ISOPERIMETRY AND PROCESSES
IN PROBABILITY IN BANACH SPACES

Probability in Banach spaces is a branch of modern Mathematics which emphasizes the geometric and functional analytic aspects of Probability theory. Its probabilistic sources may be found in the study of regularity of random processes (especially Gaussian processes) and limit theorems for sums of independent vector valued random variables which are the two main topics of this book. Banach space theory forms its functional background and Probability in Banach spaces has strong and fruitful connections with Geometry of Banach spaces and the nowadays called local theory of Banach spaces.

Probability in Banach spaces started in the early fifties with the study, by R. Fortet and E. Mourier, of the law of large numbers and the central limit theorem for sums of independent identically distributed Banach space valued random variables. Important contributions to the foundations of probability distributions on vector spaces, toward which A. N. Kolmogorov already pointed in 1935, were at the time those of L. Le Cam and Y. V. Prokhorov and the Russian school. A decisive step to the modern developments of Probability in Banach spaces was the introduction by A. Beck (1962) of a convexity condition on normed linear spaces equivalent to the validity of the extension of a classical law of large numbers of Kolmogorov. This geometric line of investigation was pursued and amplified by the Schwartz school in the early seventies. The concepts of radonifying and summing operators and the landmark work of B. Maurey and G. Pisier on type and cotype of Banach spaces considerably influenced the developments of Probability in Banach spaces. Other noteworthy achievements of the period are the early book (1968) of J.-P. Kahane, who systematically developed the crucial idea of symmetrization, and the study by J. Hoffmann-Jørgensen of sums of independent random variables. Simultaneously, the study of regularity properties of random processes, in particular Gaussian processes, saw great progress in the late sixties and early seventies with the introduction of entropy methods. Processes are understood here as random functions on some abstract index set T, in other words as a family $X = (X_t)_{t\in T}$ of random variables. In this setting of what might appear as Probability with minimal structure, the major discovery of R. Dudley (1967) was the idea of analyzing regularity properties of a Gaussian process X through the geometry of the index set T for the L_2-metric $\| X_s - X_t \|_2$ induced by X itself. These foundations of Probability in Banach spaces led to rather intense activity for the last fifteen years. In particular, the Dudley-Fernique theorems on regularity properties of Gaussian and stationary Gaussian processes allowed the definitive treatment by M.B. Marcus and G. Pisier of regularity of random Fourier series initiated in this line by J.-P. Kahane. Limit theorems for sums of

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

independent Banach space valued random variable became clearer. Under the impulse, in particular, of the local theory of Banach spaces, isoperimetric methods and concentration of measure phenomena put forward most vigorously by V.D. Milman made a strong entry in the late seventies and eighties into the probabilistic methods of investigation. Starting from Dvoretzky's theorem on Euclidian sections of convex bodies, the isoperimetric inequalities on spheres and in Gauss space proved most powerful in the study of Gaussian measures and processes, in particular through the work of C. Borell. They were useful too in the study of limit theorems through the technique of randomization. An important recent development was the discovery, motivated by these results, of a new isoperimetric inequality that is closely connected to the tail behavior of sums of independent Banach space valued random variables. It allows in particular today an almost complete description of various strong limit theorems like the strong law of large numbers and the law of the iterated logarithm. In the mean time, almost sure boundedness and continuity of general Gaussian processes has been recently completely understood with the tool of majorizing measures.

One of the fascinations of the theory of Probability in Banach spaces today is its use of a wide range of rather powerful methods. Since the field is one of the most active contact points between Probability and Analysis, it should come as no surprise that many of the techniques are not probabilistic but rather come from Analysis. The book focuses on two connected topics – the use of isoperimetric methods and the regularity of random process – where many of these techniques come into play and which encompass many (although not all) of the main aspects of Probability in Banach spaces. The purpose of this book is to give a modern and, at many places, seemingly definitive account of these topics. The book is written so as to require only basic prior knowledge of either Probability or Banach space theory, in order to make it accessible from readers of both fields as well as to non-specialists. It is moreover presented in perspective with the historical developments and strong modern interactions between Measure and Probability theory, Functional Analysis and Geometry of Banach spaces. It is essentially self-contained (with the exception that the proofs of a few deep isoperimetric results have not been reproduced), so as to be accessible to anyone starting the subject, including graduate students. Emphasis has been put in bringing forward the ideas we judge important but not on encyclopedic detail. We hope that these ideas will fruitfully serve the further developments of the field and hope their propagation will influence new areas of Mathematics.

The two parts of the book are introduced by chapters on isoperimetric background and generalities on vector valued random variables. To explain and motivate the organization of our work, let us briefly analyze one fundamental example. Let (T,d) be a compact metric space and let $X = (X)_{t\in T}$ be a Gaussian process indexed by T. If X has almost all its sample paths continuous, it defines a Gaussian Radon measure on the Banach space $C(T)$ of all continuous functions on T. Such a Gaussian measure or variable may then be studied for its own sake and shares indeed some remarkable integrability and tail behavior properties of isoperimetric nature. On the other hand, one might wonder (before) when a given Gaussian process is almost surely continuous. An analysis of the geometry of the index set T for the L_2-metric $\| X_s - X_t \|_2$ induced by the process allows a complete understanding of this property. These related but somewhat different aspects of the study of Gaussian variables, which were historically the two main streams of developments, led us to divide the book into two parts. (The logical order would have been perhaps to ask first when a given process is bounded or continuous and then investigate it for its properties as a well-defined infinite dimensional random vector; we have however chosen the other way for various pedagogical reasons.) In the first part, we study vector valued random variables, their integrability and tail behavior properties and strong limit theorems for sums of independent random variables. Sucessively, Gaussian, Rademacher series, stable and sums of independent Banach space valued random variables are investigated in this scope using isoperimetric tools. The strong law of large numbers and the law of the iterated logarithm, for which the almost sure statement is shown to reduce to the statement in probability, complete this first part with extension to infinite dimensional Banach space valued random variables of some classical real limit theorems. In the second part, tightness of sums of independent random variables and regularity properties for random processes are presented. The link with Geometry of Banach spaces through type and cotype is the subject of one chapter with applications in particular to the classical central limit theorem. General random processes are investigated and regularity properties of Gaussian processes characterized with applications to random Fourier series. The book is completed with an account on empirical processes methods and with several applications, especially to local theory of Banach spaces.

Chapter 1. Isoperimetric inequalities and the concentration of measure phenomenon

Chapter 1. Isoperimetric inequalities and the concentration of measure phenomenon

In this first chapter, we present the isoperimetric inequalities which now appear as the crucial concept in the understanding of various concentration inequalities, tail behaviors and integrabililty theorems in Probability in Banach spaces. These inequalities often arise as the final and most elaborated forms of previous, weaker (but already efficient) inequalities which will be mentioned in their framework throughout the book. In these final forms however, the isoperimetric inequalities and associated concentration of measure phenomena provide the appropriate ideas for an in-depth comprehension of some of the most important theorems of the theory.

The concentration of measure phenomenon which roughly describes how a well-behaved function is almost a constant on almost all the space can moreover be seen as the explanation for the two main parts of this work: the first one deals with "nice" functions applying isoperimetric inequalities and concentration properties, the second tries to determine conditions for a function to be "nice".

The concentration of measure phenomenon has been mainly put forward by the local theory of Banach spaces in the study of Dvoretzky's theorem on almost Euclidian sections of convex bodies. Following [G-M], [Mi-S], the basic idea may be described in the following way. Let (X,ρ,μ) be a (compact) metric space (X,ρ) with a Borel probability measure μ. The concentration function $\alpha(X,r), r>0$, is defined as

$$\alpha(X,r)=\sup\{1-\mu(A_r);\ \mu(A)\geq\frac{1}{2}, A\subset X,\ A\ \text{ Borel}\}$$

where A_r denotes the ρ-neighborhood of order r of A i.e.

$$A_r=\{x\in X;\ \rho(x,A)<r\}.$$

For many families (X,ρ,μ), the concentration function $\alpha(X,r)$ turns out to be extremely small when r increases to infinity. A typical and basic example is given by the Euclidian unit sphere S^{N-1} in $\mathbb{R}^N$ equipped with its geodesic distance ρ and normalized Haar measure σ_{N-1} for which it can be shown (see below) that

$$\alpha(S^{N-1},r)\leq\left(\frac{\pi}{8}\right)^{1/2}\exp(-(N-2)r^2/2)\ (N\geq 3).$$

Hence, the complement of the neighborhood of order r of a set of probability bigger than $1/2$ decreases extremely rapidly when r becomes large. This is what is now usually called the concentration of measure phenomenon. In presence of such a property, any nice function is very close to being a constant (median or expectation) on all but a very small set, the smallness of which depends on $\alpha(X,r)$. For example, if f is a function on S^{N-1}, denote by $\omega_f(\varepsilon)$ its modulus of continuity, $\omega_f(\varepsilon) = \sup\{|f(x)-f(y)|;\ \rho(x,y)<\varepsilon\}$, and let M_f be a median of f. Then, for every $\varepsilon>0$,

$$\sigma_{N-1}(|f-M_f|>\omega_f(\varepsilon)) \le \left(\frac{\pi}{2}\right)^{1/2} \exp(-(N-2)\varepsilon^2/2).$$

The concentration of measure phenomenon is usually derived (see however [G-M], [Mie3]) from an isoperimetric property. This chapter described some isoperimetric inequalities, which we develop here in their abstract and measure theoretical setting. Their application to Probability in Banach spaces will be one of the purposes of Part I of this book. The first section presents isoperimetric inequalities in the classical cases of the sphere, Gauss space and the cube. The main object of the second section is an isoperimetric theorem for prod uct measures (independent random variables) while the last one is devoted to some well-known and useful martingale inequalities. Proofs of some of the deepest inequalities like the isoperimetric inequality on spheres and the one for product measures are omitted and replaced by (hopefully) accurate references. Various comments and remarks try however to describe some of the ideas involved in these results as well as their consequences which will be useful in the sequel.

1.1. Some isoperimetric inequalities on the sphere, in Gauss space and on the cube

Concentration properties for families (X,ρ,μ) are thus usually established through isoperimetric inequalities. We present some of these inequalities in this section and start with the isoperimetric inequality on the spheres already alluded to above (cf. [B-Z], [F-L-M],...).

Theorem 1.1 If A is a Borel set in S^{N-1} and H a cap (i.e. a ball for the geodesic distance ρ) with the same measure $\sigma_{N-1}(H)=\sigma_{N-1}(A)$, then, for any $r>0$,

$$\sigma_{N-1}(A_r) \ge \sigma_{N-1}(H_r)$$

where we recall that $A_r=\{x\in S^{N-1};\ \rho(x,A)<r\}$ is the neighborhood of A of order r for the geodesic distance. In particular, if $\sigma_{N-1}(A)\ge 1/2$ (and $N\ge 3$), then

$$\sigma_{N-1}(A_r) \ge 1-\left(\frac{\pi}{8}\right)^{1/2} \exp(-(N-2)r^2/2).$$

The main interest in such an isoperimetric theorem is of course the possiblility of an estimate (or even explicit computation) of the measure of a cap (the neighborhood of a cap is again a cap), a particular important estimate being given in the second assertion of the theorem.

Our interest in Theorem 1.1 lies in its connections and consequences to a similar isoperimetric result for Gaussian measures. The close relationships and analogies between uniform measures on spheres and Gaussian measures have been noticed in many contexts. In this isoperimetric setting, it turns out that the isoperimetric inequality on S^{N-1} leads in the Poincaré limit when N increases to infinity to an isoperimetric inequality for Gauss measure. Poincaré's limit expresses the canonical Gaussian measure in finite dimension as the limiting distribution of projected uniform distributions on spheres of radius $\sqrt{N}$ when N tends to infinity. This is one example of the deep relations mentioned previously; it is also a way of illustrating the common belief that Wiener measure can be thought as uniform distribution on a sphere of infinite dimension and of radius square root of infinity (cf. [M-K]).

To be more precise on this observation of Poincaré, denote, for every N, by $\sigma_{N-1}^{\sqrt{N}}$ the uniform normalized measure on the sphere $\sqrt{N}S^{N-1}$ of center the origin and radius $\sqrt{N}$ in $\mathbb{R}^N$. Denote further by $\Pi_{N,d}(N \geq d)$ the projection from $\sqrt{N}S^{N-1}$ onto $\mathbb{R}^d$. Then, the sequence $\{\Pi_{N,d}(\sigma_{N-1}^{\sqrt{N}}); N \geq d\}$ of measures on $\mathbb{R}^d$ converges weakly when N goes to infinity to the canonical Gaussian measure on $\mathbb{R}^d$. To sketch the proof, simply note that by the law of large numbers $\rho_N^2/N \to 1$ almost surely (or only in probability) where $\rho_N^2 = g_1^2 + \cdots + g_N^2$ and (g_i) is a sequence of independent standard normal random variables. But, clearly, $(N^{1/2}/\rho_N) \cdot (g_1, \ldots, g_N)$ is equal in distribution to $\sigma_{N-1}^{\sqrt{N}}$; hence $(N^{1/2}/\rho_N) \cdot (g_1, \ldots, g_d) = \Pi_{N,d}(\sigma_{N-1}^{\sqrt{N}})$ and the conclusion follows. Note that we will actually need a little more than weak convergence in this Poincaré limit, namely convergence for *all* Borel sets. This can be obtained similarly with some more efforts (see [Fer9]).

Provided with this tool, it is simple to see how it is possible to derive an isoperimetric inequality for Gaussian measures from Theorem 1.1. The basic result which is obtained in this way concerns the canonical Gaussian distribution in finite dimension; as is classical however, the fundamental feature of Gaussian distributions allows then to extend these results to general finite or infinite dimensional Gaussian measures. Let us denote therefore by γ_N the canonical Gaussian probability measure on $\mathbb{R}^N$ with density

$$\gamma_N(dx) = (2\pi)^{-N/2} \exp(-|x|^2/2)dx.$$

Denote further by Φ the distribution function of this measure in dimension 1, i.e.

$$\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} \exp(-x^2/2)dx, \ t \in [\infty, +\infty];$$

Φ^{-1} is the inverse function and $\Psi = 1 - \Phi$ for which we recall the classical estimate:

$$\Psi(t) \leq \frac{1}{2} \exp(-t^2/2), \ t \geq 0.$$

The next theorem is the isoperimetric inequality for $(\mathbb{R}^N, d, \gamma_N)$ where d denotes the Euclidian distance.

Theorem 1.2. If A is a Borel set in $\mathbb{R}^N$ and if H is a half-space $\{x \in \mathbb{R}^N; < x, u > < \lambda\}$, $u \in \mathbb{R}^N$, $\lambda \in [-\infty, +\infty]$, with the same Gaussian measure $\gamma_N(H) = \gamma_N(A)$, then, for any $r > 0$, $\gamma_N(A_r) \geq \gamma_N(H_r)$ where, accordingly, A_r is the Euclidean neighborhood of order r of A. Equivalently,

$$\Phi^{-1}(\gamma_N(A_r)) \geq \Phi^{-1}(\gamma_n(A)) + r$$

and in particular, if $\gamma_N(A) \geq 1/2$,

$$1 - \gamma_N(A_r) \leq \Psi(r) \leq \frac{1}{2} \exp(-r^2/2).$$

Proof. The equivalence between the two formulations is easy since the Gaussian measure of a half-space is being computed in dimension one and thus

$$\gamma_N(H_r) = \Phi(\Phi^{-1}(\gamma_N(H)) + r) = \Phi(\Phi^{-1}(\gamma_N(A)) + r).$$

The case $\gamma_N(A) \geq 1/2$ simply follows form the fact that $\Phi^{-1}(1/2) = 0$. Turning to the proof itself, since $\Phi^{-1}(0) = -\infty$, we may assume that $a = \Phi^{-1}(\gamma_N(A)) > -\infty$. Let then $b \in] -\infty, a[$. By Poincaré's observation, for all $k (\geq N)$ large enough,

$$\sigma_{k-1}^{\sqrt{k}}(\Pi_{k,N}^{-1}(A)) > \sigma_{k-1}^{\sqrt{k}}(\Pi_{k,1}^{-1}(] -\infty, b])).$$

It is easy to see that $\Pi_{k,N}^{-1}(A_r) \supset (\Pi_{k,N}^{-1}(A))_r$ where the neighborhood of order r on the right is understood with respect to the geodesic distance on $\sqrt{k} S^{k-1}$. Since $\Pi_{k,1}^{-1}(] -\infty, b])$ is a cap on $\sqrt{k} S^{k-1}$, by the isoperimetric inequality on spheres (Theorem 1.1),

$$\begin{aligned} \sigma_{k-1}^{\sqrt{k}}(\Pi_{k,N}^{-1}(A_r)) &\geq \sigma_{k-1}^{\sqrt{k}}((\Pi_{k,1}^{-1}(A))_r) \\ &\geq \sigma_{k-1}^{\sqrt{k}}((\Pi_{k,1}^{-1}(] -\infty, b]))_r). \end{aligned}$$

Now $(\Pi_{k,1}^{-1}(]-\infty, b]))_r = \Pi_{k,1}^{-1}(]-\infty,\ b+r(k)])$ for some $r(k) \geq 0$ satisfying $\lim_{k\to\infty} r(k) = r$. In the Poincaré limit we get therefore that $\gamma_N(A_r) \geq \Phi(b+r)$, hence the result since $b < \Phi^{-1}(\gamma_N(A))$ is arbitrary.

Note that half-spaces, as caps on spheres, are extremal sets for the Gaussian isoperimetric inequality since they achieve equality in the conclusion.

The isoperimetric inequality for Gauss measure thus follows rather easily from the corresponding one on spheres. This later one however requires a quite involved proof based on extensive use of powerful symmetrization (in the sense of Steiner) techniques. It was one of the remarkable observations of the work of A. Ehrhard to show how one can introduce a similar symmetrization procedure adapted to Gaussian measure (with half-spaces as extremal sets). This allows a more intrinsic proof of Theorem 1.2. This also led him to a rather complete isoperimetric calculus in Gauss space; he obtained in particular in this way an inequality of Brunn-Minkowski's type; namely for A, B convex sets in $\mathbb{R}^N$ and $\lambda \in [0,1]$,

$$\Phi^{-1}(\gamma_N(\lambda A + (1-\lambda)B)) \geq \lambda\Phi^{-1}(\gamma_N(A)) + (1-\lambda)\Phi^{-1}(\gamma_N(B)) \tag{1.1}$$

where the sum $\lambda A + (1-\lambda)B$ is understood in the sense of Minkowski as $\{x \in \mathbb{R}^N;\ x = \lambda a + (1-\lambda)b,\ a \in A, b \in B\}$. Taking B to be the Euclidian ball of center the origin and radius $r/(1-\lambda)$ and letting λ tend to 1 , it is easily seen how (1.1) actually implies the isoperimetric inequality of Theorem 1.2 (for A convex). However, (1.1) is only known at the present time for convex sets. An inequality on which (1.1) appears as an improvement (for convex sets) but which holds for arbitrary Borel sets A, B is the so-called log-concavity of Gaussian measures:

$$\log \gamma_N(\lambda A + (1-\lambda)B) \geq \lambda \log \gamma_N(A)) + (1-\lambda)\log \gamma_N(B). \tag{1.2}$$

A proof of (1.2) may be given using again the Poincaré limit but this time on the classical Brunn-Minkowski inequality on $\mathbb{R}^N$ (see [B-Z], [Pi18],...) which states that

$$\mathrm{vol}_N(\lambda A + (1-\lambda)B) \geq (\ \mathrm{vol}_N(A))^\lambda (\ \mathrm{vol}_N(B))^{1-\lambda}$$

for λ in $[0,1]$, A, B bounded in $\mathbb{R}^N$ and where vol_N is N-dimensional volume. Let, for $k \geq N$, $P_{k,N}$ be the projection from the ball of center the origin and radius $\sqrt{k}$ in $\mathbb{R}^k$ onto $\mathbb{R}^N$. If A, B are Borel sets in $\mathbb{R}^N$ and $0 \leq \lambda \leq 1$, by convexity it is clear that

$$P_{k,N}^{-1}(\lambda A + (1-\lambda)B) \supset \lambda P_{k,N}^{-1}(A) + (1-\lambda)P_{k,N}^{-1}(B)$$

so that, by Brunn-Minkowski (in $\mathbb{R}^k$),

$$\mathrm{vol}_k(P_{k,N}^{-1}(\lambda A+(1-\lambda)B)) \geq (\mathrm{vol}_k(P_{k,N}^{-1}(A)))^\lambda (\mathrm{vol}_k(P_{k,N}^{-1}(B)))^{1-\lambda} .$$

If we subtract 1 on each side of this inequality, multiply then by k and let it tend to infinity, we see that we obtain (1.2) since it is easily checked that the Poincaré limit also indicates that

$$\lim_{k\to\infty} \mathrm{vol}_k(P_{k,N}^{-1}(A)) = \gamma_N(A)$$

for every Borel set A in $\mathbb{R}^N$. One measure further on this proof the sharpness of (1.1).

As announced, and due to the properties of Gaussian distributions, the preceding inequalities and Theorem 1.2 easily extend to general finite or infinite dimensional Gaussian measures. These extensions will usually be described below in their applications. Let us however briefly indicate one infinite dimensional result which will be useful to record here. On $\mathbb{R}^{\mathbb{N}}$, consider the measure $\gamma=\gamma_\infty$ which is the infinite product of the canonical one-dimensional Gaussian distribution on each coordinate. The isoperimetric inequality indicates similarly that for eachBorel set A in $\mathbb{R}^{\mathbb{N}}$ and $r>0$,

$$\Phi^{-1}(\gamma_*(A_r)) \geq \Phi^{-1}(\gamma(A))+r \tag{1.3}$$

where A_r is here the Euclidian or rather Hilbertian neighborhood of order r of A in $\mathbb{R}^{\mathbb{N}}$, i.e. $A_r = A+rB = \{x=a+rh; a\in A, h\in \mathbb{R}^{\mathbb{N}},\ |h|\leq 1\}$ where B is the unit ball of ℓ_2 (A_r is not necessarily measurable). This of course simply follows for Theorem 1.2 and a cylindrical approximation. Note that $\gamma(\ell_2)=0$!

As a corollary to Theorem 1.2, we now express the concentration of measure phenomenon for functions on $(\mathbb{R}^N, d, \gamma_N)$. This formulation of the isoperimetry will turn out to be a convenient tool in the applications. Let f be Lipschitzian on $\mathbb{R}^N$ with Lipschitz norm given by

$$\| f \|_{\mathrm{Lip}} = \sup\{\frac{|f(x)-f(y)|}{|x-y|};\ x,y\in\mathbb{R}^N\}.$$

Let us denote further by M_f a median of f for γ_N, i.e. M_f is a number such that $\gamma_N(f\geq M_f)$ and $\gamma_N(f\leq M_f)$ are both bigger than $1/2$. Applying the second conclusion of Theorem 1.2 to those two sets of measure bigger than $1/2$ and noticing that for $t>0$,

$$(\{f\geq M_f\}\cap\{f\leq M_f\})_t \subset \{|f-M_f|\leq t\,\| f \|_{\mathrm{Lip}}\},$$

we get that, for all $t > 0$

$$\gamma_N(|f - M_f| > t) \leq 2\Psi(t/\parallel f \parallel_{\mathrm{Lip}}) \leq \exp(-t^2/2 \parallel f \parallel^2_{\mathrm{Lip}}). \tag{1.4}$$

Hence, with very high probability, f is concentrated around its median M_f. As we will see in Chapter 3, this inequality can be used to investigate the integrability properties of Gaussian random vectors. Let us note also that the preceding argument applied only to $\{f \leq M_f\}$ shows similarly that

$$\gamma_N(f > M_f + t) \leq \frac{1}{2}\exp(-t^2/2 \parallel f \parallel^2_{\mathrm{Lip}}).$$

This inequality however appears more as a *deviation* inequality only as opposed to (1.4) which indicates a *concentration*. We shall come back to this distinction in various contexts in the sequel.

If (1.4) appears as a direct consequence of Theorem 1.2, it should be noted that inequalities of the same type can actually be established by simple direct arguments and considerations. One such approach is the following. Let f be, as before, Lipschitzian on $\mathbb{R}^N$, so that f is almost everywhere differentiable and its gradient ∇f satisfies $|\nabla f| \leq \parallel f \parallel_{\mathrm{Lip}}$. Assume now moreover that $\int f d\gamma_N = 0$. Then, for any t and $\lambda > 0$, we can write by Chebyshev's inequality,

$$\begin{aligned}\gamma_N(f > t) &\leq \exp(-\lambda t)\int \exp(\lambda f) d\gamma_N \\ &\leq \exp(-\lambda t)\int\int \exp[\lambda(f(x) - f(y))] d\gamma_N(x) d\gamma_N(y)\end{aligned}$$

where in the second inequality we have used Jensen's inequality (in y) and the mean zero property. Now, x, y being fixed in $\mathbb{R}^N$, set, for any θ in $[0, 2\pi]$,

$$x(\theta) = x\sin\theta + y\cos\theta,\ \ x'(\theta) = x\cos\theta - y\sin\theta.$$

We have

$$\begin{aligned}f(x) - f(y) &= \int_0^{\pi/2} \frac{d}{d\theta} f(x(\theta)) d\theta \\ &= \int_0^{\pi/2} < \nabla f(x(\theta)), x'(\theta) > d\theta.\end{aligned}$$

Hence, using Jensen's inequality one more time (in θ here), $\gamma_N(f > t)$ is majorized by

$$\exp(-\lambda t)\frac{2}{\pi}\int_0^{\pi/2} \left(\int\int \exp\left[\frac{\lambda\pi}{2} < \nabla f(x(\theta)), x'(\theta) >\right] d\gamma_N(x) d\gamma_N(y)\right) d\theta.$$

The fundamental rotational invariance of Gaussian measures indicates that, for any θ , the couple $(x(\theta), x'(\theta))$ has the same distribution, under $\gamma_N \otimes \gamma_N$, as the original one (x, y) . Therefore, by Fubini's theorem,

$$\gamma_N(f > t) \leq \exp(-\lambda t) \int \int \exp\left[\frac{\lambda\pi}{2} < \nabla f(x), y >\right] d\gamma_N(x) d\gamma_N(y) \ .$$

Performing integration in y ,

$$\begin{aligned} \gamma_N(f > t) &\leq \exp(-\lambda t) \int \exp\left(\frac{\lambda^2\pi^2}{8} |\nabla f|^2\right) d\gamma_N \\ &\leq \exp\left(-\lambda t + \frac{\lambda^2\pi^2}{8} \parallel f \parallel_{\mathrm{Lip}}^2\right) . \end{aligned}$$

If we then minimize in λ $(\lambda = 4t/\pi^2 \parallel f \parallel_{\mathrm{Lip}}^2)$, we finally get

$$\gamma_N(f > t) \leq \exp(-2t^2/\pi^2 \parallel f \parallel_{\mathrm{Lip}}^2)$$

for every $t > 0$. For f not necessarily of mean zero, and applying the result to $-f$ as well, it follows that, for $t > 0$,

$$\gamma_N(|f - E_f| > t) \leq 2\ \exp(-2t^2/\pi^2 \parallel f \parallel_{\mathrm{Lip}}^2) \tag{1.5}$$

where $E_f = \int f d\gamma_N$ (finite since f is Lipschitzian). This inequality is of course very close in spirit to (1.4). If (1.5) has a worse constant in the exponent, it can actually be shown, using a similar argument but with stochastic differentials with respect to Brownian motion, that we also have, for every $t > 0$,

$$\gamma_N(|f - E_f| > t) \leq 2\ \exp(-t^2/2 \parallel f \parallel_{\mathrm{Lip}}^2). \tag{1.6}$$

The preceding argument leading to (1.5) however presents the advantage to apply to more general situations as the case of vector valued functions (cf. [Pi16]). We retain in any case that concentration inequalities of the type (1.4)-(1.6) are usually easier to obtain than the rather delicate isoperimetric theorems.

(1.4) and (1.6) describe the same concentration property around a median M_f or the expectation $E_f = \int f d\gamma_N$ of a Lipschitzian function f . It is of course often easier to work with expectations rather than with medians. Actually, these are essentially of the same order here. Indeed, integrating for example (1.4) yields

$$|E_f - M_f| \ \leq\ \sqrt{2\pi} \parallel f \parallel_{\mathrm{Lip}} \ ,$$

while, given (1.6), if t is chosen such that

$$2\ \exp(-t^2/2 \parallel f \parallel^2_{\mathrm{Lip}}) < \frac{1}{2}\ ,$$

for example $t = 2 \parallel f \parallel_{\mathrm{Lip}}$, we get from (1.6) that

$$E_f - t \leq M_f \leq t + E_f.$$

However, it is not known whether it is possible to deduce *exactly* (1.4) and (1.6) from each other. Note further that (1.4) actually shows that a median M_f is necessarily *unique*. Indeed, if $M_f < M_f'$ are two distinct medians of f, letting $t = (M_f' - M_f)/2 > 0$ gives

$$\frac{1}{2} \leq \gamma_N(f \geq M_f') \leq \gamma_N(f > M_f + t) \leq \Psi(t/ \parallel f \parallel_{\mathrm{Lip}}) < \frac{1}{2}$$

which is impossible.

Uniform measures on spheres and Gaussian measures thus satisfy isoperimetric inequalities and concentration phenomena. For our purposes, there is a useful observation which allows to deduce from the Gaussian inequalities further ones for a rather large class of measures. Denote by φ a Lipschitzian map on $\mathbb{R}^N$ with values in $\mathbb{R}^N$ such that

$$|\varphi(x) - \varphi(y)| \leq c|x - y| \quad \text{for all } x, y \text{ in } \mathbb{R}^N$$

for some $c = c_\varphi > 0$. Denote by λ the image measure of γ_N by φ, i.e. $\lambda(A) = \gamma_N(\varphi^{-1}(A))$ for any Borel set A in $\mathbb{R}^N$. Then there is an isoperimetric inequality of Gaussian type for λ, namely, for A measurable in $\mathbb{R}^N$ and $r > 0$,

$$\Phi^{-1}(\lambda(A_{cr})) \geq \Phi^{-1}(\lambda(A)) + r. \tag{1.7}$$

Similarly, the corresponding inequality (1.4) for λ also holds with $c \parallel f \parallel_{\mathrm{Lip}}$ instead of $\parallel f \parallel_{\mathrm{Lip}}$ in the right hand side. For a proof of (1.7), simply note that by Theorem 1.2,

$$\Phi^{-1}(\gamma_N((\varphi^{-1}(A))_r)) \geq \Phi^{-1}(\lambda(A)) + r$$

and, by the Lipschitz property of φ, clearly $(\varphi^{-1}(A))_r \subset \varphi^{-1}(A_{cr})$ from which the result follows. Inequality (1.5) and its simple proof may be extended similarly (see [Pi16]).

It is then of course of some interest to try to describe the class of probabilities λ which can be obtained as the image of γ_N by a contraction. If a complete description is still missing, the next examples are worthwhile to be noted (and useful for the sequel) . Let λ be uniformly distributed on the cube $[0,1]^N$. Then λ is the image of γ_N by the map $\Phi^{\otimes N}$, i.e. $\varphi(x) = \varphi(x_1, \dots, x_N) = \Phi(x_1)\dots\Phi(x_N), x = (x_1, \dots, x_N) \in \mathbb{R}^N$, for which it is easily seen that $c_\varphi = (2\pi)^{-1/2}$. If one, for symmetry reasons, is rather interested by the uniform measure on $[-1/2, +1/2]^N$, choose then $\varphi = (2\Phi - 1)^{\otimes N}$ for which $c_\varphi = (2/\pi)^{1/2}$.

The preceding approach however does not allow to investigate the important case of Haar measure on $\{0,1\}^N$ (the extreme points of $[0,1]^N$), or rather on $\{-1,+1\}^N$ which we use preferably for symmetry reasons. A different way has to be taken. Denote by $\mu_N = (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1})^{\otimes N}$ the canonical probability measure (Haar measure) on the (Cantor) group $\{-1,+1\}^N$. Consider the normalized Hamming metric d on $\{-1,+1\}^N$ given by

$$d(x,y) = \frac{1}{N}\ \mathrm{Card}\{i \le N;\ x_i \ne y_i\} = \frac{1}{2N}\sum_{i=1}^N |x_i - y_i|$$

for $x, y \in \{-1,+1\}^N$. An isoperimetric theorem for the triple $(\{-1,+1\}^N, d, \mu_N)$ is known [Har] and states in particular that if $\mu_N(A) \ge 1/2$ for some $A \subset \{-1,+1\}^N$, for $r > 0$,

$$1 - \mu_N(A_r) \le \frac{1}{2}\ \exp(-2Nr^2)$$

where, as usual, $A_r = \{x \in \{-1,+1\}^N;\ d(x,A) < r\}$. Unfortunately, this result is not strong enough to yield a concentration inequality for μ_N similar to (1.4) or (1.6) since it depends on the dimension. In order to accomplish this program, we will need a stronger result (independent of N) which is the following. For a non-empty subset A of $\{-1,+1\}^N$, set, for $x \in \{-1,+1\}^N$,

$$d_A(x) = \inf\{|x-y|; y \in\ \mathrm{Conv}A\}$$

where $\mathrm{Conv}A$ is the convex hull (in $[-1,+1]^N$) of A .

Theorem 1.3. For any non-empty subset A of $\{-1,+1\}^N$,

$$\int \exp(d_A^2/8)d\mu_N \le \frac{1}{\mu_N(A)}.$$

Proof. We first consider the case where Card $A = 1$. Then

$$\begin{aligned}\int \exp(d_A^2/8)d\mu_N &= 2^{-N}\sum_{i=0}^{N}\binom{N}{i}e^{i/2}\\ &= \left(\frac{1+e^{1/2}}{2}\right)^N \le 2^N = \frac{1}{\mu_N(A)}\end{aligned}$$

since $e^{1/2} < e < 3$. This already proves the theorem when $N = 1$ since then the only case left is $A = \{-1,+1\}$ for which $d_A \equiv 0$ and the result holds.

We now prove Theorem 1.3 by induction over N. Assuming it holds for N, we prove it for $N+1$. By the preceding, it is enough to consider the case where A has at least two points. Assuming without loss of generality that these points differ on the last coordinate and identifying $\{-1,+1\}^{N+1}$ with $\{-1,+1\}^N \times \{-1,+1\}$, we can then suppose that $A = A_{-1}\times\{-1\}\cup A_{+1}\times\{+1\}$ where A_{-1}, A_{+1} are non-empty in $\{-1,+1\}^N$. We can assume for example that $\mu_N(A_{-1}) \le \mu_N(A_{+1})$ and observe moreover that $d_A((x,1)) \le d_{A_{+1}}(x)$. The crucial point of the proof is contained in the following observation: for any x in $\{-1,+1\}^N$ and $0 \le \alpha \le 1$,

$$d_A^2((x,-1)) \le 4\alpha^2 + \alpha d_{A_{+1}}^2(x) + (1-\alpha)d_{A_{-1}}^2(x). \tag{1.8}$$

Indeed, for $i = -1,+1$, let $z_i \in \mathrm{Conv}A_i$ such that $|x - z_i| = d_{A_i}(x)$. We notice that $(z_i, i) \in \mathrm{Conv}A$ so that $z = (\alpha z_{+1} + (1-\alpha)z_{-1}, -1+2\alpha) \in \mathrm{Conv}A$. Now

$$\begin{aligned}|(x,-1)-z|^2 &= 4\alpha^2 + |x - (\alpha z_{+1} + (1-\alpha)z_{-1})|^2\\ &= 4\alpha^2 + |\alpha(x - z_{+1}) + (1-\alpha)(x - z_{-1})|^2\\ &\le 4\alpha^2 + \alpha|x - z_{+1}|^2 + (1-\alpha)|x - z_{-1}|^2\end{aligned}$$

by the triangle inequality and convexity of the square function. This proves (1.8).

For $i = -1,+1$, we set $u_i = \int \exp(d_{A_i}^2/8)d\mu_N$ and $v_i = 1/\mu_N(A_i)$ so that $u_i \le v_i$ by the induction hypothesis. From $d_A((x,1)) \le dA_{+1}(x)$ and (1.8) we have, for all $0 \le \alpha \le 1$,

$$\begin{aligned}\int \exp(d_A^2/8)d\mu_{N+1} &\le \frac{1}{2}\int \exp(d_{A_{+1}}^2/8)d\mu_N\\ &\quad + \frac{1}{2}\int \exp(\alpha^2/2 + \alpha d_{A_{+1}}^2/8 + (1-\alpha)d_{A_{-1}}^2/8)\,d\mu_N\\ &\le \frac{1}{2}u_{+1} + \frac{1}{2}e^{\alpha^2/2}u_{+1}^{\alpha}u_{-1}^{1-\alpha}\\ &\le \frac{1}{2}v_{+1}\left[1 + e^{\alpha^2/2}\left(\frac{v_{-1}}{v_{+1}}\right)^{1-\alpha}\right]\end{aligned}$$

by Hölder's inequality and since $u_i \leq v_i$. The value of α which minimizes the preceding expression is $\alpha = -\log(v_{+1}/v_{-1})$ but, in order not to have to consider the case where $\alpha \geq 1$, let us take $\alpha = 1 - v_{+1}/v_{-1}$ (recall that we assume that $v_{+1} = 1/\mu_N(A_{+1}) \leq 1/\mu_N(A_{-1}) = v_{-1}$) which gives

$$\int \exp(d_A^2/8) d\mu_{N+1} \leq \frac{1}{2} v_{+1} [1 + e^{\alpha^2/2}(1-\alpha)^{\alpha-1})].$$

It is elementary to see that for $0 \leq \alpha < 1$,

$$1 + e^{\alpha^2/2}(1-\alpha)^{\alpha-1} \leq \frac{4}{2-\alpha}.$$

For the value of α,

$$\frac{1}{2} v_{+1} \left(\frac{4}{2-\alpha} \right) = \frac{2}{\frac{1}{v_{+1}} + \frac{1}{v_{-1}}} = \frac{2}{\mu_N(A_{+1}) + \mu_N(A_{-1})} = \frac{1}{\mu_{N+1}(A)}.$$

The proof of Theorem 1.3 is complete.

As announced, Theorem 1.3 contains a concentration estimate for Lipshitzian functions similar to the one we described before for Gaussian measures. This application is one of the main interests of Theorem 1.3. However, with respect to the preceding inequalities, one main additional assumption is that the inequality will only concern *convex* Lipschitzian functions f (on $\mathbb{R}^N$) with Lipschitz constant $\| f \|_{\mathrm{Lip}}$. Let M_f denote a median of f with respect to μ_N. Then, for every $t > 0$,

$$\mu_N(|f - M_f| > t) \leq 4 \exp(-t^2/8 \| f \|_{\mathrm{Lip}}^2). \tag{1.9}$$

To prove this inequality, let first $A = \{ f \leq M_f \}$. Since f is convex, $f \leq M_f$ on Conv A. Further, by the Lipschitz property, if $d_A(x) \leq t/ \| f \|_{\mathrm{Lip}}$, then $f(x) \leq M_f + t$. Hence, by Chebyshev's inequality and Theorem 1.3,

$$\begin{aligned} \mu_N(f > M_f + t) &\leq \mu_N(d_A > t/ \| f \|_{\mathrm{Lip}}) \\ &\leq \frac{1}{\mu_N(A)} \exp(-t^2/8 \| f \|_{\mathrm{Lip}}^2) \\ &\leq 2 \, \exp(-t^2/8 \| f \|_{\mathrm{Lip}}^2). \end{aligned}$$

On the other hand, let $B = \{ f < M_f - t \}$. As before we see that when $d_B(x) \leq t/ \| f \|_{\mathrm{Lip}}$, then $f(x) < M_f$; thus

$$\mu_N(d_B > t/ \| f \|_{\mathrm{Lip}}) \geq \frac{1}{2}$$

by definition of the median. But, again by Chebyshev's inequality and Theorem 1.3, we get

$$\mu_N(B) \leq 2\exp(-t^2/8 \parallel f \parallel_{\mathrm{Lip}}^2).$$

These two inequalities together then imply (1.9).

Theorem 1.3 and the subsequent concentration inequality (1.9) actually do not depend on N and easily extend to the case of Haar measure $\mu = (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1})^{\otimes \mathbb{N}}$ on $\{-1,+1\}^{\mathbb{N}}$. For example, concerning (1.9), if f on $\mathbb{R}^{\mathbb{N}}$ is convex and Lipschitzian in the sense that

$$|f(\alpha) - f(\beta)| \leq \parallel f \parallel_{\mathrm{Lip}} |\alpha - \beta|$$

for all α, β in ℓ_2, then similarly, for M_f a median of f for μ,

$$\mu(|f - M_f| > t) \leq 4\exp(-t^2/8 \parallel f \parallel_{\mathrm{Lip}}^2) \tag{1.10}$$

for all $t > 0$.

Compared with the corresponding inequalities (1.4) and (1.6) the coefficient 8 in (1.9) does not seem best possible. It is not known whether 2 can be reached, something which the argument of the proof of Theorem 1.3 cannot accomplish. Let us note further that the convexity assumption on f in (1.9) cannot be dropped. This is made clear by the following example. Let $A = \{x \in \{-1,+1\}^N;\ \sum_{i=1}^{N} x_i \leq 0\}$ and define $f(x) = \inf\{|x-y|;\ y \in A\}$. Clearly $\parallel f \parallel_{\mathrm{Lip}} = 1$ and 0 is a median of f. Assume N is an *even* integer. Then, as is easy to see, $f^2(x) = 2\left(\sum_{i=1}^{N} x_i\right)^{+}$; but then, from the central limit theorem, $\mu_N(f > cN^{1/4}) \geq 1/4$ for some $c > 0$ independent of N from which is is clear that the non-convex Lipschitzian function f cannot verify an inequality like (1.9).

Despite of these somewhat negative observations, Theorem 1.3 and the concentration inequality (1.9) will be used in Chapter 4 in the study of tail behavior and integrability properties of vector valued Rademacher series as efficiently as the Gaussian inequalities in Chapter 3.

1.2. An isoperimetric inequality for product measures

The preceding isoperimetric inequalities and concentration phenomena will be applied in the next chapters to the study of integrability properties and tail behaviors of Gaussian and Rademacher series with vector valued coefficients. In this section, we present an isoperimetric theorem for product measures which will be our key tool in the study of general sums of independent random variables (Chapter 6). Its discovery was actually motivated by these questions. The statement of the result is somewhat abstract but we will try by comments and ideas on the proof to clarify its powerful meaning.

Given a probability space (E, Σ, μ) and a fixed, but arbitrary, integer $N \geq 1$, denote by P the product measure $\mu^{\otimes N}$ on E^N. A point x in E^N has coordinates $x = (x_1, \ldots, x_N), x_i \in E$. To a subset A of E^N, we associate

$$H(A, q, k) = \{x \in E^N;\ \exists x^1, \ldots, x^q \in A \text{ such that Card } \{i \leq N;\ x_i \notin \{x_i^1, \ldots, x_i^q\}\} \leq k\}.$$

The set $H(A, q, k)$ can be thought of in an isoperimetric way as some neighborhood of A whose elements are determined by a fixed number q of points in A with at most k free coordinates. This can be made somewhat more precise in the terminology of the beginning of this chapter. For an element x in $(E^N)^q$, denote its coordinates by $x = (x^1, \ldots, x^q) = (x^\ell)_{\ell \leq q}$ where $x^\ell \in E^N$ and, as before, $x^\ell = (x_i^\ell)_{i \leq N}$. Between elements x, y in $(E^N)^q$ introduce

$$\begin{aligned} d(x, y) &= \sum_{i=1}^N I_{\{\forall \ell = 1, \ldots, q:\ x_i^\ell \neq y_i^\ell\}} \\ &= \text{ Card } \{i \leq N;\ \forall \ell = 1, \ldots, q,\ x_i^\ell \neq y_i^\ell\}. \end{aligned}$$

Then for A in $E^N, H(A, q, k)$ can simply be interpretated as the neighborhood of order k with respect to d in the sense that

$$H(A, q, k) = \{x \in E^N;\ d(\widetilde{x}, A^q) \leq k\}$$

where for x in $E^N, \widetilde{x}$ is the element of $(E^N)^q$ with coordinates $\widetilde{x} = (x, \ldots, x)$.

The isoperimetric theorem estimates the size of $H(A, q, k)$ under P in terms of $P(A), q$ and k. The main conclusion is an *exponential decay* in terms of k of the measure of the complement of $H(A, q, k)$.

Theorem 1.4. For some universal constant K,

$$P_*(H(A, q, k)) \geq 1 - \left[K\left(\frac{\log(1/P(A))}{k} + \frac{1}{q}\right)\right]^k$$

where P_* denotes inner probability.

The proof of Theorem 1.4 is isoperimetric in nature and relies on several reductions based on symmetrization (rearrangement) procedures. We below illustrate one typical argument in a particular case. It does not seem however to allow an exact solution of the isoperimetric problem which would be the determination, for any $a > 0$ of

$$\inf\{P_*(H(A,q,k));\ P(A) > a\}.$$

Note further that the use of inner probability is necessary since $H(A,q,k)$ need not be measurable when A is.

Theorem 1.4 is mostly used in application only in the typical case $P(A) \geq 1/2$ and $k \geq q$ for which we get that

$$P_*(H(A,q,k)) \geq 1 - \left(\frac{K_0}{q}\right)^k \tag{1.11}$$

where K_0 is a numerical constant. It will be convenient in the sequel to assume this constant K_0 to be an *integer*; we do this and K_0 will moreover always have the meaning of (1.11) throughout the book.

It has been remarked in [Ta11] that, in the case $P(A) \geq 1/2$ for example, (1.11) can actually be improved into

$$P_*(H(A,q,k)) \geq 1 - \left[K\left(\frac{1}{k} + \frac{1}{q\ \log\ q}\right)\right]^k. \tag{1.12}$$

The gain of the factor $\log q$ is irrelevant for the applications presented in this work. It should be noted however that this estimate is sharp. Indeed, consider the case where $E = \{0,1\}$ and $\mu = \left(1 - \frac{r}{N}\right)\delta_0 + \frac{r}{N}\delta_1$ where r is an integer less than N and N is assumed to be large. Let A in E^N be defined as

$$A = \{x \in E^N;\ \sum_{i=1}^N x_i \leq r\}.$$

Then $P(A)$ is of the order of $1/2$ and, clearly,

$$H(A,q,k) = \{x \in E^N;\ \sum_{i=1}^N x_i \leq rq + k\}.$$

Now

$$P(H(A,q,k)^c) \geq \left(\frac{r}{N}\right)^{rq+k+1} \left(1 - \frac{r}{N}\right)^{N-rq-k} \binom{N}{rq+k+1}$$
$$\geq \left(\frac{r}{N}\right)^{rq+k+1} e^{-2r} \left(\frac{N}{2(rq+k+1)}\right)^{rq+k+1}$$
$$\geq \left[\frac{1}{K}\left(\frac{1}{q+\frac{k}{r}}\right)\right]^{rq+k+1} .$$

If $k \geq q$ are fixed (large enough) and if we take r to be of the order of $k/(q \ \log \ q)$, we see that we have obtained an example for which the bound (1.12) is optimal.

As announced, we will only use (1.11) in applications. These will be mainly studied in Chapter 6 and in corollaries in Chapters 7 and 8 on strong limit theorems for sums of independent random variables. Let us therefore briefly indicate the trivial translation from the preceding abstract product measures to the setting of independent random variables. Let $\mathbb{X} = (X_i)_{i \leq N}$ be a sample of independent random variables with values in a measurable space E. By independence, they can be realized on some product probability space Ω^N, in such a way that for $\omega = (\omega_i)_{i \leq N}$ in Ω^N, $X_i(\omega)$ only depends on ω_i. Then (1.11) is simply that, when $k \geq q$ and $\mathbb{P}\{\mathbb{X} \in A\} \geq 1/2$ for some measurable set A in E^N, then

$$\mathbb{P}_*\{\mathbb{X} \in H(A,q,k)\} \geq 1 - \left(\frac{K_0}{q}\right)^k . \tag{1.13}$$

Hence, when k or q are large, the sample $\mathbb{X}$ falls with high probability into $H(A,q,k)$. On this set, $\mathbb{X}$ is entirely controlled by a finite number q of points in A provided k elements of the sample are neglected. In the applications, especially to the study of sums of independent random variables, these neglected terms can essentially be thought as the largest values of the sample. Hence, once an appropriate bound on large values have been found, a good choice of A and relations between the various parameters q and k determine sharp bounds on the tails of sums of independent random variables. This will be one of the objects of study in Chapter 6. Let us note further that this intuition about large values of the sample is justified in the special case of the proof of Theorem 1.4 we give below; the final binomial arguments exactly handle the situation as we just described.

We refer to [Ta11] for a detailed proof of Theorem 1.4. We would like however to give a proof in the simpler case where A is symmetric, i.e. invariant under permutation of the coordinates. The rearrangement part of the proof is an inequality introduced in [Ta6], in the same spirit, but easier, than the rearrangement arguments needed for Theorem 1.4. The explicit computations on the sets after appropriate rearrangement are then identical to those required to prove Theorem 1.4, and rely on classical binomial estimates. This method also yields a version of Theorem 1.4 in the case of symmetric A for $q > 1$ *not necessarily an integer.* This is another motivation for the details we are giving now. More precisely let, as before, $N \geq 1$ be an integer and let now $q > 1$; denote by N' the integer part of qN. Consider C symmetric (invariant under permutation of the coordinates) in $E^{N'}$ such that $P'(C) > 0$ where $P' = \mu^{\otimes N'}$. For each integer k, set

$$G(C,k) = \{x \in E^N;\ \exists y \in C \ \text{ such that } \ \operatorname{Card}\{i \leq N;\ x_i \notin \{y_1, \dots, y_{N'}\}\} \leq k\}.$$

For a comparison with $H(A,q,k)$ when q is an integer, let A in E^N and denote by $C \subset E^{qN}$ the set of all sequences $y = (y_i)_{i \leq qN}$ such that $\{y_1, \dots, y_{qN}\}$ can be covered by q sets of the form $\{x_1, \dots, x_N\}$ where $(x_i)_{i \leq N} \in A$. C is clearly invariant under permutation of the coordinates and $H(A,q,k) \subset G(C,k)$. On the other hand, it is not difficult to see that when A is symmetric, the converse inclusion $G(C,k) \subset H(A,q,k)$ is also satisfied at least on the subset of E^N consisting of those $x = (x_i)_{i \leq N}$ such that $x_i \neq x_j$ whenever $i \neq j$.

In these notations, we then have that there exist $K(q) < 1$ and $k(q, P'(C))$ large enough such that for all $k \geq k(q, P'(C))$,

$$P_*(G(C,k)) \geq 1 - k(q, P'(C))[K(q)]^k. \tag{1.14}$$

For simplicity, we do not indicate the explicit dependence of $K(q)$ and $k(q, P'(C))$ in function of q and $P'(C)$ but, as will be clear from the proof, these are similar to the ones explicited in Theorem 1.4.

In order to establish (1.14), we take upon the framework of [Ta11] and note in particular, to start with, that since the result is measure theoretic we might as well assume that $E = [0,1]$ and μ is Lebesgue measure λ on $[0,1]$. The main point in the proof of (1.14) is the use of Theorem 11 in [Ta6] which ensures the existence, for C symmetric, of a left-hereditary subset $\widetilde{C}$ of $]0,1[^{N'}$ such that $\lambda^{N'}(\widetilde{C}) = \lambda^{N'}(C)$ and for which, for every k,

$$\lambda_*^N(G(C,k)) \geq \lambda^N(G(\widetilde{C},k)).$$

$\widetilde{C}$ is left-hereditary in the sense that whenever $(y_i)_{i \leq n} \in \widetilde{C}$ and, for $1 \leq i \leq N', 0 < z_i < y_i$, then $(z_i)_{i \leq N'} \in \widetilde{C}$. (When $\widetilde{C}$ is left-hereditary, so is $G(\widetilde{C},k)$ which is therefore measurable, but in general $G(C,k)$ need not be measurable.) The conclusion now follows from an appropriate lower bound for

$\lambda(G(\widetilde{C},k))$ which will be obtained using binomial estimates. Convenient here is the following inequality:

$$\mathbb{P}\{B(n,\tau)\le tn\}\le\left[\left(\frac{\tau}{t}\right)^t\left(\frac{1-\tau}{1-t}\right)^{1-t}\right]^n \tag{1.15}$$

where $B(n,\tau)$ is the number of successes in a run of n Bernoulli trials with probability of success τ and $0<t<1$.

We let $q=(1+\varepsilon)^2, \varepsilon>0$, and first show that for some $\alpha=\alpha(\varepsilon,\lambda^{N'}(C))>0$ large enough, there exists $(y_i)_{i\le N'}$ in $\widetilde{C}$ such that, for every $1\le r\le N'$,

$$\text{Card }\{i\le N';\ \varepsilon_i>1-\frac{(1+\varepsilon)(r+\alpha)}{N'}\}\ge r.$$

Indeed,

$$\left(\frac{1-\tau}{1-t}\right)^{1-t}=\left(1-\frac{\tau-t}{1-t}\right)^{1-t}\le\exp(-(\tau-t))$$

so that, by (1.15),

$$\mathbb{P}\{B(n,\tau)\le tn\}\le\exp\left(-tn\left(\frac{\tau}{t}-1-\log\frac{\tau}{t}\right)\right).$$

If $u\ge 1+\varepsilon$, then $u-1-\log\ u\ge\delta^2 u$ where $\delta=\frac{1}{4}\min(\varepsilon,\frac{1}{2})$; hence, if $\tau/t\ge 1+\varepsilon$,

$$\mathbb{P}\{B(n,\tau)\le tn\}\le\exp(-\delta^2\tau n).$$

Therefore, by this inequality, for every $1\le r\le N$,

$$\begin{aligned}\lambda^{N'}((y_i)_{i\le N'};\ \text{Card }\{i\le N';\ \varepsilon_i>1-\frac{(1+\varepsilon)(r+\alpha)}{N'}\}\le r-1)\\ \le\exp(-\delta^2(1+\varepsilon)(r+\alpha))\end{aligned}$$

and

$$\sum_{r\ge1}\exp(-\delta^2(1+\varepsilon)(r+\alpha))<\lambda^{N'}(C)$$

whenever $\alpha=\alpha(\varepsilon,\lambda^{N'}(C))$ is chosen large enough. This proves the preceding claim.

Now, since $\widetilde{C}$ is left-hereditary, each sequence $(x_i)_{i\le N}$ such that for every $r>k$,

$$\text{Card }\{i\le N;\ \varepsilon_i\ge 1-\frac{(1+\varepsilon)(r-k+\alpha)}{N'}\}\le r-1$$

belongs to $G(\widetilde{C},k)$; indeed, the r -th largest element of $(x_i)_{i\le N}$ is less than $1-(1+\varepsilon)(r-k+\alpha)/N'$ and is therefore smaller than the $(r-k)$ -th largest element of $(y_i)_{i\le N'}$. Thus, by the left-hereditary property of $\widetilde{C}, (x_i)_{i\le N}\in G(\widetilde{C},k)$. The proof of (1.14) will therefore be completed if we can show that

$$\begin{aligned}\lambda^N((x_i)_{i\le N};\ \forall r>k,\ \ \text{Card}\,\{i\le N;\ \varepsilon_i>1-\frac{(1+\varepsilon)(r-k+\alpha)}{N'}\}\le r-1)\\ \ge 1-k(\varepsilon,\lambda^{N'}(C))[K(\varepsilon)]^k\end{aligned}$$

for some $K(\varepsilon)<1$ and all $k\ge k(\varepsilon,\lambda^{N'}(C))$ large enough. To this aim, note that

$$\left(\frac{\tau}{t}\right)^t\le\left(1+\frac{\tau-t}{t}\right)^t\le\exp(\tau-t)$$

and thus, by (1.15),

$$\mathbb{P}\{B(n,\tau)\le tn\}\le\exp\left(-(1-t)n\left(\frac{1-\tau}{1-t}-1-\log\frac{1-\tau}{1-t}\right)\right). \tag{1.16}$$

If $0<u<(1+\varepsilon)^{-1}$, then $u-1-\log\ u\ge\delta^2$ (where we recall that $\delta=\frac{1}{4}\min\left(\varepsilon,\frac{1}{2}\right))$. Hence, if $(1-\tau)/(1-t)\le 1/1+\varepsilon$,

$$\mathbb{P}\{B(n,\tau)\le tn\}\le\exp(-\delta^2(1-t)n).$$

Using this inequality, if $r>k$ and $k\ge k(\varepsilon,\lambda^{N'}(C))$ large enough, we get that

$$\lambda^N((x_i)_{i\le N};\ \ \text{Card}\,\{i\le N;\ \varepsilon_i\ge 1-\frac{(1+\varepsilon)(r-k+\alpha)}{N'}\}\ge r)\le\exp(-\delta^2 r).$$

As announced, this is exactly what was required to conclude the proof and therefore the isoperimetric result (1.14) is established.

1.3. Martingale inequalities

Martingale methods prove useful in order to establish various concentration results. These complement the preceding isoperimetric inequalities and will prove useful in various places throughout this work. The inequalities we state are rather classical, at least some of them, and we present them in the general spirit of concentration properties.

Recall $L_1=L_1(\Omega,\mathcal{A},\mathbb{P})$ denotes the space of all real measurable functions f on Ω such that $\mathbb{E}|f|=\int|f|d\mathbb{P}<\infty$. Assume that we are given a filtration

$$\{\phi,\Omega\}=\mathcal{A}_0\subset\mathcal{A}_1\subset\cdots\subset\mathcal{A}_N=\mathcal{A}$$

of sub- σ -algebras of $\mathcal{A}$. $\mathbb{E}^{\mathcal{A}_i}$ denotes the conditional operator with respect to $\mathcal{A}_i$. Given f in L_1 , set, for each $i = 1, \dots, N$,

$$d_i = \mathbb{E}^{\mathcal{A}_i} f - \mathbb{E}^{\mathcal{A}_{i-1}} f$$

so that $f - \mathbb{E} f = \sum_{i=1}^N d_i$. $(d_i)_{i \le N}$ defines a so-called *martingale difference sequence* characterized by the property $\mathbb{E}^{\mathcal{A}_{i-1}} d_i = 0,\ i \le N$.

One of the typical examples of martingale differences sequence we have in mind is a sequence $(X_i)_{i \le N}$ of *independent mean zero* random variables. Indeed, if $\mathcal{A}_i$ denotes the σ -algebra generated by the variables $X_1, \dots, X_i$, by independence and the mean zero property, it is clear that $\mathbb{E}^{\mathcal{A}_{i-1}} X_i = \mathbb{E} X_i = 0$. Hence, all the results we will present for $f - \mathbb{E} f = \sum_{i=1}^N d_i$ as before apply to the sum $\sum_{i=1}^N X_i$.

The first lemma is a kind of analog in this context of the concentration property for Lipschitzian functions. It expresses, in the preceding notations, high concentration of f around its expectation in terms of the size of the differences d_i .

Lemma 1.5. Let f in L_1 and let $f - \mathbb{E} f = \sum_{i=1}^N d_i$ be as before the sum of martingale differences with respect to $(\mathcal{A}_i)_{i \le N}$. Assume that $\| d_i \|_\infty < \infty$ and set $a = (\sum_{i=1}^N \| d_i \|_\infty^2)^{1/2}$. Then, for every $t > 0$,

$$\mathbb{P}\{|f - \mathbb{E} f| > t\} \le 2\ \exp(-t^2/2a^2).$$

Proof. We first note that when φ is a random variable such that $|\varphi| \le 1$ almost surely and $\mathbb{E}\varphi = 0$, then, for any real number $\lambda, \mathbb{E}\ \exp \lambda\varphi \le \exp(\lambda^2/2)$. Indeed, simply note from the convexity of $x \to \exp(\lambda x)$ and $\lambda x = \lambda(1+x)/2 - \lambda(1-x)/2$ that, for any $|x| \le 1$,

$$\exp(\lambda x) \le \ \mathrm{ch}\ \lambda + x\ \mathrm{sh}\ \lambda \le \exp(\lambda^2/2) + x\ \mathrm{sh}\ \lambda$$

and integrating yields the claim. It clearly follows that, for any $i = 1, \dots, N$,

$$\mathbb{E}^{\mathcal{A}_{i-1}} \exp \lambda d_i \le \exp(\lambda^2 \| d_i \|_\infty^2 /2).$$

Iterating by the properties of conditional expectation,

$$\begin{aligned}\mathbb{E}\exp[\lambda(f-\mathbb{E}f)] &= \mathbb{E}\exp(\lambda\sum_{i=1}^{N} d_i)\\ &= \mathbb{E}(\exp(\lambda\sum_{i=1}^{N-1} d_i)\mathbb{E}^{\mathcal{A}_{N-1}}\exp\lambda d_N)\\ &\le \mathbb{E}\exp(\lambda\sum_{i=1}^{N-1} d_i)\exp(\lambda^2 \parallel d_N \parallel_\infty^2 /2)\\ &\dots\\ &\le \exp(\lambda^2 a^2/2)\ .\end{aligned}$$

We then obtain from Chebyshev's inequality that, for $t>0$,

$$\begin{aligned}\mathbb{P}\{f-\mathbb{E}f>t\} &\le \exp(-\lambda t+\lambda^2 a^2/2)\\ &\le \exp(-t^2/2a^2)\end{aligned}$$

for the optimal choice of λ. Applying this inequality also to $-f$ then yields the conclusion of the lemma.

We should point out at this stage, and once for all, that in a statement like Lemma 1.5, it is understood that if we are interested only in $f-\mathbb{E}f$ rather than $|f-\mathbb{E}f|$, we also have

$$\mathbb{P}\{f-\mathbb{E}f>t\}\le\exp(-t^2/2a^2).$$

(This was actually explicitly extablished in the proof!) This general comment about a coefficient 2 in front of the exponential bound in order to take into account absolute values (f and $-f$) applies in many similar situations (we already mentioned it about the concentration inequalities of Section 1.1) and will be applied usually without any further comment in the sequel.

When, in addition to the bounds on d_i, some information on $\mathbb{E}^{\mathcal{A}_{i-1}}d_i^2$ is available, the preceding proof basically yields the following refinement.

Lemma 1.6. Let $f-\mathbb{E}f=\sum_{i=1}^N d_i$ be as before. Set $a=\max_{i\le N}\parallel d_i\parallel_\infty$ and let $b\ge(\sum_{i=1}^N\parallel\mathbb{E}^{\mathcal{A}_{i-1}}d_i^2\parallel_\infty)^{1/2}$. Then, for every $t>0$,

$$\mathbb{P}\{|f-\mathbb{E}f|>t\}\le 2\ \exp\left[-\frac{t^2}{2b^2}\left(2-\exp\left(\frac{at}{b^2}\right)\right)\right].$$

The proof is similar to the one of Lemma 1.5. It uses simply that

$$\begin{aligned}\mathbb{E}^{\mathcal{A}_{i-1}} \exp \lambda d_i &= 1 + \frac{\lambda^2}{2!}\mathbb{E}^{\mathcal{A}_{i-1}} d_i^2 + \frac{\lambda^3}{3!}\mathbb{E}^{\mathcal{A}_{i-1}} d_i^3 + \dots \\ &\le 1 + \frac{\lambda^2}{2} \parallel \mathbb{E}^{\mathcal{A}_{i-1}} d_i^2 \parallel_\infty \left(1 + \frac{\lambda \parallel d_i \parallel_\infty}{3} + \frac{\lambda^2 \parallel d_i \parallel_\infty^2}{3.4} + \dots\right) \\ &\le \exp\left[\frac{\lambda^2}{2} \parallel \mathbb{E}^{\mathcal{A}_{i-1}} d_i^2 \parallel_\infty e^{\lambda a}\right] .\end{aligned}$$

Turning back to Lemma 1.5, it is clear that we always have

$$|f - \mathbb{E}f| \le \sum_{i=1}^N \parallel d_i \parallel_\infty .$$

This simple observation of course suggests the possibility of some interpolation between the sum of the squares $(\sum_{i=1}^N \parallel d_i \parallel_\infty^2)^{1/2}$ which steps in in Lemma 1.5 and this trivial bound. This kind of result is described in the next two lemmas.

Lemma 1.7. Let $1 < p < 2$ and let $q = p/p-1$ denote the conjugate of p. Let further f be as before with $f - \mathbb{E}f = \sum_{i=1}^N d_i$ and set now $a = \max\limits_{i \le N} i^{1/p} \parallel d_i \parallel_\infty$. Then, for every $t > 0$,

$$\mathbb{P}\{|f - \mathbb{E}f| > t\} \le 2 \exp(-t^q/C_q a^q)$$

where $C_q > 0$ only depends on q.

Proof. By homogeneity we may and do assume that $a = 1$. For any integer m we can write

$$\begin{aligned}|f - \mathbb{E}f| &\le \sum_{i=1}^n |d_i| + |\sum_{i>m} d_i| \\ &\le \sum_{i=1}^m i^{-1/p} + |\sum_{i>m} d_i| \le q m^{1/q} + |\sum_{i>m} d_i| .\end{aligned}$$

Assume first that $t > 2q$ and denote then by m the largest integer such that $t > 2qm^{1/q}$. We can apply Lemma 1.5 to $\sum\limits_{i>m} d_i$ and thus obtain in this way, together with the preceding “interpolation” inequality,

$$\begin{aligned}\mathbb{P}\{|f - \mathbb{E}f| > t\} &\le \mathbb{P}\{|\sum_{i>m} d_i| > qm^{1/q}\} \\ &\le 2 \exp\left[-q^2 m^{2/q}/2 \sum_{i>m} \parallel d_i \parallel_\infty^2\right] .\end{aligned}$$

Now, since $a \leq 1$,

$$\sum_{i>m} \| d_i \|_\infty^2 \leq \sum_{i>m} i^{-2/p} \leq \frac{q}{q-2} m^{1-2/p}$$

so that

$$\begin{aligned} \mathbb{P}\{|f - \mathbb{E}f| > t\} &\leq 2 \ \exp(-q(q-2)m/2) \\ &\leq 2 \ \exp(-t^q / C_q^1) \end{aligned}$$

where $C_q^1 = 4(2q)^q / q(q-2)$. When $t \leq 2q$,

$$\mathbb{P}\{|f - \mathbb{E}f| > t\} \leq 1 \leq 2 \ \exp(-(2q)^q / C_q^2) \leq 2 \ \exp(-t^q / C_q^2)$$

where $C_q^2 = (2q)^q / \log \ 2$. The lemma then follows with $C_q = \max(C_q^1, C_q^2)$.

Lemma 1.8. Let $f - \mathbb{E}f = \sum_{i=1}^N d_i$ be as above and set now $a = \max\limits_{i \leq N} i \| d_i \|_\infty$. Then, for every $t > 0$,

$$\mathbb{P}\{|f - \mathbb{E}f| > t\} \leq 16 \exp[-\exp(t/4a)].$$

Proof. It is similar to the preceding one. We again assume by homogeneity that $a = 1$. When $t \leq 4$,

$$\mathbb{P}\{|f - \mathbb{E}f| > t\} \leq 16 e^{-e} \leq 16 \exp[-\exp(t/4)].$$

When $t > 4$, let m be the largest integer such that $t \geq 2 + \log \ m$. We have as before that

$$|f - \mathbb{E}f| \leq \sum_{i=1}^m |d_i| + |\sum_{i>m} d_i| \leq (1 + \log \ m) + |\sum_{i>m} d_i|.$$

Hence

$$\mathbb{P}\{|f - \mathbb{E}f| > t\} \leq \mathbb{P}\{|\sum_{i>m} d_i| > 1\} \leq 2 \exp\left(-\frac{1}{2} \sum_{i>m} i^{-2}\right) \leq 2 \exp(-m/2)$$

where we have used Lemma 1.5. Since $t < 2 + \log(m+1)$, we get

$$\mathbb{P}\{|f - \mathbb{E}f| > t\} \leq 4 \exp[-\exp(t/4)]$$

and the conclusion follows.

Notes and references

The description of the concentration of measure phenomenon is taken from the paper [G-M] by M. Gromov and V. D. Milman were further interesting examples of "Lévy' families" are discussed (see also [Mi-S], [Mi3]). The use of isoperimetric concentration properties in Dvoretzky's theorem on almost spherical sections of convex bodies (see Chapter 9) was initiated by V. D. Milman [Mi1], amplified later in [F-L-M]. Further applications to local theory of Banach spaces are presented in [Mi-S], [Pi18], [TJ2].

The isoperimetric inequality on the sphere (Theorem 1.1) is due to P. Lévy [Lé2] and E. Schmidt [Schm]. Lévy's proof, that has not been understood for a long time, has been revived and generalized by M. Gromov [Gr1], [Mi-S]. Schmidt's proof is based on deep isoperimetric symmetrizations (rearrangements in the sense of Steiner) arguments. Accounts on symmetrizations and rearrangements, geometric inequalities and isoperimetry may be found in [B-Z], [Os]. For a short proof of Theorem 1.1, we refer to [F-L-M], or [Ba-T], [Beny] (for the two point symmetrization method). Poincaré's Lemma is *not* to be found in [Po] according to [D-F]; see this paper for the history of the result. Poincaré's Lemma is nicely revisited in [MK]. The Gaussian isoperimetric Theorem 1.2 is due independently to C. Borell [Bo2] and V. N. Sudakov and B. S. Tsirel'son [S-T] with the proof sketched here. A. Ehrhard introduced Gaussian symmetrization in [Eh1] and established there inequality (1.1). See further [Eh2] and also [Eh3] where extremality of half spaces is investigated. We refer the reader to [Ta19] where a new isoperimetric inequality is presented, that improves upon certain aspects of the Gaussian isoperimetric theorem. Log-concavity of Gaussian Radon measures in locally convex spaces has been shown by C. Borell [Bo1]. The simple proof we suggest has been shown to us by J. Saint-Raymond. Inequality (1.5) is due to G. Pisier with the simple proof of B. Maurey [Pi16]. They actually deal with vector valued functions and the method of proof indeed ensures in the same way that if $f : E \to G$ is locally Lipschitzian between two Banach spaces E and G, if γ is a Gaussian Radon measure on E and if $F : G \to \mathbb{R}$ is measurable and convex, then

$$\int F(f - \int f d\gamma) d\gamma \leq \int\int F\left(\frac{\pi}{2} f'(x) \cdot y\right) d\gamma(x) d\gamma(y) .$$

(1.6) comes from B. Maurey (cf. [Pi16], [Led7]). A proof of (1.6) using Yurinskii's observation (see Chapter 6 below) and a central limit theorem for martingales has been noticed by A. de Acosta and J. Zinn (oral communication). (1.7) and the subsequent examples were observed by G. Pisier [Pi16]. The isoperimetric theorem for Haar measure on $\{-1,+1\}^N$ with respect to the Hamming metric was established by L. H. Harper [Har]. Theorem 1.3 may be found in [Ta9] where comparison with [Har] is discussed. Extensions to measures on $\{-1,+1\}^N$ with non-symmetric weights are described in [J-S2].

The isoperimetric theorem for subsets of a product of probability spaces, Theorem 1.4, is due to the second author [Ta11]. Inequality (1.14) for q not necessarily an integer is new, and is also due to the second author. The binomial computations closely follow the last step in [Ta11]. (1.15) comes from [Che].

The first two inequalities of Section 1.3 are rather classical. In this form, Lemma 1.5 is apparently due to [Azu]. Lemma 1.6 is the martingale analogue of the classical exponential inequality of Kolmogorov [Ko] (see [Sto]), in a form put forward in [Ac7]. For sums of independent random variables, and starting with Bernstein's inequality (cf. [Ho]) this type of inequality has been extensively studied leading to sharp versions in, for example, [Ben], [Ho] (see also Chapter 6). We use for simplicity Lemma 1.6 in this work but the preceding references can basically be used equivalently in our applications. Lemmas 1.6 and 1.7 are taken from [Pi16]. For applications of all these inequalities to Banach space theory, see, besides others, [Mau3], [Sch1], [Sch2], [Sch3], [J-S1], [Pi12], [B-L-M], [Mi-S], [Pi16], etc. For application of the martingale method to rather different problems, see [R-T1], [R-T2], [R-T3].

Chapter 2. Generalities on Banach space valued random variables and random processes

Chapter 2. Generalities on Banach space valued random variables and random processes

This chapter collects in a rather informal way some basic facts about processes and infinite dimensional random variables. The material we present actually only appears as the necessary background for the subsequent analysis developed in the next chapters. Only a few proofs are given and many important results are only just mentioned or even omitted. It is therefore recommended to complement if necessary these partial basis with the classical references, some of which are given at the end of the chapter.

The first section describes Radon (or separable) vector valued random variables while the second makes precise some terminology and definitions about random processes and general vector valued random variables. The third one presents some important facts about symmetric random variables, especially Lévy's inequalities and Itô-Nisio's theorem. In a last paragraph, we mention some classical and useful inequalities.

Throughout this book we deal with abstract probability spaces $(\Omega, \mathcal{A}, \mathbb{P})$ which are always assumed to be large enough in order to support all the random variables we will work with; this is legitimate by Kolmogorov's extension theorem. We also assume for convenience that $(\Omega, \mathcal{A}, \mathbb{P})$ is complete, that is the σ-algebra $\mathcal{A}$ contains the negligible sets for $\mathbb{P}$.

Throughout this book also, B denotes a Banach space, that is a vector space over $\mathbb{R}$ or $\mathbb{C}$ with norm $\|\cdot\|$ and complete with respect to it. For simplicity, we shall always consider *real* Banach spaces, but actually almost everything we will present carries over to the complex case. B' denotes the topological dual of B and $f(x) = \langle f, x\rangle$ $(\in \mathbb{R})$, $f \in B'$, $x \in B$, the duality. The norm on B' is also denoted $\|f\|$, $f \in B'$.

2.1 Banach space valued Radon random variables

A Borel random variable or vector with values in a Banach space B is a measurable map X from some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into B equipped with its Borel σ-algebra $\mathcal{B}$ generated by the open sets of B. In fact, this definition of random variable is somewhat too large for our purposes since for example the sum of two random variables is not trivially a random variable. Furthermore, if B is equipped with a different σ-algebra, like for example the coarsest one for which the linear functionals are measurable (cylindrical σ-algebra), the two definitions might not agree in general. We do not wish here to deal at length with measurability questions. One way to handle them is the concept of Radon or regular random variables, which amounts to some separability of the range.

A Borel random variable X with values in B is said to be regular with respect to compact sets, or *Radon*, or yet *tight*, if, for each $\varepsilon > 0$ there is a compact set $K = K(\varepsilon)$ in B such that

$$\mathbb{P}\{X \in K\} \geq 1 - \varepsilon \,. \tag{2.1}$$

In other words, the image of the probability $\mathbb{P}$ by X (see below) is a Radon measure on $(B, \mathcal{B})$. Equivalently X takes almost all its values in some separable (i.e. countably generated) closed linear subspace E of B. Indeed, under (2.1), there exists a sequence (K_n) of compact sets in B such that $\mathbb{P}\{X \in \bigcup_n K_n\} = 1$ so that X takes almost surely its values in some separable subspace of B. Conversely, let $\{x_i \,;\, i \in \mathbb{N}\}$ be dense in E and let $\varepsilon > 0$ be fixed. By density, for each $n \geq 1$ there exists an integer N_n such that

$$\mathbb{P}\{X \in \bigcup_{i \leq N_n} B(x_i, 2^{-n})\} \geq 1 - \varepsilon \cdot 2^{-n}$$

where $B(x_i, 2^{-n})$ denotes the closed ball of center x_i and radius 2^{-n} in B. Set then

$$K = K(\varepsilon) = \bigcap_{n \geq 1} \bigcup_{i \leq N_n} B(x_i, 2^{-n}) \,.$$

K is closed and $\mathbb{P}\{X \in K\} \geq 1 - \varepsilon$; further, K is compact since from each sequence in K one can extract a subsequence contained, for every n, in a single ball $B(x_i, 2^{-n})$; this subsequence is, therefore, a Cauchy sequence hence convergent by completness of B.

We call thus *Radon*, or *separable*, a Borel random variable X satisfying (2.1). The preceding argument shows equivalently that X is almost sure limit of step random variables of the form $\sum_{\text{finite}} x_i I_{A_i}$ where $x_i \in B$ and $A_i \in \mathcal{A}$. Note also that (2.1) is extended into

$$\mathbb{P}\{X \in A\} = \sup\{\mathbb{P}\{X \in K\} \,;\, K \text{ compact}, \ K \subset A\}$$

for every Borel set A in B. This follows from (2.1) together with the analogous property for closed sets which holds from the very definition of the Borel σ-algebra.

Since Radon random variables have separable range, it is sometimes convenient to assume the Banach space itself to be separable. We are mostly interested in this work in results concerning *sequences* of random variables. When dealing with sequences of Radon random variables, we will therefore usually assume for convenience and without any loss in generality the Banach space to be *separable*. Note further that when

B is separable all the "reasonable" definitions of σ-algebras on B coincide, in particular the Borel and cylindrical σ-algebras. Note also, and this observation will motivate parts of the next section, that if B is separable the norm can be expressed as a supremum

$$\|x\| = \sup_{f \in D} |f(x)| , \qquad x \in B ,$$

over a *countable* set D of linear functionals of norm 1 (or less than or equal to 1).

For a random variable X with values in B, the probability measure $\mu = \mu_X$ image of $\mathbb{P}$ by X is called the distribution (or law) of X; for any real bounded measurable φ on B,

$$\mathbb{E}\varphi(X) = \int_B \varphi(x) d\mu(x) .$$

The distribution of a Radon random variable is completely determined by its finite dimensional projections. More precisely, if X and Y are Radon random variables such that for all f in B', $f(X)$ and $f(Y)$ have (as real random variables) the same distribution, then $\mu_X = \mu_Y$. Indeed, we can assume B separable; the Borel σ-algebra is therefore generated by the algebra of cylinder sets. Since $\mu_{f(X)} = \mu_{f(Y)}$, μ_X and μ_Y agree on this algebra and the result follows. Note that it suffices to know that $f(X)$ and $f(Y)$ have the same law on some weakly dense subset of B'. As a consequence of the preceding, and according to the uniqueness theorem in the scalar case, the characteristic functionals on B'

$$\mathbb{E}\exp if(X) = \int_B \exp(if(x)) d\mu(x) , \qquad f \in B' ,$$

completely determines the distribution of X.

Denote by $\mathcal{P}(B)$ the space of all Radon probability measures on B. For each μ in $\mathcal{P}(B)$ consider the neighborhood

$$\{\nu \in \mathcal{P}(B) ; \, |\int_B \varphi_i d\mu - \int_B \varphi_i d\nu| < \varepsilon , \, i \leq N\}$$

where $\varepsilon > 0$ and φ_i, $i \leq N$, are real bounded continuous on B. The topology generated by these neighborhoods is called the weak topology and a sequence (μ_n) in $\mathcal{P}(B)$ converging with respect to this topology is said to converge *weakly*. Observe that $\mu_n \to \mu$ weakly if and only if

$$\lim_{n \to \infty} \int \varphi d\mu_n = \int \varphi d\mu$$

for every bounded continuous φ on B. It can be shown further that this holds if and only if

$$\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$$

for each closed set F in B, or, equivalently,

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$$

for each open set G.

The space $\mathcal{P}(B)$ equipped with the weak topology is known to be a complete metric space (separable if B is separable). Thus, in particular, in order to check that a sequence (μ_n) in $\mathcal{P}(B)$ converges weakly, it suffices to show that (μ_n) is relatively compact in the weak topology and that all possible limits are the same. The latter can be verified along linear functionals. For the former, a very useful criterion of Prokhorov characterizes relatively compact sets of $\mathcal{P}(B)$ as those which are *uniformly tight* with respect to compact sets.

Theorem 2.1. A family $(\mu_i)_{i\in I}$ in $\mathcal{P}(B)$ is relatively compact for the weak topology if and only if for each $\varepsilon > 0$ there is a compact set K in B such that

$$\mu_i(K) \geq 1 - \varepsilon \qquad \text{for all } i \in I . \tag{2.2}$$

This compactness criterion may be expressed in various manners depending on the context. For example, (2.2) holds if and only if for each $\varepsilon > 0$ there is a *finite* set A in B such that

$$\mu_i(x \in B\,;\ d(x,A) < \varepsilon) \geq 1 - \varepsilon \quad \text{for all } i \in I \tag{2.3}$$

where $d(x,A)$ denotes the distance (in B) from the point x to the set A.

Another equivalent formulation of Theorem 2.1 is based on the idea of finite dimensional approximation and is most useful in applications. It is sometimes referred to as "flat concentration". The idea is simply that bounded sets in finite dimension are relatively compact and therefore if a set of measures is concentrated near a finite dimensional subspace it should be close to be relatively compact. The following simple functional analytic lemma makes this clear. If F is a closed subspace of B, denote by $T = T_F$ the canonical quotient map $T : B \to B/F$. Then $\|T(x)\| = d(x,F)$, $x \in B$. (We denote in the same way the norm of B and the norm of B/F.)

Lemma 2.2. A subset K of B is relatively compact if and only if it is bounded and for each $\varepsilon > 0$ there is a finite dimensional subspace F of B such that if $T = T_F$, $\|T(x)\| < \varepsilon$ for every x in K (i.e. $d(x, F) < \varepsilon$ for $x \in K$).

According to this result and to Theorem 2.1, if, for each $\varepsilon > 0$, there is a bounded set L in B such that $\mu_i(L) \geq 1 - \varepsilon$ for every $i \in I$ and a finite dimensional subspace F of B such that

$$\mu_i(x \in B\,;\ d(x, F) < \varepsilon) \geq 1 - \varepsilon \quad \text{for all } i \in I\,, \tag{2.4}$$

then the family $(\mu_i)_{i \in I}$ is relatively compact.

Actually, when (2.4) holds, the existence of L is a too strong hypothesis and it is enough to assume that for all f in B', $(\mu_i \circ f^{-1})_{i \in I}$ is a weakly relatively compact family of probability measures on the line. If μ is a Borel measure on B and f a linear functional, $\mu \circ f^{-1}$ denotes the measure on $\mathbb{R}$ image of μ by f. To check the preceding claim, let F of dimension N such that (2.4) is satisfied. By the Hahn-Banach theorem, there exist linear functionals $f_1, \ldots, f_n$ in the unit ball of B' such that, whenever $x \in F$ and $a > 0$, if $\max_{i \leq n} |f_i(x)| \leq a$, then $\|x\| \leq 2a$. Therefore, if $(\mu_i \circ f^{-1})_{i \in I}$ is relatively compact for every f in B' (actually a weakly dense subset of B' would suffice), and if (2.4) holds, the family $(\mu_i)_{i \in I}$ is uniformly almost concentrated on a bounded set and Prokhorov's criterion is then fulfilled.

A sequence (X_n) of Radon random variables with values in B converges *weakly* to a Radon random variable X if the sequence of distributions (μ_{X_n}) converges weakly to μ_X. For *real* random variables, a celebrated theorem of P. Lévy indicates that (X_n) converges weakly if and only if the corresponding sequence of characteristic functions (Fourier transforms) converges pointwise (to a continuous limit). In the vector valued case, by Theorem 2.1, (X_n) converges weakly to X as soon as $(f(X_n))$ converges weakly (as a sequence of real random variables) to $f(X)$ for all f in B' (or only in a weakly dense subset) *and* the sequence (X_n) is tight in the sense that, for each $\varepsilon > 0$, there exists a compact set K in B with

$$\mathbb{P}\{X_n \in K\} \geq 1 - \varepsilon$$

for all n (or only all n large enough since the X_n's are themselves tight). By what preceeds, this can be established through (2.3) or (2.4).

The sequence (X_n) is said to converge in *probability* (in measure) to X if, for each $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\{\|X_n - X\| > \varepsilon\} = 0\,.$$

It is said to be bounded in probability (or stochastically bounded) if, for each $\varepsilon > 0$, one can find $A > 0$ such that

$$\sup_n \mathbb{P}\{\|X_n\| > A\} < \varepsilon .$$

The topology of convergence in probability is metrizable and a possible metric can be given by $\mathbb{E}\min(1, \|X - Y\|)$. Denote by $L_0(B) = L_0(\Omega, \mathcal{A}, \mathbb{P}; B)$ the vector space of all random variables (on $(\Omega, \mathcal{A}, \mathbb{P})$) with values in B equipped with the topology of convergence in probability. If (X_n) converges in $L_0(B)$, it also converges weakly and the converse holds true if the limiting distribution is concentrated on one point. If thus one has to check for example that (X_n) converges to 0 in probability, it suffice to show that the sequence (X_n) is tight and that all possible limits are 0. The L_0 and weak topologies are thus close in this sense and may be considered as *weak* statements as opposed to the *strong* almost sure properties (defined below).

If $0 < p \le \infty$, denote by $L_p(B) = L_p(\Omega, \mathcal{A}, \mathbb{P}; B)$ the space of all random variables X (on $(\Omega, \mathcal{A}, \mathbb{P})$) with values in B such that $\|X\|^p$ is integrable:

$$\mathbb{E}\|X\|^p = \int \|X\|^p d\mathbb{P} < \infty , \qquad p < \infty$$

and $\|X\|_\infty = \operatorname{ess\,sup} \|X\| < \infty$ if $p = \infty$. If $B = \mathbb{R}$, we set simply $L_p = L_p(\mathbb{R})$ ($0 \le p \le \infty$). We denote moreover, both in the scalar and vector valued cases (and without confusion), by $\|X\|_p$ the quantity $(\mathbb{E}\|X\|^p)^{1/p}$. The spaces $L_p(B)$ are Banach spaces for $1 \le p \le \infty$ (metric vector spaces for $0 \le p < 1$). If (X_n) converges to X in $L_p(B)$, it converges to X in $L_0(B)$, that is in probability, and a fortiori weakly.

Finally, a sequence (X_n) converges *almost surely* (almost everywhere) to X if

$$\mathbb{P}\{\lim_{n\to\infty} X_n = X\} = 1 .$$

The sequence (X_n) is almost surely bounded if

$$\mathbb{P}\{\sup_n \|X_n\| < \infty\} = 1 .$$

Almost sure convergence is not metrizable. It clearly implies convergence in probability which in turn implies weak convergence. Conversely, an important theorem of Skorokhod [Sk1] asserts that if (X_n) converges weakly to X, there exist, on a possibly richer probability space, random variables X_n' and X' such that $\mu_{X_n'} = \mu_{X_n}$ for every n and $\mu_X = \mu_{X'}$ and such that $X_n' \to X'$ almost surely. This property is useful in particular in convergence of moments, for example in central limit theorems.

We conclude this section with some remarks concerned with integrability. As we have already seen, a Radon random variable X on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in B belongs to $L_1(B)$, or is *strongly* or *Bochner integrable*, if the real random variable $\|X\|$ is integrable $(\mathbb{E}\|X\| < \infty)$. Suppose now we are given X such that for each f in B' the real random variable $f(X)$ is integrable. If we consider the operator

$$T : B' \to L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$$

defined by $Tf = f(X)$, T has clearly a closed graph. T is therefore a bounded operator from which we deduce that $f \to \mathbb{E}f(X)$ defines a continuous linear map on B', that is an element, let us call it z, of the bidual B'' of B. The Radon random variable X is said to be *weakly or Pettis integrable* if for each f in B', $f(X)$ is integrable, and the element z of B'' just constructed actually belongs to B. If this is the case, z is then denoted by $\mathbb{E}X$.

It is not difficult to see that if the Radon random variable X is strongly integrable it is weakly integrable and $\|\mathbb{E}X\| \le \mathbb{E}\|X\|$. Indeed, we can choose, for each $\varepsilon > 0$, a compact set K in B such that $\mathbb{E}(\|X\| I_{\{>X \notin K\}}) < \varepsilon$. Let $(A_i)_{i \le N}$ be a finite partition of K with sets of diameter less than ε and fix for each i a point x_i in A_i. Set then

$$Y(\varepsilon) = \sum_{i=1}^{N} x_i I_{\{X \in A_i\}} .$$

It is plain by construction that $\mathbb{E}\|X - Y(\varepsilon)\| < 2\varepsilon$ and that $Y(\varepsilon)$ is weakly integrable with expectation $\mathbb{E}Y(\varepsilon) = \sum_{i=1}^{N} x_i \mathbb{P}\{X \in A_i\}$. The conclusion then follows from the fact that $(\mathbb{E}Y(1/n))$ is a Cauchy sequence in B which converges therefore to an element $\mathbb{E}X$ in B satisfying $f(\mathbb{E}X) = \mathbb{E}f(X)$ for every f in B'.

