding discussion. Since $t_{n_k}^2 \sim 2LL2^k$, this establishes (3) and we have thereby completed the first part of the proof of the theorem.

We now show that the lim sup is strictly positive and follow closely ([19], pages 271–272). Recall that for $\rho > 1$ we let $n_k = \inf\{n\colon s_n > \rho^k\}$. By the first part of the proof and the zero-one law, for some finite number M,

$$\Pr\left\{ \left\| \sum_{i=1}^{n_k} f(X_i) \right\|_{\mathscr{F}} \le Ms_{n_k} t_{n_k} \text{ for all } k \text{ large enough} \right\} = 1.$$

Let $\delta > 0$ and suppose we can prove that

$$\Pr\left\{\left\|\sum_{i=n_k+1}^{n_{k+1}} f(X_i)\right\|_{\mathscr{F}} > (1-\delta)^2\left(2\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}} LL\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}}\right)^{1/2} \text{ i.o. in } k\right\} = 1. \tag{6}$$

Then, on a set of probability 1, i.o. in k,

$$\begin{aligned}\left\|\sum_{i=1}^{n_{k+1}} f(X_i)\right\|_{\mathscr{F}} &\geq \left\|\sum_{i=n_k+1}^{n_{k+1}} f(X_i)\right\|_{\mathscr{F}} - \left\|\sum_{i=1}^{n_k} f(X_i)\right\|_{\mathscr{F}} \\ &> (1-\delta)^2\left(2\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}} LL\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}}\right)^{1/2} \\ &\quad - Ms_{n_k}t_{n_k} \\ &\geq (1-\delta)^2\left[2\left(s_{n_{k+1}}^2 - s_{n_k}^2\right)LL\left(s_{n_{k+1}}^2 - s_{n_k}^2\right)\right]^{1/2} - Ms_{n_k}t_{n_k},\end{aligned}$$

and, for large k, this lower bound behaves like

$$\left[(1-\delta)^2\left(1-\frac{1}{\rho^2}\right)^{1/2} - \frac{M}{\rho}\right]s_{n_{k+1}}t_{n_{k+1}}.$$

For ρ large enough and $\delta > 0$ arbitrary, this will therefore show that the lim sup is ≥ 1 a.s.; hence the conclusion. Let us prove (6) then. For each k, let f_k in $\mathscr{F}$ be such that

$$\sum_{i=n_k+1}^{n_{k+1}} Ef_k^2(X_i) \geq (1-\delta)\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}}.$$

Thus

$$\begin{aligned}\Pr&\left\{\left\|\sum_{i=n_k+1}^{n_{k+1}} f(X_i)\right\|_{\mathscr{F}} > (1-\delta)^2\left(2\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}} LL\left\|\sum_{i=n_k+1}^{n_{k+1}} Ef^2(X_i)\right\|_{\mathscr{F}}\right)^{1/2} \text{ i.o. in } k\right\} \\ &\geq \Pr\left\{\sum_{i=n_k+1}^{n_{k+1}} f_k(X_i) > (1-\delta)\left(2\sum_{i=n_k+1}^{n_{k+1}} Ef_k^2(X_i) LL \sum_{i=n_k+1}^{n_{k+1}} Ef_k^2(X_i)\right)^{1/2} \text{ i.o. in } k\right\}\end{aligned}$$

and Kolmogorov's exponential lower inequality implies then, as in ([9], page 271), that this last probability is equal to 1. □

To conclude this section, we would like to mention that *exactly* the *same* proof can be conducted to establish a generalization of Prokhorov's law of large numbers (see [19], page 276) to empirical processes. The result, presented in the next theorem, extends in particular the recent theorem of Alt [1], as well as previous results on this question [12, 15]. Compared with Theorem 10, the almost sure convergence to 0 (i) is obtained from the corresponding convergence in probability to 0 (ii). Necessity of (iii) has already been obtained in [1]. With the preceding notation, the result is the following.

THEOREM 13. *Let $\mathscr{F}$ be a class of functions. Assume that for each $i \in \mathbb{N}$, $\|Ef^2(X_i)\|_{\mathscr{F}} < \infty$ and $Ef(X_i) = 0$, $f \in \mathscr{F}$, and that, almost surely,*

$$\| f(X_i) \|_{\mathscr{F}} \le i/LLi.$$

Then

$$\text{(i)} \qquad \lim_{n\to\infty} \frac{1}{n} \left\| \sum_{i=1}^{n} f(X_i) \right\|_{\mathscr{F}} = 0 \quad \textit{almost surely}$$

if and only if

$$\text{(ii)} \qquad \lim_{n\to\infty} \frac{1}{n} \left\| \sum_{i=1}^{n} f(X_i) \right\|_{\mathscr{F}} = 0 \quad \textit{in probability}$$

and, for each $\varepsilon > 0$,

$$\text{(iii)} \qquad \sum_{n} \exp\left(-\varepsilon 2^{2n} \Big/ \left\| \sum_{i=2^n+1}^{2^{n+1}} Ef^2(X_i) \right\|_{\mathscr{F}} \right) < \infty.$$

4. Random geometry and empirical processes. Classes satisfying the CLT are necessarily pregaussian, that is, the limiting Gaussian process has bounded and continuous paths. This pregaussian character is of course of main importance in the study of the CLT since it provides a useful tool for an in-depth random geometric approach as was shown in [10, 11, 21]. In other limit theorems like the LIL however this necessary pregaussian structure is usually not available. It is therefore of some interest to try to find a pregaussian structure in more general situations in order to be able to use ideas on the CLT. This is the main purpose of this section with particular emphasis on the LIL. More precisely, the main result, Theorem 18, extends to the LIL the random geometric description of [21] for the CLT. The whole new point will be to produce some natural associated Gaussian structure and this will be accomplished with the basic tool of majorizing measures.

The work [22] has recently provided a full description of pregaussian classes. This description involves the so-called majorizing measures: A probability measure m on a metric space (T, d) is a majorizing measure if

$$\gamma_m(T, d) = \sup_{t \in T} \int_0^\infty \left(\log \frac{1}{m(B_d(t, \varepsilon))} \right)^{1/2} d\varepsilon < \infty,$$

where $B_d(t, \varepsilon)$ is the d-ball of center t and radius $\varepsilon > 0$ in T. Moreover, set

$$\gamma(T, d) = \inf \gamma_m(T, d),$$

where the infimum is taken over all probabilities m. This functional γ is the quantity describing pregaussian classes and we use it in the sequel to *measure* the size of pregaussian classes. It is actually completed with the functional α [22]: For any metric space (T, d), let

$$\alpha(T, d) = \inf\{\gamma(U, \delta); (U, \delta) \text{ is ultrametric and } T \text{ is the image of } U \text{ by a contraction}\}.$$

Recall that a metric space (U, δ) is called ultrametric if for u, v, w in U we have

$$\delta(u, w) \le \max(\delta(u, v), \delta(v, w)).$$

Recall also that a map φ from (U, δ) onto (T, d) is a contraction if $d(\varphi(u), \varphi(v)) \le \delta(u, v)$ for u, v in U.

Let now $\mathscr{F}$ be a totally bounded class in L_2 with d_2-diameter D. Set, for every $n \ge 1$,

$$M(n) = E\left\| \sum_{i=1}^n g_i f(X_i)/\sqrt{n} \right\|_{\mathscr{F}}.$$

Let n be fixed and set $l = n^4$. Set moreover

$$M = \max(M(l), D) = \max(M(n^4), D). \tag{7}$$

Our first main objective will basically be to show that one can always find a pregaussian class $\mathscr{G} \subset \mathscr{F}$ such that $\alpha(\mathscr{G}, d_2) \le KM$ and, moreover, close enough to $\mathscr{F}$ to control some portion of $\mathscr{F}$, namely satisfying $\mathscr{F} \subset KMn^{-1/2}W_1 + \mathscr{G}$, where W_1 is the unit ball of L_1. Here, as always, K denotes some positive numerical constant, not necessarily the same at each line in what follows. Also, for classes $\mathscr{F}_1, \mathscr{F}_2$ on S, we write $\mathscr{F}_1 + \mathscr{F}_2$ to denote the class of all functions $f_1 + f_2$, $f_1 \in \mathscr{F}_1$, $f_2 \in \mathscr{F}_2$. The proof will follow closely the works [22] and [23]. Although the necessary details are provided below, familiarity with these papers will certainly help the reader. Once this step has been accomplished, it is possible to derive a random geometric characterization of the convergence in probability to 0 of $\|(P_n - P)(f)\sqrt{n/2LLn}\|_{\mathscr{F}}$ which amounts, under necessary moment conditions, to the LIL (Theorem 9). The critical value $l = n^4$ will be absorbed in the LL factor of the normalization sequence.

The next thoerem describes the result alluded to above. Instead of d_2, we will actually work with the distance d with unit ball (in L_1) $n^{-1/2}W_1 + W_2$, that is, $d(f, g) \le t$ for f, g in L_1 if $f - g = h_1 + h_2$ where $\|h_1\|_1 \le n^{-1/2}t$ and $\|h_2\|_2 \le t$. We conjecture that a similar result is actually true with $l = n$ in the definition of M in (7).

THEOREM 14. *Let $\mathscr{F}$ be a totally bounded class in L_2 and M be as defined in (7). There exist a numerical constant K and a finite class $\mathscr{G} \subset \mathscr{F}$ such that*

$$\alpha(\mathscr{G}, d) \le KM \tag{8}$$

and

$$\mathscr{F} \subset KMn^{-1/2}W_1 + \mathscr{G}. \tag{9}$$

PROOF. Let $a > 0$ be specified later. Consider $\mathscr{G}$ finite in $\mathscr{F}$ such that $d_1(f, g) > aMn^{-1/2}$ for all $f \neq g$ in $\mathscr{G}$ and such that $\mathscr{F} \subset 2aMn^{-1/2}W_1 + \mathscr{G}$ where we recall that d_1 is the L_1 distance. The first lemma evaluates the size of $\mathscr{G}$ in terms of n and the parameter a.

LEMMA 15. *For every $\varepsilon > 0$ there exists $a(\varepsilon) > 0$ large enough depending on ε only such that if $a = a(\varepsilon)$ in the definition of $\mathscr{G}$, then*

$$\operatorname{Card} \mathscr{G} \le \exp(\varepsilon n). \tag{10}$$

PROOF. Suppose this is not the case for $a = a(\varepsilon) = \max(10(\varepsilon + 2), 9K/M\sqrt{\varepsilon})$ where K (see below) is the absolute constant which appears in Sudakov's minoration inequality for Gaussian processes (cf. [9; 22; 11, Section 1.4]). There exists $\mathscr{G}' \subset \mathscr{G}$ such that $\operatorname{Card} \mathscr{G}' = [e^{\varepsilon n}] + 1$. Let $f \neq g$ in $\mathscr{G}$ and define

$$h = (f - g)I_{\{|f-g| \le u\}}, \qquad h' = (f - g) - h,$$

where $u = 2M\sqrt{n}/a$. Clearly,

$$\|h'\|_1 \le \frac{\|f - g\|_2^2}{u} \le \frac{D^2}{u} \le \frac{M^2}{u} \le \frac{aM}{2\sqrt{n}}.$$

Hence it is necessary that $\|h\|_2 \ge \|h\|_1 \ge aM/2\sqrt{n}$ since by definition of $\mathscr{G}$, $d_1(f, g) > aM/\sqrt{n}$. We then have that

$$\begin{aligned}\Pr\{d_{l,2}(f, g) \le aM/4\sqrt{n}\} &\le \Pr\left\{\left(\left(\frac{1}{l}\right)\sum_{k=1}^{l} h^2(X_k)\right)^{1/2} \le \tfrac{1}{2}\|h\|_2\right\} \\ &\le \Pr\left\{\sum_{k=1}^{l}\left(-h^2(X_k) + Eh^2\right) \ge \tfrac{3}{4}l\|h\|_2^2\right\}.\end{aligned}$$

Therefore, by Bernstein's inequality [10], (2.18),

$$\begin{aligned}&\Pr\{d_{l,2}(f, g) \le aM/4\sqrt{n}\} \\ &\quad\le \exp\left(-9l^2\|h\|_2^4/16\left(2lEh^4 + lu^2\|h\|_2^2\right)\right) \\ &\quad\le \exp\left(-9l\|h\|_2^2/48u^2\right) \\ &\quad\le \exp(-10^{-2}n^2a^4).\end{aligned}$$

By definition of $a = a(\varepsilon)$ we have that

$$(\operatorname{Card} \mathscr{G}')\exp(-10^{-2}n^2a^4) \le \tfrac{1}{2}.$$

Thus

$$\Pr\{\forall f \neq g \text{ in } \mathscr{G}',\ d_{l,2}(f, g) > aM/4\sqrt{n}\} \ge \tfrac{1}{2}.$$

We are then in a position to apply Sudakov's minoration inequality conditionally on the X_i's in the definition of $M(l)$. We namely get

$$M \geq E\left\| \sum_{i=1}^{l} g_i f(X_i)/\sqrt{l} \right\|_{\mathscr{G}'}$$
$$\geq \frac{1}{2K} \frac{aM}{4\sqrt{n}} (\log \operatorname{Card} \mathscr{G}')^{1/2} \geq \frac{aM\sqrt{\varepsilon}}{8K},$$

which leads to a contradiction by the choice of $a = a(\varepsilon)$. Lemma 15 is proved. □

The next proposition performs the main step in the proof of Theorem 14. It clearly implies the conclusion of this theorem.

PROPOSITION 16. *There exists a numerical value a_0 of the parameter a in the definition of $\mathscr{G}$ such that $\alpha(\mathscr{G}, d) \leq KM$ for some numerical constant K.*

PROOF. Let (U, δ) be an ultrametric space. For x in U and i in $\mathbb{Z}$, let $N_i(x)$ be the number of disjoint balls of radius 6^{-i-1} that are contained in the ball $B(x, 6^{-i})$ of center x and radius 6^{-i}. Set then

$$\xi(U, \delta) = \inf_{x \in U} \sum_{i \in \mathbb{Z}} 6^{-i-1} (\log N_i(x))^{1/2}.$$

By Theorem 11 of [22], there is a numerical constant K and an ultrametric space (U, δ) such that

$$\delta(f, g) \leq d(f, g) \leq 36\delta(f, g) \quad \text{for all } f, g \text{ in } U \tag{11}$$

and satisfying

$$\alpha(\mathscr{G}, d) \leq K\xi(U, \delta). \tag{12}$$

Let λ be a probability measure on (U, δ). Given a ball B of radius 6^{-i} in (U, δ) for some fixed i, denote by N the number of disjoint balls of radius 6^{-i-1} contained in B. If B_1, B_2 are two such balls (provided there are), and if b, c are positive numbers, set

$$A(B_1, B_2, b, c) = \left\{ \omega \in \Omega: \lambda \otimes \lambda\big((f, g) \in B_1 \times B_2, d_{l,2}(\omega)(f, g) \leq b6^{-i-1}\big) \geq c\lambda(B_1)\lambda(B_2) \right\}.$$

Our aim in the first step of this proof will be to show that there is a subset A of Ω of large probability, say $\Pr(A) \geq \frac{1}{2}$, such that for any ω in A and each probability measure m on U we have

$$\gamma_m(U, d_{l,2}(\omega)) \geq K^{-1}\xi(U, \delta) \tag{13}$$

for some numerical constant K. A close inspection of the arguments of [23] shows that these can be extended to our setting to obtain (13) once it can be established [see (10) of [23] with $\alpha = \infty$] that, under the previous notation,

$$\Pr\Big(A\big(B_1, B_2, 6^{-1}, (2N)^{-2}\big)\Big) \leq (2N)^{-4}. \tag{14}$$

It is the purpose of the next lemma to describe how this inequality holds.