In the same way, conditional expectation of vector valued random variables can be constructed. Let X be a Radon random variable in $L_1(\Omega, \mathcal{A}, \mathbb{P}; B)$ and let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{A}$. Then one can define $\mathbb{E}^{\mathcal{F}} X$ as the Radon random variable in $L_1(\Omega, \mathcal{F}, \mathbb{P}; B)$ such that, for any F in $\mathcal{F}$,

$$\int_F \mathbb{E}^{\mathcal{F}} X \, d\mathbb{P} = \int_F X \, d\mathbb{P} .$$

It satisfies $\|\mathbb{E}^{\mathcal{F}} X\| \le \mathbb{E}^{\mathcal{F}} \|X\|$ almost surely and $f(\mathbb{E}^{\mathcal{F}} X) = \mathbb{E}^{\mathcal{F}} f(X)$ almost surely for every f in B'. Note further that, by separability and extension of a classical martingale theorem, if X is in $L_1(\Omega, \mathcal{A}, \mathbb{P}; B)$, there exists a sequence $(\mathcal{A}_N)$ of *finite* sub-σ-algebras of $\mathcal{A}$ such that if $X_N = \mathbb{E}^{\mathcal{A}_N} X$, (X_N) converges almost surely and in $L_1(B)$ to X. If X is in $L_p(B)$, $1 < p < \infty$, the convergence also takes place in $L_p(B)$.

We would like to mention for the sequel (in particular Chapter 8) that, by analogy with the preceding, if X is a Radon random variable with values in B such that for every f in B', $\mathbb{E}f^2(X) < \infty$, the operator $Tf = f(X)$ is also bounded from B' into L_2. Furthermore, it can be shown as before that for every ξ in L_2, ξX is weakly integrable so that $\mathbb{E}(\xi X)$ is well defined as an element of B. In particular, for f, g in B', $g(\mathbb{E}(f(X)X)) = \mathbb{E}(f(X)g(X))$, which defines the so-called "covariance structure" of a random variable X weakly in L_2.

2.2 Random processes and vector valued random variables

The concept of Radon or separable random variable is a convenient concept when dealing with weak convergence or tightness properties. We will use it indeed in the typical weak convergence theorem which is the central limit theorem, and also in some related questions on the law of large numbers and the law of the iterated logarithm. This concept is also a way of taking easily into account various measurability problems.

However, Radon random variables form a somewhat too restrictive setting for other types of questions. For example, if we are given a sequence (X_n) of real random variables such that $\sup_n |X_n| < \infty$ almost surely, and if we ask (for example) for the integrability properties or tail behavior of this supremum, we are clearly faced with a random element of infinite dimension but we need not (and in general do not) have a Radon random vector. In other words, it would be convenient to have a notion of random variable with values in ℓ_∞. The space c_0 of all real sequences tending to 0 is a separable subspace of ℓ_∞ but ℓ_∞ is not separable. Recall further that every separable Banach space can be realized isometrically as a closed subspace of ℓ_∞.

On the other hand, another category of infinite dimensional random elements are the random functions or stochastic processes. Let T be a (infinite) index set which will be usually assumed to be a metric space (T, d). A *random function* or *process* $X = (X_t)_{t \in T}$ indexed by T is a collection of real random variables X_t, $t \in T$. By the distribution or law of X we mean the distribution on $\mathbb{R}^T$, equipped with the *cylindrical* σ*-algebra* generated by the cylinder sets, determined by the collection of all marginal distributions of the finite dimensional random vectors $(X_{t_1}, \ldots, X_{t_N})$, $t_i \in T$.

We often study throughout this book when a given random process is almost surely bounded and/or continuous, and, when this is the case, ask for possible integrability properties or tail behaviors of $\sup_{t \in T} |X_t|$ whenever this makes sense. These considerations of course raise some non-trivial measurability questions as soon as T is no more countable. A priori, a random process $X = (X_t)_{t \in T}$ is almost surely bounded or continuous, or has almost all its trajectories or sample paths bounded or continuous, if, for almost all ω, the path $t \to X_t(\omega)$ is bounded or continuous. However, in order to prove that a random process is almost surely bounded or continuous, and to deal with it, it is preferable and convenient to know that the sets involved in these definitions are properly measurable. It is not the focus of this work to enter these complications, but rather to try to reduce to a simple setting in order not to hide the main ideas of the theory. Let us therefore briefly indicate in this section some possible and classical arguments used to handle these annoying measurability questions. These will then mostly be used without any further comments in the sequel.

Let $X = (X_t)_{t \in T}$ be a random process. When T is not countable, the pointwise supremum $\sup_{t \in T} |X_t(\omega)|$ is usually not well defined since one has to take into account an uncountable family of negligible sets. It is therefore necessary to consider a handy notion of measurable supremum of the collection. One possible way is to understand quantities as $\sup_{t \in T} |X_t|$ (or similar ones, $\sup_{t \in T} X_t$, $\sup_{s,t \in T} |X_s - X_t|$, $\ldots$) as the essential (or lattice) supremum in L_0 of the collection of random variables $|X_t|$, $t \in T$. Even simpler, if the process X is in L_p, $0 < p < \infty$, that is, if $\mathbb{E}|X_t|^p < \infty$ for every t in T, we can simply set

$$\mathbb{E} \sup_{t \in T} |X_t|^p = \sup\{\mathbb{E} \sup_{t \in F} |X_t|^p \, ; \; F \text{ finite in } T\} .$$

This lattice supremum also works in more general Orlicz spaces than L_p-spaces and will mainly be used in Chapters 11 and 12 in order to show a process is bounded, reducing basically the various estimates to the case where T is finite.

Another possibility is the probabilistic concept of separable version which allows to deal similarly with other properties than boundedness, like for example continuity. Let (T, d) be a metric space. A random process $X = (X_t)_{t \in T}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ is said to be separable if there exist a negligible set $N \subset \Omega$ and a countable set S in T such that for every $\omega \notin N$, every $t \in T$ and $\varepsilon > 0$,

$$X_t(\omega) \in \overline{\{X_s(\omega) \, ; \; s \in S , \; d(s,t) < \varepsilon\}}$$

where the closure is taken in $\mathbb{R} \cup \{\infty\}$. If X is separable, in particular, $\sup_{t \in T} |X_t(\omega)| = \sup_{t \in S} |X_t(\omega)|$ for every $\omega \notin N$, and since S is countable there is, of course, no difficulty in dealing with this type of supremum. Note that if there exists a separable random process on (T, d), then (T, d) is separable as a metric space. If (T, d) is separable and X almost surely continuous, then X is separable.

Hence, when a random process is separable, there is no difficulty in dealing with almost sure boundedness or continuity of the trajectories since these properties are reduced along some countable parameter set. In general however, a given random process $X = (X_t)_{t \in T}$ need not be separable. But in a rather general setting, it admits a *version* which is separable. A random process $Y = (Y_t)_{t \in T}$ is said to be a version of X if, for every $t \in T$, $Y_t = X_t$ with probability one; in particular, Y has the same distribution as X. It is known that when (T, d) is separable and when $X = (X_t)_{t \in T}$ is continuous in probability, that is, for every $t_0 \in T$ and every $\varepsilon > 0$,

$$\lim_{t \to t_0} \mathbb{P}\{|X_t - X_{t_0}| > \varepsilon\} = 0 \,,$$

then X admits a separable version. Moreover, every dense sequence S in T can be taken as separable set. The preceding hypothesis will always be satisfied when we will need such a result so that we use it freely below.

Summarizing, the study of almost sure boundedness and continuity of random processes can essentially be reduced through the tools of essential supremum or separable version to the setting of a *countable* index set for which no measurability question occurs. In our first part, we will therefore basically study integrability properties and tail behaviors of supremum of bounded processes indexed by a countable set. The second examines when a given process is almost surely bounded or continuous and we use separable versions.

The purposes of the first part motivate the introduction of a slightly more general notion of random variable with vector values in order to possibly unify results on Radon random variables and on ℓ_∞-valued random variables or bounded processes. One possible definition is the following. Assume we are given a Banach space B (not necessarily separable!) such that there exists a *countable* subset D of the unit ball or sphere of the dual space B' such that

$$\|x\| = \sup_{f \in D} |f(x)| \,, \qquad\qquad x \in B \,.$$

The typical example we have of course in mind is the space ℓ_∞. Recall that separable Banach spaces possess this property. Given B, D like this, we can say that X is random variable with values in B if X is a map from some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into B such that $f(X)$ is measurable for every f in D. We can then work freely with the measurable function $\|X\|$.

This definition includes Radon random variables. It also includes almost surely bounded processes $X = (X_t)_{t \in T}$ indexed on a countable set T; take then simply $B = \ell_\infty(T)$ and $D = T$ identified with the evaluation maps. As a remark, note that when $X = (X_t)_{t\in T}$ is an almost surely continuous process on (T,d) compact, it defines a Radon random variable in the separable Banach space $C(T)$ of continuous functions on T.

When X and B are as before, we simply say that X is a *random variable* (or *vector*) with values in B, as opposed to *Radon* random variable. We will try however to recall each time it will be necessary the exact setting in which we are working, not trying to avoid repetitions in this regard. When we are dealing with a separable Banach space B, we however do not distinguish and simply speak of random variable (or Borel random variable) with values in B.

To conclude this section, let us note that for this generalized notion of random variables with values in a Banach space B we can also speak of the spaces $L_p(B)$, $0 \le p \le \infty$, as the spaces of random variables X such that $\|X\|_p = (\mathbb{E}\|X\|^p)^{1/p} < \infty$ for $0 < p < \infty$, and the corresponding concepts for $p = 0$ or ∞. Almost sure convergence of a sequence (X_n) makes sense similarly, and if we have to deal with the distribution of such a random variable X, we simply mean the one determined by its marginal distributions, i.e. distributions of the finite dimensional random vectors $(f_1(X), \ldots, f_N(X))$ where $f_1, \ldots, f_N \in D$. Again, in case of a Radon random variable, this coincides with the usual definition (choose D weakly dense in the unit ball of B').

Finally in this section, let us mention a trivial but useful observation based on independence and Jensen's inequality. If X is a random variable with values in B in the general sense just described, let us simply say that X has mean zero if $\mathbb{E}f(X) = 0$ for all f in D (we then sometimes write with some abuse that $\mathbb{E}X = 0$). Let then F be a convex function on $\mathbb{R}_+$ and let X and Y be independent random variables in B such that $\mathbb{E}F(\|X\|) < \infty$ and $\mathbb{E}F(\|Y\|) < \infty$. Then, if Y has mean zero,

$$\mathbb{E}F(\|X+Y\|) \ge \mathbb{E}F(\|X\|). \tag{2.5}$$

Indeed, this follows simply by convexity of $F(\|\cdot\|)$ and partial integration with respect to Y using Fubini's theorem.

2.3. Symmetric random variables and Lévy's inequalities

In this paragraph, B denotes a Banach space such that for some countable set D in the unit ball of B', $\|x\| = \sup_{f \in D} |f(x)|$ for all $x \in B$.

A random variable X with values in B is called symmetric if X and $-X$ have the same distribution. Equivalently, X has the same distribution as εX where ε denotes a symmetric Bernoulli or Rademacher random variable taking values ± 1 with probability $1/2$ and independent of X. (Although the name of Bernoulli is historically more appropriate, we will mostly speak of Rademacher variables since this is the most commonly used terminology in the field.) This simple observation is at the basis of *randomization* (or symmetrization, in the probabilistic sense) which is one most powerful tool in Probability in Banach spaces.

Note that for a general random variable X, there is a canonical way of generating a symmetric random variable not too "far" from X: consider indeed $\widetilde{X} = X - X'$ where X' in an independent copy of X, i.e., with the same distribution as X. In these notations, we will usually assume that X and X' are constructed on different probability spaces $(\Omega, \mathcal{A}, \mathbb{P})$ and $(\Omega', \mathcal{A}', \mathbb{P}')$.

We call by Rademacher sequence (or Bernoulli sequence) a sequence $(\varepsilon_i)_{i \in \mathbb{N}}$ of *independent Rademacher random variables* taking thus the values $+1$ and -1 with equal probability. A sequence (X_i) of random variables with values in B is called a *symmetric sequence* if, for every choice of signs ± 1, $(\pm X_i)$ has the same distribution as (X_i) (i.e. for every N, $(\pm X_1, \ldots, \pm X_N)$ has the same law as $(X_1, \ldots, X_N)$ in B^N). Equivalently, (X_i) has the same distribution as $(\varepsilon_i X_i)$ where (ε_i) is a Rademacher sequence independent of (X_i). The typical example of a symmetric sequence consists in a sequence of *independent and symmetric* random variables. In this setting of symmetric sequences, it will be convenient to denote, using Fubini's theorem, by $\mathbb{P}_\varepsilon$, $\mathbb{E}_\varepsilon$ (resp. $\mathbb{P}_X, \mathbb{E}_X$) conditional probability and expectation with respect to the sequence (X_i) (resp. (ε_i)). We hope the slight abuse in notation, ε representing (ε_i) and X, (X_i), will not get confusing in the sequel.

Partial sums of a symmetric sequence of random variables satisfy some very important inequalities known as Lévy's inequalities. They can be stated as follows. Recall they apply to the important case of independent symmetric random variables.

Proposition 2.3. Let (X_i) be a symmetric sequence of random variables with values in B. For every k, set $S_k = \sum_{i=1}^{k} X_i$. Then, for every integer N and $t > 0$,

$$\mathbb{P}\{\max_{k \le N} \|S_k\| > t\} \le 2\mathbb{P}\{\|S_N\| > t\} \tag{2.6}$$

and

$$\mathbb{P}\{\max_{i\leq N}\|X_i\| > t\} \leq 2\mathbb{P}\{\|S_N\| > t\}\,. \tag{2.7}$$

If (S_k) converges in probability to S , the inequalities extend to the limit as

$$\mathbb{P}\{\sup_k \|S_k\| > t\} \leq 2\mathbb{P}\{\|S\| > t\}$$

and similarly for (2.7). As a consequence of Proposition 2.3, note also that by integration by parts, for every $0 < p < \infty$,

$$\mathbb{E}\max_{k\leq N}\|S_k\|^p \leq 2\mathbb{E}\|S_N\|^p$$

and similarly with X_k instead of S_k .

Proof. We only detail (2.6), (2.7) being established exactly in the same way. Let $\tau = \inf\{k \leq N\,;\ \|S_k\| > t\}$. We have

$$\mathbb{P}\{\|S_N\| > t\} = \sum_{k=1}^{N}\mathbb{P}\{\|S_N\| > t\,,\ \tau = k\}\,.$$

Now, since, for every k , $(X_1,\ldots,X_k,-X_{k+1},\ldots,-X_N)$ has the same distribution as $(X_1,\ldots,X_N)$ and $\{\tau = k\}$ only depends on $X_1,\ldots,X_k$, we also have that

$$\mathbb{P}\{\|S_N\| > t\} = \sum_{k=1}^{N}\mathbb{P}\{\|S_k - R_k\| > t\,,\ \tau = k\}$$

where $R_k = S_N - S_k$, $k \leq N$. Using the triangle inequality

$$2\|S_k\| \leq \|S_k + R_k\| + \|S_k - R_k\| = \|S_N\| + \|S_k - R_k\|\,;$$

summing then the two preceding probabilities yields

$$2\mathbb{P}\{\|S_N\| > t\} \geq \sum_{k=1}^{N}\mathbb{P}\{\tau = k\} = \mathbb{P}\{\max_{k\geq N}\|S_k\| > t\}\,.$$

The proof of Proposition 2.3 is complete.

Among the consequences of Lévy's inequalities is a useful result on convergence of series of symmetric sequences. This is known as the Lévy-Itô-Nisio theorem which we present in the context of Radon random variables.

Theorem 2.4. Let (X_i) be a symmetric sequence of Borel random variables with values in a separable Banach space B. Denote, for each n, by μ_n the distribution of the n-th partial sum $S_n = \sum_{i=1}^n X_i$. The following are equivalent:

(i) the sequence (S_n) converges almost surely;

(ii) (S_n) converges in probability;

(iii) (μ_n) converges weakly;

(iv) there exists a probability measure μ in $\mathcal{P}(B)$ such that $\mu_n \circ f^{-1} \to \mu \circ f^{-1}$ weakly for every f in B'.

By a simple symmetrization argument, the equivalences (i)-(iii) can be show to also hold for sums of independent (not necessarily symmetric) random variables. We shall come back to this in Chapter 6. Observe further from the proof that the equivalence between (i) and (ii) is not restricted to the Radon setting.

Proof. (iii) $\Rightarrow$ (ii). We first show that $X_i \to 0$ in probability. By difference, (X_i) is weakly relatively compact. Hence, from every subsequence, one can extract a further one, denote it by i', such that $X_{i'}$ converges weakly to some X. Thus, along every linear functional f, $f(X_{i'}) \to f(X)$ weakly. But now $(f(S_n))$ converges in distribution as a sequence of real random variables so that, for all $\varepsilon > 0$, there is $M > 0$ such that

$$\sup_n \mathbb{P}\{|f(S_n)| > M\} < \varepsilon^2 .$$

Recall now the symmetry assumption and the preceding notations:

$$\sup_n \mathbb{P}_X \mathbb{P}_\varepsilon \{|\sum_{i=1}^n \varepsilon_i f(X_i)| > M\} < \varepsilon^2$$

where (ε_i) is a Rademacher sequence independent of (X_i). For every n, let

$$A = \{\omega\,;\ \mathbb{P}_\varepsilon\{|\sum_{i=1}^n \varepsilon_i f(X_i(\omega))| > M\} < \varepsilon\} .$$

By Fubini's theorem, $\mathbb{P}(A) = \mathbb{P}_X(A) > 1 - \varepsilon$. If $\omega \in A$, we can apply Khintchine's inequalities to the sum $\sum_{i=1}^n \varepsilon_i f(X_i(\omega))$ together with Lemma 4.2 and (4.3) below to see that, if $\varepsilon \leq 1/8$,

$$\sum_{i=1}^n f^2(X_i(\omega)) \leq 8M^2 .$$

It follows that

$$\sup_n \mathbb{P}\{\sum_{i=1}^n f^2(X_i) > 8M^2\} < \varepsilon$$

from which we get that $\sum_i f^2(X_i) < \infty$ almost surely. Thus $f(X_i) \to 0$ almost surely. Hence (X_i) is a tight sequence with only 0 as possible limit point. This shows that $X_i \to 0$ in probability.

We then deduce (ii). Indeed, if this is not the case, (S_n) is not a Cauchy sequence in probability and there exists $\varepsilon > 0$ and a strictly increasing sequence of integers (n_k) such that $T_k = S_{n_{k+1}} - S_{n_k}$ does not converge in probability to 0. Since $\sum_k T_k = \sum_i X_i$ converges weakly, we may apply the preceding step to get a contradiction.

(ii) $\Rightarrow$ (i). If $S_n \to S$ in probability, there exists a sequence (n_k) of integers such that

$$\sum_k \mathbb{P}\{\|S_{n_k} - S\| > 2^{-k}\} < \infty .$$

By Lévy's inequalities,

$$\begin{aligned}
&\mathbb{P}\{\max_{n_{k-1} < n \leq n_k} \|S_n - S_{n_{k-1}}\| > 2^{-k+1}\} \\
&\quad \leq 2\mathbb{P}\{\|S_{n_k} - S_{n_{k-1}}\| > 2^{-k+1}\} \\
&\quad \leq 2(\mathbb{P}\{\|S_{n_k} - S\| > 2^{-k}\} + \mathbb{P}\{\|S_{n_{k-1}} - S\| > 2^{-k-1}\}) .
\end{aligned}$$

By the Borel-Cantelli lemma, (S_n) is almost surely a Cauchy sequence and thus (i) holds.

We are left with the proof of (iv) $\Rightarrow$ (iii) since the other implications are obvious. By Prokhorov's criterion and (2.4) it is enough to show that for every $\varepsilon > 0$ there exists a finite dimensional subspace F of B such that, for every n,

$$\mathbb{P}\{d(S_n, F) > \varepsilon\} < \varepsilon .$$

Since μ is a Radon measure, it of course suffices to show that

$$\mathbb{P}\{d(S_n, F) > \varepsilon\} \leq 2\mu(x\,;\ d(x, F) > \varepsilon)$$

for any $n, \varepsilon > 0$ and F closed subspace in B. Now, since B is separable, for every closed subspace F in B, there is a countable subset $D = \{f_m\}$ of the unit ball of B' such that $d(x, F) = \sup_{f \in D} |f(x)|$ for every

x . For every m , $(f_1(S_n), \ldots, f_m(S_n))$ is weakly convergent in $\mathbb{R}^m$ to the corresponding marginal of μ . Hence, by Lévy's inequalities ((2.6)) in the limit, for every n ,

$$\mathbb{P}\{\max_{i \leq m} |f_i(S_m)| > \varepsilon\} \leq 2\mu(x\,;\ \max_{i \leq m} |f_i(x)| > \varepsilon)\,.$$

The conclusion is then easy:

$$\begin{aligned}
\mathbb{P}\{d(S_n, F) > \varepsilon\} &= \mathbb{P}\{\sup_{f \in D} |f(S_n)| > \varepsilon\} \\
&\leq \sup_m \mathbb{P}\{\max_{i \leq m} |f_i(S_n)| > \varepsilon\} \\
&\leq 2 \sup_m \mu(x\,;\ \max_{i \leq m} |f_i(x)| > \varepsilon) \\
&\leq 2\mu(x\,;\ \sup_{f \in D} |f(x)| > \varepsilon) \\
&\leq 2\mu(x\,;\ d(x, F) > \varepsilon)\,.
\end{aligned}$$

Theorem 2.4 is thus established.

2.4. Some inequalities for real random variables

We conclude this chapter with two elementary and classical inequalities for real random variables which will be useful to record at this stage. The first one is a version of the binomial inequalities (compare with (1.15), (1.16)) while the second is the inequality at the basis of the Borel-Cantelli lemma.

Lemma 2.5. Let (A_i) be a sequence of independent sets such that $a = \sum_i \mathbb{P}(A_i) < \infty$. Then, for every n ,

$$\mathbb{P}\{\sum_i I_{A_i} \geq n\} \leq \frac{a^n}{n!} \leq \left(\frac{ea}{n}\right)^n .$$

Proof. We have

$$\mathbb{P}\{\sum_i I_{A_i} \geq n\} \leq \sum \prod_{j=1}^n \mathbb{P}(A_{i_j})$$

where the summation is over all choices of indexes $i_1 < \cdots < i_n$. Now

$$\begin{aligned}
\sum \prod_{j=1}^n \mathbb{P}(A_{i_j}) &= \frac{1}{n!} \sum_{\substack{\text{distinct} \\ i_1, \ldots, i_n}} \prod_{j=1}^n \mathbb{P}(A_{i_j}) \\
&\leq \frac{1}{n!} \sum_{\substack{\text{all} \\ i_1, \ldots, i_n}} \prod_{j=1}^n \mathbb{P}(A_{i_j}) = \frac{a^n}{n!}\,.
\end{aligned}$$

Lemma 2.6. Let $(Z_i)_{i\leq N}$ be independent positive random variables. Then, for every $t>0$,

$$\mathbb{P}\{\max_{i\leq N} Z_i > t\} \geq \sum_{i=1}^{N} \mathbb{P}\{Z_i > t\}/(1+\sum_{i=1}^{N}\mathbb{P}\{Z_i > t\}) .$$

In particular, if $\mathbb{P}\{\max_{i\leq N} Z_i > t\} \leq \frac{1}{2}$,

$$\sum_{i=1}^{N} \mathbb{P}\{Z_i > t\} \leq 2\mathbb{P}\max_{i\leq N} Z_i > t\} .$$

Proof. For $x\geq 0$, $1-x\leq \exp(-x)$ and $1-\exp(-x)\geq x/1+x$. Thus, by independence,

$$\begin{aligned}\mathbb{P}\{\max_{i\leq N} Z_i > t\} &= 1-\prod_{i=1}^{N}(1-\mathbb{P}\{Z_i > t\} \\ &\geq 1-\exp(-\sum_{i=1}^{N}\mathbb{P}\{Z_i > t\}) \\ &\geq \sum_{i=1}^{N}\mathbb{P}\{Z_i > t\}/1+\sum_{i=1}^{N}\mathbb{P}\{Z_i > t\} .\end{aligned}$$

If $\mathbb{P}\{\max_{i\leq N} Z_i > t\} \leq \frac{1}{2}$, the preceding inequality ensures that $\sum_{i=1}^{N}\mathbb{P}\{Z_i > t\} \leq 1$ so that the second conclusion of the lemma follows from the first one.

Notes and References

The following references will hopefully complete appropriately this survey chapter.

Basics on metric spaces, infinite dimensional vector spaces, Banach spaces, etc. can be found in all classical treatrises on functional analysis like for example [Dun-S]. Informations on Banach spaces more in the spirit of this book are given in [Da], [Bea], [Li-T1], [Li-T2] as well as in the references therein.

Probability distributions on metric spaces and weak convergence are presented in [Par], [Bi]. Various accounts on random variables with values in Banach spaces may be found in [Ka1], [Schw2], [HJ3], [Ar-G2], [Li], [V-T-C]. Prokhorov's criterion comes from [Pro1]; the terminology "flat concentration" is used in [Ac1]. For Skorokhod's theorem [Sk1], see also [Du3]. The necessary elements on vector valued martingales and their convergence may be found in [Ne3].

Generalities on random processes and separability are given in [Doo], [Me], [Ne1]. More in the context of probability in Banach spaces, see also [Ba], [J-M3], [Ar-G2].

Symmetric sequences and randomization techniques were first considered by J.-P. Kahane in [Ka1] who gave a proof of Lévy's inequalities [Lé1] in this setting of vector valued random variables. See also [HJ1], [HJ2], [HJ3]. Theorem 2.4 is due to Lévy [Lé1] on the line and to Itô-Nisio [I-N] for independent Banach space valued random variables. For symmetric sequences, see J. Hoffman-Jørgensen [HJ3]. Our proof follows [Ar-G2].

The inequalities of Section 4 can be found in all classical treatrises on probability theory, like for example [Fe1].

Chapter 3. Gaussian random variables

Chapter 3. Gaussian random variables

With this chapter we really enter the subject of Probability in Banach spaces. The study of Gaussian random vectors and processes may indeed be considered as one of the fundamental topics of the theory since it inspires many other parts of the field both in the results themselves and in the techniques of investigation. The historical developments also followed this line of progress.

We shall be interested in this chapter in integrability properties and tail behavior of norms of Gaussian random vectors or bounded processes as well as in the basic comparison properties of Gaussian processes. The question of when a Gaussian process is almost surely bounded (or continuous) will be addressed and completely solved in Chapters 11 and 12. The study of the tail behavior of the norm of a Gaussian random vector is based on the isoperimetric tools introduced in the first chapter. This study will be a kind of reference for the corresponding results for other types of random vectors like Rademacher series, stable random variables, and even sums of independent random variables which will be treated respectively in the next chapters. This will be the subject of the first section of this chapter. The second examines the corresponding results for chaos. The last paragraph is devoted to the important comparison properties of Gaussian random variables. These now appear at the basis of the rather deep present knowledge on the regularity properties of sample paths of Gaussian processes (cf. Chapter 12).

We first recall the basic definitions and some classical properties of Gaussian variables. A real mean zero random variable X in $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is said to be Gaussian (or normal) if its Fourier transform satisfies

$$\mathbb{E} \exp itX = \exp(-\sigma^2 t^2/2)\,, \qquad t \in \mathbb{R}\,,$$

where $\sigma = \|X\|_2 = (\mathbb{E}X^2)^{1/2}$. When we speak of Gaussian variable we therefore always mean a *centered* Gaussian (or equivalently *symmetric*) variable. X is said to be standard if $\sigma = 1$. Throughout this work, the sequence denoted by $(g_i)_{i \in \mathbb{N}}$ will always mean a sequence of *independent standard Gaussian random variables*; we sometimes call it *orthogaussian* or *orthonormal sequence* (because orthogonality and independence are equivalent for Gaussian variables). In other words, for each N, the vector $g = (g_1, \ldots, g_N)$ follows the canonical Gaussian distribution γ_N on $\mathbb{R}^N$ with density

$$\gamma_N(dx) = (2\pi)^{-N/2} \exp(-|x|^2/2) dx\,.$$

A random vector $X = (X_1, \ldots, X_N)$ in $\mathbb{R}^N$ is Gaussian if for all real numbers $\alpha_1, \ldots, \alpha_N$, $\sum_{i=1}^{N} \alpha_i X_i$ is a real Gaussian variable. Such a Gaussian vector can always be diagonalized and regarded under the canonical distribution γ_N . Indeed, if $\Gamma = AA^t = (\mathbb{E}X_iX_j)_{1\le i,j\le N}$ denotes the symmetric (semi-) positive definite covariance matrix, Γ completely determines the distribution of X which is the same as the one of Ag where $g = (g_1, \dots, g_N)$.

One fundamental property of Gaussian distributions is their rotational invariance which may be expressed in various manners. For example, if $g = (g_1, \dots, g_N)$ is distributed according to γ_N in $\mathbb{R}^N$ and if U is an orthogonal matrix in $\mathbb{R}^N$, then Ug is also distributed according to γ_N . As a simple consequence, if (α_i) is a finite sequence of real numbers, $\sum_i g_i\alpha_i$ has the same distribution as $g_1\left(\sum_i \alpha_i^2\right)^{1/2}$. In particular, since g_1 has moments of all orders,

$$\left\|\sum_i g_i\alpha_i\right\|_p = \|g_1\|_p \left(\sum_i \alpha_i^2\right)^{1/2}$$

for any $0 < p < \infty$ so that the span of (g_i) in L_p is isometric to ℓ_2 . As another description of this invariance by rotation (which was used in the proof of (1.5) in Chapter 1), if X is Gaussian (in $\mathbb{R}^N$) and if Y is an independent copy of X , for every θ , the rotation of angle θ of the vector (X,Y) , i.e.

$$(X\sin\theta + Y\cos\theta\,,\ X\cos\theta - Y\sin\theta)\,,$$

has the same distribution as (X,Y) . These properties are trivially verified on the characteristic functionals.

A Radon random variable X with values in a Banach space B is Gaussian if for any continuous linear functional f on B , $f(X)$ is a real Gaussian variable. The typical example is given by a convergent series $\sum_i g_ix_i$ (which converges in any sense by Theorem 2.4) where (x_i) is a sequence in B . We shall actually see later on in this chapter that every Radon Gaussian vector may be represented in distribution in this way. Finally, a process $X = (X_t)_{t\in T}$ indexed by a set T is called Gaussian if each finite linear combination $\sum_i \alpha_iX_{t_i}$, $\alpha_i \in \mathbb{R}$, $t_i \in T$, is Gaussian. The covariance structure $\Gamma(s,t) = \mathbb{E}X_sX_t$, $s,t \in T$, completely determines the distribution of the Gaussian process X . Since the distributions of these infinite dimensional Gaussian variables are determined by the finite dimensional projections, the rotational invariance trivially extends. For example, if X is a gaussian process or variable (in B), and if (X_i) is a sequence of independent copies of X , then, for any finite sequence (α_i) of real numbers, $\sum_i \alpha_iX_i$ has the same distribution as $\left(\sum_i \alpha_i^2\right)^{1/2} X$.

3.1. Integrability and tail behavior

By their very definition, (norms of) real or finite dimensional Gaussian variables admit high integrability properties. One of the purposes of this section will be to show how these extend to an infinite dimensional setting. We thus investigate integrability and tail behavior of the norm $\|X\|$ of a Gaussian Radon variable or almost surely bounded process. In order to state the results in a somewhat unified way, it is convenient to adopt the terminology introduced in Section 2.2. Unless otherwise specified, we thus deal in this section with a Banach space B such that, for some countable subset D of the unit ball of B', $\|x\| = \sup_{f\in D} |f(x)|$ for all x in B. We say that X is a Gaussian random variable in B if $f(X)$ is measurable for every f in D and if every finite linear combination $\sum_i \alpha_i f_i(X)$, $\alpha_i \in \mathbb{R}$, $f_i \in D$, is Gaussian.

Let therefore X be Gaussian in B. *Two* main parameters of the distribution of X will determine the behavior of $\mathbb{P}\{\|X\| > t\}$ (when $t \to \infty$): some quantity in the (*strong*) topology of the norm, median or exceptation of $\|X\|$ for example (or some other quantity in $L_p(B)$, $0 \le p < \infty$ - see below), and *weak* variances, $\mathbb{E}f^2(X)$, $f \in D$. More precisely, let $M = M(X)$ be a median of $\|X\|$, that is a number satisfying both

$$\mathbb{P}\{\|X\| \le M\} \ge \frac{1}{2}, \qquad \mathbb{P}\{\|X\| \ge M\} \ge \frac{1}{2}.$$

Actually, for the purposes of tail behavior, integrability properties or moment equivalences, it would be enough to consider M such that $\mathbb{P}\{\|X\| \le M\} > \varepsilon > 0$ as we will see later; the concept of median is however crucial for concentration results as opposed to the preceding ones which come more under deviation inequalities. Besides M, consider

$$\sigma = \sigma(X) = \sup_{f\in D} (\mathbb{E}f^2(X))^{1/2}.$$

Note that this supremum is finite and actually controlled by M. Indeed, for every f in D, $\mathbb{P}\{|f(X)| \le M\} \ge \frac{1}{2}$; now $f(X)$ is a real Gaussian variable with variance $\mathbb{E}f^2(X)$ and this inequality implies that $(\mathbb{E}f^2(X))^{1/2} \le 2M$ since $\mathbb{P}\{|g| \le \frac{1}{2}\} < 0.4 < \frac{1}{2}$ where g is a standard normal variable. Hence $\sigma \le 2M < \infty$. Let us mention that if X is a Gaussian Radon variable with values in B, σ is really meant to be $\sup_{\|f\|\le 1} (\mathbb{E}f^2(X))^{1/2}$. This is always understood in the sequel.

The main conclusions on the behavior of $\mathbb{P}\{\|X\| > t\}$ are obtained from the Gaussian isoperimetric inequality (Theorem 1.2) and subsequent concentration results described in Section 1.1. In order to conveniently use the isoperimetric inequality, we reduce, as is usual, the distribution of X to the canonical

Gaussian distribution $\gamma = \gamma_\infty$ on $\mathbb{R}^{\mathbb{N}}$ (i.e., γ is the distribution of the orthonormal sequence (g_i)). The procedure is classical. Set $D = \{f_n\,,\ n \geq 1\}$. By the Gram-Schmidt orthonormalization procedure applied to the sequence $(f_n(X))_{n\geq 1}$ in L_2 , we can write (in distritution)

$$f_n(X) = \sum_{i=1}^{n} a_i^n g_i\,, \qquad n \geq 1\,.$$

In other words, if $x = (x_i)$ is a generic point in $\mathbb{R}^{\mathbb{N}}$, the sequence $(f_n(X))$ has the same distribution as the sequence $\left(\sum_{i=1}^{n} a_i^n x_i\right)$ under γ . For notational convenience, we then extend the meaning of f_n by letting $f_n(x) = \sum_{i=1}^{n} a_i^n x_i$ for $n \geq 1$ and $x = (x_i)$ in $\mathbb{R}^{\mathbb{N}}$. If we set further, for $x \in \mathbb{R}^{\mathbb{N}}$, $\|x\| = \sup_{n\geq 1} |f_n(x)|$, the probabilistic study of $\|X\|$ then amounts to the study of $\|x\|$ under γ . Note that in these notations

$$\sigma = \sigma(X) = \sup_{|h|\leq 1} \|h\| \tag{3.1}$$

where, as usual, $|\cdot|$ is the ℓ_2 norm. We use this simple reduction throughout the various proofs below.

As announced, the next lemma describes the isoperimetric concentration of $\|X\|$ around its median M measured in terms of σ . Recall Φ , the distribution function of a standard normal variable, Φ^{-1} , $\Psi = 1 - \Phi$ and the estimate $\Psi(t) \leq \frac{1}{2}\exp(-t^2/2)$, $t \geq 0$.

Lemma 3.1. Let X be a Gaussian variable in B with median $M = M(X)$ and supremum of weak variances $\sigma = \sigma(X)$. Then, for every $t > 0$,

$$\mathbb{P}\{|\|X\| - M| > t\} \leq 2\Psi(t/\sigma) \leq \exp(-t^2/2\sigma^2)\,.$$

Proof. We use Theorem 1.2 through the preceding reduction. Let $A = \{x \in \mathbb{R}^{\mathbb{N}}\,;\ \|x\| \leq M\}$. Then $\gamma(A) \geq \frac{1}{2}$. By Theorem 1.2, or rather (1.3) here, if A_t is the Hilbertian neighborhood of order $t > 0$ of A , $\gamma_*(A_t) \geq \Phi(t)$. But, if $x \in A_t$ $x = a + th$ where $a \in A$ and $|h| \leq 1$; hence by (3.1)

$$\|x\| \leq M + t\|h\| \leq M + t\sigma$$

and therefore $A_t \subset \{x\,;\ \|x\| \leq M + \sigma t\}$. Applying the same argument to $A = \{x\,;\ \|x\| \geq M\}$ clearly concludes the proof of Lemma 3.1.

The proof of Lemma 3.1 of course just repeats the concentration property (1.4) for Lipschitz functions as is clear from the fact that $\|x\|$ (on $\mathbb{R}^{\mathbb{N}}$) is Lipschitzian with constant σ . This observation tells us that, following (1.5), we have similarly (and at some cheaper price) a concentration of $\|X\|$ around its expectation, that is, for any $t > 0$,

$$\mathbb{P}\{|\|X\| - \mathbb{E}\|X\|| > t\} \le 2\exp(-2t^2/\pi^2\sigma^2)\,. \tag{3.2}$$

It is clear that $\mathbb{E}\|X\| < \infty$ from Lemma 3.1; this can actually also be deduced from a finite dimensional version of (3.2) together with an approximation argument. ((1.6) of course yields (3.2) with best constant $-t^2/2\sigma^2$ in the exponential.) As usual, (3.2) is interesting since it is often easier to work with expectations rather than with medians. Repeating in this framework some of the comments of Section 1.1, note that a median $M = M(X)$ of $\|X\|$ is unique and that, integrating for example the inequality of Lemma 3.1,

$$|\mathbb{E}\|X\| - M| \le \sigma(\pi/2)^{1/2}\,.$$

As already mentioned, the integrability theorems we will deduce from Lemma 3.1 (or (3.2)) actually only use half of it, that is only the deviation inequality

$$\mathbb{P}\{\|X\| > M + \sigma t\} \le \Psi(t)\,, \qquad t > 0\,,$$

(and similarly with $\mathbb{E}\|X\|$ instead of M). The concentration around M (or $\mathbb{E}\|X\|$) will however be crucial for some other questions like for example in Chapter 9 where this result and the relative weights of M and σ can be used in Geometry of Banach spaces. Actually, for the integrability theorems even only the knowledge of s such that $\mathbb{P}\{\|X\| \le s\} \ge \varepsilon > 0$ is sufficient. Indeed, if, in the proof of Lemma 3.1, we apply the isoperimetric inequality to $A = \{x\,;\ \|x\| \le s\}$, we get that, for every $t > 0$,

$$\mathbb{P}\{\|X\| > s + \sigma t\} \le \Psi(\Phi^{-1}(\varepsilon) + t)\,.$$

It follows for example that, for $t > 0$,

$$\mathbb{P}\{\|X\| > s + \sigma t\} \le \exp(\Phi^{-1}(\varepsilon)^2/2)\exp(-t^2/8)\,.$$

As we have seen, the information σ on weak moments is weaker than the one, M or $\mathbb{E}\|X\|$ for example, on the strong topology. We already noted that $\sigma \le 2M$ and we can add the trivial one $\sigma \le (\mathbb{E}\|X\|^2)^{1/2}$ (which is finite). In general, σ is much smaller. For the canonical distribution γ_N on $\mathbb{R}^N$ we already have that $\sigma = 1$ and M is of the order of $\sqrt{N}$. In the preceding inequality, σ can be replaced by one of the strong parameters yielding weaker but still interesting bounds. For example, in the context of the preceding inequality, we observe that $\sigma \leq s/\Phi^{-1}\left(\frac{1+\varepsilon}{2}\right)$ from which it follows that, for $t > 0$,

$$\mathbb{P}\{\|X\| > t\} \leq \exp\left(\frac{1}{2}\Phi^{-1}(\varepsilon)^2 + \frac{1}{8}\Phi^{-1}\left(\frac{1+\varepsilon}{2}\right)\right) \exp\left(-\frac{t^2}{32s^2}\Phi^{-1}\left(\frac{1+\varepsilon}{2}\right)^2\right). \tag{3.3}$$

This inequality seems complicated but however describes as before an exponential squared tail for $\mathbb{P}\{\|X\| > t\}$. Note further that $\Phi^{-1}\left(\frac{1+\varepsilon}{2}\right)$ becomes large when ε goes to 1. While we deduce (3.3) from isoperimetric methods, it should be noticed that such an inequality can actually be established by a direct argument which we would like to briefly sketch. Let Y denote an independent copy of X. By the rotational invariance of Gaussian measures, $(X+Y)/\sqrt{2}$ and $(X-Y)/\sqrt{2}$ are independent with the same distribution as X. Now, if for $s < t$, $\|X+Y\| \leq s\sqrt{2}$ and $\|X-Y\| > t\sqrt{2}$, we have from the triangle inequality that both $\|X\|$ and $\|Y\|$ are larger than $(t-s)/\sqrt{2}$. Hence, by independence and identical distribution,

$$\begin{aligned}
\mathbb{P}\{\|X\| \leq s\}\mathbb{P}\{\|X\| > t\} &= \mathbb{P}\{\|X+Y\| \leq s\sqrt{2},\ \|X-Y\| > t\sqrt{2}\} \\
&\leq \mathbb{P}\{\|X\| > (t-s)/\sqrt{2},\ \|Y\| > (t-s)/\sqrt{2}\} \\
&\leq \left(\mathbb{P}\{\|X\| > (t-s)/\sqrt{2}\}\right)^2.
\end{aligned}$$

Iterating this inequality with $\mathbb{P}\{\|X\| \leq s\} = \varepsilon > \frac{1}{2}$ and

$$t = t_n = (\sqrt{2^{n+1}} - 1)(\sqrt{2} + 1)s$$

easily yields that, for each $t \geq s$,

$$\mathbb{P}\{\|X\| > t\} \leq \exp\left(-\frac{t^2}{24s^2}\log\frac{\varepsilon}{1-\varepsilon}\right), \tag{3.4}$$

an inequality indeed similar in nature to (3.3).

Let us record at this stage an inequality of the preceding type in which only the strong parameter steps in and that will be convenient in the sequel. From Lemma 3.1 (for example!) and $\sigma^2 \leq \mathbb{E}\|X\|^2$, $M^2 \leq 2\mathbb{E}\|X\|^2$, we have that, for every $t > 0$,

$$\mathbb{P}\{\|X\| > t\} \leq 4\exp(-t^2/8\mathbb{E}\|X\|^2). \tag{3.5}$$

To conclude these comments, let us also mention a bound for the maximum of norms of a finite number of vector valued Gaussian variables X_i , $i \le N$. Assume first that $\max_{i\le N} \sigma(X_i) \le 1$. For any $\delta \ge 0$, we have by integration by parts and Lemma 3.1 that

$$\begin{aligned}\mathbb{E}\max_{i\le N} |\|X_i\| - M(X_i)| &\le \delta + \sum_{i=1}^{N} \int_\delta^\infty \mathbb{P}\{|\|X_i\| - M(X_i)| > t\}dt \\ &\le \delta + N\int_\delta^\infty \exp(-t^2/2)dt \\ &\le \delta + N\sqrt{\frac{\pi}{2}}\exp(-\delta^2/2)\,.\end{aligned}$$

Let then simply $\delta = (2\log N)^{1/2}$ so that we have obtained by homogeneity that

$$\mathbb{E}\max_{i\le N}\|X_i\| \le 2\max_{i\le N}\mathbb{E}\|X_i\| + 3(\log N)^{1/2}\max_{i\le N}\sigma(X_i)\,. \tag{3.6}$$

The next corollary describes applications of Lemma 3.1 and the preceding inequalities to the tail behavior and integrability properties of the norm of a Gaussian random vector.

Corollary 3.2. Let X be a Gaussian variable in B with corresponding $\sigma = \sigma(X)$. Then

$$\lim_{t\to\infty}\frac{1}{t^2}\log\mathbb{P}\{\|X\| > t\} = -\frac{1}{2\sigma^2}$$

or, equivalently,

$$\mathbb{E}\exp\left(\frac{1}{2\alpha^2}\|X\|^2\right) < \infty \quad \text{if and only if } \alpha > \sigma\,.$$

Further, all the moments of $\|X\|$ are equivalent (and equivalent to $M = M(X)$, the median of $\|X\|$): for any $0 < p, q < \infty$, there exists $K_{p,q}$ depending on p and q only such that for any Gaussian vector X

$$\|X\|_p \le K_{p,q}\|X\|_q\,.$$

In particular, $K_{p,2} = K\sqrt{p}$ ($p \ge 2$) where K is numerical.

Proof. The fact that the limit is less than or equal to $-1/2\sigma^2$ easily follows from Lemma 3.1 while the minoration simply uses $\mathbb{P}\{\|X\| > t\} \ge \mathbb{P}\{|f(X)| > t\}$ for all f in D . The equivalence with the exponential integrability is easy by Chebyshev's inequality and integration by parts. Concerning the moment equivalences, if M is the median of $\|X\|$, integrating the inequality of Lemma 3.1,

$$\begin{aligned}\mathbb{E}|\|X\| - M|^p &= \int_0^\infty \mathbb{P}\{|\|X\| - M| > t\} dt^p \\ &\leq \int_0^\infty \exp(-t^2/2\sigma^2) dt^p \leq (K\sqrt{p}\sigma)^p\end{aligned}$$

for some numerical K. Now this inequality is stronger than what we need since $\sigma \leq 2M$ and M can be majorized by $(2\mathbb{E}\|X\|^q)^{1/q}$ for all $q > 0$. The proof is complete.

As already mentioned, we use Lemma 3.1 in this proof but the inequalites discussed prior to Corollary 3.2 can be used (to some extent) similarly.

Let (X_n) be a sequence of Gaussian variables which is bounded in probability, that is, for each $\varepsilon > 0$ there exists $A > 0$ such that

$$\sup_n \mathbb{P}\{\|X_n\| > A\} < \varepsilon .$$

Then (X_n) is bounded in all the L_p-spaces. Indeed, if $M(X_n)$ is the median of $\|X_n\|$, certainly $\sup_n M(X_n) < \infty$ and the preceding equivalences of moments confirm the claim. In particular, if (X_n) is a Gaussian sequence (in the sense that $(X_{n_1}, \ldots, X_{n_N})$ is Gaussian in B^N for all $n_1, \ldots, n_N$) which converges in probability, this convergence takes place in all L_p.

Although already very precise, the previous corollary can be refined. The sharpening we describe next rests on a more elaborated use of the Gaussian isoperimetric inequality and confirm the role of the two parameters, weak and strong, used to measure the size of the distribution of the norm of a Gaussian vector. Let X be as before Gaussian in B and recall $\sigma = \sigma(X) = \sup_{f \in D} (\mathbb{E} f^2(X))^{1/2}$. Consider now

$$\tau = \tau(X) = \inf\{\lambda \geq 0 ; \ \mathbb{P}\{\|X\| \leq \lambda\} > 0\} ,$$

that is the first jump of the distribution of the norm $\|X\|$. This jump can actually be shown to be unique. In case X is Radon, $\tau = 0$. One way to prove this is to first observe that for every $\varepsilon > 0$ and x in B

$$(\mathbb{P}\{\|X - x\| \leq \varepsilon\})^2 \leq \mathbb{P}\{\|X\| \leq \varepsilon\sqrt{2}\} .$$

Indeed, if Y is an independent copy of X, by symmetry and independence,

$$\begin{aligned}(\mathbb{P}\{\|X-x\|\le\varepsilon\})^2 &= \mathbb{P}\{\|X-x\|\le\varepsilon\}\mathbb{P}\{\|Y+x\|\le\varepsilon\}\\ &\le \mathbb{P}\{\|(X-x)+(Y+x)\|\le 2\varepsilon\}\\ &\le \mathbb{P}\{\|X\|\le\varepsilon\sqrt{2}\}\end{aligned}$$

since $X+Y$ has the law of $\sqrt{2}X$. Note that this inequality can actually be improved to

$$\mathbb{P}\{\|X-x\|\le\varepsilon\}\le\mathbb{P}\{\|X\|\le\varepsilon\}$$

by the Φ^{-1} - or $\log$-concavity ((1.1) and (1.2)) but is sufficient for our modest purpose here. If X is Radon, and if we assume that $P\{\|X\|\le\varepsilon_0\}=0$ for some $\varepsilon_0>0$, there is a sequence (x_n) in B such that, by separability,

$$\mathbb{P}\{\exists n; \|X-x_n\|\le\varepsilon_0/\sqrt{2}\}=1\,.$$

But then

$$1\le\sum_n\mathbb{P}\{\|X-x_n\|\le\varepsilon_0/\sqrt{2}\}\le\sum_n(\mathbb{P}\{\|X\|\le\varepsilon_0\})^{1/2}=0$$

and thus, necessarily $\tau=\tau(X)=0$ in this case.

On the other hand, let us recall the typical example in which $\tau>0$. Consider X in ℓ_∞ given by the sequence of its coordinates $(g_n/(2\log(n+1))^{1/2})$ where (g_n) is the orthogaussian sequence. Then, by a simple use of the Borel-Cantelli lemma, it is easily seen that

$$\limsup_{n\to\infty}\frac{|g_n|}{(2\log(n+1))^{1/2}}=1\qquad\text{almost surely}\tag{3.7}$$

so that $\tau=1$ in this case.

The following theorem refines as announced Corollary 3.2 and involves both σ and τ in the integrability result.

Theorem 3.3. Let X be a Gaussian variable with values in B and let $\sigma=\sigma(X)$ and $\tau=\tau(X)$. Then, for any $\tau'>\tau$,

$$\mathbb{E}\exp\left(\frac{1}{2\sigma^2}(\|X\|-\tau')^2\right)<\infty\,.$$

Before proving this theorem, let us interpret this result as a tail behavior. We have seen in Corollary 3.2 that

$$\lim_{t\to\infty}\frac{1}{t^2}\log \mathbb{P}\{\|X\|>t\}=-\frac{1}{2\sigma^2}$$

which can be rewritten as

$$\lim_{t\to\infty}\frac{1}{t}\Phi^{-1}(\mathbb{P}\{\|X\|\le t\})=\frac{1}{\sigma}$$

($\Phi^{-1}(1-u)$ is equivalent to $(2\log\frac{1}{u})^{1/2}$ when $u\to 0$). From (1.1) we know that the function $\Phi^{-1}(\mathbb{P}\{\|X\|\le t\})$ is concave on $]\tau,\infty[$. Theorem 3.3 can therefore be described equivalently as

$$0\ge \lim_{t\to\infty}[\Phi^{-1}(\mathbb{P}\{\|X\|\le t\})-\sigma t]\ge -\frac{\tau}{\sigma}\,.$$

That is, the concave function $\Phi^{-1}(\mathbb{P}\{\|X\|\le t\})$ on $]\tau,\infty[$ has an asymptote $(t/\sigma)+\ell$ with $-\tau/\sigma\le \ell\le 0$. Note that $\tau=0$ and hence $\ell=0$ if X is Radon.

The proof of Theorem 3.3 appears as a consequence of the following (deviation) lemma of independent interest which may be compared to Lemma 3.1.

Lemma 3.4. For every $\tau'>\tau$, there is an integer N such that for every $t\ge 0$,

$$\mathbb{P}\{\|X\|>\tau'+\sigma t\}\le \mathbb{P}\left\{\left(\sum_{i=1}^N g_i^2\right)^{1/2}\right\}.$$

In particular

$$\mathbb{P}\{\|X\|>\tau'+\sigma t\}\le K(N)(1+t)^N\exp(-t^2/2)$$

where $K(N)$ is a positive constant depending on N only.

To see why the theorem follows from the lemma, let $\tau'>\tau''>\tau$ and $\varepsilon=(\tau'-\tau'')/\sigma>0$. Applying Lemma 3.4 to $\tau''>\tau$, we get that, for some integer N ,

$$\begin{aligned}\int_{\|X\|>\tau'}\exp\left(\frac{1}{2\sigma^2}(\|X\|-\tau')^2\right)d\mathbb{P}&\le 1+\int_0^\infty \mathbb{P}\{\|X\|>\tau'+\sigma t\}t\exp\left(\frac{t^2}{2}\right)dt\\&\le 1+K(N)\int_0^\infty(1+\varepsilon+t)^{N+1}\exp\left(\frac{t^2}{2}-\frac{(t+\varepsilon)^2}{2}\right)dt\\&\le 1+K(N)\int_0^\infty(1+\varepsilon+t)^{N+1}\exp(-\varepsilon t)dt<\infty\,.\end{aligned}$$

Proof of Lemma 3.4. It is based again on the Gaussian isoperimetric inequality in the form of the following lemma which we introduce with some notations. Given an integer N, a point x in $\mathbb{R}^{\mathbb{N}}$ can be decomposed in (y,z) with $y \in \mathbb{R}^N$ and $z \in \mathbb{R}^{]N,\infty[}$. By Fubini's theorem, $d\gamma(x) = d\gamma_N(y) d\gamma_{]N,\infty[}(z)$. If A is a Borel set in $\mathbb{R}^{\mathbb{N}}$, set $B = \{z \in \mathbb{R}^{]N,\infty[}\,;\ (0,z) \in A\}$. Recall for $t \geq 0$, A_t denotes the Euclidean or Hilbertian neighborhood of A.

Lemma 3.5. Under these notations, if $\gamma_{]N,\infty[}(B) \geq 1/2$, for any $t \geq 0$,

$$\gamma^*(x \in \mathbb{R}^{\mathbb{N}}\,;\ x \notin A_t) \leq \gamma_{N+1}(y' \in \mathbb{R}^{N+1}\,;\ |y'| > t)\,.$$

Proof. Since $\gamma_{]N,\infty[}(B) \geq 1/2$, Theorem 1.2 implies that $(\gamma_{]N,\infty[})_*(B_s) \geq \Phi(s)$ where B_s is of course the Hilbertian neighborhood of order $s \geq 0$ of B in $\mathbb{R}^{]N,\infty[}$. Let $y \in \mathbb{R}^N$, $t \geq |y|$ and $s = (t^2 - |y|^2)^{1/2}$. If $z \in B_s$, by definition, $z = b + sk$ where $b \in B$, $k \in \mathbb{R}^{]N,\infty[}$, $|k| \leq 1$; then

$$x = (y,z) = (0,b) + (|y|^2 + s^2)^{1/2} h = (0,b) + th$$

where $h \in \mathbb{R}^{\mathbb{N}}$ and $|h| \leq 1$. Hence $x \in A_t$. From this observation, we get that, via Fubini's theorem, for $t \geq 0$,

$$\begin{aligned}
\gamma^*(x : x \notin A_t) &\leq \gamma_N(y : |y| > t) \\
&\qquad + \gamma_N \otimes (\gamma_{]N,\infty[})^*((y,z) : |y| \leq t\,,\ z \notin B_{(t^2-|y|^2)^{1/2}}) \\
&\leq \gamma_N(y : |y| > t) \\
&\qquad + \int_{|y| \leq t} \int_{s > (t^2-|y|^2)^{1/2}} \exp\left(-\frac{s^2}{2}\right) \frac{ds}{\sqrt{2\pi}} d\gamma_N(y) \\
&\leq \int_{s^2+|y|^2>t^2} \exp\left(-\frac{s^2}{2}\right) \frac{ds}{\sqrt{2\pi}} d\gamma_N(y)
\end{aligned}$$

which is the announced result. Lemma 3.5 is thus established.

We are now in a position to prove Lemma 3.4. We of course use the reduction to $(\mathbb{R}^{\mathbb{N}}, \gamma)$ as in Lemma 3.3. Let $\tau' > \tau$ and $A = \{x; \|x\| \leq \tau'\}$. We first show that there exists an integer N such that in the notations introducing Lemma 3.5, if $B = \{z \in \mathbb{R}^{]N,\infty[}\,;\ (0,z) \in A\}$, then $\gamma_{]N,\infty[}(B) \geq 1/2$. By hypothesis $\gamma(A) > 0$. Since $\|x\| = \sup_{n \geq 1} |f_n(x)|$ and $f_n(x)$ only depends on the first n coordinates of x, there exists N such that

$$\gamma_N(y; \max_{n \leq N} |f_n(y)| \leq \tau') \leq \frac{4}{3}\gamma(A)\,.$$

By Fubini's theorem, there exists then y in $\mathbb{R}^N$ such that

$$\gamma_{]N,\infty[}(z\,;\ (y,z) \in A) \geq 3/4\,.$$

By symmetry,

$$\gamma_{]N,\infty[}(z\,;\ (y,-z) \in A) \geq 3/4$$

so that the intersection of the two preceding sets has a measure bigger than $1/2$. But if z belongs to this intersection, $\|(y,z)\| \leq \tau'$ and $\|(y,-z)\| = \|(-y,z)\| \leq \tau'$ and therefore $\|(0,z)\| \leq \tau'$. This shows that $\gamma_{]N,\infty[}(B) \geq 1/2$. From Lemma 3.5 we then get an upper bound for the complement of A_t, $t > 0$. Since

$$A_t \subset \{x\,;\ \|x\| \leq \tau' + t\sigma\}$$

the first inequality in Lemma 3.4 follows. The second one is an easy and classical consequence of the first one. The proof of Lemma 3.4 is complete.

To conclude this paragraph we briefly describe the series representation of Gaussian Radon Banach space valued random variables which is easily deduced from the integrability properties. Recall (g_i) denote an orthogaussian sequence.

Proposition 3.6. Let X be a Gaussian Radon random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a Banach space B. Then X has the same distribution as $\sum_i g_i x_i$ for some sequence (x_i) in B where this series is convergent almost surely and in all L_p's.

Proof. Let H be the closure in $L_2 = L_2(\Omega, \mathcal{A}, \mathbb{P})$ of the variables of the form $f(X)$ with f in B'. Then H may be assumed to be separable since X is Radon and is entirely formed of Gaussian variables. Let (g_i) denote an orthonormal basis of H and denote by $\mathcal{A}_N$ the σ-algebra generated by $g_1, \ldots, g_N$. It is easily seen that

$$\mathbb{E}^{\mathcal{A}_N} X = \sum_{i=1}^{N} g_i x_i$$

with $x_i = \mathbb{E}(g_i X)$. Recall now that $\mathbb{E}\|X\| < \infty$ (for example) so that by the martingale convergence theorem (cf. Section 2.1) the series $\sum_i g_i x_i$ is almost surely convergent to X. Since $\mathbb{E}\|X\|^p < \infty$, $0 < p < \infty$, it converges also in $L_p(B)$ for every p.

3.2. Integrability of Gaussian chaos

In this section, integrability and tail behavior of real and vector valued Gaussian chaos are investigated as a natural continuation of the preceding. The Gaussian variables studied so far may indeed be considered as chaos of order 1. But let us first briefly (and partially) describe the concept of chaos.

Consider the Hermite polynomials $\{h_k, k \in \mathbb{N}\}$ on $\mathbb{R}$ defined by the series expansion

$$\exp\left(\lambda x - \frac{\lambda^2}{2}\right) = \sum_{k=0}^{\infty} \frac{\lambda^k}{\sqrt{k!}} h_k(x), \qquad \lambda, x \in \mathbb{R}.$$

The Hermite polynomials form an orthonormal basis of $L_2(\mathbb{R}, \gamma_1)$. Similarly, if $\underline{k} \in \mathbb{N}^{(\mathbb{N})}$, i.e. $\underline{k} = (k_1, k_2, \dots)$, $k_i \in \mathbb{N}$, with $|\underline{k}| = \sum_i k_i < \infty$, set, for $x = (x_i) \in \mathbb{R}^{\mathbb{N}}$,

$$H_{\underline{k}}(x) = h_{k_1}(x_1) h_{k_2}(x_2) \cdots .$$

Then $\{H_{\underline{k}}\,;\ \underline{k} \in \mathbb{N}^{(\mathbb{N})}\}$ form an orthonormal basis of $L_2(\mathbb{R}^{\mathbb{N}}, \gamma)$ where we recall that γ is the canonical Gaussian product measure on $\mathbb{R}^{\mathbb{N}}$.

For $0 \le \varepsilon \le 1$, introduce the bounded linear operator $T(\varepsilon) : L_2(\gamma) \to L_2(\gamma)$ defined by

$$T(\varepsilon) H_{\underline{k}} = \varepsilon^{|k|} H_{\underline{k}}$$

for any $\underline{k}$ with $|\underline{k}| < \infty$. It is not difficult to check that $T(\varepsilon)$ extends to a positive contraction on all $L_p(\gamma)$, $1 \le p < \infty$. This is actually clear from the following integral representation of $T(\varepsilon)$: if f is in $L_2(\gamma)$ and x in $\mathbb{R}^{\mathbb{N}}$,

$$T(\varepsilon) f(x) = \int f(\varepsilon x + (1 - \varepsilon^2)^{1/2} y)\, d\gamma(y).$$

If $t \ge 0$, $T_t = T(e^{-t})$ is known as the Hermite or Ornstein-Uhlenbeck semigroup.

The operators $T(\varepsilon)$, $0 \le \varepsilon \le 1$, satisfy a very important *hypercontractivity* property which is related to the integrability properties of Gaussian variables and chaos. This property indicates that for $1 < p < q < \infty$ and ε such that $|\varepsilon| \le [(p-1)/(q-1)]^{1/2}$, $T(\varepsilon)$ maps $L_p(\gamma)$ into $L_q(\gamma)$ *with norm* 1, i.e. for any f in $L_p(\gamma)$,

$$\|T(\varepsilon) f\|_q \le \|f\|_p . \tag{3.8}$$

A function f in $L_2(\gamma)$ can be written as

$$f = \sum_{\underline{k}} H_{\underline{k}} f_{\underline{k}}$$

where $f_{\underline{k}} = \int f H_{\underline{k}} d\gamma$ and the sum runs over all $\underline{k}$ in $\mathbb{N}^{(\mathbb{N})}$. We can also write

$$f = \sum_{d=0}^{\infty} \left(\sum_{|\underline{k}|=d} H_{\underline{k}} f_{\underline{k}} \right) = \sum_{d=0}^{\infty} Q_d f .$$

$Q_d f$ is named the chaos of degree d of f. Since $h_0 = 1$, $Q_0 f$ is simply the mean of f; $h_1(x) = x$, so chaos of degree 1 are Gaussian series $\sum_i g_i \alpha_i$. Chaos of degree 2 are of the type

$$\sum_{i \neq j} g_i g_j \alpha_{ij} + \sum_i (g_i^2 - 1)\alpha_i ,$$

etc. Now, the very definition of the Hermite operators $T(\varepsilon)$ shows that the action of $T(\varepsilon)$ on a chaos of order d is simply multiplication by ε^d, that is $T(\varepsilon)Q_d f = \varepsilon^d Q_d f$. This observation together with (3.8) has some interesting consequence. If we let, in (3.8), $p = 2$, $q > 2$ and $\varepsilon = (q-1)^{-1/2}$, we see that

$$\|Q_d f\|_q \leq (q-1)^{d/2} \|Q_d f\|_2 .$$

These inequalities of course imply strong exponential integrability properties of $Q_d f$. This follows for example from the next easy lemma which is obtained by a series expansion of the exponential function.

Lemma 3.7. Let d be an integer and let Z be a positive random variable. The following are equivalent:

(i) there is a constant K such that for any $p \geq 2$

$$\|Z\|_p \leq K p^{d/2} \|Z\|_2 ;$$

(ii) for some $\alpha > 0$,

$$\mathbb{E} \exp \alpha Z^{2/d} < \infty .$$

The integrability properties of Gaussian chaos we obtain in this way extend to chaos with coefficients in a Banach space. In particular, for the case of series, this provides an alternate approach to the integrability of Gaussian variables presented in Section 3.1; note the right order of magnitude of $K_{p,2} = K\sqrt{p}$. In this paragraph, we shall actually be concerned in more precise tail behavior of Gaussian chaos similar to the ones obtained previously for chaos of order 1. This will be accomplished again with the tool of the Gaussian isoperimetric inequality. For simplicity, we only treat the case of chaos of order 2. (Let us mention that this reduction simplifies, by elementary symmetry considerations, several non-trivial polarization arguments that are necessary in general and which are thus somewhat hidden in our treatment; we refer to [Bo4], [Bo7] for details on these aspects.) With respect to the preceding description of chaos, we will study more precisely *homogeneous* Gaussian *polynomials* which basically corresponds, for the degree 2, to convergent series of the type $\sum_{i,j} g_i g_j x_{ij}$ where (x_{ij}) is a sequence in a Banach space. Following the work of C. Borell [Bo4], [Bo7], since the constant 1 belongs to the closure of the homogeneous Gaussian polynomials of degree 2 (at least if the underlying Gaussian measure is infinite dimensional), the chaos described previously (and their vector valued version as well) are limit, in probability for example, of homogeneous Gaussian polynomials of the corresponding degree. This framework therefore includes, for the results we will describe, the usual meaning of Gaussian or Wiener chaos. (At least at the order 2, some aspects of this comparison can also be made apparent through simple symmetrization arguments.) We still use the terminology of chaos in this setting.

Let us now describe the framework in which we will work. As in Section 3.1, the case of convergent quadratic sums is somewhat too restrictive. Let again B be a Banach space with $D = \{f_n, n \geq 1\}$ in the unit ball of B' such that $\|x\| = \sup_{f \in D} |f(x)|$ for all x in B. Following the reduction to the canonical Gaussian distribution on $\mathbb{R}^{\mathbb{N}}$, we say that a random variable X with values in B is a Gaussian chaos of order 2 if, for each n, there exists a bilinear form

$$Q_n(x, x') = \sum_{i,j} a_{ij}^n x_i x'_j$$

on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ such that the sequence $(f_n(X))$ has the same distribution as $\left(\sum_{i,j} a_{ij}^n g_i g_j\right)$ where we recall that (g_i) is the canonical orthogaussian sequence. Therefore, for each n, $\sum_{i,j} a_{ij}^n g_i g_j$ is almost surely (or only in probability) convergent. If we set further, by analogy, $\|Q(x,x)\| = \sup_{n \geq 1} |Q_n(x,x)|$, $x \in \mathbb{R}^{\mathbb{N}}$, we are reduced as usual to the study of the tail behavior and integrability properties of $\|Q\|$ under the canonical Gaussian product measure γ on $\mathbb{R}^{\mathbb{N}}$.

To measure the size of the tail $\mathbb{P}\{\|X\| > t\}$ we use, as in the preceding section, several parameters of the distribution of X. First, consider the "decoupled" symmetric chaos Y associated to X defined in distribution by

$$f_n(Y) = \sum_{i,j} b_{ij}^n g_i g'_j \,, \qquad n \geq 1$$

where $b_{ij}^n = a_{ij}^n + a_{ji}^n$ and (g'_j) is an independent copy of (g_i). According to our usual notations, we denote below by $\mathbb{E}'$, $\mathbb{P}'$ partial expectation and probability with respect to (g'_j). Let then M and m be such that

$$\mathbb{P}\{\|X\| \leq M\} \geq 3/4 \qquad \text{and} \qquad \mathbb{P}\{\sup_{f \in D} (\mathbb{E}' f^2(Y))^{1/2} \leq m\} \geq 3/4 \,.$$

Let further

$$\sigma = \sigma(X) = \sup_{|h| \leq 1} \|Q(h,h)\| \,.$$

If we recall the situation for Gaussian variables in the previous section, we see that σ and M correspond respectively to the weak and strong parameters. The new and important parameter m appears as some intermediate quantity involving both the weak and strong topologies. Let us show that these parameters are actually well defined. It will suffice to show that the decoupled chaos Y exists. The key observation is that, for every $t > 0$,

$$\mathbb{P}\{\|Y\| > t\} \leq 2\mathbb{P}\{\|X\| > t/2\sqrt{2}\} \,. \tag{3.9}$$

We establish this inequality for finite sums and a norm given by a finite supremum; a limiting argument then shows that the quadratic sums defining Y are convergent as soon as the corresponding ones for X are, and, by increasing the finite supremum to the norm, that (3.9) is satisfied. We reduce to $(\mathbb{R}^{\mathbb{N}}, \gamma)$ and recall that $\|Q(x,x)\| = \sup_{n \geq 1} |Q_n(x,x)|$. Let, on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$,

$$\widehat{Q}_n(x,x') = \sum_{i,j} b_{ij}^n x_i x_j'$$

and set, with some further abuse in notation, $\|\widehat{Q}(x,x')\| = \sup_{n \geq 1} |\widehat{Q}_n(x,x')|$. Then, simply note that for x, x' in $\mathbb{R}^{\mathbb{N}}$,

$$2\widehat{Q}_n(x,x') = Q_n(x+x', x+x') - Q_n(x-x', x-x')$$

and that $x+x'$ and $x-x'$ have both the same distribution under $d\gamma(x)d\gamma(x')$ than $\sqrt{2}x$ under $d\gamma(x)$. From this and the preceding comments, it follows that Y is well defined and that (3.9) is satisfied.

It might be worthwhile to note that (3.9) can essentially be reversed. Using that the couple $((x+x')/\sqrt{2}, (x-x')/\sqrt{2})$ has the same distribution as (x,x') , one easily verifies that for all $t > 0$,

$$\mathbb{P}\{\|X - X'\| > t\} \leq 3\mathbb{P}\{\|Y\| > t/3\}$$

where X' is an independent copy of X. If the diagonal terms of X are all zero, we see by Jensen's inequality that X and Y have all their moments equivalent. It follows in particular from this observation and the subsequence arguments that if X is a *real* chaos (i.e. D is reduced to one point) with all its diagonal terms zero, then the parameters M and m are equivalent (at least if M is large enough, see below), and equivalent to $\|X\|_2$ and $\|Y\|_2$. We will use this observation later on in Chapter 11, Section 11.3.

Let $\widehat{M}$ be such that $\gamma\otimes\gamma((x,x');\|\widehat{Q}(x,x')\|\le\widehat{M})\ge 7/8$. By Fubini's theorem, $\gamma(A)\ge 3/4$ where

$$A=\{x;\gamma(x';\|\widehat{Q}(x,x')\|\le\widehat{M})\ge\frac{1}{2}\}.$$

Conditionally in x, $\|\widehat{Q}(x,x')\|$ is the norm of a Gaussian variable in x'. If $x\in A$, $\widehat{M}$ is larger than the median of $\|\widehat{Q}(x,x')\|$ and thus, by what was observed early in the preceding section, the supremum of the weak variances is less than $2\widehat{M}$. But this supremum is simply

$$\sup_{n\ge1}\left(\int\widehat{Q}_n(x,x')^2d\gamma(x')\right)^{1/2}.$$

Hence we may simply take $m=2\widehat{M}$ which is therefore well defined and finite. (Notice, for later purposes, that if M is chosen to satisfy $\mathbb{P}\{\|X\|\le M\}\ge 15/16$, we can take, by (3.9), $\widehat{M}=2\sqrt{2}M$, and thus also $m=4\sqrt{2}M$.)

Concerning σ, for every $k\in\ell_2$, $|k|\le 1$, $\gamma(x;\|\widehat{Q}(x,k)\|\le m)\ge 3/4\ge 1/2$. Hence, by the same reasoning as before,

$$\sup_{|h|\le1}\|\widehat{Q}(h,k)\|=\sup_{n\ge1}\left(\int\widehat{Q}_n(x,k)^2d\gamma(x)\right)^{1/2}\le 2m.$$

Therefore,

$$\sup_{|h|,|k|\le1}\|\widehat{Q}(h,k)\|\le 2m.$$

But this supremum is easily seen to be bigger than 2σ so that $\sigma\le m<\infty$.

As for series in the previous paragraph, notice that it is usually easier to work with expectations and moments rather than medians or quantiles. Actually, by independence and the results of Section 3.1 it is not difficult to see that $\mathbb{E}\|Y\|^2<\infty$ (actually $\mathbb{E}\|Y\|^p<\infty$ for every p) and that we have the following hierarchy in the parameters:

$$(\mathbb{E}\|Y\|^2)^{1/2}\ge(\mathbb{E}\sup_{f\in D}\mathbb{E}'f^2(Y))^{1/2}\ge 2\sigma.$$

After these long preliminaries in order to properly describe the various parameters we are using, we now state and prove the lemma which described the tail $\mathbb{P}\{\|X\| > t\}$ of a Gaussian chaos X in terms of these parameters. We shall not be interested here in concentration properties; some, however, can be obtained at the expense of some complications. Recall, also that for *real* chaos, the parameters M and m are equivalent so that the inequality of the lemma may be formulated only in terms of M and σ. This will be used in Section 11.3.

Lemma 3.8. Let X be a Gaussian chaos of order 2 as just defined with corresponding parameters M, m and σ. Then, for every $t > 0$,

$$\mathbb{P}\{\|X\| > M + mt + \sigma t^2\} \leq \exp(-t^2/2) .$$

Proof. It is a simple consequence of Theorem 1.2. We use the preceding notations. Let

$$A_1 = \{x\,;\ \|Q(x,x)\| \leq M\} ,$$
$$A_2 = \{x\,;\ \sup_{|k|\leq 1} \|\widehat{Q}(x,k)\| \leq m\}$$

and $A = A_1 \cap A_2$ so that, by definition of M and m, $\gamma(A) \geq 1/2$. By Theorem 1.2, more precisely (1.3), for any $t > 0$, $\gamma_*(A_t) \geq \Phi(t)$. But if $x \in A_t$, $x = a + th$ where $a \in A$, $|h| \leq 1$. Thus, for every n,

$$Q_n(x,x) = Q_n(a,a) + t\widehat{Q}_n(a,h) + t^2 Q_n(h,h)$$

and therefore

$$A_t \subset \{x\,;\ \|Q(x,x)\| \leq M + tm + t^2\sigma\} .$$

Lemma 3.8 is therefore established.

The next corollary is the qualitative result drawn from the preceding lemma concerning integrability and tail behavior of Gaussian chaos. It corresponds to Corollary 3.2 in the first section.

Corollary 3.9. Let X be a Gaussian chaos of order 2 with corresponding $\sigma = \sigma(X)$. Then

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}\{\|X\| > t\} = -\frac{1}{2\sigma}$$

or, equivalently,

$$\mathbb{E}\exp\left(\frac{1}{2\alpha}\|X\|\right) < \infty \qquad \text{if and only if } \ \alpha > \sigma .$$

Further all moments of X are equivalent.

Proof. That the limit is $\leq -1/2\sigma$ follows from Lemma 3.8; this also implies that $\mathbb{E}\exp(\|X\|/2\alpha) < \infty$ for $\alpha > \sigma$. To prove the converse assertions, let $\varepsilon > 0$ and choose $|h| \leq 1$ such that $\|Q(h,h)\| \geq \sigma - \varepsilon$. Given this $h = (h_i)_{i\geq 1}$, there exists an orthonormal basis $(h^i)_{i\geq 1}$ of ℓ_2 such that $h_1^i = h_i$ for every $i \geq 1$. By the rotational invariance of Gaussian measures, the distribution of $y = (\langle x, h^i \rangle)$ under γ is the same as the distribution of x. If we then set $\widetilde{y} = (\langle \widetilde{x}, h^i \rangle)$ where $\widetilde{x} = (0, x_2, x_3, \dots)$, we can write $y = x_1 h + \widetilde{y}$ and x_1 and $\widetilde{y}$ are independent. Since, for each n,

$$Q_n(y,y) = x_1^2 Q_n(h,h) + x_1 \widehat{Q}_n(\widetilde{y}, h) + Q_n(\widetilde{y}, \widetilde{y})$$

(where $\widehat{Q}_n$, Q were introduced prior to Lemma 3.8) and since $x_1 h - \widetilde{y}$ is distributed as y, a simple symmetry argument shows that one can find M such that

$$\gamma(x; \|Q(\widetilde{y}, \widetilde{y})\| \leq M\,, \quad \|\widehat{Q}(\widetilde{y}, h)\| \leq M) \geq \frac{1}{2}\,.$$

We then deduce from Fubini's theorem that for every $t > 0$

$$\begin{aligned}\gamma(x; \|Q(x,x)\| > t) &= \gamma(x; \|Q(y,y)\| > t) \\ &\geq \frac{1}{2}\gamma(x; x_1^2(\sigma - \varepsilon) > t + |x_1|M + M)\,.\end{aligned}$$

The proof of the first claim in Corollary 3.9 is then easily completed. That $\mathbb{E}\exp(\|X\|/2\sigma) = \infty$ can be established in the same way.

We have seen before Lemma 3.8 that if M is chosen to satisfy $\mathbb{P}\{\|X\| \leq M\} \geq 15/16$, then we can take $m \leq 4\sqrt{2}M$ and σ satisfies $\sigma \leq 4\sqrt{2}M$. Hence, if $p > 0$, integrating the inequality of Lemma 3.8 immediately yields

$$\|X\|_p \leq K_p M$$

for some constant K_p depending only on p. If we then have $q > 0$, simply take $M = (16\mathbb{E}\|X\|^q)^{1/q}$ from which the equivalence of the moments of X follows. Note that K_p is of the order of p when $p \to \infty$, in accordance with what we observed in the beginning through the hypercontractive estimates. The proof of Corollary 3.9 is thus complete.

We conclude this section with a refinement of the previous corollary along the lines of what we obtained in Theorem 3.3. Let X be as before a Gaussian chaos variable of order 2 with values in B and recall the symmetric decoupled chaos Y associated to X. Set then

$$\tau = \tau(X) = \inf\{\lambda > 0 : \mathbb{P}\{\sup_{f \in D}(\mathbb{E}' f^2(Y))^{1/2} \le \lambda\} > 0\}.$$

We can then state:

Theorem 3.10. Let X be a Gaussian chaos with corresponding $\sigma = \sigma(X)$ and $\tau = \tau(X)$. Then, for every $\tau' > \tau$,

$$\mathbb{E} \exp \frac{1}{2}\left(\left(\frac{\|X\|}{\sigma}\right)^{1/2} - \frac{\tau'}{2\sigma}\right)^2 < \infty .$$

Proof. We show that for every $\tau' > \tau$ there exists an integer N and a real number t_0 such that for every $t \ge t_0$

$$\mathbb{P}\{\|X\| > \tau' t + \sigma t^2\} \le \mathbb{P}\left\{\left(\sum_{i=1}^N g_i^2\right)^{1/2} > t\right\}. \tag{3.10}$$

This will be sufficient to establish the theorem. Indeed, if $\tau' > \tau'' > \tau$, let $\varepsilon = (\tau' - \tau'')/2\sigma > 0$. Then, applying (3.10) to τ'', for some integer N,

$$\begin{aligned}
&\int_{(\frac{\|X\|}{\sigma})^{1/2} > \frac{\tau'}{2\sigma}} \exp \frac{1}{2}\left(\left(\frac{\|X\|}{\sigma}\right)^{1/2} - \frac{\tau'}{2\sigma}\right)^2 d\mathbb{P} \\
&\quad \le \exp\left(\frac{t_0^2}{2}\right) + \int_{t_0}^{\infty} \mathbb{P}\left\{\left(\frac{\|X\|}{\sigma}\right)^{1/2} > \frac{\tau'}{2\sigma} + t\right\} t \exp\left(\frac{t^2}{2}\right) dt \\
&\quad \le \exp\left(\frac{t_0^2}{2}\right) + \int_{t_0}^{\infty} \mathbb{P}\left\{\left(\frac{\|X\|}{\sigma}\right)^{1/2} > \frac{\tau''}{2\sigma} + (t+\varepsilon)\right\} t \exp\left(\frac{t^2}{2}\right) dt \\
&\quad \le \exp\left(\frac{t_0^2}{2}\right) + \int_{t_0}^{\infty} \mathbb{P}\{\|X\| > \tau''(t+\varepsilon) + \sigma(t+\varepsilon)^2\} t \exp\left(\frac{t^2}{2}\right) dt \\
&\quad \le \exp\left(\frac{t_0^2}{2}\right) + K(N) \int_0^{\infty} (1+\varepsilon+t)^{N+1} \exp\left(\frac{t^2}{2} - \frac{(t+\varepsilon)^2}{2}\right) dt < \infty .
\end{aligned}$$

Let us therefore prove (3.10). Recall the notations of the proof of Lemma 3.4. Given an integer N, a point x in $\mathbb{R}^{\mathbb{N}}$ is decomposed in (y,z) with $y \in \mathbb{R}^N$ and $z \in \mathbb{R}^{]N,\infty[}$. Further $\gamma = \gamma_N \otimes \gamma_{]N,\infty[}$ and if A is a Borel set in $\mathbb{R}^{\mathbb{N}}$, we set $B = \{z \in \mathbb{R}^{]N,\infty[} ;\ (0,z) \in A\}$.

If $\tau' > \tau$, $\gamma(A_2) > 0$ where

$$A_2 = \{x\,;\ \sup_{|k|\le 1} \|\widehat{Q}(x,k)\| \le \tau'\}$$

and where $\widehat{Q}$ is the quadratic form associated to the decoupled chaos Y (cf. the proof of Lemma 3.8). Choose M large enough such that if

$$A_1 = \{x\,;\ \|Q(x,x)\| \le M\}\,,$$

and $A_3 = A_1 \cap A_2$, then $\gamma(A_3) > 0$. There exists an integer ℓ such that if

$$A_3' = \{x\,;\ \max_{n\le \ell} |Q_n(x,x)| \le M\,,\ \max_{n\le \ell}\sup_{|k|\le 1} |\widehat{Q}_n(x,k)| \le \tau'\}\,,$$

then $\gamma(A_3') < 4\gamma(A_3)/3$. By a simple approximation, replacing M and τ' by $M+\varepsilon$ and $\tau'+\varepsilon$ if necessary, we may assume the bilinear forms Q_n , $\widehat{Q}_n$, $n \le \ell$, only depend on a finite number of coordinates. Hence, for some integer N , $A_3' = L \times \mathbb{R}^{]N,\infty[}$ with $L \subset \mathbb{R}^N$. By Fubini's theorem, there exists then y in $\mathbb{R}^N$ such that

$$\gamma_{]N,\infty[}(z; (y,z) \in A_3) \ge \frac{3}{4}\,.$$

By symmetry we also have

$$\gamma_{]N,\infty[}(z; (y,-z) \in A_3) \ge \frac{3}{4}\,.$$

The intersection of these two sets of measure bigger than $3/4$ has therefore measure bigger than $1/2$. Let z belong to this intersection. By convexity,

$$\sup_{|k|\le 1} \|\widehat{Q}((0,z),k)\| \le \tau'\,.$$

Moreover, since

$$\|Q((y,z),(y,z))\|\,,\ \|Q((y,-z),(y,-z))\| \le M\,,$$

summing, we get that

$$\|Q((0,z),(0,z))\| \le M + \sup_{n\ge 1}\left|\sum_{i,j=1}^{N} a_{ij}^n y_i y_j\right| \le M + \sigma |y|^2\,.$$

Hence, if we set $M' = M + \sigma|y|^2$ and

$$A = \{x\,;\ \|Q(x,x)\| \le M'\,,\ \sup_{|k|\le 1} \|\widehat{Q}(x,k)\| \le \tau'\}\,,$$

it follows that $B = \{z \in \mathbb{R}^{]N,\infty[}\,;\ (0,z) \in A\}$ satisfies $\gamma_{]N,\infty[}(B) \ge 1/2$. We are then in a position to apply Lemma 3.5 from which we get that for every $t \ge 0$:

$$\gamma^*(x\,;\ x \notin A_t) \le \gamma_{N+1}(y'\,;\ |y'| > t)\,.$$

Now it has simply to be observed that

$$A_t \subset \{x\,;\ \|Q(x,x)\| \le M' + t\tau' + t^2\sigma\}\,.$$

Indeed, if $x \in A_t$, $x = a + th$ with $a \in A$, $|h| \le 1$, then

$$\|Q(x,x)\| \le \|Q(a,a)\| + t\|\widehat{Q}(a,y)\| + t^2\|Q(h,h)\|\,.$$

(3.10) then easily follows from the preceding playing with $\tau' > \tau$ and t to be large enough. The proof of Theorem 3.10 is complete.

To conclude, let us briefly mention the formal corresponding results for the chaos of degree $d > 2$. If X is such a chaos, and if σ is defined analogously, Corollary 3.9 reads:

$$\lim_{t\to\infty} \frac{1}{t^{2/d}} \log \mathbb{P}\{\|X\| > t\} = -\frac{1}{2\sigma^{2/d}}\,.$$

Theorem 3.10 is somewhat more difficult to translate. τ is defined appropriately from the associated d-decoupled symmetric chaos on which one takes weak moments on $d-1$ coordinates and the strong parameter on the remaining one. We then get that for all $\tau' > \tau$

$$\mathbb{E}\exp \frac{1}{2}\left(\left(\frac{\|X\|}{\sigma}\right)^{1/d} - \frac{\tau'}{d\sigma}\right)^2 < \infty\,.$$

One could possibly imagine further refinements involving $d+1$ parameters.

3.3 Comparison theorems

In the last part of this chapter we investigate the Gaussian comparison theorems which, together with integrability, are very important and useful tools in the Probability calculus in Banach spaces. These results, which may be considered as geometrical, first step in with the so-called Slepian's lemma on which the further results are variations. We present here some of these statements in the scope of their further applications.

Assume we are given two Gaussian random vectors $X = (X_1, \dots, X_N)$ and $Y = (Y_1, \dots, Y_N)$ in $\mathbb{R}^N$. In order to describe the question we would like to study, let us assume first, as an example, that the covariance structure of X dominates that of Y in the (strong) sense that for every α in $\mathbb{R}^N$

$$\mathbb{E}\langle \alpha, Y\rangle^2 \le \mathbb{E}\langle \alpha, X\rangle^2 .$$

Then, for any convex set C in $\mathbb{R}^N$

$$\mathbb{P}\{Y \notin C\} \le 2\mathbb{P}\{Y \notin C\} . \tag{3.11}$$

Indeed, we may of course assume for the proof that X and Y are independent. If Z is a Gaussian variable independent from Y with covariance $\mathbb{E}Z_iZ_j = \mathbb{E}X_iX_j - \mathbb{E}Y_iY_j$ (which is positive definite from the assumption), X has the same distribution as $Y + Z$. By independence and symmetry, X has also the same distribution as $Y - Z$. Hence, by convexity,

$$\begin{aligned} \mathbb{P}\{Y \notin C\} &= \mathbb{P}\left\{\frac{1}{2}[(Y+Z) + (Y-Z)] \notin C\right\} \\ &\le \mathbb{P}\{Y + Z \notin C\} + \mathbb{P}\{Y - Z \notin C\} \end{aligned}$$

which is (3.11).

It should be noted that deeper tools can actually yield (3.11) without the numerical constant 2 for all convex and symmetric (with respect to the origin) sets C. Indeed, by Fubini's theorem,

$$\mathbb{P}\{X \in C\} = \mathbb{P}\{Y + Z \in C\} = \int \mathbb{P}\{Y + z \in C\} d\mathbb{P}_Z(z) .$$

Now, the concavity inequalities (1.1) or (1.2) and symmetry ensure that, for every z in $\mathbb{R}^N$,

$$\mathbb{P}\{Y \in C - z\} \le \mathbb{P}\{Y \in C\} ,$$

hence the announced claim.