Lemma 17. *There is a numerical value a_0 of a such that inequality* (14) *is satisfied.*

Proof. Let i_0 be the smallest integer i satisfying $6^{-i+3} \le aMn^{-1/2}$. Then, for any $f \ne g$ in U ($\subset \mathscr{G}$), $\delta(f, g) > 6^{-i_0}$. Indeed, if $\delta(f, g) \le 6^{-i_0}$, then $d(f, g) \le 6^{-i_0+2}$ [by (11)] which means that $f - g = h_1 + h_2$ with

$$\|h_1\|_1 \le n^{-1/2}6^{-i_0+2}, \qquad \|h_2\|_2 \le 6^{-i_0+2}.$$

It follows that

$$d_1(f, g) \le (n^{-1/2} + 1)6^{-i_0+2} \le 6^{-i_0+3} \le aMn^{-1/2},$$

which is impossible by the very definition of $\mathscr{G}$. Hence the balls of radius 6^{-i_0} in (U, δ) are reduced to one point and (14) has only to be checked for $i < i_0$. We now proceed much as in the proof of Lemma 15. Let $f \in B_1$, $g \in B_2$. Then $\delta(f, g) > 6^{-i-1}$, which implies that

$$f \notin \{g\} + 6^{-i-1}n^{-1/2}W_1 + 6^{-i-1}W_2. \tag{15}$$

Set

$$h = (f - g)I_{\{|f-g| \le u\}}, \qquad h' = (f - g) - h,$$

with $u = 6^{i+1}M^2\sqrt{n}$ so that

$$\|h'\|_1 \le \frac{\|f - g\|_2^2}{u} \le \frac{M^2}{u} \le \frac{6^{-i-1}}{\sqrt{n}}.$$

Hence $\|h\|_2 > 6^{-i-1}$ by (15). As in the proof of Lemma 15 we then get from Bernstein's inequality that

$$\begin{aligned}\Pr\{d_{l,2}(f, g) \le 6^{-i-2}\} &\le \exp(-9l\|h\|_2^2/48u^2) \\ &\le \exp(-9n^3 6^{-4i-4}/48M^4).\end{aligned}$$

But now, since $i < i_0$,

$$6^{-i+3} \ge 6^{-i_0+3} \ge \frac{aM}{\sqrt{n}}$$

from which we get that

$$\Pr\{d_{l,2}(f, g) \le 6^{-i-2}\} \le \exp(-6^{-17}na^4).$$

We can now make use of the conclusion of Lemma 15 to see that for some numerical $a = a_0$,

$$\Pr\{d_{l,2}(f, g) \le 6^{-i-2}\} \le (2\,\mathrm{Card}\,\mathscr{G})^{-6} \le (2N)^{-6}, \tag{16}$$

where we have used that, trivially, $N \le \mathrm{Card}\,U \le \mathrm{Card}\,\mathscr{G}$.

We get from (16) the conclusion of the proof. Indeed, let $A(f, g) = \{\omega \in \Omega;\ d_{l,2}(\omega)(f, g) \le 6^{-i-2}\}$. (16) means

$$E(I_{A(f,g)}) = \Pr(A(f, g)) \le (2N)^{-6}.$$

Hence

$$E\left(\iint_{B_1 \times B_2} I_{A(f,g)}\, d\lambda(f)\, d\lambda(g)\right) \le (2N)^{-6}\lambda(B_1)\lambda(B_2)$$

and since

$$A\left(B_1, B_2, 6^{-1}, (2N)^{-2}\right)$$
$$= \left\{\omega: \iint_{B_1 \times B_2} I_{A(f,g)}(\omega)\, d\lambda(f)\, d\lambda(g) \ge (2N)^{-2}\lambda(B_1)\lambda(B_2)\right\},$$

it follows that

$$\Pr\left(A\left(B_1, B_2, 6^{-1}, (2N)^{-2}\right)\right) \le (2N)^2 (2N)^{-6} = (2N)^{-4}$$

and the lemma is proved. □

We can now conclude the proof of Proposition 16. We first have to use the main result of [22] conditionally on the X_i's: For each ω in Ω, there is a probability measure m_ω on $(U, d_{l,2}(\omega))$ such that

$$\gamma_{m_\omega}(U, d_{l,2}(\omega)) \le K E_g \left\| \sum_{i=1}^{l} g_i f(X_i(\omega))/\sqrt{l} \right\|_{\mathscr{F}}.$$

Using then (13), there is a set A in Ω, depending only on the X_i's, such that $\Pr(A) \ge \frac{1}{2}$ and for each ω in A,

$$\xi(U, \delta) \le K E_g \left\| \sum_{i=1}^{l} g_i f(X_i(\omega))/\sqrt{l} \right\|_{\mathscr{F}}.$$

Together with (12) we thus get that for ω in A,

$$\alpha(\mathscr{G}, d) \le K E_g \left\| \sum_{i=1}^{l} g_i f(X_i(\omega))/\sqrt{l} \right\|_{\mathscr{F}}.$$

To conclude the proof simply note that

$$M \ge E \left\| \sum_{i=1}^{l} g_i f(X_i)/\sqrt{l} \right\|_{\mathscr{F}}$$
$$\ge \int_A E_g \left\| \sum_{i=1}^{l} g_i f(X_i)/\sqrt{l} \right\|_{\mathscr{F}} dP_X \ge K^{-1}\Pr(A)\alpha(\mathscr{G}, d). \qquad \square$$

Having described a natural pregaussian structure associated to any totally bounded class in L_2, we are now in a position to make use of the ideas developed for pregaussian classes. These have been studied previously in the context of the CLT [10, 11, 21] as we presented it in the introduction. Our aim will now be to set these results together with the preceding conclusion and apply it to the LIL to get a characterization of the behavior of $\|I_n(f)\|_{\mathscr{F}}$ along a class of functions through random geometric conditions. Let us recall that, under necessary moment conditions, Theorem 9 has reduced the LIL to this behavior in probability, a statement similar therefore to the CLT. Further, $\|I_n(f)\|_{\mathscr{F}} \to 0$ in probability if and only if

$$\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{2nLLn}\right\|_{\mathscr{F}} \to 0 \quad \text{or} \quad \left\|\sum_{i=1}^{n} g_i f(X_i)/\sqrt{2nLLn}\right\|_{\mathscr{F}} \to 0$$

in probability or in L_1 (see [16], Section 2), which justifies the formulation of Theorem 18 below.

To ease the notation, we say in the following that a metric space (T', d') is k-Lipschitz to another metric space (T, d) if T' is the image of T by a map φ such that

$$d'(\varphi(s), \varphi(t)) \le kd(s,t) \quad \text{for all } s, t \text{ in } T.$$

These relations trivially imply that $\gamma(T', d') \le K\gamma(T, d)$.

Theorem 18. *Let $\mathscr{F}$ be totally bounded in L_2 with d_2-diameter D and let n be a fixed integer. Let also $u > 0$ be such that*

$$u \ge \sup_{l \ge n} E\left\|\sum_{i=1}^{l} g_i f(X_i)/\sqrt{2lLLl}\right\|_{\mathscr{F}}$$

and $u \ge D(LLn)^{-1/2}$. Then there exist $\mathscr{G} \subset \mathscr{F}$, an ultrametric structure δ on $\mathscr{G}$ and classes $\mathscr{F}_1$, $\mathscr{F}_2$ in L_2 (depending on n) such that

(i) $$\gamma(G, \delta) \le Ku\sqrt{LLn} \quad \textit{and} \quad \mathscr{F} \subset \mathscr{F}_1 + \mathscr{F}_2$$

with the following properties: For each $\varepsilon > 0$ there is $k = k(\varepsilon)$ such that

(ii) $$\Pr\left\{(\mathscr{F}_2, d_{n,2}) \textit{ is k-Lipschitz to } (\mathscr{G}, \delta) \quad \textit{and} \quad \sup_{f \in \mathscr{F}_2} \|f\|_{n,2} \le ku\sqrt{LLn}\right\} \ge 1 - \varepsilon$$

and

(iii) $$\Pr\left\{\left\|\sum_{i=1}^{n} |f(X_i)|/\sqrt{2nLLn}\right\|_{\mathscr{F}_1} \le ku\right\} \ge 1 - \varepsilon.$$

Conversely, if such a decomposition of $\mathscr{F}$ exists satisfying (i), (ii) and (iii) for some $u > 0$, then, for each $\varepsilon > 0$, there is a $k = k(\varepsilon)$ such that

(iv) $$\Pr\left\{\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{2nLLn}\right\|_{\mathscr{F}} \le ku\right\} \ge 1 - \varepsilon.$$

Proof. It is instructive to prove first the converse assertion: If $\mathscr{F} \subset \mathscr{F}_1 + \mathscr{F}_2$,

$$\left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)\right\|_{\mathscr{F}} \le \left\|\sum_{i=1}^{n} |f(X_i)|\right\|_{\mathscr{F}_1} + \left\|\sum_{i=1}^{n} \varepsilon_i f(X_i)\right\|_{\mathscr{F}_2},$$

so that by (iii), it is enough to show (iv) for $\mathscr{F}_2$ instead of $\mathscr{F}$. But, on the set of ω's for which $(\mathscr{F}_2, d_{n,2}(\omega))$ is k-Lipschitz to $(\mathscr{G}, \delta)$ and $\sup_{f \in \mathscr{F}_2} \|f\|_{n,2}(\omega) \le$

$ku\sqrt{LLn}$, by the majorizing measure bound (see [9; 22; 11, Section 1.4]),

$$E_\varepsilon\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i(\omega))/\sqrt{n} \right\|_{\mathscr{F}_2} \le K\Big(\sup_{f\in\mathscr{F}_2} \| f \|_{n,2}(\omega) + \gamma(\mathscr{F}_2, d_{n,2}(\omega))\Big)$$
$$\le k'u\sqrt{LLn}$$

for some k' and the conclusion follows.

We turn to the direct part. By Theorem 14, each f in $\mathscr{F}$ can be written as a sum $f = g + h$ where

$$\|h\|_1 \le KMn^{-1/2} \le Kun^{-1/2}\sqrt{2LLn^4} \le Ku(LLn/n)^{1/2} \tag{17}$$

and g belongs to a finite class $\mathscr{G} \subset \mathscr{F}$ such that

$$\alpha(\mathscr{G}, d) \le KM \le Ku\sqrt{LLn}\,.$$

Note here the smoothing properties of the LL function which imply that $LLn^4 \sim LLn$. Also K is, as always, a numerical constant not always the same each time it appears. We also know from the definition of α and [22] that there is an ultrametric distance $\delta \ge d$ on $\mathscr{G}$ such that $\gamma(\mathscr{G}, \delta) \le Ku\sqrt{LLn}$. Let q_0 be the largest element of $\mathbb{Z}$ such that $2^{-q_0} \ge \operatorname{diam}(\mathscr{G}, \delta)$ and denote by $\mathscr{B}_q$ the family of all balls of radius 2^{-q} of $(\mathscr{G}, \delta)$, $q \ge q_0$. For each B in $\mathscr{B}_q$, fix $x(B) \in B$. Define $\pi_q g = x(B(g, 2^{-q}))$ for g in $\mathscr{G}$. Consider now a majorizing measure μ on $\mathscr{G}$ such that $\gamma_\mu(\mathscr{G}, \delta) \le Ku\sqrt{LLn}$ and let

$$m = \sum_{q=q_0}^{\infty} 2^{q_0-q-1} \sum_{B\in\mathscr{B}_q} \delta_{x(B)}\mu(B).$$

It is easily seen that

$$\sup_{g\in\mathscr{G}} \sum_{q=q_0}^{\infty} 2^{-q}\left(\log\frac{1}{m(\{\pi_q g\})}\right)^{1/2} \le Ku\sqrt{LLn}\,. \tag{18}$$

Moreover $d(g, \pi_q g) \le \delta(g, \pi_q g) \le 2^{-q}$ for all g and q and note that $\pi_{q_0} g$ can be chosen independently of g so that we denote it simply by π_{q_0}. Since $\operatorname{diam}(\mathscr{G}, \delta) \le K\gamma(\mathscr{G}, \delta)$, we also have that $2^{-q_0} \le Ku\sqrt{LLn}$.

Once this has been obtained, we write the usual chaining of an element g of $\mathscr{G}$ so that any f in $\mathscr{F}$ can be written as

$$f = h + \pi_{q_0} + \sum_{q=q_0+1}^{\infty} (\pi_q g - \pi_{q-1} g).$$

Define now, for g in $\mathscr{G}$ and $q > q_0$,

$$a_{g,q} = \sqrt{n}\,2^{-q}\left(\log\frac{2^{q-q_0}}{m(\{\pi_{q-1}g\})m(\{\pi_q g\})}\right)^{-1/2}.$$

Since $d(\pi_q g, \pi_{q-1} g) \le 3.2^{-q}$, we have $\pi_q g - \pi_{q-1} g = v(g, q) + w(g, q)$ where

$\|v(g,q)\|_1 \le n^{-1/2}3.2^{-q}$ and $\|w(g,q)\|_2 \le 3.2^{-q}$. We then write

$$\begin{aligned} f &= \left(h + \sum_{q=q_0+1}^{\infty} w(g,q) I_{\{|w(g,q)| > a_{g,q}\}} + \sum_{q=q_0+1}^{\infty} v(g,q)\right) \\ &\quad + \left(\pi_{q_0} + \sum_{q=q_0+1}^{\infty} w(g,q) I_{\{|w(g,q)| \le a_{g,q}\}}\right) \\ &= f_1 + f_2 \end{aligned} \tag{19}$$

and let $\mathscr{F}_1 = \{f_1;\ f \in \mathscr{F}\}$ and $\mathscr{F}_2 = \{f_2;\ f \in \mathscr{F}\}$ with the obvious abuse in notation. $\mathscr{F}_1$ and $\mathscr{F}_2$ are the classes of the expected decomposition and we would like to show that they satisfy (iii) and (ii) respectively.