Typically (3.11) (or the preceding) is used to show a property like the following one: if (X_n) is a sequence of Gaussian Radon random variables with values in a Banach space B such that $\mathbb{E}f^2(X_n) \leq \mathbb{E}f^2(X)$ for all f in B', all n and some Gaussian Radon variable X in B, then the sequence (X_n) is tight. Indeed, since X is Radon, for every $\varepsilon > 0$, there exists a compact set C which may be chosen convex such that $\mathbb{P}\{X \in C\} \geq 1 - \varepsilon$. Since B may be assumed to be separable, there is a sequence (f_k) in B such that $x \in K$ whenever $f_k(x) \leq 1$ for all k. The conclusion is then immediate from (3.11) which implies that $\mathbb{P}\{X_n \in C\} \geq 1 - 2\varepsilon$ for all n.

The first comparison property we now present is the abstract statement from which we will deduce many consequences of interest. Its quite easy proof clearly describes the Gaussian properties which enter these questions. The result is similar in nature to (3.11) but under weaker conditions on the covariance structures.

Theorem 3.11. Let $X = (X_1, \ldots, X_N)$ and $Y = (Y_1, \ldots Y_N)$ be Gaussian random variables in $\mathbb{R}^N$. Assume that

$$\begin{aligned} \mathbb{E}X_iX_j &\leq \mathbb{E}Y_iY_j \qquad \text{if } (i,j) \in A\,, \\ \mathbb{E}X_iX_j &\geq \mathbb{E}Y_iY_j \qquad \text{if } (i,j) \in B\,, \\ \mathbb{E}X_iX_j &= \mathbb{E}Y_iY_j \qquad \text{if } (i,j) \notin A \cup B \end{aligned}$$

where A and B are subsets of $\{1, \ldots, N\} \times \{1, \ldots, N\}$. Let f be a function on $\mathbb{R}^N$ such that its second derivatives in the sense of distributions satisfy

$$\begin{aligned} D_{ij}f &\geq 0 \qquad \text{if } (i,j) \in A\,, \\ D_{ij}f &\leq 0 \qquad \text{if } (i,j) \in B\,. \end{aligned}$$

Then

$$\mathbb{E}f(X) \leq \mathbb{E}f(Y)\,.$$

Proof. As before we may assume X and Y to be independent. Set, for $t \in [0,1]$, $X(t) = (1-t)^{1/2}X + t^{1/2}Y$ and $\varphi(t) = \mathbb{E}f(X(t))$. We have

$$\varphi'(T) = \sum_{i=1}^{N} \mathbb{E}(D_i f(X(t))X_i'(t))\,.$$

Let now t and i be fixed. It is easily seen that, for every j,

$$\mathbb{E}X_j(t)X_i'(t) = \frac{1}{2}\mathbb{E}(Y_jY_i - X_jX_i)\,.$$

The hypotheses of the theorem then indicate that we can write

$$X_j(t) = \alpha_j X_i'(t) + Z_j$$

where Z_j is orthogonal to $X_i'(t)$ and $\alpha_j \geq 0$ if $(i,j) \in A$, $\alpha_j \leq 0$ if $(i,j) \in B$, $\alpha_j = 0$ if $(i,j) \notin A \cup B$. If we now examine $\mathbb{E}(D_i f(X(t))X_i'(t))$ as a function of the α_j 's (for $(i,j) \in A \cup B$), the hypotheses on f show that this function is increasing of those α_j 's such that $(i,j) \in B$. But, by the orthogonality and therefore independence, this function vanishes when all the α_j 's are 0 since

$$\mathbb{E}(D_i f(Z)X_i'(t)) = \mathbb{E}(D_i f(Z))\mathbb{E}X_i'(t) = 0\,.$$

Hence $\mathbb{E}(D_i f(X(t))X_i'(t)) \geq 0$, $\varphi'(t) \geq 0$ and therefore $\varphi(0) \leq \varphi(1)$ which is the conclusion of the theorem.

As a first corollary, we state Slepian's lemma. It is simply obtained by taking in the theorem $A = \{(i,j)\,;\, i \neq j\}$, $B = \emptyset$ and $f = I_G$ where G is a product of half-lines $]-\infty, -\lambda_i]$ for which the hypotheses on the second derivatives are immediately verified. The claim concerning expectations of maxima in the next statement simply follows from the integration by parts formula

$$\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > t\}\,dt - \int_0^\infty \mathbb{P}\{X < -t\}\,dt\,.$$

Corollary 3.12 Let X and Y be Gaussian in $\mathbb{R}^N$ such that

$$\begin{cases} \mathbb{E}X_iX_j \leq \mathbb{E}Y_iY_j & \text{for all } i \neq j\,, \\ \mathbb{E}X_i^2 = \mathbb{E}Y_i^2 & \text{for all } i\,. \end{cases}$$

Then, for all real numbers λ_i, $i \leq N$,

$$\mathbb{P}\{\bigcup_{i=1}^N (Y_i > \lambda_i)\} \leq \mathbb{P}\{\bigcup_{i=1}^N (X_i > \lambda_i)\}\,.$$

In particular, by integration by parts,

$$\mathbb{E}\max_{i \leq N} Y_i \leq \mathbb{E}\max_{i \leq N} X_i\,.$$

The next corollary relies on a more elaborated use of Theorem 3.11. It will be useful in various applications both in this chapter yet and in Chapters 9 and 15.

Corollary 3.13. Let $X = (X_{i,j})$ and $Y = (Y_{i,j})$, $1 \le i \le n$, $1 \le j \le m$, be Gaussian random vectors such that

$$\begin{cases} \mathbb{E}X_{i,j}X_{i,k} \le \mathbb{E}Y_{i,j}Y_{i,k} & \text{for all } i,j,k\,, \\ \mathbb{E}X_{i,j}X_{\ell,k} \ge \mathbb{E}Y_{i,j}Y_{\ell,k} & \text{for all } i \ne \ell \text{ and } j,k\,, \\ \mathbb{E}X_{i,j}^2 = \mathbb{E}Y_{i,j}^2 & \text{for all } i,j\,. \end{cases}$$

Then, for all real numbers $\lambda_{i,j}$,

$$\mathbb{P}\{\bigcap_{i=1}^{n}\bigcup_{j=1}^{m}(Y_{i,j} > \lambda_{i,j})\} \le \mathbb{P}\{\bigcap_{i=1}^{n}\bigcup_{j=1}^{m}(X_{i,j} > \lambda_{i,j})\}\,.$$

In particular,

$$\mathbb{E}\min_{i\le n}\max_{j\le m} Y_{i,j} \le \mathbb{E}\min_{i\le n}\max_{j\le m} X_{i,j}\,.$$

Proof. Let $N = mn$. For $I \in \{1,\dots,N\}$ let $i = i(I)$, $j = j(I)$ be the unique $1 \le i \le n$, $1 \le n \le m$ such that $I = m(i-1)+j$. Consider then X and Y as random vectors in $\mathbb{R}^N$ indexed in this way, i.e. $X_I = X_{i(I),j(I)}$. Let

$$A = \{(I,J)\,;\ i(I) = i(J)\}\,,$$

$$B = \{(I,J)\,;\ i(I) \ne i(J)\}\,.$$

Then the first set of hypotheses of Theorem 3.11 is fulfilled. Taking further f to be the indicator function of the set

$$\bigcup_{i=1}^{n}\bigcap_{I/i(I)=i}\{x \in \mathbb{R}^N\,;\ x_I > \lambda_{i,j(I)}\}\,,$$

Theorem 3.11 implies the conclusion by taking complements.

In the preceding results the comparison was made possible by assumptions on the respective covariances of the Gaussian vectors with, especially, conditions of equality on the "diagonal". In practice, it is often more convenient to deal with the corresponding L_2-metrics $\|X_i - X_j\|_2$ which do not require those special conditions on the diagonal. The next statement is a simple consequence of Corollary 3.12 in this direction.

Corollary 3.14. Let $X = (X_1,\dots,X_N)$ and $Y = (Y_1,\dots,Y_N)$ be Gaussian variables in $\mathbb{R}^N$ such that for every i,j

$$\mathbb{E}|Y_i - Y_j|^2 \le \mathbb{E}|X_i - X_j|^2\,.$$

Then

$$\mathbb{E}\max_{i\le N} Y_i \le 2\mathbb{E}\max_{i\le N} X_i\,.$$

Proof. Replacing $X=(X_i)_{i\le N}$ by $(X_i-X_1)_{i\le N}$ we may and do assume that $X_1=0$ and similarly $Y_1=0$. Let $\sigma=\max_{i\le N}(\mathbb{E}X_i^2)^{1/2}$ and consider the Gaussian variables $\overline{X}$ and $\overline{Y}$ in $\mathbb{R}^N$ defined by, $i\le N$,

$$\begin{aligned}\overline{X}_i &= X_i + g(\sigma^2+\mathbb{E}Y_i^2-\mathbb{E}X_i^2)^{1/2}\,,\\ \overline{Y}_i &= Y_i+g\sigma\end{aligned}$$

where g is standard normal independent from X and Y. It is easily seen that

$$\mathbb{E}\overline{Y}_i^2=\mathbb{E}\overline{X}_i^2=\sigma^2+\mathbb{E}Y_i^2$$

while

$$\mathbb{E}|\overline{Y}_i-\overline{Y}_j|^2=\mathbb{E}|Y_i-Y_j|^2\le\mathbb{E}|X_i-X_j|^2\le\mathbb{E}|\overline{X}_i-\overline{X}_j|^2$$

so that $\mathbb{E}\overline{X}_i\overline{X}_j\le\mathbb{E}\overline{Y}_i\overline{Y}_j$ for all $i\ne j$. We are thus in the hypotheses of Corollary 3.12 and therefore

$$\mathbb{E}\max_{i\le N}\overline{Y}_i\le\mathbb{E}\max_{i\le N}\overline{X}_i\,.$$

Now, clearly, $\mathbb{E}\max_{i\le N}\overline{Y}_i=\mathbb{E}\max_{i\le N}Y_i$ while

$$\mathbb{E}\max_{i\le N}\overline{X}_i\le\mathbb{E}\max_{i\le N}X_i+\sigma\mathbb{E}g^+$$

where we have used that $\mathbb{E}Y_i^2\le\mathbb{E}X_i^2$ (since $X_1=Y_1=0$). But now,

$$\begin{aligned}\sigma=\max_{i\le N}(\mathbb{E}X_i^2)^{1/2} &= \frac{1}{\mathbb{E}|g|}\max_{i\le N}\mathbb{E}|X_i|\\ &\le\frac{1}{\mathbb{E}|g|}\mathbb{E}\max_{i,j\le N}|X_i-X_j|\\ &=\frac{2}{\mathbb{E}|g|}\mathbb{E}\max_{i\le N}X_i=\frac{1}{\mathbb{E}g^+}\mathbb{E}\max_{i\le N}X_i\end{aligned}$$

where we have used again, in the first inequality, that $X_1=0$. This bound on σ and the preceding finish the proof.

If $X=(X_1,\dots X_N)$ is Gaussian in $\mathbb{R}^N$, by symmetry,

$$\mathbb{E}\max_{i,j}|X_i-X_j|=\mathbb{E}\max_{i,j}(X_i-X_j)=2\mathbb{E}\max_i X_i\,.$$

The comparison theorems usually deal with $\max_i X_i$ or $\max_{i,j} |X_i - X_j| = \max_{i,j}(X_i - X_j)$ rather than $\max_i |X_i|$. Of course, for every $i_0 \leq N$,

$$\begin{aligned}\mathbb{E}\max_i X_i \leq \mathbb{E}\max_i |X_i| &\leq \mathbb{E}|X_{i_0}| + \mathbb{E}\max_{i,j}|X_i - X_j| \\ &\leq \mathbb{E}|X_{i_0}| + 2\mathbb{E}\max_i X_i\,.\end{aligned}$$

But in general the comparison results do not apply directly to $\mathbb{E}\max_i |X_i|$ (see however [Si1], [Si2]); take for example $Y_i = X_i + cg$ where g is standard normal independent from X in Corollary 3.14 and let c tend to infinity. Actually, one convenient feature of $\mathbb{E}\max_i X_i$ is that for any real mean zero random variable Z, $\mathbb{E}\max_i(X_i + Z) = \mathbb{E}\max_i X_i$.

The numerical constant 2 in the preceding corollary is not best possible and can be improved to 1 with however a somewhat more complicated proof. On the other hand, under the hypotheses of Corollary 3.12, we also have that, for all $\lambda > 0$,

$$\mathbb{P}\{\max_{i,j}|Y_i - Y_j| > \lambda\} \leq 2\mathbb{P}\{\max_i X_i > \frac{\lambda}{2}\}\,.$$

Following the proof of Corollary 3.14 we can then obtain that if X and Y are Gaussian vectors in $\mathbb{R}^N$ such that for all i,j, $\|Y_i - Y_j\|_2 \leq \|X_i - X_j\|_2$, then, $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\{\max_{i,j}|Y_i - Y_j| > \lambda\} &\leq 2\mathbb{P}\{\max_{i,j}|X_i - X_j| > \frac{\lambda}{4}\} \\ &\quad + 2\mathbb{P}\{\max_{i,j}(\mathbb{E}|X_i - X_j|^2)^{1/2} g^+ > \frac{\lambda}{4}\}\,.\end{aligned} \tag{3.12}$$

This inequality can of course be integrated by parts. This observation suggests that the functional $\max_{i,j}|X_i - X_j|$ is perhaps more natural in comparison theorems. The next result (due to X. Fernique [Fer4]) which we state without proof completely answers these questions.

Theorem 3.15. Let X and Y be Gaussian random vectors in $\mathbb{R}^N$ such that for every i,j

$$\mathbb{E}|Y_i - Y_j|^2 \leq \mathbb{E}|X_i - X_j|^2\,.$$

Then, for every non-negative convex increasing function F on $\mathbb{R}_+$

$$\mathbb{E}F(\max_{i,j}|Y_i - Y_j|) \leq \mathbb{E}F(\max_{i,j}|X_i - X_j|)\,.$$

There is also a version of Corollary 3.13 with conditions on L_2-distances. Again the proof is more involved so that we only state the result. It is due to Y. Gordon [Gor1], [Gor3].

Theorem 3.16. Let $X = (X_{i,j})$ and $Y = (Y_{i,j})$, $1 \le i \le n$, $1 \le j \le m$ be Gaussian random vectors such that

$$\begin{cases} \mathbb{E}|Y_{i,j} - Y_{i,k}|^2 \le \mathbb{E}|X_{i,j} - X_{i,k}|^2 & \text{for all } i, j, k\,, \\ \mathbb{E}|Y_{i,j} - Y_{\ell,k}|^2 \ge \mathbb{E}|Y_{i,j} - Y_{\ell,k}|^2 & \text{for all } i \neq \ell \text{ and } j, k\,. \end{cases}$$

Then

$$\mathbb{E} \min_{i \le n} \max_{j \le m} Y_{i,j} \le \mathbb{E} \min_{i \le n} \max_{j \le m} X_{i,j}\,.$$

Among the various consequences of the preceding comparison properties, let us start with an elementary one which we only present in the scope of the next chapter where a similar result for Rademacher averages is obtained. We use Theorem 3.15 for convenience but a somewhat weaker result can be obtained through (3.12).

Corollary 3.17. Let T be bounded in $\mathbb{R}^N$ and consider the Gaussian process indexed by T defined as $\sum_{i=1}^N g_i t_i$, $t = (t_1, \ldots, t_N) \in T \subset \mathbb{R}^N$. Let further $\varphi_i : \mathbb{R} \to \mathbb{R}$, $i \le N$, be contractions with $\varphi_i(0) = 0$. Then for any non-negative convex increasing function F on $\mathbb{R}_+$,

$$\mathbb{E}F\left(\frac{1}{2}\sup_{t\in T}\left|\sum_{i=1}^N g_i\varphi_i(t_i)\right|\right) \le \mathbb{E}F\left(2\sup_{t\in T}\left|\sum_{i=1}^N g_i t_i\right|\right)\,.$$

Proof. Let $u \in T$. We can write by convexity:

$$\begin{aligned} \mathbb{E}F\left(\frac{1}{2}\sup_{t\in T}\left|\sum_{i=1}^N g_i\varphi_i(t_i)\right|\right) &\le \frac{1}{2}\mathbb{E}F\left(\sup_{t\in T}\left|\sum_{i=1}^N g_i(\varphi_i(t_i) - \varphi_i(u_i))\right|\right) + \frac{1}{2}\mathbb{E}F\left(\left|\sum_{i=1}^N g_i\varphi_i(u_i)\right|\right) \\ &\le \frac{1}{2}\mathbb{E}F\left(\sup_{s,t}\left|\sum_{i=1}^N g_i(\varphi_i(s_i) - \varphi_i(t_i))\right|\right) + \frac{1}{2}\mathbb{E}F\left(\left|\sum_{i=1}^N g_i u_i\right|\right) \end{aligned}$$

where we have use that $|\varphi_i(u_i)| \le |u_i|$ since $\varphi_i(0) = 0$. Now, by Theorem 3.15 (and a trivial approximation reducing to finite supremum), the preceding is further majorized by

$$\frac{1}{2}\mathbb{E}F\left(\sup_{s,t}\left|\sum_{i=1}^N g_i(s_i - t_i)\right|\right) + \frac{1}{2}\mathbb{E}F\left(\left|\sum_{i=1}^N g_i u_i\right|\right)$$

since by contraction, for every s, t,

$$\sum_{i=1}^N |\varphi_i(s_i) - \varphi_i(t_i)|^2 \le \sum_{i=1}^N |s_i - t_i|^2\,.$$

Corollary 3.17 is therefore established.

A most important consequence of the comparison theorems is the so-called Sudakov's minoration. We shall come back to it in Chapter 12 when we investigate regularity of Gaussian processes but it is fruitful to already record it at this stage. To introduce it, let us first observe the following easy facts. Let $X = (X_1, \ldots, X_N)$ be Gaussian in $\mathbb{R}^N$. Then

$$\mathbb{E}\max_{i\le N} X_i \le 3(\log N)^{1/2}\max_{i\le N}(\mathbb{E}X_i^2)^{1/2}. \tag{3.13}$$

Indeed, assume by homogeneity that $\max_{i\le N}\mathbb{E}X_i^2 \le 1$; then, for every $\delta \ge 0$, by integration by parts,

$$\begin{aligned}\mathbb{E}\max_{i\le N} X_i \le \mathbb{E}\max_{i\le N}|X_i| &\le \delta + N\int_\delta^\infty \mathbb{P}\{|g|>t\}\,dt \\ &\le \delta + N\sqrt{\frac{\pi}{2}}\exp(-\delta^2/2)\end{aligned}$$

where g is a standard normal. Choose then simply $\delta = (2\log N)^{1/2}$. Note the comparison with (3.6).

The preceding inequality is *two-sided* for the *canonical* Gaussian vector $(g_1, \ldots, g_N)$ where we recall that g_i are independent standard normal variables. Namely, for some numerical constant K,

$$K^{-1}(\log N)^{1/2} \le \mathbb{E}\max_{i\le N} g_i \le K(\log N)^{1/2}. \tag{3.14}$$

Indeed, since $\mathbb{E}\max(g_1, g_2) \ge 1/3$ (for example), we may assume N to be large enough. Note that, by independence and identical distribution, for every $\delta > 0$,

$$\begin{aligned}\mathbb{E}\max_{i\le N}|g_i| &\ge \int_0^\delta [1-(1-\mathbb{P}\{|g|>t\})^N]\,dt \\ &\ge \delta[1-(1-\mathbb{P}\{|g|>\delta\})^N].\end{aligned}$$

Now

$$\mathbb{P}\{|g|>\delta\} = \sqrt{\frac{2}{\pi}}\int_\delta^\infty \exp(-t^2/2)\,dt \ge \sqrt{\frac{2}{\pi}}\exp(-(\delta+1)^2/2).$$

Choose then for example $\delta = (\log N)^{1/2}$ (N large) so that $\mathbb{P}\{|g|>\delta\} \ge 1/N$ and hence

$$\mathbb{E}\max_{i\le N}|g_i| \ge \delta\left[1-\left(1-\frac{1}{N}\right)^N\right] \ge \delta\left(1-\frac{1}{e}\right).$$

Since $\mathbb{E}\max_{i\leq N}|g_i| \leq \mathbb{E}|g| + 2\mathbb{E}\max_{i\leq N} g_i$, this proves the lower bound in (3.14) since N is assumed to be large enough. The upper bound has been established before.

If (T,d) is a metric or pseudo-metric space (d need not separate points of T), denote by $N(T,d;\varepsilon)$ the minimal number of open balls of radius $\varepsilon > 0$ in the metric d necessary to cover T (the results we present are actually identical with closed balls). Of course $N(T,d;\varepsilon)$ need not be finite in which case we agree that $N(T,d;\varepsilon) = \infty$. $N(T,d;\varepsilon)$ is finite for each $\varepsilon > 0$ if and only if (T,d) is totally bounded.

Let $X = (X_t)_{t\in T}$ be a Gaussian process indexed by a set T . As will be further and deeply investigated in Chapter 12, one fruitful way to analyze the regularity properties of the process X is to study the "geometric" properties of T with respect to the L_2 pseudo-metric d_X induced by X defined as

$$d_X(s,t) = \|X_s - X_t\|_2 , \qquad s,t \in T .$$

The next theorem is an estimate of the size of $N(T,d_X;\varepsilon)$ for each $\varepsilon > 0$ in terms of the supremum of X . In the statement, we simply let

$$\mathbb{E}\sup_{t\in T} X_t = \sup\{\mathbb{E}\sup_{t\in F} X_t \,;\ F \ \text{ finite in } T\} .$$

Theorem 3.18. Let $X = (X_t)_{t\in T}$ be a Gaussian process with L_2 -metric d_X . Then, for each $\varepsilon > 0$,

$$\varepsilon(\log N(T,d_X;\varepsilon))^{1/2} \leq K\mathbb{E}\sup_{t\in T} X_t$$

where K is a numerical constant. In particular, if X is almost surely bounded, (T,d_X) is totally bounded.

Proof. Let N be such that $N(T,d_X;\varepsilon) \geq N$. There exists $U \subset T$ with $\mathrm{Card}U = N$ and such that $d_X(u,v) > \varepsilon$ for all $u \neq v$ in U . Let $(g_u)_{u\in U}$ be standard normal independent variables and consider $X'_u = \frac{\varepsilon}{\sqrt{2}} g_u$, $u \in U$. Then, clearly, for all u,v ,

$$\|X'_u - X'_v\|_2 = \varepsilon < d_X(u,v) .$$

Therefore, by Corollary 3.14,

$$\mathbb{E}\sup_{u\in U} X'_u \leq 2\mathbb{E}\sup_{u\in U} X_u .$$

If we now recall (3.14) we see that

$$\mathbb{E}\sup_{u\in U} X'_u \geq K^{-1}\frac{\varepsilon}{\sqrt{2}}(\log\ \mathrm{Card}U)^{1/2} = K^{-1}\frac{\varepsilon}{\sqrt{2}}(\log N)^{1/2} .$$

The conclusion follows.

Theorem 3.18 admits a slight strengthening when the process is continuous.

Corollary 3.19. If the Gaussian process $X = (X_t)_{t \in T}$ has a version with almost all bounded and continuous sample paths on (T, d_X), then

$$\lim_{\varepsilon \to 0} \varepsilon (\log N(T, d_X; \varepsilon))^{1/2} = 0 .$$

Proof. We denote by X itself the bounded continuous (and therefore separable) version. By the integrability properties of Gaussian vectors and compactness of (T, d_X) (since X is also bounded),

$$\lim_{\delta \to 0} \mathbb{E} \sup_{d_X(s,t) < \delta} |X_s - X_t| = 0 .$$

For every $\eta > 0$, let $\delta > 0$ be small enough that

$$\mathbb{E} \sup_{d_X(s,t) < \delta} |X_s - X_t| < \eta .$$

Let A be finite in T such that the balls of radius δ with centers in A cover T (such an A exists by Theorem 3.18). Let $\varepsilon > 0$. By Theorem 3.18, for every s in A there exists $A_s \subset T$ satisfying

$$\varepsilon (\log \ \mathrm{Card} A_s)^{1/2} \leq K\eta$$

and such that if $t \in T$ and $d_X(s,t) < \delta$ there exists u in A_s with $d_X(u,t) < \varepsilon$. Let then $B = \bigcup_{s \in A} A_s$. Each point of T is within distance ε of an element of B; hence

$$N(T, d_X; \varepsilon) \leq \ \mathrm{Card} B \leq \ \mathrm{Card} A \max_s \ \mathrm{Card} A_s .$$

Therefore

$$\varepsilon (\log N(T, d_X; \varepsilon))^{1/2} \leq \varepsilon (\log \ \mathrm{Card} A)^{1/2} + K\eta .$$

Letting ε, and then η, tend to 0 concludes the proof.

Sudakov's minoraton is the occasion of a short digression on some *dual* formulation. Let T be a convex body in $\mathbb{R}^N$, i.e. T is bounded convex symmetric about the origin with non-empty interior in $\mathbb{R}^N$ (T is a Banach ball). Consider the Gaussian process $X = (X_t)_{t \in T}$ defined as

$$X_t = \sum_{i=1}^{N} g_i t_i , \qquad t = (t_1, \ldots, t_N) \in T \subset \mathbb{R}^N .$$

Set

$$\ell(T) = \mathbb{E}\sup_{t\in T}|X_t| = \mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^N g_i t_i\right| = \int \sup_{t\in T}|\langle x,t\rangle| d\gamma_N(x)\,.$$

The L_2-metric of X is of course simply the Euclidean metric in $\mathbb{R}^N$. If A, B are sets in $\mathbb{R}^N$, denote by $N(A,B)$ the minimal number of translates of B by elements of A necessary to cover A. For example, $N(T, d_X; \varepsilon) = N(T, \varepsilon B_2)$ where B_2 is the Euclidean (open, for consistency!) unit ball of $\mathbb{R}^N$. Sudakov's minoration indicates that the rate of growth of $N(T,\varepsilon B_2)$ when $\varepsilon \to 0$ is controlled by $\ell(T)$. It might be interesting to point out here that the dual version of this result is also true, namely that the sup-norm of X controls in the same way $N(B_2, \varepsilon T^0)$ where $T^0 = \{x \in \mathbb{R}^N\,;\ \langle x,y\rangle \le 1$ for all y in $T\}$ is the polar of T; more precisely, for some numerical constant K,

$$\sup_{\varepsilon>0}\varepsilon(\log N(B_2,\varepsilon T^0))^{1/2} \le K\ell(T)\,. \tag{3.15}$$

The proof of (3.15) is rather simple, simpler than the proof of Sudakov's minoration. Let $a = 2\ell(T)$. Then

$$\gamma_N(aT^0) = \mathbb{P}\{\sup_{t\in T}|X_t| \le a\} \ge \frac{1}{2}$$

where γ_N is the canonical Gaussian measure on $\mathbb{R}^N$. Let now $\varepsilon > 0$ and n be such that

$$N(B_2,\varepsilon T^0) = N(\frac{a}{\varepsilon}B_2, aT^0) \ge n\,.$$

There exist $z_1,\ldots,z_n$ in $\frac{a}{\varepsilon}B_2$ such that, for all $i \ne j$, $(z_i + aT^0)\cap(z_j + aT^0) = \emptyset$. Hence,

$$1 \ge \gamma_N\left(\bigcup_{i=1}^n (z_i + aT^0)\right) = \sum_{i=1}^n \gamma_N(z_i + aT^0)\,.$$

For any z in $\mathbb{R}^N$, a change of variables indicates that

$$\gamma_N(z + aT^0) = \exp(-|z|^2/2)\int_{aT^0}\exp\langle z,x\rangle d\gamma_N(x)\,,$$

and thus, by Jensen's inequality and symmetry of T^0,

$$\gamma_N(z + aT^0) \ge \exp(-|z|^2/2)\gamma_N(aT^0)\,.$$

Therefore, since $z_i \in \frac{a}{\varepsilon}B_2$, $i \in N$, and $\gamma_N(aT^0) \ge 1/2$, we finally get that $2 \ge n\exp(-a^2/2\varepsilon^2)$ which is exactly (3.15).

It is worthwhile noting that Theorem 3.18 and its dual version (3.15) can easily be shown to be equivalent. Let us sketch for example how Sudakov's minoration can be deduced from (3.15) using a simple duality argument. Observe first that for every $\varepsilon > 0$, $2T \cap (\frac{\varepsilon^2}{2}T^0) \subset \varepsilon B_2$. Indeed, if $t \in 2T$ and $t \in \frac{\varepsilon^2}{2}T^0$,

$$|t|^2 = \langle t, t\rangle \le \|t\|_T \|t\|_{T^0} \le 2 \cdot \frac{\varepsilon^2}{2} = \varepsilon^2 ,$$

where $\|\cdot\|_T$ (respectively $\|\cdot\|_{T^0}$) is the Banach norm (gauge) induced by T (T^0). It follows that

$$N(T, \varepsilon B_2) \le N(T, 2T \cap (\frac{\varepsilon^2}{2}T^0)) = N(T, \frac{\varepsilon^2}{2}T^0) .$$

By homogeneity, and elementary properties of entropy numbers

$$\begin{aligned} N(T, \frac{\varepsilon^2}{2}T^0) &\le N(T, 2\varepsilon B_2) N(2\varepsilon B_2, \frac{\varepsilon^2}{2}T^0) \\ &\le N(T, 2\varepsilon B_2) N(B_2, \frac{\varepsilon}{4}T^0) . \end{aligned}$$

Thus, for every $\varepsilon > 0$,

$$\varepsilon(\log N(T, \varepsilon B_2))^{1/2} \le \varepsilon(\log N(T, 2\varepsilon B_2))^{1/2} + 4M$$

where $M = \sup_{\varepsilon > 0} \varepsilon(\log N(B_2, \varepsilon T^0))^{1/2}$. One then easily deduces that

$$\sup_{\varepsilon > 0} \varepsilon(\log N(T, \varepsilon B_2))^{1/2} \le 8M . \tag{3.16}$$

The converse inequality may be shown similarly. By duality, $B_2 \subset \mathrm{Conv}(\frac{\varepsilon}{2}T^0, \frac{2}{\varepsilon}T) \subset \frac{\varepsilon}{2}T^0 + \frac{2}{\varepsilon}T$. Then

$$\begin{aligned} N(B_2, \varepsilon T^0) &\le N(\frac{\varepsilon}{2}T^0 + \frac{2}{\varepsilon}T, \varepsilon T^0) \\ &= N(\frac{2}{\varepsilon}T, \varepsilon T^0) \\ &\le N(\frac{2}{\varepsilon}T, \frac{1}{4}B_2) N(\frac{1}{4}B_2, \frac{\varepsilon}{2}T^0) \end{aligned}$$

and we can conclude as before that

$$\sup_{\varepsilon > 0} \varepsilon(\log N(B_2, \varepsilon T^0))^{1/2} \le 4M' \tag{3.17}$$

where $M' = \sup_{\varepsilon > 0} \varepsilon(\log N(T, \varepsilon B_2))^{1/2}$.

The last application of the comparison theorems that we present concern tensor products of Gaussian measures. For simplicity, we only deal with Radon random variables. Let E and F be two Banach spaces. If $x \in E$ and $y \in F$, $x \otimes y$ is the bilinear form on $E' \times F'$ which maps $(f, h) \in E' \times F'$ into $f(x)h(y)$. The linear tensor product $E \otimes F$ consists of all finite sums $u = \sum_i x_i \otimes y_i$ with $x_i \in E$, $y_i \in F$. On $E \otimes F$, consider the injective tensor product norm

$$\|u\|_\vee = \sup\left\{ \left| \sum_i f(x_i)h(y_i) \right| \; ; \; \|f\| \leq 1 \, , \; \|h\| \leq 1 \right\} ,$$

i.e. the norm of u as a bounded bilinear form on $E' \times F'$. The completion of $E \otimes F$ with respect to this norm is called the injective tensor product of E and F and denoted by $E \check{\otimes} F$.

Consider now $X = \sum_i g_i x_i$ (resp. $Y = \sum_j g_j y_j$) a Gaussian random variable with values in E (resp. F). Here (g_i) denotes as usual an orthogaussian sequence. Let further $(g_{i,j})$ be a doubly-indexed orthogaussian sequence. Given convergent series X and Y like this with values in E and F respectively, one might wonder whether

$$\sum_{i,j} g_{i,j} x_i \otimes y_j$$

is almost surely convergent in the injective tensor product space $E \check{\otimes} F$. This question has a positive answer and this is the conclusion of the next theorem. Recall that $\sigma(X) = \sup_{\|f\| \leq 1} (\mathbb{E} f^2(X))^{1/2} = \sup_{\|f\| \leq 1} (\sum_i f^2(x_i))^{1/2}$, $\sigma(Y)$ being defined similarly.

Theorem 3.20. Let $X = \sum_i g_i x_i$ and $Y = \sum_j g_j y_j$ be convergent Gaussian series with values in E and F respectively and with corresponding $\sigma(X)$ and $\sigma(Y)$. Then $G = \sum_{i,j} g_{i,j} x_i \otimes y_j$ is almost surely convergent in the injective tensor product $E \check{\otimes} F$ and the following inequality holds:

$$\max(\sigma(X)\mathbb{E}\|Y\| \, , \; \sigma(Y)\mathbb{E}\|X\|) \leq \mathbb{E}\|G\|_\vee \leq \sigma(X)\mathbb{E}\|Y\| + \sigma(Y)\mathbb{E}\|X\| \, .$$

Proof. To prove that G is convergent it is enough to establish the right side of the inequality of the theorem for finite sums and use a limiting argument. By definition of the tensor product norm, the left side is easy; note that it indicates in the same way that the convergence of G implies the convergence of X and Y. In the sequel we therefore only deal with finite sequences (x_i) and (y_j). The idea of the proof is to compare G, considered as a Gaussian process indexed by the product of the unit balls of E' and F', to another built as a kind of (independent) "sum" of X and Y. Consider namely the Gaussian process $\widetilde{G}$ indexed by $E' \times F'$:

$$\widetilde{G}(f,h) = \sigma(X)\sum_j g'_j h(y_j) + (\sum_j h^2(y_j))^{1/2} \sum_i g_i f(x_i)\,, \qquad f \in E'\,,\ h \in F'$$

where (g_i), (g'_j) are independent orthogaussian sequences. Actually, rather than to compare G to $\widetilde{G}$ it is convenient to replace G by

$$\widehat{G}(f,h) = \sum_{i,j} g_{i,j} f(x_i)h(y_j) + \sigma(X)(\sum_j h^2(y_j))^{1/2} g\,, \qquad f \in E'\,,\ h \in F'$$

where g is a standard normal variable independent of $(g_{i,j})$. Clearly, by Jensen's inequality and independence,

$$\mathbb{E}\|G\|_\vee \le \mathbb{E} \sup_{\|f\|=1} \sup_{\|h\|=1} \widehat{G}(f,h)\,.$$

The reason for the introduction of $\widehat{G}$ is that we will use Corollary 3.12 where we need special information on the diagonal of the covariance structures, something which is given by $\widehat{G}$. Indeed, it is easily verified that for f, $f' \in E'$, h, $h' \in F'$,

$$\begin{aligned} &\mathbb{E}\widehat{G}(f,h)\widehat{G}(f',h') - \mathbb{E}\widetilde{G}(f,h)\widetilde{G}(f',h') \\ &\qquad = (\sigma(X)^2 - \sum_i f(x_i)f'(x_i))[(\sum_j h^2(y_j))^{1/2}(\sum_j h'^{\,2}(y_j))^{1/2} - \sum_j h(y_j)h'(y_j)]\,. \end{aligned}$$

Hence this difference is always positive and is equal to 0 when $h = h'$. We are thus in a position to apply Corollary 3.12 to the Gaussian processes $\widehat{G}$ and $\widetilde{G}$. After an approximation and compactness argument, it implies that

$$\mathbb{E} \sup_{\|h\|=1} \sup_{\|f\|=1} \widehat{G}(f,h) \le \mathbb{E} \sup_{\|h\|=1} \sup_{\|f\|=1} \widetilde{G}(f,h)\,.$$

To get the inequality of the theorem we need simply note that

$$\mathbb{E} \sup_{\|h\|=1} \sup_{\|f\|=1} \widetilde{G}(f,h) \le \sigma(X)\mathbb{E}\|Y\| + \sigma(Y)\mathbb{E}\|X\|\,.$$

The proof of Theorem 3.20 is therefore complete. Notice that Theorem 3.15 yields a somewhat simplified proof.

As a consequence of Corollary 3.13, we also have that

$$\mathbb{E} \inf_{\|h\|=1} \sup_{\|f\|=1} \widetilde{G}(f,h) \leq \mathbb{E} \inf_{\|h\|=1} \sup_{\|f\|=1} \widehat{G}(f,h) .$$

Now notice that

$$\mathbb{E} \inf_{\|h\|=1} \sup_{\|f\|=1} \widetilde{G}(f,h) = \sigma(X)\mathbb{E} \inf_{\|h\|=1} h(Y) + \inf_{\|h\|=1} (\sum_j h^2(y_j))^{1/2}\mathbb{E}\|X\|$$

so that we have the following lower bound:

$$\mathbb{E} \inf_{\|h\|=1} \sup_{\|f\|=1} \widehat{G}(f,h) \geq \sigma(X)\mathbb{E} \inf_{\|h\|=1} h(Y) + \inf_{\|h\|=1} (\sum_j h^2(y_j))^{1/2}\mathbb{E}\|X\| . \tag{3.18}$$

This inequality has some interesting consequences together with the one in Theorem 3.20. One application is the following corollary which will be of interest in Chapter 9.

Corollary 3.21. Let X be a Gaussian Radon random variable with values in a Banach space B. Let also (X_i) be independent copies of X. Then, for every N, if $\alpha = (\alpha_1, \ldots, \alpha_N)$ is a generic point in $\mathbb{R}^N$,

$$\mathbb{E} \sup_{|\alpha|=1} \left\| \sum_{i=1}^N \alpha_i X_i \right\| \leq \mathbb{E}\|X\| + \sigma(X)\sqrt{N}$$

and

$$\mathbb{E} \inf_{|\alpha|=1} \left\| \sum_{i=1}^N \alpha_i X_i \right\| \geq \mathbb{E}\|X\| - \sigma(X)\sqrt{N} .$$

Proof. X can be represented as $X = \sum_i g_i x_i$ for some sequence (x_i) in B. Consider

$$Y = \sum_{j=1}^N g_j e_j$$

where (e_j) is the canonical basis of ℓ_2^N. Then Theorem 3.20 in the tensor space $B \check{\otimes} \ell_2^N$ immediately yields

$$\mathbb{E} \sup_{|\alpha|=1} \left\| \sum_{i=1}^N \alpha_i X_i \right\| \leq \mathbb{E}\|X\| + \sigma(X)\mathbb{E} \left(\left(\sum_{j=1}^N g_j^2 \right)^{1/2} \right)$$

and thus the first inequality of the corollary follows since obviously $\mathbb{E}((\sum_{j=1}^N g_j^2)^{1/2}) \leq \sqrt{N}$. For the second use (3.18); in this case indeed

$$\mathbb{E} \inf_{|\alpha|=1} \sup_{\|f\|=1} \left(\sum_{i,j} g_{i,j}\alpha_j f(x_i) + \sigma(X) \left(\sum_j \alpha_j^2 \right)^{1/2} g \right) = \mathbb{E} \inf_{|\alpha|=1} \sup_{\|f\|=1} \left(\sum_{j=1}^N \alpha_j f(X_j) + \sigma(X) g \right)$$

$$= \mathbb{E} \inf_{|\alpha|=1} \left\| \sum_{j=1}^N \alpha_j X_j \right\|$$

and thus

$$\mathbb{E} \inf_{|\alpha|=1} \left\| \sum_{j=1}^N \alpha_j X_j \right\| \geq \mathbb{E}\|X\| - \sigma(X)\mathbb{E}\left(\left(\sum_{j=1}^N g_j^2 \right)^{1/2} \right) .$$

The proof is therefore complete.

Notes and references

Some general references on Gaussian processes and measures (for both this chapter and Chapters 11 and 12) are [Ne2], [B-C], [Fer4], [Ku], [Su4], [J-M3], [Fer9], [V-T-C], [Ad], [K-L]. The interested reader may find there (completed with the papers [Bo3], [Ta1]) various topics on Gaussian measures on abstract spaces, like for example zero-one laws, reproducing kernel Hilbert space, etc., not developed in this book.

The history of integrability properties of norms of infinite dimensional Gaussian random vectors starts with the papers [L-S] and [Fer2]. Fernique's simple argument is the one leading to (3.4) and applies to the rather general setting of measurable seminorms. The proof by H. J. Landau and L. A. Shepp is isoperimetric and led eventually to the Gaussian isoperimetric inequality. Skorokhod [Sk2] had an argument to show that $\mathbb{E}\exp \alpha\|X\| < \infty$ (using the strong Markov property of Brownian motion); J. Hoffmann-Jørgensen indicates in [HJ3] a way to get from this partial conclusion the usual exponential square integrability. Corollary 3.2 is due to M. B. Marcus and L. A. Shepp [M-S2]. Our description of the integrability properties and tail behavior is isoperimetric and follows C. Borell [Bo2] (and the exposition of [Eh4]). The concentration in Lemma 3.1 is issued from Chapter 1 . Theorem 3.3 was established in [Ta2] where examples describing its optimality are given. In particular, $\Phi^{-1}(\mathbb{P}\{\|X\| \leq t\})$ approaches its asymptote as slowly as one wishes it. Lemma 3.5 in a slight improved form was used in [Go]; see also [G-K1], [G-K2]. Theorem 3.3 has some interpretation in large deviations; indeed, the limit in Corollary 3.2

$$\lim_{t\to\infty} \frac{1}{t^2} \log \mathbb{P}\{\|X\| > t\} = -\frac{1}{2\sigma^2}$$

is of course a large deviation result for complements of balls centered at the origin. Theorem 3.3 improves this limit into

$$\lim_{t\to\infty}\left(\frac{1}{t}\log \mathbb{P}\{\|X\|>t\}+\frac{t}{2\sigma^2}\right)=0$$

(for Radon variables) which appears as a "normal" deviation result for complements of balls; similar results for different sets might hold as well (on large deviations, see e.g. [Az], [Str], [Ja3], ...). That $\tau=\tau(X)=0$ for a Gaussian Radon measure was recorded in [D-HJ-S] and that τ is the unique jump of the distribution of $\|X\|$ is due to B. S. Tsirelson [Ts].

Homogeneous chaos were introduced by N. Wiener [Wie] and are presented, e.g., in [Ne2]. Their order of integrability was first investigated in [Schr] and [Var]. Hypercontractivity of the Hermite semigroup has been discovered by E. Nelson [Nel]. L. Gross [Gro] translated this property into a logarithmic Sobolev inequality and uses a two point inequality and the central limit theorem to provide an alternate proof (see also Chapter 4 and cf. [Bec] for further deep results in Fourier Analysis along these lines). The relevance of hypercontractivity to integrability of Gaussian chaos (and its extension to the vector valued case) was noticed by C. Borell [Bo4], [Bo6]; the deep work [Bo4] however develops the isoperimetric approach that we closely follow here (and that is further developed in [Bo7]). The introduction of decoupled chaos is motivated by [Kw4] (following [Bo6], [Bo7]). Theorem 3.10 is perhaps new.

Inequality (3.11) with its best constant 1 (for C symmetric) is due to T. W. Anderson [An]. Slepian's Lemma appeared in [Sl]. Its geometric meaning makes it probably more ancient as was noted by several authors [Su1], [Su4], [Gr2]. Related to this lemma, let us mention its "two-sided" analogue studied by Z. Šidák [Si1], [Si2] which expresses that if $X=(X_1,\ldots,X_N)$ is a Gaussian vector in $\mathbb{R}^N$, for any positive numbers λ_i, $i\le N$,

$$\mathbb{P}\left\{\bigcup_{i=1}^{N}(|X_i|\le\lambda_i)\right\}\ge\prod_{i=1}^{N}\mathbb{P}\{|X_i|\le\lambda_i\}\,.$$

We refer to [To] for more inequalities on Gaussian distributions in finite dimension. Slepian's lemma was first used in the study of Gaussian processes in [Su1], [M-S1] and [M-S2] where Corollary 3.14 is established. Theorem 3.15 was announced by V. N. Sudakov in [Su2] (see also [Su4]) and established in this form by X. Fernique [Fer4]; credit is also due to S. Chevet (unpublished). Y. Gordon [Gor1], [Gor2] discovered Corollary 3.13 and Theorem 3.16 motivated by Dvoretzky's theorem (cf. Chapter 9). [Gor3] contains a more general and simplified proof of Theorem 3.16 with applications. Our exposition of the inequalities by Slepian and Gordon is based on Theorem 3.11 of J.-P. Kahane [Ka2]. Sudakov's minoration was observed in [Su1], [Su3].

Its dual version (3.15) appeared in the context of local theory of Banach spaces and duality of entropy numbers [P-TJ]. The consideration of $\ell(T)$ (with this notation) goes back to [L-P]. [The simple proof of (3.15) presented here is due to the second author. This proof was communicated in particular to the author of [Go] (where a probabilistic application is obtained) who gives a strickingly creative acknowledgement of the fact. Further applications of the method are presented in [Ta19].] The equivalence between Sudakov's minoration and its dual version ((3.16) and (3.17)) is due to N. Tomczak-Jaegermann [TJ1] (and her argument actually shows a closer relationshiop between the entropy numbers and the dual entropy numbers, cf. [TJ1]; this will partially be used below in Section 15.5). Tensor products of Gaussian measures were initiated by S. Chevet [Ch1], [Ch2]; see also [Car]. The best constants in Theorem 3.20 and inequality (3.18) follow from [Gor1], [Gor2] from where Corollary 3.21 is also taken.

Chapter 4. Rademacher averages

Chapter 4. Rademacher averages

This chapter is devoted to Rademacher averages $\sum_i \varepsilon_i x_i$ with vector valued coefficients as a natural analog of the Gaussian averages $\sum_i g_i x_i$. The properties we examine are entirely similar to the ones investigated in the Gaussian case. We will see in this way how isoperimetric methods can be used to yield strong integrability properties of convergent Rademacher series and chaos. This is studied in Sections 4.3 and 4.4. Some comparison results are also available in the form, for example, of a version of Sudakov's minoration presented in Section 4.5. We however start in the first two sections with some basic facts on Rademacher averages with real coefficients as well as on the so-called contraction principle, a most valuable tool in Probability in Banach spaces.

We thus assume we are given on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ a sequence (ε_i) of independent random variables taking the values ± 1 with probability $1/2$, that is symmetric Bernoulli or Rademacher random variables. We usually call (ε_i) a Rademacher sequence. If (ε_i) is considered alone, one might take, as a concrete example, Ω to be the Cantor group $\{-1,+1\}^{\mathbb{N}}$, $\mathbb{P}$ its canonical product probability measure (Haar measure) $\mu = \left(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}\right)^{\otimes \mathbb{N}}$ and ε_i the coordinate maps. We thus investigate finite or convergent sums $\sum_i \varepsilon_i x_i$ with vector valued coefficients x_i . As announced the first paragraph is devoted to some preliminaries in the real case.

4.1. Real Rademacher averages

If (α_i) is a sequence of real numbers, a trivial application of the three series theorem (or Lemma 4.2 below) indicates that the series $\sum_i \varepsilon_i \alpha_i$ is almost surely (or in probability) convergent if and only if $\sum_i \alpha_i^2 < \infty$. Actually the sum $\sum_i \varepsilon_i \alpha_i$ has remarkable properties in connection with the sum of the squares $\sum_i \alpha_i^2$ and it is the purpose of this paragraph to recall some of these.

Since we will only be interested in estimates which easily extend to infinite sums, there is no loss in generality to assume, as is usual, for simplicity in the exposition, that we deal with finite sequences (α_i) , i.e. finitely many α_i 's only are non-zero.

The first main observation is the classical *subgaussian* estimate which draws its name from the Gaussian type tail. We can obtain it as a consequence of Lemma 1.5 since $\sum_i \varepsilon_i \alpha_i$ clearly defines a mean zero sum of martingale differences or directly by the same argument: indeed, given thus a finite sequence (α_i) of real numbers, for all $\lambda > 0$,

$$\mathbb{E}\exp\left(\lambda\sum_i \varepsilon_i\alpha_i\right) = \prod_i \mathbb{E}\exp(\lambda\varepsilon_i\alpha_i) \le \prod_i \exp\left(\frac{\lambda^2}{2}\alpha_i^2\right) = \exp\left(\frac{\lambda^2}{2}\sum_i \alpha_i^2\right),$$

and hence, by Chebyshev's inequality, for every $t > 0$,

$$\mathbb{P}\left\{\left|\sum_i \varepsilon_i\alpha_i\right| > t\right\} \le 2\exp\left(-t^2/2\sum_i \alpha_i^2\right). \tag{4.1}$$

In particular, a convergent Rademacher series $\sum_i \varepsilon_i\alpha_i$ satisfy exponential squared integrability properties exactly as Gaussian variables.

This simple inequality (4.1) is extremely useful. It is moreover sharp in various instances and we have in particular the following converse which we record at this stage for further purposes: there is a numerical constant $K \ge 1$ such that if (α_i) and t satisfy $t \ge K(\sum_i \alpha_i^2)^{1/2}$ and $t\max_i |\alpha_i| \le K^{-1}\sum_i \alpha_i^2$, then

$$\mathbb{P}\{\sum_i \varepsilon_i\alpha_i > t\} \ge \exp(-Kt^2/\sum_i \alpha_i^2). \tag{4.2}$$

This inequality, actually in a more precise form concerning the choice of the constants, can be deduced for example from the more general Kolmogorov minoration inequality given below as Lemma 8.1. Let us however give a direct proof of (4.2). Assume that $(\alpha_i)_{i\ge1}$ is such that, by homogeneity, $\sum_i \alpha_i^2 = 1$ and that $t \ge 2$, $|\alpha_i| \le 1/16t$ for all i. Define n_j, $j \le k$ ($n_0 = 0$) by

$$n_j = \inf\left\{n > n_{j-1}\,;\ \sum_{i=n_{j-1}+1}^{n} \alpha_i^2 > \frac{1}{16^2t^2}\right\}.$$

Since $\sum_i \alpha_i^2 = 1$, $k \le 16^2t^2$. On the other hand, for each $j \le k$,

$$\sum_{i=n_{j-1}+1}^{n_j} \alpha_i^2 \le \sum_{i=n_{j-1}+1}^{n_j-1} \alpha_i^2 + \frac{1}{16^2t^2} \le \frac{2}{16^2t^2}$$

so that $k \ge 4\cdot 16t^2$ since k being the last one means that

$$\sum_{i>n_k} \alpha_i^2 \le \frac{1}{16^2t^2} \le \frac{1}{2}.$$

Set $I_j = \{n_{j-1}+1, \ldots, n_j\}$, $1 \le j \le k$. We can then write by independence

$$\begin{aligned} \mathbb{P}\left\{\sum_i \varepsilon_i \alpha_i > t\right\} &\ge \prod_{j \le k} \mathbb{P}\left\{\sum_{i \in I_j} \varepsilon_i \alpha_i > \frac{1}{4 \cdot 16t}\right\} \\ &\ge \prod_{j \le k} \mathbb{P}\left\{\sum_{i \in I_j} \varepsilon_i \alpha_i > \frac{1}{4}\left(\sum_{i \in I_j} \alpha_i^2\right)^{1/2}\right\} . \end{aligned}$$

Now, (4.3) and Lemma 4.2 below indicate together that

$$\mathbb{P}\left\{\sum_i \varepsilon_i \beta_i > \frac{1}{4}\left(\sum_i \beta_i^2\right)^{1/2}\right\} \ge \frac{1}{16} .$$

It follows that

$$\mathbb{P}\left\{\sum_i \varepsilon_i \alpha_i > t\right\} \ge \left(\frac{1}{16}\right)^k \ge \exp(-16^2 t^2 \log 16)$$

which is the result (with e.g. $K = 16^2 \log 16$).

The subgaussian inequality (4.1) can be used to yield a simple proof of the classical Khintchine inequalities.

Lemma 4.1. For any $0 < p < \infty$, there exist positive finite constants A_p and B_p depending on p only such that for any finite sequence (α_i) of real numbers

$$A_p\left(\sum_i \alpha_i^2\right)^{1/2} \le \left\|\sum_i \varepsilon_i \ alpha_i\right\|_p \le B_p\left(\sum_i \alpha_i^2\right)^{1/2} .$$

Proof. By homogeneity assume that $\sum_i \alpha_i^2 = 1$. Then, by the integration by parts formula and (4.1),

$$\begin{aligned} \mathbb{E}\left|\sum_i \varepsilon_i \alpha_i\right|^p &= \int_0^\infty \mathbb{P}\left\{\left|\sum_i \varepsilon_i \alpha_i\right| > t\right\} dt^p \\ &\le 2\int_0^\infty \exp(-t^2/2)\, dt^p = B_p^p . \end{aligned}$$

For the left hand side inequality, it is enough, by Jensen's inequality, to consider the case $p < 2$. By mean of Hölder's inequality, we get

$$
\begin{aligned}
1 = \mathbb{E}\left|\sum_i \varepsilon_i \alpha_i\right|^2 &= \mathbb{E}\left(\left|\sum_i \varepsilon_i \alpha_i\right|^{2p/3}\left|\sum_i \varepsilon_i \alpha_i\right|^{2-2p/3}\right) \\
&\leq \left(\mathbb{E}\left|\sum_i \varepsilon_i \alpha_i\right|^p\right)^{2/3}\left(\mathbb{E}\left|\sum_i \varepsilon_i \alpha_i\right|^{6-2p}\right)^{1/3} \\
&\leq \left(\mathbb{E}\left|\sum_i \varepsilon_i \alpha_i\right|^p\right)^{2/3} B_{6-2p}^{2-2p/3}
\end{aligned}
$$

from which the conclusion follows.

The best possible constants A_p and B_p in Khinchine's inequalities are known [Ha]. We retain from the preceding proof that $B_p \leq K\sqrt{p}$ ($p \geq 1$) for some numerical constant K. We will use also the known fact [Sz] that $A_1 = 2^{-1/2}$ (in order to deal with a specific value), i.e.

$$
\left(\sum_i \alpha_i^2\right)^{1/2} = \left\|\sum_i \varepsilon_i \alpha_i\right\|_2 \leq \sqrt{2}\left\|\sum_i \varepsilon_i \alpha_i\right\|_1 . \tag{4.3}
$$

Khintchine's inequalities show how the Rademacher sequence defines a basic unconditional sequence and spans ℓ_2 in the spaces L_p, $0 < p < \infty$. We could also add in a sense $p = 0$ in this claim as is shown by the simple (but useful) following lemma. The interest of this lemma goes beyond this application and it will be mentioned many times throughout this book.

Lemma 4.2. Let Z be a positive random variable such that for some $q > p > 0$ and some constant C

$$
\|Z\|_q \leq C\|Z\|_p .
$$

Then, if $t > 0$ is such that $\mathbb{P}\{Z > t\} \leq (2C^p)^{q/(p-q)}$, we have

$$
\|Z\|_p \leq 2^{1/p} t \qquad \text{and} \qquad \|Z\|_q \leq 2^{1/p} C t .
$$

Proof. By Hölder's inequality

$$
\mathbb{E}Z^p \leq t^p + \int_{\{Z>t\}} Z^p d\mathbb{P} \leq t^p + \|Z\|_q^p (\mathbb{P}\{Z > t\})^{1-p/q} \leq 2t^p
$$

where the last inequality is obtained from the choice of t.

Note, as a consequence of this lemma, that if (X_n) is a sequence of random variables (real or vector valued) such that for some $q > p > 0$ and $C > 0$, $\|X_n\|_q \le C\|X_n\|_p$ for all n, and if (X_n) converges in probability to some variable X, then, since $\sup_n \|X_n\|_q < \infty$ by Lemma 4.2, (X_n) also converges to X in $L_{q'}$ for all $q' < q$.

While the subspace generated by (ε_i) in L_p for $0 \le p < \infty$ is ℓ_2, in L_∞ however this subspace is isometric to ℓ_1. Indeed, for any finite sequence (α_i) of real numbers, there exists a choice of signs $\varepsilon_i = \pm 1$ such that $\varepsilon_i \alpha_i = |\alpha_i|$ for all i. Hence

$$\left\| \sum_i \varepsilon_i \alpha_i \right\|_\infty = \sum_i |\alpha_i| .$$

It might therefore be of some interest to try to have an idea of the span of the Rademacher sequence in spaces "between" L_p for $p < \infty$ and L_∞. Among the possible intermediary spaces we may consider Orlicz spaces with exponential rates. A Young function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is convex, increasing with $\lim_{t\to\infty} \psi(T) = \infty$ and $\psi(0) = 0$. Denote by $L_\psi = L_\psi(\Omega, \mathcal{A}, \mathbb{P})$ the Orlicz space of all real random variables X (defined on $(\Omega, \mathcal{A}, \mathbb{P})$) such that $\mathbb{E}\psi(|X|/c) < \infty$ some $c > 0$. Equipped with the norm

$$\|X\|_\psi = \inf\{c > 0 \, ; \; \mathbb{E}\psi(|X|/c) \le 1\} ,$$

L_ψ defines a Banach space. If for example $\psi(x) = x^p$, $1 \le p < \infty$, then $L_\psi = L_p$. We shall be interested here in the exponential functions

$$\psi_q(x) = \exp(x^q) - 1 , \qquad 1 \le q < \infty .$$

To handle the small convexity problem when $0 < q < 1$, we can let $\psi_q(x) = \exp(x^q) - 1$ for $x \ge x(q)$ large enough, and take ψ_q to be linear on $[0, x(q)]$.

The first observation is that (ε_i) still spans a subspace isomorphic to ℓ_2 in ψ_q whenever $q \le 2$. We assume that $1 \le q \le 2$ for simplicity, but the proof is similar when $0 < q < 1$. Letting $X = \sum_i \varepsilon_i \alpha_i$ where (α_i) is a finite sequence of real numbers such that, by homogeneity, $\sum_i \alpha_i^2 = 1$, we have, by integration by parts and (4.1),

$$\begin{aligned} \mathbb{E}\psi_q\left(\frac{|X|}{c}\right) &= \int_0^\infty \mathbb{P}\{|X| > ct\} d(e^{t^q} - 1) \\ &\le 2q \int_0^\infty \exp\left(-\frac{c^2 t^2}{2} + t^q\right) t^{q-1} dt . \end{aligned}$$

Hence, when $q \leq 2$ and $c \geq B'_q$ large enough, $\mathbb{E}\psi_q(|X|/c) \leq 1$ so that $\|X\|_{\psi_q} \leq B'_q$. On the other hand, since $e^x - 1 \geq x$,

$$\mathbb{E}\psi_q\left(\frac{|X|}{c}\right) \geq \mathbb{E}\left(\frac{|X|^q}{c^q}\right) \geq \left(\frac{A_q}{c}\right)^q > 1$$

for $c \leq A'_q$ small enough so that $\|X\|_{\psi_q} \geq A'_q$. Hence, as claimed, when $q \leq 2$,

$$A'_q\left(\sum_i \alpha_i^2\right)^{1/2} \leq \left\|\sum_i \varepsilon_i \alpha_i\right\|_{\psi_q} \leq B'_q\left(\sum_i \alpha_i^2\right)^{1/2}$$

for any sequence (α_i) of real numbers.

For $q > 2$, the span of (ε_i) in L_{ψ_q} is no more isomorphic to ℓ_2. This span actually appears as some interpolation space between ℓ_2 and ℓ_1. To see it, we simply follow the observation which led to Lemma 1.7. Recall that for $0 < p < \infty$ we denote by $\ell_{p,\infty}$ the space of all real sequences $(\alpha_i)_{i \geq 1}$ such that

$$\|(\alpha_i)\|_{p,\infty} = (\sup_{t>0} t^p \text{ Card}\{i\,;\ |\alpha_i| > t\})^{1/p} < \infty\,.$$

Equivalently $\|(\alpha_i)\|_{p,\infty} = \sup_{i \geq 1} i^{1/p}\alpha_i^*$ where (α_i^*) is the non-increasing rearrangement of the sequence $(|\alpha_i|)$. These spaces are known to show up in interpolations of ℓ_p-spaces. The functional $\|\cdot\|_{p,\infty}$, equivalent to a norm when $p > 1$, may be compared to the ℓ_p-norms as follows: for $r < p$,

$$\|(\alpha_i)\|_{p,\infty} \leq \left(\sum_i |\alpha_i|^p\right)^{1/p} \leq \left(\frac{p}{p-r}\right)^{1/p} \|(\alpha_i)\|_{r,\infty}\,.$$

Let now $2 < q < \infty$ and denote by p the conjugate of $q : 1/p + 1/q = 1$, $1 < p < \infty$. The next lemma describes how in this case the span of (ε_i) in L_{ψ_q} is isomorphic to $\ell_{p,\infty}$.

Lemma 4.3. Let $2 < q < \infty$ and $p = q/q-1$. There exist positive finite constants A'_q, B'_q depending only on q (p) such that for any finite sequence (α_i) of real numbers

$$A'_q\|(\alpha_i)\|_{p,\infty} \leq \left\|\sum_i \varepsilon_i \alpha_i\right\|_{\psi_q} \leq B'_q\|(\alpha_i)\|_{p,\infty}\,.$$

Proof. By symmetry $(\varepsilon_i\alpha_i)$ has the same distribution as $(\varepsilon_i|\alpha_i|)$. Further, by identical distribution and definition of $\|(\alpha_i)\|_{p,\infty}$ we may assume that $|\alpha_1| \geq \cdots \geq |\alpha_i| \geq \cdots$. The martingale inequality of Lemma 1.7 then yields, for every $t > 0$,

$$\mathbb{P}\left\{\left|\sum_i \varepsilon_i\alpha_i\right| > t\right\} \leq 2\exp(-t^q/C_q\|(\alpha_i)\|_{p,\infty}^q)\,. \tag{4.4}$$

As before, one then deduces the right side of the inequality of Lemma 4.3 from a simple integration by parts. Turning to the left side, we make use of the contraction principle in the form of Theorem 4.4 below. It implies indeed, by symmetry and monotonicity of $(|\alpha_i|)$, that for every m

$$\mathbb{E}\exp\left|\sum_i \varepsilon_i \alpha_i\right|^q \geq \mathbb{E}\exp\left(vert\alpha_m|^q \left|\sum_{i=1}^m \varepsilon_i\right|^q\right).$$

Hence

$$\left\|\sum_i \varepsilon_i \alpha_i\right\|_{\psi_q} \geq |\alpha_m \ vert \left\|\sum_{i=1}^m \varepsilon_i\right\|_{\psi_q}.$$

Now, since $\sum_{i=1}^m \varepsilon_i = m$ with probability 2^{-m}, it is easily seen that

$$\left\|\sum_{i=1}^m \varepsilon_i\right\|_{\psi_q} \geq (1+\log 2)^{-1/q} m^{1/p}.$$

Summarizing, we have obtained that

$$\left\|\sum_i \varepsilon_i \alpha_i\right\|_{\psi_q} \geq (1+\log 2)^{-1/q} \sup_{m\geq 1} m^{1/p}|\alpha_m|$$

which is the result.

We should point out further that the inequality corresponding to Lemma 1.8 states in this setting that for any finite sequence (α_i) and any $t>0$

$$\mathbb{P}\left\{\left|\sum_i \varepsilon_i \alpha_i\right| > t\right\} \leq 16\exp[-\exp(t/4\|(\alpha_i)\|_{1,\infty})]. \tag{4.5}$$

The rest of this chapter is mainly devoted to extensions of the previous classical results to Rademacher averages with Banach space valued coefficients. Of course, this vector valued setting is characterized by the lack of the *orthogonality* property

$$\mathbb{E}\left|\sum_i \varepsilon_i \alpha_i\right|^2 = \sum_i \alpha_i^2.$$

Various substitutes have therefore to be investigated but the extension program will basically be fulfilled. Classification of Banach spaces according to the preceding orthogonality property is at the origin of the notions of type and cotype of Banach spaces which will be discussed later on in Chapter 9. We study here integrability and comparison theorems for Rademacher averages with vector valued coefficients.

4.2. The contraction principle

It is plain from Khintchine's inequalities that if (α_i) and (β_i) are two sequences of real numbers such that $|\alpha_i| \leq |\beta_i|$ for all i, one can compare $\|\sum_i \varepsilon_i \alpha_i\|_p$ and $\|\sum_i \varepsilon_i \beta_i\|_p$ for all p. This comparison is thus based on the sum of the squares and orthogonality. It however extends to vector valued coefficients and even in an improved form. This property is known as the *contraction principle* to which this paragraph is devoted. The main result is expressed in the following fundamental theorem.

Theorem 4.4. Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex. For any finite sequence (x_i) in a Banach space B and any real numbers (α_i) such that $|\alpha_i| \leq 1$ for all i, we have

$$\mathbb{E}F\left(\left\|\sum_i \alpha_i \varepsilon_i x_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_i \varepsilon_i x_i\right\|\right) . \tag{4.6}$$

Further, for any $t > 0$,

$$\mathbb{P}\left\{\left\|\sum_i \alpha_i \varepsilon_i x_i\right\| > t\right\} \leq 2\mathbb{P}\left\{\left\|\sum_i \varepsilon_i x_i\right\| > t\right\} . \tag{4.7}$$

Proof. The function on $\mathbb{R}^N$

$$(\alpha_1, \ldots, \alpha_N) \to \mathbb{E}F\left(\left\|\sum_{i=1}^N \alpha_i \varepsilon_i x_i\right\|\right)$$

is convex. Therefore, on the compact convex set $[-1, +1]^N$, it attains its maximum at an extreme point, that is a point $(\alpha_i)_{i \leq N}$ such that $\alpha_i = \pm 1$. But for such values of α_i, by symmetry, both terms in (4.6) are equal. This proves (4.6). Concerning (4.7), replacing α_i by $|\alpha_i|$ we may assume by symmetry that $\alpha_i \geq 0$. Further, by identical distribution, we suppose that $\alpha_1 \geq \cdots \geq \alpha_N \geq \alpha_{N+1} = 0$. Set $S_k = \sum_{i=1}^k \varepsilon_i x_i$. Then

$$\sum_{i=1}^N \alpha_i \varepsilon_i x_i = \sum_{k=1}^N \alpha_k (S_k - S_{k-1}) = \sum_{k=1}^N (\alpha_k - \alpha_{k+1}) S_k .$$

It follows that

$$\left\|\sum_{i=1}^N \alpha_i \varepsilon_i x_i\right\| \leq \max_{k \leq N} \|S_k\| .$$

We conclude by Lévy's inequalities (Proposition 2.3).

As a simple consequence of inequality (4.7), notice the following fact. Let (α_i) and (β_i) be two sequences of real numbers such that $|\beta_i| \leq |\alpha_i|$ for all i; then, if the series with vector coefficients $\sum_i \alpha_i \varepsilon_i x_i$ converges almost surely or equivalently in probability, then the same holds for the series $\sum_i \beta_i \varepsilon_i x_i$.

Theorem 4.4 admits several easy generalizations which will be used mostly without further comments in the sequel. Let us briefly indicate some of these generalizations. Recall that a sequence (η_i) of real random variables is called a symmetric sequence when (η_i) has the same distribution, as a sequence, as $(\varepsilon_i \eta_i)$ where (ε_i) is independent from (η_i). It is then clear by independence and Fubini's theorem that Theorem 4.4 also applies to (η_i) in place of (ε_i). Further, as is easy also, if the α_i's are now random variables independent of (ε_i) (or (η_i)), such that $\|\alpha_i\|_\infty \leq 1$ for all i, the conclusion of Proposition 4.4 still holds true. Moreover, as we will see in Chapter 6, the fixed points x_i can also be replaced by vector valued random variables independent of (ε_i).

The next lemma is another extension and formulation of the contraction principle.

Lemma 4.5. Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and let (η_i) be a symmetric sequence of real random variables such that $\mathbb{E}|\eta_i| < \infty$ for every i. Then, for any finite sequence (x_i) in a Banach space,

$$\mathbb{E}F\left(\inf_i \mathbb{E}|\eta_i| \left\|\sum_i \varepsilon_i x_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_i \eta_i x_i\right\|\right) .$$

Proof. By the symmetry asumption, (η_i) has the same distribution as $(\varepsilon_i|\eta_i|)$ where, as usual, (ε_i) is independent from (η_i). Using first Jensen's inequality and partial integration, and then the contraction principle (4.6), we get that

$$\begin{aligned}
\mathbb{E}F\left(\left\|\sum_i \eta_i x_i\right\|\right) &= \mathbb{E}F\left(\left\|\sum_i \varepsilon_i |\eta_i| x_i\right\|\right) \\
&\geq \mathbb{E}F\left(\left\|\sum_i \varepsilon_i \mathbb{E}|\eta_i| x_i\right\|\right) \\
&\geq \mathbb{E}F\left(\inf_i \mathbb{E}|\eta_i| \left\|\sum_i \varepsilon_i x_i\right\|\right) .
\end{aligned}$$

Note that in case the η_i's have a common distribution the inequality reduces to the application of Jensen's inequality.

An example of a sequence (η_i) of particular interest is given by the orthogaussian sequence (g_i) consisting of independent standard normal variables. Since $\mathbb{E}|g_i| = (2/\pi)^{1/2}$ we have from the previous lemma and its notation that

$$\mathbb{E}F\left(\left\|\sum_i \varepsilon_i x_i\right\|\right) \leq \mathbb{E}F\left(\left(\frac{\pi}{2}\right)^{1/2}\left\|\sum_i g_i x_i\right\|\right). \tag{4.8}$$

Hence Gaussian averages always dominate the corresponding Rademacher ones. In particular, from the integrability properties of Gaussian series (Corollary 3.2), if the series $\sum\limits_i g_i x_i$ is convergent, so is $\sum\limits_i \varepsilon_i x_i$. One might wonder whether a converse inequality or implication hold true. Letting $F(t) = t$ for simplicity, the contraction principle applied conditionally on (g_i) assumed to be independent from (ε_i) yields

$$\mathbb{E}_\varepsilon \left\|\sum_{i=1}^N \varepsilon_i g_i x_i\right\| \leq \max_{i\leq N}|g_i|\mathbb{E}_\varepsilon\left\|\sum_{i=1}^N \varepsilon_i x_i\right\|$$

for any finite sequence $(x_i)_{i\leq N}$ in a Banach space B where we recall that $\mathbb{E}_\varepsilon$ is partial integration with respect to (ε_i). If we now recall from (3.13) that $\mathbb{E}\max\limits_{i\leq N}|g_i| \leq K(\log(N+1))^{1/2}$ for some numerical constant K, we see by integrating the previous inequality that

$$\mathbb{E}\left\|\sum_{i=1}^N g_i x_i\right\| \leq K(\log(N+1))^{1/2}\mathbb{E}\left\|\sum_{i=1}^N \varepsilon_i x_i\right\|. \tag{4.9}$$

This is *not* the converse of inequality (4.8) since the constant depends on the number of elements x_i which are considered. (4.8) is actually best possible in general Banach spaces as is shown by the example of the canonical basis of ℓ_∞ together with the left hand side of (3.14). This example is however extremal in the sense that if a Banach space does not contain subspaces isomorphic to finite dimensional subspaces of ℓ_∞, then (4.9) holds with a constant independent of N (but depending on the Banach space). We shall come back to this later on in Chapter 9 (see (9.12)). So far, we retain that Gaussian averages dominate Rademacher ones and that the converse is not true in general or only holds in the form (4.9).

The next lemma is yet another form of the contraction principle under comparison in probability.

Lemma 4.6. Let $F:\mathbb{R}_+\to\mathbb{R}_+$ be convex. Let further (η_i) and (ξ_i) be two symmetric sequences of real random variables such that for some constant $K\geq 1$ and all i and $t>0$

$$\mathbb{P}\{|\eta_i|>t\} \leq K\mathbb{P}\{|\xi_i|>t\}.$$

Then, for any finite sequence (x_i) in a Banach space,

$$\mathbb{E}F\left(\left\|\sum_i \eta_i x_i\right\|\right) \le \mathbb{E}F\left(K\left\|\sum_i \xi_i x_i\right\|\right) .$$

Proof. Let (δ_i) be independent of (η_i) such that $\mathbb{P}\{\delta_i - 1\} = 1 - \mathbb{P}\{\delta_i = 0\} = 1/K$ for all i ; then, for every $t > 0$,

$$\mathbb{P}\{|\delta_i \eta_i| > t\} \le \mathbb{P}\{|\xi_i| > t\} .$$

Taking inverses of the distribution functions, it is easily seen and classical that the sequences $(\delta_i \eta_i)$ and (ξ_i) can be realized on some probability space in such a way that, almost surely,

$$|\delta_i \eta_i| \le |\xi_i| \qquad \text{for all } i .$$

From the contraction principle and the symmetry assumption it follows that

$$\mathbb{E}F\left(\left\|\sum_i \delta_i \eta_i x_i\right\|\right) \le \mathbb{E}F\left(\left\|\sum_i \xi_i x_i\right\|\right) .$$

The proof is then completed via Jensen's inequality with respect to the sequence (δ_i) since $\mathbb{E}\delta_i = 1/K$.

Notice that if we have only in the preceding lemma that

$$\mathbb{P}\{|\eta_i| > t\} \le K\mathbb{P}\{|\xi_i| > t\}$$

for all $t \ge t_0 > 0$, then the conclusion is somewhat weakened: we have

$$\mathbb{E}F\left(\left\|\sum_i \eta_i x_i\right\|\right) \le \frac{1}{2}\mathbb{E}F\left(2Kt_0\left\|\sum_i \varepsilon_i x_i\right\|\right) + \frac{1}{2}\mathbb{E}F\left(2K\left\|\sum_i \xi_i x_i\right\|\right) .$$

Indeed, simply note that if

$$\xi_i' = t_0 \varepsilon_i I_{\{|\xi_i| \le t_0\}} + \xi_i I_{\{|\xi_i| > t_0\}}$$

where (ε_i) is an independent Rademacher sequence, the couple (η_i, ξ_i') satisfies the hypothesis of Lemma 4.6. Use then convexity and the contraction principle to get rid of the indicator functions.

4.3. Integrability and tail behavior of Rademacher series

On the basis of the scalar results described in the first paragraph, we now investigate integrability properties of Rademacher series with vector valued coefficients. The typical object of study is a convergent series $\sum_i \varepsilon_i x_i$ where (x_i) is a sequence in a Banach space B. This defines a Radon random variable in B. Motivated by the Gaussian study of the previous chapter, there is a somewhat larger setting corresponding to what could be called almost surely bounded *Rademacher processes.* That is, for some set T assumed to be countable for simplicity, let (x_i) be a sequence of functions on T such that for all t, $\sum_i \varepsilon_i x_i(t)$ is almost surely (or in probability) convergent; in other words, $\sum_i x_i(t)^2 < \infty$ for all t. Assuming that

$$\sup_{t\in T}\left|\sum_i \varepsilon_i x_i(t)\right| < \infty \qquad \text{almost surely},$$

we are interested in the integrability properties and tail behavior of this almost surely finite supremum.

As in the previous chapter, in order to unify the exposition, we assume that we are given a Banach space B such that for some countable subset D in the unit ball of B', $\|x\| = \sup_{f\in D}|f(x)|$. We deal with a random variable X with values in B such that there exists a sequence (x_i) of points in B such that $\sum_i f^2(x_i) < \infty$ for every f in D and for which $(f_1(X), \dots, f_N(X))$ has the same distribution as $(\sum_i \varepsilon_i f_1(x_i), \dots, \sum_i \varepsilon_i f_N(x_i))$ for every finite subset $\{f_1, \dots, f_N\}$ of D. We then speak of X as a vector valued Rademacher series (although this terminology is somewhat improper) or almost surely bounded Rademacher process. For such an X we investigate integrability and tail behavior of $\|X\|$. The size of the tail $\mathbb{P}\{\|X\| > t\}$ will be measured in terms of *two* parameters similar to the ones used in the Gaussian case. As for Gaussian random vectors, we consider indeed

$$\sigma = \sigma(X) = \sup_{f\in D}(\mathbb{E}f^2(X))^{1/2} = \sup_{f\in D}\left(\sum_i f^2(x_i)\right)^{1/2} = \sup_{|h|\le 1}\sup_{f\in D}\left|\sum_i h_i f(x_i)\right|$$

(where $h = (h_i) \in \ell_2$). Recall that if $X = \sum_i \varepsilon_i x_i$ is almost surely convergent in a (arbitrary) Banach space B, defining thus a Radon random variable, we can let simply

$$\sigma(X) = \sup_{\|f\|\le 1}(\mathbb{E}f^2(X))^{1/2}.$$

It is easy to see that σ is finite, actually controlled by some quantity associated to the L_0 topology of the norm $\|X\|$: if M is for example such that $\mathbb{P}\{\|X\| > M\} \le 1/8$, we have that $\sigma \le 2\sqrt{2}M$. To see this, recall first that by (4.3), for any f in D,

$$\left(\mathbb{E}\left|\sum_i \varepsilon_i f(x_i)\right|^2\right)^{1/2} \leq \sqrt{2}\,\mathbb{E}\left|\sum_i \varepsilon_i f(x_i)\right| .$$

It then follows from Lemma 4.2 and definition of M that

$$\left(\sum_i f^2(x_i)\right)^{1/2} = \left(\mathbb{E}\left|\sum_i \varepsilon_i f(x_i)\right|^2\right)^{1/2} \leq 2\sqrt{2}M$$

and thus our preceding claim. The tail behavior and integrability properties of $\|X\|$ are measured in terms of this number σ, supremum of weak moments, and some quantity, median or expectation, related to the L_0 topology of the norm (strong moments). The main ingredient is the isoperimetric Theorem 1.3 and related concentration inequality. The next statement summarizes more or less the various results around this question.

Theorem 4.7. Let X be a Rademacher series in B as defined before with corresponding $\sigma = \sigma(X)$. Let moreover $M = M(X)$ denote a median of $\|X\|$. Then, for every $t > 0$,

$$\mathbb{P}\{|\|X\| - M| > t\} \leq 4\exp(-t^2/8\sigma^2) . \tag{4.10}$$

In particular, there exists $\alpha > 0$ such that $\mathbb{E}\exp\alpha\|X\|^2 < \infty$ and all moments of X are equivalent: that is, for any $0 < p, q < \infty$, there is a constant $K_{p,q}$ depending on p, q only such that

$$\|X\|_p \leq K_{p,q}\|X\|_q .$$

Proof. Recall Haar measure μ on $\{-1,+1\}^{\mathbb{N}}$. The function φ on $\mathbb{R}^{\mathbb{N}}$ defined by $\varphi(\alpha) = \sup_{f\in D}|\sum_i \alpha_i f(x_i)|$, $\alpha = (\alpha_i)$, is μ-almost everywhere finite by definition of X. It is convex and Lipschitzian on ℓ_2 with Lipschitz constant σ since

$$|\varphi(\alpha) - \varphi(\beta)| \leq \sup_{f\in D}\left|\sum_i (\alpha_i - \beta_i) f(x_i)\right| \leq \sigma|\alpha - \beta| .$$

Inequality (4.10) is then simply the concentration inequality (1.10) issued from Theorem 1.3 applied to this convex Lipschitzian function on $(\mathbb{R}^{\mathbb{N}}, \mu)$. Alternatively, one may obtain (4.10) directly form Theorem 1.3

by a simple finite dimensional approximation, first through a finite supremum in f, then through finite sums. Recall, as is usual in this setting, that we also have, for all $t > 0$,

$$\mathbb{P}\{\|X\| > M + t\} \leq 2\exp(-t^2/8\sigma^2). \tag{4.11}$$

An integration by parts on the basis of (4.10) already ensures that $\mathbb{E}\exp\alpha\|X\|^2 < \infty$ for all $\alpha > 0$ small enough, namely less than $1/8\sigma^2$. Concerning the moment equivalences, if M' is chosen to satisfy $\mathbb{P}\{\|X\| > M'\} \leq 1/8$, we know that $\sigma \leq 2\sqrt{2}M'$, and, from (4.11) and $M \leq M'$,

$$\mathbb{P}\{\|X\| > M' + t\} \leq 2\exp(-t^2/8\sigma^2)$$

for all $t > 0$. Integrating by parts, for any $0 < p < \infty$,

$$\begin{aligned}\mathbb{E}|\|X\| - M'|^p &\leq M'^{\,p} + \int_0^\infty \mathbb{P}\{\|X\| > M' + t\}\,dt^p \\ &\leq M'^{\,p} + K_p\sigma^p \leq K'_p M'^{\,p}.\end{aligned}$$

Since $M' \leq (8\mathbb{E}\|X\|^q)^{1/q}$ for every $0 < q < \infty$, the claim of the theorem is established. Note that we can take $K_{p,2} = K\sqrt{p}$ ($p \geq 2$) for some numerical constant K.

It is worthwhile mentioning that the moment equivalences in Theorem 4.7 (due to J.-P. Kahane [Ka1]) provide an alternate proof of Khintchine's inequalities (with the right order of magnitude of the constants when $p \to \infty$). They are therefore sometimes refered to in the literature as the *Khintchine-Kahane inequalities*. Note also that (4.10) (or rather (4.11)) implies the weaker but sometimes convenient inequality which corresponds perhaps more directly to the subgaussian inequality (4.1): for every $t > 0$,

$$\mathbb{P}\{\|X\| > t\} \leq 2\exp(-t^2/32\mathbb{E}\|X\|^2). \tag{4.12}$$

For the proof, use (4.11) together with the fact that $M \leq (2\mathbb{E}\|X\|^2)^{1/2}$ and $\sigma^2 \leq \mathbb{E}\|X\|^2$.

For an almost surely convergent series $X = \sum_i \varepsilon_i x_i$ in B, it is very easy to see that the exponential square integrability result of Theorem 4.7 can be refined into $\mathbb{E}\exp\alpha\|X\|^2 < \infty$ *for all* $\alpha > 0$. Set indeed $X_N = \sum_{i=1}^N \varepsilon_i x_i$ for each N. X_N converges to X almost surely and in $L_2(B)$ by Theorem 4.7 so that in particular $\sigma(X_N - X) \to 0$. Let then $\alpha > 0$ and choose an integer N such that the median of $\|X - X_N\|$ is less than 1 (say) and such that $8\sigma(X - X_N)^2\alpha < 1$. It follows from (4.11) that for every $t > 0$

$$\mathbb{P}\{\|X - X_N\| > t + 1\} \leq 2\exp(-t^2/8\sigma(X - X_N)^2).$$

Hence, by the choice of N, $\mathbb{E}\exp\alpha\|X-X_N\|^2<\infty$ from which the claim follows since $\|X_N\|$ is bounded.

It is still true that the preceding observation holds in the more general setting of almost surely bounded Rademacher processes. The proof is however somewhat more complicated and makes use of a variation on the converse subgaussian inequality (4.2).

Theorem 4.8. Let X be a vector valued Rademacher series as in Theorem 4.7. Then

$$\mathbb{E}\exp\alpha\|X\|^2<\infty \qquad \text{for all } \alpha>0\,.$$

Proof. It relies on the following lemma.

Lemma 4.9. There is a numerical constant K with the following property. Let $(\alpha_i)_{i\geq 1}$ be a decreasing sequence of positive numbers such that $\sum_i \alpha_i^2<\infty$. If $t\geq K(\sum_i \alpha_i^2)^{1/2}$, define n to be the smallest integer such that $\sum_{i=1}^n \alpha_i>t$. Then

$$\mathbb{P}\left\{\left|\sum_i \varepsilon_i\alpha_i\right|>t\right\}\geq\frac{1}{2}\exp\left(-Kt^2/\sum_{i\geq n}\alpha_i^2\right).$$

Proof. Let $\sigma=(\sum_{i\geq n}\alpha_i^2)^{1/2}$ and let $K\geq 1$ be the numerical constant of (4.2). We distinguish between two cases. If $n\leq 2Kt^2/\sigma^2$, by definition of n,

$$\mathbb{P}\left\{\sum_{i=1}^n \varepsilon_i\alpha_i>t\right\}\geq 2^{-n}\geq\exp(-2Kt^2/\sigma^2)\,.$$

If $n>2Kt^2/\sigma^2$, by definition of n and since $\alpha_i\leq t$ for all i,

$$\max_{i\geq n}\alpha_i=\alpha_n\leq\frac{1}{n}\sum_{i=1}^n\alpha_i\leq\frac{2t}{n}\leq\frac{K\sigma^2}{t}\,.$$

Therefore, since K is the numerical constant of (4.2),

$$\mathbb{P}\left\{\sum_{i\geq n}\varepsilon_i\alpha_i>t\right\}\geq\exp(-Kt^2/\sigma^2)\,.$$

Lemma 4.9 follows by Lévy's inequality (2.6), changing K into $2K$.

To establish Theorem 4.8 we show a quantitative version of the result. Namely, for every $\alpha > 0$, there exists $\varepsilon = \varepsilon(\alpha) > 0$ such that if M satisfies $\mathbb{P}\{\|X\| > M\} < \varepsilon$, for all $t > 0$,

$$\mathbb{P}\{\|X\| > KM(t+1)\} \leq 2\exp(-\alpha t^2) \tag{4.13}$$

for some numerical K. If $\mathbb{P}\{\|X\| > M\} \leq 1/8$ we know that $\sigma = \sigma(X) \leq 2\sqrt{2}M$. Let then $M' = 2\sqrt{2}KM$ where $K \geq 1$ is the constant of Lemma 4.9. We thus assume that $\varepsilon < 1/8$. By this lemma (applied to the non-increasing rearrangement of $(|f(x_i)|)$ with $t = M'$), there exist, for each f in D, sequences $(u_i(f))$ and $(v_i(f))$ such that, for all i, $f(x_i) = u_i(f) + v_i(f)$, with the following properties:

$$\sum_i |u_i(f)| \leq M', \qquad \exp\left(-KM'^2 / \sum_i v_i(f)^2\right) \leq 2\varepsilon .$$

In particular

$$\|X\| \leq M' + \sup_{f \in D}\left|\sum_i \varepsilon_i v_i(f)\right|$$

and the Rademacher process $(\sum_i \varepsilon_i v_i(f))_{f \in D}$ satisfies the following properties:

$$\mathbb{P}\left\{\sup_{f \in D}\left|\sum_i \varepsilon_i v_i(f)\right| > 2M'\right\} < \varepsilon < \frac{1}{2}$$

and

$$\sup_{f \in D}\sum_i v_i(f)^2 \leq KM'^2\left(\log\frac{1}{2\varepsilon}\right)^{-1} .$$

Given $\alpha > 0$, choose $\varepsilon = \varepsilon(\alpha) > 0$ small enough in order that $K(\log\frac{1}{2\varepsilon})^{-1}$ be smaller than $1/8\alpha$. Then, from (4.11), for all $t > 0$,

$$\mathbb{P}\left\{\sup_{f \in D}\left|\sum_i \varepsilon_i v_i(f)\right| > M'(t+2)\right\} \leq 2\exp(-\alpha t^2) .$$

Hence

$$\mathbb{P}\{\|X\| > M'(t+3)\} \leq 2\exp(-\alpha t^2)$$

which gives (4.13).

We can now easily conclude the proof of Theorem 4.8. For each N, set $Z_N = \sup_{f \in D} |\sum_{i>N} \varepsilon_i f(x_i)|$. (Z_N) defines a reverse submartingale with respect to the family of σ-algebras generated by $\varepsilon_{N+1}, \varepsilon_{N+2}, \ldots$, $N \in \mathbb{N}$. By Theorem 4.7,

$$\sup_N \mathbb{E} Z_N \leq \mathbb{E}\|X\| < \infty .$$

(Z_N) is therefore almost surely convergent. Its limit is measurable with respect to the tail algebra, and therefore degenerated. Hence, there exists $M < \infty$ such that for all $\varepsilon > 0$ one can find N large enough such that $\mathbb{P}\{Z_N > M\} < \varepsilon$. Apply then (4.13) to Z_N to easily conclude the proof of Theorem 4.8 since $\|\sum_{i=1}^N \varepsilon_i x_i\|$ is bounded.