We start with (ii). It is easy to see that this property will be satisfied once we have shown that for every $\varepsilon > 0$ there is a $k = k(\varepsilon)$ such that

$$\Pr\left\{\forall\, q > q_0, \forall\, g \in \mathscr{G},\ \|w(g,q) I_{\{|w(g,q)| \le a_{g,q}\}}\|_{n,2} \le k 2^{-q}\right\} \ge 1 - \varepsilon. \tag{20}$$

Indeed, by definition of the family $\{\pi_q g;\ g \in \mathscr{G},\ q > q_0\}$, $\delta(g,g') \ge 2^{-q-1}$ if $q = \sup\{l:\ \pi_j g = \pi_j g',\ \forall\, j \le l\}$. Hence, under (20), on a set of large probability, $(\mathscr{F}_2, d_{n,2})$ is $4k$-Lipschitz to $(\mathscr{G}, \delta)$. $\sup_{f \in \mathscr{F}} \|f\|_{n,2}$ is also estimated by (20) apart perhaps from π_{q_0}. But as an element of $\mathscr{G} \subset \mathscr{F}$, the definition of u controls by itself $\|\pi_{q_0}\|_{n,2}$. We establish (20) using Bernstein's inequality: Since $\|w(g,q)\|_2 \le 3.2^{-q}$, recentering, this inequality yields, for each g in $\mathscr{G}$ and $q > q_0$,

$$\Pr\left\{\|w(g,q) I_{\{|w(g,q)| \le a_{g,q}\}}\|_{n,2} > k 2^{-q}\right\} \le \exp\left(-kn 2^{-2q} a_{g,q}^{-2}\right)$$

for $k \ge k_0$ large enough. By the choice of $a_{g,q}$, this probability is estimated by

$$\exp\left(-k \log\left(\frac{2^{q-q_0}}{m(\{\pi_{q-1}g\}) m(\{\pi_q g\})}\right)\right).$$

For each $\varepsilon > 0$, there is a $k = k(\varepsilon)$ sufficiently large such that, for all g and $q > q_0$,

$$\exp\left(-k \log\left(\frac{2^{q-q_0}}{m(\{\pi_{q-1}g\}) m(\{\pi_q g\})}\right)\right) \le \varepsilon 2^{q_0 - q} m(\{\pi_{q-1}g\}) m(\{\pi_q g\}).$$

Hence the probability of the complement of the event in (20) is smaller than

$$\sum_{q=q_0+1}^{\infty} \sum_{\{\pi_{q-1}g\}} \sum_{\{\pi_q g\}} \varepsilon 2^{q_0 - q} m(\{\pi_{q-1}g\}) m(\{\pi_q g\}),$$

which is less than ε since m is a probability. This shows (ii).

The main observation to establish (iii) is that for each g and $q > q_0$,

$$\begin{aligned} \left\| w(g,q) I_{\{|w(g,q)| > a_{g,q}\}} \right\|_1 &\le E|w(g,q)|^2 a_{g,q}^{-1} \\ &\le 9 n^{-1/2} 2^{-q} \left(\log \frac{2^{q-q_0}}{m(\{\pi_{q-1}g\}) m(\{\pi_q g\})}\right)^{1/2}. \end{aligned}$$

Therefore, by (18) and the fact that

$$\|h\|_1 + \sum_{q=q_0+1}^{\infty} \|v(g,q)\|_1 \le Ku(LLn/n)^{1/2} + 3.2^{-q_0}n^{-1/2}$$
$$\le Ku(LLn/n)^{1/2}$$

in decomposition (19), each f in $\mathscr{F}_1$ satisfies

$$E|f| = \|f\|_1 \le K_0 u(LLn/n)^{1/2}$$

for some numerical constant K_0. It is then easy to see how (iii) holds. Since $\mathscr{F}_1 \subset \mathscr{F} - \mathscr{F}_2$, for each $\varepsilon > 0$ there is a $k = k(\varepsilon)$ such that

$$P_X\left\{ E_\varepsilon \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{2nLLn} \right\|_{\mathscr{F}_1} > ku \right\} \le \varepsilon.$$

We used here the information we already have on $\mathscr{F}_2$ (see the above proof of the converse portion) and the fact that Gaussian averages dominate Rademacher averages. It follows then from the comparison properties of Section 2 (Proposition 1) and elementary arguments that

$$\Pr\left\{ \left\| \sum_{i=1}^{n} \varepsilon_i |f(X_i)|/\sqrt{2nLLn} \right\|_{\mathscr{F}_1} > \frac{2k}{\varepsilon} u \right\} \le 2\varepsilon.$$

Since $E|f| \le K_0 u(LLn/n)^{1/2}$ for all f in $\mathscr{F}_1$,

$$\sup_{f \in \mathscr{F}_1} \Pr\left\{ \left| \sum_{i=1}^{n} (|f(X_i)| - E|f|)/\sqrt{2nLLn} \right| > \frac{2k}{\varepsilon} u \right\} \le \varepsilon$$

whenever k is large enough to be greater than K_0. It follows then from the symmetrization lemma (Lemma 6) that

$$\Pr\left\{ \left\| \sum_{i=1}^{n} (|f(X_i)| - E|f|)/\sqrt{2nLLn} \right\|_{\mathscr{F}_1} > \frac{6k}{\varepsilon} u \right\} \le 5\varepsilon$$

for k large enough. Using one more time that $E|f| \le K_0 u(LLn/n)^{1/2}$ for $f \in \mathscr{F}_1$ yields then the conclusion. The proof of Theorem 18 is complete. □

The next corollary summarizes in the limit the description obtained in Theorem 18 and is ready for possible concrete applications. Recall that $I_n(f) = (P_n - P)(f)\sqrt{n/2LLn}$.

COROLLARY 19. *Let $\mathscr{F}$ be totally bounded in L_2. In order that $\|I_n(f)\|_{\mathscr{F}} \to 0$ in probability, it is necessary and sufficient that for each n, $\mathscr{F}$ can be decomposed in $\mathscr{F} \subset \mathscr{F}_1 + \mathscr{F}_2$, where $\mathscr{F}_1$ and $\mathscr{F}_2$ are classes in L_2 depending on n and satisfying*

$$\left\| \sum_{i=1}^{n} |f(X_i)|/\sqrt{2nLLn} \right\|_{\mathscr{F}_1} \to 0 \quad \textit{in probability}$$

and

$$E_g\left\| \sum_{i=1}^{n} g_i f(X_i)/\sqrt{2nLLn} \right\|_{\mathscr{F}_2} \to 0 \quad \textit{in probability}$$

$$\big(\textit{or } \gamma(\mathscr{F}_2, d_{n,2})/\sqrt{2LLn} \to 0 \quad \textit{in probability}\big).$$

5. Bracketing and local Lipschitz conditions. In the recent paper [2] (see also [3]), several sharp nonrandom conditions have been obtained for the CLT and the LIL for empirical processes. These conditions, called bracketing (cf. [7, 17]) or local Lipschitz conditions [2], have been basically introduced to control classes under the $L_1(P_n)$ norm, which appeared to be crucial in what we presented in the preceding section. We therefore present these results and their proofs following [2] as a natural continuation of the preceding discussion in the form of a general statement (Theorem 20 below) that collects the main ideas and conclusions. The somewhat technical formulation of this theorem is adapted to the next corollaries, in particular, (iii) is the local Lipschitz condition and (i) the majorizing measure condition used to control the $L_2(P_n)$ portion.

THEOREM 20. *Let $\mathscr{F}$ be a class of functions and d a distance on $\mathscr{F}$. Let also n be a fixed integer, $s \geq 1$, $v > 0$ numbers (to depend on n in applications). Let q_0 be such that $2^{-q_0} \leq s$ and denote by q_1 the largest integer such that $2^{-q_1} \geq sn^{-1/2}$. We assume the following conditions hold:*

(i) *There exist a subset $\mathscr{G}$ of $\mathscr{F}$ such that $d(f, \mathscr{G}) \leq 2^{-q_1}$ for each $f \in \mathscr{F}$ and a probability measure m on $\mathscr{F}$ such that*

$$\sup_{g\in\mathscr{G}} \sum_{q=q_0}^{q_1} 2^{-q}\left(\log\frac{1}{m(B_d(g, 2^{-q}))}\right)^{1/2} \leq s.$$

(ii) *For all g, g' in $\mathscr{G}$,*

$$\left\| \frac{|g-g'|}{d(g,g')} \wedge v\sqrt{n} \right\|_2 \leq v.$$

(iii) *For each g in $\mathscr{G}$, $\alpha > 0$ and $t \leq v\sqrt{n}$,*

$$P\left\{\frac{1}{\alpha}M(g,\alpha) > t\right\} \leq \frac{v^2}{t^2},$$

where $M(g, \alpha) = \sup\{|f-g|;\ f \in \mathscr{F},\ d(f,g) \leq \alpha\}$ in L_0.

Then there exist a finite subset $\{\pi_{q_0} f;\ f \in \mathscr{F}\}$ of $\mathscr{F}$ and for each $\varepsilon > 0$ a number $k = k(\varepsilon)$ such that

$$\Pr\left\{ \left\| \sum_{i=1}^{n} \varepsilon_i (f - \pi_{q_0} f)(X_i)/\sqrt{n} \right\|_{\mathscr{F}} \leq 2kvs \right\}$$

$$\geq 1 - \varepsilon - \Pr\left\{ 2\sum_{i=1}^{n} \|f(X_i)\|_{\mathscr{F}} I_{\{\|f(X_i)\|_{\mathscr{F}} > c\}}/\sqrt{n} > kvs \right\},$$

where $c = v\sqrt{n}\,2^{-q_0-1}(\log 3N(\mathscr{G}, d;\ 2^{-q_0-r_0}))^{1/2}$. *Here* $N(\mathscr{G}, d;\ \alpha)$ *is the smallest number of d-balls of radius* $\alpha > 0$ *that cover* $\mathscr{G}$ *and* r_0 *is some fixed integer.*

PROOF. The first step in the proof deals with condition (i) that we modify in an appropriate way for its later use. Let $\mathscr{G}_1$ be maximal in $\mathscr{G}$ satisfying $d(g, g') > 2^{-q_1+1}$ for $g, g' \in \mathscr{G}_1$. Let $\varphi\colon \mathscr{F} \to \mathscr{G}_1$ be such that $d(f, \varphi(f)) = d(f, \mathscr{G}_1)$, $f \in \mathscr{F}$, and denote $m' = \varphi(m)$. Since $m'(B_d(g, 2^{-q_1})) \geq m(B_d(g, 2^{-q_1}))$ and $\varphi(B_d(g, 2^{-q})) \subset B_d(g, 2^{-q+1})$ for $q < q_1$, it follows from (i) that

$$\sup_{g \in \mathscr{G}_1} \sum_{q=q_0}^{q_1} 2^{-q}\left(\log \frac{1}{m'(B_d(g, 2^{-q}))}\right)^{1/2} \leq 3s.$$

We discretize this majoring measure condition by Lemma 2.1 of [2]: There exists a family $\{\pi_q g;\ g \in \mathscr{G}_1,\ q_0 \leq q \leq q_1\}$ such that $d(g, \pi_q g) \leq 2^{-q}$ for all g and q and a discrete probability measure μ such that

$$\sup_{g \in \mathscr{G}_1} \sum_{q=q_0}^{q_1} 2^{-q}\left(\log \frac{1}{\mu(\{\pi_q g\})}\right)^{1/2} \leq Ks,$$

where K is a numerical constant (possibly changing from line to line in what follows). The construction of [2] in one of its various versions (see Remarks 2.2 in [2]) shows that it is possible to choose the π_q's such that $\pi_q \circ \pi_r = \pi_{q \wedge r}$ and, moreover, as follows from this construction, such that $\operatorname{Card}\{\pi_q g;\ g \in \mathscr{G}_1\} \leq N(\mathscr{G}, d; 2^{-q-r_0+1})$ (for some appropriate fixed integer r_0). We extend the π_q's to $\mathscr{F}$ by setting $\pi_q f = \pi_q \varphi(f)$. Therefore, $d(f, \pi_q f) \leq 2^{-q+1}$ for all f and q. We set finally $\nu = \frac{1}{2}\mu + \frac{1}{2}\lambda$ where λ is uniformly distributed on $\{\pi_{q_0+1} f;\ f \in \mathscr{F}\}$ and note that we still have

$$\sup_{f \in \mathscr{F}} \sum_{q=q_0}^{q_1} 2^{-q}\left(\log \frac{1}{\nu(\{\pi_q f\})}\right)^{1/2} \leq Ks. \tag{21}$$

After these preliminaries, we can now start the proof of the theorem. We first decompose each f in $\mathscr{F}$ along a partition of S in the following way: Define

$$a_{f,q} = v\sqrt{n}\,2^{-q+1}\left(\log \frac{2^{q-q_0}}{\nu(\{\pi_{q-1} f\})\nu(\{\pi_q f\})}\right)^{-1/2}, \qquad q_0 < q \leq q_1,$$

$$a_{f,q_1+1} = vsn^{-1/2}.$$

Set then, for $q_0 < q \leq q_1$,

$$B_{f,q} = \left\{\forall\, l = q_0+1, \ldots, q;\ M(\pi_{l-1} f, 6.2^{-l+1}) \leq a_{f,l}\right\},$$

$$A_{f,q} = B_{f,q} \setminus B_{f,q+1} (q_0 < q < q_1),\ A_{f,q_1} = B_{f,q_1}.$$

Clearly $\{A_{f,q};\ q_0 < q \leq q_1\}$ defines a partition of B_{f,q_0+1}. On this set, since

$\{B_{f,q};\ q_0 < q \le q_1\}$ is decreasing, we have

$$(22)\qquad f - \pi_{q_0 f} = \sum_{q=q_0+1}^{q_1} (f - \pi_q f) I_{A_{f,q}} + \sum_{q=q_0+1}^{q_1} (\pi_q f - \pi_{q-1} f) I_{B_{f,q}} = f_1 + f_2.$$

For each f, $B^c_{f,q_0+1} \subset \{2\|f\|_{\mathscr{F}} > \inf_{f\in\mathscr{F}} a_{f,q_0+1}\}$. By definition of ν and λ,

$$\begin{aligned} a_{f,q_0+1} &\ge v\sqrt{n}\,2^{-q_0}\left(\log\frac{8}{\lambda(\{\pi_{q_0 f}\})\lambda(\{\pi_{q_0+1}f\})}\right)^{-1/2} \\ &\ge v\sqrt{n}\,2^{-q_0}\left(\log 8N^2(\mathscr{G}, d; 2^{-q_0-r_0})\right)^{-1/2}. \end{aligned}$$

From this observation and decomposition (22), it will thus be enough, in order to establish the theorem, to prove that for each $\varepsilon > 0$, there is a $k = k(\varepsilon)$ such that

$$\Pr\left\{\left\|\sum_{i=1}^{n} \varepsilon_i f_j(X_i)/\sqrt{n}\right\|_{\mathscr{F}} \le kvs\right\} \ge 1 - \varepsilon, \qquad j = 1,2.$$

The main idea of the proof will be to study the terms f_2 like class $\mathscr{F}_2$ in Theorem 18 using the majorizing measure condition (i), and the f_1's through assumption (iii) for which the summation by parts (22) has provided the right form.