Going back to the moment equivalences of Theorem 4.7, it is interesting to point out that these inequalities contain the corresponding ones for Gaussian averages. This can be seen from a simple argument involving the central limit theorem. Denote by (ε_{ij}) a doubly indexed Rademacher sequence; for any finite sequence (x_i) in B and any n, $0 < p, q < \infty$,

$$\left\| \sum_i \left(\sum_{j=1}^n \frac{\varepsilon_{ij}}{\sqrt{n}} \right) x_i \right\|_p \leq K_{p,q} \left\| \sum_i \left(\sum_{j=1}^n \frac{\varepsilon_{ij}}{\sqrt{n}} \right) x_i \right\|_q .$$

When n tends to infinity, $(\sum_{j=1}^n \varepsilon_{ij}/\sqrt{n})$ converges in distribution to a standard normal. It follows that

$$\left\| \sum_i g_i x_i \right\|_p \leq K_{p,q} \left\| \sum_i g_i x_i \right\|_q .$$

From the right order of magnitude of the constants $K_{p,q}$ in function of p and q also follows the exponential square integrability of Gaussian random vectors (Lemma 3.7).

As yet another remark, we would like to outline a different approach to the conclusions of Theorem 4.7 based on the Gaussian isoperimetric inequality and contractions of Gaussian measures (see (1.7) in Chapter 1). Let (u_i) denote a sequence of independent random variables uniformly distributed on $[-1, +1]$. As described by (1.7), there is an isoperimetric inequality of Gaussian type for these measures. As in the previous chapter (cf. (3.2)), we can then obtain, for example, that for any (finite) sequence (x_i) in B and any $t > 0$

$$\mathbb{P}\left\{ \left\| \sum_i u_i x_i \right\| > 2\mathbb{E}\left\| \sum_i u_i x_i \right\| + t \right\} \leq \exp(-t^2/\pi\sigma^2) \tag{4.14}$$

where σ is as before i.e. $\sigma = \sup\limits_{f\in D}(\sum\limits_i f^2(x_i))^{1/2}$. We now would like to apply the contraction principle in order to replace the sequence (u_i) in (4.14) by the Rademacher sequence. To perform this we need to translate (4.14) in some moment inequalities in order to be able to apply Theorem 4.4 and Lemma 4.5. We make use of the following easy equivalence (in the spirit of Lemma 3.7).

Lemma 4.10. Let Z be a positive random variable and α, β be positive numbers. The following are equivalent:

(i) there is a constant $K > 0$ such that for all $t > 0$,

$$\mathbb{P}\{Z > K(\beta + \alpha t)\} \le K \exp(-t^2/K)\,;$$

(ii) there is a constant $K > 0$ such that for all $p \ge 1$

$$\|Z\|_p \le K(\beta + \alpha\sqrt{p})\,.$$

Further, the constants in (i) and (ii) only depend on each other.

From (4.14) and this lemma it follows that for some numerical constant K

$$\Big\|\sum_i u_i x_i\Big\|_p \le K\left(\Big\|\sum_i u_i x_i\Big\|_1 + \sigma\sqrt{p}\right)$$

for all $p \ge 1$. The contraction principle (Theorem 4.4 and Lemma 4.5) applies to give

$$\Big\|\sum_i \varepsilon_i x_i\Big\|_p \le 2K\left(\Big\|\sum_i \varepsilon_i x_i\Big\|_1 + \sigma\sqrt{p}\right)$$

$(\mathbb{E}|u_i| = 1/2)$. Going back to a tail estimate through Lemma 4.10 we get

$$\mathbb{P}\left\{\Big\|\sum_i \varepsilon_i x_i\Big\| > K\left(\mathbb{E}\,\|\ sum_i \varepsilon_i x_i\| + t\right)\right\} \le K \exp(-t^2/K\sigma^2) \tag{4.15}$$

for some numerical constant $K > 0$ and all $t > 0$.

This inequality easily extends to infinite sums and bounded Rademacher process (start with a norm given by a finite supremum and pass to the limit). It is possible to obtain from it all the conclusions of Theorem 4.7. The most important difference however is that (4.10) expresses a concentration property. However, each time (4.10) is only used as a deviation inequality (i.e. in the form of (4.11)), then (4.15) can be used equivalently.

4.4. Integrability of Rademacher chaos

Let us consider the canonical representation of the Rademacher functions (ε_i) as the coordinate maps on $\Omega = \{-1,+1\}^{\mathbb{N}}$ equipped with the natural product probability measure $\mu = (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1})^{\otimes \mathbb{N}}$. For any finite subset A of $\mathbb{N}$, define $w_A = \prod_{i \in A} \varepsilon_i$ ($w_\emptyset = 1$). It is known that the Walsh system $\{w_A\,;\ A \subset \mathbb{N},\ |A| < \infty\}$ defines an orthonormal basis of $L_2(\mu)$.

For $0 \le \varepsilon \le 1$, introduce the operator $T(\varepsilon) : L_2(\mu) \to L_2(\mu)$ defined by

$$T(\varepsilon) w_A = \varepsilon^{|A|} w_A$$

for $A \subset \mathbb{N}$, $|A| < \infty$. Since, as in the Gaussian case, $T(\varepsilon)$ is a convolution operator, it extends to a positive contraction on all $L_p(\mu)$, $1 \le p < \infty$. One striking property of the operator $T(\varepsilon)$ is the hypercontractivity property similar to the one observed for the Hermite semigroup in the preceding chapter. Namely, for $1 < p < q < \infty$ and $\varepsilon \le [(p-1)/(q-1)]^{1/2}$, $T(\varepsilon)$ maps $L_p(\mu)$ into $L_q(\mu)$ *with norm* 1, i.e. for all f in $L_p(\mu)$

$$\|T(\varepsilon) f\|_q \le \|f\|_p \,. \tag{4.16}$$

This property can be deduced from a sharp two point inequality together with a convolution argument. It implies moreover the corresponding Gaussian hypercontractivity using the central limit theorem ([Gro], [Bec]).

A function f in $L_2(\mu)$ can be written as $f = \sum_A w_A f_A$ where $f_A = \int f w_A \, d\mu$ and the sum runs over all $A \subset \mathbb{N}$, $|A| < \infty$. Regrouping, we can write

$$f = \sum_{d=0}^{\infty} \left(\sum_{|A|=d} w_A f_A \right) = \sum_{d=0}^{\infty} Q_d f \,.$$

$Q_d f$ is called the chaos of degree or order d of f. Chaos of degree 1 are simply Rademacher series $\sum_i \varepsilon_i \alpha_i$, chaos of degree 2 quadratic series of the type $\sum_{i \ne j} \varepsilon_i \varepsilon_j \alpha_{ij}$, etc. Chaos of degree d are characterized by the fact that the action of the operator $T(\varepsilon)$ is multiplication by ε^d, that is

$$T(\varepsilon) Q_d f = \varepsilon^d Q_d f \,.$$

Using (4.16) with $p = 2$, $q > 2$ and $\varepsilon = (q-1)^{-1/2}$, we have that

$$\|Q_d f\|_q \le (q-1)^{d/2} \|Q_d f\|_2 .$$

As we know, this type of inequalities implies strong integrability properties of $Q_d f$; recalling Lemma 3.7, it follow indeed as in the Gaussian case that

$$\mathbb{E} \exp \alpha |Q_d f|^{2/d} < \infty$$

for some (actually all) $\alpha > 0$.

This approach extends to chaos with values in a Banach space providing in particular a different proof of some of the results of the preceding section for chaos of order 1. As for Gaussian chaos, we however would like to complete these integrability results with some more precise tail estimates. This is the object of what follows where we one more time make use of isoperimetric methods. For simplicity we only deal with the chaos of degree, or order, 2 in a setting similar to the one developed in the Gaussian case.

We keep the setting of the preceding paragraph with a Banach space B for which there exists a countable subset D of the unit ball of B' such that $\|x\| = \sup_{f \in D} |f(x)|$ for all x in B. We say that a random variable X with values in B is a Rademacher chaos, of order 2 thus, if there is a sequence (x_{ij}) in B such that $\sum_{i,j} \varepsilon_i \varepsilon_j f(x_{ij})$ is almost surely converent (or only in probability) for all f in D and such that $(f(X))_{f \in D}$ has the same distribution as $(\sum_{i,j} \varepsilon_i \varepsilon_j f(x_{ij}))_{f \in D}$. We assume for simplicity that the diagonal terms x_{ii} are zero in which case the results are more complete. We briefly discuss the general case at the end of the section.

The idea of this study will be to follow the approach to Gaussian chaos using isoperimetric methods in the form of Theorem 1.3. However, with respect to the Gaussian case, some convexity and decoupling results appear to be slightly more complicated here and we will try to detail some of these difficulties.

Let thus X be a Rademacher chaos (of order 2) as just defined. Our aim will be to try to find estimates of the tail $\mathbb{P}\{\|X\| > t\}$ in terms of some parameters of the distribution of X. These are similar to the ones used in the Gaussian setting. Let us consider first the "decoupled" chaos $Y = (\sum_{i,j} \varepsilon_i \varepsilon'_j f(y_{ij}))_{f \in D}$ where (ε'_j) is an independent copy of (ε_i) and $y_{ij} = x_{ij} + x_{ji}$. Let further M be a number such that $\mathbb{P}\{\|X\| \le M\}$ is large enough, for example

$$\mathbb{P}\{\|X\| \le M\} > \frac{63}{64} .$$

Let also m be such that

$$\mathbb{P}\left\{\sup_{|h|\le 1}\sup_{f\in D}\left|\sum_{i,j}\varepsilon_i h_j f(y_{ij})\right|\le m\right\}>\frac{15}{16}$$

and set

$$\sigma=\sigma(X)=\sup_{|h|\le 1}\sup_{f\in D}\left|\sum_{i,j}h_i h_j f(x_{ij})\right|.$$

It might be worthwhile to already mention at this stage that these parameters as well as the decoupled chaos Y are well defined. To this aim we use a *decoupling* argument which will also be useful in the proof of the main result below.

Let us first assume we deal with a norm $\|\cdot\|$ given by a *finite* supremum. Once the right estimates are established we will simply need to increase these supremum to the norm. For each N, let us set further $X_N=\sum_{i,j=1}^{N}\varepsilon_i\varepsilon_j x_{ij}$. If M' is such that $\mathbb{P}\{\|X\|\le M'\}>127/128$, for all N large enough (recall the norm is a finite supremum) $\mathbb{P}\{\|X_N\|\le M'\}\ge 127/128$. Let $I\subset\{1,\dots,N\}$ and let (η_i) be defined as $\eta_i=-1$ if $i\in I$, $\eta_i=+1$ if $i\notin I$. By symmetry

$$\mathbb{P}\left\{\|X_N\|\le M,\left\|\sum_{i,j=1}^{N}\varepsilon_i\eta_i\varepsilon_j\eta_j x_{ij}\right\|\le M\right\}\ge\frac{63}{64}$$

and thus by difference

$$\mathbb{P}\left\{\left\|\sum_{i\in I,j\notin I}\varepsilon_i\varepsilon_j y_{ij}\right\|\le M\right\}\ge\frac{63}{64}.$$

Of course, we are now in a position to "decouple" in the sense that $\sum_{i\in I,j\notin I}\varepsilon_i\varepsilon_j y_{ij}$ has the same distribution as $\sum_{i\in I,j\notin I}\varepsilon_i\varepsilon'_j y_{ij}$. Let us assume for clarity that (ε_i) and (ε'_j) are constructed on different probability spaces $(\Omega,\mathcal{A},\mathbb{P})$ and $(\Omega',\mathcal{A}',\mathbb{P}')$ respectively. We claim that, for some numerical constant K,

$$\mathbb{E}\left\|\sum_{i\in I,j\notin I}\varepsilon_i\varepsilon'_j y_{ij}\right\|^2\le KM'^{\,2}. \tag{4.17}$$

This actually simply follows from Fubini's theorem and the integrability results for one-dimensional chaos, i.e. series. Let indeed

$$A=\left\{\omega:\mathbb{P}'\left\{\left\|\sum_{i\in I,j\notin J}\varepsilon_i(\omega)\varepsilon'_j y_{ij}\right\|\le M'\right\}\ge 7/8\right\}.$$

Then $\mathbb{P}(A) \geq 7/8$. For ω in A, Theorem 4.7 applied to the sum in ε'_j implies that

$$\mathbb{E}' \left\| \sum_{i \in I, j \notin J} \varepsilon_i(\omega) \varepsilon'_j y_{ij} \right\|^2 \leq K M'^{\,2}$$

for some numerical K. But now, the same result applied to the sum in ε_i but in $L_2(\Omega', \mathbb{P}'; B)$ implies the announced claim (4.17) since $\mathbb{P}(A) \geq 7/8$. We note then that

$$\sum_{i,j=1}^{N} \varepsilon_i \varepsilon'_j y_{ij} = 4 \cdot \frac{1}{2^N} \sum_{I \subset \{1, \ldots, N\}} \left(\sum_{i \in I, j \notin J} \varepsilon_i \varepsilon'_j y_{ij} \right) .$$

Therefore, from (4.17),

$$\mathbb{E} \left\| \sum_{i,j=1}^{N} \varepsilon_i \varepsilon'_j y_{ij} \right\|^2 \leq K M'^{\,2} .$$

From a Cauchy sequence argument it then easily follows that for each f in D, $\sum_{i,j} \varepsilon_i \varepsilon'_j f(y_{ij})$ is convergent in L_2 (and almost surely by Fubini). Hence Y is well defined and, *increasing the finite supremum to the norm*, also satisfies $\mathbb{E}\|Y\|^2 \leq K M'^{\,2}$. To control m and σ, observe that, by independence,

$$\begin{aligned} \mathbb{E}\|Y\|^2 &\geq \mathbb{E} \sup_{|h| \leq 1} \sup_{f \in D} \left| \sum_{i,j} \varepsilon_i h_j f(y_{ij}) \right|^2 \\ &\geq \sup_{|h|, |k| \leq 1} \sup_{f \in D} \left| \sum_{i,j} h_i k_j f(y_{ij}) \right|^2 \geq 4\sigma^2 . \end{aligned}$$

Further, by Theorem 4.7 again (although for a different norm), we may take m to be equivalent to

$$\left(\mathbb{E} \sup_{|h| \leq 1} \sup_{f \in D} \left| \sum_{i,j} \varepsilon_i h_j f(y_{ij}) \right|^2 \right)^{1/2}$$

and thus $\sigma \leq Km \leq K^2 M'$ for some numerical constant K. In particular moreover, we will be allowed to deal with finite sums in the estimates we are looking for since the preceding scheme and these inequalities justify all the necessary approximations.

After having described the parameters, decoupling and approximation arguments we will use, we are now ready to state and prove the tail behavior of $\mathbb{P}\{\|X\| > t\}$ where X is a chaos as before.

Theorem 4.11. Let $X = (\sum_{i,j} \varepsilon_i \varepsilon_j f(x_i))_{f \in D}$ be a Rademacher chaos with $x_{ii} = 0$ for all i and let M, m, σ be its parameters as just described. Then, for every $t > 0$,

$$\mathbb{P}\{\|X\| > 2(M + mt + \sigma t^2)\} \leq 20 \exp(-t^2/144).$$

Moreover, $\mathbb{E} \exp \alpha \|X\| < \infty$ for all $\alpha > 0$.

Proof. As announced, we can assume that we deal with a finite sum $X = \sum_{i,j=1}^{N} \varepsilon_i \varepsilon_j x_{ij}$ for the proof of the tail estimate. Recall that

$$X = 2 \cdot \frac{1}{2^N} \sum_{I \subset \{1,\ldots,N\}} \left(\sum_{i \in I, j \notin I} \varepsilon_i \varepsilon_j y_{ij} \right).$$

We first estimate, for all $I \subset \{1, \ldots, N\}$, the tail probability of $\| \sum_{i \in I, j \notin I} \varepsilon_i \varepsilon_j y_{ij} \|$. For simplicity in the notation, we thus assume for this step that $y_{ij} = 0$ if $i \notin I$ or $j \in I$. Recall that, by decoupling and difference,

$$\mathbb{P}\left\{ \left\| \sum_{i,j} \varepsilon_i \varepsilon_j y_{ij} \right\| \leq M \right\} \geq \frac{31}{32}.$$

In the same way, we have that

$$\mathbb{P}\left\{ \sup_{|h| \leq 1} \left\| \sum_{i,j} \varepsilon_i h_j y_{ij} \right\| \leq m \right\} \geq \frac{7}{8}.$$

Let

$$B = \left\{ \omega\,;\ \mathbb{P}' \left\{ \left\| \sum_{i,j} \varepsilon_i(\omega) \varepsilon'_j y_{ij} \right\| \leq M \right\} \geq \frac{3}{4} \ \text{ and } \ \sup_{|h| \leq 1} \left\| \sum_{i,j} \varepsilon_i(\omega) h_j y_{ij} \right\| \leq m \right\}.$$

Thus $\mathbb{P}(B) \geq 3/4$. If $\omega \in B$, we see that we control the one-dimensional parameters in the summation with respect to ε'_j. It therefore follows from Theorem 1.3 that, for $t > 0$,

$$\mathbb{P}(A_1) \geq \mathbb{P}(B)(1 - 2\exp(-t^2/8))$$

where

$$A_1 = \left\{ (\omega, \omega')\,;\ \omega \in B\,,\ \left\| \sum_{i,j} \varepsilon_i(\omega) \varepsilon'_j(\omega') y_{ij} \right\| \leq M + mt \right\}.$$

Also from Theorem 1.3, we have that, for $t > 0$,

$$\mathbb{P}'\left\{\sup_{|h|\le 1}\left\|\sum_{i,j} h_i\varepsilon'_j y_{ij}\right\| > m + 6\sigma t\right\} \le 2\exp(-t^2/8)$$

where we have used the simple fact that

$$\sup_{|h|,|k|\le 1}\left\|\sum_{i,j} h_i k_j y_{ij}\right\| \le 6\sigma\,.$$

Hence, if we let for every $t > 0$

$$A = \left\{(\omega,\omega')\,;\ \omega\in B\,,\ \left\|\sum_{i,j}\varepsilon_i(\omega)\varepsilon'_j(\omega')y_{ij}\right\| \le M + mt\right.$$
$$\left.\text{and}\ \sup_{|h|\le 1}\left\|\sum_{i,j} h_i\varepsilon'_j y_{ij}\right\| \le m + 4\sigma t\right\}$$

we have that

$$\mathbb{P}(A) \ge \mathbb{P}(B)(1 - 4\exp(-t^2/18))\,.$$

Let then

$$B' = \left\{\omega'\,;\ \mathbb{P}\left\{\omega\in B\,;\ \left\|\sum_{i,j}\varepsilon_i(\omega)\varepsilon'_j(\omega')y_{ij}\right\| \le M + mt\right\} \ge \frac{1}{2}\ ,\right.$$
$$\left.\sup_{|h|\le 1}\left\|\sum_{i,j} h_i\varepsilon'_j(\omega')y_{ij}\right\| \le m + 4\sigma t\right\}\,.$$

By Fubini's theorem, $\mathbb{P}'(B') \ge 1 - 8\exp(-t^2/18)$. If ω' is in B', we are in a position to apply the one-dimensional integrability results for the sum in ε_i since we control the corresponding parameters; this gives that

$$\mathbb{P}\left\{\left\|\sum_{i,j}\varepsilon_i\varepsilon'_j y_{ij}\right\| \le M + 2mt + 4\sigma t^2\right\} = \mathbb{P}\left\{\left\|\sum_{i,j}\varepsilon_i\varepsilon'_j y_{ij}\right\| \le M + mt + (m + 4\sigma t)t\right\}$$
$$\ge 1 - 10\exp(-t^2/18)\,.$$

Summarizing, we have obtained that for every $I \subset \{1,\dots,N\}$ and $t > 0$,

$$\mathbb{P}\left\{\left\|\sum_{i\in I,j\notin I}\varepsilon_i\varepsilon_j y_{ij}\right\| > M + tm + \sigma t^2\right\} \le 10\exp\left(-\frac{t^2}{72}\right)\,. \tag{4.18}$$

If we now recall that
$$\sum_{i,j=1}^{N} \varepsilon_i \varepsilon_j x_{ij} = 2 \cdot \frac{1}{2^N} \sum_{I \subset \{1,\dots,N\}} \left(\sum_{i \in I, j \notin I} \varepsilon_i \varepsilon_j y_{ij} \right) ,$$

we are basically left to show that the preceding tail estimate is stable by convex combination. This is however easily proved. Indeed, let $u = u(t) = \sigma t^2 + mt$ and denote by $t = t(u)$ the inverse function (on $\mathbb{R}_+$). The function $\psi(u) = \exp(t(u)^2/144) - 1$ is convex increasing with $\psi(0) = 0$ (elementary computation). By (4.18) and integration by parts we have that
$$\mathbb{E}\psi \left(\left(\left\| \sum_{i \in I, j \notin I} \varepsilon_i \varepsilon_j y_{ij} \right\| - M \right)^+ \right) \leq 10 .$$

Hence by convexity
$$\mathbb{E}\psi \left(\left(\frac{1}{2} \|X\| - M \right)^+ \right) \leq 10 .$$

Thus, for every $t > 0$,
$$\mathbb{P} \left\{ \left(\frac{1}{2} \|X\| - M \right)^+ > mt + \sigma t^2 \right\} \leq \frac{10}{\exp(t^2/144) - 1}$$

from which the tail estimate of the theorem follows. Note that the somewhat unsatisfactory factor 2 came in only at the very end from the decoupling formula.

By the preceding tail estimate, we already know that $\mathbb{E} \exp \alpha \|X\| < \infty$ for some $\alpha > 0$. To establish that this hold for every $\alpha > 0$ it clearly suffices by the decoupling argument developed above (and Fatou's lemma) to show this integrability result for the decoupled chaos Y. We make use of the proof of Theorem 4.8 and (4.13). Let $Z_N = \sup_{f \in D} \| \sum_{i,j > N} \varepsilon_i \varepsilon_j' y_{ij} \|$. By the reverse submartingale theorem, Z_N converges almost surely to some degenerate distribution. Using then (4.13) and Fubini's theorem as in the previous proof of the tail estimate, we find that there exists $M > 0$ such that for every $\alpha > 0$ one can find N large enough with
$$\mathbb{P}\{Z_N > KM(t+1)\} \leq K \exp(-\alpha t)$$

for all $t > 0$, K numerical. By the one-dimensional integrability properties, the proof of Theorem 4.11 is easily completed.

We conclude this part on Rademacher chaos with a few words on the case where the diagonal elements are non-zero. Assume we are given a finite sequence (x_{ij}) in a Banach space B. We just learned that there is a numerical constant K such that for all $p \geq 1$

$$\Big\| \sum_{i \neq j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_p \leq Kp \Big\| \sum_{i \neq j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_1$$

(cf. Lemma 3.7). Denote by (ε_i') an independent Rademacher sequence. Then, by independence and Jensen's inequality,

$$\begin{aligned} \Big\| \sum_{i \neq j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_1 &\leq \Big\| \sum_{i \neq j} (\varepsilon_i \varepsilon_j - \varepsilon_i' \varepsilon_j') x_{ij} \Big\|_1 \\ &= \Big\| \sum_{i,j} (\varepsilon_i \varepsilon_j - \varepsilon_i' \varepsilon_j') x_{ij} \Big\|_1 \\ &\leq 2 \Big\| \sum_{i,j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_1 . \end{aligned}$$

Hence, by difference,

$$\Big\| \sum_i x_{ii} \Big\| \leq 3 \Big\| \sum_{i,j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_1 .$$

Therefore, for every $p \geq 1$,

$$\Big\| \sum_{i,j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_p \leq (3 + 2Kp) \Big\| \sum_{i,j} \varepsilon_i \varepsilon_j x_{ij} \Big\|_1 ,$$

from which we deduce similar integrability properties for chaos with non-zero diagonal terms. We do not know however whether the tail estimate of Theorem 4.11 extends to this setting.

4.5. Comparison theorems

The norm of a (finite) Rademacher sum $\| \sum_{i=1}^N \varepsilon_i x_i \|$ with coefficients in a Banach space B is the supremum of a Rademacher process. For the purposes of this study, one convenient representation is

$$\Big\| \sum_{i=1}^N \varepsilon_i x_i \Big\| = \sup_{t \in T} \Big| \sum_{i=1}^N \varepsilon_i t_i \Big|$$

where T is the (bounded) subset of $\mathbb{R}^N$ defined by $T = \{t = (f(x_i))_{i \le N} \, ; \ f \in B' , \ \|f\| \le 1\}$. We therefore present the results on comparison for Rademacher processes of this type, i.e. $\sum_{i=1}^{N} \varepsilon_i t_i$, $t = (t_1, \ldots, t_N) \in T$, T (bounded) subset of $\mathbb{R}^N$.

We learned in the preceding chapter how Gaussian processes can be compared by mean of their L_2-metrics. While this is not completely possible for Rademacher processes, one can however investigate analogs of some of the usual consequences of the Gaussian comparison theorems. More precisely, we establish a comparison theorem for Rademacher averages when coordinates are contracted and we prove a version of Sudakov's minoration inequality in this context. Both results are due to the second author.

We start with the comparison theorem, analogous to Corollary 3.17. A map $\varphi : \mathbb{R} \to \mathbb{R}$ is called a contraction when $|\varphi(s) - \varphi(t)| \le |s - t|$ for all $s, t \in \mathbb{R}$. If h is a map on some set T, we set for simplicity (and with some abuse) $\|h(t)\|_T = \|h\|_T = \sup_{t \in T} |h(t)|$.

Theorem 4.12. Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing. Let further $\varphi_i : \mathbb{R} \to \mathbb{R}$, $i \le N$, be contractions such that $\varphi_i(0) = 0$. Then, for any bounded subset T in $\mathbb{R}^N$

$$\mathbb{E}F\left(\frac{1}{2}\left\|\sum_{i=1}^{N} \varepsilon_i \varphi_i(t_i)\right\|_T\right) \le \mathbb{E}F\left(\left\|\sum_{i=1}^{N} \varepsilon_i t_i\right\|_T\right) .$$

Before turning to the proof, note the following. The numerical constant $\frac{1}{2}$ is optimal as can be seen from the example of the subset T of $\mathbb{R}^2$ consisting of the points $(1,1)$ and $(-1,-1)$ with $\varphi_1(x) = x$, $\varphi_2(x) = -|x|$ and $F(x) = x$. One typical appication of Theorem 4.12 is of course given when $\varphi_i(x) = |x|$ for all i . Another one which will be useful in the sequel is the following. If $(x_i)_{i \le N}$ are points in a Banach space, then

$$\mathbb{E}\left(\sup_{\|f\| \le 1}\left|\sum_{i=1}^{N} \varepsilon_i f^2(x_i)\right|\right) \le 4\mathbb{E}\left\|\sum_{i=1}^{N} \varepsilon \|x_i\| x_i\right\| . \tag{4.19}$$

(Recall that by the contraction principle we can replace the right hand side by $4 \max_{i \le N} \|x_i\| \mathbb{E}\| \sum_{i=1}^{N} \varepsilon_i x_i \|$.) To deduce (4.19) from the theorem, let simply T be as before, i.e. $T = \{t = (f(x_i))_{i \le N}; \ \|f\| \le 1\}$, and take

$$\varphi_i(s) = \min\left(\frac{s^2}{2\|x_i\|^2} , \ \frac{\|x_i\|^2}{2}\right) , \qquad s \in \mathbb{R} ; \ i \le N .$$

As we mentioned before, theorems like Theorem 4.12 for Gaussian averages follow from the Gaussian comparison properties. The Rademacher case involves independent (and conceptually simpler) proofs to which we now turn.

Proof of Theorem 4.12. We first show that if $G : \mathbb{R} \to \mathbb{R}$ is convex and increasing

$$\mathbb{E}G\left(\sup_{t\in T}\sum_{i=1}^{N}\varepsilon_i\varphi_i(t_i)\right) \leq \mathbb{E}G\left(\sup_{t\in T}\sum_{i=1}^{N}\varepsilon_i t_i\right) . \tag{4.20}$$

By conditioning and iteration, it suffices to show that if T is a subset of $\mathbb{R}^2$ and φ a contraction on $\mathbb{R}$ such that $\varphi(0) = 0$, then

$$\mathbb{E}G(\sup_{t\in T}(t_1 + \varepsilon_2\varphi(t_2))) \leq \mathbb{E}G(\sup_{t\in T}(t_1 + \varepsilon_2 t_2))$$

($t = (t_1, t_2)$). We show that for all t and s in T, the right hand side is always larger than

$$I = \frac{1}{2}G(t_1 + \varphi(t_2)) + \frac{1}{2}G(s_1 - \varphi(s_2)) .$$

We may assume that

$$t_1 + \varphi(t_2) \geq s_1 + \varphi(s_2) \tag{$*$}$$

and

$$s_1 - \varphi(s_2) \geq t_1 - \varphi(t_2) . \tag{$**$}$$

We distinguish between the following cases.

1 $^{\text{st}}$ case. $t_2 \geq 0$, $s_2 \geq 0$. Assume to begin with that $s_2 \leq t_2$. We show that

$$2I \leq G(t_1 + t_2) + G(s_1 - s_2) .$$

Set $a = s_1 - \varphi(s_2)$, $b = s_1 - s_2$, $a' = t_1 + t_2$, $b' = t_1 + \varphi(t_2)$ so that we would like to prove that

$$G(a) - G(b) \leq G(a') - G(b') . \tag{4.21}$$

Since φ is a contraction with $\varphi(0) = 0$ and $s_2 \geq 0$, $|\varphi(s_2)| \leq s_2$. Hence, $a \geq b$ and, by $(*)$, $b' \geq b$. Further, again by contraction and $s_2 \leq t_2$,

$$a - b = s_2 - \varphi(s_2) \leq t_2 - \varphi(t_2) = a' - b' .$$

Since G is convex and increasing, for all positive x, the map $G(\cdot + x) - G(\cdot)$ is increasing. Thus, for $x = a - b \geq 0$ and with $b \leq b'$, we get that

$$G(a) - G(b) \leq G(b' + (a - b)) - G(b')\,.$$

Using that $b' + a - b \leq a'$ yields then the announced claim (4.21). When $s_2 \geq t_2$, the argument is similar changing s into t and φ into $-\varphi$.

2^{nd} case. $t_2 \leq 0$, $s_2 \leq 0$. It is completely similar to the preceding one.

3^{rd} case. $t_2 \geq 0$, $s_2 \leq 0$. Since $\varphi(t_2) \leq t_2$, $-\varphi(s_2) \leq -s_2$, we have that

$$2I \leq G(t_1 + t_2) + G(s_1 - s_2)$$

and the result follows.

4^{th} case. $t_2 \leq 0$, $s_2 \geq 0$. Similar to the 3^{rd} case. This completes the proof of (4.20).

We now conclude the proof of the theorem. By convexity,

$$\begin{aligned}\mathbb{E}F\left(\frac{1}{2}\left\|\sum_{i=1}^N \varepsilon_i\varphi_i(t_i)\right\|_T\right) &\leq \frac{1}{2}\left(\mathbb{E}F\left(\sup_{t\in T}\left(\sum_{i=1}^N \varepsilon_i\varphi_i(t_i)\right)^+\right) + \mathbb{E}F\left(\sup_{t\in T}\left(\sum_{i=1}^N \varepsilon_i\varphi_i(t_i)\right)^-\right)\right)\\ &\leq \mathbb{E}F\left(\sup_{t\in T}\left(\sum_{i=1}^N \varepsilon_i\varphi_i(t_i)\right)^+\right)\end{aligned}$$

where we have used in the second step that, by symmetry, $(-\varepsilon_i)$ has the same distribution as (ε_i) and $(-\cdot)^- = (\cdot)^+$. Applying (4.20) to $F((\cdot)^+)$ which is convex and increasing on $\mathbb{R}$ yields then immediately the conclusion. The proof of Theorem 4.12 is complete.

After comparison properties, we describe in the last part of this chapter a version of the Sudakov minoration inequality (Theorem 3.18) for Rademacher processes.

If T is a (bounded) subset of $\mathbb{R}^N$, set

$$r(T) = \mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^N \varepsilon_i t_i\right|\,.$$

Denote by $d_2(s,t) = |s - t|$ the Euclidean metric on $\mathbb{R}^N$ and recall $N(T, d_2; \varepsilon)$ the minimal number of (open) balls of radius $\varepsilon > 0$ in the metric d_2 sufficient to cover T. Equivalently $N(T, d_2; \varepsilon) = N(T, \varepsilon B_2)$

where we denote by $N(A,B)$ the minimal number of translates of B by elements of A necessary to cover A, and where B_2 is the Euclidean unit ball (open). (We do not specify later if the balls are open or closed since this distinction clearly does not affect the various results.) The next result is the main step in the proof of Sudakov's minoration inequality for Rademacher processes.

Proposition 4.13. There is a numerical constant K such that for any $\varepsilon > 0$, if T is a subset of $\mathbb{R}^N$ such that $\max\limits_{i\le N}|t_i| \le \varepsilon^2/Kr(T)$ for any $t\in T$, then

$$\varepsilon(\log N(T,d_2;\varepsilon))^{1/2} \le Kr(T)\,.$$

Proof. As an intermediary step, we first show that when $T\subset B_2$ and $\max\limits_{i\le N}|t_i| \le 1/Kr(T)$ for all $t\in T$, then

$$\log N(T,\frac{1}{2}B_2))^{1/2} \le Kr(T) \tag{4.23}$$

where K is a numerical constant.

Let g be a standard normal variable. For $s>0$, set $h = gI_{\{|g|>s\}}$. The first simple observation of this proof is that, whenever $\lambda \le s/4$,

$$\mathbb{E}\exp\lambda h \le 1 + 16\lambda^2\exp(-s^2/32) \le \exp[16\lambda^2\exp(-s^2/32)]\,. \tag{4.24}$$

For a proof, consider for example $f(\lambda) = \mathbb{E}(\exp\lambda h) - 1 - 16\lambda^2\exp(-s^2/32)$ for $\lambda\ge 0$. Since $f(0) = f'(0) = 0$, it will be sufficient to check that $f''(\lambda)\le 0$ when $\lambda\le s/4$. Now

$$f''(\lambda) = \mathbb{E}(h^2\exp\lambda h) - 32\exp(-s^2/32)\,.$$

By definition of h and change of variables,

$$\begin{aligned}\mathbb{E}(h^2\exp\lambda h) &= \int_{|x|>s} x^2\exp\left(\lambda x - \frac{x^2}{2}\right)\frac{dx}{\sqrt{2\pi}}\\ &= \exp\left(\frac{\lambda^2}{2}\right)\int_{|x+\lambda|>s}(x+\lambda)^2\exp\left(-\frac{x^2}{2}\right)\frac{dx}{\sqrt{2\pi}}\,.\end{aligned}$$

When $\lambda\le s/4\le s/2$, $s<|x+\lambda|\le|x|+s/2\le 2|x|$ so that

$$\begin{aligned}\mathbb{E}(h^2\exp\lambda h) &\le 4\exp(\frac{\lambda^2}{2})\int_{|x|>s/2}x^2\exp(-\frac{x^2}{2})\frac{dx}{\sqrt{2\pi}}\\ &\le 16\exp(\frac{\lambda^2}{2})\int_{|x|>s/2}\exp(-\frac{x^2}{4})\frac{dx}{\sqrt{2\pi}}\\ &\le 32\exp(\frac{\lambda^2}{2}-\frac{s^2}{16})\le 32\exp(-\frac{s^2}{32})\end{aligned}$$

which gives the result.

Let us now show (4.23). There is nothing to prove it $T \subset \frac{1}{2}B_2$ so that we may assume that there is an element t of T with $|t| \geq 1/2$; then, by (4.3), $r(T) \geq 1/2\sqrt{2}$. By definition of $N(T, \frac{1}{2}B_2)$ there exists a subset U of T of cardinality $N(T, \frac{1}{2}B_2)$ such that $d_2(u,v) \geq 1/2$ whenever u, v are distinct elements in U. Let us then consider the Gaussian process $(\sum_{i=1}^{N} g_i u_i)_{u \in U}$ where (g_i) is an orthogaussian sequence. As a consequence of Sudakov's minoration inequality and Gaussian integrability properties (Theorem 3.18 and Corollary 3.2), there is a numerical constant $K' \geq 1$ such that

$$\mathbb{P}\left\{\sup_{u \in U}\left|\sum_{i=1}^{N} g_i u_i\right| > (K')^{-1}(\log\ \mathrm{Card} U)^{1/2}\right\} \geq \frac{1}{2}.$$

We use a kind of *a priori* estimate argument. Let $K = (100K')^2$ and assume that $\max_{i \leq N}|t_i| \leq 1/Kr(T)$ for every $t \in T$. We claim that whenever $\alpha \geq 1$ satisfies $(\log\ \mathrm{Card} U)^{1/2} \leq \alpha K r(T)$, then

$$\mathbb{P}\left\{\sup_{u \in U}\left|\sum_{i=1}^{N} g_i u_i\right| > \alpha\frac{K}{2K'}r(T)\right\} < \frac{1}{2}$$

so that, intersecting the probabilities, $(\log\ \mathrm{Card} U)^{1/2} \leq \alpha K r(T)/2$. This of course ensures that $(\log\ \mathrm{Card} U)^{1/2} \leq Kr(T)$ which is the conclusion (4.23).

For all i, set $h_i = g_i I_{\{|g_i| > s\}}$, $k_i = g_i - h_i$. By the triangle inequality and the contraction principle, since $|k_i| \leq s$,

$$\mathbb{P}\left\{\sup_{u \in U}\left|\sum_{i=1}^{N} g_i u_i\right| > \alpha\frac{K}{2K'}r(T)\right\} \leq \frac{4K's}{\alpha K} + \mathbb{P}\left\{\sup_{u \in U}\left|\sum_{i=1}^{N} h_i u_i\right| > \alpha\frac{K}{4K'}r(T)\right\}.$$

By (4.24), for any $\lambda \leq sKr(T)/4$, and since $U \subset T \subset B_2$,

$$\begin{aligned}\mathbb{P}\left\{\sup_{u \in U}\left|\sum_{i=1}^{N} h_i u_i\right| > \alpha\frac{K}{4K'}r(T)\right\} &\leq 2\ \mathrm{Card} U \exp[-\lambda\alpha\frac{K}{4K'}r(T) + 16\lambda^2\exp(-\frac{s^2}{32})] \\ &\leq 2\exp[\alpha^2K^2r(T)^2 - \lambda\alpha\frac{K}{4K'}r(T) + 16\lambda^2\exp(-\frac{s^2}{32})].\end{aligned}$$

Let $s = \alpha K/10K'$ and $\lambda = \alpha K^2 r(T)/40K'$. Then

$$\mathbb{P}\{\sup_{u \in U}|\sum_{i=1}^{N} g_i u_i| > \alpha\frac{K}{4K'}r(T)\} \leq \frac{2}{5} + 2\exp\left[\alpha^2K^2r(T)^2(1 - \frac{K}{160(K')^2} + \frac{16K^2}{(40K')^2}\exp(-\frac{\alpha^2K^2}{32(10K')^2}))\right].$$

If we recall that $\alpha \geq 1$, $r(T) \geq 1/2\sqrt{2}$ and $K = (100K')^2 \geq 10^4$, it is clear that the preceding probability is made less than $1/2$ which was the announced claim.

To reach the full conclusion of the proposition, we use a simple iteration procedure. For each t and $\delta > 0$, denote by $B_2(t,\delta)$ the Euclidean ball of center t and radius δ. Let $\varepsilon > 0$ and k be an integer such that $2^{-k} < \varepsilon \leq 2^{-k+1}$. Then

$$N(T, d_2; \varepsilon) = N(T, \varepsilon B_2) \leq N(T, 2^{-k} B_2) .$$

Clearly

$$N(T, 2^{-k} B_2) \leq \prod_{\ell \leq k} \sup_{t \in T} N(T \cap B_2(t, 2^{-\ell+1}), 2^{-\ell} B_2) .$$

By homogeneity, (4.23) tells us that

$$N(T \cap B_2(t, 2^{-\ell+1}), 2^{-\ell} B_2) \leq \exp(K 2^{2\ell-2} r(T)^2) .$$

Hence

$$N(T, d_2; \varepsilon) \leq \exp\left(K \sum_{\ell \leq k} 2^{2\ell-2} r(T)^2\right) \leq \exp(4K \frac{r(T)^2}{\varepsilon^2}) .$$

Proposition 4.13 is therefore established.

The previous proposition yields a first version for Rademacher processes of Sudakov's minoration which however involves a factor depending on the dimension. It can be stated as follows.

Corollary 4.14. Let T be a subset of $\mathbb{R}^N$; for all $\varepsilon > 0$

$$\varepsilon (\log N(T, d_2; \varepsilon))^{1/2} \leq K r(T) \left(\log(2 + \frac{\varepsilon\sqrt{N}}{r(T)})\right)^{1/2}$$

where K is some numerical constant.

Proof. As before, let us first assume that $T \subset B_2$ and estimate $N(T, \frac{1}{2} B_2)$. Denote by K_1 the numerical constant in Proposition 4.13. We can write that

$$\begin{aligned} N(T, \frac{1}{2} B_2) &\leq N(T, \frac{1}{K_1 r(T)} B_\infty) N((T-T) \cap \frac{1}{K_1 r(T)} B_\infty\,,\ \frac{1}{2} B_2) \\ &\leq N(B_2, \frac{1}{K_1 r(T)} B_\infty) N((T-T) \cap \frac{1}{K_1 r(T)} B_\infty\,,\ \frac{1}{2} B_2) \end{aligned}$$

where B_∞ is the unit ball for the sup-norm in $\mathbb{R}^N$. It is known that (see [Schü, Theorem 1])

$$\left(\log N(B_2\,,\ \frac{1}{K_1 r(T)}B_\infty)\right)^{1/2} \leq K_2 r(T)\left(\log(2+\frac{\sqrt{N}}{r(T)})\right)^{1/2}$$

where K_2 is numerical (it can be assumed that $r(T)$ is bounded below). Combining with Proposition 4.13 we get that for some K_3 numerical

$$(\log N(T,\frac{1}{2}B_2))^{1/2} \leq K_3 r(T)\left(\log(2+\frac{\sqrt{N}}{r(T)})\right)^{1/2} .$$

We can then use an iteration argument similar to the one used in Proposition 4.13 to obtain the inequality of the corollary. The proof is complete.

We now turn to another version of Sudakov's minoration. The example of T consisting of the canonical basis of $\mathbb{R}^N$ for which clearly $r(T) = 1$ and $N(T, B_2) = N$ indicates that Sudakov's minoration for Gaussians cannot extend litterally to Rademachers. On the other hand, note that if $T \subset B_1$, the ℓ_1^N-ball, $r(T) \leq 1$. This suggests the possibility of some interpolation and of a minoration involving *both* B_2 and B_1, the unit balls of ℓ_2^N and ℓ_2^N respectively. This is the conclusion of the next statement.

Theorem 4.15. Let T be a (bounded) subset of $\mathbb{R}^N$ and let $r(T) = \mathbb{E}\sup_{t\in T}|\sum_{i=1}^N \varepsilon_i t_i|$. There exists a numerical constant K such that if $\varepsilon > 0$ and if $D = Kr(T)B_1 + \varepsilon B_2$, then

$$\varepsilon(\log N(T,D))^{1/2} \leq Kr(T)$$

where we recall that $N(T,D)$ is the minimal number of translates of D by elements of T necessary to cover T.

Proof. The idea is to use Proposition 4.13 by changing the strange balls D into ℓ_2 balls for another T. Let K_1 be the constant of the conclusion of Proposition 4.13 and set $K = 3K_1$ which we would like to fit the statement of Theorem 4.15. Set $a = \varepsilon^2/Kr(T)$ and let M be an integer bigger than $r(T)/a$. Define a map φ in the following way:

$$\varphi : [-Ma\,,\ +Ma] \to \mathbb{R}^{[-M,M]}$$

$$u \to (\varphi(u)_j)_{-M\leq j\leq M}\ ;$$

$\varphi(u)_j$ is defined according to the following rule: if $u \in [0, Ma]$, $k = k(u)$ is the integer part of u/a; we then set

$$\begin{aligned} \varphi(u)_j &= a \qquad &&\text{for } 1 \le j \le k \\ \varphi(u)_{k+1} &= u - ka && \\ \varphi(u)_j &= 0 &&\text{for all other values of } j\,. \end{aligned}$$

If $u \in [-Ma, 0]$, we let $\varphi(u)_j = \varphi(-u)_{-j}$. We mention some elementary properties of φ. First, for every u, u' in $[-Ma, Ma]$,

$$\sum_{j=-M}^{M} |\varphi(u)_j - \varphi(u')_j| = |u - u'|\,. \tag{4.25}$$

Another elementary property is the following. Suppose we are given u, u' in $[-Ma, Ma]$ and assume $u' < u$. Let us define $v' \le v$ in the following way: if $u \ge 0$, let $v = k(u)a$ and $v' = (k(u') + 1)a$ or $v' = k(u')a$ according whether $u' \ge 0$ or $u' < 0$; if $u < 0$, we let $v = (k(u) - 1)a$ and $v' = k(u')a$. We then have that

$$\sum_{j=-M}^{M} |\varphi(u)_j - \varphi(u')_j|^2 = |u - v|^2 + |u' - v'|^2 + a|v - v'|\,. \tag{4.26}$$

Once these properties have been observed, let

$$\begin{aligned} \psi : T &\to (\mathbb{R}^{[-M,M]})^N \\ t = (t_i)_{i \le N} &\to ((\varphi(t_i)_j)_{-M \le j \le M})_{i \le N}\,. \end{aligned}$$

ψ is of course well defined since if $t \in T$, then $|t_i| \le r(T) \le Ma$ for every $i = 1, \ldots, N$. Consider now a doubly indexed Rademacher sequence (ε_{ij}) and another one (ε'_i) which we assume to be independent. Then, by symmetry,

$$r(\psi(T)) = \mathbb{E} \sup_{t \in T} \left| \sum_{i=1}^{N} \sum_{j=-M}^{M} \varepsilon_{ij} \varphi(t_i)_j \right| = \mathbb{E} \sup_{t \in T} \left| \sum_{i=1}^{N} \varepsilon'_i \Big(\sum_{j=-M}^{M} \varepsilon_{ij} \varphi(t_i)_j \Big) \right|\,.$$

Now, for every choice of (ε_{ij}), every i and t, t' in T, by (4.25),

$$\left| \sum_{j=-M}^{M} \varepsilon_{ij} \varphi(t_i)_j - \sum_{j=-M}^{M} \varepsilon_{ij} \varphi(t'_i)_j \right| \le |t_i - t'_i|\,.$$

This means that, with respect to (ε'_i), we are in a position to apply the comparison Theorem 4.12 from which we get that $r(\psi(T)) \le 3r(T)$. Now, by construction, for every t, i, j,

$$|\varphi(t_i)_j| \le a = \frac{\varepsilon^2}{3K_1 r(T)} \le \frac{\varepsilon^2}{K_1 r(\psi(T))} .$$

Hence, by Proposition 4.13 applied to $\psi(T)$ in $(\mathbb{R}^{[-M,M]})^N$,

$$\varepsilon(\log N(\psi(T), \varepsilon B_2))^{1/2} \le K_1 r(\psi(T)) \le K r(T) .$$

Now this implies the conclusion; indeed, by (4.26), if t, t' are such that $|\psi(t) - \psi(t')| \le \varepsilon$, then

$$t \in t' + K r(T) B_1 + \varepsilon B_2 .$$

Hence $N(T, D) \le N(\psi(T), \varepsilon B_2)$ and the proof of Theorem 4.15 is therefore complete.

We conclude this chapter with a remark on tensorization of Rademacher series. As in the Gaussian case, (and we follow here the notations introduced in Section 3.3) we ask ourselves if, given (x_i) and (y_i) in Banach spaces E and F respectively such that $\sum_i \varepsilon_i x_i$ and $\sum_i \varepsilon_i y_i$ are both almost surely convergent, this also true for $\sum_{i,j} \varepsilon_{ij} x_i \otimes y_j$ in the injective tensor product $E \check{\otimes} F$. To investigate this question, recall first (4.8). For a large class of Banach spaces (which will be described in Chapter 9 as the spaces having a finite cotype), convergence of Rademacher series $\sum_i \varepsilon_i x_i$ and corresponding Gaussian series $\sum_i g_i x_i$ are equivalent. Therefore, according to Theorem 3.20, if E and F have this property, the answer to the preceding question is yes. What we would like to briefly point out here is that this is not the case in general.

Let $(x_i)_{i \le N}$ (resp. $(y_i)_{i \le N}$) be a finite sequence in E (resp. F) and set $X = \sum_{i=1}^N \varepsilon_i x_i$ (resp. $Y = \sum_{i=1}^N \varepsilon_i y_i$). Recall $\sigma(X) = \sup_{\|f\| \le 1} (\sum_{i=1}^N f^2(x_i))^{1/2}$ ($\sigma(Y) = \sup_{\|f\| \le 1} (\sum_{i=1}^N f^2(y_i))^{1/2}$). Then for the Rademacher average $\sum_{i,j=1}^N \varepsilon_{ij} x_i \otimes y_j$ in $E \check{\otimes} F$ we have:

$$\text{(4.27)} \qquad \mathbb{E} \left\| \sum_{i,j=1}^N \varepsilon_{ij} x_i \otimes y_j \right\|_{\vee} \le K (\log(N+1))^{1/2} (\sigma(X) \mathbb{E}\|Y\| + \sigma(Y) \mathbb{E}\|X\|)$$

where K is numerical. This inequality is an immediate consequence of, respectively, (4.8), Theorem 3.20 and (4.9). The point of this observation is that (4.27) is best possible in general. To see it, let $E = \ell_\infty^N$, x_i be the elements of the canonical basis, and $F = \mathbb{R}$, $y_i = N^{-1/2}$, $i = 1, \dots, N$. Then clearly $\sigma(X) = \sigma(Y) = 1$, $\mathbb{E}\|X\| = 1$, $\mathbb{E}\|Y\| \leq 1$. However, by definition of the tensor product norm

$$\mathbb{E}\left\|\sum_{i,j=1}^{N} \varepsilon_{ij} x_i \otimes y_j\right\|_{\vee} = \mathbb{E}\left(\max_{i \leq N} \frac{1}{\sqrt{N}}\left|\sum_{j=1}^{N} \varepsilon_{ij}\right|\right)$$

and this quantity turns out to be of the order of $(\log N)^{1/2}$. Indeed, by (4.2), for some numerical $K > 0$,

$$\mathbb{P}\left\{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varepsilon_i > K^{-1}(\log N)^{1/2}\right\} \geq \frac{1}{N} \tag{4.28}$$

(at least for all N large enough). But then

$$\begin{aligned}
\mathbb{E}\left(\max_{i \leq N} \frac{1}{\sqrt{N}}\left|\sum_{i=1}^{N} \varepsilon_{ij}\right|\right) &\geq \int_0^{K^{-1}(\log N)^{1/2}} \mathbb{P}\left\{\max_{i \leq N} \frac{1}{\sqrt{N}}\left|\sum_{j=1}^{N} \varepsilon_{ij}\right| > t\right\} dt \\
&\geq \int_0^{K^{-1}(\log N)^{1/2}} \left[1 - \left(1 - \mathbb{P}\left\{\frac{1}{\sqrt{N}}\left|\sum_{i=1}^{N} \varepsilon_i\right| > t\right\}\right)^N\right] dt \\
&\geq \int_0^{K^{-1}(\log N)^{1/2}} [1 - (1 - \frac{1}{N})^N] \, dt \\
&\geq K^{-1}(1 - \frac{1}{e})(\log N)^{1/2}
\end{aligned}$$

which proves our claim.

Notes and references

The name of Bernoulli is historically more appropriate for a random variable taking the values ± 1 with equal probability. Strictly speaking, the Rademacher sequence is the sequence on $[0,1]$ defined by $r_i(t) = \sin(2\pi i t)$ ($i \geq 1$). We decided to use the terminology of Rademacher sequence since it is the commonly used one in the field as well as in Geometry of Banach spaces.

The best constants in the (real) Khintchine's inequality were obtained by U. Haagerup [Ha]. See [Sz] for the case $p = 1$ (4.3). Lemma 4.2 is in the spirit of the Paley-Zygmund inequality (cf. [Ka1]). Lemma 4.3 has been observed in [R-S] and [Pi10] and used in Probability in Banach spaces in [M-P2] (cf. Chapter 13).

The contraction principle has been discovered by J.-P. Kahane [Ka1]. Some further extensions have been obtained by J. Hoffmann-Jørgensen [HJ1], [HJ2], [HJ3]. Lemma 4.6 is taken from [J-M2].

In [Ka1] (first edition), J.-P. Kahane showed that an almost surely convergent Rademacher series $X = \sum_i \varepsilon_i x_i$ with coefficients in a Banach space satisfies $\mathbb{E}\exp\alpha\|X\| < \infty$ for all $\alpha > 0$ and has all its moments equivalent. Using Lemma 4.5, S. Kwapień [Kw3] improved this integrability to $\mathbb{E}\exp\alpha\|X\|^2 < \infty$ for some (and also all) $\alpha > 0$ (Theorem 4.7). The proof of this result presented here is different and is based on isoperimetry. Theorem 4.8 on bounded Rademacher processes is perhaps new; its proof uses Lemma 4.9 which was noticed independently in [MS2].

The hypercontractivity inequality (4.16) was established by L. Gross (as a logarithmic Sobolev inequality) and W. Beckner (as a two point inequality). Its interest for integrability of Rademacher chaos was pointed out by C. Borell [Bo5]. Complete details may be found in [Pi4] where the early contribution of A. Bonami [Bon] is pointed out. The decoupling argument used in the proof of Theorem 4.11 is inspired from [B-T1]. General results on the decoupling tool may be found in [K-S], [Kw4], [MC-T1], [MC-T2], [Zi3], etc.

The comparison Theorem 4.12 is due to the second named author and first appeared in [L-T4] (the proof presented here being simpler). Proposition 4.13 and Theorem 4.15 are recent results of the second author while Corollary 4.14 is essentially in [C-P] (see also [Pa] for some earlier related results). (4.27) and the fact that it is best possible belongs to the folklore. Theorem 4.15 is in particular applied in [Ta18].

Chapter 5. Stable random variables

Chapter 5. Stable random variables

After Gaussian variables and Rademacher series, we investigate in this chapter another important class of random variables and vectors, namely stable random variables. Stable random variables appear as fundamental in Probability Theory and, as will be seen later, also play a role in structure theorems of Banach spaces. The literature is rather extensive on this topic and we only concentrate here on the parts of the theory which will be of interest and use to us in the sequel. In particular, we do not attempt to study stable measures in the natural more general setting of infinitely divisible distributions. We refer to [Ar-G2] and [Li] for such a study. We only concentrate here on the aspects of stable distributions analogous to those developed in the preceding chapters on Gaussian and Rademacher variables. In particular, our study is based on a most useful representation of stable random variables detailed in the first paragraph. The second one examines integrability properties and tail behavior of norms of infinite dimensional stable random variables. Finally, the last section is devoted to some comparison theorems.

We recall that, for $0 < p < \infty$, $L_{p,\infty} = L_{p,\infty}(\Omega, \mathcal{A}, \mathbb{P})$ denotes the space of all real random variables X on $(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$\|X\|_{p,\infty} = (\sup_{t>0} t^p \mathbb{P}\{|X| > t\})^{1/p} < \infty .$$

$\|\cdot\|_{p,\infty}$ is only a quasi-norm but is equivalent to a norm when $p > 1$; take for example

$$N_p(X) = \sup\{\mathbb{P}(A)^{-1/q} \int_A |X| d\mathbb{P} \,;\ A \in \mathcal{A} ,\ \mathbb{P}(A) > 0\} \tag{5.1}$$

where $q = p/p-1$ is the conjugate of p, for which we have that, for all X , $\|X\|_{p,\infty} \leq N_p(X) \leq q\|X\|_{p,\infty}$ (integration by parts). A random variable X in L_p is of course in $L_{p,\infty}$, satisfying even $\lim_{t\to\infty} t^p \mathbb{P}\{|X| > t\} = 0$. Conversely, the space of all random variables having this limit 0 is the closure in the $L_{p,\infty}$ -norm of the step random variables. Recall also the comparisons with the L_r -norms: for every $r > p$ and X

$$\|X\|_{p,\infty} \leq \|X\|_p \leq \left(\frac{r}{r-p}\right)^{1/p} \|X\|_{r,\infty} .$$

Finally, if B is a Banach space, we denote by $L_{p,\infty}(B)$ the space of all random variables X in B such that $\|X\| \in L_{p,\infty}$; we let simply $\|X\|_{p,\infty} = \| \, \|X\| \, \|_{p,\infty}$.

As in the Gaussian case, we only consider *symmetric* stable random variables. A real valued (symmetric) random variable X is called p-stable, $0 < p \leq 2$, if, for some $\sigma \geq 0$, its Fourier transform is of the form

$$\mathbb{E} \exp itX = \exp(-\sigma^p |t|^p/2) , \quad t \in \mathbb{R} .$$

$\sigma = \sigma_p = \sigma_p(X)$ is called the *parameter* of the stable random variable X with *index* p. A 2-stable random variable with parameter σ is just Gaussian with variance σ^2. If $\sigma = 1$, X is called *standard.* As for Gaussians, when we speak of a *standard* p*-stable sequence* (θ_i), we always mean a sequence of independent standard p-stable random variables θ_i (the index p will be clear from the context).

Despite the analogy in the definition, the case $p = 2$ corresponding to Gaussian variables and the case $0 < p < 2$ present some quite important differences. For example, while stable distributions have densities, these cannot be easily expressed in general. Further, if Gaussian variables have exponential moments, a non-zero p-stable random vari able X, $0 < p < 2$, is not even in L_p. It can however be shown that $\|X\|_{p,\infty} < \infty$ and actually that

$$\lim_{t\to\infty} t^p \mathbb{P}\{|X| > t\} = c_p^p \sigma^p \tag{5.2}$$

where σ is the parameter of X and $c_p > 0$ only depends on p (cf. e.g. [Fe1]). X has therefore moments of order r for every $r < p$ and $\|X\|_r = c_{p,r}\sigma$ where $c_{p,r}$ depends on p and r only.

Stable random variables are characterized by their fundamental "stability" property (from which they draw their name): if (θ_i) is a standard p-stable sequence, for any finite sequence (α_i) of real numbers, $\sum\limits_i \alpha_i\theta_i$ has the same distribution as $\left(\sum\limits_i |\alpha_i|^p\right)^{1/p} \theta_1$. By what preceeds, in particular, for any $r < p$,

$$\left\| \sum_i \alpha_i\theta_i \right\|_r = c_{p,r} \left(\sum_i |\alpha_i|^p \right)^{1/p}$$

so that the span in L_r, $r < p$, of (θ_i) is isometric to ℓ_p. This property, which is analogous to what we learned in the Gaussian case but with a smaller spectrum in r, is of fundamental interest in the study of ℓ_p^n-subspaces of Banach spaces (cf. Chapter 9). In this order of ideas and among the consequences of (5.2) , we would like to note further that if (θ_i) is a standard p-stable sequence with $0 < p < 2$ and (α_i) a sequence of real numbers such that $\sup\limits_i |\alpha_i\theta_i| < \infty$ almost surely, then $\sum\limits_i |\alpha_i|^p < \infty$. Indeed, we have from the Borel-Cantelli lemma (cf. Lemma 2.6) that for some $M' > 0$

$$\sum_i \mathbb{P}\{|\alpha_i\theta_i| > M'\} < \infty .$$

It already follows that (α_i) is bounded, i.e. $\sup\limits_i |\alpha_i| \le M''$ for some M''. By (5.2) there exists t_0 such that for all $t \ge t_0$

$$\mathbb{P}\{|\theta_1| > t\} \ge \frac{1}{2c_p^p t^p} .$$

Hence, letting $M = \max(M', t_0 M'')$, we have that

$$\infty > \sum_i \mathbb{P}\{|\alpha_i \theta_i| > M\} \geq \frac{1}{2M^p c_p^p} \sum_i |\alpha_i|^p .$$

This kind of result is of course completely different in the case $p = 2$ for which we recall (see (3.7)), as an example and for the matter of comparison, that if (g_i) is an orthogaussian sequence,

$$\limsup_{i \to \infty} \frac{|g_i|}{(2 \log(i+1))^{1/2}} = 1 \quad \text{almost surely} .$$

A random variable $X = (X_1, \ldots, X_N)$ with values in $\mathbb{R}^N$ is p-stable if each linear combination $\sum_{i=1}^N \alpha_i X_i$ is a real p-stable variable. A random process $X = (X_t)_{t \in T}$ indexed by a set T is called p-stable if for every $t_1, \ldots, t_N$ in T , $(X_{t_1}, \ldots, X_{t_N})$ is a p-stable random vector. Similarly, a Radon random variable X with values in a Banach space B is p-stable if $f(X)$ is p-stable for every f in B' . By their very definition, all these p-stable random vectors satisfy the fundamental and characteristic stability property of stable distributions: if, and only if, X is p-stable, if X_i are independent copies of X , then

$$\sum_i \alpha_i X_i \ \text{ has the same distribution as } \left(\sum_i |\alpha_i|^p \right)^{1/p} X$$

for every finite sequence (α_i) of real numbers.

It will almost always be assumed in the sequel that $0 < p < 2$. The case $p = 2$ corresponding to Gaussian variables was investigated previously.

5.1. Representation of stable random variables

p-stable ($0 < p < 2$) finite or infinite dimensional random variables can be given (in distribution) a series representation. This representation can be though as some central limit theorem with stable limits and we will actually have the opportunity to verify this observation in the sequel. This representation is a most valuable tool in the study of stable random variables; it almost allows to think of stable variables as sums of nicely behaved independent random variables for which a large variety of tools is available. We use it almost automatically each time we deal with stable distributions.

To introduce this representation we first investigate the scalar case. We need some notations. Let (λ_i) be independent random variables with common exponential distribution $\mathbb{P}\{\lambda_i > t\} = e^{-t}$, $t \geq 0$. Set

$\Gamma_j = \sum_{i=1}^{j} \lambda_i$, $j \geq 1$, which will always have the same meaning throughout the book. The sequence $(\Gamma_j)_{j\geq 1}$ defines the successive times of the jumps of a standard Poisson process (cf. [Fe1]). As is easy to see,

$$\mathbb{P}\{\Gamma_j \leq t\} = \int_0^t \frac{x^{j-1}}{(j-1)!} e^{-x}\, dx\,, \quad t \geq 0\,.$$

In particular,

$$\mathbb{P}\{\Gamma_j^{-1/p} > t\} \leq \frac{1}{j!} \cdot \frac{1}{t^{pj}}$$

for all $0 < p < \infty$, $j \geq 1$ and $t > 0$. It already follows that while

$$\lim_{t\to\infty} t^p \mathbb{P}\{\Gamma_1^{-1/p} > t\} = 1\,, \tag{5.3}$$

for $j \geq 2$ we have that

$$\|\Gamma_j^{-1/p}\|_p < \infty \tag{5.4}$$

(actually $\|\Gamma_j^{-1/p}\|_r < \infty$ for any $r < pj$), hence $\lim_{t\to\infty} t^p \mathbb{P}\{\Gamma_j^{-1/p} > t\} = 0$.

By the strong law of large numbers, $\Gamma_j/j \to 1$ almost surely. A powerful method will be to replace Γ_j by j which is non random. We already quote at this stage a first observation: for any $\alpha > 0$ and $j > \alpha$,

$$\mathbb{E}(\Gamma_j^{-\alpha}) = \frac{\Gamma(j-\alpha)}{\Gamma(j)} \sim \frac{1}{j^\alpha} \qquad \text{for } j > \alpha \tag{5.5}$$

as can easily be seen from Stirling's formula (Γ is the gamma function).

Provided with these easy observations, the series representation of p-stable random variables can be formulated as follows.

Theorem 5.1. Let $0 < p < 2$ and let η be a symmetric real random variable such that $\mathbb{E}|\eta|^p < \infty$. Denote further by (η_j) independent copies of η assumed to be independent from the sequence (Γ_j) . Then, the almost surely convergent series

$$X = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \eta_j$$

defines a p-stable random variable with parameter $\sigma = c_p^{-1} \|\eta\|_p$ (where c_p has been introduced in (5.2)).

Proof. Let us first convince ourselves that the sum defining X is almost surely convergent. To this aim, we prove a little bit more than necessary but which will be useful later in this proof. Let us show indeed that

$$\lim_{j_0\to\infty}\sup_{N\ge j_0}\left\|\sum_{j=j_0}^{N}\Gamma_j^{-1/p}\eta_j\right\|_{p,\infty}=0\,. \tag{5.6}$$

If this is satisfied, the sum $\sum\limits_{j=1}^{\infty}\Gamma^{-1/p}\eta_j$ is in particular convergent in all L_r, $r<p$, and thus in probability. Since $(\Gamma_j^{-1/p}\eta_j)$ is a symmetric sequence, the series converges almost surely by Itô-Nisio's theorem. Let us thus establish (5.6). An alternate proof making use of some of the material introduced later in this chapter is given at the end of Section 5.2. For every $t>0$, $2\le j_0\le N$,

$$\mathbb{P}\left\{\left|\sum_{j=j_0}^{N}\Gamma_j^{-1/p}\eta_j\right|>t\right\}\le\mathbb{P}\{\exists j\ge j_0\,,\ |\eta_j|>tj^{1/p}\}+\frac{1}{t^2}\sum_{j\ge j_0}\mathbb{E}(\Gamma_j^{-2/p}\eta_j^2 I_{\{|\eta_j|\le tj^{1/p}\}})\,.$$

Clearly

$$\mathbb{P}\{\exists j\ge j_0\,,\ |\eta_j|>tj^{1/p}\}\le\sum_{j\ge j_0}\mathbb{P}\{|\eta|>tj^{1/p}\}\le\frac{1}{t^p}\mathbb{E}(|\eta|^p I_{\{|\eta|>tj_0^{1/p}\}})$$

while, by (5.5), provided j_0 is large enough, for some constants C,C',

$$\begin{aligned}\frac{1}{t^2}\sum_{j\ge j_0}\mathbb{E}(\Gamma_j^{-2/p}\eta_j^2 I_{\{|\eta_j|\le tj^{1/p}\}})&\le\frac{C}{t^2}\mathbb{E}\left(|\eta|^2\sum_{j\ge j_0}j^{-2/p}I_{\{j\ge|\eta|^p/t^p\}}\right)\\&\le\frac{C'}{t^2}\cdot\frac{1}{j_0^{(2/p)-1}}\mathbb{E}(|\eta|^2 I_{\{|\eta|\le tj_0^{1/p}\}})+C'\le vert^p\mathbb{E}(|\eta|^p I_{\{|\eta|>tj_0^{1/p}\}})\,.\end{aligned}$$

The conclusion now follows: for every $\varepsilon>0$, $a>0$ and j_0 large enough, we have obtained that

$$\sup_{N\ge j_0}\left\|\sum_{j=j_0}^{N}\Gamma_j^{-1/p}\eta_j\right\|_{p,\infty}^{p}\le\varepsilon^p+\frac{C'a^2}{\varepsilon^{2-p}j_0^{(2/p)-1}}+C'\mathbb{E}(|\eta|^p I_{\{|\eta|>a\}})+(C'+1)\mathbb{E}(|\eta|^p I_{\{|\eta|>\varepsilon j_0^{1/p}\}})\,.$$

Since $\mathbb{E}|\eta|^p<\infty$, we can let j_0 tend to infinity, then a also, and then ε to 0 to get the conclusion.

In order to establish the theorem, we show that X satisfies the characteristic property of p-stable random variables, namely that if X_1 and X_2 are independent copies of X, for all real numbers α_1,α_2,

$$\alpha_1X_1+\alpha_2X_2\ \text{ has the same distribution as }(|\alpha_1|^p+|\alpha_2|^p)^{1/p}X\,.$$

Write $X_i = \sum_{j=1}^{\infty} \Gamma_{ji}^{-1/p} \eta_{ji}$, $i = 1, 2$, where $\{(\Gamma_{ji})_{j \geq 1}, (\eta_{ji})_{j \geq 1}\}$ for $i = 1, 2$ are independent copies of $\{(\Gamma_j)_{j \geq 1}, (\eta_j)_{j \geq 1}\}$. Set $a_i = |\alpha_i|^p$. Consider the non-decreasing rearrangement $\{\gamma_j^* \, ; \, j \geq 1\}$ of the countable set $\{\Gamma_{ji}/a_i \, ; \, j \geq 1 \, , \, i = 1, 2\}$. The sequence $\{\Gamma_{ji}/a_i \, ; \, j \geq 1\}$ corresponds to the successive times of the jumps of a Poisson process N^i of parameter a_i, $i = 1, 2$. It is easily seen that $\{\gamma_j^* \, ; \, j \geq 1\}$ corresponds then to the sequence of the successive times of the jumps of the process $N^1 + N^2$. But $N^1 + N^2$ is a Poisson process of parameter $a_1 + a_2$. Hence $\{\gamma_j^* \, ; \, j \geq 1\}$ has the same distribution as the sequence $\{\Gamma_j/(a_1 + a_2) \, ; \, j \geq 1\}$. We have therefore the following equalities in distribution:

$$\alpha_1 X_1 + \alpha_2 X_2 = \sum_{j=1}^{\infty} (\gamma_j^*)^{-1/p} \eta_j = (a_1 + a_2)^{1/p} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \eta_j = (|\alpha_1|^p + |\alpha_2|^p)^{1/p} X \, .$$

Hence X is p-stable. The final step of the proof consists in showing that X has parameter $c_p^{-1} |\eta|_p$. To this aim we identify the limit (5.2) and make use of the fact established in the first part of this proof. We have indeed from (5.6) and (5.4) that

$$\lim_{t \to \infty} t^p \mathbb{P} \left\{ \left| \sum_{j=2}^{\infty} \Gamma_j^{-1/p} \eta_j \right| > t \right\} = 0 \, .$$

Now from (5.3) and independence, we see that

$$\lim_{t \to \infty} t^p \mathbb{P} \{ \Gamma_1^{-1/p} |\eta_1| > t \} = \mathbb{E} |\eta|^p \, .$$

Hence combining these observations yields

$$\lim_{t \to \infty} t^p \mathbb{P} \{ |X| > t \} = \mathbb{E} |\eta|^p$$

and by comparison with (5.2) we indeed get that X has parameter $c_p^{-1} |\eta|_p$. The proof of Theorem 5.1 is complete.

After the scalar case we now attack the case of infinite dimensional stable random vectors and processes. The key tool in this investigation is the concept of *spectral measure.* The spectral measure of a stable variable arises from the general theory of Lévy-Khintchine representations of infinitely divisible distributions. We do no follow here this approach but rather outline, for the modest purposes of our study, a somewhat weaker but simple description of spectral measures of stable distributions in infinite dimension. It will cover the applications we have in mind and explains the flavor of the result. We refer to [Ar-G2] and [Li] for the more general infinitely divisible theory.

We state and prove the existence of a spectral measure of a stable distribution in the context of a random process $X = (X_t)_{t\in T}$ indexed by a countable set T in order to avoid measurability questions. This is anyway the basic result from which the necessary corollaries are easily deduced.

Theorem 5.2. Let $0 < p < 2$ an d let $X = (X_t)_{t\in T}$ be a p-stable process indexed by a countable set T. There exists a positive finite measure m on $\mathbb{R}^T$ (equipped with its cylindrical σ-algebra) such that for every finite sequence (α_j) of real numbers

$$\mathbb{E}\exp i\sum_j \alpha_j X_{t_j} = \exp\left(-\frac{1}{2}\int_{\mathbb{R}^T}\left|\sum_j \alpha_j x_{t_j}\right|^p dm(x)\right).$$

m is called a spectral measure of X (it is not necessarily unique).

Before the proof, let us just mention that in the case $p = 2$ we can simply take for m the distribution of the Gaussian process X.

Proof. In a first step, assume that T is a finite set $\{t_1, \ldots, t_N\}$. Recall that if θ is real p-stable with parameter σ, and $r < p$, then $\|\theta\|_r = c_{p,r}\sigma$. It follows that for every $\alpha = (\alpha_1, \ldots, \alpha_N)$ in $\mathbb{R}^N$,

$$\begin{aligned}\mathbb{E}\exp i\sum_{j=1}^N \alpha_j X_{t_j} &= \exp\left(-\frac{1}{2}\sigma(\alpha)^p\right)\\ &= \exp\left[-\frac{1}{2c_{p,r}^p}\left(\mathbb{E}\left|\sum_{j=1}^N \alpha_j X_{t_j}\right|^r\right)^{p/r}\right]\end{aligned}$$

where $\sigma(\alpha)$ denotes the parameter of $\sum_{j=1}^N \alpha_j X_{t_j}$. For every $r < p$, define then a positive finite measure m_r on the unit sphere S for the sup norm $\|\cdot\|$ on $\mathbb{R}^N$ by setting for every bounded measurable function φ on S

$$\int_S \varphi(y)dm_r(y) = \frac{1}{c_{p,r}^r}\int_{\mathbb{R}^N}\varphi\left(\frac{x}{\|x\|}\right)\|x\|^r d\mathbb{P}_X(x)$$

where $\mathbb{P}_X$ is the law of $X = (X_{t_1}, \ldots, X_{t_N})$. Hence, for any $\alpha = (\alpha_1, \ldots, \alpha_N)$ in $\mathbb{R}^N$,

$$\mathbb{E}\exp i\sum_{j=1}^N \alpha_j X_{t_j} = \exp\left[-\frac{1}{2}\left(\int_S\left|\sum_{j=1}^N \alpha_j x_j\right|^r dm_r(x)\right)^{p/r}\right].$$

Now, the total mass $|m_r|$ of m_r is easily seen to be majorized by

$$|m_r| \leq \left(\inf_{x \in S} \sum_{j=1}^{N} |x_j|^r \right)^{-1} \sum_{j=1}^{N} \sigma(e_j)^r$$

where e_j , $1 \leq j \leq N$, are the unit vectors of $\mathbb{R}^N$. Therefore $\sup_{r<p} |m_r| < \infty$. Let then m be a cluster point (in the weak-star sense) of $(m_r)_{r<p}$; m is a positive finite measure which is clearly a spectral measure of X . This proves the theorem in this finite dimensional case.

Assume now that $T = \{t_1, t_2, \dots\}$. It is not difficult to see that we may assume the stable process X be almost surely bounded. Indeed, if this is not the case, by the integrability property (5.2) and the Borel-Cantelli lemma, there exists a sequence $(a_t)_{t \in T}$ of positive numbers such that if $Z_t = a_t X_t$, the p-stable process $Z = (Z_t)_{t \in T}$ satisfies $\sup_{t \in T} |Z_t| < \infty$ almost surely. If we can then construct a spectral measure m' on $\mathbb{R}^T$ for Z , define m in such a way that for any bounded measurable function φ on $\mathbb{R}^T$ depending only on finitely many coordinates,

$$\int_{\mathbb{R}^T} \varphi(x) dm(x) = \int_{\mathbb{R}^T} \varphi \left(\left(\frac{x_t}{a_t} \right) \right) dm'(x)$$

where $x = (x_t) \in \mathbb{R}^T$. Then the positive finite measure m on $\mathbb{R}^T$ is a spectral measure for the p-stable process X . We therefore assume that X is almost surely bounded. For each N , the preceding finite dimensional step provides us with a spectral measure m_N concentrated on the unit sphere of ℓ_∞^N of the random vector $(X_{t_1}, \dots, X_{t_N})$ in $\mathbb{R}^N$. Denote by (Y_j^N) independent random variables distributed like $m_N / |m_N|$. If we recall the sequence (Γ_j) of the representation and if we let (ε_j) denote a Rademacher sequence, then, the sequences (Y_j^N) , (Γ_j) , (ε_j) being assumed independent, Theorem 5.1 indicates that $(X_{t_1}, \dots, X_{t_N})$ has the same distribution as

$$c_p |m_N|^{1/p} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j Y_j^N .$$

Our next step is to show that $\sup_N |m_N| < \infty$. Since X is assumed to be almost surely bounded, we can choose to this aim a finite number u such that $\mathbb{P}\{\sup_{t \in T} |X_t| > u\} \leq 1/4$. By Lévy's inequality (2.7) and the preceding representations, it follows that, for every N ,

$$\begin{aligned} \frac{1}{2} \geq 2\mathbb{P}\{\sup_{t \in T} |X_t| > u\} &\geq 2\mathbb{P}\{\max_{i \leq N} |X_{t_i}| > u\} \\ &\geq \mathbb{P}\{c_p |m_N|^{1/p} \Gamma_1^{-1/p} > u\} . \end{aligned}$$

We deduce that $|m_N| \le c_p^{-p} u^p$ and thus that $\sup_N |m_N| < \infty$ since u has been chosen independently of N. As before, if m denotes then a cluster point (in the weak-star sense) of the bounded sequence (m_N) of positive measures, m is immediately seen to fulfil the conclusion of Theorem 5.2; m is a spectral measure of X and the proof is complete.

As an immediate corollary to Theorems 5.1 and 5.2 we can now state the following:

Corollary 5.3. Let $0 < p < 2$ and let $X = (X_t)_{t \in T}$ be a p-stable random process indexed by a countable set T with spectral measure m. Let (Y_j) be a sequence of independent random variables distributed like $m/|m|$ (in $\mathbb{R}^T$). Let further (η_j) be real independent symmetric random variables with the same law as η where $\mathbb{E}|\eta|^p < \infty$ and assume the sequences (Y_j), (η_j), (Γ_j) to be independent. Then, the random process $X = (X_t)_{t \in T}$ has the same distribution as

$$\left(c_p \|\eta\|_p^{-1} |m|^{1/p} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \eta_j Y_j(t) \right)_{t \in T} .$$

Remark 5.4. If m is a spectral measure of an *almost surely bounded* p-stable process $X = (X_t)_{t \in T}$, then necessarily

$$\int \|x\|^p dm(x) < \infty$$

where we have denoted by $\|\cdot\|$ the $\ell_\infty(T)$-norm. This can be seen for example from the preceding representation; indeed, by Lévy's inequalities applied conditionally on (Γ_j), we must have that $\sup_{j \ge 1} \|Y_j\| / \Gamma_j^{1/p} < \infty$ almost surely. Now recall that from the strong law of large numbers $\Gamma_j / j \to 1$ with probability one and therefore we also have that $\sup_{j \ge 1} \|Y_j\| / j^{1/p} < \infty$. The claim thus follows from the Borel-Cantelli lemma and the fact that the independent random variables Y_j are distributed like $m/|m|$. Actually, a close inspection of the proof of Theorem 5.2 shows that for a bounded process we directly constructed a spectral measure concentrated on the unit ball of $\ell_\infty(T)$. It is in fact convenient in many problems to work with a spectral measure concentrated on the unit sphere (or only ball) of $\ell_\infty(T)$. To this aim, observe that if m is a spectral measure of X satisfying $\int \|x\|^p dm(x) < \infty$, let m_1 be the image of the measure $\|x\|^p dm(x)$ by the map $x \to x/\|x\|$. Then m_1 is a spectral measure of X concentrated on the unit sphere of $\ell_\infty(T)$ with total mass

$$|m_1| = \int \|x\|^p dm(x) .$$

It can be shown that a *symmetric* spectral measure *on the unit sphere* of $\ell_\infty(T)$ is *unique* and this actually follows from (5.10) below (at least in the Radon case). This uniqueness is however rather irrelevant for our purposes. By analogy with the scalar case, we can however define the *parameter* of X by

$$\sigma_p(X) = |m_1|^{1/p} = \left(\int \|x\|^p dm(x)\right)^{1/p} . \tag{5.7}$$

Note that by uniqueness of m_1, $\sigma_p(X)$ is well defined (cf. also (5.11)). The terminology of parameter extends the real case. This parameter plays sometimes roles analogous to the σ's encountered in the study of Gaussian and Rademacher variables; it is however quite different in nature as will become apparent later on.

We now inspect the consequences of the preceding results in the case of a p-stable Radon random variable in a Banach space.

Corollary 5.5. Let X be a p-stable ($0 < p < 2$) Radon random variable with values in a Banach space B. Then there is a positive finite Radon measure m on B satisfying $\int \|x\|^p dm(x) < \infty$ such that for every f in B'

$$\mathbb{E}\exp if(X) = \exp\left(-\frac{1}{2}\int_B |f(x)|^p dm(x)\right) .$$

Further, if (Y_j) is a sequence of independent random variables distributed like $m/|m|$, and (η_j) a sequence of real symmetric random variables with the same law as η where $\mathbb{E}|\eta|^p < \infty$, the series

$$c_p|\eta|_p^{-1}|m|^{1/p}\sum_{j=1}^{\infty}\Gamma_j^{-1/p}\eta_j Y_j$$

converges almost surely in B where, as usual, the sequences (Γ_j), (η_j) and (Y_j) are assumed to be independent from each other and is distributed as X.

Proof. We may and do assume B to be separable. Let D be countable weakly dense in the unit ball of B'. By Corollary 5.3, there exists a positive finite measure M on the unit ball of $\ell_\infty(D)$ such that if $(Y_j^f)_{f\in D}$ are independent and distributed like $m/|m|$, and independent of (Γ_j) and (η_j), $(f(X))_{f\in D}$ has the same distribution as

$$\left(c_p\|\eta\|_p^{-1}|m|^{1/p}\sum_{j=1}^{\infty}\Gamma_j^{-1/p}\eta_j Y_j^f\right)_{f\in D} .$$

Let (x_n) be a dense sequence in B and denote, for each n, by F_n the subspace generated by $x_1, \dots, x_n$. By Lévy's inequality (2.7) applied to the norms $\inf_{z \in F_n} \sup_{f \in D} |\cdot - f(z)|$, and conditionally on the sequence (Γ_j), we get that for all n and $\varepsilon > 0$,

$$2\mathbb{P}\{\inf_{z \in F_n} \|X - z\| > \varepsilon\} \geq \mathbb{P}\{c_p \|\eta\|_p^{-1} |m|^{1/p} \inf_{z \in F_n} \sup_{f \in D} |\Gamma_1^{-1/p} \eta_1 Y_1^f - z| > \varepsilon\}.$$

Since X is a Radon variable in B, the left hand side of this inequality can be made, for every $\varepsilon > 0$, arbitrarily small for all n large enough. It easily follows that we can define a random variable with values in B, call it Y, such that $f(Y) = Y_1^f$ almost surely for all f in D. The same argument as before through Lévy's inequalities indicates that Y is Radon. By density of D, the law of Y is, up to a multiplicative factor, a spectral measure of X and we have, by Remark 5.4, that $\mathbb{E}\|Y\|^p < \infty$. The convergence of the series representation indeed takes place in B by the Itô-Nisio theorem (Theorem 2.4). Corollary 5.5 is therefore established.

As in (5.7) we define the parameter $\sigma_p(X) = \left(\int \|x\|^p dm(x)\right)^{1/p}$. We can choose further (following Remark 5.4) the spectral measure to be symmetrically distributed on the unit sphere of B; it is then unique (cf. (5.10) below). A typical example of a Radon p-stable random variable with values in a Banach space B is of course given by an almost sure convergent series $X = \sum_i \theta_i x_i$ where (θ_i) is a standard p-stable sequence and (x_i) a sequence in B. In this case, the spectral measure is discrete and can be explicitely described. For example, since necessarily $\sup_i \|\theta_i x_i\| < \infty$ almost surely, we learned in the beginning of this chapter that, when $0 < p < 2$, $\sum_i \|x_i\|^p < \infty$. Let then m be given by

$$m = \sum_i \frac{\|x_i\|^p}{2} (\delta_{\frac{x_i}{\|x_i\|}} + \delta_{-\frac{x_i}{\|x_i\|}}).$$

Then m is a spectral measure for X (symmetric and concentrated on the unit sphere of B). We note in this case that the parameter $\sigma_p(X)$ of X (cf. (5.7)) is simply

$$\sigma_p(X) = \left(\sum_i \|x_i\|^p\right)^{1/p} < \infty.$$

This property induces a rather deep difference with the Gaussian situation in which $\sum_i \|x_i\|^2$ is not necessarily finite if $\sum_i g_i x_i$ converges. As yet another difference, note that if convergent series $\sum_i g_i x_i$ completely describe the class of Gaussian Radon random vectors, this is no more the case when $p < 2$ (as soon as the spectral measure is no more discrete). The series representation might then be thought as a kind of substitute to this property.

The representation theorems were described with a sequence (η_j) of independent identically distributed real symmetric random variables with $\mathbb{E}|\eta_j|^p < \infty$. Two choices are of particular interest. First is the simple case of a Rademacher sequence. A second choice is an orthogaussian sequence. It then appears that p-stable vectors and processes may be seen as conditionally Gaussian. Various Gaussian results can then be used to yield, after integration, similar consequences for stables. This main idea, and the two preceding choices, will be used extensively in the subsequent study of p-stable random variables and processes.

To conclude this section we would like to come back to the comparison of (Γ_j) with (j) initiated in the beginning. The representation now clearly indicates what kind of properties would be desirable.

First, as an easy consequence of $\Gamma_j/j \to 1$ almost surely and the contraction principle (conditionally) in the form of Theorem 4.4, it is plain that, in the previous notations, $\sum_{j=1}^{\infty} \Gamma_j^{-1/p}\eta_j Y_j$ converges almost surely if and only if $\sum_{j=1}^{\infty} j^{-1/p}\eta_j Y_j$ does. The next two observations will be used as quantitative versions of this result. As a simple consequence of the expression of $\mathbb{P}\{\Gamma_j \le t\}$ and Stirling's formula, we have that

$$\left\| \sup_{j\ge 1} \left(\frac{j}{\Gamma_j}\right)^{1/p} \right\|_{p,\infty} \le K_p \tag{5.8}$$

for some $K_p < \infty$ and similarly with $(\Gamma_j/j)^{1/p}$ (for which actually all moments exist). More important perhaps is the following:

$$\sum_{j\ge 2} \|\Gamma_j^{-1/p} - j^{-1/p}\|_p^{p\wedge 1} < \infty \,. \tag{5.9}$$

Note that the sum in (5.9) is up from 2 in order for $\Gamma_j^{-1/p}$ to be in L_p. It suffices to show that for all j large enough

$$\mathbb{E}|\Gamma_j^{-1/p} - j^{-1/p}|^p \le K_p \frac{1}{j^{p/2+1}} \,.$$

We have that

$$\mathbb{E}|\Gamma_j^{-1/p} - j^{-1/p}|^p \le \frac{1}{j}\mathbb{P}\{\Gamma_j - j > j\} + \int_{\{\Gamma_j \le 2j\}} \Gamma_j^{-1} \left| 1 - \left(\frac{\Gamma_j}{j}\right)^{1/p} \right|^p d\mathbb{P} \,.$$

Now, if $0 \le x \le 2$, $|1-x^{1/p}| \le K_p|1-x|$, so that, by Chebyshev's and Hölder's inequalities the preceding is bounded by

$$\frac{1}{j^2} + K_p^p \left(\mathbb{E}\left|1 - \frac{\Gamma_j}{j}\right|^2 \right)^{p/2} (\mathbb{E}(\Gamma_j^{2/(p-2)}))^{(2-p)/2} = \frac{1}{j^2} + K_p^p \cdot \frac{1}{j^{p/2}} (\mathbb{E}(\Gamma_j^{2/(p-2)}))^{(2-p)/2} .$$

By (5.5), the claim follows.

5.2. Integrability and tail behavior

We investigate the integrability properties of p-stable Radon random variables and almost sure bounded processes, $0 < p < 2$. We already know from the real case the severe limitations compared to the Gaussian case $p = 2$. This study could be based entirely on the representation and the results of the next chapter on sums of independent random variables substituting, as we just learned it, j to Γ_j. There is however a first a priori simple result which will be convenient to record. This is Proposition 5.6 below. We then use the representation, combined with some results of the next chapter, for some precise information on tail behaviors.