We first study the terms f_2. First note that $d(\pi_q f, \pi_{q-1} f) \le 6.2^{-q}$, so that, on $B_{f,q}$, $|\pi_q f - \pi_{q-1} f| \le a_{f,q} \le v\sqrt{n}\,2^{-q}$, $q_0 < q \le q_1$. Therefore, by (ii), for each q such that $q_0 < q \le q_1$,

$$(23)\qquad \begin{aligned} \left\|(\pi_q f - \pi_{q-1} f) I_{B_{f,q}}\right\|_2 &\le 6.2^{-q}\left\|\frac{|\pi_q f - \pi_{q-1} f|}{6.2^{-q}} \wedge v\sqrt{n}\right\|_2 \\ &\le 6.2^{-q}\left\|\frac{|\pi_q f - \pi_{q-1} f|}{d(\pi_q f, \pi_{q-1} f)} \wedge v\sqrt{n}\right\|_2 \\ &\le 6.2^{-q} v. \end{aligned}$$

Let now

$$B = \left\{\forall\, q_0 < q \le q_1, \forall\, f \in \mathscr{F}, \left\|(\pi_q f - \pi_{q-1} f) I_{B_{f,q}}\right\|_{n,2} \le kv2^{-q}\right\}.$$

By (23), we are in a position to apply Bernstein's inequality [10], (2.18) to get that for each f and q,

$$\Pr\left\{\left\|(\pi_q f - \pi_{q-1} f) I_{B_{f,q}}\right\|_{n,2} > kv2^{-q}\right\} \le \exp\left(-knv^2 2^{-2q} a_{f,q}^{-2}\right)$$

for $k \ge k_0$ large enough. Thus, by definition of $a_{f,q}$, for $k = k(\varepsilon)$ large enough depending on $\varepsilon > 0$, the preceding probability is smaller than

$$\exp\left(-k\log\frac{2^{q-q_0}}{\nu(\{\pi_{q-1} f\})\nu(\{\pi_q f\})}\right) \le \varepsilon 2^{q_0-q-1}\nu(\{\pi_{q-1} f\})\nu(\{\pi_q f\}).$$

Since ν is a probability, it follows that

$$(24)\quad \Pr(B^c) \le \sum_{q=q_0+1}^{q_1} \sum_{\{\pi_{q-1} f\}} \sum_{\{\pi_q f\}} \varepsilon 2^{q_0-q-1}\nu(\{\pi_{q-1} f\})\nu(\{\pi_q f\}) \le \varepsilon/2.$$

Now if $U = \{\pi_q f;\ f \in \mathscr{F}\}$ is equipped with the ultrametric structure

$$\delta(\pi_q f, \pi_q f') = 2^{-q} \quad \text{if} \quad q = \sup\{l;\ \pi_l f = \pi_l f'\},$$

which defines a metric by the composition properties $\pi_q \circ \pi_r = \pi_{q \wedge r}$, it follows from (21) that $\gamma(U, \delta) \le Ks$. On the set B, the class of all f_2's, $f \in \mathscr{F}$, is $2kv$-Lipschitz to (U, δ). Hence, by the majorizing measure bound (cf. [9; 23; 11, Section 1.4]), it follows that

$$P_\varepsilon\left\{\left\| \sum_{i=1}^{n} \varepsilon_i f_2(X_i)/\sqrt{n} \right\|_{\mathscr{F}} \le kvs \right\} \ge 1 - \varepsilon/2$$

for $k = k(\varepsilon)$ sufficiently large, which, together with (24), completes the proof concerning the f_2's.

Let us now come to the terms $f_1 = \sum_{q=q_0+1}^{q_1}(f - \pi_q f)I_{A_{f,q}}$ of the decomposition (22). Note that on $A_{f,q}$, $q_0 < q < q_1$,

$$M_{q,f} \equiv M(\pi_q f, 6.2^{-q}) \ge a_{f,q+1}.$$

On the other hand, since $M_{q,f} \le 2M_{q-1,f}$, $q_0 < q < q_1$, we also have that, on $A_{f,q}$, $q_0 < q \le q_1$, $M_{q,f} \le 2a_{f,q}$. Hence, in particular $M_{q,f} \le a'_{f,q}$, where

$$a'_{f,q} = v\sqrt{n}\,2^{-q}\left(\log \frac{2^{q-q_0}}{\nu(\{\pi_q f\})}\right)^{-1/2}.$$

To treat the terms f_1, we form the following crude estimate:

$$\begin{aligned} f_1 &= \sum_{q=q_0+1}^{q_1} (f - \pi_q f) I_{A_{f,q}} \\ &\le \sum_{q=q_0+1}^{q_1} M_{q,f} I_{\{a_{f,q+1} < M_{q,f} \le a'_{f,q}\}} + a_{f,q_1+1} \\ &= f_3 + vsn^{-1/2}. \end{aligned}$$

We would like to show that, with large probability,

$$\left\| \sum_{i=1}^{n} f_3(X_i)/\sqrt{n} \right\|_{\mathscr{F}} \le kvs, \tag{25}$$

from which the conclusion concerning the f_1's clearly follows from the preceding inequality. To establish (25), we estimate the probability of the set

$$A = \left\{\forall\, q_0 < q \le q_1, \forall\, f \in \mathscr{F},\ \sup_{t>0} t \sum_{i=1}^{n} M_{q,f} I_{\{t < M_{q,f} \le a'_{f,q}\}}(X_i) \le knv^2 2^{-2q}\right\}.$$

If we can prove that for every $\varepsilon > 0$, there is a $k = k(\varepsilon)$ such that $\Pr(A) \ge 1 - \varepsilon$, (25) will hold since, on A, for each f,

$$\sum_{i=1}^{n} f_3(X_i)/\sqrt{n} \le \sum_{q=q_0+1}^{q_1} k\sqrt{n}\, v^2 2^{-2q} a_{f,q+1}^{-1} \le 16kvs.$$

$\Pr(A)$ has been estimated in ([2], proof of Lemma 7.16). Here we use a somewhat different method. We write, for each f and q,

$$\Pr\left\{\sup_{t>0} t\sum_{i=1}^{n} M_{q,f} I_{\{t<M_{f,q}\le a'_{f,q}\}}(X_i) > knv^2 2^{-2q}\right\}$$
$$\le \sum_j \Pr\left\{\sup_{2^{-j-1}\le t\le 2^{-j}} t\sum_{i=1}^{n} M_{q,f} I_{\{t<M_{q,f}\le a'_{f,q}\}}(X_i) > knv^2 2^{-2q}\right\}$$
$$\le \sum_j \Pr\left\{\sum_{i=1}^{n} M_{q,f} I_{\{2^{-j-1}<M_{q,f}\le a'_{f,q}\}}(X_i) > knv^2 2^{-2q} 2^{j}\right\},$$

where the sum is extended over all $j \in \mathbb{Z}$ such that $2^{-j} \le a'_{f,q}$. To estimate the preceding probabilities, we use Bernstein's inequality. By (iii), the centerings are nicely controlled:

$$E\left(M_{q,f} I_{\{2^{-j-1}<M_{q,f}\le a'_{f,q}\}}\right) = \int_0^{a'_{f,q}} \Pr\left\{M_{q,f} > t,\, M_{q,f} > 2^{-j-1}\right\} dt \le 144 v^2 2^{-2q} 2^{j}.$$

Hence, for $k \ge k_0$ large enough,

$$\Pr\left\{\sum_{i=1}^{n} M_{q,f} I_{\{2^{-j-1}<M_{q,f}\le a'_{f,q}\}}(X_i) > knv^2 2^{-2q_2 j}\right\} \le \exp\left(-knv^2 2^{-2q} 2^{j} a'^{-1}_{f,q}\right).$$

It is now easy to see, again for $k \ge k_0$ large enough, that

$$\sum_{j:\, 2^{-j}\le a'_{f,q}} \exp\left(-knv^2 2^{-2q} 2^{j} a'^{-1}_{f,q}\right) \le 2\exp\left(-2knv^2 2^{-2q} a'^{-2}_{f,q}\right).$$