As usual, in order to unify our statement on Radon random variables and bounded processes, let us assume we are given a Banach space B with a countable subset D in the unit ball of B' such that $\|x\| = \sup_{f \in D} |f(x)|$ for all x in B. X is a random variable with values in B if $f(X)$ is measurable for every f in D. It is p-stable if each finite linear combination $\sum_i \alpha_i f_i(X)$, $\alpha_i \in \mathbb{R}$, $f_i \in D$, is a real p-stable random variable.

Proposition 5.6. Let $0 < p < 2$ and let X be a p-stable random variable in B. Then $\|X\|_{p,\infty} < \infty$. Furthermore, all the moments of X of order $r < p$ are equivalent, and equivalent to $\|X\|_{p,\infty}$, i.e. for every $r < p$, there exists $K_{p,r}$ such that for every p-stable variable X

$$K_{p,r}^{-1} \|X\|_r \le \|X\|_{p,\infty} \le K_{p,r} \|X\|_r .$$

Proof. As in the previous chapters, we show that the moments of X are controled by some parameter in the $L_0(B)$ topology. Let indeed t_0 be such that $\mathbb{P}\{\|X\| > t_0\} \le 1/4$. Let (X_i) be independent copies of X. Since, for each N,

$$\frac{1}{N^{1/p}} \sum_{i=1}^N X_i \quad \text{has the same distribution as } X ,$$

we get from Lévy's inequality (2.7) that

$$\frac{1}{4} \geq \mathbb{P}\{\|X\| > t_0\} \geq \mathbb{P}\left\{\left\|\sum_{i=1}^{N} X_i\right\| > t_0 N^{1/p}\right\}$$
$$\geq \frac{1}{2}\mathbb{P}\{\max_{i\leq N}\|X_i\| > t_0 N^{1/p}\}\,.$$

By Lemma 2.6 and identical distribution it follows that

$$\mathbb{P}\{\|X\| > t_0 N^{1/p}\} \leq \frac{1}{N}$$

which therefore holds for every $N \geq 1$. By a trivial interpolation, $\|X\|_{p,\infty} \leq 2^{1/p}t_0$. To show the moment equivalences, simply note that for $0 < r < p$, if $t_0 = (4\mathbb{E}\|X\|^r)^{1/r}$, then

$$\mathbb{P}\{\|X\| > t_0\} \leq \frac{1}{t_0^r}\mathbb{E}\|X\|^r \leq \frac{1}{4}\,.$$

As a consequence of this proposition, we see that if (X_n) is a p-stable sequence of random variables converging almost surely to X, or only in probability, then (X_n) also converges in $L_{p,\infty}$ and therefore in L_r for every $r < p$. This follows from the preceding moment equivalences together with Lemma 4.2 or directly from the proof of Proposition 5.6; indeed, for every $\varepsilon > 0$, $\mathbb{P}\{\|X_n - X\| > \varepsilon\}$ can be made smaller than $1/4$ for all n large enough and then $\|X_n - X\|_{p,\infty} \leq 2^{1/p}\varepsilon$.

Integrability properties of infinite dimensional p-stable random vectors are thus similar to the finite dimensional ones. This observation can be pushed further to obtain that (5.2) also extends. The proof is based on the representation and mimicks the last argument in the proof of Theorem 5.1. For notational convenience and simplicity in the exposition, we present this result in the setting of Radon random variables but everything goes through to the case of almost surely bounded stable processes.

Let therefore X be a p-stable Radon random variable with values in a Banach space B. According to Corollary 5.5 (and Remark 5.4), let m be a spectral measure of X symmetrically distributed on the unit sphere of B. Then, for every measurable set A in the unit sphere of B such that $m(\partial A) = 0$ where ∂A is the boundary of A,

$$\lim_{t\to\infty} t^p \mathbb{P}\left\{\|X\| > t\,,\ \frac{X}{\|X\|} \in A\right\} = c_p^p m(A)\,. \tag{5.10}$$

This shows in particular uniqueness of such a spectral measure as announced in Remark 5.4. If we recall the parameter $\sigma_p(X) = |m|^{1/p}$ of X (cf. (5.7)), we have in particular that

$$\lim_{t\to\infty} t^p \mathbb{P}\{\|X\| > t\} = c_p^p \sigma_p(X)^p \,. \tag{5.11}$$

To better describe the idea of the proof of (5.10), let us first establish the particular case (5.11). We use a result of the next chapter on sums of independent random variables. Let (Y_j) be independent random variables distributed like $m/|m|$. According to Corollary 5.5, X has the same distribution as $c_p|m|^{1/p} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} Y_j$. The main observation is that $\mathbb{E}\|\sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j\|^p < \infty$. By (5.9), it is enough to have $\mathbb{E}\|\sum_{j=2}^{\infty} j^{-1/p} Y_j\|^p < \infty$. But then we are dealing with a sum of independent random variables. Anticipating on the next chapter, we may invoke Theorem 6.11 to see that indeed $\mathbb{E}\|\sum_{j=2}^{\infty} j^{-1/p} Y_j\|^p < \infty$. Hence it follows that

$$\lim_{t\to\infty} t^p \mathbb{P}\left\{ \left\| \sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j \right\| > t \right\} = 0 \,. \tag{5.12}$$

Since Y_1 is concentrated on the unit sphere of B, combining with (5.3) we get that

$$\lim_{t\to\infty} t^p \mathbb{P}\left\{ \left\| \sum_{j=1}^{\infty} \Gamma_j^{-1/p} Y_j \right\| > t \right\} = 1 \,,$$

hence the result.

We next turn to (5.10). By homogeneity, assume that $c_p|m|^{1/p} = 1$. We only establish that for every closed subset set F of the unit sphere of B

$$\limsup_{t\to\infty} t^p \mathbb{P}\left\{ \|X\| > t \,,\ \frac{X}{\|X\|} \in F \right\} \le \mathbb{P}\{Y_1 \in F\} \,.$$

The corresponding lower bound for open sets is established similarly yielding thus (5.10) since Y_1 is distributed like $m/|m|$. For each $\varepsilon > 0$, we set $F_\varepsilon = \{x \in B \,;\, \exists y \in F \,,\ \|x-y\| \le \varepsilon\}$. Set also $Z = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} Y_j$ which has the same distribution as X. For every $\varepsilon, t > 0$

$$\begin{aligned} \mathbb{P}\left\{ \|X\| > t \,,\ \frac{X}{\|X\|} \in F \right\} &\le \mathbb{P}\left\{ \|Z\| > t \,,\ \frac{Z}{\|Z\|} \in F \,,\ \|\sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j\| \le \varepsilon t \right\} \\ &\quad + \mathbb{P}\left\{ \left\| \sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j \right\| > \varepsilon t \right\} \,. \end{aligned}$$

By (5.12) we will need only concentrate on the first probability on the right of this inequality. Assume thus that $\|Z\| > t$, $Z/\|Z\| \in F$ and $\left\| \sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j \right\| \leq \varepsilon t$. Since

$$Z = \Gamma_1^{-1/p} Y_1 + \sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j ,$$

we deduce from the triangle inequality that $\Gamma_1^{-1/p} Y_1 / \|Z\| \in F_\varepsilon$. Further

$$\left\| \frac{\Gamma_1^{-1/p} Y_1}{\|Z\|} - Y_1 \right\| \leq \left| 1 - \frac{\Gamma_1^{-1/p}}{\|Z\|} \right|$$

and since $(1-\varepsilon)t < \|Z\| - \varepsilon t \leq \Gamma_1^{-1/p} \leq \|Z\| + \varepsilon t$, we get that $Y_1 \in F_{2\varepsilon}$. Summarizing,

$$\mathbb{P}\left\{ \|X\| > t \,,\ \frac{X}{\|X\|} \in F \right\} \leq \mathbb{P}\{\Gamma_1^{-1/p} > (1-\varepsilon)t \,,\ Y_1 \in F_{2\varepsilon}\} + \mathbb{P}\left\{ \left\| \sum_{j=2}^{\infty} \Gamma_j^{-1/p} Y_j \right\| > \varepsilon t \right\} .$$

By independence of Γ_1 and Y_1 , and (5.3), (5.12), it follows that for every $\varepsilon > 0$

$$\limsup_{t \to \infty} t^p \mathbb{P}\left\{ \|X\| > t \,,\ \frac{X}{\|X\|} \in F \right\} \leq (1+\varepsilon) \mathbb{P}\{Y_1 \in F_\varepsilon\} + \varepsilon \,.$$

Since $\varepsilon > 0$ is arbitrary, this proves the claim.

Along the same line of ideas, it is possible to obtain from the representation a concentration inequality of $\|X\|$ around its expectation ($p > 1$). The argument relies on a concentration idea for sums of independent random variables presented in Section 6.3 but we already explain here the result. For simplicity, we deal as before with Radon random variables and the case $p > 1$. The case $0 < p \leq 1$ can be discussed similarly with $\mathbb{E}\|X\|^r$, $r < p$, instead of $\mathbb{E}\|X\|$.

Proposition 5.7. Let $1 < p < 2$ and let X be a p-stable Radon random variable with values in a Banach space B . Then, $\sigma_p(X)$ denoting the parameter of X , for all $t > 0$,

$$\mathbb{P}\{ |\, \|X\| - \mathbb{E}\|X\| \,| > t \} \leq C_p \frac{\sigma_p(X)^p}{t^p}$$

where $C_p > 0$ only depends on p .

Proof. Let (Y_j) be independent identically distributed random variables in the unit sphere of B such that X has the law of $c_p\sigma_p(X)Z$ where $Z = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} Y_j$ is almost surely convergent in B (Corollary 5.5). By the triangle inequality,

$$
\begin{aligned}
\big| \|Z\| - \mathbb{E}\|Z\| \big| \leq & \left| \left\| \sum_{j=1}^{\infty} j^{-1/p} Y_j \right\| - \mathbb{E} \left\| \sum_{j=1}^{\infty} j^{-1/p} Y_j \right\| \right| \\
& + \sum_{j=1}^{\infty} |\Gamma_j^{-1/p} - j^{-1/p}| + \sum_{j=1}^{\infty} \mathbb{E}|\Gamma_j^{-1/p} - j^{-1/p}| \, .
\end{aligned}
$$

By (5.9), $\sum_{j=1}^{\infty} \mathbb{E}|\Gamma_j^{-1/p} - j^{-1/p}| < \infty$, and, while only $\|\Gamma_1^{-1/p} - 1\|_{p,\infty} < \infty$,

$$
\left\| \sum_{j=2}^{\infty} |\Gamma_j^{-1/p} - j^{-1/p}| \right\|_p \leq \sum_{j=2}^{\infty} \|\Gamma_j^{-1/p} - j^{-1/p}\|_p < \infty \, .
$$

In order to estimate $\left| \| \sum_{j=1}^{\infty} j^{-1/p} Y_j \| - \mathbb{E} \| \sum_{j=1}^{\infty} j^{-1/p} Y_j \| \right|$, we can use the (martingale) quadratic inequality (6.11) to see that

$$
\mathbb{E} \left| \left\| \sum_{j=1}^{\infty} j^{-1/p} Y_j \right\| - \mathbb{E} \left\| \sum_{j=1}^{\infty} j^{-1/p} Y_j \right\| \right|^2 \leq \sum_{j=1}^{\infty} j^{-2/p} < \infty \, .
$$

Combining these various estimates, the proof is easily completed.

As a parenthesis, note that if X is a p-stable random variable with parameter $\sigma_p(X)$, applying Lévy's inequalities on the representation yields

$$
\sigma_p(X) \leq K_p \|X\|_{p,\infty} \tag{5.13}
$$

for some constant K_p depending only on p . (It is also a consequence of (5.11).) Thus, the "strong" norm of a p-stable variable always dominates its parameter $\sigma_p(X)$. Let us already mention that (5.13) is two-sided when $0 < p < 1$. This follows again from the representation together with the fact that $\sum_{j=1}^{\infty} j^{-1/p} < \infty$ when $p < 1$. We will see later how this is of course no more the case when $1 \leq p < 2$.

We conclude this paragraph with some useful inequalities for real valued independent random variables which do not seem at first sight to be connected with stable distributions. They will however be much helpful later on in the study of various questions involving stable distributions like in Chapters 9 and 13. They also allow to evalute some interesting norms of p-stable variables.

Recall that for $0 < p < \infty$ and $(\alpha_i)_{i\geq 1}$ a sequence of real numbers, we set

$$\|(\alpha_i)\|_{p,\infty} = (\sup_{t>0} t^p \text{ Card}\{i\,;\ |\alpha_i| > t\})^{1/p} = \sup_{i\geq 1} i^{1/p}\alpha_i^*$$

where $(\alpha_i^*)_{i\geq 1}$ is the non-increasing rearrangement of the sequence $(|\alpha_i|)_{i\geq 1}$. The basic inequality we present is contained in the following lemma.

Lemma 5.8. Let $0 < p < \infty$. Let (Z_i) be independent positive random variables. Then

$$\sup_{t>0} t^p \mathbb{P}\{\|(Z_i)\|_{p,\infty} > t\} \leq 2e \sup_{t>0} t^p \sum_i \mathbb{P}\{Z_i > t\}\,.$$

Proof. By homogeneity (replacing Z_i by Z_i^p) it suffices to deal with the case $p = 1$. If $(Z_i^*)_{i\geq 1}$ denotes the non-increasing rearrangement of the sequence $(Z_i)_{i\geq 1}$, $Z_n^* > u$ if and only if $\sum\limits_i I_{\{Z_i > u\}} \geq n$. Hence, if $a = \sum\limits_i \mathbb{P}\{Z_i > u\}$, by Lemma 2.5, for all $n \geq 1$,

$$\mathbb{P}\{Z_n^* > u\} \leq \left(\frac{ea}{n}\right)^n .$$

Now

$$\begin{aligned}
\mathbb{P}\{\|(Z_i)\|_{1,\infty} > t\} &= \mathbb{P}\{\sup_{n\geq 1} nZ_n^* > t\} \\
&\leq \sum_{n=1}^{\infty} \mathbb{P}\left\{Z_n^* > \frac{t}{n}\right\} \\
&\leq \sum_{n=1}^{\infty} \left(\frac{e}{n}\sum_i \mathbb{P}\left\{Z_i > \frac{t}{n}\right\}\right)^n .
\end{aligned}$$

Assuming by homogeneity that $\sup\limits_{u>0} u \sum\limits_i \mathbb{P}\{Z_i > u\} \leq 1$, we see that if $t > 2e$

$$\mathbb{P}\{\|(Z_i)\|_{1,\infty} > t\} \leq \sum_{n=1}^{\infty} \left(\frac{e}{t}\right)^n \leq \frac{2e}{t}$$

while this inequality is trivial for $t \leq 2e$. Lemma 5.7 is therefore established.

Note that the t^{-p} tail of $\mathbb{P}\{\sup\limits_{n\geq 1} n^{1/p} Z_n^* > t\}$ is actually given by the *largest* term Z_1^*. The next terms are smaller and the preceding proof indicates indeed that for all $t > 0$ and integer $k \geq 1$,

$$\mathbb{P}\{\sup_{n\geq k} n^{1/p} Z_n^* > t\} \leq \left(\frac{2e}{t^p}\right)^k \sup_{u>0} u^p \sum_i \mathbb{P}\{Z_i > u\} \tag{5.14}$$

Motivated by the representation of stable variables, the preceding lemma has an interesting consequence to the case where $Z_i = Y_i/i^{1/p}$ where (Y_i) is a sequence of independent identically distributed random variables.

Corollary 5.9. Let $0 < p < \infty$ and let Y be a positive random variable such that $\mathbb{E}Y^p < \infty$. Let (Y_i) be independent copies of Y and set $Z_i = Y_i/i^{1/p}$, $i \geq 1$. Then, if (Z_i^*) is as usual the non-increasing rearrangement of the sequence (Z_i), for every $t > 0$ and integer $n \geq 1$

$$\mathbb{P}\{n^{1/p} Z_n^* > e^{1/p} \|Y\|_p t\} \leq \frac{1}{t^{np}} .$$

Further

$$\mathbb{P}\{\sup_{n \geq k} n^{1/p} Z_n^* > (2e)^{1/p} \|Y\|_p t\} \leq \frac{1}{t^{kp}}$$

for all $k \geq 1$ and $t > 0$.

Proof. Note that for every $u > 0$

$$\sum_{i=1}^{\infty} \mathbb{P}\{Z_i > u\} = \sum_{i=1}^{\infty} \mathbb{P}\{Y > u i^{1/p}\} \leq \frac{1}{u^p} \mathbb{E}Y^p .$$

Hence the first inequality of the lemma simply follows from $\mathbb{P}\{Z_n^* > t\} \leq (ea/n)^n$ (Lemma 2.5) with $u = (e/n)^{1/p} \|Y\|_p t$. The second inequality follows similarly from Lemma 5.8 and (5.14).

Let us note that the previous simple statements yield an alternate proof of (5.6). Let us simply indicate how to establish that $\left\| \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \eta_j \right\|_{p,\infty} < \infty$ whenever $\mathbb{E}|\eta|^p < \infty$ with these tools. For every $t > 0$,

$$\mathbb{P}\left\{ \left| \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \eta_j \right| > 3t \right\} \leq \mathbb{P}\left\{ \left| \sum_{j=1}^{\infty} \eta_j (\Gamma_j^{-1/p} - j^{-1/p}) \right| > t \right\} + \mathbb{P}\left\{ \left| \sum_{j=1}^{\infty} j^{-1/p} \eta_j \right| > 2t \right\} .$$

By (5.3) and (5.9) the first probability on the right of this inequality is of the order of t^{-p}. We are thus left with $\mathbb{P}\left\{ \left| \sum_{j=1}^{\infty} j^{-1/p} \eta_j \right| > 2t \right\}$. Set $Z_j = \eta_j / j^{1/p}$, $j \geq 1$, and let ¿ (Z_j^*) be the non-increasing rearrangement of the sequence $(|Z_j|)$. Denote further by (ε_j) a Rademacher sequence independent of (η_j). Then, by symmetry and identical distribution,

$$\begin{aligned}\mathbb{P}\left\{\left|\sum_{j=1}^{\infty} j^{-1/p}\eta_j\right| > 2t\right\} &= \mathbb{P}\left\{\left|\sum_{j=1}^{\infty} \varepsilon_j Z_j^*\right| > 2t\right\} \\ &\le \mathbb{P}\{Z_1^* > t\} + \mathbb{P}\left\{\left|\sum_{j=2}^{\infty} \varepsilon_j Z_j^*\right| > t\right\} \\ &\le \mathbb{P}\{Z_1^* > t\} + \mathbb{P}\{\sup_{j\ge 2} j^{1/p} Z_j^* > \sqrt{t}\} \\ &\quad + \mathbb{P}\left\{\left|\sum_{j=2}^{\infty} \varepsilon_j Z_j^*\right| > t\,,\ \sup_{j\ge 2} j^{1/p} Z_j^* \le \sqrt{t}\right\}.\end{aligned}$$

Recall that $\mathbb{E}|\eta_j|^p < \infty$ and $Z_j = \eta_j/j^{1/p}$. By Corollary 5.9, the first two terms in the last estimate are of the order of t^{-p}. Concerning the third one, we can use for example the subgaussian inequality (4.1) conditionally on the sequence (Z_j) and find in this way a bound of the order of $\exp(-K_p t)$. This shows indeed that $\left\|\sum_{j=1}^{\infty} \Gamma_j^{-1/p}\eta_j\right\|_{p,\infty} < \infty$. The limit (5.6) can be established in the same way.

Lemma 5.8 has some further interesting consequences to the evaluation of the norm of certain stable vectors. Consider a standard p-stable sequence (θ_i) $(0 < p < 2)$ and (α_i) a finite (for simplicity) sequence of real numbers. We might be interested in $\left\|\left(\sum_i |\alpha_i\theta_i|^r\right)^{1/r}\right\|_{p,\inf ty}$, $0 < r \le \infty$, as a function of the α_i 's. In other words, we would like to examine the p-stable random variable in ℓ_r whose coordinates are $(\alpha_i\theta_i)$. Lemma 5.8 indicates to start with that

$$\| \|(\alpha_i\theta_i)\|_{p,\infty}\|_{p,\infty} \le (2e)^{1/p}\|\theta_1\|_{p,\infty}\left(\sum_i |\alpha_i|^p\right)^{1/p}\,,. \tag{5.15}$$

This inequality is in fact two-sided; we have indeed that

$$\| \sup_i |\alpha_i\theta_i|\, \|_{p,\infty} \sim \left(\sum_i |\alpha_i|^p\right)^{1/p} \tag{5.16}$$

where the equivalence sign means a two-sided inequality up to a constant K_p depending on p only. The right side follows from (5.15) while for the left one we may assume by homogeneity that $\sum_i |\alpha_i|^p = 1$; then by Lemma 2.6 and (5.2), if t is large enough and satisfies $\mathbb{P}\{\sup_i |\alpha_i\theta_i| > t\} \le 1/2$,

$$\mathbb{P}\{\sup_i |\alpha_i\theta_i| > t\} \ge \frac{1}{2}\sum_i \mathbb{P}\{|\alpha_i\theta_i| > t\} \ge (K_p t^p)^{-1}$$

from which (5.16) clearly follows.

Since for $r > p$

$$\sup_i |\alpha_i\theta_i| \leq \left(\sum_i |\alpha_i\theta_i|^r\right)^{1/r} \leq \left(\frac{r}{r-p}\right)^{1/r} \|(\alpha_i\theta_i)\|_{p,\infty} ,$$

it is plain that (5.16) extends for $r > p$ into

$$\left\| \left(\sum_i |\alpha_i\theta_i|^r\right)^{1/r} \right\|_{p,\infty} \sim \left(\sum_i |\alpha_i|^p\right)^{1/p} \tag{5.17}$$

where the equivalence is up to $K_{p,r}$ depending on p, r only. When $r < p$, we can simply use Proposition 5.6 and the moment equivalences to see that, by Fubini,

$$\left\| \left(\sum_i |\alpha_i\theta_i|^r\right)^{1/r} \right\|_{p,\infty} \sim \left(\sum_i |\alpha_i|^r\right)^{1/r} . \tag{5.18}$$

We are thus left with the slightly more complicated case $r = p$. It states that

$$\left\| \left(\sum_i |\alpha_i\theta_i|^p\right)^{1/p} \right\|_{p,\infty} \sim \left(\sum_i |\alpha_i|^p \left[1 + \log\left(\frac{(\sum_i |\alpha_i|^p)^{1/p}}{|\alpha_i|}\right)\right]\right)^{1/p} . \tag{5.19}$$

Assume by homogeneity that $\sum_i |\alpha_i|^p = 1$. For the upper bound, we note that for every $t > 0$,

$$\mathbb{P}\left\{\left(\sum_i |\alpha_i\theta_i|^p\right)^{1/p} > t\right\} \leq \mathbb{P}\left\{\sum_i |\alpha_i\theta_i|^p I_{\{|\alpha_i\theta_i| \leq t\}} > t^p\right\} + \mathbb{P}\{\sup_i |\alpha_i\theta_i| > t\} .$$

By (5.16) we need only be concerned with the first probability on the right of this inequality. By integration by parts, it is easily seen that there exists a constant K_p large enough such that, if $t^p \geq K_p \sum_i |\alpha_i|^p \left(1 + \log \frac{1}{|\alpha_i|}\right)$, then

$$\sum_i \mathbb{E}(|\alpha_i\theta_i|^p I_{\{|\alpha_i\theta_i| \leq t\}}) \leq \frac{t^p}{2} .$$

Therefore, for such t, by Chebyshev's inequality,

$$\begin{aligned} \mathbb{P}\left\{\sum_i |\alpha_i\theta_i|^p I_{\{|\alpha_i\ theta_i| \leq t\}} > t^p\right\} &\leq \mathbb{P}\left\{\sum_i |\alpha_i\theta_i|^p I_{\{|\alpha_i\theta_i| \leq t\}} - \mathbb{E}(|\alpha_i\theta_i|^p I_{\{|\alpha_i\theta_i| \leq t\}}) > \frac{t^p}{2}\right\} \\ &\leq \frac{4}{t^{2p}} \sum_i \mathbb{E}(|\alpha_i\theta_i|^{2p} I_{\{|\alpha_i\theta_i| \leq t\}}) . \end{aligned}$$

Integrating again by parts, this quantity is seen to be less than $8\|\theta_1\|_{p,\infty}^p t^{-p}$. If we set together all these informations we see that these yield the upper bound in (5.19). The lower bound is proved similarly and we thus leave it to the interested reader.

5.3. Comparison theorems

In this section, we are concerned with comparisons of stable proceses analogous to the ones described for Gaussian and Rademacher processes. It will appear that these cannot in general be extended to the stable setting. However, several interesting results are stil available. All of them are based on the observation at the end of Section 5.1, namely that stable variables can be represented as conditionally Gaussian. Gaussian techniques can then be used to yield some positive consequences for p-stable variables, $0 < p < 2$. This line of investigation will also be the key idea in Section 12.2.

We would like to start with inequality (5.13). We have noticed that this inequality is two-sided for $0 < p < 1$. In some sense therefore the study of p-stable variables with $0 < p < 1$ is not really interesting since the parameter, that is the mere existence of a spectral measure m satisfying $\int \|x\|^p dm(x) < \infty$ (cf. Remark 5.4), completely describes boundedness and size of the variable. Things are quite different when $1 \leq p < 2$ as the following example shows.

Assume that $1 < p < 2$, although the case $p = 1$ can be treated completely similarly. Consider the p-stable random variable X in $\mathbb{R}^N$ equipped with the sup-norm given by the representation $\sum_{j=1}^{\infty} \Gamma_j^{-1/p} Y_j$ where Y_j are independent and distributed like the Haar measure μ_N on $\{-1,+1\}^N$. Then $\sigma_p(X)$ of (5.7) is 1. Let us show however that $\|X\|_{p,\infty}$ is of the order of $(\log N)^{1/q}$ (for N large) where q is the conjugate of p ($\log\log N$ when $p = 1$). We only prove the lower bound, the upper bound being similar. First note that by the contraction principle

$$\mathbb{E}\left\|\sum_{j=1}^{\infty} j^{-1/p} Y_j\right\| \leq \mathbb{E}\sup_{j\geq 1}\left(\frac{\Gamma_j}{j}\right)^{1/p} \mathbb{E}\|X\| .$$

By (5.8) we know that $\mathbb{E}\sup_{j\geq 1}(\Gamma_j/j)^{1/p} \leq K_p$ for some K_p depending on p only. Let now Z be the real random variable $\sum_{j=1}^{\infty} j^{-1/p}\varepsilon_j$ where (ε_j) is a Rademacher sequence and denote by $Z_1,\dots,Z_N$ independent copies of Z. By definition, and since we consider $\mathbb{R}^N$ with the sup-norm,

$$\mathbb{E}\left\|\sum_{j=1}^{\infty} j^{-1/p} Y_j\right\| = \mathbb{E}\max_{i\leq N}|Z_i|$$

so that we have simply to bound below this maximum. For $t > 0$, let ℓ be the smallest integer such that $(\ell + 1)^{1/q} > t$. By Lévy's inequality (2.6),

$$\mathbb{P}\{|Z| > t\} \geq \frac{1}{2}\mathbb{P}\left\{\left|\sum_{j=1}^{\ell+1} j^{-1/p}\varepsilon_j\right| > t\right\}.$$

With probability $2^{-\ell}$,

$$\left|\sum_{j=1}^{\ell+1} j^{-1/p}\varepsilon_j\right| = \sum_{j=1}^{\ell+1} j^{-1/p} \geq (\ell + 1)^{1/q} > t$$

so that

$$\mathbb{P}\{|Z| > t\} \geq 2^{-\ell-1} \geq \frac{1}{2}\exp(-t^q).$$

Let then $t = 2\mathbb{E}\max\limits_{i \leq N}|Z_i|$ so that in particular $\mathbb{P}\{\max\limits_{i \leq N}|Z_i| > t\} \leq 1/2$. By Lemma 2.6 we have $N\mathbb{P}\{|Z| > t\} \leq 1$. From the preceding lower bound it follows that ¿ $t \geq K_p^{-1}(\log N)^{1/q}$ for some $K_p > 0$. Therefore we have obtained that $\|X\|_{p,\infty} \geq K_p^{-1}(\log N)^{1/q}$ while $\sigma_p(X) = 1$. This clearly indicates what the differences can be between $\|X\|_{p,\infty}$ and the parameter $\sigma_p(X)$ of a infinite dimensional p-stable random variable X for $1 \leq p < 2$.

According to the previous observations, the next results on comparison theorems and Sudakov's minoration for p-stable random variables are restricted to the case $1 \leq p < 2$.

We first address the question of comparison properties in the form of Slepian's lemma for stable random vectors. Consider two p-stable, $1 \leq p < 2$, random vectors $X = (X_1, \ldots, X_N)$, $Y = (Y_1, \ldots, Y_N)$ in $\mathbb{R}^N$. By analogy with the Gaussian case denote by $d_X(i,j)$ (resp. $d_Y(i,j)$) the parameter of the real p-stable variable $X_i - X_j$ (resp. $Y_i - Y_j$), $1 \leq i$, $j \leq N$. One of the ideas of the Gaussian comparison theorems was that if one can compare d_X and d_Y, then one should be able to compare the distributions or averages of $\max\limits_{i \leq N} X_i$ and $\max\limits_{i \leq N} Y_i$ (cf. Corollary 3.14 and Theorem 3.15). In the stable case with $p < 2$, the following simple example furnishes a very negative result to begin with. Let X be as in the preceding example and take Y to be the canonical p-stable vector in $\mathbb{R}^N$ given by $Y = (\theta_1, \ldots, \theta_N)$ ((θ_i) is a standard p-stable sequence). It is easily seen that $d_X(i,j) = 2^{1/q}$ while $d_Y(i,j) = 2^{1/p}$, $i \neq j$. Thus d_X and d_Y are equivalent. However, assuming for example that $p > 1$, we know that

$$\mathbb{E}\max_{i \leq N} X_i \leq K_p(\log N)^{1/q}$$

while, as a consequence of (5.16),

$$\mathbb{E}\max_{i\le N}\theta_i \ge K_p^{-1}N^{1/p}$$

(at least for all N large enough - compare $\mathbb{E}\max_{i\le N}\theta_i$ and $\mathbb{E}\max_{i\le N}|\theta_i|$). One can thus measure on this example the gap which may arise in comparison theorems for p-stable vectors when $p<2$.

Nevertheless some positive results remain. This is for example the case for Sudakov's minoration (Theorem 3.18). If $X=(X_t)_{t\in T}$ is a p-stable process $(1\le p<2)$ indexed by some set T, denote as before by $d_X(s,t)$ the parameter of the p-stable real random variable X_s-X_t, $s,t\in T$. Since $\|X_s-X_t\|_r=c_{p,r}d_X(s,t)$, $r<p$, d_X (d_X^r if $p=1$) defines a pseudo-metric on T. Recall $N(T,d_X;\varepsilon)$ is the smallest number of open balls of radius $\varepsilon>0$ in the metric d_X which cover T (possibly infinite). We then have the following extension of Sudakov's minoration. (This minoration will be improved in Section 12.2.) The idea of the proof is to represent X as conditionally Gaussian and apply then the Gaussian inequalities. As is usual in similar contexts, we simply let here

$$\|\sup_{t\in T}|X_t|\,\|_{p,\infty}=\sup\{\|\sup_{t\in F}|X_t|\,\|_{p,\infty}\,;\ F \text{ finite in } T\}\,.$$

Theorem 5.10. Let $X=(X_t)_{t\in T}$ be a p-stable random process with $1\le p<2$ and associated pseudo-metric d_X. There is a constant K_p depending only on $p>1$ such that if q is the conjugate of p, for every $\varepsilon>0$,

$$\varepsilon(\log N(T,d_X;\varepsilon))^{1/q}\le K_p\|\sup_{t\in T}|X_t|\,\|_{p,\infty}\,.$$

When $p=1$, the lower bound has to be replaced by $\varepsilon\log\log^+ N(T,d_X;\varepsilon)$. In both cases, if X is almost surely bounded, (T,d_X) is totally bounded.

Proof. We only show the result when $1<p<2$. The case $p=1$ seems to require independent deeper tools; we refer to [Ta8] for this investigation. Let $N(T,d_X;\varepsilon)\ge N$; there exists $U\subset T$ with cardinality N such that $d_X(s,t)>\varepsilon$ for $s\ne t$ in U. Consider the p-stable process $(X_t)_{t\in U}$ and let m be a spectral measure of this random vector (in $\mathbb{R}^N$ thus). Let (Y_j) be independent and distributed like $m/|m|$ and let further (g_j) denote an orthogaussian sequence. As usual in the representation, the sequences (Y_j), (Γ_j), (g_j) are independent. From Corollary 5.3, $(X_t)_{t\in U}$ has the same distribution as

$$c_p\|g_1\|_p^{-1}|m|^{1/p}\sum_{j=1}^{\infty}\Gamma_j^{-1/p}g_jY_j\,.$$

Related to this representation, we introduce *random* distances (on U). Denote by ω randomness in the sequences (Y_j) and (Γ_j) and set, for each such ω and s,t in U,

$$\begin{aligned} d_\omega(s,t) &= \left(\mathbb{E}_g \left| c_p \|g_1\|_p^{-1} |m|^{1/p} \sum_{j=1}^{\infty} \Gamma_j(\omega)^{-1/p} g_j (Y_j(\omega,s) - Y_j(\omega,t)) \right|^2 \right)^{1/2} \\ &= c_p \|g_1\|_p^{-1} |m|^{1/p} \left(\sum_{j=1}^{\infty} \Gamma_j(\omega)^{-2/p} |Y_j(\omega,s) - Y_j(\omega,t)|^2 \right)^{1/2} \end{aligned}$$

where $\mathbb{E}_g$ denotes partial integration with respect to the Gaussian sequence (g_i). Accordingly, note that for all $\lambda \in \mathbb{R}$ and s,t

$$\mathbb{E} \exp i\lambda (X_s - X_t) = \exp\left(-\frac{1}{2} |\lambda|^p d_X^p(s,t) \right) = \mathbb{E} \exp\left(-\frac{1}{2} \lambda^2 d_\omega^2(s,t) \right) .$$

It follows that for every u and $\lambda > 0$ and s,t in U,

$$\begin{aligned} \mathbb{P}\{ d_\omega(s,t) \le u d_X(s,t) \} &= \mathbb{P}\left\{ \exp\left(-\frac{\lambda^2}{2} d_\omega^2(s,t) \right) \ge \exp\left(-\frac{\lambda^2}{2} u^2 d_X^2(s,t) \right) \right\} \\ &\le \exp\left(\frac{\lambda^2}{2} u^2 d_X^2(s,t) - \frac{\lambda^p}{2} d_X^p(s,t) \right) . \end{aligned}$$

Minimizing over $\lambda > 0$, namely taking $\lambda = [(2u^2)^{\alpha/2p} d_X(s,t)]^{-1}$ where $\frac{1}{\alpha} = \frac{1}{p} - \frac{1}{2}$ yields, for all $u > 0$,

$$\mathbb{P}\{ d_\omega(s,t) \le u d_X(s,t) \} \le \exp(-c_\alpha u^{-\alpha}) \tag{5.20}$$

where $c_\alpha = (4 \cdot 2^{\alpha/2})^{-1}$.

Now recall that $d_X(s,t) > \varepsilon$ for $s \ne t$ in U. From (5.20) we thus get that for all $u > 0$

$$\mathbb{P}\{ \exists s \ne t \text{ in } U ,\ d_\omega(s,t) \le \varepsilon u \} \le (\operatorname{Card} U)^2 \exp(-c_\alpha u^{-\alpha}) .$$

Choose $u > 0$ in order this probability is less than $1/2$ (say), more precisely take

$$u = c_\alpha^{1/\alpha} (\log(2N^2))^{-1/\alpha}$$

where $N = \operatorname{Card} U$. Hence, on a set Ω_0 of ω's of probability bigger than $1/2$, $d_\omega(s,t) > \varepsilon u$ for all $s \ne t$ in U. By Sudakov's minoration for Gaussian processes (Theorem 3.18), for some numerical constant K and all ω in Ω_0

$$c_p \|g_1\|_p^{-1} |m|^{1/p} \mathbb{E}_g \sup_{t \in U} \left| \sum_{j=1}^{\infty} \Gamma_j(\omega)^{-1/p} g_j Y_j(\omega,t) \right| \ge \frac{1}{K} \cdot \frac{\varepsilon u}{2} (\log N)^{1/2}$$

(since $N(U, d_\omega; \varepsilon u/2) \geq N$). Now by partial integration

$$\begin{aligned}\mathbb{E} \sup_{t \in U} |X_t| &\geq c_p \|g_1\|_p^{-1} |m|^{1/p} \int_{\Omega_0} \mathbb{E}_g \sup_{t \in U} \left| \sum_{j=1}^{\infty} \Gamma_j^{-1/p} g_j Y_j(t) \right| d\mathbb{P} \\ &\geq \frac{1}{4K} \cdot \varepsilon u (\log N)^{1/2} .\end{aligned}$$

By the choice of u and Proposition 5.6, $\| \sup_{t \in U} |X_t| \|_{p,\infty} \geq K_p^{-1} \varepsilon (\log N)^{1/q}$. If we now recall that $N \leq N(T, d_X; \varepsilon)$ was arbitrary the proof is seen to be complete.

There is also an extension of Corollary 3.19 whose proof is completely similar. That is, if $X = (X_t)_{t \in T}$ is a p-stable random process, $1 \leq p < 2$, with almost all trajectories bounded and continuous on (T, d_X) (or having a version with these properties), then

$$\lim_{\varepsilon \to 0} \varepsilon (\log N(T, d_X; \varepsilon))^{1/q} = 0 , \quad p > 1 \tag{5.21}$$

(and

$$\lim_{\varepsilon \to 0} \varepsilon \log^+ \log N(T, d_X; \varepsilon) = 0 , \quad p = 1) .$$

In the last part of this chapter, we briefly investigate tensorization of stable random variables analogous to the ones studied for Gaussian and Rademacher series with vector valued coefficients. One corresponding question would be the following: if (x_i) is a sequence in a Banach space E and (y_j) a sequence in a Banach space F such that $\sum_i \theta_i x_i$ and $\sum_j \theta_j y_j$ are both almost surely convergent, where (θ_i) is a standard p-stable sequence, is it the same for $\sum_{i,j} \theta_{ij} x_i \otimes y_j$, where (θ_{ij}) is a doubly indexed standard p-stable sequence, in the injective tensor product $E \check{\otimes} F$ of the Banach spaces E and F ? Theorem 3.20 has provided a positive answer in the case $p = 2$ and the object of what follows will be to show that this remains valid when $p < 2$. We have however to somewhat widen this study; we know indeed that, contrary to the Gaussian case, not all p-stable Radon random variable in a Banach space can be represented as a convergent series of the type $\sum_i \theta_i x_i$. We use instead spectral measures. Let thus $0 < p < 2$ and let U and V be p-stable Radon random variables with values in E and F respectively. Denote by m_U (resp. m_V) the symmetric spectral measure of U (resp. V) concentrated on the unit sphere of E (resp. F). One can then define naturally the symmetric measure $m_U \otimes m_V$ on the unit sphere of $E \check{\otimes} F$. Is this measure the spectral measure of some p-stable random variable with values in $E \check{\otimes} F$? The next theorem describes the positive answer to this question.

Theorem 5.11. Let $0 < p < 2$ and let U and V be p-stable Radon random variables with values in Banach spaces E and F respectively. Let m_U and m_V be the respective symmetric spectral measures on the unit spheres of E and F. Then, there exists a p-stable Radon random variable W with values in $E\check{\otimes}F$ with spectral measure $m_U \otimes m_V$. Moreover, for some constant K_p depending on p only,

$$\|W\|_{p,\infty} \leq K_p(\sigma_p(U)\|V\|_{p,\infty} + \sigma_p(V)\|U\|_{p,\infty}) .$$

Proof. The idea is again to use Gaussian randomization and to benefit conditionally of the Gaussian comparison theorems. Let (Y_j) (resp. (Z_j)) be independent with values in E (resp. F) and distributed like $m_U/|m_U|$ (resp. $m_V/|m_V|$). From Corollary 5.5,

$$U' = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} g_j Y_j \qquad \text{and} \qquad V' = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} g_j Z_j$$

are almost surely convergent in E and F respectively where (g_j) is an orthogaussian sequence and, as usual, (Γ_j), (g_j), (Y_j) and (Z_j) are independent. Our aim is to show that

$$W' = \sum_{j=1}^{\infty} \Gamma_j^{-1/p} g_j Y_j \otimes Z_j$$

is almost surely convergent in $E\check{\otimes}F$ and satisfies

$$\|W'\|_{p,\infty} \leq K_p(\|U'\|_{p,\infty} + \|V'\|_{p,\infty}) . \tag{5.22}$$

W' induces a p-stable Radon random variable W with values in $E\check{\otimes}F$ and spectral measure $m_U \otimes m_V$. Since $\sigma_p(U) = |m_U|^{1/p}$, $\sigma_p(V) = |m_V|^{1/p}$, $\sigma_p(W) = |m_U|^{1/p}|m_V|^{1/p}$, homogeneity and the normalization in Corollary 5.5 lead then easily to the conclusion.

We establish inequality (5.22) for sums U', V', W' as before but only for finitely many terms in the summations, simply indicated by $\sum_j$. Convergence will follow by a simple limiting argument from (5.22).

Let f, f' in the unit ball of E', h, h' in the unit ball of F'. Since $\|Y_j\| = \|Z_j\| = 1$, for every j, with probability one,

$$|f(Y_j)h(Z_j) - f'(Y_j)h'(Z_j)| \leq |f(Y_j) - f'(Y_j)| + |h(Z_j) - h'(Z_j)| .$$

Let (g'_j) be another orthogaussian sequence independent from (g_j) and denote by $\mathbb{E}_g$ conditional integration with respect to those sequences. By the preceding inequality

$$\left(\mathbb{E}_g \left|\sum_j \Gamma_j^{-1/p} g_j [f \otimes h(Y_j \otimes Z_j) - f' \otimes h'(Y_j \otimes Z_j)]\right|^2\right)^{1/2}$$
$$\leq \left(2\mathbb{E}_g \left|\sum_j \Gamma_j^{-1/p} g_j (f(Y_j) - f'(Y'_j)) + \sum_j \Gamma_j^{-1/p} g'_j (h(Z_j) - h'(Z_j))\right|^2\right)^{1/2} .$$

It therefore follows from the Gaussian comparison theorems in the form for example of Corollary 3.14 (or Theorem 3.15) that, almost surely in (Γ_j), (Y_j), (Z_j),

$$\mathbb{E}_g \left\|\sum_j \Gamma_j^{-1/p} g_j Y_j \otimes Z_j\right\| \leq 2\sqrt{2}\mathbb{E}_g \left\|\sum_j \Gamma_j^{-1/p} g_j Y_j\right\| + 2\sqrt{2}\mathbb{E}_g \left\|\sum_j \Gamma_j^{-1/p} g'_j Z_j\right\| .$$

Integrating and using the moment equivalences of both Gaussian and stable random vectors conclude in this way the proof of Theorem 5.11.

We mention to conclude some comments and open questions on Theorem 5.11. While $\|W\|_{p,\infty}$ always dominates $\sigma_p(U)\sigma_p(V)$, it is not true in general that the inequality of Theorem 5.11 can be reversed. The works [G-M-Z] and [M-T] have actually shown a variety of examples with different sizes of $\|W\|_{p,\infty}$. One may introduce weak moments similar to the Gaussian case; a natural lower bound for $\|W\|_{p,\infty}$ would then be, besides $\sigma_p(U)\sigma_p(V)$,

$$\lambda_p(U)\|V\|_{p,\infty} + \lambda_p(V)\|U\|_{p,\infty}$$

where

$$\lambda_p(U) = \sup_{\|f\| \leq 1} \left(\int |f(x)|^p dm_U(x)\right)^{1/p}$$

and similarly for V. However, neither this lower bound nor the upper bound of the theorem seem to provide the exact weight of $\|W\|_{p,\infty}$. The definitive result, if any, should be in between. This question is still under study.

Notes and references

As announced, our exposition of stable distributions and random variables in infinite dimensional spaces is quite restricted. More general expositions based on infinitely divisible distributions, Lévy measures, Lévy-Khintchine representations and related central limit theorems for triangular arrays may be found in the treatrises [Ar-G2] and [Li] to which we actually also refer for more accurate references and historcial background. See also the work [A-A-G], the paper [M-Z], etc. The survey article [Wer] presents a sample of the topics studied in the rather extensive literature on stable distributions. More on $L_{p,\infty}$-spaces (and interpolation spaces $L_{p,q}$) may be found e.g. in [S-W].

The theory of stable laws was constructed by P. Lévy [Lé1]. The few facts presented here as an introduction may be found in the classical books on Probability theory, like [Fe1].

Representation of p-stable variables, $0 < p < 2$, goes back to the work by P. Lévy and was revived recently by R. LePage, M. Woodroofe and J. Zinn [LP-W-Z]; see [LP] for the history of this representation. For a recent and new representation, see [Ro3]. The proof of Theorem 5.1 is taken from [Pi16] (see also [M-P2]). Theorem 5.2 and the existence of spectral measures is due to P. Lévy [Lé1] (who actually dealt with the Euclidean sphere); our exposition follows [B-DC-K]. Remark 5.4 and uniqueness of the symmetric spectral measure concentrated on the unit sphere follow from the more general results about uniqueness of Lévy measures for Banach space valued random variables, and started with [Ja1] and [Kue1] in Hilbert space and was then extended to more general spaces by many authors (cf. [Ar-G2], [Li] for the details). (5.9) was noticed in [Pi12].

Proposition 5.6 is due to A. de Acosta [Ac2]; a prior result for $\sum_i \theta_i x_i$ was established by J. Hoffmann-Jørgensen [HJ1] (see also [HJ3] and Chapter 6). A. de Acosta [Ac3] also established the limit (5.11) (by a different method however) while the full conclusion (5.10) was proved by A. Araujo and E. Giné [Ar-G1]. Proposition 5.7 is taken from [G-M-Z]. Lemma 5.8 is due to M.B. Marcus and G. Pisier [M-P2] with a simplified proof by J. Zinn (cf. [M-Zi], [Pi16]). The equivalences (5.16)-(5.19) were described by L. Schwartz [Schw1].

Comparison theorems for stable random variables intrigued many people and it was probably known for a long time that Slepian's lemma does not extend to p-stable variables with $0 < p < 2$. The various introductory comments collect informations taken from [E-F], [M-P2], [Li], [Ma3],... Our exposition follows the work by M. B. Marcus and G. Pisier [M-P2]; Theorem 5.10 is theirs. Theorem 5.11 on tensor product of stable distributions was established by E. Giné, M. B. Marcus and J. Zinn [G-M-Z], and further investigated in [M-T].

Chapter 6. Sums of independent random variables

Chapter 6. Sums of independent random variables

Sums of independent random variables already appeared in the preceding chapters in concrete situations (Gaussian and Rademacher averages, representation of stable random variables). On the intuitive basis of central limit theorems which approximate normalized sums of independent random variables by smooth limiting distributions (Gaussian, stable), one would expect that results similar to those presented previously should hold in a sense or another for sums of independent random variables. The results presented in this chapter go in this direction and the reader will recognize in this general setting the topics covered before: integrability properties, moment equivalences, concentration, tail behavior, etc. We will mainly describe ideas and techniques which go from simple but powerful observations like symmetrization (randomization) techniques to more elaborated results like those obtained from the isoperimetric inequality for product measures of Theorem 1.4. Section 6.1 is concerned with symmetrization, Section 6.2 with Hoffmann-Jørgensen's inequalities and moment equivalences of sums of independent random variables. In the last and main section, martingale and isoperimetric methods are developed in this context. Many results presented in this chapter will be of basic use in the study of limit theorems later.

Let us emphasize that the infinite dimensional setting is characterized by the *lack* of the orthogonality property $\mathbb{E}|\sum_i X_i|^2 = \sum_i \mathbb{E}|X_i|^2$, where (X_i) is a finite sequence of independent mean zero real random variables. This type of identity or equivalence extends to finite dimensional random vectors and even to Hilbert space valued random variables, but *does not* in general for arbitrary Banach space valued random variables (cf. Chapter 9). With respect to the classical theory which is developed under this orthogonality property, the study of sums of independent Banach space valued random variables undertaken here requires in particular to circumvent this difficulty. Besides the difficult control in probability (that will be discussed further in this book), the various tools introduced in this chapter allow a more than satisfactory extension of the classical theory of sums of independent random variables. Actually, many ideas clarify the real case (for example, the systematic use of symmetrization-randomization) and, as for the isoperimetric approach to exponential inequalities (Section 6.3), go beyond the known results.

Since we need not be concerned here with tightness properties, we present the various results in the setting introduced in Chapter 2 and already used in the previous chapters. That is, let B be a Banach space such that there exists a countable subset D of the unit ball of the dual space such that $\|x\| = \sup_{f\in D}|f(x)|$ for all x in B. We say that a map X from some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into B is a random variable if $f(X)$ is measurable for each f in D. We recall that this definition covers the case of Radon random variables or equivalently Borel random variables taking their values in a separable Banach space.

6.1 Symmetrization and some inequalities on sums of independent random variables

One simple but basic idea in the study of sums of independent random variables is the concept of symmetrization. If X is a random variable, one can construct a *symmetric* random variable which is "near" X by looking at $\widetilde{X} = X - X'$ where X' denotes an independent copy of X (constructed on some different probability space $(\Omega', \mathcal{A}', \mathbb{P}')$). The distributions of X and $X - X'$ are indeed closely related; for example, for any $t, a > 0$, by independence and identical distribution,

$$\mathbb{P}\{\|X\| \le a\}\mathbb{P}\{\|X\| > t + a\} \le \mathbb{P}\{\|X - X'\| > t\}. \tag{6.1}$$

This is of particular interest when for example a is chosen such that $\mathbb{P}\{\|X\| \le a\} \ge 1/2$ in which case it follows that

$$\mathbb{P}\{\|X\| > t + a\} \le 2\mathbb{P}\{\|X - X'\| > t\}.$$

It also follows in particular that $\mathbb{E}\|X\|^p < \infty$ $(0 < p < \infty)$ if and only if $\mathbb{E}\|X - X'\|^p < \infty$.

Actually (6.1) is somewhat too crude in various applications and it is worthwhile mentioning here the following improvements: for $t, a > 0$,

$$\inf_{f \in D} \mathbb{P}\{|f(X)| \le a\}\mathbb{P}\{\|X\| > t + a\} \le \mathbb{P}\{\|X - X'\| > t\}. \tag{6.2}$$

For the proof, let ω be such that $\|X(\omega)\| > t + a$; then, for some h in D, $|h(X(\omega))| > t + a$. Hence

$$\inf_{f \in D} \mathbb{P}'\{|f(X')| \le a\} \le \mathbb{P}\{\|X(\omega) - X'\| > t\}.$$

Integrating with respect to ω then yields (6.2). Similarly, one can show that for $t, a > 0$,

$$\mathbb{P}\{\|X\| > t + a\} \le \mathbb{P}\{\|X - X'\| > t\} + \sup_{f \in D} \mathbb{P}\{|f(X)| > a\}. \tag{6.3}$$

When dealing with a sequence $(X_i)_{i \in \mathbb{N}}$ of independent random variables, we construct an associated sequence of independent and symmetric random variables by setting, for each i, $\widetilde{X}_i = X_i - X_i'$ where (X_i') is an independent copy of the sequence (X_i). Recall that $(\widetilde{X}_i)$ is then a symmetric sequence in the sense that it has the same distribution as $(\epsilon_i \widetilde{X}_i)$ where (ϵ_i) is a Rademacher sequence independent of (X_i) and (X_i'). That is, we can *randomize* by independent choices of signs symmetric sequences (X_i). Accordingly and following Chapter 2, we denote by $\mathbb{E}_\epsilon, \mathbb{P}_\epsilon$ (resp. $\mathbb{E}_X, \mathbb{P}_X$) partial integration with respect to (ϵ_i) (resp. (X_i)).

The fact that the symmetric sequence $(\widetilde{X}_i)$ built over (X_i) is useful in the study of (X_i) can be illustrated in different ways. Let us start for example with the Lévy-Itô-Nisio theorem for independent but not necessarily symmetric variables (cf. Theorem 2.4) where symmetrization proves its efficiency. Since weak convergence is involved in this statement we restrict for simplicity to the case of Radon random variables. The equivalence between (i) and (ii) however holds in our general setting.

Theorem 6.1. Let (X_i) be a sequence of independent Borel random variables with values in a separable Banach space B. Set $S_n = \sum_{i=1}^{n} X_i, n \geq 1$. The following are equivalent:

(i) the sequence (S_n) converges almost surely;

(ii) (S_n) converges in probability;

(iii) (S_n) converges weakly.

Proof. Suppose that (S_n) converges weakly to some random variable S. On some different probability space $(\Omega', \mathcal{A}', \mathbb{P}')$, consider a copy (X_i') of the sequence (X_i) and set $S_n' = \sum_{i=1}^{n} X_i'$; (S_n') converges weakly to S' which has the same distribution as S. Set $\widetilde{S}_n = S_n - S_n'$, $\widetilde{S} = S - S'$ defined on $\Omega \times \Omega'$. Since $\widetilde{S}_n \to \widetilde{S}$ weakly, by the result for symmetric sequences (Theorem 2.4), $\widetilde{S}_n \to \widetilde{S}$ almost surely. In particular, there exists by Fubini's theorem ω' in Ω' such that

$$S_n - S_n'(\omega') \to S - S'(\omega') \text{ almost surely.}$$

On the other hand $S_n \to S$ weakly. By difference, it follows from these two observations that $(S_n'(\omega'))$ is a relatively compact sequence in B. Further, taking characteristic functionals, for every f in B',

$$\exp(if(S_n'(\omega'))) \to \exp(if(S'(\omega'))).$$

Hence $f(S_n'(\omega')) \to f(S'(\omega'))$ and thus $S_n'(\omega)$ converges in B to $S'(\omega')$. Therefore $S_n \to S$ almost surely and the theorem is proved.

While Lévy's inequalities (Proposition 2.3) are one of the main ingredients in the proof of the Itô-Nisio's theorem in the symmetrical case, one can actually also prove directly the preceding statement using instead a similar inequality known as Ottaviani's inequality. Its proof follows the pattern of the proof of Lévy's inequalities.

Lemma 6.2. Let $(X_i)_{i\leq N}$ be independent random variables in B and set $S_k = \sum_{i=1}^{k} X_i, k \leq N$. Then, for every $s,t>0$,

$$\mathbb{P}\{\max_{k\leq N}\|S_k\| > s+t\} \leq \frac{\mathbb{P}\{\|S_N\|>t\}}{1-\max_{k\leq N}\mathbb{P}\{\|S_N-S_k\|>s\}}.$$

Proof. Let $\tau = \inf\{k\leq N; \|S_k\|>s+t\}$ ($+\infty$ if no such k exists). Then, as usual, $\{\tau=k\}$ only depends on $X_1,\ldots,X_k$ and $\sum_{k=1}^{N}\mathbb{P}\{\tau=k\} = \mathbb{P}\{\max_{k\leq N}\|S_k\|>s+t\}$. When $\tau=k$ and $\|S_N-S_k\|\leq s$, then $\|S_N\|>t$. Hence, by independence,

$$\begin{aligned}\mathbb{P}\{\|S_N\|>t\} &= \sum_{k=1}^{N}\mathbb{P}\{\tau=k, \|S_N\|>t\}\\ &\geq \sum_{k=1}^{N}\mathbb{P}\{\tau=k, \|S_N-S_k\|\leq s\}\\ &\geq \inf_{k\leq N}\mathbb{P}\{\|S_N-S_k\|\leq s\}\sum_{k=1}^{N}\mathbb{P}\{\tau=k\}\end{aligned}$$

which is the result.

The symmetrization procedure is illustrated further in the next trivial lemma. As always, (ε_i) is a Rademacher sequence independent of (X_i). Recall that X is centered if $\mathbb{E}f(X)=0$ for all f in D.

Lemma 6.3. Let $F:\mathbb{R}_+\to\mathbb{R}_+$ be convex. Then for any finite sequence (X_i) of independent mean zero random variables in B such that $\mathbb{E}F(\|X_i\|)<\infty$ for all i,

$$\mathbb{E}F\left(\frac{1}{2}\left\|\sum_i\varepsilon_iX_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_i X_i\right\|\right) \leq \mathbb{E}F\left(2\left\|\sum_i\varepsilon_iX_i\right\|\right).$$

Proof. Recall $\widetilde{X}_i = X_i - X_i'$ and let (ε_i) be a Rademacher sequence independent from (X_i) and (X_i'). Then, by Fubini, Jensen's inequality and zero mean (cf. (2.5)), and then by convexity, we have

$$\mathbb{E}F\left(\left\|\sum_i X_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_i \widetilde{X}_i\right\|\right) = \mathbb{E}F\left(\left\|\sum_i \varepsilon_i\widetilde{X}_i\right\|\right) \leq \mathbb{E}F\left(2\left\|\sum_i\varepsilon_iX_i\right\|\right).$$

Conversely, by the same arguments,

$$\mathbb{E}F\left(\frac{1}{2}\left\|\sum_i\varepsilon_iX_i\right\|\right) \leq \mathbb{E}F\left(\frac{1}{2}\left\|\sum_i\varepsilon_i\widetilde{X}_i\right\|\right) = \mathbb{E}F\left(\frac{1}{2}\left\|\sum_i\widetilde{X}_i\right\|\right) \leq \mathbb{E}F\left(\left\|\sum_i X_i\right\|\right).$$

The lemma is proved.

Note that when the variables X_i are not centered, we have similarly that

$$\mathbb{E}F\left(\sup_{f\in D}\left|\sum_i f(X_i)-\mathbb{E}f(X_i)\right|\right)\leq \mathbb{E}F\left(2\left\|\sum_i \varepsilon_i X_i\right\|\right)$$

and also

$$\mathbb{E}F\left(\sup_{f\in D}\left|\sum_i \varepsilon_i(f(X_i)-\mathbb{E}f(X_i))\right|\right)\leq \mathbb{E}F\left(2\left\|\sum_i X_i\right\|\right) .$$

Symmetrization thus indicates how results on symmetric random variables can be transferred to general ones. In the sequel of this chapter, we therefore mainly concentrate only on symmetrical distributions for which results are usually clearer and easier to state. We leave it if necessary to the interested reader to extend them to the case of general (or mean zero) independent random variables by the techniques just presented.

Before turning to the main object of this chapter, we would like to briefly mention in passing a concentration inequality often useful when for example Lévy's inequalities do not readily apply. It is due to M. Kanter [Kan].

Proposition 6.4. Let (X_i) be a finite sequence of independent symmetric random variables with values in B. Then, for any x in B and $t>0$,

$$\mathbb{P}\left\{\left\|\sum_i X_i - x\right\|\leq t\right\}\leq \frac{3}{2}\left(1+\sum_i \mathbb{P}\{\|X_i\|>t\}\right)^{-\frac{1}{2}} .$$

Since symmetric sequences of random variables can be randomized by an independent Rademacher sequence, the contraction and comparison properties described for Rademacher averages in Chapter 4 can be extended to this more general setting. The next two lemmas are easy instances of this procedure. The second one will prove extremely useful in the sequel.

Lemma 6.5. Let (X_i) be a finite symmetric sequence of random variables with values in B. Let further (ξ_i) and (ζ_i) be a real random variables such that $\xi_i=\varphi_i(X_i)$ where $\varphi_i:\mathbb{R}\to\mathbb{R}$ is symmetric (even), and similarly for ζ_i. Then, if $|\xi_i|\leq|\zeta_i|$ almost surely for all i, for any convex function $F:\mathbb{R}_+\to\mathbb{R}_+$ (and under appropriate integrability),

$$\mathbb{E}F(\|\sum_i \xi_i X_i\|)\leq \mathbb{E}F(\|\sum_i \xi_i X_i\|).$$

We also have that, for every $t > 0$,

$$\mathbb{P}\{\|\sum_i \xi_i X_i\| > t\} \le 2\mathbb{P}\{\|\sum_i \zeta_i X_i\| > t\}.$$

These inequalities in particular apply when $\xi_i = I_{\{X_i \in A_i\}} \le 1 \equiv \zeta_i$ where the sets A_i are symmetric in B (in particular $A_i = \{\|x\| \le a_i\}$).

Proof. The sequence (X_i) has the same distribution as $(\varepsilon_i X_i)$. By the symmetry assumption on the φ_i 's and Fubini's theorem

$$\mathbb{E}F(\|\sum_i \xi_i X_i\|) = \mathbb{E}_X \mathbb{E}_\epsilon F(\|\sum_i \varepsilon_i \xi_i X_i\|).$$

By the contraction principle (Theorem 4.4)

$$\mathbb{E}_\varepsilon F(\|\sum_i \varepsilon_i \zeta_i X_i\|)\ le \mathbb{E}_\varepsilon F(\|\sum_i \varepsilon_i \zeta_i X_i\|)$$

from which the first inequality of the lemma follows. The second is established similarly using (4.7).

Lemma 6.6. Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing. Let (X_i) be arbitrary random variables in B. Then, if $\mathbb{E}F(\|X_i\|) < \infty$,

$$\mathbb{E}F\left(\frac{1}{2}\sup_{f\in D}\left|\sum_i \varepsilon_i |f(X_i)|\right|\right) \le \mathbb{E}F(\|\sum_i \varepsilon_i X_i\|). \tag{6.4}$$

When the X_i 's are independent and symmetric in $L_2(B)$, we also have that

$$\mathbb{E}\left(\sup_{f\in D}(\sum_i f^2(X_i))\right) \le \sup_{f\in D}\sum_i \mathbb{E}f^2(X_i) + 8\mathbb{E}\left\|\sum_i X_i\|X_i\|\right\|. \tag{6.5}$$

Proof. (6.4) is simply Theorem 4.12 applied conditionally. In order to establish (6.5), write

$$\mathbb{E}\left(\sup_{f\in D}\left(\sum_i f^2(X_i)\right)\right) \le \sup_{f\in D}\sum_i \mathbb{E}f^2(X_i) + \mathbb{E}\left(\sup_{f\in D}\left|\sum_i f^2(X_i) - \mathbb{E}f^2(X_i)\right|\right).$$

Lemma 6.3 shows that

$$\mathbb{E}\left(\sup_{f\in D}\left|\sum_i f^2(X_i) - \mathbb{E}f^2(X_i)\right|\right) \le 2\mathbb{E}\left(\sup_{f\in D}\left|\sum_i \varepsilon_i f^2(X_i)\right|\right)$$

and, by (4.19),

$$\mathbb{E}\left(\sup_{f\in D}\left|\sum_i \varepsilon_i f^2(X_i)\right|\right) \le 4\mathbb{E}\left\|\sum_i \varepsilon_i X_i\|X_i\|\right\|.$$

The lemma is thus established.

6.2 Integrability of sums of independent random variables

Integrability of sums of independent vector valued random variables are based on various inequalities. While isoperimetric methods, which are the most powerful ones, will be described in the next section, we present here some more classical and easier ideas. An important result is a set of inequalities due to J. Hoffmann-Jørgensen which is the content of the next statement. Some of its various consequences are presented in the subsequent theorems.

Proposition 6.7. Let $(X_i)_{i\leq N}$ be independent random variables with values in B. Set $S_k = \sum_{i=1}^{k} X_i$, $k \leq N$. For every $s,t>0$,

$$\mathbb{P}\{\max_{k\leq N}\|S_k\| > 3t+s\} \leq (\mathbb{P}\{\max_{k\leq N}\|S_k\| > t\})^2 + \mathbb{P}\{\max_{i\leq N}\|X_i\| > s\}. \tag{6.6}$$

If the variables are symmetric, for $s,t>0$,

$$\mathbb{P}\{\|S_N\| > 2t+s\} \leq 4(\mathbb{P}\{\|S_N\| > t\})^2 + \mathbb{P}\{\max_{i\leq N}\|X_i\| > s\}. \tag{6.7}$$

Proof. Let $\tau = \inf\{j\leq N; \|S_j\| > t\}$. By definition, $\{\tau = j\}$ only depends on the random variables $X_1,\dots,X_j$ and $\{\max_{k\leq N}\|S_k\| > t\} = \sum_{j=1}^{N}\{\tau=j\}$ (disjoint union). On $\{\tau=j\}$, $\|S_k\|\leq t$ if $k<j$ and when $k\geq j$

$$\|S_k\| \leq t + \|X_j\| + \|S_k - S_j\|$$

so that in any case

$$\max_{k\leq N}\|S_k\| \leq t + \max_{i\leq N}\|X_i\| + \max_{j<k\leq N}\|S_k - S_j\|.$$

Hence, by independence,

$$\begin{aligned}
&\mathbb{P}\{\tau=j, \max_{k\leq N}\|S_k\| > 3t+s\}\\
&\quad\leq \mathbb{P}\{\tau=j, \max_{i\leq N}\|X_i\| > s\} + \mathbb{P}\{\tau=j\}\mathbb{P}\{\max_{j<k\leq N}\|S_k - S_j\| > 2t\}.
\end{aligned}$$

Since $\max_{j<k\leq N}\|S_k - S_j\| \leq 2\max_{k\leq N}\|S_k\|$, a summation over $j=1,\dots,N$ yields (6.6).

Concerning (6.7), for every $j=1,\dots,N$,

$$\|S_N\| \leq \|S_{j-1}\| + \|X_j\| + \|S_N - S_j\|,$$

so that

$$\mathbb{P}\{\tau = j, \|S_N\| > 2t + s\} \le \mathbb{P}\{\tau = j, \max_{i \le N} \|X_i\| > s\} + \mathbb{P}\{\tau = j\}\mathbb{P}\{\|S_N - S_j\| > t\}.$$

Making use of Lévy's inequality (2.6) for symmetric variables and summing over j yields then (6.7). The proposition is proved.

The preceding inequalities are mainly used with $s = t$. Their main interest and usefulness stem from the squared probability which make them close in a sense to exponential inequalities (see below).

As a first consequence of the preceding inequalities, the next proposition is still a technical step before the integrability statements. It however already expresses in this general context of sums of independent random variables a property similar to the one presented in the preceding chapters on Gaussian, Rademacher and stable vector valued random variables. Namely, if the sums are controlled in probability (L_0) they are also controlled in $L_p, p > 0$, provided the same holds for the maximum of the individual summands. Applied to Gaussian, Rademacher or stable averages, the next proposition actually gives rise to *new proofs* of the moment equivalences for these particular sums of independent random variables.

Proposition 6.8. Let $0 < p < \infty$ and let $(X_i)_{i \le N}$ be independent random variables in $L_p(B)$. Set $S_k = \sum_{i=1}^{k} X_i, k \le N$. Then, for $t_0 = \inf\{t > 0; \mathbb{P}\{\max_{k \le N} \|S_k\| > t\} \le (2 \cdot 4^p)^{-1}\}$,

$$\mathbb{E} \max_{k \le N} \|S_k\|^p \le 2.4^p \mathbb{E} \max_{i \le N} \|X_i\|^p + 2(4t_0)^p. \tag{6.8}$$

If the X_i's are moreover symmetric, and $t_0 = \inf\{t > 0; \mathbb{P}\{\|S_N\| > t\} \le (8.3^p)^{-1}\}$, then

$$\mathbb{E}\|S_N\|^p \le 2 \cdot 3^p \mathbb{E} \max_{i \le N} \|X_i\|^p + 2(3t_0)^p. \tag{6.9}$$

Proof. We only show (6.9), the proof of (6.8) being similar using (6.6). Let $u > t_0$. By integration by parts and (6.7),

$$\begin{aligned}
\mathbb{E}\|S_N\|^p &= 3^p \int_0^\infty \mathbb{P}\{\|S_N\| > 3t\} dt^p \\
&= 3^p \left(\int_0^u + \int_u^\infty \right) \mathbb{P}\{\|S_N\| > 3t\} dt^p \\
&\le (3u)^p + 4 \cdot 3^p \int_u^\infty (\mathbb{P}\{\|S_N\| > t\})^2 dt^p + 3^p \int_u^\infty \mathbb{P}\{\max_{i \le N} \|X_i\| > t\} dt^p \\
&\le (3u)^p + 4 \cdot 3^p \mathbb{P}\{\|S_N\| > u\} \int_0^\infty \mathbb{P}\{\|S_N\| > t\} dt^p + 3^p \mathbb{E} \max_{i \le N} \|X_i\|^p \\
&\le 2(3u)^p + 2 \cdot 3^p \mathbb{E} \max_{i \le N} \|X_i\|^p
\end{aligned}$$

since $4 \cdot 3^p \mathbb{P}\{\|S_N\| > u\} \leq 1/2$ by the choice of u. Since this holds for arbitrary $u > t_0$ the proposition is established.

It is actually possible to obtain a true equivalence of moments for sums of independent symmetric random variables. The formulation is however somewhat technical since it involves truncation. Before introducing this result we need a simple lemma on moments of maximum of independent random variables which is of independent interest.

Lemma 6.9. Let $p > 0$ and let (Z_i) be a finite sequence of positive random variables in L_p. Given $\lambda > 0$, let $\delta_0 = \inf\{t > 0; \sum\limits_i \mathbb{P}\{Z_i > t\} \leq \lambda\}$. Then

$$\begin{aligned} \lambda(1+\lambda)^{-1}\delta_0^p + (1+\lambda)^{-1} \sum \lim its_i \int_{\delta_0}^{\infty} \mathbb{P}\{Z_i > t\} dt^p \\ \leq \mathbb{E} \max_i Z_i^p \leq \delta_0^p + \sum_i \int_{\delta_0}^{\infty} \mathbb{P}\{Z_i > t\} dt^p. \end{aligned}$$

Proof. Use integration by parts. The right hand side is trivial (and actually holds for *any* $\delta_0 > 0$). Turning to the left side we use Lemma 2.6: the definition of δ_0 then indicates that

$$\mathbb{P}\{\max_i Z_i > t\} \geq \begin{cases} (1+\lambda)^{-1} \sum\limits_i \mathbb{P}\{Z_i > t\} & \text{if } t > \delta_0 \\ \lambda(1+\lambda)^{-1} & \text{if } t \leq \delta_0. \end{cases}$$

Lemma 6.9 then clearly follows.

The announced equivalence of moments is as follows.

Proposition 6.10. Let $0 < p, q < \infty$. Let (X_i) be a finite sequence of independent and symmetric random variables in $L_p(B)$. Then, for some constant $K_{p,q}$ depending on p, q only,

$$\|\sum_i X_i\|_p \underset{K_{p,q}}{\sim} \|\max_i \|X_i\|\,\|_p + \|\sum_i X_i I_{\{\|X_i\| \leq \delta_0\}}\|_q$$

where $\delta_0 = \inf\{t > 0; \sum\limits_i \mathbb{P}\{\|X_i\| > t\} \leq (8.3^p)^{-1}\}$ and where the sign $a \underset{K_{p,q}}{\sim} b$ means that $K_{p,q}^{-1} b \leq a \leq K_{p,q} b$.

Proof. By the triangle inequality

$$\mathbb{E}\|\sum_i X_i\|^p \leq 2^p \mathbb{E}\|\sum_i X_i I_{\{\|X_i\| \leq \delta_0\}}\|^p + 2^p \mathbb{E}\|\sum_i X_i I_{\{\|X_i\| > \delta_0\}}\|^p.$$

If we apply (6.9) to the second term of the right side of this inequality we see that, by definition of δ_0 , we can take $t_0 = 0$ there so that

$$\mathbb{E}\|\sum_i X_i I_{\{\|X_i\|>\delta_0\}}\|^p \le 2\cdot 3^p \mathbb{E}\max_i \|X_i\|^p.$$

Turning to the first term and applying again Proposition 6.8, we can take

$$t_0 = (8\cdot 3^p \mathbb{E}\|\sum_i X_i I_{\{\|X_i\|\le\delta_0\}}\|^q)^{1/q}.$$

The first half of the proposition follows. To prove the reverse inequality note that, by (6.9) again,

$$\mathbb{E}\|\sum_i X_i I_{\{\|X_i\|\le\delta_0\}}\|^q \le 2\cdot 3^q \delta_0^q + 2.(3t_0)^q$$

where we can choose as t_0 ,

$$t_0 = (8\cdot 3^q \mathbb{E}\|\sum_i X_i\|^p)^{1/p}.$$

If we then draw from Lemma 6.9 the fact that

$$\mathbb{E}\max_i \|X_i\|^p \ge \lambda(1+\lambda)^{-1}\delta_0^p$$

with $\lambda = (8.3^p)^{-1}$, the proof will be complete since we know from Lévy's inequality (2.7) that

$$\mathbb{E}\max_i \|X_i\|^p \le 2\mathbb{E}\|\sum_i X_i\|^p.$$

We now summarize in various integrability theorems for sums of independent vector valued random variables the preceding powerful inequalities and arguments.

Theorem 6.11. Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent random variables with values in B . Set, as usual, $S_n = \sum_{i=1}^n X_i, n\ge 1$. Let also $0<p<\infty$. Then, if $\sup_n \|S_n\| < \infty$ almost surely, we have equivalence between:

(i) $\mathbb{E}\sup_n \|S_n\|^p < \infty$;

(ii) $\mathbb{E}\sup_n \|X_n\|^p < \infty$.

Further, if (S_n) converges almost surely, (i) and (ii) are also equivalent to

(iii) $\mathbb{E}\|\sum_i X_i\|^p < \infty$

and in this case (S_n) converges also in L_p.

Proof. That (i) implies (ii) is obvious. Let N be fixed. By Proposition 6.8, t_0 being defined there,

$$\mathbb{E}\max_{n\le N}\|S_n\|^p \le 2\cdot 4^p \mathbb{E}\max_{i\le N}\|X_i\|^p + 2(4t_0)^p.$$

Since $\mathbb{P}\{\sup_n \|S_n\| < \infty\} = 1$, there is $M > 0$ such that $t_0 \le M$ independently of N. Letting N tend to infinity shows (ii) $\Rightarrow$ (i). The assertion relative to (iii) follows from Lévy's inequalities for symmetric random variables, and from an easy symmetrization argument based on (6.1) in general.

Corollary 6.12. Let (a_n) be an increasing sequence of positive numbers tending to infinity. Let (X_i) be independent random variables with values in B and set, as usual, $S_n = \sum_{i=1}^{n} X_i, n \ge 1$. Then, if $\sup_n \|S_n\|/a_n < \infty$ almost surely, for any $0 < p < \infty$, the following are equivalent:

(i) $\mathbb{E}\sup_n \left(\frac{\|S_n\|}{a_n}\right)^p < \infty$;

(ii) $\mathbb{E}\sup_n \left(\frac{\|X_n\|}{a_n}\right)^p < \infty$.

Proof. We define a new sequence (Y_i) of independent random variables with values in the Banach space $\ell_\infty(B)$ of all bounded sequences $x = (x_n)$ with sup-norm $\|x\| = \sup_n \|x_n\|$ by setting

$$Y_i = \left(0, \dots, 0, \frac{X_i}{a_i}, \frac{X_i}{a_{i+1}}, \frac{X_i}{a_{i+2}}, \dots\right)$$

where there are $i-1$ zeroes to start with. Clearly $\|Y_i\| = \|X_i\|/a_i$ for all i, and

$$\sup_n \|\sum_{i=1}^{n} Y_i\| = \sup_n \frac{\|S_n\|}{a_n}.$$

Apply then Theorem 6.11 to the sequence (Y_i) in $\ell_\infty(B)$.

Remark 6.13. It is amusing to note that Hoffmann-Jørgensen's inequalities and Theorem 6.11 describe well enough independence to contain the Borel-Cantelli lemma! Indeed, if (A_i) is a sequence of independent sets, we get from Theorem 6.11 that if $\sum_i I_{A_i}$ converges almost surely, then $\mathbb{E}(\sum_i I_{A_i}) = \sum_i \mathbb{P}(A_i) < \infty$; this corresponds to the independent portion of the Borel-Cantelli lemma.

Provided with the preceding material, let us now consider an almost surely convergent series $S = \sum_i X_i$ of independent symmetric or only mean zero uniformly bounded random variables with values in B and let us try to investigate the integrability properties of $\|S\|$. Assume more precisely that the X_i 's are symmetric and that $\|X_i\|_\infty \le a < \infty$ for all i. For each N, set $S_N = \sum_{i=1}^{N} X_i$. By (6.7), for every $t > 0$,

$$\mathbb{P}\{\|S_N\| > 2t + a\} \le (2\mathbb{P}\{\|S_N\| > t\})^2.$$

Let t_0 to be specified in a moment and define the sequence $t_n = 2^n(t_0 + a) - a$. The preceding inequality indicates that, for every n,

$$\mathbb{P}\{\|S_N\| > t_n\} \le (2\mathbb{P}\{\|S_N\| > t_{n-1}\})^2.$$

Iterating, we get that

$$\mathbb{P}\{\|S_N\| > t_n\} \le 2^{2^{n+1}-2}(\mathbb{P}\{\|S_N\| > t_0\})^{2^n}.$$

If $S = \sum_i X_i$ is almost surely convergent, there exists t_0 such that for every N, $\mathbb{P}\{\|S_N\| > t_0\} \le 1/8$. Summarizing, for every N and n,

$$\mathbb{P}\{\|S_N\| > 2^n(t_0 + a)\} \le 2^{-2^n}.$$

It easily follows that for some $\lambda > 0$, $\sup_N \mathbb{E}\exp\lambda\|S_N\| < \infty$, and thus, by Fatou's lemma, that $\mathbb{E}\exp\lambda\|S\| < \infty$. By convergence, this can easily be improved into the same property for *all* $\lambda > 0$. Note actually that $\sup_N \mathbb{E}\exp\lambda\|S_N\| < \infty$ as soon as the sequence (S_N) is stochastically bounded.

The preceding iteration procedure may be compared to the proof of (3.5) in the Gaussian case. To complement it, let us present a somewhat neater argument for the same result, which, if only a small variation, completes our hability with the technique. The argument is the following; applied to power functions, it yields an alternate proof of Proposition 6.8 (cf. Remark 6.15 below).

Proposition 6.14. Let $(X_i)_{i\le N}$ be independent and symmetric random variables with values in B, $S_k = \sum_{i=1}^{k} X_i, k \le N$. Assume that $\|X_i\|_\infty \le a$ for all $i \le N$. Then, for every $\lambda, t > 0$,

$$\mathbb{E}\exp\lambda\|S_N\| \le \exp(\lambda t) + 2\exp(\lambda(t+a))\mathbb{P}\{\|S_N\| > t\}\mathbb{E}\exp\lambda\|S_N\|.$$

Proof. Let as usual $\tau = \inf\{k \le N; \|S_k\| > t\}$. We can write

$$\mathbb{E}\exp\lambda\|S_N\| \le \exp(\lambda t) + \sum_{k=1}^{N}\int_{\{\tau=k\}}\exp\lambda\|S_N\|d\mathbb{P}.$$

On the set $\{\tau = k\}$,

$$\|S_N\| \le \|S_{k-1}\| + \|X_k\| + \|S_N - S_k\| \le t + a + \|S_N - S_k\|$$

so that, by independence,

$$\int_{\{\tau=k\}} \exp \lambda\|S_N\| d\mathbb{P} \le \exp(\lambda(t+a))\mathbb{P}\{\tau = k\}\mathbb{E}\exp\lambda\|S_N - S_k\|.$$

By Jensen's inequality and mean zero, $\mathbb{E}\exp\lambda\|S_N - S_k\| \le \mathbb{E}\exp\lambda\|S_N\|$ (cf. (2.5)), and, summing over k ,

$$\sum_{k=1}^{N} \mathbb{P}\{\tau = k\} = \mathbb{P}\{\max_{k\le N}\|S_k\| > t\} \le 2\mathbb{P}\{\|S_N\| > t\}$$

where we have used Lévy's inequality (2.6). The proof is complete. (Note that there is an easy analog when the variables are only centered.)

Remark 6.15. As announced, the proof of Proposition 6.14 applied to power functions yields an alternate proof of the inequalities of Proposition 6.8. It actually yields an inequality in the form of Kolmogorov's converse inequality. That is, under the assumption of the last proposition, for all $t > 0$ and every $1 \le p < \infty$,

$$\mathbb{P}\{\|S_N\| > t\} \ge \frac{1}{2^p}\left[1 - \frac{2^{2p}(t^p + \mathbb{E}\max_{i\le N}\|X_i\|^p)}{\mathbb{E}\|S_N\|^p}\right].$$

Let (X_i) be a sequence of independent symmetric (or only mean zero) uniformly bounded by a such that $S = \sum_i X_i$ converges almost surely. Then, as a consequence of Proposition 6.14, we recover the fact that $\mathbb{E}\exp\lambda\|S\| < \infty$ for some (actually all) $\lambda > 0$. Indeed, if we choose t in Proposition 6.14 satisfying $\mathbb{P}\{\|S_N\| > t\} \le (2e)^{-1}$ for all N and let $\lambda = (t+a)^{-1}$, we simply get that

$$\sup_N \mathbb{E}\exp\lambda\|S_N\| \le 2\exp(\lambda t) < \infty.$$

This exponential integrability result is not quite satisfactory since it is known that for real random variables, $\mathbb{E}\exp\lambda|S|\log^+|S| < \infty$ for some $\lambda > 0$. This result on the line is one instance of the Poisson behavior of general sums of independent (bounded) random variables as opposed to the normal behavior of more specialized ones, like Rademacher averages. It originates in the sharp *quadratic* real exponential inequalities (see for example (6.10) below). These real results can however be extended to the vector valued case. We use to this aim isoperimetric methods to obtain sharp exponential estimates for sums of independent random variables, even improving at some places the scalar case.

6.3. Concentration and tail behavior

This section is mainly devoted to applications of isoperimetric methods (Theorem 1.4) to integrability and tail behavior of sums of independent random variables. One of the objectives will be to try to extend to the infinite dimensional setting the classical (quadratic) exponential inequalities like those of Bernstein, Kolmogorov (Lemma 1.6), Prokhorov, Bennett, Hoeffding, etc. (Of course, the lack of orthogonality forces to investigate new arguments, like therefore isoperimetry.) To state one, and for the matter of comparison, let us consider Bennett's inequality. Let (X_i) be a finite sequence of independent mean zero *real* random variables such that $\|X_i\|_\infty \le a$ for all i ; then, if $b^2 = \sum_i \mathbb{E}X_i^2$, for all $t > 0$,

$$\mathbb{P}\{\sum_i X_i > t\} \le \exp\left[\frac{t}{a} - \left(\frac{t}{a} + \frac{b^2}{a^2}\right)\log\left(1 + \frac{at}{b^2}\right)\right]. \tag{6.10}$$

This inequality is rather typical of the tail behavior of sums of independent random variables. This behavior is variable depending on the relative sizes of t and the ratio b^2/a . Since $\log(1+x) \ge x - \frac{1}{2}x^2$ when $0 \le x \le 1$, if $t \le b^2/a$, (6.10) implies that

$$\mathbb{P}\{\sum_i X_i > t\} \le \exp\left(-\frac{t^2}{2b^2} + \frac{at^3}{2b^4}\right),$$

which is further less than $\exp(-t^2/4b^2)$ if $t \le b^2/2a$ (for example). On the other hand, for all $t > 0$

$$\mathbb{P}\{\sum_i X_i > t\} \le \exp\left[-\frac{t}{a}\left(\log\left(1 + \frac{at}{b^2}\right) - 1\right)\right]$$

which is sharp for large values of t (bigger than b^2/a). These two inequalities actually describe the classical *normal* and *Poisson* types behavior of sums of independent random variables according to the size of t with respect to b^2/a .

Before turning to isoperimetric argument in this study, we would like to present some results based on martingales. Although these do not seem to be powerful enough in general for the integrability and tail behavior questions we have in mind, they are however rather simple and quite useful in many situations. They also present the advantage of being formulated as concentration inequalities, some of which will be useful in this form in Chapter 9.

The key observation in order to use (real) martingale inequalities in the study of sums of independent vector valued random variables relies on the following simple but extremely useful observation of V. Yurinskii.

Let $(X_i)_{i\le N}$ be integrable random variable in B. Denote by $\mathcal{A}_i$ the σ-algebra generated by the variables $X_1,\dots,X_i$, $i\le N, \mathcal{A}_0$ the trivial algebra. Write as usual $S_N=\sum_{i=1}^N X_i$ and set, for each i,

$$d_i=\mathbb{E}^{\mathcal{A}_i}\|S_N\|-\mathbb{E}^{\mathcal{A}_{i-1}}\|S_N\|.$$

$(d_i)_{i\le N}$ defines a real martingale difference sequence ($\mathbb{E}^{\mathcal{A}_{i-1}}d_i=0$) and $\sum_{i=1}^N d_i=\|S_N\|-\mathbb{E}\|S_N\|$. We then have:

Lemma 6.16. Assume the random variables X_i are independent. Then, in the preceding notations, almost surely for every $i\le N$,

$$|d_i|\le\|X_i\|+\mathbb{E}\|X_i\|.$$

Further, if the X_i's are in $L_2(B)$ we also have that

$$\|\mathbb{E}^{\mathcal{A}_{i-1}}d_i^2\|_\infty\le\mathbb{E}\|X_i\|^2.$$

Proof. Independence ensures that

$$d_i=(\mathbb{E}^{\mathcal{A}_i}-\mathbb{E}^{\mathcal{A}_{i-1}})(\|S_N\|-\|S_N-X_i\|)$$

and the first inequality of the lemma already follows from the triangle inequality since

$$|d_i|\le(\mathbb{E}^{\mathcal{A}_i}+\mathbb{E}^{\mathcal{A}_{i-1}})(\|X_i\|)=\|X_i\|+\mathbb{E}\|X_i\|.$$

Since conditional expectation is a projection in L_2 the same argument leads to the second inequality of the lemma.

The philosophy of the preceding observation is that the deviation of the norm of a sum S_N of independent random vectors $X_i, i\le N$, from its expectation $\mathbb{E}\|S_N\|$ can be written as a *real* martingale whose differences d_i are nearly exactly controlled by the norms of the corresponding individual summands X_i of S_N. In a sense therefore, up to $\mathbb{E}\|S_N\|, \|S_N\|$ is as good as a real martingale with differences comparable to $\|X_i\|$. Typical in this regard is the following quadratic inequality, immediate consequence of Lemma 6.16 and orthogonality of martingale differences:

$$\mathbb{E}\big|\,\|S_N\|-\mathbb{E}\|S_N\|\big|^2\le\sum_{i=1}^N\mathbb{E}\|X_i\|^2. \tag{6.11}$$

Of course, on the line or even in Hilbert space, if the variables are centered, this inequality (actually an equality) holds true without the centering factor $\mathbb{E}\|S_N\|$ by orthogonality.

This remark is one rather important feature of the study of sums of independent Banach space valued random variables (we will find it also in the isoperimetric approach developed later). It shows how, when a control in *expectation* (or only in *probability* by Proposition 6.8) of a sum of independent random variables is given, then one can expect, using some of the classical *real* arguments, an *almost sure* control. This was already the line of the integrability theorems discussed in Section 6.2 and those of this section will be similar. In the next two chapters on strong limit theorems for sums of independent random variables, this will lead to various equivalences between almost sure and in probability limiting properties under necessary (and classical) moment assumptions on the individual summands. As in the Gaussian, Rademacher and stable cases, there is then to know how to control in probability (or weakly) sums of independent random variables. On the line or in finite dimensional spaces this is easily accomplished by orthogonality and moment conditions. This is much more difficult in the infinite dimensional setting and may be considered a main problem of the theory. It will be studied in some instances in the second part of this work, in Chapters 9, 10 and 14 in particular.