Thus, using that $M_{q,f}$ only depends on $\pi_q f$, we obtain that, for each $\varepsilon > 0$, there is a $k = k(\varepsilon)$ such that

$$\Pr(A^c) \le 2 \sum_{q=q_0+1}^{q_1} \sum_{\{\pi_q f\}} \exp\left(-2knv^2 2^{-2q} a'^{-2}_{f,q}\right)$$
$$\le \varepsilon \sum_{q=q_0+1}^{q_1} \sum_{\{\pi_q f\}} 2^{q_0-q} \nu\left(\{\pi_q f\}\right) \le \varepsilon.$$

The proof of Theorem 20 is complete. □

The main interest of Theorem 20 lies in applications for which one can easily play on the various parameters involved in its statement. For example, the (normal) CLT and LIL under local Lipschitz conditions of [2, 3] are contained in this result. We do not detail here the CLT which is basically obtained by taking $s = v = 1$, but rather present the illustration concerning the LIL which reduces, thanks to the characterization of Theorem 9, to the convergence in probability to 0 of $\|I_n(f)\|_{\mathscr{F}}$. The next two statements of [2] and [3] (Theorem 21 below is actually somewhat stronger, up to measurability, than the corresponding result in [2]) describe sufficient conditions for this to hold. Note that, as stated, Theorem 21 also contains a result of [4], as was communicated to us by E. Giné and J. Zinn.

THEOREM 21 [2]. *Let $\mathscr{F}$ be a class in L_2 such that there exists a probability measure m on $\mathscr{F}$ satisfying*

$$\text{(i)} \qquad \lim_{\varepsilon\to 0} \sup_{f\in\mathscr{F}} \sup_{0<\delta<\varepsilon} \left(LL\frac{1}{\delta}\right)^{-1/2} \int_\delta^\varepsilon \left(\log\frac{1}{m(B_{d_2}(f,t))}\right)^{1/2} dt = 0.$$

Assume further that there exists another metric ρ on $\mathscr{F}$ satisfying the same property (i) and such that, moreover, for every α, $t > 0$ and g in $\mathscr{F}$

$$\text{(ii)} \qquad P\left\{\frac{1}{\alpha}\sup\{|f-g|;\ f\in\mathscr{F}, \rho(f,g)\le\alpha\} > t\right\} \le \frac{1}{t^2}.$$

Then $\|I_n(f)\|_{\mathscr{F}} \to 0$ in probability.

THEOREM 22 [3]. *Let $\mathscr{F}$ be a class in L_2 such that*

$$\lim_{t\to\infty} (t^2/LLt)P(\|f\|_{\mathscr{F}} > t) = 0.$$

Assume there is a metric ρ on $\mathscr{F}$ satisying the following properties:

(i) *There is a probability measure m on $\mathscr{F}$ such that*

$$\lim_{\varepsilon\to 0}\sup_{f\in\mathscr{F}} \int_0^\varepsilon \left(\log\frac{1}{m(B_\rho(f,t))}\right)^{1/2} dt = 0.$$

$$\text{(ii)} \qquad \sup_{f,g\in\mathscr{F}} E\left(\frac{|f-g|^2}{\rho(f,g)^2}\Big/ LL\frac{|f-g|}{\rho(f,g)}\right) \le 1.$$

(iii) *For every α, $t > 0$ and g in $\mathscr{F}$,*

$$P\left\{\frac{1}{\alpha}\sup\{|f-g|;\ f\in\mathscr{F}, \rho(f,g)\le\alpha\} > t\right\} \le \frac{LLt}{t^2}.$$

Then $\|I_n(f)\|_{\mathscr{F}} \to 0$ in probability.

These results are easily deduced from Theorem 20 for appropriate choices of the parameters d, s, v; basically $d = d_2 + \rho$, $s = \sqrt{LLn}$, $v = 1$ in Theorem 21 and $d = \rho$, $s = 1$, $v = \sqrt{LLn}$ in Theorem 22. We should, however, point out that Theorem 20 actually contains more information which interpolates in some sense between these extremes and also that its various hypotheses seem the best possible and almost necessary for some of them. Note also that, in the same way, the ideas contained in the proof of Theorem 20 can be used for stable central limit theorems as in [3]; however, the majorizing measure condition has to be adapted to the stable case.

REFERENCES

[1] ALT, J.-C. (1987). La loi des grands nombres de Prohorov dans les espaces de type *p*. *Ann. Inst. H. Poincaré Probab. Statist.* **23** 561–574.

[2] ANDERSEN, N. T., GINÉ, E., OSSIANDER, M. and ZINN, J. (1988). The central limit theorem and the law of the iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields* **77** 271–305.

[3] ANDERSEN, N. T., GINÉ, E. and ZINN, J. (1988). The central limit theorem under local conditions: The case of Radon infinitively divisible limits without Gaussian component. *Trans. Amer. Math. Soc.* **309** 1–34.
[4] BORISOV, I. S. (1985). Exponential estimates for distributions of sums of independent random fields. *Siberian Math. J.* **26** 7–15.
[5] DONOGHUE, W. F., JR. (1969). *Distributions and Fourier Transforms*. Academic, New York.
[6] DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.
[7] DUDLEY, R. M. (1984). A course on empirical processes. *École d'Été de Probabilités de Saint-Flour XII—1982. Lecture Notes in Math.* **1097** 2–142. Springer, Berlin.
[8] DUDLEY, R. M. and KUELBS, J. (1980). Log log laws for empirical measures. *Ann. Probab.* **8** 405–418.
[9] FERNIQUE, X. (1975). Régularité des trajectoires des fonctions aléatoires gaussiennes. *École d'Été de Probabilités de Saint-Flour IV—1974. Lecture Notes in Math.* **480** 1–96. Springer, Berlin.
[10] GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.
[11] GINÉ E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Probability and Banach Spaces, Zaragoza 1985. Lecture Notes in Math.* **1221** 50–113. Springer, Berlin.
[12] HEINKEL, B. (1984). Une extension de la loi des grands nombres de Prohorov. *Z. Wahrsch. verw. Gebiete* **67** 349–362.
[13] HOFFMANN-JØRGENSEN, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–186.
[14] KUELBS, J. (1977). Kolmogorov's law of the iterated logarithm for Banach space valued random variables. *Illinois J. Math.* **21** 784–800.
[15] KUELBS, J. and ZINN, J. (1979). Some stability results for vector valued random variables. *Ann. Probab.* **7** 75–84.
[16] LEDOUX, M. and TALAGRAND, M. (1988). Characterization of the law of the iterated logarithm in Banach spaces. *Ann. Probab.* **16** 1242–1264.
[17] OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2-bracketing. *Ann. Probab.* **15** 897–919.
[18] PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. *Probability and Analysis, Varenna 1985. Lecture Notes in Math.* **1206** 167–241. Springer, Berlin.
[19] STOUT, W. F. (1974). *Almost Sure Convergence*. Academic, New York.
[20] TALAGRAND, M. (1986). Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Ann. Probab.* To appear.
[21] TALAGRAND, M. (1987). Donsker classes and random geometry. *Ann. Probab.* **15** 1327–1338.
[22] TALAGRAND, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** 99–149.
[23] TALAGRAND, M. (1988). Necessary conditions for sample boundedness of p-stable processes. *Ann. Probab.* **16** 1584–1595.

INSTITUT DE RECHERCHE
MATHÉMATIQUE AVANCÉE
LABORATOIRE ASSOCIÉ AU C.N.R.S.
UNIVERSITÉ LOUIS PASTEUR
7, RUE RENÉ DESCARTES
F-67084 STRASBOURG
FRANCE

ÉQUIPE D'ANALYSE
ASSOCIÉE AU C.N.R.S.
UNIVERSITÉ DE PARIS VI
F-75230 PARIS
FRANCE
AND
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
231 WEST 18TH AVENUE
COLUMBUS, OHIO 43210-1174