From Lemma 6.16, the various martingale inequalities of Chapter 1, Section 1.3, can be applied to $\|S_N\| - \mathbb{E}\|S_N\|$ yielding concentration properties for norms of sums $S_N = \sum_{i=1}^N X_i$ of independent random variables around their expectation in terms of the size of the summands X_i . Let us record some of them at this stage. Lemma 1.6 (together thus with Lemma 6.16) shows that if $a = \max_{i \le N} \|X_i\|_\infty$ and $b \le (\sum_{i=1}^N \mathbb{E}\|X_i\|^2)^{1/2}$, for all $t > 0$,

$$\mathbb{P}\{|\,\|S_N\| - \mathbb{E}\|S_N\|\,| > t\} \le 2\exp\left[-\frac{t^2}{2b^2}\left(2 - \exp\left(\frac{2at}{b^2}\right)\right)\right]. \tag{6.12}$$

One can also prove a martingale version of (6.10) which, applied to $\|S_N\| - \mathbb{E}\|S_N\|$, yields

$$\mathbb{P}\{|\,\|S_N\| - \mathbb{E}\|S_N\|\,| > t\} \le \exp\left[\frac{t}{2a} - \left(\frac{t}{2a} + \frac{b^2}{4a^2}\right)\log\left(1 + \frac{2at}{b^2}\right)\right] \tag{6.13}$$

Lemma 1.7 indicates in the same way that if $1 < p < 2,\ q = p/p-1$ and $a = \sup_{i \ge 1} i^{1/p}\|X_i\|_\infty$ assumed to be finite, for all $t > 0$,

$$\mathbb{P}\{|\,\|S_N\| - \mathbb{E}\|S_N\|\,| > t\} \le 2\exp(-t^q/C_q a^q) \tag{6.14}$$

where $C_q > 0$ only depends on q. (6.14) is of particular interest when $X_i = \epsilon_i x_i$ where (ϵ_i) is a Rademacher sequence and (x_i) a finite sequence in B. We then have, for all $t > 0$,

$$\mathbb{P}\{|\,\|\sum_i \varepsilon_i x_i\| - \mathbb{E}\|\sum_i \epsilon_i x_i\|\,| > t\} \leq 2\exp(-t^q / C_q \|(x_i)\|_{p,\infty}^q) \tag{6.15}$$

where $\|(x_i)\|_{p,\infty} = \|(\|x_i\|)\|_{p,\infty}$. This inequality may of course be compared to the concentration property of Theorem 4.7 as well as to the real inequalities described in the first section of Chapter 4. The previous two inequalities will be helpful both in this chapter and in Chapter 9 where in particular concentration will be required for the construction of ℓ_p^n-subspaces, $1 < p < 2$, of Banach spaces. Note that we already used (6.11) in the preceding chapter to prove a concentration inequality for stable random variables (Proposition 5.7).

We now turn to the main part of this chapter with the applications to sums of independent random variables of the isoperimetric inequality for product measures of Theorem 1.4. This isoperimetric inequality appears as a powerful tool which will be shown to be efficient in many situations. It will allow in particular to complete the study of integrability properties of sums of independent random variables started in the preceding section and to investigate almost sure limit theorems in the next chapters. This isoperimetric approach, a priori very different from the classical tools (although similarities can and will be detected), seems to subsume in general the usual arguments.

Let us first briefly recall Theorem 1.4 and (1.13). Let N be an arbitrary but fixed integer and let $\mathbb{X} = (X_i)_{i \leq N}$ be a sample of independent random variables with values in B. (In this setting, we may simply equip B with the σ-algebra generated by the linear functionals $f \in D$.) For A measurable in the product space $B^N = \{x \in B; x = (x_i)_{i \leq N}, x_i \in B\}$ and for integers q, k, set

$$H(A, q, k) = \{x \in B^N; \exists x^1, \ldots, x^q \in A,\ \mathrm{Card}\{i \leq N; x_i \notin \{x_i^1, \ldots, x_i^q\}\} \leq k\}.$$

Then, if $\mathbb{P}\{\mathbb{X} \in A\} \geq 1/2$ and $k \geq q$,

$$\mathbb{P}_*\{\mathbb{X} \in H(A, q, k)\} \geq 1 - \left(\frac{K_0}{q}\right)^k. \tag{6.16}$$

Recall that for convenience the numerical constant K_0 is assumed to be an integer.

On $H(A, q, k)$, the sample $\mathbb{X}$ is controlled by a finite number q of points in A provided k values are neglected. The isoperimetric inequality (6.16) precisely estimates, with an exponential decay in k, the probability that this is realized. In applications to sums of independent random variables, the k values for which the isoperimetric inequality does not provide any control may be thought as the largest elements (in norm) of the sample. This observation is actually the conducting rod to the subsequent developments. We will see how, up to the control of the large values, the isoperimetric inequality provides optimal estimates of the tail behavior of sums of independent random variables. As for vector valued Gaussian variables and Rademacher series, several parameters are used to measure tail of sums of independent random variables. These involve some quantity in the L_0 topology (median or expectation), information on weak moments, and thus, as is clear from the isoperimetric approach, estimates on large values.

Let us now state and prove the tail estimate on sums of independent random variables we draw from the isoperimetric inequality (6.16). It may be compared to the real inequalities presented at the beginning of the section although we do not explicit this comparison since the vector valued case induces several complications. Various arguments in both this chapter and the next one however provide the necessary methodology and tools towards this goal. It deals with symmetric variables since Rademacher randomization and conditional use of tail estimates for Rademacher averages are an essential complement to the approach. If $(X_i)_{i\le N}$ is a finite sequence of random variables, we denote by $(\|X_i\|^*)_{i\le N}$ the non-increasing rearrangement of $(\|X_i\|)_{i\le N}$.

Theorem 6.17. Let $(X_i)_{i\le N}$ be independent and symmetric random variables with values in B . Then, for any integers $k\ge q$ and real numbers $s,t>0$,

$$\begin{aligned}(6.17)\qquad & \mathbb{P}\left\{\left\|\sum_{i=1}^N X_i\right\| > 8qM+2s+t\right\} \\ &\le \left(\frac{K_0}{q}\right)^k + \mathbb{P}\left\{\sum_{i=1}^k \|X_i\|^* > s\right\} + 2\exp\left(-\frac{t^2}{128qm^2}\right)\end{aligned}$$

where

$$M=\mathbb{E}\left\|\sum_{i=1}^N u_i\right\|, \quad m=\mathbb{E}\left(\sup_{f\in D}\left(\sum_{i=1}^N f^2(u_i)\right)^{1/2}\right)$$

and $u_i = X_i I_{\{\|X_i\|\le s/k\}}, i\le N.$

Before turning to the proof of Theorem 6.17, let us make some comments in order to clarify the statement. First note that M and m defined from truncated random variables u_i are easily majorized (using the contraction principle for M) by the same expressions with X_i instead of u_i (provided $X_i\in L_2(B)$). (6.17) is often good enough for applications in this simpler form. Actually, note also that when we need not be concerned with truncations, for example if we deal with bounded variables, then in (6.17) the parameter $2s$ in the left hand side can be improved to s. This is completely clear from the proof below and is sometimes useful as we will see in Theorem 6.19. The reader recognizes in the first two terms on the right of (6.17) the isoperimetric bound (6.16) and the largest values. M corresponds to a control in probability of the sum, m to weak moment estimates. Actually m can be used in several ways. These are summarized in the following main three estimates. First, by (6.5) and the contraction principle

$$m^2 \le \sup_{f \in D} \sum_{i=1}^{N} \mathbb{E} f^2(u_i) + 8\frac{Ms}{k}. \tag{6.18}$$

Then, by (4.3) and symmetry,

$$m^2 \le 2M^2, \tag{6.19}$$

while, trivially,

$$m^2 \le \sum_{i=1}^{N} \mathbb{E}\|u_i\|^2. \tag{6.20}$$

(6.18) corresponds basically to the sharpest estimate and will prove efficient in limit theorems for example. The two others, especially (6.19), are convenient when no weak moments assumptions need to be taken into account; both include the real valued situation.

Let us now show how Theorem 6.17 is obtained from the isoperimetric inequality.

Proof. We decompose the proof step by step.

1 $^{\text{st}}$ step. This is an elementary observation on truncation and large values. Recall that if no truncations are needed, this step can be omitted. If $(X_i)_{i \le N}$ are (actually arbitrary) random variables and if $s \ge \sum_{i=1}^{k} \|X_i\|^*$, then

$$\|\sum_{i=1}^{N} X_i\| \le s + \|\sum_{i=1}^{N} u_i\|$$

where we recall that $u_i = X_i I_{\{\|X_i\| \le s/k\}}, i \le N$. Indeed, if J denotes the set of integers $i \le N$ such that $\|X_i\| > s/k$, then $\operatorname{Card} J \le k$ since if not this would contradict $\sum_{i=1}^{k} \|X_i\|^* \le s$. Then

$$\left\|\sum_{i=1}^{N} X_i\right\| \le \left\|\sum_{i \in J} X_i\right\| + \left\|\sum_{i \notin J} X_i\right\| \le \sum_{i=1}^{k} \|X_i\|^* + \|\sum_{i=1}^{N} u_i\|$$

which gives the result.

2 nd step. Application of the isoperimetric inequality. By symmetry and independence recall that the sample $\mathbb{X} = (X_i)_{i \le N}$ has the same distribution as $(\varepsilon_i X_i)_{i \le N}$ where (ε_i) is a Rademacher sequence independent of $\mathbb{X}$. Suppose that we are given A in B^N such that $\mathbb{P}\{\mathbb{X} \in A\} \ge 1/2$. If then $\mathbb{X} \in H = H(A,q,k)$, there exist by definition $j \le k$ and $x^1, \ldots, x^q \in A$ such that

$$\{1, \ldots, N\} = \{i_1, \ldots, i_j\} \cup I$$

where $I = \bigcup_{\ell=1}^{q} \{i \le N; X_i = x_i^\ell\}$. Together with the first step we can then write, if $s \ge \sum_{i=1}^{k} \|X_i\|^*$,

$$\begin{aligned} \|\sum_{i=1}^{N} \varepsilon_i X_i\| &\le s + \|\sum_{i=1}^{N} \varepsilon_i u_i\| \\ &\le s + \|\sum_{\ell=1}^{j} \varepsilon_{i_\ell} u_{i_\ell}\| + \|\sum_{i \in I} \varepsilon_i u_i\| \\ &\le 2s + \|\sum_{i \in I} \varepsilon_i u_i\|. \end{aligned}$$

From the isoperimetric inequality (6.16) we then clearly get that, for $k \ge q$, $s, t > 0$,

$$\begin{aligned} &\mathbb{P}\{\|\sum_{i=1}^{N} X_i\| > 2s + t\} \\ &\le \left(\frac{K_0}{q}\right)^k + \mathbb{P}\{\sum_{i=1}^{k} \|X_i\|^* > s\} + \int_{\{\mathbb{X} \in H\}} \mathbb{P}_\varepsilon\{\|\sum_{i \in > I} \varepsilon_i u_i\| > t\} d\mathbb{P}_X. \end{aligned} \tag{6.21}$$

There is here a slight abuse in notation since $\{\mathbb{X} \in H\}$ need not be measurable and we should be somewhat more careful in this application of Fubini's theorem. This is however irrelevant and for simplicity we skip these details.

3 rd step. Choice of A and conditional estimates on Rademacher averages. On the basis of (6.21), we are now interested in conditional estimates of $\mathbb{P}_\varepsilon\{\|\sum_{i \in I} \varepsilon_i u_i\| > t\}$ with some appropriate choice for A. This is the place where randomization appears to be crucial. The tail inequality on Rademacher averages we use is inequality (4.11) in the following form: if (x_i) is a finite sequence in B and $\sigma = \sup_{f \in D} (\sum_i f^2(x_i))^{1/2}$, for any $t > 0$,

$$\mathbb{P}\{\|\sum_i \varepsilon_i x_i\| > 2\mathbb{E}\|\sum_i \varepsilon_i x_i\| + t\} \le 2\exp(-t^2/8\sigma^2). \tag{6.22}$$

(With some worse constants, we could also use (4.15).) Of course, this kind of inequality is simpler on the line (cf. the subgaussian inequality (4.1)) and *the interested reader is perhaps invited to consider this simpler case to start with.* Provided with this inequality, let us consider $A = A_1 \cap A_2$ where

$$A_1 = \{x = (x_i)_{i \le N}; \mathbb{E}\|\sum_{i=1}^N \varepsilon_i x_i I_{\{\|x_i\| \le s/k\}}\| \le 4M\},$$
$$A_2 = \{x = (x_i)_{i \le N}; \sup_{f \in D} (\sum_{i=1}^N f^2(x_i) I_{\{\|x_i\| \le s/k\}})^{1/2} \le 4m\}.$$

The very definitions of M and m clearly show that $\mathbb{P}\{\mathbb{X} \in A\} \ge 1/2$ so that we are in a position to apply (6.21). Observe now that Rademacher averages are *monotone* in the sense that $\mathbb{E}\|\sum_{i \in J} \epsilon_i x_i\|$ is an increasing function of $J \subset \mathbb{N}$. We use this property in the following way. By definition of I, for each $i \in I$, we can fix $1 \le \ell(i) \le q$ with $X_i = x_i^{\ell(i)}$. Let $I_\ell = \{i; \ell(i) = \ell\}$, $1 \le \ell \le q$. We have that

$$\mathbb{E}_\varepsilon \left\| \sum_{i \in I} \varepsilon_i u_i \right\| = \mathbb{E}_\varepsilon \left\| \sum_{\ell=1}^q \sum_{i \in I_\ell} \varepsilon_i u_i \right\| \le \sum_{\ell=1}^q \mathbb{E} \left\| \sum_{i \in I_\ell} \varepsilon_i x_i^\ell I_{\{\|x_i^\ell\| \le s/k\}} \right\| .$$

Then, by monotonicity of Rademacher averages, and definition of A (recall $x^1, \ldots, x^q$ belong to A), it follows that

$$\mathbb{E}_\varepsilon \left\| \sum_{i \in I} \varepsilon_i u_i \right\| \le \sum_{\ell=1}^q \mathbb{E}_\varepsilon \left\| \sum_{i=1}^N \varepsilon_i x_i^\ell I_{\{\|x_i^\ell\| \le s/k\}} \right\| \le 4qM .$$

Similarly, but with sums of squares (that are also monotone),

$$\sup_{f \in D} \sum_{i \in I} f^2(u_i) \le 16qm^2.$$

Theorem 6.17 now clearly follows from these observations combined with the estimate on Rademacher averages (6.22) and (6.21).

Remark 6.18. One of the key observations in the third step of the preceding proof is the monotonicity of Rademacher averages. It might be interesting to describe what the isoperimetric inequality directly produces when applied to conditional averages $\mathbb{E}_\varepsilon\|\sum_{i=1}^N \varepsilon_i X_i\|$ which thus satisfy this basic unconditionality property. Let therefore $(X_i)_{i \le N}$ be independent symmetric variables in $L_1(B)$ and set $M = \mathbb{E}\|\sum_{i=1}^N X_i\|$. Then, for every $k \ge q$ and $s > 0$,

$$\mathbb{P}_X\{\mathbb{E}_\varepsilon\|\sum_{i=1}^N \varepsilon_i X_i\| > 2qM + s\} \le \left(\frac{K_0}{q}\right)^k + \mathbb{P}\{\sum_{i=1}^k \|X_i\|^* > s\}. \tag{6.23}$$

For the proof we let

$$A = \{x = (x_i)_{i \le N} ; \mathbb{E}\|\sum_{i=1}^{N} \varepsilon_i x_i\| \le 2M\}$$

so that, by symmetry, $\mathbb{P}\{\mathbb{X} \in A\} \ge 1/2$. If $\mathbb{X} \in H(A,q,k)$, there exist $j \le k$ and $x^1, \ldots, x^q$ in A such that $\{1, \ldots, N\} = \{i_1, \ldots, i_j\} \cup I$ where $I = \bigcup_{\ell=1}^{q} \{i \le N ; X_i = x_i^\ell\}$. Then, as in the third step,

$$\begin{aligned}
\mathbb{E}_\varepsilon \|\sum_{i=1}^{N} \varepsilon_i X_i\| &\le \sum_{i=1}^{k} \|X_i\|^* + \mathbb{E}_\varepsilon \|\sum_{i \in I} \varepsilon_i X_i\| \\
&\le \sum_{i=1}^{k} \|X_i\|^* + \sum_{\ell=1}^{q} \mathbb{E}_\varepsilon \|\sum_{i=1}^{N} \varepsilon_i x_i^\ell\| \\
&\le \sum_{i=1}^{k} \|X_i\|^* + 2qM
\end{aligned}$$

by monotonicity of Rademacher averages and definition of A. (6.23) then simply follows from (6.16). Note that the same applies to sums of independent real *positive* random variables since these share this monotonicity property, and this case is actually one first instructive example for the understanding of the technique.

We next turn to several applications of Theorem 6.17. We first solve the integrability question of almost surely convergent series of independent centered bounded random variables. Proposition 6.14 provided an exponential integrability. It is however known, from (6.10) or Prokhorov's arcsinh inequality e.g. (cf. [Sto]), that in the real case these series are integrable with respect to the function $\exp(x \log^+ x)$. This order of integrability is furthermore best possible. Take indeed $(X_i)_{i \ge 1}$ to be a sequence of independent random variables such that, for each $i \ge 1, X_i = \pm 1$ with equal probability $(2i^2)^{-1}$ and $X_i = 0$ with probability $1 - i^{-2}$. Then $\sum_i \mathbb{E}X_i^2 < \infty$. However, if $S_N = \sum_{i=1}^{N} X_i$ and $\alpha, \epsilon > 0$, for every N,

$$\begin{aligned}
\mathbb{E}\exp(\alpha |S_N| (\log^+ |S_N|)^{1+\epsilon}) &= \sum_{k=0}^{N} \exp(\alpha k (\log^+ k)^{1+\epsilon}) \mathbb{P}\{|S_N| = k\} \\
&\ge \exp(\alpha N (\log N)^{1+\epsilon}) \prod_{i=1}^{N} i^{-2}
\end{aligned}$$

which thus goes to infinity with N.

The following theorem extends to the vector valued case this strong integrability property.

Theorem 6.19. Let (X_i) be independent and symmetric random variables with values in B such that $S = \sum_i X_i$ converges almost surely and $\|X_i\|_\infty \leq a$ for all i. Then, for every $\lambda < 1/a$,

$$\mathbb{E} \exp \lambda \|S\| \log^+ \|S\| < \infty.$$

If the X_i's are merely centered, this holds for all $\lambda < 1/2a$.

Proof. For every N, set $S_N = \sum_{i=1}^N X_i$. We show that

$$\sup_N \mathbb{E} \exp \lambda \|S_N\| \log^+ \|S_N\| < \infty$$

which is enough by Fatou's lemma. We already know from Theorem 6.11 that $\sup_N \mathbb{E}\|S_N\| = M < \infty$ (this can also be deduced directly from the isoperimetric inequality so that this proof is actually self-contained). We use (6.17) together with (6.19) and the comment after Theorem 6.17 concerning truncation to see that, for all integers $k \geq q$ and all real numbers $s, t > 0$,

$$\mathbb{P}\|S_N\| > 8qM + s + t\} \leq \left(\frac{K_0}{q}\right)^k + \mathbb{P}\{\sum_{i=1}^k \|X_i\|^* > s\} + 2\exp\left(-\frac{t^2}{256qM^2}\right).$$

Since, almost surely, $\sum_{i=1}^k \|X_i\|^* \leq ka$,

$$\mathbb{P}\{\|S_N\| > 8qM + ka + t\} \leq \left(\frac{K_0}{q}\right)^k + 2\exp\left(-\frac{t^2}{256qM^2}\right).$$

Let $\epsilon > 0$. For $u > 0$ large enough, set $t = \epsilon u$, $k = [(1-2\epsilon)a^{-1}u]$ (integer part) and

$$q = \left[\frac{\epsilon^2 a}{256M^2} \cdot \frac{u}{\log u}\right].$$

Then, for $u \geq u_0(M, a, \epsilon)$ large enough, $k \geq q$ and

$$\begin{aligned} \mathbb{P}\{\|S_N\| > u\} &\leq \mathbb{P}\{\|S_N\| > 8qM + ka + t\} \\ &\leq \exp(-(1-3\epsilon)a^{-1}u\log u) + 2\exp(-a^{-1}u\log u). \end{aligned}$$

Since this is uniform in N the conclusion follows. If the random variables are only centered, use for example the symmetrization Lemma 6.3. Note again that $\sup_N \mathbb{E}\exp\lambda\|S_N\|\log^+\|S_N\| < \infty$ as soon as the sequence (S_N) is bounded in probability.

Theorem 6.19 is closely related to the best order of growth in function of p of the constants in the L_p-inequalities of J. Hoffmann-Jørgensen (Proposition 6.8). This is the content of the next statement.

Theorem 6.20. There is a universal constant K such that for all $p > 1$ and all finite sequences (X_i) of independent mean zero random variables in $L_p(B)$,

$$\|\sum_i X_i\|_p \leq K\frac{p}{\log p}(\|\sum_i X_i\|_1 + \|\max_i \|X_i\|\|_p).$$

Proof. We may and do assume the X_i's, $i \leq N$, to be symmetric. If r is an integer, we set $X_N^{(r)} = \|X_j\|$ whenever $\|X_j\|$ is the r-th maximum of the sample $(\|X_i\|)_{i\leq N}$ (ties being broken by the index and $X_N^{(r)} = 0$ if $r > N$). Then, if $M = \|\sum_{i=1}^N X_i\|_1$, Theorem 6.17 and (6.19) indicate that, for $k \geq q$ and $s, t > 0$,

$$\mathbb{P}\{\|\sum_{i=1}^N X_i\| > 8qM + 2s + t\} \leq \left(\frac{K_0}{q}\right)^k + \mathbb{P}\{\sum_{r=1}^k X_N^{(r)} > s\} + 2\exp\left(-\frac{t^2}{256qM^2}\right). \tag{6.24}$$

To establish Theorem 6.20, we may assume by homogeneity that $M \leq 1$ and $\|X_N^{(1)}\|_p \leq 1$ ($X_N^{(1)} = \max_{i\leq N}\|X_i\|$). In particular therefore $\mathbb{P}\{X_N^{(1)} > k\} \leq u^{-p}$ for all $u > 0$. By induction over r, for all $u > 0$, one easily sees that

$$\mathbb{P}\{X_N^{(r)} > u\} \leq \mathbb{P}\{\max_{i\leq N}\|X_i\| > u\}\mathbb{P}\{X_N^{(r-1)} > u\},$$

from which we deduce, iterating, that

$$\mathbb{P}\{X_N^{(r)} > u\} \leq (\mathbb{P}\{X_N^{(1)} > u\})^r \leq u^{-rp}.$$

(With some worse irrelevant numerical constants, the same result may be obtained, perhaps more simply, from the successive application of Lemmas 2.5 and 2.6.) Let $u \geq 1$ be fixed. We have $\mathbb{P}\{X_N^{(2)} > u^{2/3}\} \leq u^{-4p/3}$. Further, if ℓ is the smallest integer such that $2^\ell \geq u^2$,

$$\mathbb{P}\{X_N^{(\ell)} > 2\} \leq 2^{-\ell p} \leq u^{-2p} \leq u^{-4p/3}.$$

Hence, the complement of the set $\{X_N^{(1)} \leq u,\ X_N^{(2)} \leq u^{2/3},\ X_N^{(\ell)} \leq 2\}$ has a probability smaller than $\mathbb{P}\{X_N^{(1)} > u\} + 2u^{-4p/3}$. We now apply (6.24). Let k be the smallest integer $\geq u$. On the set $\{X_N^{(1)} \leq u,\ X_N^{(2)} \leq u^{2/3},\ X_N^{(\ell)} \leq 2\}$,

$$\sum_{r=1}^k X_N^{(r)} \leq u + \ell u^{2/3} + 2(k-1) \leq Cu$$

for some constant C. If we now take in (6.24) q to be the smallest integer $\geq \sqrt{u}, s = Cu, t = u$, it follows from the preceding that

$$\begin{aligned} &\mathbb{P}\{\|\sum_{i=1}^{N} X_i\| > 2(C+10)u\} \\ &\leq \exp\left(-u \log\left(\frac{\sqrt{u}}{K_0}\right)\right) + \mathbb{P}\{X_N^{(1)} > u\} + 5u^{-4p/3} + 2\exp\left(-\frac{u^{3/2}}{256}\right). \end{aligned}$$

Standard computations using the integration by parts formula give then the constant $p/\log p$ in the inequality of the theorem. The proof is complete.

The preceding moment inequalities can also be investigated for exponential functions. Recall that for $0 < \alpha < \infty$ we let $\psi_\alpha = \exp(x^\alpha) - 1$ (linear near the origin when $0 < \alpha < 1$ in order for ψ_α to be convex) and denote by $\|\cdot\|_{\psi_\alpha}$ the norm of the Orlicz space L_{ψ_α}. We then have

Theorem 6.21. There is a constant K_α depending on α only such that for all finite sequences (X_i) of independent mean zero random variables in $L_{\psi_\alpha}(B)$, if $0 < \alpha \leq 1$,

$$\|\sum_i X_i\|_{\psi_\alpha} \leq K_\alpha(\|\sum_i X_i\|_1 + \|\max_i \|X_i\|\|_{\psi_\alpha}), \tag{6.25}$$

and if $1 < \alpha \leq 2$,

$$\|\sum_i X_i\|_{\psi_\alpha} \leq K_\alpha((\|\sum_i X_i\|_1 + (\sum_i \|X_i\|_{\psi_\alpha}^\beta)^{1/\beta}) \tag{6.26}$$

where $1/\alpha + 1/\beta = 1$.

Proof. We only give the proof of (6.26). Similar (and even simpler when $0 < \alpha < 1$) arguments are used for (6.25) (cf. [Ta11]). By Lemma 6.3 we reduce to symmetric random variables $(X_i)_{i \leq N}$. We set $d_i = \|X_i\|_{\psi_\alpha}$ so that $\mathbb{P}\{\|X_i\| > u\} \leq 2\exp(-(u/d_i)^\alpha)$. By homogeneity we can assume that $\|\sum_{i=1}^{N} X_i\|_1 \leq 1, \sum_{i=1}^{N} d_i^\beta \leq 1$, and there is no loss in generality to assume the sequence $(d_i)_{i \leq N}$ decreasing. Hence $\sum_{2^i \leq N} 2^i d_{2^i}^\beta \leq 2$. We can find a sequence $\gamma_i \geq 2^i d_{2^i}^\beta$ such that $\gamma_i \geq \gamma_{i+1}/\sqrt{2} \geq \gamma_i/2$ for $i \geq 1$ and $\sum_i \gamma_i \leq 10$ (e.g. $\gamma_i = \sum_{j \geq 1} 2^{-|j-i|/2} 2^j d_{2^j}^\beta$). Let $c_i = (2^{-i}\gamma_i)^{1/\beta}$ so that $c_{i+1}^\alpha \leq c_i^\alpha 2^{-\alpha/2\beta}, c_i^\alpha \leq c_{i+1}^\alpha 2^{3\alpha/2\beta}$. Then $\sum_i 2^i c_i^\beta \leq 10$ and $d_j \leq c_i$ for $j \geq 2^i$. We observe that $\sum_{\ell \geq 2} 4c_\ell (4 \log 4^{\ell+1})^{1/\alpha} \leq C_\alpha < \infty$.

It will be enough to show that for some constants a, c depending on α only, if $u \geq c$ then

$$\mathbb{P}\{\sum_{r \leq u^\alpha} X_N^{(r)} > au\} \leq \exp(-u^\alpha). \tag{6.27}$$

Indeed, if we then take, in (6.24), q to be $2K_0, k$ of the order of u^α, s and t of the order of u, we obtain the tail estimate corresponding to (6.26). In order to establish (6.27), since $\mathbb{P}\{X_N^{(1)} > u\} \leq 2\exp(-u^\alpha)$, it is actually enough to find s, a, c such that when $u \geq c$

$$\mathbb{P}\{\sum_{2^s \leq r \leq u^\alpha} X_N^{(r)} > au\} \leq \exp(-u^\alpha).$$

Fix u (large enough) and denote by n the largest integer such that $2^n \leq u^\alpha$. Suppose that $\sum_{r \leq u^\alpha} X_N^{(r)} > au$ so that $\sum_{\ell=0}^{n} 2^\ell X_N^{(2^\ell)} > a$. Let

$$L = \{s \leq \ell \leq n; X_N^{(2^\ell)} > 2c_\ell(4\log 4^{\ell+1})^{1/\alpha}\}.$$

Then $\sum_{\ell \in L} 2^\ell X_N^{(2^\ell)} > au - C_\alpha > au/2$ for u large enough. For $\ell \in L$, we can find a number a_ℓ of the form $2^m (m \in \mathbb{Z})$ such that $X_N^{(2^\ell)} > a_\ell$ and $a_\ell \geq c_\ell(4\log 4^{\ell+1})^{1/\alpha}$ and $\sum_{\ell \in L} 2^\ell a_\ell > au/4$. There exist disjoint subsets $J_\ell, \ell \in L$, of $\{1, \ldots, N\}$ such that $\mathrm{Card} J_\ell \geq 2^{\ell-1}$ and $\|X_i\| > a_\ell$ for $i \in J_\ell$. Set $I_\ell = J_\ell \backslash \{1, \ldots, 2^{\ell-2}\}$. We have

$$\mathbb{P}\{\forall \ell \in L, X_N^{(2^\ell)} > a_\ell\} \leq \sum \prod_{\ell \in L} \prod_{i \in I_\ell} \mathbb{P}\{\|X_i\| > a_\ell\}$$

where the summation is taken over all possible choices of $(I_\ell)_{\ell \in L}$. Hence

$$\mathbb{P}\{\forall \ell \in L, X_N^{(2^\ell)} > a_\ell\} \leq \prod_{\ell \in L} \left(\sum_{2^{\ell-2} < i \leq N} \mathbb{P}\{\|X_i\| > a_\ell\} \right)^{2^{\ell-2}}.$$

We know that $\mathbb{P}\{\|X_i\| > a_\ell\} \leq 2\exp(-(a_\ell/d_i)^\alpha)$, so

$$\sum_{2^{\ell-2} < i \leq N} \mathbb{P}\{\|X_i\| > a_\ell\} \leq \sum_{j \geq \ell-2} 2^{j+1} \exp(-(a_\ell/c_j)^\alpha).$$

Since $\ell \in L$, we have $(a_\ell/c_\ell)^\alpha \geq (a_\ell/c_\ell)^\alpha/2 + \log 4^{j+2}$ and since $c_{j+1}^\alpha \leq c_j^\alpha 2^{-\alpha/2\beta}, c_j^\alpha \leq c_{j+1}^\alpha 2^{3\alpha/2\beta}$ we clearly have that

$$(a_\ell/c_j)^\alpha \geq (a_\ell/c_\ell)^\alpha/4 + \log 4^{j+2}$$

for $j \geq \ell - 2$ provided s has been taken large enough. It follows that

$$\begin{aligned} \sum_{j \geq \ell-2} 2^{j+1} \exp(-(a_\ell/c_j)^\alpha) &\leq \sum_{j \geq \ell-2} 2^{-j-1} \exp\left(-\frac{1}{4}\left(\frac{a_\ell}{c_\ell}\right)^\alpha\right) \\ &\leq \exp\left(-\frac{1}{4}\left(\frac{a_\ell}{c_\ell}\right)^\alpha\right). \end{aligned}$$

Hence

$$\mathbb{P}\{\forall \ell \le n, X_N^{(2^\ell)} > a_\ell\} \le \exp\left(-\frac{1}{8}\sum_{\ell \in L} 2^\ell \left(\frac{a_\ell}{d_\ell}\right)^\alpha\right).$$

By Hölder's inequality,

$$au/4 \le \sum_{\ell \in L} 2^\ell a_\ell \le \left(\sum_{\ell \in L} 2^\ell \left(\frac{a_\ell}{d_\ell}\right)^\alpha\right)^{1/\alpha} \left(\sum_{\ell \in L} 2^\ell c_\ell^\beta\right)^{1/\beta}$$

so that $\sum\limits_{\ell \in L} 2^\ell (a_\ell/d_\ell)^\alpha \ge (au/40)^\alpha$, and, if we take for example $a = 640$, we get that

$$\mathbb{P}\{\forall \ell \le n, X_N^{(2^\ell)} > a_\ell\} \le \exp(-2t^\alpha).$$

The number of possible choices for the sequence $(a_\ell)_{\ell \le n}$ is less than $\exp u^\alpha$ for c large enough. From this, (6.27) follows and the proof of (6.26) is complete.

Remark 6.22. The preceding proof shows that the constant K_α in (6.26) is bounded in any interval $[1+\epsilon, 2]$. Applying (6.26) to independent copies of a Gaussian random variable, one can then deduce from this observation (and a finite dimensional approximation) that vector valued Gaussian variables are in $L_{\psi_2}(B)$.

To conclude this chapter, we would like to mention that the previous statements are only a few examples of the seemingly large number of variations that one can try with the isoperimetric approach and estimates on large values. As an illustration, let us mention a few more ideas.

The first remark is that in the third step of the proof of Theorem 6.17, possibly other inequalities on Rademacher averages can be used. One may think for example of (6.15). Actually, from the more general (6.14) directly (which was obtained through martingale methods), we already get an interesting inequality in the spirit of Theorem 6.21 which deals with the case $\alpha > 2$. Namely, in the hypotheses of Theorem 6.21, for $2 < \alpha < \infty$ and $1/\alpha + 1/\beta = 1$,

$$\|\sum_i X_i\|_{\psi_\alpha} \le K_\alpha(\|\sum_i X_i\|_1 + \|(\|X_i\|_\infty)\|_{\beta,\infty}). \tag{6.28}$$

Using estimates on large values similar to the ones put forward in the proof of Theorem 6.21, it is possible to improve (6.28) by the isoperimetric method into

$$\|\sum_i X_i\|_{\psi_\alpha} \le K_\alpha(\|\sum_i X_i\|_1 + \|(\|X_i\|_{\psi_s})\|_{\beta,\infty}) \tag{6.29}$$

for $2 < \alpha < s \leq \infty$ (See [Ta11]).

Another remark concerns the possibility in the three step procedure of the proof of Theorem 6.17 to let s, and at some point t too, be *random*. One random bound for the largest values is given by the inequaliity

$$\sum_{i=1}^{k} \|X_i\|^* \leq \beta k^{1/\beta} \|(X_i)\|_{\alpha,\infty},$$

$1 < \alpha < \infty$, $\beta = \alpha/\alpha - 1$. With this choice of s, one can show the following inequalities; as usual, (X_i) is a finite sequence of independent symmetric vector valued random variables. If $2 \leq \alpha < \infty$, for some constant K_α depending only on α and all $t > 0$,

$$\mathbb{P}\{\|\sum_i X_i\| > t(\|\sum_i X_i\|_1 + \|(X_i)\|_{\alpha,\infty})\} \leq K_\alpha \exp(-t^\beta / K_\alpha)\,; \tag{6.30}$$

if $1 < \alpha < 2$,

$$\mathbb{P}\{\|\sum_i X_i\| > K_\alpha \|\sum_i X_i\|_1 + t\|(X_i)\|_{\alpha,\infty}\} \leq K_\alpha \exp(-t^\beta / K_\alpha)\,. \tag{6.31}$$

To illustrate these inequalities, let us prove the second one; it uses (6.15) as opposed to the first one which uses the usual quadratic inequality (Theorem 4.7). We start with (6.21), actually without truncated variables since we need not be concerned with truncation here. We set there s to be random and equal to $\beta k^{1/\beta} \|(X_i)\|_{\alpha,\infty}$. Take further $t = 2qM + k^{1/\beta}\|(X_i)\|_{\alpha,\infty}$, also random therefore, where $M = \|\sum_i X_i\|_1$. It follows that, for $k \geq q$,

$$\begin{aligned} &\mathbb{P}\{\|\sum_i X_i\| > 2qM + (2\beta+1)k^{1/\beta}\|(X_i)\|)_{\alpha,\infty}\} \\ &\leq \left(\frac{K_0}{q}\right)^k + \int_{\{\mathbb{X} \in H\}} \mathbb{P}_\epsilon\{\|\sum_{i\in I} \epsilon_i X_i\| > 2qM + k^{1/\beta}\|(X_i)\|_{\alpha,\infty}\} d\mathbb{P}_X\,. \end{aligned}$$

Let $A = \{(x_i); \mathbb{E}\|\sum_i \epsilon_i x_i\| \leq 2M\}$ so that $\mathbb{P}\{\mathbb{X} \in A\} \geq 1/2$. By definition of I and monotonicity of Rademacher averages,

$$\mathbb{E}_\epsilon \|\sum_{i\in I} \epsilon_i X_i\| \leq 2qM.$$

Now, by inequality (6.15) conditionally on the X_i's,

$$\mathbb{P}_\varepsilon\{\|\sum_{i\in I} \varepsilon_i X_i\| > 2qM + k^{1/\beta}\|(X_i)\|_{\alpha,\infty}\} \leq 2\exp(-k/C_\alpha).$$

Letting for example $q = 2K_0$ we then clearly deduce (6.31). As already mentioned, the proof of (6.30) is similar.

We conclude here these applications although many different techniques might now be combined from what we learned to yield some new useful inequalities. We hope the interested reader has now enough informations to establish from the preceding ideas the tools he will need in its own study.

Notes and references

The study of sums of independent Banach space valued random variables and the introduction of symmetrization ideas were initiated by the work of J.-P. Kahane [Ka1] and J. Hoffmann-Jørgensen [HJ1], [HJ2]. Sections 6.1 and 6.2 basically follow their work (cf. [HJ3]).

Inequality (6.2) has been noticed in [Ale1] and [V-C1]. Theorem 6.1 is due to K. Itô and M. Nisio [I-N] (cf. Chapter 2), extending the classical result of P. Lévy on the line. Lemma 6.2 is usually referred to as Ottaviani's inequality and is part of the folklore. Proposition 6.4 was established by M. Kanter [Kan] as an improvement of various previous concentration inequalities of the same type. It turns out to be rather useful in some cases (cf. Section 10.3). Lemmas 6.5 and 6.6 are easy variations and extensions of the contraction principle and comparison properties (see further [J-M2], [HJ3]).

Proposition 6.7 is due to J. Hoffmann-Jørgensen [HJ2] who used it to establish Theorem 6.11 and Corollary 6.12. Proposition 6.8 is implicit in his proof. Applied to a sum $\sum_i \theta_i x_i$ where (θ_i) is a standard p-stable sequence, it yielded historically the first integrability and moment equivalences results for vector valued stable random variables. Lemma 6.9 is classical and is taken in this form from [G-Z1]. In this paper, E. Giné and J. Zinn establish the moment equivalences of sums of independent (symmetric) random variables of Proposition 6.10. Remark 6.13 has been mentioned to us by J. Zinn. Integrability of sums of independent bounded vector valued random variables has been studied by many authors. Iteration of Hoffmann-Jørgensen's inequality has been used in [J-M2], [Pi3] and [Kue5] for example. Proposition 6.14 is due to A. de Acosta [Ac4] while the content of Remark 6.15 is taken from the paper [A-S].

Among the numerous real (quadratic) exponential inequalities, let us mention the ones by S. Bernstein (cf. [Ho]), A. N. Kolmogorov [Ko1] (cf. [Sto]), Y. Prokhorov [Pro2] (cf. [Sto]), G. Bennett [Ben], etc. We refer to the interesting paper by W. Hoeffding [Ho] for comparison between these inequalities and further developments. See also an inequality by D. K. Fuk and S. V. Nagaev [F-N] (cf. [Na2], [Yu2]). The key

Lemma 6.16 is due to V. Yurinskii [Yu1] (see also [Yu2]). Inequality (6.11) (and some others) has been put forward in [Ac6]. Inequalities (6.12), (6.13) and (6.14) may be found respectively in [K-Z], [Ac5] and [Pi16] (see also [Pi12]). A. de Acosta [Ac5] used (6.13) to establish Theorem 6.19 in cotype 2 spaces. There is also a version of Prokhorov's $\operatorname{arcsinh}$ inequality noticed in [J-S-Z] which may be applied similarly to $\|S_N\| - \mathbb{E}\|S_N\|$.

In the contribution [Led6] (see [L-T2]), in the context of the law of the iterated logarithm, a Gaussian randomization argument is used to decompose the study of sums of independent random variables into two parts: one for which the Gaussian (or Rademacher) concentration properties can be applied conditionally, and a second one which is enriched by an unconditionality property (monotonicity of Rademacher averages). This kind of decomposition is one important feature of the isoperimetric approach (cf. Remark 6.18) and Theorem 1.4 was motivated by this unconditionality property. Our exposition here follows [Ta11]. Theorem 6.19 is in [Ta11] and extends the prior result of [Ac5]. Recently, an alternate proof of Theorem 6.20 has been given by S. Kwapień and J. Szulga [K-S] using very interesting hypercontractive estimates in the spirit of what we discussed in the Gaussian and Rademacher cases (Sections 3.2 and 4.4). On the line, Theorem 6.20 was obtained in [J-S-Z] where a thorough relation to exponential integrability is given. Theorem 6.21 is taken from [Ta11] and improves upon [Kue5], [Ac5]. Inequality (6.29) is also in [Ta11]. Inequalities (6.30) and (6.31) extend to the vector valued case various ideas of [He7], [He8] (see further Lemma 14.4).

Chapter 7. The strong law of large numbers

Chapter 7. The strong law of large numbers

In this chapter and the next one, we present the strong law of large numbers and the law of the iterated logarithm respectively for sums of independent Banach space valued random variables. In this study, the isoperimetric approach of Section 6.3 demonstrates its efficiency. We only investigate extensions to vector valued random variables of some of the classical limit theorems like here the laws of large numbers of Kolmogorov and Prokhorov.

One main feature of the results we present is the equivalence, under *classical* moment conditions, of the *almost sure* limit theorem and the corresponding property in *probability*. In a sense, this can be seen as yet another instance in which the theory is broken into two parts: under a statement in probability (weak statement), prove almost sure properties; then try to understand the weak statement. It is one of the main difficulties of the vector valued setting to control boundedness in probability or tightness of a sum of independent random variables. On the line, this is usually done with orthogonality and moment conditions. In general spaces, one has to either put conditions on the Banach space or to use empirical process methods. Some of these questions will be discussed in the sequel of the work, starting with Chapter 9, especially in the context of the central limit theorem which forms the typical example of a weak statement. As announced, in this chapter and the next one, almost sure limit properties are investigated under assumptions in probability.

In the first part of this chapter, we study a general statement for almost sure limit theorems for sums of independent random variables. It is directly drawn from the isoperimetric approach of Section 6.3 and already presents some interest for real random variables. We introduce it with generalities on strong limit theorems like symmetrization (randomization) and blocking arguments. The second paragraph is devoted to applications to concrete examples like the independent and identically distributed (iid) Kolmogorov-Marcinkiewicz-Zygmund strong laws of large numbers and the laws of Kolmogorov, Brunk, Prokhorov, etc. for independent random variables.

Apart of one example where Radon random variables will be useful, we can adopt the setting of the last chapter and deal with a Banach space B for which there is a countable subset D of the unit ball of the dual space B' such that $\|x\| = \sup_{f \in D} |f(x)|$ for all $x \in B$. X is a random variable with values in B if $f(X)$ is measurable for all f in D. When $(X_i)_{i \in \mathbb{N}}$ is a sequence of (independent) random variables in B we set, as usual, $S_n = X_1 + \cdots + X_n, n \geq 1$.

7.1. A general statement for strong limit theorems

Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent random variables with values in B. Let also (a_n) be a sequence of positive numbers increasing to infinity. We study the *almost sure* behavior of the sequence (S_n/a_n). As described before, such a study in the infinite dimensional setting can only be developed reasonably if one assumes some (*necessary*) boundedness or convergence in *probability.* Recall that a sequence (Y_n) is bounded in probability if for every $\varepsilon > 0$ there exists $A > 0$ such that for all n

$$\mathbb{P}\{\|Y_n\| > A\} < \varepsilon.$$

This kind of hypothesis will be common to all limit theorems discussed here. It allows in particular a simple symmetrization procedure summarized in the next trivial lemma.

Lemma 7.1. Let $(Y_n), (Y_n')$ be independent sequences of random variables such that the sequence $(Y_n - Y_n')$ is almost surely bounded (resp. convergent to 0) and (Y_n) is bounded (resp. convergent to 0) in probability. Then (Y_n) is almost surely bounded (resp. convergent to 0). More quantitatively, if for some numbers M and A,

$$\limsup_{n\to\infty} \|Y_n - Y_n'\| \leq M \quad \text{almost surely}$$

and

$$\limsup_{n\to\infty} \mathbb{P}\{\|Y_n\| > A\} < 1,$$

then

$$\limsup_{n\to\infty} \|Y_n\| \leq 2M + A \quad \text{almost surely.}$$

Given (X_i), let then (X_i') denote an independent copy of the sequence (X_i) and set, for each i, $\widetilde{X}_i = X_i - X_i'$ defining thus independent and symmetric random variables. Lemma 7.1 tells us that under appropriate assumptions in probability on (S_n/a_n), it is enough to study $(\sum_{i=1}^{n} \widetilde{X}_i/a_n)$, reducing to symmetric random variables. From now on, we therefore only describe the various results and the general theorem we have in mind in the *symmetrical* case. This avoids several unnecessary complications about centerings, but, with some care, it would however be possible to study the general case.

As we learned, properties in probability are often equivalent to properties in L_p which are usually more convenient. For sums of independent random varibles, this is shown by Hoffmann-Jørgensen's inequalities (Proposition 6.8) on which relies the following useful lemma.

Lemma 7.2. Let (X_i) be independent and symmetric random variables with values in B. If the sequence (S_n/a_n) is bounded (resp. convergent to 0) in probability, for any $p > 0$ and any bounded sequence (c_n) of positive numbers, the sequence

$$(\mathbb{E}\|\sum_{i=1}^{n} X_i I_{\{\|X_i\|\leq c_n a_n\}}/a_n\|^p)$$

is bounded (resp. convergent to 0).

Proof. We only show the convergence statement. Let (c_n) be bounded by c. By inequality (6.9), for each n,

$$\mathbb{E}\|\sum_{i=1}^{n} X_i I_{\{\|X_i\|\leq c_n a_n\}}\|^p \leq 2.3^p \mathbb{E}(\max_{i\leq n}\|X_i\|^p I_{\{\|X_i\|\leq c_n a_n\}}) + 2(3t_0(n))^p$$

where

$$t_0(n) = \inf\{t > 0 : \mathbb{P}\{\|\sum_{i=1}^{n} X_i I_{\{\|X_i\|\leq c_n a_n\}}\| > t\} \leq (8.3^p)^{-1}\}.$$

When (S_n/a_n) converges to 0 in probability, using the contraction principle in the form of Lemma 6.5, for each $\varepsilon > 0$, $t_0(n)$ is seen to be smaller than εa_n for all n large enough. Concerning the maximum term, by integration by parts and Lévy's inequality (2.7),

$$\begin{aligned}\mathbb{E}(\max_{i\leq n}\|X_i\|^p I_{\{\|X_i\|\leq c_n a_n\}}) &\leq \int_0^{ca_n} \mathbb{P}\{\max_{i\leq n}\|X_i\| > t\}dt^p \\ &\leq 2a_n^p \int_0^{c} \mathbb{P}\{\|S_n\| > ta_n\}dt^p.\end{aligned}$$

The conclusion follows by dominated convergence.

A classical and important observation in the study of strong limit theorems for sums S_n of independent random variables is that it can be developed in quite general situations through *blocks of exponential size.* More precisely, assume there exists a subsequence (a_{m_n}) of (a_n) such that for each n

$$ca_{m_n} \leq a_{m_{n+1}} \leq Ca_{m_n+1} \tag{7.1}$$

where $1 < c \leq C < \infty$. This hypothesis is by no mean restrictive since it can be shown (cf. [Wi]) that for any fixed $M > 1$ one can find a strictly increasing sequence (m_n) of integers such that the preceding holds with $c = M, C = M^3$. We thus assume *throughout this section* that (7.1) holds for some subsequence (m_n) and define, for each n, $I(n)$ as the set of integers $\{m_{n-1}+1, \ldots, m_n\}$. The next lemma then describes the reduction to blocks in the study of the almost sure behavior of (S_n/a_n).

Lemma 7.3. Let (X_i) be independent and symmetric random variables. The sequence (S_n/a_n) is almost surely bounded (resp. convergent to 0) if and only if the same holds for $(\sum_{i\in I(n)} X_i/a_{m_n})$.

Proof. We only show the convergence statement. By the Borel-Cantelli lemma and the Lévy inequality for symmetric random variables (2.6), the sequence

$$(\sup_{k\in I(n)} \| \sum_{i=m_{n-1}+1}^{k} X_i/a_{m_n}\|)$$

converges almost surely to 0. Hence, for almost all ω and for all $\varepsilon > 0$, there exists ℓ_0 such that for all $\ell \geq \ell_0$,

$$\sup_{k\in I(\ell)} \| \sum_{i=m_{\ell-1}+1}^{k} X_i(\omega)\| \leq \varepsilon a_{m_\ell}.$$

Let now n and $j \geq \ell_0$ be such that $m_{j-1} < n \leq m_j$. Then

$$\begin{aligned}
\|S_n(\omega)\| &\leq \|S_{m_{\ell_0-1}}(\omega)\| + \sum_{\ell=\ell_0}^{j-1} \| \sum_{i\in I(\ell)} X_i(\omega)\| + \| \sum_{i=m_{j-1}+1}^{n} X_i(\omega)\| \\
&\leq \|S_{m_{\ell_0-1}}(\omega)\| + \varepsilon \sum_{\ell=1}^{j} a_{m_\ell} \\
&\leq \|S_{m_{\ell_0-1}}(\omega)\| + \varepsilon a_n C \sum_{\ell=0}^{\infty} c^{-\ell}
\end{aligned}$$

where we have used (7.1). Since $c > 1$ and $C < \infty$ the conclusion follows.

As a corollary to Lemmas 7.1 and 7.3 we can state the following equivalence for general independent variables.

Corollary 7.4. Let (X_i) be independent random variables with values in B. Then $S_n/a_n \to 0$ almost surely if and only if $S_n/a_n \to 0$ in probability and $S_{m_n}/a_{m_n} \to 0$ almost surely, and similarly for boundedness.

After these preliminaries, we are in a position to describe the general result about the almost sure behavior of (S_n/a_n). Recall we assume (7.1). By symmetrization and Lemma 7.3, we have to study, thanks to the Borel-Cantelli lemma, convergence of series of the type

$$\sum_n \mathbb{P}\{\| \sum_{i\in I(n)} X_i\| > \varepsilon a_n\}$$

for some, or all, $\varepsilon > 0$. The sufficient conditions we describe for this to hold are obtained by the isoperimetric inequality for product measures in the form of Theorem 6.17. They are of various types. There is the usual assumption in probability on the sequence (S_n/a_n). If the sequence (S_n/a_n) is almost surely bounded, this is also the case for $(\max_{i\leq n}\|X_i\|/a_n)$. This necessary condition on the *norm* of the individual summands X_i is unfortunately not powerful enough in general and has to be complemented with informations on the successive maximum of the sample $(\|X_1\|,\dots\|X_n\|)$. Once this is given, the last sufficient condition deals with *weak* moments assumptions which are kind of optimal.

The announced result can now be stated. Recall K_0 is the absolute constant of the isoperimetric inequality (6.16). If r is an integer, we set $X_{I(n)}^{(r)} = \|X_j\|$ whenever $\|X_j\|$ is the r-th maximum of the sample $(\|X_i\|)_{i\in I(n)}$ (breaking ties by priority of index, and setting $X_{I(n)}^{(r)} = 0$ if $r >$ Card$I(n)$).

Theorem 7.5. Let (X_i) be a sequence of independent and symmetric random variables with values in B. Assume there exist an integer $q \geq 2K_0$ and a sequence (k_n) of integers such that the following hold:

$$\sum_n \left(\frac{K_0}{q}\right)^{k_n} < \infty, \tag{7.2}$$

$$\sum_n \mathbb{P}\{\sum_{r=1}^{k_n} X_{I(n)}^{(r)} > \varepsilon a_{m_n}\} < \infty \tag{7.3}$$

for some $\varepsilon > 0$. Set then, for each n,

$$M_n = \mathbb{E}\|\sum_{i\in I(n)} X_i I_{\{\|X_i\|\leq \varepsilon a_{m_n}/k_n\}}\|,$$
$$\sigma_n = \sup_{f\in D}\Big(\sum_{i\in I(n)} \mathbb{E}(f^2(X_i) I_{\{\|X_i\|\leq\varepsilon a_{m_n}/k_n\}})\Big)^{1/2}.$$

Then, if $L = \limsup_{n\to\infty} M_n/a_{m_n} < \infty$, and, for some $\delta > 0$,

$$\sum_n \exp(-\delta^2 a_{m_n}^2/\sigma_n^2) < \infty, \tag{7.4}$$

we have

$$\sum_n \mathbb{P}\{\|\sum_{i\in I(n)} X_i\| > 10^2\alpha(\varepsilon,\delta,q,L) a_{m_n}\} < \infty \tag{7.5}$$

where $\alpha(\varepsilon,\delta,q,L) = \varepsilon + qL + (\varepsilon L + \delta^2)^{1/2}(q\log\frac{q}{K_0})^{1/2} \leq \varepsilon + qL + q(\varepsilon L+\delta^2)^{1/2}$. Conversely, if (7.5) holds for some (resp. all) $\alpha > 0$ and (7.2) and (7.3) are satisfied, then $L < \infty$ (resp. $L = 0$) and (7.4) holds for some (resp. all) $\delta > 0$.

Proof. There is nothing to prove concerning the sufficiency part of this theorem which readily follows from the inequality of Theorem 6.17 together with (6.18) applied to the sample $(X_i)_{i\in I(n)}$ with $k = k_n$ ($\geq q$ for n large enough), $s = \varepsilon a_{m_n}$ and

$$t = 10^2(\varepsilon L + \delta^2)^{1/2}\left(q\log\frac{q}{K_0}\right)^{1/2} a_{m_n}.$$

(Of course, the numerical constant 10^2 is not the best one, just a convenient number.) The necessity part concerning L is contained in Lemma 7.2. The necessity of (7.4) is based on Kolmogorov's exponential minoration inequality (Lemma 8.1 below). Assume that for all $\alpha > 0$ (the case for some $\alpha > 0$ is similar)

$$\sum_n \mathbb{P}\{\|\sum_{i\in I(n)} X_i\| > \alpha a_{m_n}\} < \infty.$$

Set, for simplicity, $X_i^n = X_i I_{\{\|X_i\|\leq \varepsilon a_{m_n}/k_n\}}, i \in I(n)$, and choose, for each n, f_n in D such that

$$\sigma_n^2 \leq 2\sum_{i\in I(n)} \mathbb{E}f_n^2(X_i^n) \qquad (\leq 2\sigma_n^2).$$

By the contraction principle (second part of Lemma 6.5), we still have that

$$\sum_n \mathbb{P}\{\sum_{i\in I(n)} f_n(X_i^n) > \alpha a_{m_n}\} < \infty.$$

Let $\delta > 0$. If $\eta > 0$ is such that $\delta^2\eta \geq \log(q/K_0)$, by (7.2) it is enough to check (7.4) for the integers n satisfying $a_{m_n}^2 \leq \eta k_n \sigma_n^2$. Let us therefore assume this holds. Recall Lemma 8.1 below and the parameter and constants $\gamma, \varepsilon(\gamma), K(\gamma)$ therein, γ being arbitrary but fixed for our purposes. Let $\alpha > 0$ be small enough in order that $(1+\gamma)\alpha^2 \leq \delta^2$ and $2\alpha\varepsilon\eta \leq \varepsilon(\gamma)$ so that

$$(\alpha a_{m_n})(\varepsilon a_{m_n}/k_n) \leq \alpha\varepsilon\eta\sigma_n^2 \leq \varepsilon(\gamma)\sum_{i\in I(n)} \mathbb{E}f_n^2(X_i^n).$$

Since $L = 0$, it follows from Lemma 7.2 and orthogonality that $\sum_{i\in I(n)} \mathbb{E}f_n^2(X_i^n)/a_{m_n}^2 \to 0$; thus, for all n large enough,

$$\alpha a_{m_n} \geq K(\gamma)(\sum_{i\in I(n)} \mathbb{E}f_n^2(X_i^n))^{1/2}.$$

Lemma 8.1 then exactly implies that

$$\mathbb{P}\{\sum_{i\in I(n)} f_n(X_i^n) > \alpha a_{m_n}\} \geq \exp(-(1+\gamma)\alpha^2 a_{m_n}^2/\sigma_n^2)$$

which gives the result since $(1+\gamma)\alpha^2 \leq \delta^2$. This completes the proof of Theorem 7.5.

Theorem 7.5 expresses that, under (7.2) and (7.3), some necessary and sufficient conditions involving the behavior in probability or expectation and weak moments can be given to describe the almost sure behavior of (S_n/a_n) . Conditions (7.2) and (7.3) look however rather technical and it would be desirable to find if possible simple, or at least easy to be handled, hypotheses on (X_i) in order these conditions be fulfilled. There could be many ways to do this. We suggest a possible one in terms of the probabilities $\mathbb{P}\{\max_{i\in I(n)} \|X_i\| > t\}$ (or $\sum_{i\in I(n)} \mathbb{P}\{\|X_i\| > t\}$). No vector valued structure is of course involved here.

Lemma 7.6. In the notations of Theorem 7.5, assume that, for some $u > 0$,

$$\sum_n \mathbb{P}\{\max_{i\in I(n)} \|X_i\| > ua_{m_n}\} < \infty, \tag{7.6}$$

and that, for some $v > 0$, all n and $t, 0 < t \leq 1$,

$$\mathbb{P}\{\max_{i\in I(n)} \|X_i\| > tva_{m_n}\} \leq \delta_n \exp\left(\frac{1}{t}\right) \tag{7.7}$$

where $\sum_n \delta_n^s < \infty$ for some integer s . Then, for each $q > K_0$, there exists a sequence (k_n) of integers such that $\sum_n (K_0/q)^{k_n} < \infty$ and satisfying

$$\sum_n \mathbb{P}\left\{\sum_{r=1}^{k_n} X_{I(n)}^{(r)} > 2s\left(u + v\left(\log\frac{q}{K_0}\right)^{-1}\right) a_{m_n}\right\} < \infty.$$

Proof. The idea is simply that if the largest element of the sample $(X_i)_{i\in I(n)}$ is exactly estimated by (7.6), the $2s$-th largest one is already small enough so that quite a big number of values after it are under control. If $\mathbb{P}\{\max_{i\in I(n)} \|X_i\| > tva_{m_n}\} \leq 1/2$, and since $X_{I(n)}^{(2s)} > tva_{m_n}$ if and only if $\mathrm{Card}\{i \in I(n)\,;\ \|X_i\| > tva_{m_n}\} \geq 2s$, we have by Lemmas 2.5 and 2.6

$$\begin{aligned}\mathbb{P}\{X_{I(n)}^{(2s)} > tva_{m_n}\} &\leq \Big(\sum_{i\in I(n)} \mathbb{P}\{\|X_i\| > tva_{m_n}\}\Big)^{2s} \\ &\leq \big(2\mathbb{P}\{\max_{i\in I(n)} \|X_i\| > tva_{m_n}\}\big)^{2s} \\ &\leq \left(2\delta_n \exp\left(\frac{1}{t}\right)\right)^{2s}\end{aligned}$$

where we have used hypothesis (7.7) in the last inequality (one can also use the small trick on large values shown in the proof of Theorem 6.20). The choice of $t = t(n) = (\log 1/\sqrt{\delta_n})^{-1}$ bounds the previous probability by $(2\delta_n)^s$ which, by hypothesis, is the term of a summable series. Define then k_n, for each n, to be the integer part of

$$2s\left(\log\frac{q}{K_0}\right)^{-1}\log\frac{1}{\sqrt{\delta_n}}.$$

It is plain that $\sum_n (K_0/q)^{k_n} < \infty$. Now, we have that

$$\sum_{r=1}^{k_n} X_{I(n)}^{(r)} \le 2sX_{I(n)}^{(1)} + k_n X_{I(n)}^{(2s)}$$

from which it follows that, for every n,

$$\begin{aligned}&\mathbb{P}\left\{\sum_{r=1}^{k_n} X_{I(n)}^{(r)} > 2s\left(u + v\left(\log\frac{q}{K_0}\right)^{-1}\right)a_{m_n}\right\}\\ &\le \mathbb{P}\{X_{I(n)}^{(1)} > ua_{m_n}\} + \mathbb{P}\{X_{I(n)}^{(2s)} > t(n)va_{m_n}\}.\end{aligned}$$

The lemma is therefore established.

Remark 7.7. Assume that (7.7) of Lemma 7.6 is strengthened into

$$\mathbb{P}\{\max_{i\in I(n)}\|X_i\| > tva_{m_n}\} \le \frac{1}{t^p}\delta_n$$

for some $v > 0$, all n and $t, 0 < t \le 1$, and some $p > 0$. Then the preceding proof can be easily improved to yield that Lemma 7.6 holds with a sequence (k_n) satisfying

$$\sum_n \frac{1}{k_n^{2ps}} < \infty$$

(or even $\sum_n k_n^{-p's} < \infty$ for $p' > p$). This observation is sometimes useful.

When the independent random variables X_i are identically distributed, and the normalizing sequence (a_n) is regular enough, the first condition in Lemma 7.6 actually implies the second. This is the purpose of the next lemma. The subsequence (m_n) is chosen to be $m_n = 2^n$ for all n. The regularity condition on (a_n) contains the cases of $a_n = n^{1/p}$, $0 < p < \infty$, $a_n = (2nLLn)^{1/2}$, etc, which are the basic examples we have in mind for applications.

Lemma 7.8. Let (a_n) be such that for some $p > 0$ and all $k \le n$, $a_{2^n}^p \ge 2^{n-k}a_{2^k}^p$. Assume (X_i) is a sequence of independent and identically distributed (like X) random variables. Then, if for some $u > 0$,

$$\sum_n 2^n\mathbb{P}\{\|X\| > ua_{2^n}\} < \infty,$$

for all n and $0 < t \leq 1$

$$\mathbb{P}\{\max_{i \in I(n)} \|X_i\| > tua_{2^n}\} \leq 2^n \mathbb{P}\{\|X\| > tua_{2^n}\} \leq \frac{1}{t^{2p}} \delta_n$$

where $\sum_n \delta_n < \infty$.

Proof. For each n set $\gamma_n = 2^n \mathbb{P}\{\|X\| > ua_{2^n}\}$. There exists a sequence (β_n) such that $\beta_n \geq \gamma_n$ and $\beta_n \leq 2\beta_{n+1}$ for every n and satisfying $\sum_n \beta_n < \infty$. Let $0 < t \leq 1$, and $k \geq 1$ be such that $2^{-k} \leq t^p \leq 2^{-k+1}$. If $k < n$,

$$2^n \mathbb{P}\{\|X\| > tua_{2^n}\} \leq 2^k \gamma_{n-k} \leq 2^k \beta_{n-k} \leq 2^{2k} \beta_n \leq 4t^{-2p} \beta_n.$$

If $k \geq n$,

$$2^n \mathbb{P}\{\|X\| > tua_{2^n}\} \leq 2^n \leq 4t^{-2p} 2^{-n}.$$

The conclusion follows with $\delta_n = 4\max(\beta_n, 2^{-n})$.

Notice that Lemma 7.8 enters the setting of Remark 7.7.

7.2. Examples of laws of large numbers

This section is devoted to applications to some classical strong laws of large numbers for Banach space valued random variables of the preceding general Theorem 7.5. Issued from sharp isoperimetric methods, this general result is actually already of interest on the line as will be clear from some of the statements we will obtain here. The results we present follow rather easily from Theorem 7.5.

We do not seek the greatest generality in normalizing sequences and (almost) only deal with the classical strong law of large numbers given by $a_n = n$. That is, we usually say that a sequence (X_i) satisfies the strong law of large numbers (in short SLLN) if $S_n/n \to 0$ almost surely . We sometimes speak of the weak law of large numbers meaning that $S_n/n \to 0$ in probability. When thus $a_n = n$, we may simply take $m_n = 2^n$ as the blocking subsequence. The applications we present deal with the independent and identically distributed (idd) SLLN and the SLLN of Kolmogorov and Prokhorov as typical and classical examples. Further applications can easily be imagined.

The first theorem is the vector valued version of the SLLN of Kolmogorov and Marcinkiewicz-Zygmund for iid random variables. Although this result can be deduced from Theorem 7.5, there is a simpler argument based on Lemma 6.16 and the martingale representation of $\|S_n\| - \mathbb{E}\|S_n\|$. In order however to demonstrate the universal character of the isoperimetric approach, we present the two proofs.

Theorem 7.9. Let $0 < p < 2$. Let (X_i) be a sequence of iid random variables distributed like X with values in B. Then

$$\frac{S_n}{n^{1/p}} \to 0 \text{ almost surely}$$

if and only if

$$\mathbb{E}\|X\|^p < \infty \text{ and } \frac{S_n}{n^{1/p}} \to 0 \text{ in probability.}$$

Proof. Necessity is obvious; indeed, if $S_n/n^{1/p} \to 0$ almost surely, then $X_n/n^{1/p} \to 0$ almost surely from which it follows by the Borel-Cantelli lemma and identical distribution that

$$\sum_n \mathbb{P}\{\|X\| > n^{1/p}\} < \infty$$

which is equivalent to $\mathbb{E}\|X\|^p < \infty$. Turning to sufficiency, it is enough, by Lemma 7.1, to prove the conclusion for the symmetric variable $X - X'$ where X' denotes an independent copy of X. Since $X - X'$ satisfies the same conditions as X, we can assume without loss of generality X itself to be symmetric. By Lemma 7.3, or directly by Lévy's inequality (2.6), it suffices to show that for every $\varepsilon > 0$

$$\sum_n \mathbb{P}\{\|\sum_{i \in I(n)} X_i\| > \varepsilon 2^{n/p}\} < \infty$$

where $I(n) = \{2^{n-1}+1, \dots, 2^n\}$.

1 st proof. For each n, set $u_i = u_i(n) = X_i I_{\{\|X_i\| \le 2^{n/p}\}}, i \in I(n)$. We have:

$$\sum_n \mathbb{P}\{\exists i \in I(n) : u_i \neq X_i\} \le \sum_n 2^n \mathbb{P}\{\|X\| > 2^{n/p}\}$$

which is finite under $\mathbb{E}\|X\|^p < \infty$. It therefore suffices to show that for every $\varepsilon > 0$

$$\sum_n \mathbb{P}\{\|\sum_{i \in I(n)} u_i\| > \varepsilon 2^{n/p}\} < \infty.$$

By Lemma 7.2, we know that

$$\lim_{n\to\infty} \frac{1}{2^{n/p}} \mathbb{E}\|\sum_{i \in I(n)} u_i\| = 0.$$

Hence, it is sufficient to prove that for all $\varepsilon > 0$

$$\sum_n \mathbb{P}\{\|\sum_{i\in I(n)} u_i\| - \mathbb{E}\|\sum_{i\in I(n)} u_i\| > \varepsilon 2^{n/p}\} < \infty.$$

By the quadratic inequality (6.11) and identical distribution, the preceding sum is less than

$$\begin{aligned}\frac{1}{\varepsilon^2}\sum_n \frac{1}{2^{2n/p}}\sum_{i\in I(n)} \mathbb{E}\|u_i\|^2 &\le \frac{1}{\varepsilon^2}\sum_n \frac{1}{2^{n(2/p-1)}}\mathbb{E}(\|X\|^2 I_{\{\|X\|\le 2^{n/p}\}}) \\ &\le \frac{1}{\varepsilon^2}\mathbb{E}(\|X\|^2\sum_n \frac{1}{2^{n(2/p-1)}} I_{\{2^n\ge\|X\|^p\}})\end{aligned}$$

which is finite under $\mathbb{E}\|X\|^p < \infty$. This concludes the 1 st proof.

2 nd proof. We apply the general Theorem 7.5 with Lemmas 7.6 and 7.8 for the control of the large values. $\mathbb{E}\|X\|^p < \infty$ is equivalent to say that for every $\varepsilon > 0$

$$\sum_n 2^n \mathbb{P}\{\|X\| > \varepsilon 2^{n/p}\} < \infty.$$

Let $\varepsilon > 0$ be fixed. By Lemmas 7.6 and 7.8, there is a sequence (k_n) of integers such that $\sum_n 2^{-k_n} < \infty$ and

$$\sum_n \mathbb{P}\{\sum_{r=1}^{k_n} X_{I(n)}^{(r)} > 5\varepsilon 2^{n/p}\} < \infty.$$

Apply then Theorem 7.5, with $q = 2K_0$. As before, we can take $L = 0$ by Lemma 7.2. To check condition (7.4) note that

$$\sigma_n^2 \le 2^n \mathbb{E}(\|X\|^2 I_{\{\|X\|\le 2^{n/p}\}})$$

(at least for all n large enough). The same computation as in the first proof shows that $\sum_n 2^{-2n/p}\sigma_n^2 < \infty$ under $\mathbb{E}\|X\|^p < \infty$ so that (7.4) will hold for every $\delta > 0$. The conclusion of Theorem 7.5 then tells us that, for all $\delta > 0$,

$$\sum_n \mathbb{P}\left\{\left\|\sum_{i\in I(n)} X_i\right\| > 10^2(5\varepsilon + 2K_0\delta)\right\} < \infty.$$

The proof is therefore complete.

At this point, we open a digression on the hypothesis $S_n/n^{1/p} \to 0$ in probability in the preceding theorem on which we shall actually come back in Chapter 9 on type and cotype of Banach spaces. Let us assume we deal here with a Borel random variable X with values in a separable Banach space B. It is known and very easy to show that in finite dimensional spaces, when $\mathbb{E}\|X\|^p < \infty, 0 < p < 2$ (and $\mathbb{E}X = 0$ for $1 \le p < 2$), then

$$\frac{S_n}{n^{1/p}} \to 0 \text{ in probability.}$$

Let us briefly sketch the proof for real random variables. For all $\varepsilon, \delta > 0$

$$\mathbb{P}\{|S_n| > 2\varepsilon n^{1/p}\} \le n\mathbb{P}\{|X| > \delta n^{1/p}\} + \mathbb{P}\left\{\left|\sum_{i=1}^n X_i I_{\{|X_i| \le \delta n^{1/p}\}}\right| > 2\varepsilon n^{1/p}\right\}.$$

If $0 < p < 1$,

$$\left|\sum_{i=1}^n \mathbb{E}(X_i I_{\{|X_i| \le \delta n^{1/p}\}})\right| \le n\mathbb{E}|X|^p(\delta n^{1/p})^{1-p};$$

choose then $\delta = \delta(\varepsilon) > 0$ such that $\mathbb{E}|X|^p \delta^{1-p} = \varepsilon$. If $1 \le p < 2$, by centering,

$$\left|\sum_{i=1}^n \mathbb{E}(X_i I_{\{|X_i| \le \delta n^{1/p}\}})\right| \le n\mathbb{E}(|X| I_{\{|X| > \delta n^{1/p}\}})$$

which can be made smaller than $\varepsilon n^{1/p}$ for all n large enough. Hence, in any case, we can center and write that, for n large,

$$\begin{aligned}
&\mathbb{P}\left\{\left|\sum_{i=1}^n X_i I_{\{|X_i| \le \delta n^{1/p}\}}\right| > 2\varepsilon n^{1/p}\right\} \\
&\le \mathbb{P}\left\{\left|\sum_{i=1}^n X_i I_{\{|X_i| \le \delta n^{1/p}\}} - \mathbb{E}(X_i I_{\{|X_i| \le \delta n^{1/p}\}})\right| > \varepsilon n^{1/p}\right\} \\
&\le \frac{1}{\varepsilon^2 n^{2/p}} \sum_{i=1}^n \mathbb{E}(|X_i|^2 I_{\{|X_i| \le \delta n^{1/p}\}})
\end{aligned}$$

by Chebyshev's inequality. But now, by integration by parts,

$$n^{1-2/p}\mathbb{E}(|X|^2 I_{\{|X| \le \delta n^{1/p}\}}) \le \int_0^\delta nt^p \mathbb{P}\{|X| > tn^{1/p}\} \frac{dt^2}{t^p}$$

so that the conclusion follows by dominated convergence since $\lim_{u\to\infty} u^p \mathbb{P}\{|X| > u\} = 0$ under $\mathbb{E}|X|^p < \infty$.

Note indeed that if X is real symmetric (for example), then $S_n/n^{1/p} \to 0$ in probability as soon as $\lim_{t\to\infty} t^p \mathbb{P}\{|X| > t\} = 0$ (which is actually necessary and sufficient). We shall come back to this and extensions to the vector valued setting in Chapter 9.

From the preceding observation, a finite dimensional approximation argument shows that in arbitrary separable Banach spaces, *when* $0 < p \le 1$, the integrability condition $\mathbb{E}\|X\|^p < \infty$ (and X has mean zero for $p=1$) also implies that $S_n/n^{1/p} \to 0$ in probability. Indeed, since X is Radon and $\mathbb{E}\|X\|^p < \infty$, we can choose for every $\varepsilon > 0$ a finite valued random variable Y (with mean zero when $p=1$ and $\mathbb{E}X = 0$) such that $\mathbb{E}\|X-Y\|^p < \varepsilon$. Letting T_n denote the partial sums associated to independent copies of Y, we have, *by the triangle inequality and since* $p \leq 1$,

$$\mathbb{E}\|S_n - T_n\|^p \leq n\mathbb{E}\|X-Y\|^p \leq \varepsilon n.$$

Y being a finite dimensional random variable, $T_n/n^{1/p} \to 0$ in probability and the claim follows immediately. Theorem 7.9 therefore states in this case (i.e. $0 < p \leq 1$) that

$$\frac{S_n}{n^{1/p}} \to 0 \text{ almost surely if and only if } \mathbb{E}\|X\|^p < \infty$$

(and $\mathbb{E}X = 0$ for $p=1$). In particular, we recover the extension to separable Banach space of the classical iid SLLN of Kolmogorov.

Corollary 7.10. Let X be a Borel random variable with values in a separable Banach space B. Then

$$\frac{S_n}{n} \to 0 \text{ almost surely}$$

if and only if $\mathbb{E}\|X\| < \infty$ and $\mathbb{E}X = 0$.

The proof of this result presented as a corollary of Theorems 7.5 and 7.9 is of course too complicated and a rather elementary direct proof can be given (see e.g. [HJ3]). It is however instructive to deduce it from general methods.

The preceding elementary approximation argument *does not* extend to the case $1 < p < 2$. It would require an inequality like

$$\mathbb{E}\left\|\sum_i Y_i\right\|^p \leq C \sum_i \mathbb{E}\|Y_i\|^p$$

for every finite sequence (Y_i) of independent centered random variables with values in B, C depending on B and p. Such an inequality does not hold in general Banach space and actually defines the spaces of *type* p discussed later in Chapter 9. Mimicking the preceding argument we can however already announce that in (separable) spaces of type $p, 1 < p < 2$, $S_n/n^{1/p} \to 0$ almost surely if and only if $\mathbb{E}\|X\|^p < \infty$ and $\mathbb{E}X = 0$. We shall see how this property is actually characteristic of type p spaces (see Theorem 9.21 below).

The following example completes this discussion, showing in particular that $S_n/n^{1/p}, 1 < p < 2$, cannot in general tend to 0 even if X has very strong integrability properties. The example is adapted for further purposes (the constructed random variable is moreover pregaussian in the sense of Section 9.3).

Example 7.11. In c_0 , the separable Banach space of all real sequences tending to 0 equipped with the sup-norm, there exists, for all decreasing sequences (α_n) of positive numbers tending to 0 , an almost surely bounded and symmetric random variable X such that $(S_n/n\alpha_n)$ does not tend to 0 in probability.

Proof. Let $(\xi_k)_{k\geq 1}$ be independent with distribution

$$\mathbb{P}\{\xi_k = +1\} = \mathbb{P}\{\xi_k = -1\} = \frac{1}{2}(1 - \mathbb{P}\{\xi_k = 0\}) = \frac{1}{\log(k+1)}.$$

Define $\beta_k = \sqrt{\alpha_n}$ whenever $2^{n-1} \leq k < 2^n$ and take X to be the random variable in c_0 with coordinates $(\beta_k \xi_k)_{k\geq 1}$. Then X is clearly symmetric and almost surely bounded. However $(S_n/n\alpha_n)$ does not tend to 0 in probability. Indeed, denote by $(\xi_{k,i})_i$ independent copies of (ξ_k) . Set further, for each $k, n \geq 1$,

$$E_{n,k} = \bigcap_{i=1}^{n} \{\xi_{k,i} = 1\}, \ A_n = \bigcup_{k\leq 2^n} E_{n,k}.$$

Clearly $\mathbb{P}(E_{n,k}) = (\log(k+1))^{-n}$ and

$$\mathbb{P}(A_n) = 1 - \prod_{k\leq 2^n} \mathbb{P}(E_{n,k}^c) = 1 - \prod_{k\leq 2^n} \left(1 - \frac{1}{(\log(k+1))^n}\right).$$

Therefore, as is easily seen, $\mathbb{P}(A_n) \to 1$. Now, if (e_k) is the canonical basis of c_0 ,

$$\frac{S_n}{n\alpha_n} = \sum_{k=1}^{\infty} \frac{1}{n\alpha_n} \left(\sum_{i=1}^{n} \xi_{k,i}\right) \beta_k e_k.$$

On A_n ,

$$\frac{\|S_n\|}{n\alpha_n} \geq \max_{k\leq 2^n} \frac{1}{n\alpha_n} \beta_k \left|\sum_{i=1}^{n} \xi_{k,i}\right| \geq \frac{\sqrt{\alpha_n}}{\alpha_n} = \frac{1}{\sqrt{\alpha_n}}.$$

Hence, since $\alpha_n \to 0$, for every $\varepsilon > 0$,

$$\liminf_{n\to\infty} \mathbb{P}\left\{\frac{\|S_n\|}{n\alpha_n} > \varepsilon\right\} \geq \liminf_{n\to\infty} \mathbb{P}(A_n) = 1$$

which establishes the claim.

After having extended to the vector valued setting the iid SLLN, we turn to the SLLN for independent but not necessarily identically distributed random variables. Here again, we restrict to classical statements like the SLLN of Kolmogorov. This SLLN states that if (X_i) is a sequence of independent mean zero real random variables such that

$$\sum_i \frac{\mathbb{E}X_i^2}{i^2} < \infty,$$

then the SLLN holds, i.e. $S_n/n \to 0$ almost surely. From this result, together with a truncation argument, Kolmogorov deduced his idd statement. The next theorem points toward a general extension of this result already of interest on the line. It is characterized by a careful balance between conditions on the norm of the X_i 's and assumptions on their weak moments. The subsequent corollary is perhaps a more practical result. We take again our framework of non necessarily Radon random variables.

Theorem 7.12. Let (X_i) be a sequence of independent random variables with values in B. Assume that

$$\frac{X_i}{i} \to 0 \text{ almost surely} \tag{7.8}$$

and

$$\frac{S_n}{n} \to 0 \text{ in probability.} \tag{7.9}$$

Assume further that for some $v > 0$, all n and $t, 0 < t \leq 1$,

$$\mathbb{P}\{\max_{i \in I(n)} \|X_i\| > tv2^n\} \leq \delta_n \exp\left(\frac{1}{t}\right) \tag{7.10}$$

where $\sum_n \delta_n^s < \infty$ for some $s > 0$, and that, for each $\delta > 0$,

$$\sum_n \exp\left(-\delta 2^{2n} / \sup_{f \in D} \sum_{i \in I(n)} \mathbb{E}(f^2(X_i) I_{\{\|X_i\| \leq 2^n\}})\right) < \infty \tag{7.11}$$

(where we recall that $I(n) = \{2^{n-1}+1, \ldots, 2^n\}$). Then the SLLN holds, i.e.

$$\frac{S_n}{n} \to 0 \text{ almost surely.}$$

Proof. We simply apply Theorem 7.5 and Lemma 7.6. If we define $Y_i = X_i I_{\{\|X_i\| \leq i\}}$, by (7.8), almost surely $Y_i = X_i$ for every i large enough so that it clearly suffices to prove the result for the sequence (Y_i) instead of (X_i). We therefore assume that $\|X_i\| \leq i$ almost surely. If (X_i') is an independent copy of the sequence (X_i), the sequence of symmetric variables $(X_i - X_i')$ will satisfy the same kind of hypotheses as (X_i). By (7.9) and Lemma 7.1, we can therefore reduce to the case of a symmetric sequence (X_i). In Theorem 7.5, whatever the choice of (k_n) will be, we can take $L = 0$ thanks to (7.9) (Lemma 7.2). Since $X_i/i \to 0$ almost surely, for every $u > 0$,

$$\sum_n \mathbb{P}\{\max_{i\in I(n)} \|X_i\| > u2^n\} < \infty.$$

Summarizing the conclusions of Lemma 7.6 and Theorem 7.5, for all $u, \delta > 0$ and all $q \geq 2K_0$, and for s assumed to be an integer,

$$\sum_n \mathbb{P}\left\{\left\|\sum_{i\in I(n)} X_i\right\| > 10^2\left[2s\left(u + v\left(\log\frac{q}{K_0}\right)^{-1}\right) + q\delta\right]2^n\right\} < \infty.$$

It follows obviously that

$$\sum_n \mathbb{P}\left\{\left\|\sum_{i\in I(n)} X_i\right\| > \varepsilon 2^n\right\} < \infty$$

for every $\varepsilon > 0$, hence the conclusion by Lemma 7.3.

Corollary 7.13. Under the hypotheses of Theorem 7.10 but with (7.10) replaced by

$$\sum_n \left(\frac{1}{2^{np}}\sum_{i\in I(n)} \mathbb{E}\|X_i\|^p\right)^s < \infty \tag{7.10$'$}$$

for some $p > 0$ and $s > 0$, the SLLN is satisfied.

Proof. Simply note that, for every n,

$$\begin{aligned}\sum_{i\in I(n)} \mathbb{P}\{\|X_i\| > tv2^n\} &\leq \frac{1}{(tv)^p}\frac{1}{2^{np}}\sum_{i\in I(n)} \mathbb{E}\|X_i\|^p \\ &\leq C(p)\exp\left(\frac{1}{t}\right)\frac{1}{v^p}\frac{1}{2^{np}}\sum_{i\in I(n)} \mathbb{E}\|X_i\|^p\end{aligned}$$

from which (7.10) of Theorem 7.12 follows. Note that the sums $\sum_{i\in I(n)} \mathbb{E}\|X_i\|^p$ in (7.10$'$) can also be replaced, if one wishes it, by expressions of the type

$$\sup_{t>0} t^p \sum_{i\in I(n)} \mathbb{P}\{\|X_i\| > t\}.$$

In Theorem 7.12 (and Corollary 7.13), conditions (7.8) and (7.9) are of course necessary, (7.9) describing the usual assumption in probability on the sequence (S_n). For *real* centered random variables, this condition

(7.9) is *automatically satisfied under* (7.11) and the real statement such obtained is sharp. Note also that, under (7.8), it is legitimate (and we used it in the proof of Theorem 7.12) to assume that $\|X_i\|_\infty \le i$ for all i. This is sometimes convenient when various statements have to be compared; for example, (7.10) (via (7.10′)) holds then under the stronger condition

$$\sum_i \frac{\mathbb{E}\|X_i\|^p}{i^p} < \infty$$

which is in this case seen to be weaker and weaker as p increases. Then this condition implies (7.8) and (7.11) provided $p \le 2$ and we have therefore the following corollary. (Since no weak moments are involved, it might be obtained simpler from Lemma 6.16, see [K-Z].)

Corollary 7.14. Let (X_i) be a sequence of independent random variables with values in B. If for some $1 \le p \le 2$

$$\sum_i \frac{\mathbb{E}\|X_i\|^p}{i^p} < \infty,$$

the SLLN holds, i.e. $S_n/n \to 0$ almost surely, if and only if the weak law of large numbers holds, i.e. $S_n/n \to 0$ in probability.

Along the same line of ideas, Theorem 7.12 also contains extensions of Brunk's SLLN. Brunk's theorem in the real case states that if (X_i) are independent with mean zero satisfying

$$\sum_i \frac{\mathbb{E}|X_i|^p}{i^{p/2+1}} < \infty \quad \text{for some } p \ge 2,$$

then the SLLN holds. To include this result, simply note that for $p \ge 2$

$$\frac{1}{2^{2n}} \sum_{i \in I(n)} \mathbb{E} f^2(X_i) \le \left(\frac{1}{2^{n(p/2+1)}} \sum_{i \in I(n)} \mathbb{E}|f(X_i)|^p \right)^{2/p}$$

for every n and f in D.

One feature of the isoperimetric approach at the basis of Theorems 7.5 and 7.12 is a common treatment of the SLLN of Kolmogorov and Prokhorov. As easily as we obtained Theorem 7.12 from the preceding section, we get an extension of Prokhovov's theorem to the vector valued case. We still work under conditions (7.9) and (7.11) but reinforce (7.8) in

$$\|X_i\|_\infty \le i/LLi \quad \text{for each } i$$

where $LLt = L(Lt)$ and $Lt = \max(1, \log t), t \geq 0$. This boundedness assumption provides the exact bound on large values and actually fits (7.10) of Theorem 7.12. Indeed, for each n and t, $0 < t \leq 1$,

$$\mathbb{P}\{\max_{i \in I(n)} \|X_i\| > 2t2^n\} \leq \delta_n \exp\left(\frac{1}{t}\right)$$

with $\delta_n = \exp(-2LL2^n)$ which is summable. We thus obtain as a last corollary the following version of Prokhorov's SLLN. Note that under the preceding boundedness assumption on the X_i's, condition (7.11) becomes necessary; the proof follows the necessity portion in Theorem 7.5 and we therefore do not detail it.

Corollary 7.15. Let (X_i) be a sequence of independent random variables with values in B. Assume that, for every i,

$$\|X_i\|_\infty \leq i/LLi.$$

Then the SLLN is satisfied if and only if

$$\frac{S_n}{n} \to 0 \text{ in probability}$$

and

$$\sum_n \exp\left(-\delta 2^{2n} / \sup_{f \in D} \sum_{i \in I(n)} \mathbb{E} f^2(X_i)\right) < \infty$$

for every $\delta > 0$ (where $I(n) = \{2^{n-1}+1, \ldots, 2^n\}$).

Notes and references

Various expositions on the strong laws of large numbers (SLLN) for sums of independent real random variables may be found in the classical works e.g. [Lé1], [Gn-K], [Re], [Sto], [Pe] etc. In particular, Lemmas 7.1 and 7.3 are clearly presented in [Sto] and the vector valued situation does not make any difference.

This chapter based on the isoperimetric approach of [Ta11], [Ta9] presented in Section 6.3 follows the paper [L-T5]. In particular, Theorem 7.5 and Lemma 7.6 are taken from there.

The extension of the Marcinkiewicz-Zygmund SLLN (Theorem 7.9) is due independently to A. de Acosta [Ac6] and T.A. Azlarov and N. A. Volodin [A-V], with the first proof. The classical iid SLLN of Kolmogorov in separable Banach spaces (Corollary 7.10) was established by E. Mourier back in the early fifties [Mo] (see also [F-M1], [F-M2]). A simple proof may be found in [HJ3]. The non-separable version of this result which is not discussed in this text has given recently rise to many developments related to measure theory; see for example [HJ5], [Ta3], and, in the context of empirical processes, [V-C1], [V-C2], [G-Z2]. Example 7.11 is taken from [C-T1].

Theorem 7.12 comes from [L-T5]. For further developments, cf. [Al2]. The real valued statement, at least in the form of Corollary 7.13, can be obtained as a consequence of the Fuk-Nagaev inequality (cf. [F-N], [Yu2]). Let us mention that a suitable vector valued version of the SLLN of S. V. Nagaev [Na1], [Na2] seems still to be found; cf. [Al3] in this regard. Corollary 7.14 is due to J. Kuelbs and J. Zinn [K-Z] which extended results of A. Beck [Be] and J. Hoffmann-Jørgensen and G. Pisier [HJ-P] in special classes of spaces (cf. Chapter 9). The work of J. Kuelbs and J. Zinn was important in realizing that under an assumption in probability no conditions have to be imposed on the spaces. In a special class of Banach spaces however, see also [He5]. Brunk's SLLN appeared in [Br] and was first investigated in Banach spaces in [Wo3]. Extensions of Prokhorov's SLLN [Pro2] was undertaken in [K-Z], [He6], [Al1] (where in particular necessity was shown) and the final result obtained in [L-T4] (see also [L-T5]). Applications of the isoperimetric method to strong limit theorems for trimmed sums of iid random variables are further described in [L-T5].

Chapter 8. The law of the iterated logarithm

Chapter 8. The law of the iterated logarithm

This chapter is devoted to the classical laws of the iterated logarithm of Kolmogorov and Hartman-Wintner-Strassen in the vector valued setting. These extensions both enlighten the scalar statements and describe various new interesting phenomena in the infinite dimensional setting. As in the previous chapter on the strong law of large numbers, the isoperimetric approach proves to be an efficient tool in this study. The main results described here show again how the strong *almost sure* statement of the law of the iterated logarithm reduces to the corresponding (necessary) one in *probability*, under moment conditions similar to the ones of the scalar case.

As the law of large numbers and the central limit theorem, the law of the iterated logarithm (in short LIL) is a vast subject in Probability theory. We only concentrate here on the classical (but typical) forms of the LIL for sums of independent Banach space valued random variables. We first describe , starting from the real case, the extension of Kolmogorov's LIL. In Section 8.2, we describe the Hartman-Wintner-Strassen form of the (iid) LIL in Banach space and characterize the random variables which satisfy it. A last survey paragraph is devoted to a discussion on various results and questions about identification of the limits in the vector valued LIL.

In all this chapter, if $(X_i)_{i\in\mathbb{N}}$ is a sequence of random variables, we set, as usual, $S_n = X_1 + \cdots + X_n$, $n \geq 1$. Recall also that LL denotes the iterated logarithm function, that is, $LLt = L(Lt)$ and $Lt = \max(1, \log t)$, $t \geq 0$.

8.1. The law of the iterated logarithm of Kolmogorov

Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent real mean zero random variables such that $\mathbb{E}X_i^2 < \infty$ for all i. Set, for each n, $s_n = (\sum_{i=1}^{n} \mathbb{E}X_i^2)^{1/2}$.

Assume the sequence (s_n) increases to infinity. Assume further that for some sequence (η_i) of positive number tending to 0,

$$\|X_i\|_\infty \leq \eta_i s_i/(LLs_i^2)^{1/2} \qquad \text{for every } i\,.$$

Then, *Kolmogorov's* LIL *state that*, with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{(2s_n^2 LLs_n^2)^{1/2}} = 1\,. \tag{8.1}$$

The proof of the upper bound in (8.1) is based on the exponential inequality of Lemma 1.6 applied to the sums S_n of independent mean zero random variables. The lower bound, somewhat more complicated, relies on Kolmogorov's converse exponential inequality described in the following lemma. Its proof (cf. [Sto]) is a precise amplification of the argument leading to (4.2).

Lemma 8.1. Let (X_i) be a finite sequence of independent mean zero real random variables such that $\|X_i\|_\infty \le a$ for all i. Then, for every $\gamma > 0$, there exist positive numbers $K(\gamma)$ (large enough) and $\varepsilon(\gamma)$ (small enough) depending on γ only such that for every t satisfying $t \ge K(\gamma)b$ and $ta \le \varepsilon(\gamma)b^2$ where $b = (\sum_i \mathbb{E}X_i^2)^{1/2}$,

$$\mathbb{P}\left\{\sum_i X_i > t\right\} \ge \exp[-(1+\gamma)t^2/2b^2].$$

The next theorem presents the extension to Banach space valued random variables of the LIL of Kolmogorov. This extension involves a careful balance between conditions on the norms of the random variables and weak moment assumptions. Since tightness properties are unessential in this first section, we describe the result in our setting of a Banach space B for which there exists a countable subset D of the unit ball of the dual space such that $\|x\| = \sup_{f\in D} |f(x)|$ for all x in B. X is a random variable with values in B if $f(X)$ is measurable for every f in D.

Theorem 8.2. Let B be as before and let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent random variables with values in B such that $\mathbb{E}f(X_i) = 0$ and $\mathbb{E}f^2(X_i) < \infty$ for each i and f in D. Set, for each n, $s_n = \sup_{f\in D}(\sum_{i=1}^n \mathbb{E}f^2(X_i))^{1/2}$, assumed to increase to infinity. Assume further that for some sequence (η_i) of positive numbers tending to 0 and all i

$$\|X_i\|_\infty \le \eta_i s_i/(2LLs_i^2)^{1/2}. \tag{8.2}$$

Then, if the sequence $(S_n/(2s_n^2 LLs_n^2)^{1/2})$ converges to 0 in probability, with probability one,

$$\limsup_{n\to\infty} \frac{\|S_n\|}{(2s_n^2 LLs_n^2)^{1/2}} = 1. \tag{8.3}$$

This type of statement clearly shows in what direction the extension of the real result has to be understood. The proof of Theorem 8.2 could seem to be somewhat involved. Let us mention however, and this will be accomplished in a first step, that the proof that the $\limsup$ in (8.3) is finite (less than some numerical constant) is rather easy on the basis of the isoperimetric approach of Section 6.3. The fact that it is actually less than 1 requires then some technicalities. The lower bound reproduces the real case.

Proof. To simplify the notations, let us set, for each n, $u_n = (2LLs_n^2)^{1/2}$. As announced, we first show that

$$\limsup_{n\to\infty} \frac{\|S_n\|}{s_n u_n} \leq M \tag{8.4}$$

almost surely for some numerical constant M. To this aim, replacing (X_i) by $(X_i - X_i')$ where (X_i') is an independent copy of the sequence (X_i), and since (by centering, cf. (2.5))

$$s_n \leq \sup_{f\in D}\left(\sum_{i=1}^n \mathbb{E}f^2(X_i - X_i')\right)^{1/2} \leq 2s_n\,,$$

we can assume by Lemma 7.1 that we deal with symmetric variables. For each n, define m_n as the smallest integer m such that $s_m > 2^n$. It is easily seen that

$$\frac{s_{n+1}}{s_n} \sim 1\,, \qquad s_{m_n} \sim 2^n\,, \qquad \frac{s_{m_{n+1}}}{s_{m_n}} \sim 2\,.$$

By the Borel-Cantelli lemma, we need show that

$$\sum_n \mathbb{P}\{\max_{m_{n-1}<m\leq m_n} \frac{\|S_n\|}{s_m u_m} > M\} < \infty\,.$$

Using the preceding and Lévy's inequality (2.6) (increasing M), it suffices to prove that

$$\sum_n \mathbb{P}\{\|S_{m_n}\| > M s_{m_n} u_{m_n}\} < \infty\,.$$

We make use of the isoperimetric inequality of Theorem 6.17 together with (6.18) which we apply to the sample of independent and symmetric random variables $(X_i)_{i\leq m_n}$ for each n. Assuming for simplicity the sequence (η_i) be bounded by 1, we take there $q = 2K_0$, $k = [u_{m_n}^2]+1$, $s = t = 20\sqrt{q}\, s_{m_n} u_{m_n}$. For these choices, by (8.2),

$$\sum_{i=1}^k \|X_i\|^* \leq (u_{m_n}^2 + 1)\frac{s_{m_n}}{u_{m_n}} \leq s\,.$$

Since $(s_n/s_n u_n)$ converges to 0 in probability, by Lemma 7.2, $\mathbb{E}\|S_n\|/s_n u_n \to 0$. Hence, at least for n large enough, m^2 in Theorem 6.17 can be bounded, using (6.18), by $2s_{m_n}^2$. It thus follows from (6.17) that, for large n,

$$\mathbb{P}\{\|S_{m_n}\| > (60\sqrt{2K_0}+1)s_{m_n}u_{m_n}\} \leq 2^{-u_{m_n}^2} + 2\exp(-u_{m_n}^2)$$

which gives the result since $u_{m_n}^2 \sim 2LL4^n$.

Note that this proof shows (8.4) already when the sequence $(S_n/s_n u_n)$ is only bounded in probability, which is of course necessary.

We now turn to the more delicate proof that the $\limsup$ is actually equal to 1 . We begin by showing it is less than 1 . Since $S_n/s_n u_n \to 0$ in probability, observe first that by symmetrization, Lemma 7.2 and centering (Lemma 6.3), we have both that

$$\lim_{n\to\infty} \frac{1}{s_n u_n} \mathbb{E}\|S_n\| = \lim_{n\to\infty} \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i X_i\right\| = 0 \tag{8.5}$$

where as usual (ε_i) denotes a Rademacher sequence independent of (X_i) .

For $\rho > 1$, let m_n for each n be the smallest m such that $s_m > \rho^n$. As before for $\rho = 2$ we have that

$$s_{m_n} \sim \rho^n , \qquad \frac{s_{m_{n+1}}}{s_{m_n}} \sim \rho .$$

To establish the claim it will be sufficient to show that for every $\varepsilon > 0$ and $\rho > 1$,

$$\sum_n \mathbb{P}\{\|S_{m_n}\| > (1+\varepsilon) s_{m_n} u_{m_n}\} < \infty . \tag{8.6}$$

Indeed, in order that $\limsup_{n\to\infty} \|S_n\|/s_n u_n \le 1$ almost surely, it suffices, by the Borel-Cantelli lemma, that for every $\delta > 0$ there is an increasing sequence (m_n) of integers such that

$$\sum_n \mathbb{P}\{\max_{m_{n-1} < m \le m_n} \frac{\|S_m\|}{s_m u_m} > (1+2\delta)\} < \infty .$$

A simple use of Ottaviani's inequality (Lemma 6.2) together with (8.5) shows that, for all n large enough,

$$\begin{aligned} \mathbb{P}\{\max_{m_{n-1} < m \le m_n} \frac{\|S_m\|}{s_m u_m} > (1+2\delta)\} &\le \mathbb{P}\{\max_{m_{n-1} < m \le m_n} \|S_m\| > (1+2\delta) s_{m_{n-1}} u_{m_{n-1}}\} \\ &\le 2\mathbb{P}\{\|S_{m_n}\| > (1+\delta) s_{m_{n-1}} u_{m_{n-1}}\} . \end{aligned}$$

Now, for $\delta > 0$, there exist $\varepsilon > 0$ and $\rho > 1$ such that if m_n is defined from ρ as before, for all large n ,

$$(1+\delta) s_{m_{n-1}} u_{m_{n-1}} \ge (1+\varepsilon) s_{m_n} u_{m_n} .$$

So, it is enough to show (8.6).

The proof is based on a finite dimensional approximation argument through some *entropy* estimate. Let now $\varepsilon > 0$ and $\rho > 1$ be fixed. For f, g in D, and every n, set

$$d_2^n(f,g) = \left(\frac{1}{s_{m_n}^2} \sum_{i=1}^{m_n} \mathbb{E}(f-g)^2(X_i) \right)^{1/2} .$$

Recall $N(D, d_2^n; \varepsilon)$ denote the minimal number of elements g in D such that for every f in D there exists such a g with $d_2^n(f,g) < \varepsilon$. For every n, define

$$\alpha_n = \frac{1}{s_{m_n} u_{m_n}} \mathbb{E} \left\| \sum_{i=1}^{m_n} \varepsilon_i X_i \right\|$$

which tends to 0 when n goes to infinity by (8.5).

Lemma 8.3. For every n large enough,

$$N(D, d_2^n; \varepsilon) \le \exp(\alpha_n u_{m_n}^2) .$$

Proof. Suppose this is not the case. Then, infinitely often in n, there exists $D_n \subset D$ such that for any $f \neq g$ in D_n, $d_2^n(f,g) \ge \varepsilon$ and

$$\mathrm{Card} D_n = [\exp(\alpha_n u_{m_n}^2)] + 1 .$$

By Lemma 1.6 and (8.2), for $h = f - g$, $f \neq g$ in D_n, and n large,

$$\mathbb{P}\left\{ \frac{1}{s_{m_n}^2} \sum_{i=1}^{m_n} h^2(X_i) < \frac{\varepsilon^2}{2} \right\} \le \mathbb{P}\left\{ \frac{1}{s_{m_n}^2} \sum_{i=1}^{m_n} (-h^2(X_i) + \mathbb{E}h^2(X_i)) > \frac{\varepsilon^2}{2} \right\} \le \exp(-u_{m_n}^2) .$$

For n large enough,

$$\mathrm{Card} D_n \exp(-u_{m_n}^2) < \frac{1}{4} .$$

It follows that, infinitely often in n,

$$\mathbb{P}\left\{ \forall f \neq g \text{ in } D_n , \ \frac{1}{s_{m_n}^2} \sum_{i=1}^{m_n} (f-g)^2(X_i) \ge \frac{\varepsilon^2}{2} \right\} \ge \frac{3}{4} .$$

We would like to apply, conditionally on the X_i's, the Sudakov type minoration inequality put forward in Proposition 4.13. To this aim, note first that by (8.5), with high probability, for example bigger than $3/4$,

$$\frac{1}{s_{m_n}}\mathbb{E}_\varepsilon\left\|\sum_{i=1}^{m_n}\varepsilon_i X_i\right\| \leq \frac{\varepsilon^2}{K}\ \frac{u_{m_n}}{\max\limits_{i\leq m_n}\eta_i}$$

for all n large enough, and thus, by (8.2), $i \leq m_n$,

$$\frac{\|X_i\|_\infty}{s_{m_n}} \leq \frac{\max\limits_{j\leq m_n}\eta_j}{u_{m_n}} \leq \varepsilon^2\left(\frac{K}{s_{m_n}}\mathbb{E}_\varepsilon\left\|\sum_{i=1}^{m_n}\varepsilon_i X_i\right\|\right)^{-1}$$

where K is the numerical constant of Proposition 4.13. This proposition then shows that, with probability bigger than $1/2$,

$$\frac{1}{s_{m_n}}\mathbb{E}_\varepsilon\left\|\sum_{i=1}^{m_n}\varepsilon_i X_i\right\| \geq \frac{1}{K}\cdot\frac{\varepsilon}{\sqrt{2}}\ (\log\ \mathrm{Card}D_n)^{1/2} \geq \frac{\varepsilon\sqrt{\alpha_n}u_{m_n}}{\sqrt{2}\,K}\,.$$

Therefore, integrating, infinitely often in n, $\alpha_n \geq \varepsilon\sqrt{\alpha_n}/2\sqrt{2}K$ which leads to a contradiction since $\alpha_n \to 0$. The proof of Lemma 8.3 is complete.

We can now establish (8.6). According to Lemma 8.3, we denote, for each n (large enough) and f in D, by $g_n(f)$ an element of D such that $d_2^n(f, g_n(f)) < \varepsilon$ in such a way that the set D_n of all $g_n(f)$ has a cardinality less than $\exp(\alpha_n u_{m_n}^2)$. We write that

$$\left\|\sum_{i=1}^{m_n} X_i\right\| \leq \sup_{g\in D_n}\left|\sum_{i=1}^{m_n} g(X_i)\right| + \sup_{h\in D_n'}\left|\sum_{i=1}^{m_n} h(X_i)\right|$$

where $D_n' = \{f - g_n(f); f\in D\}$. The main observation concerning D_n' is that

$$\sup_{h\in D_n'}\left(\sum_{i=1}^{m_n}\mathbb{E}h^2(X_i)\right)^{1/2} \leq \varepsilon s_{m_n}^2\,.$$

It is then an easy exercise to see how the proof of the first part can be reproduced (and adapted to the case of a norm of the type $\sup\limits_{h\in D_n'}|h(\cdot)|$, thus depending on n) to yield that for some numerical constant M,

$$\sum_n \mathbb{P}\left\{\sup_{h\in D_n'}\left|\sum_{i=1}^{m_n} h(X_i)\right| > M\varepsilon s_{m_n}u_{m_n}\right\} < \infty\,.$$

The proof of (8.6) will therefore be complete if we show that

$$\sum_n \mathbb{P}\left\{\sup_{g\in D_n}\left|\sum_{i=1}^{m_n} g(X_i)\right| > (1+\varepsilon)s_{m_n}u_{m_n}\right\} < \infty\,.$$

But now, as in the real case, we have by Lemma 1.6 that for all n large enough (in order to efficiently use (8.2) and $\eta_i \to 0$, first neglect the first terms of the summation),

$$\mathbb{P}\left\{\sup_{g\in D_n}\left|\sum_{i=1}^{m_n} g(X_i)\right| > (1+\varepsilon)s_{m_n}u_{m_n}\right\} \le 2\ \mathrm{Card}D_n \exp(-(1+\varepsilon)LLs_{m_n}^2)\,,$$

hence the result since $\mathrm{Card}D_n \le \exp(2\alpha_n LLs_{m_n}^2)$, $\alpha_n \to 0$ and $s_{m_n} \sim \rho^n$ ($\rho > 1$).

In the last part of this proof, we show that

$$\limsup_{n\to\infty} \frac{\|S_n\|}{s_n u_n} \ge 1$$

almost surely, reporducing simply (more or less) the real case based on the exponential minoration inequality of Lemma 8.1. Recall that for $\rho > 1$ we let $m_n = \inf\{m; s_m > \rho^n\}$. By the zero-one law (cf. [Sto]), the lim sup we study is almost surely non-random; by the upper bound we just established, we have that (for example)

$$\mathbb{P}\{\|S_{m_n}\| \le 2s_{m_n}u_{m_n} \text{ for all } n \text{ large enough}\} = 1\,.$$

Suppose it can be proved that for all $\varepsilon > 0$ and all $\rho > 1$ large enough

$$\begin{aligned}\mathbb{P}\Bigg\{\Bigg\|\sum_{i\in I(n)} X_i\Bigg\| > (1-\varepsilon)^2 \Bigg(2\sup_{f\in D}\sum_{i\in I(n)}\mathbb{E}f^2(X_i)\Bigg)& \\ \cdot LL\Bigg(\sup_{f\in D}\sum_{i\in I(n)}\mathbb{E}f^2(X_i)\Bigg)^{1/2} \quad \text{i.o. in } n\Bigg\} = 1&\end{aligned} \tag{8.7}$$

where $I(n)$ denotes the set of integers between $m_{n-1}+1$ and m_n. Then, on a set of probability one, i.o. in n,

$$\begin{aligned}\|S_{m_n}\| &\ge \Bigg\|\sum_{i\in I(n)} X_i\Bigg\| - \|S_{m_{n-1}}\| \\ &\ge (1-\varepsilon)^2\Bigg(\sup_{f\in D}\sum_{i\in I(n)}\mathbb{E}f^2(X_i)\Bigg) LL\Bigg(\sup_{f\in D}\sum_{i\in I(n)}\mathbb{E}f^2(X_i)\Bigg)^{1/2} - 2s_{m_{n-1}}u_{m_{n-1}} \\ &\ge (1-\varepsilon)^2[2(s_{m_n}^2 - s_{m_{n-1}}^2)LL(s_{m_n}^2 - s_{m_{n-1}}^2)]^{1/2} - 2s_{m_{n-1}}u_{m_{n-1}}\end{aligned}$$

and, for n large, this lower bound behaves like

$$\left[(1-\varepsilon)^2 \left(1 - \frac{1}{\rho^2} \right)^{1/2} - \frac{2}{\rho} \right] s_{m_n} u_{m_n} .$$

For ρ large enough and $\varepsilon > 0$ arbitrarily small, this will therefore show that the $\limsup$ is ≥ 1 almost surely, hence the conclusion. Let us prove (8.7) then. For each n, let f_n in D be such that

$$\sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) \geq (1-\varepsilon) \sup_{f \in D} \sum_{i \in I(n)} \mathbb{E} f^2(X_i) .$$

Thus, the probability in (8.7) is bigger than

$$\mathbb{P} \left\{ \sum_{i \in I(n)} f_n(X_i) > (1-\varepsilon) \left(2 \sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) LL \left(\sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) \right) \right)^{1/2} \text{ i.o. in } n \right\} .$$

We can now apply Lemma 8.1 to the independent centered real random variables $f_n(X_i)$, $i \in I(n)$. Taking there $\gamma = \varepsilon/1-\varepsilon$, for all n large enough,

$$\begin{aligned} \mathbb{P} \left\{ \sum_{i \in I(n)} f_n(X_i) > (1-\varepsilon) \left(2 \sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) LL \left(\sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) \right) \right)^{1/2} \right\} \\ \geq \exp \left[-(1-\varepsilon) LL \left(\sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) \right) \right] \end{aligned}$$

where it has been used that

$$\begin{aligned} s_{m_n} &\leq \sup_{f \in D} \left(\sum_{i \in I(n)} \mathbb{E} f^2(X_i) \right)^{1/2} + s_{m_{n-1}} \\ &\leq \left(\frac{1}{1-\varepsilon} \sum_{i \in I(n)} \mathbb{E} f_n^2(X_i) \right)^{1/2} + s_{m_{n-1}} \end{aligned}$$

and that the ratio $s_{m_{n-1}}/s_{m_n}$ is small for large $\rho > 1$. This observation yields then also the conclusion with the Borel-Cantelli lemma (independent case). Theorem 8.2 has been established.

8.2. The law of the iterated logarithm of Hartman-Wintner-Strassen

Having described the fundamental LIL of Kolmogorov for sums of independent random variables and its extension to the vector value case, we now turn to the independent and identically distributed (iid) LIL and the results of Hartman-Wintner and Strassen. Let X be a random variable, and in the rest of the chapter $(X_i)_{i\in\mathbb{N}}$ will always denote a sequence of independent *copies* of X. The basic normalization sequence is here

$$a_n = (2nLLn)^{1/2}$$

which is seen to correspond to the sequence in (8.1) when $\mathbb{E}X^2 = 1$. The LIL of *Hartman and Wintner states that*, if X is a real random variable such that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = \sigma^2 < \infty$, the sequence (a_n) stabilizes the partial sums S_n in such a way that, with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{a_n} = -\liminf_{n\to\infty} \frac{S_n}{a_n} = \sigma \, . \tag{8.8}$$

Conversely, if the sequence (S_n/a_n) is almost surely bounded, then $\mathbb{E}X^2 < \infty$ (and $\mathbb{E}X = 0$, trivially by the SLLN). P. Hartman and W. Wintner deduced their result from Kolmogorov's LIL using a (clever) truncation argument. When $\mathbb{E}X^2 < \infty$, one can find a sequence (η_i) of positive numbers tending to 0 such that

$$\sum_i \frac{1}{LLi}\mathbb{P}\{|X| > \eta_i(i/LLi)^{1/2}\} < \infty \, . \tag{8.9}$$

Set then, for each i,

$$Y_i = X_i I_{\{|X_i|\le\eta_i(i/LLi)^{1/2}\}} - \mathbb{E}(X_i I_{\{|X_i|\le\eta_i(i/LLi)^{1/2}\}}) \, ,$$

and $Z_i = X_i - Y_i$. Since the Y_i's are bounded at a level corresponding to the application of Kolmogorov's LIL, (8.1) already gives that

$$\limsup_{n\to\infty} \frac{1}{a_n}\sum_{i=1}^n Y_i = \sigma$$

almost surely. The proof then consists in showing that the contribution of the Z_i's is negligible. To this aim, simply observe that by Cauchy-Schwarz's inequality,

$$\frac{1}{a_n}\left|\sum_{i=1}^n X_i I_{\{|X_i|>\eta_i(i/LLi)^{1/2}\}}\right| \le \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{1/2} \left(\frac{1}{2LLn}\sum_{i=1}^n I_{\{|X_i|>\eta_i(i/LLi)^{1/2}\}}\right)^{1/2} .$$

The first root on the right of this inequality defines an almost surely bounded (convergent!) sequence by the SLLN and $\mathbb{E}X^2 < \infty$; the second one converges to 0 by Kronecker's lemma (cf. e.g. [Sto]) and (8.9). Since the centerings in (Z_i) are taken into account similarly, it follows that

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{i=1}^{n} Z_i = 0 \text{ almost surely}$$

and therefore (8.8) holds.

The necessity of $\mathbb{E}X^2 < \infty$ can be obtained from a simple symmetry argument. By symmetrization, we may and do assume X to be symmetric. Let $c > 0$ and define

$$\overline{X} = XI_{\{|X|\le c\}} - XI_{\{|X|>c\}} .$$

Since X is symmetric, $\overline{X}$ has the same distribution as X . Assume now the sequence (S_n/a_n) is almost surely bounded. By the zero-one law, there is a finite number M such that, with probability one, $\limsup_{n\to\infty} |S_n|/a_n = M$. Now $2XI_{\{|X|\le c\}} = X + \overline{X}$ and since $\overline{X}$ has the same law as X ,

$$2\limsup_{n\to\infty} \frac{1}{a_n} \left| \sum_{i=1}^{n} X_i I_{\{|X_i|\le c\}} \right| \le 2M$$

almost surely. By the LIL of Hartman-Wintner (8.8), since $\mathbb{E}(X^2 I_{\{|X|\le c\}}) < \infty$ and X is symmetric, it follows that

$$\mathbb{E}(X^2 I_{\{|X|\le c\}}) \le M^2 .$$

Letting c tend to infinity implies indeed that $\mathbb{E}X^2 < \infty$.

While P. Hartman and A. Wintner used the rather deep result of A. N. Kolmogorov, the case of iid random variables should a priori appear as easier. Since then, simpler proofs of (8.8), which even produce more, have been obtained. As an illustration, we would like to intuitively describe in a direct way why the $\limsup$ in (8.8) should be *finite* when $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$. The idea is based on randomization by Rademacher random variables, a tool extensively used throughout this book. It explains rather easily the common steps and features of LIL's results like for example the fundamental use of exponential bounds of Gaussian type (Lemma 1.6 in Kolmogorov's LIL) and the study of (S_n) through blocks of exponential size. It suffices for this (modest) purpose to treat the case of a symmetric random variable so that we may assume as usual that

(X_i) has the same distribution as $(\varepsilon_i X_i)$ where (ε_i) is a Rademacher sequence independent of (X_i). By Lévy's inequality (2.6) for sums of independent symmetric random variables and the Borel-Cantelli lemma, it is enough to find a finite number M such that

$$\sum_n \mathbb{P}\left\{\left|\sum_{i=1}^{2^n} \varepsilon_i X_i\right| > M a_{2^n}\right\} < \infty .$$

Since $\mathbb{E}X^2 < \infty$, X^2 satisfies the law of large numbers and hence, by the Borel-Cantelli lemma again (independent case),

$$\sum_n \mathbb{P}\left\{\sum_{i=1}^{2^n} X_i^2 > 2^{n+1}\mathbb{E}X^2\right\} < \infty .$$

We now simply write that, for every n,

$$\begin{aligned}\mathbb{P}\left\{\left|\sum_{i=1}^{2^n} \varepsilon_i X_i\right| > M a_{2^n}\right\} &\leq \mathbb{P}\left\{\sum_{i=1}^{2^n} X_i^2 > 2^{n+1}\mathbb{E}X^2\right\} \\ &\quad + \mathbb{P}\left\{\left|\sum_{i=1}^{2^n} \varepsilon_i X_i\right| > M a_{2^n} ,\ \sum_{i=1}^{2^n} X_i^2 \leq 2^{n+1}\mathbb{E}X^2\right\} .\end{aligned}$$

The classical subgaussian estimate (4.1) applies conditionally on the sequence (X_i) to show that the second probability on the right of the preceding inequality is less than

$$\int_{\left\{\sum\limits_{i=1}^{n} X_i^2 \leq 2^{n+1}\mathbb{E}X^2\right\}} 2\exp\left(-M^2 2^n LL2^n \Big/ \sum_{i=1}^{2^n} X_i^2\right) d\mathbb{P} \leq 2\exp(-M^2 LL2^n/2\mathbb{E}X^2) .$$

If we then choose $M^2 > 2\mathbb{E}X^2$, the claim is established.

This simple approach, which reduces in a sense the iid LIL to the SLLN through Gaussian exponential estimates, can be pushed further in order to show the necessity of $\mathbb{E}X^2 < \infty$ when the sequence (S_n/a_n) is almost surely bounded. One can use the converse subgaussian inequality (4.2). Alternatively, and without going into the details, it is not difficult to see that if (g_i) is an orthogaussian sequence independent of (X_i), the two non-random $\limsup$'s

$$\limsup_{n\to\infty} \frac{1}{a_n}\left|\sum_{i=1}^{n} X_i\right| \qquad \text{and} \qquad \limsup_{n\to\infty} \frac{1}{a_n}\left|\sum_{i=1}^{n} g_i X_i\right|$$

are equivalent. Independence and the stability properties of (g_i) expressed for example by

$$\limsup_{n\to\infty} \frac{|g_n|}{(2\log n)^{1/2}} = 1 \quad \text{almost surely}$$

can then easily be used to check the necessity of $\mathbb{E}X^2 < \infty$ (cf. [L-T2]).

These rather easy ideas describe some basic facts in the study of the LIL like exponential estimates of Gaussian type, blocking arguments, the connection of the LIL with the law of large numbers (for squares) and of course the central limit theorem through the introduction of Gaussian randomization. In fact, the LIL can be thought of as some almost sure form of the central limit theorem. The framework of these elementary observations will lead later to the infinite dimensional LIL.

The preceding sketchy proof of Hartman-Wintner's LIL of course only provides qualitative results and not the exact value of the $\limsup$ in (8.8). Simple proofs of (8.8) have been given in the literature; they include the more precise and so-called *Strassen's form of the* LIL which states that, if and only if $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$,

$$\lim_{n\to\infty} d\left(\frac{S_n}{a_n}\,,\ [-\sigma,\sigma]\right) = 0 \tag{8.10}$$

and

$$C\left(\frac{S_n}{a_n}\right) = [-\sigma,\sigma] \tag{8.11}$$

almost surely, where $d(x,A) = \inf\{|x-y|\,;\ y \in A\}$ is the distance of the point x to the set A and where $C(x_n)$ denotes the set of limit points of the sequence (x_n), i.e. $C(x_n) = \{x \in \mathbb{R}\,;\ \liminf\limits_{n\to\infty} |x_n - x| = 0\}$. (8.10) and $C(S_n/a_n) \subset [-\sigma,\sigma]$ follow rather easily from the LIL of Hartman-Wintner. The full property (8.11) is more delicate and various arguments can be used in order to establish it; we will obtain (8.10) and (8.11) in the more general context of Banach space valued random variables below. Strassen's approach used Brownian motion and the Skorohod embedding of a sequence of iid random variables in Brownian paths.

Our objective in the sequel of this chapter will now be to investigate the iid LIL for vector valued random variables. We deal until the end of the chapter with Radon variables and even, for more convenience, with separable Banach spaces, although various conclusions still hold in our usual more general setting; these will be indicated in remarks. For Radon random variables, the picture is probably the most complete and satisfactory and we adopt this framework in order not to obscure the main scheme.

Let therefore B denote a separable Banach space. We start by describing what can be understood by a LIL for independent and identically distributed Banach space valued random variables. Let X be a Borel random variable with values in B, (X_i) a sequence of independent copies of X. As usual, $S_n = X_1 + \cdots + X_n$,

$n \geq 1$. According to the LIL of Hartman-Wintner, we can say that X satisfies the LIL with respect to the classical normalizing sequence $a_n = (2nLLn)^{1/2}$ if the non-random limit (zero-one law)

$$\Lambda(X) = \limsup_{n \to \infty} \frac{\|S_n\|}{a_n} \tag{8.12}$$

is finite. (If X is degenerate, $\Lambda(X) > 0$ by the scalar case.) We will actually define this property as the *bounded* LIL. Indeed, we might as well say that X satisfies the LIL whenever the sequence (S_n/a_n) is almost surely relatively compact in B since this means the same in finite dimensional spaces. We will say then that X satisfies the *compact* LIL and it will turn out that in infinite dimension, bounded and compact LIL are *not* equivalent. Actually, Strassen's formulation (8.10) and (8.11) even suggests a third definition: X satisfies the LIL if there is a compact convex symmetric set K in B such that, almost surely,

$$\lim_{n \to \infty} d\left(\frac{S_n}{a_n}\,,\ K\right) = 0 \tag{8.13}$$

and

$$C\left(\frac{S_n}{a_n}\right) = K \tag{8.14}$$

where $d(x, K) = \inf\{\|x - y\|\,;\ y \in K\}$ and $C(S_n/a_n)$ denotes the set of cluster points of the sequence (S_n/a_n) . It is a nontrivial result that the compact LIL and this definition actually coincide. Before describing precisely this result, we would like to study what the limit set K should be. It will turn out to be the unit ball of the so-called reproducing kernel Hilbert space associated to the covariance structure of X . We sketch its construction and properties, some of which go back to the Gaussian setting as described in Chapter 3.

X is therefore a fixed Borel random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in the separable Banach space B . Recall the separability allows to assume $\mathcal{A}$ countably generated and thus $L_2(\Omega, \mathcal{A}, \mathbb{P})$ separable. Suppose that for all f in B' , $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$. Let us observe, as a remark, that these hypotheses are natural in the context of the LIL since if for example X satisfies the bounded LIL, for each f in B' , $(f(S_n/a_n))$ is almost surely bounded and therefore $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$ by the scalar case. Under these hypotheses,

$$\sigma(X) = \sup_{\|f\| \leq 1} (\mathbb{E}f^2(X))^{1/2} < \infty\,. \tag{8.15}$$

Indeed, if we consider the operator $A = A_X$ defined as $A : B' \to L_2 = L_2(\Omega, \mathcal{A}, \mathbb{P})$, $Af = f(X)$, then $\|A\| = \sigma(X)$ and A is bounded by an easy closed graph argument. Let $A^* = A_X^*$ denote the adjoint of

A . Note first that since X defines a Radon random variable, A^* actually maps L_2 into $B \subset B''$ (cf. Section 2.1). Indeed, there exists a sequence (K_n) of compact sets in B such that $\mathbb{P}[X \notin K_n\} \to 0$. If ξ is in L_2 , $A^*(\xi I_{\{X \in K_n\}})$ belongs to B since it can be identified with the expectation (in the strong sense) $\mathbb{E}(\xi X I_{[X \in K_n\}})$. But $\mathbb{E}(\xi X I_{\{X \in K_n\}})$ converges to $\mathbb{E}(\xi X)$ (weak integral) in B'' since

$$\sup_{\|f\| \leq 1} f(\mathbb{E}(\xi X I_{\{X \notin K_n\}})) \leq \sigma(X)(\mathbb{E}(\xi^2 I_{\{X \notin K_n\}}))^{1/2} \to 0 .$$

Hence $A^*\xi = \mathbb{E}(\xi X)$ belongs to B .

On the image $A^*(L_2) \subset B$ of L_2 by A^* , consider the scalar product $\langle \cdot, \cdot \rangle_X$ transferred from L_2 : if $\xi, \zeta \in L_2$,

$$\langle A^*\xi, A^*\zeta \rangle_X = \langle \xi, \zeta \rangle_{L_2} = \int \xi\zeta \, d\mathbb{P} .$$

Denote by $H = H_X$ the (separable) Hilbert space $A^*(L_2)$ equipped with $\langle \cdot, \cdot \rangle_X$. H is called the *reproducing kernel Hilbert space* associated to the covariance structure of X . The word "reproducing" stems from the fact that H reproduces the covariance of X in the sense that for f, g in B' , if $x = A^*(g(X)) \in H$, $f(x) = \mathbb{E} f(X) g(X)$. In particular, if X and Y are random variables with the same covariance structure, i.e. $\mathbb{E} f(X) g(X) = \mathbb{E} f(Y) g(Y)$ for all f, g in B' , this reproducing property implies that $H_X = H_Y$. Note that since $A(B')^\perp = \mathrm{Ker} A^*$, we also have that H is the completion, with respect to the scalar product $\langle \cdot, \cdot \rangle_X$, of the image of B' by the composition $S = A^*A : B \to B'$. Observe further that for any x in H ,

$$\|x\| \leq \sigma(X) \langle x, x \rangle_X . \tag{8.16}$$

Denote by $K = K_X$ the closed unit ball of H , i.e. $K = \{x \in B \, ; \; x = \mathbb{E}(\xi X) , \; \|\xi\|_2 \leq 1\}$, which thus defines a bounded convex symmetric set in B . By the Hahn-Banach theorem, we also have that $K = \{x \in B \, ; \; f(x) \leq \|f(X)\|_2$ for all f in $B'\}$, and by separability this can be achieved by taking only a (well-chosen) sequence (f_k) in B' . As the image of the unit ball of L_2 by A^* , K is weakly compact and therefore also closed for the topology of the norm on B . K is separable in B by (8.16). Further, it is easily verified that for any f in B' ,

$$\|f(X)\|_2 = \sup_{x \in K} f(x) , \qquad \sigma(X) = \sup_{x \in K} \|x\| .$$

While K is weakly compact, it is not always compact. The next easy lemma describes for further references equivalent instances for K to be compact.

Lemma 8.4. The following are equivalent:

(i) K is compact;

(ii) A (resp. A^*) is compact;

(iii) $S = A^*A$ is compact;

(iv) the covariance function $T(f,g) = \mathbb{E}f(X)g(X)$ is weakly sequentially continuous;

(v) the family of real random variables $\{f^2(X)\,;\ f \in B',\ \|f\| \le 1\}$ is uniformly integrable.

Proof. (i) and (ii) are clearly equivalent and imply (iii). To see that (iv) holds under (iii), it suffices to show that $\|f_n(X)\|_2 \to 0$ when $f_n \to 0$ weakly in B'. By the uniform boundedness principle, we may assume that $\|f_n\| \le 1$ for all n. The compactness of S ensures that we can extract from the sequence (x_n) defined by $x_n = \mathbb{E}(Xf_n(X))$ a subsequence, still denoted with n, convergent to some x. But then

$$\mathbb{E}f_n^2(X) = f_n(x_n) \le \|x_n - x\| + |f_n(x)| \to 0\,.$$

Assume (v) is not satisfied. Then, there exist $\varepsilon > 0$ and a sequence (c_n) of positive numbers increasing to infinity such that for every n

$$\sup_{\|f\|\le 1} \int_{\{\|X\|>c_n\}} f^2(X)\,d\mathbb{P} \ge \sup_{\|f\|\le 1} \int_{\{|f(X)|>c_n\}} f^2(X)d\mathbb{P} > \varepsilon\,.$$

Hence, for every n, one can find f_n, $\|f_n\| \le 1$, such that

$$\int_{\{\|X\|>c_n\}} f_n^2(X)d\mathbb{P} > \varepsilon\,.$$

Extract then from the sequence (f_n) in the unit ball of B' a weakly convergent subsequence, still denoted (f_n), convergent to some f. By (iv), $f_n(X) \to f(X)$ in L_2 and this clearly reaches a contradiction since

$$\lim_{n\to\infty} \int_{\{\|X\|>c_n\}} f^2(X)d\mathbb{P} = 0\,.$$

Finally, (v) easily implies (ii); indeed, if (f_n) is a sequence in the unit ball of B', for some subsequence and some f, $f_n \to f$ weakly, so $f_n(X) \to f(X)$ almost surely and hence in L_2 by uniform integrability; A is therefore compact. The proof of Lemma 8.4 is complete.

Note that when $\mathbb{E}\|X\|^2 < \infty$ (in case of Gaussian variables for example), K is compact.

As simple examples, if $B = \mathbb{R}^N$ and the covariance matrix of X is the identity matrix, K is simply the Euclidean unit ball of $\mathbb{R}^N$. When X follows the Wiener distribution on $C[0,1]$, H can be identified with the so-called Cameron-Martin Hilbert space of the absolutely continuous elements x in $C[0,1]$ such that $x(0) = 0$ and $\int_0^1 x'(t)^2 dt < \infty$, and K is known in this case as Strassen's limit set.

Having described the natural limit set in (8.13) and (8.14), we now present the theorem, due to J. Kuelbs, connecting the definition of the compact LIL with (8.13) and (8.14).

Theorem 8.5. Let X be a Borel random variable with values in a separable Banach space B. If the sequence (S_n/a_n) is almost surely relatively compact in B, then, with probability one,

$$\lim_{n\to\infty} d\left(\frac{S_n}{a_n}\,,\ K\right) = 0 \qquad \text{(and)} \qquad C\left(\frac{S_n}{a_n}\right) = K$$

where $K = K_X$ is the unit ball of the reproducing kernel Hilbert space associated to the covariance structure of X and K is compact. Conversely, if the preceding holds for some compact set K, then X satisfies the compact LIL and $K = K_X$.

According to this theorem, when we speak of compact LIL we mean one of the equivalent properties of this statement.

Proof. As we have seen, when X satisfies the bounded LIL, which is always the case under one of the properties of Theorem 8.5, $K = K_X$ is well defined. Let us first show that, with probability one, $C(S_n/a_n) \subset K$. As was observed in the definition of K, there is a sequence (f_k) in B' such that a point x belongs to K as soon as

$$f_k(x) \le \|f_k(X)\|_2 \tag{8.17}$$

for all k. Denote by Ω_0 the set of full probability (by the scalar LIL) of the ω's such that for every k

$$\limsup_{n\to\infty} \frac{|f_k(S_n(\omega))|}{a_n} \le \|f_k(X)\|_2\,.$$

So if $x \in C(S_n(\omega)/a_n)$ and $\omega \in \Omega_0$, x clearly satisfies (8.17) and therefore belongs to K.

This first property easily implies that

$$\lim_{n\to\infty} d\left(\frac{S_n}{a_n}\,,\ K\right) = 0 \tag{8.18}$$

with probability one. Indeed, if this is not the case, one can find, by relative compactness of (S_n/a_n), a subsequence of (S_n/a_n) converging to some point exterior to K and this is impossible as we have just seen.

We are thus left with the proof that $C(S_n/a_n) = K$. To this aim, it suffices, by density, to show that any x in K belongs almost surely to $C(S_n/a_n)$.

Let us assume first that B is finite dimensional. Since the covariance matrix of X is symmetric positive definite, it may be diagonalized in some orthonormal basis. We are therefore reduced to the case where B is Hilbertian and K is its unit ball. Let then x be in B with $|x| = 1$ and let $\varepsilon > 0$. By (8.18), for n large enough, $|S_n/a_n|^2 \le 1+\varepsilon$. By Hartman-Wintner's LIL (8.8), along a subsequence,

$$\langle \frac{S_n}{a_n}\,,\ x\rangle \ge \|\langle x, X\rangle\|_2 - \varepsilon = 1-\varepsilon\,.$$

Hence, along this subsequence and for n large,

$$\left|\frac{S_n}{a_n} - x\right|^2 = \left|\frac{S_n}{a_n}\right|^2 + |x|^2 - 2\langle\frac{S_n}{a_n}\,,\ x\rangle$$

$$\le 1+\varepsilon+1-2+2\varepsilon = 3\varepsilon$$

and therefore $x \in C(S_n/a_n)$ almost surely. To reach the interior points of K, we climb in dimension and consider the random variable in $B\times\mathbb{R}$ given by $Y = (X,\varepsilon)$ where ε is a Rademacher variable independent of X. Let x in K with $|x| = \theta < 1$; then $y = (x,(1-\theta^2)^{1/2})$ belongs to the unit sphere of $B\times\mathbb{R}$ and thus, by the preceding step, to the cluster set associated to Y. By projection, $x \in C(S_n/a_n)$.

We now complete the proof of Theorem 8.5 and use this finite dimensional result to get the full conclusion concerning the cluster set $C(S_n/a_n)$. When X satisfies the LIL, it also satisfies the strong law of large numbers and therefore $\mathbb{E}\|X\| < \infty$. There exists an increasing sequence $(\mathcal{A}_N)$ of finite σ-algebras of X such that $X^N = \mathbb{E}^{\mathcal{A}_N}X$ converges almost surely and in $L_1(B)$ to X. Note that if (S_n/a_n) is almost surely relatively compact, property (iv) of Lemma 8.4 is fulfilled; indeed, if $f_k \to 0$ weakly, by compactness,

$$\lim_{k\to\infty}\sup_n \left|f_k\left(\frac{S_n}{a_n}\right)\right| = 0$$

with probability one. By (8.8), it follows that $\|f_k(X)\|_2 \to 0$ which gives (iv). By Lemma 8.4, K is therefore compact, or equivalently, $\{f^2(X)\,;\ f \in B'\,,\ \|f\| \le 1\}$ is uniformly integrable. There exists therefore (Lemma of La Vallé-Poussin, cf. e.g. [Me]) a positive convex function ψ on $\mathbb{R}_+$ with $\lim_{t\to\infty}\psi(t)/t = \infty$ such that

$\sup_{\|f\|\le 1} \mathbb{E}\psi(f^2(X)) < \infty$. By Jensen's inequality, it follows that the family $\{f^2(X - X^N)\,;\ \|f\| \le 1\,,\ N \in \mathbb{N}\}$ is also uniformly integrable. This can be used to obtain that $\sigma(X - X^N) \to 0$ when $N \to \infty$ (where $\sigma(\cdot)$ is defined in (8.15)). Let now x in $K : x = \mathbb{E}(\xi X)$, $\|\xi\|_2 \le 1$. If $x^N = \mathbb{E}(\xi X^N)$, $\|x - x^N\| \le \sigma(X - X^N) \to 0$. Further, by (8.18), $\Lambda(X) \le \sup_{x \in K} \|x\| = \sigma(X)$, and since the X^N 's are finite dimensional they also satisfy the compact LIL and therefore $\Lambda(X - X^N) \le \sigma(X - X^N)$ for every N . Now, we simply write by the triangle inequality, that for every N ,

$$\begin{aligned}\liminf_{n\to\infty} \left\| \frac{S_n}{a_n} - x \right\| &\le \liminf_{n\to\infty} \left\| \frac{S_n^N}{a_n} - x^N \right\| + \Lambda(X - X^N) + \|x - x^N\| \\ &\le \liminf_{n\to\infty} \left\| \frac{S_n^N}{a_n} - x^N \right\| + 2\sigma(X - X^N)\end{aligned}$$

where S_n^N are the partial sums associated to X^N . By the previous step in finite dimension, for each N ,

$$\liminf_{n\to\infty} \left\| \frac{S_n^N}{a_n} - x^N \right\| = 0$$

almost surely. Letting N tend to infinity then shows that $x \in C(S_n/a_n)$ which completes the proof of Theorem 8.5.

The preceding discussion and Theorem 8.5 present the definitions of the LIL of Hartman-Wintner-Strassen for Banach space valued random variables. We now would like to turn to the crucial question of knowning *when* a random variable X with values in a Banach space B satisfies the bounded or compact LIL in terms of minimal conditions, depending if possible only on the distribution of X .

If B is the line or, more generally, a finite dimensional space, X satisfies the bounded or compact LIL if and only if $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. However, already in Hilbert space, while these conditions are sufficient, the integrability $\mathbb{E}\|X\|^2 < \infty$ is no longer necessary. It happens conversely that in some spaces, bounded mean zero random variables do not satisfy the LIL. Further, examples disprove the equivalence between bounded and compact LIL in the infinite dimensional setting. All these examples will actually become clear on the final characterization. They however pointed out historically the difficulty in finding what this characterization should be. The issue is based in particular on a careful examination of the necessary conditions for a Banach space valued random variable to satisfy the LIL which we now would like to describe.

Assume first that X satisfies the bounded LIL in B . Then clearly, for each f in B' , $f(X)$ satisfies the scalar LIL and thus $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$. These *weak* integrability conditions are complemented by a necessary integrability property on the *norm*; indeed, it is necessary that the sequence (X_n/a_n) is bounded almost surely and thus, by independence, the Borel-Cantelli lemma and identical distribution, for some finite M,

$$\sum_n \mathbb{P}\{\|X\| > Ma_n\} < \infty .$$

As is easily seen, this turns out to be equivalent to the integrability condition

$$\mathbb{E}(\|X\|^2/LL\|X\|) < \infty .$$

These are the best moment conditions which can be deduced from the bounded LIL. As we mentioned it however, there are almost surely bounded (mean zero) random variables which do not satisfy the LIL. This unfortunate fact forces, in order to expect some characterization, to complete the preceding integrability conditions, which depend only on the distribution of X, by some condition involving the laws of the partial sums S_n instead only of X. As we will see, this can be avoided in some spaces, but it is necessary to proceed along these lines in general as actually could have been expected from the previous chapters. The third necessary condition is then simply (and trivially) that the sequence (S_n/a_n) should be bounded in probability.

In finite dimension, the weak L_2 integrability of course implies the strong L_2 integrability, and therefore $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$, as well as the stochastic boundedness of (S_n/a_n) (as is easily seen, for example, from the central limit theorem). It is remarkable that this easily obtained set of necessary conditions is also sufficient for X to satisfy the bounded LIL. Before stating this characterization, let us complete the discussion on necessary conditions by the case of the compact LIL. We keep that $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$. By Theorem 8.5, $K = K_X$ should be compact, or equivalently $\{f^2(X)\,;\ f \in B',\ \|f\| \le 1\}$ uniformly integrable (this can also be proved directly as is clear from the last part of the proof of Theorem 8.5). Finally, under the compact LIL, it is necessary that the sequence (S_n/a_n) is not only bounded in probability, but convergent to 0; indeed, the sequence of the laws of S_n/a_n is necessarily tight with 0 as only possible limit point since $\mathbb{E}f(X) = 0$, $\mathbb{E}f^2(X) < \infty$ for all f in B'. Let us note that the stochastic boundedness (and a fortiori convergence to 0) of the sequence (S_n/a_n) also contains the fact that X is centered. (To see this, use the analog of Lemma 10.1 for the normalization a_n.)

We can now present the characterization of random variables satisfying the bounded or compact LIL. As typical in Probability in Banach spaces, it reduces in a sense, under necessary and natural moment conditions, the *almost sure* behavior of the sequence (S_n/a_n) to its behavior in *probability*.

Theorem 8.6. Let X be a Borel random variable with values in a separable Banach space B. In order that X satisfy the bounded LIL, it is necessary and sufficient that the following conditions are fulfilled:

(i) $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$;

(ii) for each f in B', $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$;

(iii) the sequence (S_n/a_n) is bounded in probability.

In order that X satisfy the compact LIL, it is necessary and sufficient that (i) holds and that (ii) and (iii) are replaced by

(ii') $\mathbb{E}X = 0$ and $\{f^2(X)\,;\ f \in B',\ \|f\| \leq 1\}$ is uniformly integrable;

(iii') $S_n/a_n \to 0$ in probability.

Proof. Necessity has been discussed above. The proof of the sufficiency for the bounded LIL is the main point of the whole theorem. We will show indeed that for some numerical constant M, for all symmetric random variables X satisfying (i), we have

$$\Lambda(X) = \limsup_{n\to\infty} \frac{\|S_n\|}{a_n} \leq M(\sigma(X) + L(X)) \tag{8.19}$$

where $\sigma(X) = \sup_{\|f\|\leq 1} (\mathbb{E}f^2(X))^{1/2}$ and

$$L(X) = \limsup_{n\to\infty} \frac{1}{a_n} \mathbb{E} \left\| \sum_{i=1}^{n} X_i I_{\{\|X_i\| \leq a_n\}} \right\| .$$

Since $\sigma(X) < \infty$ under (ii) ((8.15)) and since (iii) implies that $L(X) < \infty$ by Lemma 7.2, this inequality (8.19) contains the bounded LIL, at least for symmetric random variables but actually also in general by symmetrization and Lemma 7.1. From (8.19) also follows the compact version. Indeed, by Lemma 7.2, $L(X)$ can be chosen to be 0 by (iii') and, by symmetrization and Lemma 7.1, the inequality also holds in this form (with thus $L(X) = 0$) for non-necessarily symmetric random variables satisfying (i) and (iii'). This estimate applied to quotient norms by finite dimensional subspaces then yields the conclusion. More precisely, if F denotes a finite dimensional subspace of B and $T = T_F$ the quotient map $B \to B/F$, we get from (8.19) that $\Lambda(T(X)) \leq M\sigma(T(X))$. Under (ii'), $\sigma(T(X))$ can be made arbitrarily small with large enough F (show for example, as in the proof of Theorem 8.5, that $\sigma(X - X^N) \to 0$ where $X^N = \mathbb{E}^{\mathcal{A}_N} X$). The sequence (S_n/a_n) then appears as being arbitrarily close to some bounded set in the finite dimensional subspace F, and is therefore relatively compact (Lemma 2.2). Hence X satisfies the compact LIL under (i), (ii') and (iii').

We are thus left with the proof of (8.19). To this aim, we use the isoperimetric approach as developed for example in the preceding chapter on the SLLN. We intend more precisely to apply Theorem 7.5. In order to verify the first set of conditions there, we employ Lemmas 7.6 and 7.8. The integrability condition $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$ is equivalent to say that for every $\varepsilon > 0$

$$\sum_n 2^n \mathbb{P}\{\|X\| > \varepsilon a_{2^n}\} < \infty .$$

Let $\varepsilon > 0$ be fixed. Setting together the conclusions of Lemmas 7.6 and 7.8 we see that (taking $q = 2K_0$) there exists a sequence of integers (k_n) such that $\sum_n 2^{-k_n} < \infty$ for which

$$\sum_n \mathbb{P}\left\{\sum_{r=1}^{k_n} \|X_{2^{n-1}}^{(r)}\| > 5\varepsilon a_{2^n}\right\} < \infty$$

where $X_{2^{n-1}}^{(r)}$ is the r-th largest element of the sample $(\|X_i\|)_{i \le 2^{n-1}}$. L in Theorem 7.5 is less than $L(X)$ (contraction principle) and, for each n, $\sigma_n^2 \le 2^n \sigma(X)^2$ so that (7.4) is satisfied with for example $\delta = \sigma(X)$. The conclusion of Theorem 7.5, with $q = 2K_0$, is then that

$$\sum_n \mathbb{P}\{\|S_{2^{n-1}}\| > 10^2[\varepsilon + 2K_0(\sigma(X) + L(X) + (5\varepsilon L(X))^{1/2})]a_{2^n}\} < \infty .$$

Since $\varepsilon > 0$ is arbitrary and $a_{2^n} \sim \sqrt{2} a_{2^{n-1}}$, inequality (8.19) follows from the Borel-Cantelli lemma and the maximal inequality of Lévy (2.6). Theorem 8.6 is thus established.

While Theorem 8.6 provides a rather complete characterization of random variables satisfying the LIL, hypotheses on the distribution of the partial sums rather than only on the variable have to be used. It is worthwhile to point out at this stage that in a special class of Banach spaces, it is possible to get rid of these assumptions and state the characterization in terms only of the moment conditions (i) and (ii) (or (ii')). Anticipating on the next chapter as we did for the law of large numbers, say a Banach space B is of type 2 if there is a constant C such that for any finite sequence (Y_i) of independent mean zero random variables with values in B, we have

$$\mathbb{E}\left\|\sum_i Y_i\right\|^2 \le C \sum_i \mathbb{E}\|Y_i\|^2 .$$

Hilbert spaces are clearly of type 2; further examples are discussed in Chapter 9. In type 2 spaces the integrability conditi on $\mathbb{E}(\|X\|^2/LL\|X\|)<\infty$ implies, if $\mathbb{E}X=0$, that $S_n/a_n\to 0$ in probability and hence the nicer form of Theorem 8.6 in this case. Let us prove this implication.

Lemma 8.7. Let X be a mean zero random variable such that $\mathbb{E}(\|X\|^2/LL\|X\|)<\infty$ with values in a type 2 Banach space B. Then $S_n/a_n\to 0$ is probability.

Proof. We show that if X is symmetric and $\mathbb{E}(\|X\|^2/LL\|X\|)<\infty$, then $\mathbb{E}\|S_n\|/a_n\to 0$ which, by Lemma 6.3, implies the lemma. For each n,

$$\mathbb{E}\|S_n\|\le \mathbb{E}\left\|\sum_{i=1}^n X_i I_{\{\|X_i\|\le a_n\}}\right\| + n\mathbb{E}(\|X\|I_{\{\|X\|>a_n\}})\,.$$

A simple integration by parts shows that, under $\mathbb{E}(\|X\|^2/LL\|X\|)<\infty$,

$$\lim_{n\to\infty}\frac{n}{a_n}\mathbb{E}(\|X\|I_{\{\|X\|>a_n\}})=0\,.$$

By the type 2 inequality (and symmetry),

$$\frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^n X_i I_{\{\|X_i\|\le a_n\}}\right\| \le \left(\frac{C}{2LLn}\mathbb{E}(\|X\|^2 I_{\{\|X\|\le a_n\}})\right)^{1/2}\,.$$

For each $t>0$, the right hand side squared of this inequality is seen to be smaller than

$$\frac{Ct^2}{2LLn}+\frac{C}{2LLn}\mathbb{E}(\|X\|^2 I_{\{t<\|X\|\le a_n\}})\le \frac{Ct^2}{2LLn}+C\mathbb{E}\left(\frac{\|X\|^2}{LL\|X\|}I_{\{\|X\|>t\}}\right)\,.$$

Letting n and then t go to infinity concludes the proof.

As announced, Lemma 8.7 implies the next corollary.

Corollary 8.8. Let X be a Borel random variable with values in a separable type 2 Banach space B. Then X satisfies the bounded (resp. compact) LIL if and only if $\mathbb{E}(\|X\|^2/LL\|X\|)<\infty$ and $\mathbb{E}f(X)=0$, $\mathbb{E}f^2(X)<\infty$ for all f in B' (resp. $\{f^2(X)\,;\ f\in B',\ \|f\|\le 1\}$ is uniformly integrable).

Remark 8.9. Theorem 8.6 has been presented in the context of Radon random variables. Its proof however clearly indicates some possible extensions to more general settings of random variables as the one we usually adopt in this text. This is in particular completely obvious for the bounded version which, as in the case of Kolmogorov's LIL, does not require any approximation argument. With some precautions, extensions of the compact LIL to this setting can also be imagined. We leave this to the interested reader.

8.3. On the identification of the limits

In this last paragraph, we would like to describe various results and examples on the limits of the sequence (S_n/a_n) in the bounded form of the LIL for Banach space valued random variables. We learned from Theorem 8.5 that when the sequence (S_n/a_n) is almost surely relatively compact in B, then, with probability one,

$$\lim_{n\to\infty} d\left(\frac{S_n}{a_n}\,,\ K\right) = 0 \tag{8.20}$$

and

$$C\left(\frac{S_n}{a_n}\right) = K \tag{8.21}$$

where $K = K_X$ is the unit ball of the reproducing kernel Hilbert space associated to the covariance structure of X and K is compact in this case. In particular also,

$$\Lambda(X) = \limsup_{n\to\infty} \frac{\|S_n\|}{a_n} = \sigma(X) = \sup_{\|f\|\le 1} (\mathbb{E}f^2(X))^{1/2} \tag{8.22}$$

(recall $\sigma(X) = \sup_{x\in K} \|x\|$). One might now be interested in knowning if these properties still hold, or what they become, when for example X only satisfies the bounded LIL and not the compact one, or even, for (8.21), if X is just such that $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$ for all f in B' (in order for K to be well defined). To put the question in clearer perspective, let us mention the example (cf. [Kue6]) of a bounded random variable satisfying the bounded LIL but for which the cluser set $C(S_n/a_n)$ is empty. Further examples of pathological situations have been observed in the literature. We would like here to briefly describe some positive results as well as some problems left open.

We start with the remarkable results of K. Alexander [Ale3] on the cluster set. Let X be a Borel random variable with values in a separable Banach space B such that $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$ for all f in B'. As we have seen in the first part of the proof of Theorem 8.5, almost surely $C(S_n/a_n) \subset K$ where $K = K_X$. From an easy zero-one law, it can be shown that the cluster set $C(S_n/a_n)$ is almost surely non-random. It can be K, and can also be empty as alluded to above. As a main result, it is shown in [Ale3] that $C(S_n/a_n)$ can actually only be empty or αK for some α in $[0,1]$, and examples are given in [Ale4] showing that every value of α can indeed occur. Moreover, a series condition involving the laws of the partial sums S_n determines the value of α. More precisely we have the following theorem which we state without proof refering to [Ale3].

Theorem 8.10. Let X be a Borel random variable with values in a separable Banach space B such that $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$ for every f in B'. Let

$$\alpha^2 = \sup\{\beta \geq 0\,;\ \sum_n n^{-\beta}\mathbb{P}\{\|S_m/a_m\| < \varepsilon \ \text{ for some } 2^{n-1} < m \leq 2^n\} = \infty \ \text{ for all } \varepsilon > 0\}$$

whenever this set is not empty. Then $C(S_n/a_n) = \alpha K$, or $\emptyset$ when this set is empty. In particular, $\alpha = 1$ when $S_n/a_n \to 0$ in probability.

These results settle the nature of the cluster set $C(S_n/a_n)$. Similar questions can of course be asked concerning (8.20) and (8.22). Although the results are less complete here, one positive fact is available. We have of couse to assume here that X satisfies the bounded LIL, that is, by Theorem 8.6, that $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$, $\sigma(X) < \infty$ and (S_n/a_n) is bounded in probability. It turns out that when this last condition is strengthened into $S_n/a_n \to 0$ in probability, one can prove (8.20) and (8.22) with $K = K_X$, compact or not. This is the object of the following theorem, whose proof amplifies some of the techniques of the proof of Theorem 8.2 and which provides a rather complete description of the limits in this case. As we will see, the situation may be quite different in general.

Theorem 8.11. Let X be a Borel random variable with values in a separable Banach space B. Assume that $\mathbb{E}X = 0$, $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$, $\sigma(X) = \sup_{\|f\|\leq 1}(\mathbb{E}f^2(X))^{1/2} < \infty$ and that $S_n/a_n \to 0$ in probability. Then we have

$$\Lambda(X) = \limsup_{n\to\infty}\frac{\|S_n\|}{a_n} = \sigma(X) \ \text{ almost surely}. \tag{8.23}$$

Moreover,

$$\lim_{n\to\infty} d\left(\frac{S_n}{a_n},\ K\right) = 0 \qquad \text{and} \qquad C\left(\frac{S_n}{a_n}\right) = K \tag{8.24}$$

with probability one where $K = K_X$ is the unit ball of the reproducing kernel Hilbert space associated to the covariance structure of X.

Proof. It is enough to prove (8.23); indeed, replacing the norm of B by the gauge of $K + \varepsilon B_1$ where B_1 is the unit ball of B, it is easily seen that $d(S_n/a_n, K) \to 0$. Identification of the cluster set follows from Theorem 8.10 since $S_n/a_n \to 0$ in probability. To establish (8.23), by homogeneity and the real LIL, we need only show that $\Lambda(X) \leq 1$ when $\sigma(X) = 1$. As in the proof of Theorem 8.2 (see (8.6)), by the

Borel-Cantelli lemma and Ottaviani's inequality (Lemma 6.2), it suffices to prove that for all $\varepsilon > 0$ and $\rho > 1$

$$\sum_n \mathbb{P}\left\{ \left\| \sum_{i=1}^{m_n} X_i \right\| > (1+\varepsilon)a_{m_n} \right\} < \infty$$

where $m_n = [\rho^n]$, $n \geq 1$ (integer part).

Let then $0 < \varepsilon \leq 1$ and $\rho > 1$ be fixed. For every integer n and f, h in the unit ball U of B', set

$$d_2^n(f,h) = (\mathbb{E}(f-h)^2(X) I_{\{\|X\| \leq a_{m_n}\}})^{1/2} .$$

Let further $N(U, d_2^n; \varepsilon)$ be the minimal number of elements h in U such that for every f in U there exists such a h with $d_2^n(f,h) < \varepsilon$. As in the proof of Theorem 8.2, we first need an estimate of the size of these entropy numbers when n varies. However, with respect to Lemma 8.3, it does not seem possible to use the Sudakov minoration for Rademacher processes since the truncations do not appear to fit the right levels. Instead rather, we will use the Gaussian minoration through an improved randomization property which is made possible by the fact that we are working with identically distributed random variables. Let respectively (ε_i) and (g_i) be Rademacher and standard Gaussian sequences independent of (X_i). Under the assumptions of the theorem, we have that

$$\lim_{n\to\infty} \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^{n} \varepsilon_i X_i \right\| = \lim_{n\to\infty} \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^{n} g_i X_i \right\| = 0 . \tag{8.25}$$

If X is symmetric, the limit on the left of (8.25) is seen to be 0 using Lemma 7.2 (since $S_n/a_n \to 0$ in probability) and the elementary fact that $\lim_{n\to\infty} n a_n^{-1} \mathbb{E}(\|X\| I_{\{\|X\| > a_n\}}) = 0$ under $\mathbb{E}(\|X\|^2 / LL\|X\|) < \infty$. By symmetrization, the left of (8.25) also holds in general since X is centered (Lemma 6.3). One can also use for this result Corollary 10.2 for a_n. Concerning Gaussian randomization, we refer to Proposition 10.4 below and the comments thereafter. Using the latter property and the Gaussian Sudakov minoration, the proof of Lemma 8.3 is trivially modified to this setting to yield the existence of a sequence (α_n) of positive numbers tending to 0 such that for all n large enough

$$N(U, d_2^n; \varepsilon) \leq \exp(\alpha_n LL m_n) . \tag{8.26}$$

According to this result, we denote, for each n and f in U, by $h_n(f)$ an element of U such that $d_2^n(f, h_n(f)) < \varepsilon$ in such a way that the set U_n of all $h_n(f)$ has a cardinality less than $\exp(\alpha_n LL m_n)$. We can write that

$$\left\| \sum_{i=1}^{m_n} X_i \right\| \leq \sup_{f \in U_n} \left| \sum_{i=1}^{m_n} f(X_i) \right| + \sup_{h \in V_n} \left| \sum_{i=1}^{m_n} h(X_i) \right|$$

where $V_n = \{f - h_n(f)\,;\ f \in U\} \subset 2U$. The main observation concerning V_n is that $\mathbb{E}(h^2(X) I_{\{\|X\| \le a_{m_n}\}}) \le \varepsilon^2$ for all h in V_n and all n . Although the proof of Theorem 8.6 and (8.19) through Theorem 7.5 is described in the setting of a single true norm of a Banach space, it is clear that it also applies to more general norms which might depend on n on the block of size m_n . In this way, it is just a mere exercise to verify that for some numerical constant C ,

$$\sum_n \mathbb{P}\left\{\sup_{h \in V_n} \left|\sum_{i=1}^{m_n} h(X_i)\right| > C\varepsilon a_{m_n}\right\} < \infty\,.$$

We are thus left to show that, for some $C > 0$,

$$\sum_n \mathbb{P}\left\{\sup_{f \in U_n} \left|\sum_{i=1}^{m_n} f(X_i)\right| > (1 + C\varepsilon) a_{m_n}\right\} < \infty\,. \tag{8.27}$$

Let $\delta = \delta(\varepsilon) > 0$ to be specified in a moment and set, for each n , $c_n = \delta m_n / a_{m_n}$. Define, for each n , $i \le m_n$ and f in U ,

$$Y_i(f,n) = \max(-c_n\,,\ \min(f(X_i)\,,\ c_n)) - \mathbb{E}(\max(-c_n\,,\ \min(f(X_i)\,,\ c_n)))\,.$$

Note that $|Y_i(f,n)| \le 2c_n$ and $\mathbb{E}(Y_i(f,n)^2) \le 1$. By Lemma 1.6 applied to the sum of the independent mean zero random variables $Y_i(f,n)$, $i \le m_n$, it follows that

$$\mathbb{P}\left\{\sup_{f \in U_n} \left|\sum_{i=1}^{m_n} Y_i(f,n)\right| > (1+\varepsilon) a_{m_n}\right\} \le 2\ \mathrm{Card} U_n \exp(-(1+\varepsilon) LL m_n)$$

provided $\delta = \delta(\varepsilon) > 0$ is small enough in order that $2 - \exp(2(1+\varepsilon)\delta) \ge (1+\varepsilon)^{-1}$. By (8.26), it thus already follows that

$$\sum_n \mathbb{P}\left\{\sup_{f \in U_n} \left|\sum_{i=1}^{m_n} Y_i(f,n)\right| > (1+\varepsilon) a_{m_n}\right\} < \infty\,. \tag{8.28}$$

Consider now $Z_i(f,n) = f(X_i) - Y_i(f,n)$, $i \le m_n$, $f \in U$, $n \ge 1$. Note that, by centering of $f(X_i)$,

$$\mathbb{E}|Z_i(f,n)| \le 2\mathbb{E}(\|X\| I_{\{\|X\| > c_n\}})\,.$$

The integrability condition $\mathbb{E}(\|X\|^2 / LL\|X\|) < \infty$ is equivalent to say that $\sum_n \beta_n < \infty$ where

$$\beta_n = \frac{m_n}{(LLm_n)^2} \mathbb{P}\{\|X\| > c_n\}$$

(elementary verification). There exists a sequence (γ_n) such that $\gamma_n \ge \beta_n$, $\sum_n \gamma_n < \infty$ and satisfying the regularity property $\gamma_{n+1} \le \rho^{1/3}\gamma_n$ for every n (recall that $\rho > 1$). It is then easily seen that for all n,

$$\begin{aligned}
\mathbb{E}(\|X\| I_{\{\|X\|>c_n\}}) &\le \sum_{\ell \ge n} c_{\ell+1} \mathbb{P}\{\|X\| > c_\ell\} \\
&\le \sum_{\ell \ge n} \frac{c_{\ell+1}(LLm_\ell)^2}{m_\ell} \gamma_\ell \\
&\le C_1(\rho,\delta)\gamma_n \sum_{\ell \ge n} \frac{(LL\rho^\ell)^{3/2}}{\rho^{\ell/2}} \rho^{(\ell-n)/3} \\
&\le C_2(\rho,\delta)\gamma_n \frac{(LLm_n)^{3/2}}{m_n^{1/2}}
\end{aligned}$$

for some $C_1(\rho,\delta)$, $C_2(\rho,\delta) > 0$. Consider the set of integers $L = \{n\,;\ 2C_2(\rho,\delta)\gamma_n LLm_n \le \varepsilon\}$. The preceding estimate indicates that for all $n \in L$, $f \in U$ and $i \le m_n$,

$$\mathbb{E}|Z_i(f,n)| \le \varepsilon a_{m_n}/m_n\,.$$

We now use this property to show that if $n \in L$ is large enough,

$$\mathbb{E}\left(\sup_{f\in U} \sum_{i=1}^{m_n} |Z_i(f,n)|\right) \le 2\varepsilon a_{m_n}\,. \tag{8.29}$$

Indeed, from Theorem 4.12, (8.25) implies that

$$\lim_{n\to\infty} \frac{1}{a_{m_n}} \mathbb{E}\left(\sup_{f\in U}\left|\sum_{i=1}^{m_n} \varepsilon_i |Z_i(f,n)|\right|\right) = 0\,;$$

hence, by Lemma 6.3,

$$\lim_{n\to\infty} \frac{1}{a_{m_n}} \mathbb{E}\left(\sup_{f\in U}\left|\sum_{i=1}^{m_n} |Z_i(f,n)| - \mathbb{E}|Z_i(f,n)|\right|\right) = 0$$

from which the announced property (8.29) follows.

The main interest in the introduction of the absolute values in (8.29) is that it allows a simple use of the isoperimetric inequality (6.16). It provides us indeed with the crucial monotonicity property (cf. Remark 6.18 about positive random variables). More precisely, let $n \in L$ and set

$$A = \left\{\omega \in \Omega\,;\ \sup_{f\in U} \sum_{i=1}^{m_n} |Z_i(f,n)(\omega)| \le 4\varepsilon a_{m_n}\right\}.$$

Then $\mathbb{P}(A) \geq 1/2$ by (8.29) (at least for n large). Now, if, for $\omega \in \Omega$, there exist $\omega^1, \ldots, \omega^q$ in A such that $X_i(\omega) \in \{X_i(\omega^1), \ldots, X_i(\omega^q)\}$ except perhaps for at most k values of $i \leq m_n$, then

$$\begin{aligned}\sup_{f \in U} \sum_{i=1}^{m_n} |Z_i(f,n)(\omega)| &\leq \sum_{i=1}^{k} \|Z_i(\omega)\|^* + \sum_{\ell=1}^{q} \sup_{f \in U} \sum_{i=1}^{m_n} |Z_i(f,n)(\omega^\ell)| \\ &\leq \sum_{i=1}^{k} \|Z_i(\omega)\|^* + 4q\varepsilon a_{m_n}\end{aligned}$$

where $(\|Z_i\|^*)$ denotes the non-increasing rearrangement of $(\|X_i\| + \mathbb{E}\|X_i\|)_{i \leq m_n}$. Hence, the isoperimetric inequality (6.16) ensures that for $k \geq q$

$$\mathbb{P}\left\{\sup_{f \in U} \sum_{i=1}^{m_n} |Z_i(f,n)| > (1+4q)\varepsilon a_{m_n}\right\} \leq \left(\frac{K_0}{q}\right)^k + \mathbb{P}\left\{\sum_{i=1}^{k} \|Z_i\|^* > \varepsilon a_{m_n}\right\}.$$

If we now choose $q = 2K_0$ and $k = k_n$ as in the proof of Theorem 8.6 using the integrability condition $\mathbb{E}(\|X\|^2 / LL\|X\|) < \infty$, we get that

$$\sum_{n \in L} \mathbb{P}\left\{\sup_{f \in U} \sum_{i=1}^{m_n} |Z_i(f,n)| > (1+8K_0)\varepsilon a_{m_n}\right\} < \infty. \tag{8.30}$$

Combining (8.28) and (8.30), we see that in order to establish (8.27) and conclude the proof of the theorem, we have to show that for some numerical $C > 0$,

$$\sum_{n \notin L} \mathbb{P}\left\{\sup_{f \in U_n} \left|\sum_{i=1}^{m_n} f(X_i)\right| > C\varepsilon a_{m_n}\right\} < \infty. \tag{8.31}$$

We follow very much the pattern of the case $n \in L$. Let now $c'_n = m_n / 4\varepsilon a_{m_n}$, and define $Y'_i(f,n)$, $Z'_i(f,n)$ as $Y_i(f,n)$, $Z_i(f,n)$ before but with c'_n instead of c_n. We observe, now because $\sigma(X) \leq 1$, that

$$\mathbb{E}|Z'_i(f,n)| \leq 8\varepsilon a_{m_n} / m_n,$$

$i \leq m_n$. Exactly as what we described before for $Z_i(f,n)$, we can get from the isoperimetric approach that

$$\sum_{n} \mathbb{P}\left\{\sup_{f \in U} \sum_{i=1}^{m_n} |Z'_i(f,n)| > C\varepsilon a_{m_n}\right\} < \infty. \tag{8.32}$$

Concerning $Y'_i(f,n)$, the exponential inequality of Lemma 1.6 shows that

$$\begin{aligned}\mathbb{P}\left\{\sup_{f \in U_n} \left|\sum_{i=1}^{m_n} Y'_i(f,n)\right| > \varepsilon a_{m_n}\right\} &\leq 2 \ \mathrm{Card} U_n \exp(-\varepsilon^2(2-\sqrt{e})LLm_n) \\ &\leq 2\exp\left(-\left(\frac{\varepsilon^2}{4} - \alpha_n\right) LLm_n\right)\end{aligned}$$

where we have used (8.26). Now, if $n \notin L$, $LLm_n > \varepsilon(2C_2(\rho,\delta)\gamma_n)^{-1}$ where $\sum_n \gamma_n < \infty$. Since $\alpha_n \to 0$, we clearly get that

$$\sum_{n \notin L} \mathbb{P}\left\{ \sup_{f \in U_n} \left| \sum_{i=1}^{m_n} Y_i'(f,n) \right| > \varepsilon a_{m_n} \right\} < \infty$$

from which, together with (8.32), (8.31) follows. This completes the proof of Theorem 8.11.

Theorem 8.11 settles the question of the identification of the limits when $S_n/a_n \to 0$ in probability. Very few however is known at the present time about the limit (8.20), or just only (8.22), in case of the bounded LIL, that is, in the setting of Theorem 8.11, when (S_n/a_n) is only bounded in probability (cf. Theorem 8.6). The limsup $\Lambda(X)$ need not be equal to $\sigma(X)$ and has to take into account the stochastic boundedness of (S_n/a_n), for example through $\Gamma(X) = \limsup_{n\to\infty} \mathbb{E}\|S_n/a_n\|$ (cf. [A-K-L]). One might wonder what function of $\sigma(X)$ and $\Gamma(X)$ (or some other quantity equivalent to $\Gamma(X)$) $\Lambda(X)$ could be. We believe that this study could lead to some rather intricated situations as suggested by the following example with which we close this chapter. We construct here an example showing that the condition $S_n/a_n \to 0$ in probability in Theorem 8.11 is *not* necessary for the limit $\lim_{n\to\infty} d(S_n/a_n, K)$ to be almost surely 0. Let us note that K. Alexander [Ale4] showed that when $\lim_{n\to\infty} d(S_n/a_n, K) = 0$, then necessarily $C(S_n/a_n) = K$.

Example 8.12. There exists a random variable X satisfying the bounded LIL in the Banach space c_0 such that

$$\lim_{n\to\infty} d\left(\frac{S_n}{a_n}, K\right) = 0$$

with probability one, where $K = K_X$ the unit ball of the reproducing kernel Hilbert space associated to the covariance structure of X, but for which S_n/a_n does not converge in probability to 0.

The construction of this example is based on the following preliminary study which appears to be kind of canonical and could possibly be useful for related constructions. It will be convenient for this study, as well as for the example itself, to use the language and notations of empirical processes (cf. Chapter 14).

Let I be a subinterval of $[0,1]$ of length b divided into p equal subintervals. Consider the class $\mathcal{G}$ of functions on $[0,1]$ defined by $\mathcal{G} = \{I_A ; A$ is the union of h subintervals (of I)$\}$. It is implicit that p and h are large enough for all the computations we will develop. With some abuse, we denote by (X_i) a sequence of independent variables uniformly distributed on $[0,1]$. We study here, for every n,

$$\mathbb{E}\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}}$$

where $\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} = \sup_{f \in \mathcal{G}} \left| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right|$ and, as usual, (ε_i) is a Rademacher sequence independent of (X_i). Note first that, obviously, $\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} \leq \operatorname{Card}\{i \leq n\,;\ X_i \in I\}$ so that

$$\mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} \leq bn\,. \tag{8.33}$$

Let us now try to improve this general estimate for relative values of n. To this aim, we use Bennett's inequality (6.10) from which we easily deduce that, for all f in $\mathcal{G}$ and all $t > 0$,

$$\mathbb{P}\left\{ \left| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right| > t \right\} le 2 \exp\left[-\frac{t}{C_1} \theta\left(\frac{t}{n\sigma^2} \right) \right]$$

where $\sigma^2 = hb/p$, $\theta(u) = u$ when $0 < u \leq e$, $\theta(u) = e \log u$ when $u \geq e$ and where C_1 is some (large) numerical constant. Consider now t_0 such that $t_0 \theta(t_0/n\sigma^2)/C_1 \geq h \log p$ (≥ 1). Since $\operatorname{Card}\mathcal{G} \leq p^h = \exp(h \log p)$ and θ is increasing, for all $t \geq t_0$,

$$\mathbb{P}\left\{ \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} > t \right\} \leq 2 \exp\left[h \log p - \frac{t}{C_1} \theta\left(\frac{t_0}{n\sigma^2} \right) \right] .$$

Now, by integration by parts and definition of t_0, it follows that $\mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} \leq 3t_0$. It is easily verified that when $n \geq C_1 p \log p / e^2 b$, one may take t_0 to be $\sqrt{C_1}\sqrt{n} h (b \log p / p)^{1/2}$ whereas when $n \leq C_1 p \log p / e^2 b$ we can take

$$t_0 = \frac{C_1 e h \log p}{\log\left(\frac{C_1 e p \log p}{bn} \right)} .$$

We have thus obtained, combining with (8.33), that:

$$\frac{1}{\sqrt{n}} \mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} \leq b\sqrt{n} \qquad \text{if } n \leq \frac{h}{b}\,, \tag{8.34}$$

$$\frac{1}{\sqrt{n}} \mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} \leq \frac{3 C_1 e h \log p}{\sqrt{n} \log\left(\frac{C_1 e p \log p}{bn} \right)} \qquad \text{if } n \leq \frac{C_1}{e^2}\ \frac{p \log p}{b}\,, \tag{8.35}$$

$$\frac{1}{\sqrt{n}} \mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{G}} \leq 3\sqrt{C_1} h \left(\frac{b \log p}{p} \right)^{1/2} \qquad \text{if } n \geq \frac{C_1}{e^2}\ \frac{p \log p}{b}\,. \tag{8.36}$$

Since the right hand side of (8.35) is decreasing in n (for the values of n considered), we obtain a first consequence of this investigation. C denotes below some numerical constant possibly varying from line to line.

Corollary 8.13. If $m \geq h/b$,

$$\sup_{n\geq m} \frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}} \leq C\max\left(\frac{h\log p}{\sqrt{LL\frac{h}{b}}\sqrt{m}\log\left(\frac{C_1 ep\log p}{m}\right)}\,,\ \frac{h}{\sqrt{LL\frac{h}{b}}}\left(\frac{b\log p}{p}\right)^{1/2}\right).$$

Corollary 8.14. If $m \geq h/b$ and, in addition, $C_1 p \geq m^2$,

$$\sup_{n\geq m} \frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}} \leq C\max\left(\frac{h}{\sqrt{LL\frac{h}{b}}\sqrt{m}}\,,\ \frac{h}{\sqrt{LL\frac{h}{b}}}\left(\frac{b\log p}{p}\right)^{1/2}\right).$$

To obtain a control which is uniform in n, consider the bound (8.34) for $r = h/b$ and the previous corollary.

Corollary 8.15. If $C_1 p \geq h^2$, then

$$\sup_{n} \frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}} \leq C\frac{\sqrt{hb}}{\sqrt{LL\frac{h}{b}}}.$$

Provided with this preliminary study and the preceding statements, we now start the construction of Example 8.12. Consider increasing sequences of integers $(n(q))$, $(p(q))$, $(s(q))$ to be specified later on. Let I_q, J_q, $q \in \mathbb{N}$, be disjoint intervals in $[0,1]$ where, for each q, I_q has length $b(q) = LLn(q)/n(q)$ and J_q, $b(q)/s(q)$. We divide I_q in $p(q)$ equal subintervals and denote by $\mathcal{I}_q$ the family of these subintervals. We set

$$a(q) = a_{n(q)} = (2n(q)LLn(q))^{1/2} \qquad \text{and} \qquad c(q) = \frac{1}{10}\,\frac{a(q)}{LLn(q)}.$$

We set further, always for all q,

$$\mathcal{G}_q = \{c(q)I_A\,;\ A \text{ is union of } [n(q)b(q)] \text{ intervals of } \mathcal{I}_q\}$$

where $[\cdot]$ is the integer part function. We note that for every f in $\mathcal{G}_q$, $\|f\|_2 = d(q)$ where

$$d(q)^2 = c(q)^2[n(q)b(q)]\frac{b(q)}{p(q)}$$

which is equivalent, for q large, to $2LLn(q)/10^2p(q)$. Let T_q be the affine map from J_q into I_q. For f with support in I_q and constant on the intervals of J_q, set

$$U_q(f) = f + \left(\frac{1-d(q)^2}{d(q)^2}\right)^{1/2} \sqrt{s(q)} f_\circ T_q$$

so that $\|U_q(f)\|_2 = \|f\|_2/d(q)$. We set further $\mathcal{F}_q = \{U_q(f)\,;\ f \in \mathcal{G}_q\}$, $\mathcal{G}'_q = \{U_q(f) - f\,;\ f \in \mathcal{G}_q\}$. In particular $\|f\|_2 = 1$ if $f \in \mathcal{F}_q$. Let $\mathcal{F} = \bigcup\limits_q \mathcal{F}_q$. We will show that one can choose appropriately the sequences $(n(q)), (p(q)), (s(q))$ such that

$$\frac{1}{a_{n(q)}} \left\| \sum_{i=1}^{n(q)} \varepsilon_i f(X_i) \right\|_{\mathcal{F}} \not\to 0 \ \text{ in probability} \tag{8.37}$$

and, for all $u > 0$,

$$\limsup_{n\to\infty} \frac{1}{a_n} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_u} \le 1 \ \text{ almost surely} \tag{8.38}$$

where $\mathcal{F}_u = \{f \in u\ \mathrm{Conv}\mathcal{F}\,;\ \|f\|_2 \le 1\}$. (The notation $\|\cdot\|_{\mathcal{F}}$ has the same meaning as in the preliminary study.)

Before turning to the somewhat technical details of the construction, let us show why (8.37) and (8.38) give raise to Example 8.12. Observe that $\mathcal{F}$ is countable. Let X be the map from $[0,1] \times \{-1,+1\}$ in $c_0(\mathcal{F})$ defined by $X(x,\varepsilon) = (\varepsilon f(x))_{f\in\mathcal{F}}$. Since the intervals I_q, J_q are disjoint, X actually takes its values in the space of finite sequences. That $S_n/a_n \not\to 0$ in probability follows from (8.37). To establish that $\lim\limits_{n\to\infty} d(S_n/a_n, K) = 0$ almost surely where $K = K_X$, it suffices to show (as in Theorem 8.11) that for every $\varepsilon > 0$, $\limsup\limits_{n\to\infty} |||S_n/a_n||| \le 1$ with probability one where $|||\cdot|||$ is the gauge of $K + \varepsilon B_1$, B_1 the unit ball of $c_0(\mathcal{F})$. But the unit ball of the dual norm is $V = \{g \in \ell_1(\mathcal{F})\,;\ \|g\| \le 1/\varepsilon\,,\ \mathbb{E}g(X)^2 \le 1\}$ so that it thus suffices to have that

$$\limsup_{n\to\infty} \sup_{g\in V} \frac{1}{a_n} \left| \sum_{i=1}^{n} \varepsilon_i g(X_i) \right| \le 1 \ \text{ almost surely}.$$

Let V_0 be the elements of V with finite support. Then (8.38) exactly means that

$$\limsup_{n\to\infty} \sup_{g\in V_0} \frac{1}{a_n} \left| \sum_{i=1}^{n} \varepsilon_i g(X_i) \right| \le 1 \ \text{ almost surely},$$

and since V_0 is easily seen to be dense in norm in V, the conclusion follows.

Let us now turn to the construction and proof of (8.37) and (8.38). We will actually construct (by induction) the sequences $(p(q))$ and $(s(q))$ from sequences $(r(q))$ and $(m(q))$ such that $r(q-1) < n(q) < m(q) < r(q)$ (where each strict inequality actually means much bigger). The case $q = 1$, similar to the general case is left to the reader. Let us therefore assume that $r(q-1)$ has been constructed. We then take $n(q)$ large enough in order that

$$LLn(q) \geq 2^{4q}\,, \qquad LLc(q) \geq 2^q\,, \qquad b(q) \leq 2^q p(q-1) \tag{8.39}$$

(which is possible since $c(q) = a(q)/10LLn(q)$ and $b(q) = LLn(q)/n(q)$) and

$$\frac{r(q-1)}{a_{r(q-1)}} \cdot \frac{a(q)}{n(q)} \leq 2^{-q}\,. \tag{8.40}$$

We then take $p(q)$ sufficiently large such that

$$p(q) \geq n(q)^4\,. \tag{8.41}$$

Set $m(q) = [(2^{-q}p(q))^{1/2}]$ and choose then $s(q)$ large enough so that

$$s(q) \geq 2^q m(q) b(q)\,, \qquad s(q)d(q)^2 \geq 1 \tag{8.42}$$

and (what is actually a stronger condition)

$$LL\sqrt{s(q)} \geq p(q)\,. \tag{8.43}$$

We are left with the choice of $r(q)$. To this aim, set $\mathcal{H}_q = \{g \in 2^q \operatorname{Conv}(\bigcup_{\ell \leq q} \mathcal{F}_\ell)\,;\ \|g\|_2 \leq 1\}$. $\mathcal{H}_q$ is a convex set of finite dimension. There exists a finite set $\mathcal{H}'_q$ such that $\mathcal{H}_q \subset \operatorname{Conv}\mathcal{H}'_q$ and $\|g\|_2 \leq (1-2^{-q-1})^{-1/2}$ for $g \in \mathcal{H}'_q$. The exponential inequality of Kolmogorov (Lemma 1.6), applied to each g in $\mathcal{H}'_q$, easily implies that one can find $r(q)$ such that, for all $n \geq r(q)$ and all t with $a_n/2 \leq t \leq 2a_n$,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathcal{H}_q} > t\right\} \leq 2\exp\left[-\frac{t^2}{2n}(1-2^{-q})\right] \tag{8.44}$$

and

$$\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{H}_q} \leq C\sqrt{n} \tag{8.45}$$

where C is numerical. This completes the construction (by induction) of the various sequences of integers we will work with. We can now turn to the proofs of (8.37) and (8.38).

Let T_q^1 be the event

$$T_q^1 = \{\forall i \leq m(q)\,,\ \text{ each interval of } \mathcal{I}_q \text{ contains at most one } X_i\}\,.$$

Clearly, if $T_q^{1,c}$ is the complement of T_q^1,

$$\mathbb{P}(T_q^{1,c}) \leq \frac{1}{2}m(q)(m(q)-1)p(q)\left(\frac{b(q)}{p(q)}\right)^2$$

and thus, by definition of $m(q)$ and the fact that $b(q) \leq 1$, $\mathbb{P}(T_q^{1,c}) \leq 2^{-q}$. Similarly, let

$$T'_q = \{\forall i \leq m(q)\,,\ X_i \notin J_q\}\,.$$

Then $\mathbb{P}(T_q^{\prime,c}) \leq m(q)b(q)/s(q) \leq 2^{-q}$ by (8.42). We can then already show (8.37). From these estimates indeed, for all q large enough, on a set of probability bigger than $1/3$ (for example), there are at least $[n(q)b(q)]$ points X_i, $i \leq n(q)$, in I_q which are in different intervals of $\mathcal{I}_q$ and not in J_q. There exists therefore a union A of $[n(q)b(q)]$ intervals of $\mathcal{I}_q$ such that $\left|\sum_{i=1}^{n(q)}\varepsilon_i I_A(X_i)\right| \geq \frac{1}{2}[n(q)b(q)]$. Since $c(q)n(q)b(q) = a(q)/10$, it follows that, for all q large enough,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^{n(q)}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_q} \geq a(q)/40\right\} \geq \frac{1}{3}$$

for which (8.37) clearly follows.

We now turn to (8.38) which is of course the most difficult part. Let us fix $u > 0$ and recall that $\mathcal{F}_u = \{f \in u\mathrm{C}\,onv\mathcal{F}\,;\ \|f\|_2 \leq 1\}$. As in the proof of Theorem 8.11, it suffice to show that for all $\xi > 1$, $\rho > 1$,

$$\limsup_{n\to\infty}\frac{1}{a_{\rho(n)}}\left\|\sum_{i=1}^{\rho(n)}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_u} \leq \xi \quad \text{almost surely} \tag{8.46}$$

where $\rho(n) = [\rho^n]$. We set $\mathbb{N}_1 = \bigcup_q [r(q-1), m(q)]$ and $\mathbb{N}_2 = \bigcup_q [m(q), r(q)]$ (as subsets of integers) and study separately $\limsup_{\rho(n)\in\mathbb{N}_1}$ and $\limsup_{\rho(n)\in\mathbb{N}_2}$. Let us first consider the limsup when $\rho(n) \in \mathbb{N}_2$. We make use for it of the proof of Theorem 8.11 from which we know that we will get a limsup less than 1 under the conditions $\mathbb{E}(g^2/LLg) < \infty$ where $g = \|f\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |f|$ and

$$\lim_{n\in\mathbb{N}_2} \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} = 0 \,. \tag{8.47}$$

To check the integrability condition, let $g_q = \sup_{f\in\mathcal{F}_q} |f|$ which have disjoint supports. Moreover

$$\begin{aligned} g_q &= c(q) I_{I_q} + c(q) \left(\frac{1-d(q)^2}{d(q)^2} \right)^{1/2} \sqrt{s(q)} I_{J_q} \\ &\le c(q) I_{I_q} + \frac{c(q)}{d(q)} \sqrt{s(q)} I_{J_q} \,. \end{aligned}$$

It follows that, for all large q,

$$\begin{aligned} \mathbb{E}(g_q^2/LLg_q^2) &\le \frac{c(q)^2 b(q)}{LLc(q)} + \frac{c(q)^2 b(q)}{d(q)^2 LL\sqrt{s(q)}} \\ &\le \frac{1}{LLc(q)} + \frac{p(q)}{LLn(q) LL\sqrt{s(q)}} \end{aligned}$$

which gives raise to a summable series by (8.39) and (8.43). Turning to (8.47), let $n \in [m(q), r(q)]$. Note that $\mathcal{F} \subset \mathcal{H}_q \cup \mathcal{F}_q \cup \bigcup_{\ell>q} \mathcal{F}_\ell$ so that

$$\begin{aligned} \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} &\le \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{H}_q} + \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_q} \\ &\quad + \frac{1}{a_n} \mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\bigcup_{\ell>q} \mathcal{F}_\ell} \\ &\le \text{(I)} + \text{(II)} + \text{(III)} \,. \end{aligned}$$

By (8.45), (I) has a limit zero. Concerning (III), for $\ell > q$ we have that (cf. (8.33))

$$\mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{G}_\ell} \le nc(\ell) b(\ell) \le n \frac{a(\ell)}{n(\ell)}$$

so that, since $n \le r(q)$,

$$\frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}_\ell} \le \frac{n}{a_n}\cdot\frac{a(\ell)}{n(\ell)} \le \frac{r(q)}{a_{r(q)}}\cdot\frac{a(\ell)}{n(\ell)} \le 2^{-\ell}$$

by (8.40). On the other hand, by (8.42),

$$\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}'_\ell} \le n\frac{c(\ell)}{d(\ell)}\cdot\frac{b(\ell)}{\sqrt{s(\ell)}} \le nc(\ell)b(\ell) \le n\frac{a(\ell)}{n(\ell)}\,,$$

and thus, as before, $\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}'_\ell} \le a_n 2^{-\ell}$. It follows that $\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_\ell} \le a_n 2^{-\ell+1}$ and therefore that (III) $\le 2^{-q+2}$. Hence, (III) also has a limit which is 0 . We are left with (II). As for (III), we write that

$$\text{(III)} \le \frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}_q} + \frac{1}{a_n}\mathbb{E}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{G}'_q} = \text{(IV)} + \text{(V)}\,.$$

We evaluate (IV) and (V) by the preliminary study. For (IV), we let there $b = b(q)$, $h = [n(q)b(q)]$, $p = p(q)$ and $m = m(q)$. For q large enough, the definition of $m(q)$ $(m(q) = [(2^{-q}p(q))^{1/2}])$ shows that we are in a position to apply Corollary 8.14. Since $b(q) \le 1$ and $c(q)n(q)b(q) = a(q)/10$, it follows that, for some numerical constant C and all q large enough,

$$\text{(IV)} \le C\max\left(\left(\frac{n(q)}{m(q)}\right)^{1/2}\,,\ \left(\frac{n(q)}{p(q)}\log p(q)\right)^{1/2}\right)\,.$$

By the choice of $p(q) \ge n(q)^4$, we have that $m(q) \ge 2^{-q-1}n(q)^2$, from which, together with (8.39), we deduce that (IV) tends to 0 . Using Corollary 8.15 with $b = b(q)/s(q)$, $h = [n(q)b(q)]$ and $p = p(q)$, the control of (V) is similar by (8.43). We have therefore established in this way that (8.47) holds and thus (8.46) along $\rho(n) \in \mathbb{N}_2$.

In the last part of this proof, we establish (8.46) when $\rho(n) \in \mathbb{N}_1$. For each q , consider $T_q = \bigcap_{i=1}^{5} T_q^i$ where

$T_q^1 = \{\forall i \le m(q)\,,$ each interval of $\mathcal{I}_q$ contains at most one $X_i\}$,

$T_q^2 = \{\forall i \le m(q)\,,\ X_i \notin J_q \cup \bigcup_{\ell > q}(I_\ell \cup J_\ell\}$,

$T_q^3 = \{$ for $2^{-2q}n(q) \le n \le m(q)\,,\ \mathrm{Card}\{i \le n\,;\ X_i \in I_q\} \le 2e^2 nb(q)\}$,

$T_q^4 = \{$ for $2^{-q}n(q)/LLn(q) \le n \le 2^{-2q}n(q)\,,\ \ \mathrm{Card}\{i \le n\,;\ X_i \in I_q\} \le e^2 2^{-q+1}a_n/c(q)\}$,

$T_q^5 = \{\forall i \le 2^{-q}n(q)/LLn(q)\,,\ X_i \notin I_q\}$.

We would like to show that $\sum_q \mathbb{P}(T_q^c) < \infty$ for which it suffices to prove that $\sum_q \mathbb{P}(T_q^{i,c}) < \infty$, $i = 1, \ldots, 5$, where $T_q^{i,c}$ is the complement of T_q^i . We have already seen that $\mathbb{P}(T_q^{1,c}) \le 2^{-q}$. For $i = 2$, note that, when $\ell \ge q$,

$$\mathbb{P}\{\exists i \le m(q)\,,\ X_i \in J_\ell\} \le m(q)\frac{b(\ell)}{s(\ell)} \le m(\ell)\frac{b(\ell)}{s(\ell)} \le 2^{-\ell}$$

by (8.42). When $\ell > q$, by (8.39),

$$\mathbb{P}\{\exists i \le m(q)\,,\ X_i \in I_\ell\} \le m(q)b(\ell) \le p(q)b(\ell) \le p(\ell-1)b(\ell) \le 2^{-\ell}\,.$$

Hence $\mathbb{P}(T_q^{2,c}) \le 3 \cdot 2^{-q}$. Concerning T_q^5 , one need simply note that $|I_q| = b(q) = LLn(q)/n(q)$. For $i = 3, 4$, we use the binomial bound (1.16) which implies in particular that

$$\mathbb{P}\{B(n,\tau) \ge tn\} \le \exp\left[-tn\left(\frac{\tau}{t} - 1 - \log\frac{\tau}{t}\right)\right] \le \exp(-tn) \tag{8.48}$$

when $t \ge e^2\tau$ (for example), where $B(n,\tau)$ is the number of successes in a run of n Bernoulli trials with probability of success τ and $0 < t < 1$. We have that

$$T_q^3 \subset \bigcap_{\ell \ge 0} \{\text{ Card}\{i \le 2^{-2q+\ell}n(q)\,;\ X_i \in I_q\} \le e^2 2^{-2q+\ell} b(q)\}\,.$$

Then, taking in (8.48) $\tau = b(q)$ and $t = e^2\tau$ (q large), we have that

$$\mathbb{P}(T_q^{3,c}) \le \sum_{\ell \ge 0} \exp(-e^2 2^{-2q+\ell} n(q)b(q))\,.$$

Since $n(q)b(q) = LLn(q) \ge 2^{4q}$ by (8.39), it follows that $\sum_q \mathbb{P}(T_q^3) < \infty$. Concerning T_q^4 , set $v(\ell) = [2^{-q+\ell}n(q)/LLn(q)]$. Then

$$T_q^4 \subset \bigcap \{\text{Card}\{i \le v(\ell)\,;\ X_i \in I_q\} \le e^2 2^{-q} a_{v(\ell)}/c(q)\}$$

where the intersection is over all $\ell \ge 0$ such that $v(\ell) \le 2^{-2q}n(q)$, i.e. $2^\ell \le 2^{-q}LLn(q)$. Take then in (8.48), $t = e^2 2^{-q} a_{v(\ell)}/v(\ell)c(q)$, $\tau = b(q)$ and $n = v(\ell)$. Since $2^\ell \le 2^{-q}LLn(q)$, one verifies that, at least for q large enough (by (8.39)), $t \ge e^2\tau$. Hence

$$\mathbb{P}(T_q^{4,c}) \le \sum_{\ell \ge 0} \exp(-e^2 2^{-q} a_{v(\ell)}/c(q))\,.$$

Since $a_{v(\ell)}/c(q) \geq 2^{-q/2+\ell/2}(LLn(q))^{1/2}$, one concludes as for $i = 3$ that $\sum_q \mathbb{P}(T_q^{4,c}) < \infty$. We have thus established that $\sum_q \mathbb{P}(T_q^c) < \infty$.

For every q (large enough), set now

$$P_q = \sum_{r(q-1) \leq \rho(n) \leq m(q)} \mathbb{P}\left\{ \left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_u} > \xi a_{\rho(n)} \; ; \; T_q \right\} .$$

The proof will be completed once we will have shown that $\sum_q P_q < \infty$. Take q large enough so that $2^q > u$. On T_q , none of the X_i 's, $i \leq \rho(n) \leq m(q)$, is in J_ℓ for $\ell \geq q$ or I_ℓ for $\ell > q$ (definition of T_q^2). On the other hand, the restriction of a function of $\mathcal{F}_u$ to $I_q \cap J_q$ belongs to $\mathcal{F}_{q,u} = \{f \in u \operatorname{Conv}\mathcal{F}_q \; ; \; \|f\|_2 \leq 1\}$ and its restriction to $\bigcup_{\ell < q} I_\ell \cup J_\ell$ is in $\mathcal{H}_{q-1}$. We thus have that

$$\left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_u} \leq \left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{H}_{q-1}} + \left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_{q,u}} . \tag{8.49}$$

Lemma 8.16. For every $\varepsilon > 0$, there exists $q_0 = q_0(\varepsilon)$ such that for all $q \geq q_0$ and $r(q-1) \leq \rho(n) \leq m(q)$, one can find numbers $\xi(n,q)$ with the following properties:

$$\text{on } T_q , \qquad \left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_{u,q}} \leq \xi(n,q) \; ; \tag{8.50}$$

$$\text{there exists } M = M(\varepsilon, q_0) \text{ such that } \operatorname{Card}\{n \; ; \; \xi(n,q) > \varepsilon a_{\rho(n)}\} \leq M \; ; \tag{8.51}$$

$$\xi(n,q) \leq \frac{2}{3} a_{\rho(n)} . \tag{8.52}$$

Before proving this lemma, let us show how to conclude from it. By (8.49) and (8.50)

$$\mathbb{P}\left\{ \left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_u} > \xi a_{\rho(n)} \; ; \; T_q \right\} \leq \mathbb{P}\left\{ \left\| \sum_{i=1}^{\rho(n)} \varepsilon_i f(X_i) \right\|_{\mathcal{H}_{q-1}} > \xi a_{\rho(n)} - \xi(n,q) \right\} .$$

Using that $\rho(n) \geq r(q-1)$, we may now use (8.44) and estimate the latter probability by

$$2 \exp\left[-\frac{1}{2\rho(n)} (\xi a_{\rho(n)} - \xi(n,q))^2 (1 - 2^{-q+1}) \right] .$$

In (8.51), take $\varepsilon > 0$ to be such that $\xi - \varepsilon > 1$. If $\xi(n,q) \le \varepsilon a_{\rho(n)}$, $\xi a_{\rho(n)} - \xi(n,q) \ge (\xi - \varepsilon) a_{\rho(n)}$ and the preceding gives raise to a summable series. For each q ($\ge q_0$), the number of $\xi(n,q) > \varepsilon a_{\rho(n)}$ is controlled by M and thus, using (8.52), there is a contribution of at most

$$M \exp\left(-\frac{1}{18} LLn(q-1)\right)$$

which is summable by (8.39).

Proof of Lemma 8.16. The crucial point is the following. If $g \in \mathcal{F}_{q,u}$, let f be the restriction of g to I_q. Then $g = U_q f$. Therefore, $1 \ge \|U_q f\|_2 = \|f\|_2 / d(q)$ so that $\|f\|_2 \le d(q)$. (The first reader that will have comprehensively reached this point of this construction is invited, first to courageously pursue his effort until the end, and then to contact the authors (the second one!) for a reward as a token of his perseverance.) Let n be fixed, $n \le m(q)$, and assume we are on T_q. Let $N = \operatorname{Card}\{i \le n\,;\ X_i \in I_q\}$ so that by Cauchy-Schwartz $\sum_{i=1}^n |f(X_i)| \le \left(N \sum_{i=1}^n f(X_i)^2\right)^{1/2}$. Since on T_q (T_q^1) the X_i's are in distinct intervals of $\mathcal{I}_q$, we have that

$$\begin{aligned} \sum_{i=1}^n f(X_i)^2 \le \sum_{I \in \mathcal{I}_q} (\text{value of } f \text{ on } I)^2 = \|f\|_2^2 \frac{p(q)}{b(q)} \\ \le d(q)^2 \frac{p(q)}{b(q)} \le \frac{2}{100} \cdot \frac{LLn(q)}{b(q)}. \end{aligned}$$

When $n \ge 2^{-2q} n(q)$, $N \le 2e^2 n b(q) \le 20 n b(q)$ (T_q^3) so that

$$\sum_{i=1}^n |f(X_i)| \le \left(\frac{2}{5} n LLn(q)\right)^{1/2}$$

and therefore in this case

$$\frac{1}{a_n} \left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_{u,q}} \le \left(\frac{1}{5} \frac{LLn(q)}{LLn}\right)^{1/2}.$$

Therefore, when $n \ge 2^{-2q} n(q)$ and $n(q)$ is so large that $LL(2^{-2q} n(q)) \ge \frac{1}{2} LLn(q)$, we have that

$$\frac{1}{a_n} \left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_{u,q}} \le \frac{2}{3}. \tag{8.53}$$

On the other hand, obviously (cf. (8.33)),

$$\left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}_{u,q}} \le u \left\| \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathcal{F}} \le u c(q) n(q) b(q) \le \frac{u}{10} a(q)$$

and thus, when $n \geq n(q)$,

$$\frac{1}{a_n}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_{u,q}} \leq \frac{u}{10}\left(\frac{n(q)}{n}\right)^{1/2} . \tag{8.54}$$

Finally, again

$$\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_{u,q}} \leq u\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}} \leq uc(q)\ \mathrm{Card}\{i \leq n\,;\ X_i \in I_q\} .$$

When $n \geq 2^{-2q}n(q)$ (by T_q^3),

$$\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_{u,q}} \leq 20unc(q)b(q) = 2un\frac{a(q)}{n(q)}$$

and therefore

$$\frac{1}{a_n}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_{u,q}} \leq 2u\left(\frac{n}{n(q)}\right)^{1/2} \tag{8.55}$$

(where it is assumed as before that q is large enough so that $LL(2^{-q}n(q)) \geq \frac{1}{2}LLn(q))$. When $n \leq 2^{-2q}n(q)$, by T_q^3 , T_q^4 , $\mathrm{Card}\{i \leq N\,;\ X_i \in I_q\} \leq 20 \cdot 2^{-q}a_n/c(q)$ and therefore

$$\frac{1}{a_n}\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}_{u,q}} \leq 20u2^{-q} . \tag{8.56}$$

Recall u is fixed. We can then simply take

$$\xi(n,q) = \begin{cases} 20u2^{-q}a_{\rho(n)} & \text{if } \rho(n) \leq 2^{-2q}n(q)\,, \\ \min\left(\frac{2}{3}a_{\rho(n)}, 2u\left(\frac{\rho(n)}{n(q)}\right)^{1/2} a_{\rho(n)}\right) & \text{if } 2^{-2q}n(q) \leq \rho(n) \leq n(q)\,, \\ \min\left(\frac{2}{3}a_{\rho(n)}\,,\ \frac{u}{10}\left(\frac{n(q)}{\rho(n)}\right)^{1/2} a_{\rho(n)}\right) & \text{if } \rho(n) \geq n(q)\,. \end{cases}$$

By (8.53)-(8.56), the numbers $\xi(n,q)$ satisfy all the required properties. This completes the proof of Lemma 8.16 and therefore of Example 8.12.

Notes and references

Before reaching the rather complete description we give in this chapter, the study of the law of the iterated logarithm (LIL) for Banach space valued random variables went through several stages of partial results and understanding. We do not try here to give a detailed account of the contributions of all the authors but rather concentrate only on the main steps of the history of these results. The exposition of this chapter is basically taken from the papers [L-T2] and [L-T5]. Let us mention that the LIL is a vast subject of Probability theory from which only one particular aspect is developed here. For a survey of various LIL topics, we refer to [Bin].

The LIL of Kolmogorov appeared in 1929 [Ko] and is extraordinary of accuracy for the time. A. N. Kolmogorov used sharp both upper and lower exponential inequalities (carefully described in [Sto]) presented here as Lemma 1.6 and Lemma 8.1. The extension to the vector valued setting in the form of Theorem 8.2 first appeared in [L-T4] with a limsup only finite. The best result presented here is new. A first form of the vector valued extension is due to J. Kuelbs [Kue4].

The independent and identically distributed scalar LIL is due to P. Hartman and A. Wintner [H-W] (with the proof sketched in the text) who extended previous early results in particular by Hardy and Littlewood and by Khintchine. Necessity was established by V. Strassen [St2]; the simple argument presented here is taken from [Fe2]. The simple qualitative proof by randomization seems to be part of the folklore. For the converse using Gaussian randomization, we refer to [L-T2]. Strassen's paper [St1] is a milestone in the study of the LIL and strongly influenced the infinite dimensional developments. Various simple proofs of the Hartman-Wintner-Strassen scalar LIL are now available, cf. e.g. [Ac7] and the references therein.

The setting up of the framework of the study of the LIL for Banach space valued random variables was undertaken in the early seventies by J. Kuelbs (cf. e.g. [Kue2]). Theorem 8.5 belongs to him [Kue3] (with however a somewhat too strong hypothesis removed in [Pi3]). The definition of the reproducing kernel Hilbert space of a weak-L_2 random variable and Lemma 8.4 on compactness of its unit ball combine observations of [Kue3], [Pi3] and [G-K-Z]. The progresses leading to the final characterization of Theorem 8.6 were numerous. To mention some, R. LePage first showed the result for Gaussian random variables and G. Pisier [Pi3] established the LIL for square integrable random variables satisfying the central limit theorem, a condition weakened later into the boundedness or convergence to 0 in probability of (S_n/a_n) by J. Kuelbs [Kue4]. The first real characterization of the LIL in some infinite dimensional spaces is due to V. Goodman, J. Kuelbs and J. Zinn [G-K-Z] in Hilbert space after a preliminary investigation by G. Pisier and J. Zinn [P-Z]. Then, succesively, several authors extended the conclusion to various classes of smooth normed spaces [A-K], [Led2], [Led5]. The final characterization was obtained in [L-T2] using a Gaussian randomization technique already suggested in [Pi2] and put forward in the unpublished manuscript [Led6] where the case of type 2 spaces (Corollary 8.8) was settled. Lemma 8.7 is taken from [G-K-Z]. The short proof given here based on the isoperimetric approach of Section 6.3 is taken from [L-T5]. Its consequence to the relation between LIL and CLT is discussed in Chapter 10, with in particular results of [Pi3], [G-K-Z], [He2].

The remarkable results on the cluster set $C(S_n/a_n)$ presented here as Theorem 8.10 are due to K. Alexander [Ale3], [Ale4]. Among other results, he further showed that, when $\mathbb{E}\|X\|^2 < \infty$, $C(S_n/a_n)$ is almost surely empty if it does not contain 0 or, equivalently, if (and only if) $\liminf_{n\to\infty} \mathbb{E}\|S_n\|/a_n > 0$. Preliminary observations appeared in [Kue6], [G-K-Z] and in [A-K] where it was first shown that $C(S_n/a_n) = K$ when $S_n/a_n \to 0$ in probability (a result to which the first author of this monograph had some contribution). Theorem 8.11 is taken from [L-T5]. Some observations on possible values of $\Lambda(X)$ in case of the bounded LIL may be found in [A-K-L]. Example 8.12, in the spirit of [Ale4], is new and due to the second author.

Chapter 9. Type and cotype of Banach spaces

Chapter 9. Type and cotype of Banach spaces

The notion of type of a Banach space already appeared in the last chapters on the law of large numbers and the law of the iterated logarithm. We observed there that, in quite general situations, almost sure properties can be reduced to properties in probability or in $L_p, 0 \leq p < \infty$. Starting with this chapter, we will now study the possibility of a control in probability or in the weak topology of probability distributions of sums of independent random variables. On the line or in finite dimensional spaces, such a control is usually easily verified through moment conditions by the orthogonality property

$$\mathbb{E}\left|\sum_i X_i\right|^2 = \sum_i \mathbb{E}|X_i|^2$$

where (X_i) is a finite sequence of independent mean zero real random variables (in L_2). This property extends to Hilbert space with norm instead of absolute values but *does not* extend to general Banach spaces.

This observation already indicates the difficulties in showing tightness or boundedness in probability of sums of independent vector valued random variables. This will be in particular illustrated in the next chapter on the central limit theorem, which is indeed one typical example of this tightness question.

In some classes of Banach spaces, this general question has however reasonable answers. This classification of Banach spaces is based on the concept of type and cotype. These notions have their origin in the preceding orthogonality property which they extend in various ways. They are closely related to geometric properties of Banach spaces of containing or not subspaces isomorphic to ℓ_p^n. Starting with Dvoretzky's theorem on spherical sections of convex bodies, we describe these relations in the first paragraph of this chapter as a geometrical background. A short exposition of some general properties on type and cotype is given in Section 9.2. In the last section, we come back to our starting question and investigate some results on sums of independent random variables in spaces with some type or cotype. In particular, we complete some results on the law of large numbers as announced in Chapter 7. Pregaussian random variables and stable random variables in spaces of stable type are also discussed. We also briefly discuss spaces which do not contain c_0.

9.1 ℓ_p^n-subspaces of Banach spaces

Given $1 \leq p \leq \infty$, recall that ℓ_p^n denotes $\mathbb{R}^n$ equipped with the norm $\left(\sum\limits_{i=1}^n |\alpha_i|^p\right)^{1/p}$ ($\max\lim its_{i\leq n}|\alpha_i|$ if $p = \infty$), $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$. When $1 \leq p \leq \infty$, n an integer and $\varepsilon > 0$, a Banach space B

is said to *contain a subspace* $(1+\varepsilon)$*-isomorphic to* ℓ_p^n if there exist $x_1, \dots, x_n$ in B such that for all $\alpha = (\alpha_1, \dots, \alpha_n)$ in $\mathbb{R}^n$

$$\left(\sum_{i=1}^n |\alpha_i|^p\right)^{1/p} \leq \left\|\sum_{i=1}^n \alpha_i x_i\right\| \leq (1+\varepsilon)\left(\sum_{i=1}^n |\alpha_i|^p\right)^{1/p}$$

($\max_{i \leq n} |\alpha_i|$ if $p = \infty$). B contains ℓ_p^n 's *uniformly* if it contains subspaces $(1+\varepsilon)$-isomorphic to ℓ_p^n for all n and $\varepsilon > 0$.

The purpose of this paragraph is to present some results which relate the set of p 's for which a Banach space contains ℓ_p^n 's to some probabilistic inequalities satisfied by the norm of B. The fundamental result at the basis of this theory is the following theorem of A. Dvoretzky [Dv].

Theorem 9.1. Any infinite dimensional Banach space B contains ℓ_2^n 's uniformly.

It should be mentioned that the various subspaces isomorphic to ℓ_2^n do not form a net and therefore cannot in general be patched together to form an infinite dimensional Hilbertian subspace of B. We shall give two different proofs of Theorem 9.1, both of them based on Gaussian variables and their properties described in Chapter 3. The first one uses isoperimetric and concentration inequalities (Section 3.1), the second comparison theorems (Section 3.3). They actually yield a stronger finite dimensional version of Theorem 9.1 which may be interpreted geometrically as a result on almost spherical sections of convex bodies (cf. [F-L-M], [Mi-S], [Pi18]). This finite dimensional statement is the following.

Theorem 9.2. For each $\varepsilon > 0$ there exists $\eta(\varepsilon) > 0$ such that every Banach space B of dimension N contains a subspace $(1+\varepsilon)$-isomorphic to ℓ_2^n where $n = [\eta(\varepsilon) \log N]$.

Theorem 9.1 clearly follows from this result. In the two proofs of Theorem 9.2 we will give, we use a crucial intermediate result known as the Dvoretzky-Rogers lemma. It will be convenient to immediately interpret this result in terms of Gaussian variables. Recall from Chapter 3 that if X is a Gaussian Radon random variable with values in a Banach space B, we set $\sigma(X) = \sup_{\|f\| \leq 1} (\mathbb{E} f^2(X))^{1/2}$, and X has strong moments of all orders (Corollary 3.2). We may then consider the "dimension" (or "concentration dimension") of a Gaussian variable X as the ratio

$$d(X) = \frac{\mathbb{E}\|X\|^2}{\sigma(X)^2}.$$

Note that $d(X)$ depends on both X and the norm of B. Since all moments of X are equivalent and equivalent to the median $M(X)$ of X, the replacement of $\mathbb{E}\|X\|^2$ by $(\mathbb{E}\|X\|)^2$ or $M(X)^2$ in $d(X)$ gives rise to a dimension equivalent, up to numerical constants, to $d(X)$. We use freely this observation below. We already mentioned in Chapter 3 that the strong moment of a Gaussian vector valued variable is usually much bigger than the weak moments $\sigma(X)$. Recall for example that if X follows the canonical distribution in $\mathbb{R}^N$ and if $B = \ell_2^N$, then $\sigma(X) = 1$ and $\mathbb{E}\|X\|^2 = N$ so that $d(X) = N$. Note that if $B = \ell_\infty^N$, then $\sigma(X)$ is still 1 but $\mathbb{E}\|X\|^2$ is of the order of $\log N$ (cf. (3.14)). One of the conclusions of Dvoretzky-Rogers' lemma is that the case of ℓ_∞^N is extremal. Let us state without proof the result (cf. [D-R], [F-L-M], [Mi-S], [Pi16], [TJ]).

Lemma 9.3. Let B be a Banach space of dimension N and let $\overline{N} = [N/2]$ (integer part). There exist points $(x_i)_{i \le \overline{N}}$ in B such that $\|x_i\| \le 1/2$ for all $i \le \overline{N}$ and satisfying

$$\left\| \sum_{i=1}^{\overline{N}} \alpha_i x_i \right\| \le \left(\sum_{i=1}^{N} |\alpha_i|^2 \right)^{1/2} \tag{9.1}$$

for all $\alpha = (\alpha_1, \ldots, \alpha_{\overline{N}})$ in $\mathbb{R}^{\overline{N}}$. In particular, there exists a Gaussian random vector X with values in B whose dimension $d(X)$ is larger than $c \log N$ where $c > 0$ is some numerical constant.

It is easily seen how the second assertion of the lemma follows from the first one. Let indeed $X = \sum_{i=1}^{\overline{N}} g_i x_i$ where (g_i) is orthogaussian. By (9.1),

$$\sigma(X) = \sup_{\|f\| \le 1} (\mathbb{E} f^2(X))^{1/2} = \sup_{\|f\| \le 1} \left(\sum_{i=1}^{\overline{N}} f^2(x_i) \right)^{1/2} = \sup_{|\alpha| \le 1} \left\| \sum_{i=1}^{\overline{N}} \alpha_i x_i \right\| \le 1 .$$

On the other hand, since $\|x_i\| \le 1/2$ for all $i \le \overline{N}$, by Lévy's inequality (2.7) and (3.14),

$$\mathbb{E}\|X\|^2 \ge \frac{1}{8} \mathbb{E} \max_{i \le \overline{N}} |g_i|^2 \ge c \log \overline{N}.$$

Provided with this tool, we can now attack the proofs of Theorem 9.2.

1^{st} **proof of Theorem 9.2.** It is based on the concentration properties of the norm of a Gaussian random vector around its median or mean and stability properties of Gaussian distributions. We use Lemma 3.1 but the simpler inequality (3.2) can be used completely similarly. In a first step, we need two technical lemmas to discretize the problem of finding ℓ_2^n-subspaces. We state them in a somewhat more general setting for further purposes. A δ-net $(\delta > 0)$ of a set A in $(B, \|\cdot\|)$ is a set S such that for every x in A one can find y in S with $\|x - y\| \le \delta$.

Lemma 9.4. For each $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ with the following property. Let n be an integer and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Let further S be a δ-net of the unit sphere of $(\mathbb{R}^n, \|\cdot\|)$. Then, if for some $x_1, \ldots, x_n$ in a Banach space B, we have $1 - \delta \leq > \|\sum_{i=1}^n \alpha_i x_i\| \leq 1 + \delta$ for all α in S, then

$$(1+\varepsilon)^{-1/2}\|\alpha\| \leq \left\|\sum_{i=1}^n \alpha_i x_i\right\| \leq (1+\varepsilon)^{1/2}\|\alpha\|$$

for all α in $\mathbb{R}^n$.

Proof. By homogeneity, we can and do assume that $\|\alpha\| = 1$. By definition of S, there exists α^0 in S such that $\|\alpha - \alpha^0\| \leq \delta$, hence $\alpha = \alpha^0 + \lambda_1 \beta$ with $\|\beta\| = 1$ and $|\lambda_1| \leq \delta$. Iterating the procedure, we get that $\alpha = \sum_{j=0}^{\infty} \lambda_j \alpha^j$ with $\alpha^j \in S$, $|\lambda_j| \leq \delta^j$ for all $j \geq 0$. Hence

$$\left\|\sum_{i=1}^n \alpha_i x_i\right\| \leq \sum_{j=0}^{\infty} \delta^j \left\|\sum_{i=1}^n \alpha_i^j x_i\right\| \leq \frac{1+\delta}{1-\delta}$$

$(\delta < 1)$. In the same way, $\|\sum_{i=1}^n \alpha_i x_i\| \geq (1 - 3\delta)/(1-\delta)$. It therefore suffices to choose appropriately δ in function of $\varepsilon > 0$ only in order to get the conclusion.

The size of δ-net of spheres in finite dimension is easily estimated in terms of $\delta > 0$ and the dimension, as shown in the next lemma which follows from a simple volume argument.

Lemma 9.5. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. There is a δ-net S of the unit sphere of $(\mathbb{R}^n, \|\cdot\|)$ of cardinality less than $(1 + 2/\delta)^n \leq \exp(2n/\delta)$.

Proof. Let U denote the unit ball of $(\mathbb{R}^n, \|\cdot\|)$ and let $(x_i)_{i \leq m}$ be maximal in the unit sphere of U under the relations $\|x_i - x_j\| > \delta$ for $i \neq j$. Then the balls $x_i + (\delta/2)U$ are disjoint and are all contained in $(1+2/\delta)U$. By comparing the volumes, we get that $m(\delta/2)^n \leq (1+2/\delta)^n$ from which the lemma follows.

We are now in a position to perform the first proof of Theorem 9.2. The main argument is the concentration property of Gausian vectors. Let B of dimension N. By Lemma 9.3 there exists a Gaussian variable X with values in B whose dimension is larger than $c \log N$. Let $M = M(X)$ denote the median of X. Since (Corollary 3.2) M is equivalent to $\|X\|_2$, we also have that $M \geq c(\log N)^{1/2}\sigma(X)$ where $c > 0$ is numerical (possibly different from the preceding). Let n to be specified and consider independent copies $X_1, \ldots, X_n$ of X. With positive probability, the sample $(X_1, \ldots, X_n)$ will be shown to span a subspace almost isometric to ℓ_2^n. More precisely, the basic rotational invariance property of Gaussian distributions indicates that if $\sum_{i=1}^n \alpha_i^2 = 1$, then $\sum_{i=1}^n \alpha_i X_i$ has the same distribution as X. In particular, by Lemma 3.1, for all $t > 0$ and $\alpha_1, \dots, \alpha_n$ with $\sum_{i=1}^n \alpha_i^2 = 1$,

$$\mathbb{P}\left\{ \left| \left\| \sum_{i=1}^n \alpha_i X_i \right\| - M \right| > t \right\} \leq \exp(-t^2/2\sigma(X)^2).$$

Let now $\varepsilon > 0$ be fixed and choose $\delta = \delta(\varepsilon) > 0$ according to Lemma 9.4. Let furthermore S be a δ-net of the unit sphere of ℓ_2^n which can be chosen of cardinality less than $\exp(2n/\delta)$ (Lemma 9.5). Let $t = \delta M$; the preceding inequality implies that

$$\begin{aligned} \mathbb{P}\left\{ \exists \alpha \in S : \left| \left\| \sum_{i=1}^n \alpha_i \frac{X_i}{M} \right\| - 1 \right| > \delta \right\} &\leq (\mathrm{Card}\ S) \exp\left(-\frac{\delta^2 M^2}{2\sigma(X)^2} \right) \\ &\leq \exp\left(\frac{2n}{\delta} - \frac{\delta^2 c^2}{2} \log N \right) \end{aligned}$$

since $M \geq c(\log N)^{1/2}\sigma(X)$. Assuming N is large enough (otherwise there is nothing to prove), choose $n = [\eta \log N]$ for $\eta = \eta(\delta) = \eta(\varepsilon)$ small enough in order the preceding probability is strictly less than 1. It follows that there exists an ω such that for all α in S

$$1 - \delta \leq \left\| \sum_{i=1}^n \alpha_i \frac{X_i(\omega)}{M} \right\| \leq 1 + \delta.$$

Hence, for $x_i = X_i(\omega)/M$, $i \leq n$, we are in the hypotheses of Lemma 9.4 so that the conclusion readily follows.

The second proof is shorter and neater but relies on the somewhat more involved tool of Theorem 3.16 and Corollary 3.21. Indeed, while isoperimetric arguments were used in the first proof, this was only through concentration inequalities which we know, following e.g. (3.2), can actually be established rather easily. The discretization lemmas are not necessary in this second approach.

2^{nd} **proof of Theorem 9.2.** If X is a Gaussian random variable with values in B, denote by $X_1, \dots, X_n$ independent copies of X. We apply Corollary 3.21. Set $\varphi = \inf_{|\alpha|=1} \| \sum_{i=1}^n \alpha_i X_i \|, \Phi = \sup_{|\alpha|=1} \| \sum_{i=1}^n \alpha_i X_i \|$. Clearly, $0 \leq \varphi \leq \Phi$ and, by Corollary 3.21,

$$\mathbb{E}\|X\| - \sqrt{n}\sigma(X) \leq \mathbb{E}\varphi \leq \mathbb{E}\Phi \leq \mathbb{E}\|X\| + \sqrt{n}\sigma(X).$$

It follows that for some ω, if $\mathbb{E}\|X\| > \sqrt{n}\sigma(X)$,

$$\frac{\Phi(\omega)}{\varphi(\omega)} \leq \frac{\mathbb{E}\|X\| + \sqrt{n}\sigma(X)}{\mathbb{E}\|X\| - \sqrt{n}\sigma(X)}.$$

Choose now X in B according to Lemma 9.3 so that $\mathbb{E}\|X\| > c(\log N)^{1/2}\sigma(X)$ for some numerical $c > 0$. Then, if $\varepsilon > 0$ and $n = [\eta(\varepsilon)\log N]$ where $\eta(\varepsilon) > 0$ can be chosen depending on $\varepsilon > 0$ only,

$$\frac{\mathbb{E}\|X\| + \sqrt{n}\sigma(X)}{\mathbb{E}\|X\| - \sqrt{n}\sigma(X)} \leq \frac{c(\log N)^{1/2} + \sqrt{n}}{c(\log N)^{1/2} - \sqrt{n}} \leq 1 + \varepsilon,$$

so that $\Phi(\omega) \leq (1+\varepsilon)\varphi(\omega)$. But now, by definition of φ and Φ, for every α in $\mathbb{R}^n$ with $|\alpha| = 1$,

$$\varphi(\omega) \leq \left\| \sum_{i=1}^{n} \alpha_i X_i(\omega) \right\| \leq \Phi(\omega).$$

Hence, setting $x_i = X_i(\omega)/\varphi(\omega)$, $i \leq n$, for every $|\alpha| = 1$,

$$1 \leq \left\| \sum_{i=1}^{n} \alpha_i x_i \right\| \leq 1 + \varepsilon$$

which means, by homogeneity, that B contains a subspace $(1+\varepsilon)$-isomorphic to ℓ_2^n. The proof is complete.

Let us note that this second proof provides a dependence of $\eta(\varepsilon)$ in function of ε in Theorem 9.2 of the order of $c\varepsilon^2$ where c is numerical. This is best possible. Recently, G. Schechtman [Sch5] has shown how the more classical isoperimetric approach of the first proof can be modified so to yield also this dependence.

According thus to Theorem 9.1, every infinite dimensional Banach contains ℓ_2^n's uniformly. Clearly, this does not extend to ℓ_p^n's for $p \neq 2$ as can be seen from the example of Hilbert space. Related to this question, note that if, for $0 < p \leq 2$, (θ_i) denotes a sequence of independent standard p-stable random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then, when $p = 2$, (θ_i) (which is then simply an orthogaussian sequence) spans ℓ_2 in $L_q = L_q(\Omega, \mathcal{A}, \mathbb{P})$ for all q, whereas for $p < 2$, (θ_i) spans ℓ_p in L_q only for $q < p$.

It is remarkable that, at least in the case $1 \leq p < 2$, the set of p's for which a Banach space B contains ℓ_p^n's uniformly can be characterized through a probabilistic inequality satisfied by the norm of B. This is what we would like to describe now in the rest of this section. The case $p > 2$ will be shortly presented once the notion of cotype has been introduced in the next paragraph.

Let (θ_i) denote as usual a sequence of independent standard p-stable random variables. For $1 \leq p \leq 2$, a Banach space B is said to be of *stable type* p if there is a constant C such that for every finite sequence (x_i) in B

$$\left\| \sum_i \theta_i x_i \right\|_{p,\infty} \leq C \left(\sum_i \|x_i\|^p \right)^{1/p}. \tag{9.2}$$

The integrability properties and moment equivalences of stable random vectors (Chapter 5) tell us that an equivalent definition of the stable type property is obtained when $\|\cdot\|_{p,\infty}$ is replaced by any $\|\cdot\|_r, r < p$. Further, in terms of infinite sequences and using a closed graph argument, B is of stable type p if and only if $\sum_i \theta_i x_i$ converges almost surely whenever $\sum_i \|x_i\|^p < \infty$. In other words, the existence of the spectral measure of the stable variable $\sum_i \theta_i x_i$ determines its convergence. (As we know, cf. (5.13), this is automatic when $0 < p < 1$ and this is why the range of the stable type is $1 \le p \le 2$.) A Banach space B of stable type p is also of stable type p' for every $p' < p$. This is contained in Proposition 9.12 below but we may briefly anticipate one argument in this regard here. We may assume that $p > 1$. Then, as a consequence of the contraction principle (Lemma 4.5), for some C, $r > 0$ and all finite sequences (x_i) in B,

$$\left\|\sum_i \varepsilon_i x_i\right\|_r \le C\left(\sum_i \|x_i\|^p\right)^{1/p}. \tag{9.3}$$

In particular, since $p' < p$,

$$\mathbb{E}\left\|\sum_i \varepsilon_i x_i\right\|^r \le C'\|(x_i)\|_{p',\infty}^r$$

where $C' = C'(r,p,p')$. Let now (θ_i) be a standard p'-stable sequence. Since (θ_i) has the same distribution as $(\varepsilon_i\theta_i)$, this inequality applied conditionally yields

$$\mathbb{E}\left\|\sum_i \theta_i x_i\right\|^r \le C'\mathbb{E}\|(\theta_i x_i)\|_{p',\infty}^r$$

choose $(r < p')$. Using now Lemma 5.8 and the basic fact that $\|\theta_1\|_{p',\infty} < \infty$, B is seen to be of stable type p'.

As a consequence of the preceding, there is some interest to consider the number

$$p(B) = \sup\{p; B \text{ is of stable type } p\}. \tag{9.4}$$

Examples of spaces of stable type p will be given in the next section once the general theory of type and cotype has been developed. Let us however mention an important example at this stage. Let $1 \le q < 2$ and let $L_q = L_q(S,\Sigma,\mu)$ on some measure space (S,Σ,μ). Then L_q is of stable type p for all $p < q$. This can be seen for example from the preceding; indeed, a simple use of Fubini's theorem together with Khintchine's inequalities shows that (9.3) holds with $r = p = q$. It then follows that L_q is of stable type p. Unless finite dimensional, L_q is however *not* of stable type q. Indeed, according to (5.19), the canonical basis of

ℓ_q cannot satisfy the q-stable type inequality (9.2). Since the stable type property clearly only depends on the collection of the finite dimensional subspaces of a given Banach space, we have similarly that a Banach space containing ℓ_p^n's uniformly $(1 \leq p < 2)$ is *not* of stable type p. The following theorem expresses the striking fact that the converse also holds.

Theorem 9.6. Let $1 \leq p < 2$. A Banach space B contains ℓ_p^n's uniformly if and only if B is not of stable type p.

Before turning to its proof, let us first state some important and useful consequences of Theorem 9.6. The first one expresses that the set of p's for which a Banach space is of stable type p is open. The second answers the question addressed at the light of Dvoretzky's theorem for the p's such that $1 \leq p \leq 2$. Recall $p(B)$ of (9.4).

Corollary 9.7. Let $1 \leq p < 2$. If a Banach space is of stable type p, it is also of stable type p_1 for some $p_1 > p$.

Proof. It is rather elementary to check that the set of p's in $[1,2]$ for which a Banach space contains ℓ_p^n's uniformly is a closed subset of $[1,2]$. Therefore its complement is open and Theorem 9.6 allows to conclude.

Corollary 9.8. The set of p's of $[1,2]$ for which an infinite dimensional Banach space B contains ℓ_p^n's uniformly is equal to $[p(B),2]$.

Proof. If $p(B) = 2$, B contains ℓ_2^n's uniformly by Theorem 9.1 and if $p < 2$, B is of stable type p and Theorem 9.6 can be applied. By definition of $p(B)$ and Corollary 9.7, if $p(B) \leq p < 2$, B is not of stable type p and therefore contains ℓ_p^n's uniformly whereas for $p < p(B)$, B is of stable type p and Theorem 9.6 applies again.

We next turn the proof of Theorem 9.6 and, as for Theorem 9.1, will deduce this result from some stronger finite dimensional version. If B is a Banach space and $1 < p < 2$, denote by $ST_p(B)$ the smallest constant C for which (9.2) with the L_1 norm on the left (yielding as we know an equivalent condition) holds.

Theorem 9.9. Let $1 < p < 2$ and $q = p/p-1$ be the conjugate of p. For each $\varepsilon > 0$ there exists $\eta_p(\varepsilon) > 0$ such that every Banach space B of stable type p contains a subspace $(1+\varepsilon)$-isomorphic to ℓ_p^n where $n = [\eta_p(\varepsilon) ST_p(B)^q]$.

This statement clearly contains Theorem 9.6. Indeed, if $1 < p < 2$ and B is not of stable type p, then $ST_p(B) = \infty$ and one can find finite dimensional subspaces of B with corresponding stable type constant arbitrarily large and hence, by Theorem 9.9, B contains ℓ_p^n's uniformly. Since the set of p's of $[1,2]$ for which B contains ℓ_p^n's uniformly is closed, we reach the case $p = 1$ as well. That Banach spaces containing ℓ_p^n's uniformly are not of stable type p has been discussed prior to Theorem 9.6.

Proof of Theorem 9.9. It will follow the pattern of the first proof of Theorem 9.2 and will rely, through the series representation of stable random vectors, on the concentration inequality (6.14) for sums of independent random variables. Let $S = \frac{1}{2} ST_p(B)$; by definition of $ST_p(B)$, there exist non-zero $x_1, \dots, x_N$ in B such that

$$\sum_{i=1}^{N} \|x_i\|^p = 1 \textit{ and } \mathbb{E}\left\|\sum_{i=1}^{N} \theta_i x_i\right\| > S$$

(recall (θ_i) is a standard p-stable sequence). (This is in a sense the step corresponding to Lemma 9.3 in Theorem 9.2 but whereas Lemma 9.3 holds true in any Banach space, the stable type constant enters the question here.) Let Y with values on the unit sphere of B be defined as $Y = \pm x_i / \|x_i\|$ with probability $\|x_i\|^p/2,\ i \leq N$. Let (Y_j) be independent copies of Y. Then, as a consequence of Corollary 5.5, $X = c_p^{-1/p} \sum_{i=1}^{N} \theta_i x_i$ has the same distribution as $\sum_{j=1}^{\infty} \Gamma_j^{-1/p} Y_j$. Consider now

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

and let (X_i) (resp. (Z_i)) denote independent copies of X (resp. Z). Let furthermore $\alpha = (\alpha_1, \dots, \alpha_n)$ in $\mathbb{R}^n$ be such that $\sum_{i=1}^{n} |\alpha_i|^p = 1$. We first claim that inequality (6.14) implies that for every $t > 0$,

$$\mathbb{P}\left\{\left|\left\|\sum_{i=1}^{n} \alpha_i Z_i\right\| - \mathbb{E}\left\|\sum_{i=1}^{n} \alpha_i Z_i\right\|\right| > t\right\} \leq 2\exp(-t^q/C_q). \tag{9.5}$$

Indeed, if we have independent copies $(Y_{j,i})_i$ of the sequence (Y_j),

$$\sum_{i=1}^{n} \alpha_i Z_i = \sum_{j=1}^{\infty} \sum_{i=1}^{n} \alpha_i j^{-1/p} Y_{j,i}$$

which, since the $Y_{j,i}$ are iid and symmetric, has the same distribution as

$$\sum_{k=1}^{\infty} \beta_k Y_k$$

where $(\beta_k)_{k\geq 1}$ is the non-increasing rearrangement of the doubly-indexed collection $\{|\alpha_i|j^{-1/p}\,;\ j\geq 1\,,$ $1\leq i\leq n\}$. It is easily seen, since $\sum_{i=1}^{n}|\alpha_i|^p=1$, that $\beta_k\leq k^{-1/p}$, so that (6.14) applied to the preceding sum of independent random variables indeed yields (9.5).

From (9.5), the idea of the proof is exactly the same as in the proof of Theorem 9.2 and the subspace isomorphic to ℓ_p^n will be found at random. However, while the X_i's are stable random variables and therefore, by the fundamental stability property, for $\sum_{i=1}^{n}|\alpha_i|^p=1$,

$$\mathbb{E}\left\|\sum_{i=1}^{n}\alpha_i X_i\right\|=\mathbb{E}\|X\|>c_p^{-1/p}S, \tag{9.6}$$

there is no more exactly the case for the Z_i's. This would however be needed in view of (9.5). But Z is actually close enough to X so that this stability property can almost be safed for the Z_i's. Indeed, we know from (5.8) that

$$\mathbb{E}\left\|\sum_{j=1}^{\infty}Y_j(\Gamma_j^{-1/p}-j^{-1/p})\right\|\leq\sum_{j=1}^{\infty}\mathbb{E}|\Gamma_j^{-1/p}-j^{-1/p}|=D_p$$

where D_p is a finite constant depending only on p. Hence, by the triangle inequality and Hölder's inequality, for $\sum_{i=1}^{n}|\alpha_i|^p=1$,

$$\left|\mathbb{E}\left\|\sum_{i=1}^{n}\alpha_i X_i\right\|-\mathbb{E}\left\|\sum_{i=1}^{n}\alpha_i Z_i\right\|\right|\leq D_p\sum_{i=1}^{n}|\alpha_i|\leq D_p n^{1/q}.$$

Let now $\delta>0$ and impose, as a first condition on n, that

$$D_p n^{1/q}\leq\delta S/2c_p^{1/p}. \tag{9.7}$$

By (9.6), setting $M=\mathbb{E}\|X\|$, we see that

$$\left|\mathbb{E}\left\|\sum_{i=1}^{n}\alpha_i Z_i\right\|-M\right|\leq\frac{\delta M}{2}.$$

Hence (9.5) for $t=\delta M/2$ yields

$$\begin{aligned}\mathbb{P}\left\{\left|\left\|\sum_{i=1}^{n}\alpha_i Z_i\right\|-M\right|>\delta M\right\}&\leq 2\exp(-\delta^q M^q/2^q C_q)\\&\leq 2\exp(-\delta^q S^q/2^q c_p^{q/p}C_q).\end{aligned}$$

The proof is almost complete. Let R be a δ-net of the unit sphere of ℓ_p^n which can be chosen, according to Lemma 9.5, with a cardinality less than $\exp(2n/\delta)$. Then

$$\mathbb{P}\left\{\forall\alpha\in R;\left|\left\|\sum_{i=1}^n \alpha_i Z_i\right\| - M\right|\le \delta M\right\}\ge 1-2\exp\left(\frac{2n}{\delta}-\frac{\delta^q S^q}{2^q c_p^{q/p} C_q}\right).$$

Given $\varepsilon>0$ choose $\delta=\delta(\varepsilon)>0$ small enough according to Lemma 9.4. Take then $\eta=\eta(\delta)=\eta(\varepsilon)>0$ such that if $n=[\eta ST_p(B)^q]$ ($ST_p(B)$ being assumed large enough otherwise there is nothing to prove), (9.7) holds and the preceding probability is strictly positive. It follows then that one can find an ω such that for every α in $\mathbb{R}^n$

$$(1+\varepsilon)^{-1/2}\left(\sum_{i=1}^n |\alpha_i|^p\right)^{1/p}\le\left\|\sum_{i=1}^n \alpha_i\frac{Z_i(\omega)}{M}\right\|\le(1+\varepsilon)^{1/2}\left(\sum_{i=1}^n |\alpha_i|^p\right)^{1/p}$$

which gives the conclusion. Theorem 9.9 is thus established.

9.2 Type and cotype.

In the preceding section, we discovered how the probabilistic conditions of stable type are related to some geometric properties of Banach spaces. We start in this paragraph a systematic study of related probabilistic conditions named type and cotype (or Rademacher type and cotype).

As usual (ε_i) denotes a Rademacher sequence. Let $1\le p<\infty$. A Banach space B is said to be of *type* p (or *Rademacher type* p) if there is a constant C such that for all finite sequences (x_i) in B,

$$\left\|\sum_i \varepsilon_i x_i\right\|_p \le C\left(\sum_i \|x_i\|^p\right)^{1/p}. \tag{9.8}$$

From the triangle inequality, every Banach space is of type 1. On the other hand, Khintchine's inequalities indicate that the definition makes sense only for $p\le 2$. Note moreover that the Khintchine-Kahane inequalities in the form of moment equivalences of Rademacher series (Theorem 4.7) show that replacing the p-th moment of $\sum_i \varepsilon_i x_i$ by any other moment leads to an equivalent definition. Furthermore, by a closed graph argument, B is of type p if and only if $\sum_i \varepsilon_i x_i$ converges almost surely when $\sum_i \|x_i\|^p<\infty$.

Let $1\le q\le\infty$. A Banach space B is called of *(Rademacher) cotype* q if there is a constant C such that for all finite sequences (x_i) in B

$$\left(\sum_i \|x_i\|^q\right)^{1/q}\le C\left\|\sum_i \varepsilon_i x_i\right\|_q \tag{9.9}$$

$$\left(\sup_i \|x_i\|\le C\left\|\sum_i \varepsilon_i x_i\right\|_1 \quad \text{when } q=\infty\right).$$

By Lévy's inequalities (Proposition 2.7, (2.7)), or actually some easy direct argument based on the triangle inequality, every Banach space is of infinite cotype whereas, by Khintchine's inequalities, the definition of cotype q reduces actually to $q \geq 2$. The same comments as for the type apply: any moment of the Rademacher average in (9.9) leads to an equivalent definition and B is of cotype q if and only if the almost sure convergence of the series $\sum_i \varepsilon_i x_i$ implies $\sum_i \|x_i\|^q < \infty$.

It is clear from the preceding comments that a Banach space of type p (resp. cotype q) is also of type p' for every $p' \leq p$ (resp. of cotype q' for every $q' \geq q$). Thus the "best" possible spaces in terms of the type and cotype conditions are the spaces of type 2 and cotype 2. Hilbert spaces have this property since by *orthogonality*

$$\mathbb{E}\left\|\sum_i \varepsilon_i x_i\right\|^2 = \sum_i \|x_i\|^2.$$

It is a basic result of the theory due to S. Kwapień [Kw1] that the converse (up to isomorphism) also holds.

Theorem 9.10. A Banach space is of type 2 and cotype 2 if only if it is isomorphic to a Hilbert space.

If a Banach space is of type p and cotype q so are all its subspaces. Actually, the type and cotype properties are seen to depend only on the collection of the finite dimensional subspaces. It is not difficult to verify that quotients of a Banach space of type p are also of type p, with the same constant. This is no more true however for the cotype as is clear for example from the fact that every Banach space can be realized as a quotient of L_1 and that (see below) L_1 is of best possible cotype, cotype 2.

To mention examples, let (S, Σ, μ) be a measure space and let $L_p = L_p(S, \Sigma, \mu)$, $1 \leq p \leq \infty$. Fubini's theorem and Khintchine's inequalities can easily be used to determine the type and cotype of the L_p-spaces. Assume that $1 \leq p < \infty$. Then L_p is of type p when $p \leq 2$ and of type 2 for $p \geq 2$. Let us briefly check these assertions. Let (x_i) be a finite sequence in L_p. Using Lemma 4.1, we can write that

$$\mathbb{E}\left\|\sum_i \varepsilon_i x_i\right\|^p = \int_S \mathbb{E}\left|\sum_i \varepsilon_i x_i(s)\right|^p d\mu(s) \leq B_p^p \int_S \left(\sum_i |x_i(s)|^2\right)^{p/2} d\mu(s).$$

If $p \leq 2$,

$$\int_S \left(\sum_i |x_i(s)|^2\right)^{p/2} d\mu(s) \leq \int_S \sum_i |x_i(s)|^p d\mu(s) = \sum_i \|x_i\|^p,$$

whereas, when $p \geq 2$, by the triangle inequality,

$$\left(\int_S \left(\sum_i |x_i(s)|^2\right)^{p/2} d\mu(s)\right)^{2/p} \leq \sum_i \left(\int_S |x_i(s)|^p d\mu(s)\right)^{2/p} = \sum_i \|x_i\|^2.$$

By considering the canonical basis of ℓ_p one can easily show that the preceding result cannot be improved, i.e., if L_p is infinite dimensional, it is of no type $p' > p$. It can be shown similary that L_p is of cotype p for $p \geq 2$ and cotype 2 for $p \leq 2$ (and nothing better). Note that L_1 is of cotype 2 as mentioned previously. We are left with the case $p = \infty$. It is obvious on the canoncial basis that ℓ_1 is of no type $p > 1$ and c_0 (or ℓ_∞) of no cotype $q < \infty$. Since L_∞ contains isometrically any separable Banach space, in particular ℓ_1 and c_0, L_∞ is of type 1 and cotype ∞ and nothing more, and similarly c_0. In the same way, $C(S)$ the space of continuous functions on a compact metric space S with the sup norm has no non-trivial type or cotype.

Using the moment equivalences of vector valued Rademacher averages (Theorem 4.7) instead of Khintchine's inequalities, the preceding examples can easily be generalized. Let B be a Banach space of type p and cotype q. Then, for $1 \leq r \leq \infty, L_r(S, \Sigma, \mu; B)$ is of type $\min(r,p)$ and of cotype $\max(r,q)$.

Let us mention further that the type and cotype properties appear as dual notions. Indeed, if a Banach space B is of type p, its dual space B' is of cotype $q = p/p-1$. To check this, let (x_i') be a finite sequence in B'. For each $\varepsilon > 0$ and each i, let x_i in B, $\|x_i\| = 1$, such that $x_i'(x_i) = \langle x_i', x_i \rangle \geq (1-\varepsilon)\|x_i'\|$ where $\langle .,. \rangle$ is duality. We then have:

$$\begin{aligned}
(1-\varepsilon)\sum_i \|x_i'\|^q &\leq \sum_i \langle x_i', x_i \rangle \|x_i'\|^{q-1} \\
&= \mathbb{E}\left(\sum_{i,j} \varepsilon_i \varepsilon_j \langle x_j', x_i \rangle \|x_i'\|^{q-1}\right) \\
&= \mathbb{E}\left(\left\langle \sum_i \varepsilon_i x_i \|x_i'\|^{q-1}, \sum_j \varepsilon_j x_j' \right\rangle\right).
\end{aligned}$$

Hence, by Hölder's inequality (assuming $p > 1$) and the type p property of B,

$$\begin{aligned}
(1-\varepsilon)\sum_i \|x_i'\|^q &\leq \left\|\sum_i \varepsilon_i x_i \|x_i'\|^{q-1}\right\|_p \left\|\sum_i \varepsilon_i x_i'\right\|_q \\
&\leq C\left(\sum_i \|x_i'\|^{p(q-1)}\right)^{1/p} \left\|\sum_i \varepsilon_i x_i'\right\|_q .
\end{aligned}$$

It follows that B' is of cotype q (with constant C). The converse assertion is not true in general since ℓ_1 is of cotype 2 but ℓ_∞ is of no type $p > 1$. A deep result of G. Pisier [Pi11] implies that the cotype is dual to the type if the Banach space does not contain ℓ_1^n's uniformly (i.e., by Theorem 9.6, if it is of some non-trivial type).

After these preliminaries and examples on the notions of type and cotype, we now examine several questions concerning the replacement in the definitions (9.8) and (9.9) of the Rademacher sequence by some other sequences of random variables. We start with a general elementary result. For reasons that will become clearer in the next section, we only deal with Radon random variables.

Proposition 9.11. Let B be a Banach space of type p and cotype q with type constant C_1 and cotype constant C_2. Then, for every finite sequence (X_i) of independent mean zero Radon random variables in $L_p(B)$ (resp. $L_q(B)$),

$$\mathbb{E}\left\|\sum_i X_i\right\|^p \leq (2C_1)^p \sum_i \mathbb{E}\|X_i\|^p$$

and

$$\mathbb{E}\left\|\sum_i X_i\right\|^q \geq (2C_2)^{-q} \sum_i \mathbb{E}\|X_i\|^q$$

Proof. We show the assertion relative to the type. By Lemma 6.3,

$$\mathbb{E}\left\|\sum_i X_i\right\|^p \leq 2^p \mathbb{E}\left\|\sum_i \varepsilon_i X_i\right\|^p$$

where (ε_i) is a Rademacher sequence independent of (X_i). Applying (9.8) conditionally on the X_i 's then immediately yields the result.

We now investigate the case where the Rademacher sequence (ε_i) is replaced by a p-stable standard sequence (θ_i) $(1 \leq p \leq 2)$. This will in particular clarify the close relationship between the Rademacher type (9.8) and the stable type discussed in the preceding section. Let $1 \leq p \leq 2$. Recall a Banach space B is said to be of stable type p if there is a constant C such that for all finite sequences (x_i) in B,

$$\left\|\sum_i \theta_i x_i\right\|_{p,\infty} \leq C\left(\sum_i \|x_i\|^p\right)^{1/p}$$

where (θ_i) is a sequence of independent standard p-stable variables.

Proposition 9.12. Let $1 \leq p < 2$ and let B be a Banach space. Then, we have:

(i) If B is of stable type p, it is of type p.

(ii) If B is of type p, it is of stable type p' for every $p' < p$.

(iii) B is of stable type p if and only if there is a constant C such that for all finite sequences (x_i) in B,

$$\mathbb{E}\left\|\sum_i \varepsilon_i x_i\right\| \leq C\|(x_i)\|_{p,\infty}.$$

Proof. Both (i) and (ii) follow from the more difficult claim (iii) but however can be given simple proofs. Indeed, concerning (i), we may assume $p > 1$ so that $\mathbb{E}|\theta_i| < \infty$ and (i) follows from Lemma 4.5. For (ii), let $1 < p' < p$ and (θ_i) denote a standard p'-stable sequence. Recall that since $p > p'$,

$$\left(\sum_i |\alpha_i|^p\right)^{1/p} \leq \left(\frac{p}{p-p'}\right)^{1/p} \|(\alpha_i)\|_{p',\infty}.$$

Applying conditionally the type p inequality and the preceding, for (x_i) a finite sequence in B and some constant C not necessarily the same at each line,

$$\begin{aligned}\mathbb{E}\left\|\sum_i \theta_i x_i\right\| = \mathbb{E}\left\|\sum_i \varepsilon_i \theta_i x_i\right\| &\leq C\mathbb{E}\left(\left(\sum_i |\theta_i|^p \|x_i\|^p\right)^{1/p}\right) \\ &\leq C\mathbb{E}(\|(|\theta_i|\|x_i\|)\|_{p',\infty}).\end{aligned}$$

We then conclude by Lemma 5.8 since $p' > 1$ and $\|\theta_i\|_{p',\infty} < \infty$.

The "if" part of (iii) reproduces the proof of (ii) we just gave (with $p' = p$). Conversely, if B is of stable type p, by Corollary 9.7, B is also of stable type p_1 for some $p_1 > p$, hence of type p_1 by (i). By the comparison between the ℓ_{p_1} and $\ell_{p,\infty}$ norms, the proof is complete.

Note, as a consequence to this proposition, that $p(B)$ introduced in (9.4) is also given by

$$p(B) = \sup\{p;\ B \text{ is of type } p\}.$$

Further, we know that ℓ_1 is of no type $p > 1$; by Proposition 9.12 and Theorem 9.6, we actually have that a Banach space B is of some type $p > 1$, or equivalently of stable type 1 or of stable type p for some $p > 1$, *if and only if* B does not contain ℓ_1^n's uniformly.

As a consequence of Proposition 9.12, note the following version of Proposition 9.11 for the stable type.

Proposition 9.13. Let $1 \leq p < 2$. A Banach space B is of stable type p if and only if there is a constant C such that for every finite sequence (X_i) of independent symmetric Radon random variable in $L_{p,\infty}(B)$,

$$\left\|\sum_i X_i\right\|_{p,\infty}^p \leq C\sup_{t>0} t^p \sum_i \mathbb{P}\{\|X_i\| > t\},$$

or, equivalently, if and only if,

$$\left\|\sum_i X_i\right\|_{p,\infty}^p \le C\sum_i \|X_i\|_{p,\infty}^p .$$

Proof. The "if" part follows by simply letting $X_i = \theta_i x_i$ in the second inequality. We establish the first inequality (which clearly implies the second) in spaces of stable type p. Assume by homogeneity that $\sup_{t>0} t^p \sum_i \mathbb{P}\{\|X_i\| > t\} \le 1$. Let $t > 0$. Since $t^p\mathbb{P}\{\max_i \|X_i\| > t\} \le 1$,

$$t^p\mathbb{P}\left\{\left\|\sum_i X_i\right\| > t\right\} \le 1 + t^p\mathbb{P}\left\{\left\|\sum_i X_i I_{\{\|X_i\|\le t\}}\right\| > t\right\}.$$

Since B is of stable type p, it is also of type p' for some $p' > p$. Hence by Proposition 9.11, for some C,

$$\begin{aligned} t^p\mathbb{P}\left\{\left\|\sum_i X_i\right\| > t\right\} &\le 1 + t^{p-p'}\mathbb{E}\left\|\sum_i X_i I_{\{\|X_i\|\le t\}}\right\|^{p'} \\ &\le 1 + Ct^{p-p'}\sum_i \mathbb{E}(\|X_i\|^{p'} I_{\{\|X_i\|\le t\}}). \end{aligned}$$

Now

$$\begin{aligned} \sum_i \mathbb{E}(\|X_i\|^{p'} I_{\{\|X_i\|\le t\}}) &\le \int_0^t \sum_i \mathbb{P}\{\|X_i\| > s\} ds^{p'} \\ &\le \int_0^t \frac{ds^{p'}}{s^p} = \frac{p'}{p'-p}\, t^{p'-p} \end{aligned}$$

and the conclusion follows.

We next turn to the case of a 2-stable standard sequence, that is an orthogaussian sequence (g_i), which will lead us by the same way to some questions analogous to the preceding ones in the context of the cotype. We first complete the case of type and orthogaussian sequences which is the simplest overall. Indeed, if B is of type p, by Proposition 9.11,

$$\mathbb{E}\left\|\sum_i g_i x_i\right\|^p \le C\sum_i \|x_i\|^p$$

for some constant C. Conversely, since Gaussian averages always dominate Rademacher averages ((4.8)), such an inequality implies that B must be of type p. In particular, stable type 2 and Rademacher type 2 are equivalent notions.

We have seen however in a discussion after Lemma 4.5 that conversely, Rademacher averages do not always dominate the corresponding Gaussian ones, in particular in ℓ_∞. That is, if in a Banach space B, for some constant C and all finite sequences (x_i),

$$\sum_i \|x_i\|^q \leq c\mathbb{E}\left\|\sum_i g_i x_i\right\|^q , \tag{9.10}$$

this inequality does not readily imply that B is of cotype q. This is however true and we now would like to describe some of the deep steps leading to this conclusion, mainly without proofs. The next proposition already covers various applications. Recall that for a (real) random variable ξ, we set

$$\|\xi\|_{q,1} = \int_0^\infty (\mathbb{P}\{|\xi| > t\})^{1/q} dt .$$

Note that if $s > q$, $\|\xi\|_{q,1} \leq q/(s-q)\|\xi\|_s$.

Proposition 9.14. Let $r \geq 1$ and let (ξ_i) be a sequence of independent symmetric real random variables distributed like ξ. If B is a Banach space of cotype $q_0 < \infty$ and if $q = \max(r, q_0)$, there is a constant C such that for all finite sequences (x_i) in B,

$$\left\|\sum_i \xi_i x_i\right\|_r \leq C\|\xi\|_{q,1}\left\|\sum_i \varepsilon_i x_i\right\|_r .$$

Proof. On some (rich enough) probability space, let A be measurable and such that $\mathbb{P}(A) > 0$. Set $\varphi = I_A$; consider independent copies (φ_i) of φ and assume first that $\xi_i = \varepsilon_i\varphi_i$ where (ε_i) is an independent Rademacher sequence. We show that in this case, for some constant C,

$$\left\|\sum_i \xi_i x_i\right\|_r \leq C(\mathbb{P}(A))^{1/q}\left\|\sum_i \varepsilon_i x_i\right\|_r . \tag{9.11}$$

By an easy approximation and the contraction principle, we may and do assume that $\mathbb{P}(A) = 1/N$ for some integer N. Let then $\{A^1, \dots, A^N\}$ be a partition of the probability space into sets of probability $1/N$ with $A^1 = A$. Let further $(\varphi_i^j)_i$ be independent copies of I_{A^j} for all $j \leq N$. Using that $L_r(B)$ is of cotype q, with constant C say, we see that

$$\left(\sum_{j=1}^N \left\|\sum_i \varepsilon_i\varphi_i^j x_i\right\|_r^q\right)^{1/q} \leq C\left\|\sum_{j=1}^N \varepsilon'_j\left(\sum_i \varepsilon_i\varphi_i^j x_i\right)\right\|_r$$

where (ε'_j) is another independent Rademacher sequence. The left hand side of this inequality is just $N^{1/q}\|\sum_i \xi_i x_i\|_r$. Since $|\sum_{j=1}^N \varepsilon'_j\varphi_i^j| = 1$ for every i, by symmetry the right side is $C\|\sum_i \varepsilon_i x_i\|_r$. Thus inequality (9.11) holds.

We can then conclude the proof of the proposition. Note that

$$|\xi_i| = \int_0^\infty I_{\{|\xi_i|>t\}}\,dt.$$

For every $t > 0$, by (9.11),

$$\Big\| \sum_i \varepsilon_i I_{\{|\xi_i|>t\}} x_i \Big\|_r \le C(\mathbb{P}\{|\xi| > t\})^{1/q} \Big\| \sum_i \varepsilon_i x_i \Big\|_r .$$

Therefore, by the triangle inequality,

$$\Big\| \sum_i \varepsilon_i |\xi_i| x_i \Big\|_r \le C \left(\int_0^\infty (\mathbb{P}\{|\xi| > t\})^{1/q} dt \right) \Big\| \sum_i \varepsilon_i x_i \Big\|_r$$

from which the conclusion follows since $(\varepsilon_i|\xi_i|)$ has the same distribution as (ξ_i).

Before turning back to the discussion leading to Proposition 9.14, let us note that this result has a dual version for the type. Namely

Proposition 9.15. Let $r \ge 1$ and let (ξ_i) be a sequence of independent symmetric real random variables distributed like ξ. If B is a Banach space of type p_0 and if $p = \min(r, p_0)$, there is a constant C such that for all finite sequences (x_i) in B,

$$\|\xi\|_{p,\infty} \Big\| \sum_i \varepsilon_i x_i \Big\|_r \le C \Big\| \sum_i \xi_i x_i \Big\|_r .$$

This result thus appears as an improvement, in spaces of some type, of the usual contraction principle in which we get $\|\xi\|_1$ on the left (Lemma 4.5). The proof is entirely similar to the proof of Proposition 9.14; in the last step, simply use the contraction principle to see that for every $t > 0$,

$$\Big\| \sum_i \varepsilon_i I_{\{|\xi_i|>t\}} x_i \Big\|_r \le \frac{1}{t} \Big\| \sum_i \varepsilon_i |\xi_i| x_i \Big\|_r .$$

It should be noticed that Propositions 9.14 and 9.15 are optimal in the sense that they characterize cotype q_0 and type p_0 whenever $r = q_0$, resp. $r = p_0$. Indeed, if $x_1, \ldots, x_N$ are points in a Banach space and if A is such that $\mathbb{P}(A) = 1/N$, let (φ_i) be independent copies of I_A and $\xi_i = \varepsilon_i \varphi_i$. Then clearly

$$\begin{aligned} \mathbb{E} \Big\| \sum_{i=1}^N \xi_i x_i \Big\|^r = \mathbb{E} \Big\| \sum_{i=1}^N \varepsilon_i \varphi_i x_i \Big\|^r &= \sum_{i=1}^n \int_{\{\varphi_j = 1\,;\ \forall i \ne j\,,\ \varphi_i = 0\}} \Big\| \sum_{i=1}^N \varepsilon_i \varphi_i x_i \Big\|^r dP \\ &= \sum_{j=1}^N \|x_j\|^r \left[\frac{1}{N} \left(1 - \frac{1}{N} \right)^{N-1} \right] \end{aligned}$$

which is of the order of $\frac{1}{N}\sum_{i=1}^{N}\|x_i\|^r$. This easily shows the above claim.

Turning back to the question behind Proposition 9.14, we therefore know that if B has some finite cotype, inequality (9.10) will imply that B is of cotype q . Inequality (9.10) actually easily implies that B cannot contain ℓ_∞^n 's uniformly (simply because it cannot hold for the canonical basis of ℓ_∞). A deep result of B. Maurey and G. Pisier [Mau-Pi] shows that this last property characterizes spaces having a non-trivial cotype. The theorem, which is the counterpart for the cotype of the results detailed previously for the type, can be stated as follows. It completes the ℓ_p^n -subspaces question for $p > 2$ although the set of $p > 2$ for which a given Banach space contains ℓ_p^n 's uniformly seems to be rather arbitrary. A more probabilistic proof of this theorem is perhaps still to be found. We refer to [Mau-Pi], [Mi-Sh], [Mi-S].

Theorem 9.16. A Banach space B is of cotype q for some $q < \infty$ if and only if B does not contain ℓ_∞^n 's uniformly. More precisely, if

$$q(B) = \inf\{q; B \text{ is of cotype q}\}$$

and B is infinite dimensional, then B contains ℓ_q^n 's uniformly for $q = q(B)$.

Summarizing in particular some of the (dual) conclusions of Corollary 9.8 and Theorem 9.16, we retain that an (infinite dimensional) Banach space has some non-trivial type (resp. cotype) if and only if it does not contain ℓ_1^n 's (resp. ℓ_∞^n 's) uniformly. Further, combining Theorem 9.16 with Proposition 9.14, if a Banach space B does not contain ℓ_∞^n 's uniformly, there is a constant C such that for all finite sequences (x_i) in B ,

$$\mathbb{E}\left\|\sum_i g_i x_i\right\| \le C\mathbb{E}\left\|\sum_i \varepsilon_i x_i\right\| . \tag{9.12}$$

This is thus an improvement in those spaces over the, in general, best possible inequality (4.9). Conversely thus, if (9.12) holds in a Banach space B , B does not contain ℓ_∞^n 's uniformly. By Proposition 9.14, this characterization easily extends to more general sequences of independent random variables than the orthogaussian sequence.

To conclude this section, we would like to briefly indicate the (easy) extension of the notions of type and cotype to operators between Banach spaces. A linear operator $u : E \to F$ between two Banach spaces E

and F is said to be of (Rademacher) type $p, 1 \le p \le 2$, if there is a constant C such that for all finite sequences (x_i) in E,

$$\left\| \sum_i \varepsilon_i u(x_i) \right\|_p \le C \left(\sum_i \|x_i\|^p \right)^{1/p} .$$

Similarly, u is said to be of cotype q if

$$\left(\sum_i \|x_i\|^q \right)^{1/q} \le C \left\| \sum_i \varepsilon_i u(x_i) \right\|_q .$$

Some of the easy properties of type and cotype clearly extend without modifications to operators. This is in particular trivially the case for Proposition 9.11 which we use freely below. One can also consider operators of stable type but, on the basis for example of Proposition 9.12, one may consider (possibly) different definitions. We can say that $u : E \to F$ is of stable type p, $1 \le p < 2$ if for some constant C and all finite sequences (x_i) in E,

$$\left\| \sum_i \theta_i u(x_i) \right\|_{p,\infty} \le C \left(\sum_i \|x_i\|^p \right)^{1/p} \tag{9.13}$$

where (θ_i) is a standard p-stable sequence. We can also say that it is p-stable if

$$\left\| \sum_i \varepsilon_i u(x_i) \right\|_p \le C \|(x_i)\|_{p,\infty} . \tag{9.14}$$

(9.13) and (9.14) are thus equivalent for the identity operator of a given Banach space but, from the lack of a geometrical characterization analogous to Theorem 9.6, these two definitions are actually different in general. We refer to [P-R1] for a discussion on this difference as well as on related definitions.

9.3. Some probabilistic statements in presence of type and cotype

In this last paragraph, we try to answer some of the questions we started with. We will establish namely tightness and convergence in probability of various sums of independent random variables taking their values in Banach spaces having some type or cotype. As we know, this question is motivated by the strong limit theorems which were reduced in the preceding chapters to weak statements as well as by the central limit theorem investigated in next chapter. We thus revisit now the strong laws of Kolmogorov and of Marcinkiewicz-Zygmund. Type 2 and cotype 2 will also be examined in their relations to pregaussian random variables, as well as spectral measures of stable distributions in spaces of stable type. Finally, but however not directly related to type and cotype, we present some results on almost sure boundedness and convergence of sums of independent random variables in spaces which do not contain isomorphic copies of c_0 .

As announced, since we will be dealing with tightness and weak convergence properties, we only consider in this chapter *Radon* random variables. Equivalently, we may assume we are given a *separable* Banach space.

We start with the SLLN of Kolmogorov for independent random variables. We have seen in Corollary 7.14 that if (X_i) is a sequence of independent random variables with values in a Banach space such that for some $1 \le p \le 2$,

$$\sum_i \frac{\mathbb{E}\|X_i\|^p}{i^p} < \infty, \tag{9.15}$$

then the SLLN holds, i.e. $S_n/n \to 0$ almost surely, if and only if the weak law of large numbers holds, i.e. $S_n/n \to 0$ in probability. In type p spaces, and actually only in them, the series condition (9.15) implies the weak law $S_n/n \to 0$ in probability (provided the variables are centered). This is the conclusion of the next theorem.

Theorem 9.17. Let $1 \le p \le 2$. A Banach space B is of type p if and only if for every sequence (X_i) of independent mean zero (or only symmetric) Radon random variables with values in B , the condition $\sum_i \mathbb{E}\|X_i\|^p/i^p < \infty$ implies the SLLN.

Proof. Assume first that B is of type p . As we have seen, by Corollary 7.14 we need only show that $S_n/n \to 0$ in probability when $\sum_i \mathbb{E}\|X_i\|^p/i^p < \infty$ and the X_i 's have mean zero. But this is rather trivial under the type p condition. Indeed, by Proposition 9.11, for some constant C and all n ,

$$\mathbb{E}\left\|\frac{S_n}{n}\right\|^p \le C\frac{1}{n^p}\sum_{i=1}^n \mathbb{E}\|X_i\|^p.$$

The result follows from the classical Kronecker's lemma (cf. [Sto]). To prove the converse, we assume the SLLN property for random variables of the type $X_i = \varepsilon_i x_i$ where $x_i \in B$ and (ε_i) is a Rademacher sequence. Then, if $\sum_i \|x_i\|^p/i^p < \infty$, we know that $\sum_{i=1}^n \varepsilon_i x_i/n \to 0$ almost surely, and also in $L_1(B)$ (or

$L_r(B)$ for any $r < \infty$) by Theorem 4.7 together with Lemma 4.2. Hence, by the closed graph theorem, for some constant C ,

$$\sup_n \frac{1}{n}\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\| \leq C\left(\sum_i \frac{\|x_i\|^p}{i^p}\right)^{1/p}$$

for every sequence (x_i) in B . Given $y_1,\ldots,y_m$ in B , apply this inequality to the sequence (x_i) defined by $x_i = 0$ if $i \leq m$ or $i > 2m$, $x_{m+1} = y_1$, $x_{m+2} = y_2,\ldots,x_{2m} = y_m$. We get that

$$\mathbb{E}\left\|\sum_{i=1}^m \varepsilon_i y_i\right\| \leq 2C\left(\sum_{i=1}^m \|y_i\|^p\right)^{1/p}$$

and B is therefore of type p . Theorem 9.17 is thus established.

As a consequence of the preceding and Corollary 9.8 we can state

Corollary 9.18. A Banach space B is of type p for some $p > 1$ if and only if every sequence (X_i) of independent symmetric uniformly bounded Radon random variables with values in B satisfies the SLLN.

Proof. Necessity follows from Theorem 9.17. Conversely, it suffices to prove that if B is of no type $p > 1$ there exists a bounded sequence (x_i) such that $(\sum_{i=1}^n \varepsilon_i x_i/n)$ does not converge almost surely or in $L_1(B)$. By Corollary 9.8, together with Proposition 9.12, if B has no non-trivial type, then B contains ℓ_1^n 's uniformly. Hence, for every n , there exist $y_1^n,\ldots,y_{2^n}^n$ in B such that $\|y_i^n\| \leq 1$, $i = 1,\ldots,2^n$, and

$$\mathbb{E}\left\|\sum_{i=1}^{2^n} \varepsilon_i y_i^n\right\| \geq 2^{n-1}.$$

Define now (x_i) by letting $x_i = y_j^n$, $j = i+1-2^n$, $2^n \leq i < 2^{n+1}$. It follows by Jensen's inequality that if $2^n \leq m < 2^{n+1}$,

$$\frac{1}{m}\mathbb{E}\left\|\sum_{i=1}^m \varepsilon_i x_i\right\| \geq \frac{1}{2^{n+1}}\mathbb{E}\left\|\sum_{i=1}^{2^n} \varepsilon_i y_i^n\right\| \geq \frac{1}{4}.$$

This proves Corollary 9.18.

Further results in Chapter 7 can be interpreted similarly under type conditions. We do not detail everything but would like to describe the independent and identically distributed (iid) case following the discussion next to Theorem 7.9. To this aim, it is convenient to record that Proposition 9.11 is still an equivalence when the random variables X_i have the *same* distribution. This is, for the type, the content of the following statement.

Proposition 9.19. Let $1 \leq p \leq 2$ and let B be a Banach space. Assume there is a constant C such that for every finite sequence $(X_1, \ldots, X_N)$ of iid symmetric Radon random variables in $L_p(B)$,

$$\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| \leq CN^{1/p}(\mathbb{E}\|X_1\|^p)^{1/p}.$$

Then B is of type p.

Proof. Let $(\varphi_j)_{j \leq N}$ be real symmetric random variables with disjoint supports such that for every $j = 1, \ldots, N$,

$$\mathbb{P}\{\varphi_j = 1\} = \mathbb{P}\{\varphi_j = -1\} = \frac{1}{2}(1 - \mathbb{P}\{\varphi_j = 0\}) = \frac{1}{2N}.$$

Let $(x_j)_{j \leq N}$ be points in B. Then $X = \sum_{j=1}^{N} \varphi_j x_j$ is such that $\mathbb{E}\|X\|^p = \sum_{j=1}^{N} \|x_j\|^p/N$ so that it is enough to show that if $X_1, \ldots, X_N$ are independent copies of X,

$$\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| \geq c\mathbb{E}\left\|\sum_{j=1}^{N} \varepsilon_j x_j\right\|$$

for some $c > 0$. To this aim, denote by $(\varphi_j^i), i \leq N$, independent copies of (φ_j), assumed to be independent from a Rademacher sequence (ε_i). By symmetry

$$\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| = \mathbb{E}\left\|\sum_{j=1}^{N} \varepsilon_j \left(\sum_{i=1}^{N} \varphi_j^i\right) x_j\right\|$$

and therefore, by Lemma 4.5 for symmetric sequences,

$$\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| \geq \left(\mathbb{E}\left|\sum_{i=1}^{N} \varphi_1^i\right|\right) \mathbb{E}\left\|\sum_{j=1}^{N} \varepsilon_j x_j\right\|.$$

But now, by symmetry and Khintchine's inequalities ((4.3)),

$$\mathbb{E}\left|\sum_{i=1}^{N} \varphi_1^i\right| = \mathbb{E}\left|\sum_{i=1}^{n} \varepsilon_i \varphi_1^i\right| \geq \frac{1}{\sqrt{2}}\mathbb{E}\left(\left(\sum_{i=1}^{N} |\varphi_1^i|^2\right)^{1/2}\right) = \frac{1}{\sqrt{2}}\mathbb{E}\left(\left(\sum_{i=1}^{N} |\varphi_1^i|\right)^{1/2}\right).$$

Since

$$\mathbb{P}\left\{\sum_{i=1}^{N} |\varphi_1^i| = 0\right\} = \left(1 - \frac{1}{N}\right)^N \leq \frac{1}{e},$$

we get that

$$\mathbb{E}\left|\sum_{i=1}^{N} \varphi_1^i\right| \geq \frac{1}{\sqrt{2}}\left(1 - \frac{1}{e}\right) \geq \frac{1}{3},$$

hence the announced claim with $c = 1/3$. The proof of Proposition 9.19 is complete.

There is an analogous statement for the cotype but the proof involves the deeper tool of Theorem 9.16. The main idea of the proof is the so-called "Poissonization" technique.

Proposition 9.20. Let $2 \le q < \infty$ and let B be a Banach space. Assume there is a constant C such that for every finite sequence $(X_1, \ldots, X_N)$ of iid symmetric Radon random variables in $L_q(B)$,

$$N\mathbb{E}\|X_1\|^q \le C\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\|^q .$$

Then B is of cotype q.

Proof. Let $x_1, \ldots, x_N$ be points in B and let X with distribution $(2N)^{-1}\sum_{i=1}^{N}(\delta_{x_i} + \delta_{-x_i})$. Then

$$N\mathbb{E}\|X\|^q = \sum_{i=1}^{N} \|x_i\|^q.$$

Take $(X_1, \ldots, X_N)$ to be independent copies of X and let us consider $\sum_{i=1}^{N} X_i$. Let $(N_i)_{i \le N}$ be independent Poisson random variables with parameter 1, independent of the X_i's. Let further $(X_{i,j})_{1 \le i,j \le N}$ be independent copies of X and set $X_{i,0} = 0$ for each i. Then, as is easy to check on characteristic functionals,

$$\sum_{i=1}^{N}\sum_{j=0}^{N_i} X_{i,j} \text{ has the same distribution as } \sum_{i=1}^{N} \widetilde{N}_i x_i$$

where $\widetilde{N}_i = N_i(1/2) - N_i'(1/2)$ and $N_i(1/2), N_i'(1/2), i \le N$, are independent Poisson random variables with parameter $1/2$. Now, by Jensen's inequality conditionally on (N_i) (cf. (2.5)),

$$\mathbb{E}\left\|\sum_{i=1}^{N}\sum_{j=0}^{N_i} X_{i,j}\right\|^q \ge \mathbb{E}\left\|\sum_{i=1}^{N}\sum_{j=0}^{N_i \wedge 1} X_{i,j}\right\|^q .$$

Further, $\mathbb{P}\{N_i \wedge 1 = 0\} = 1 - \mathbb{P}\{N_i \wedge 1 = 1\} = e^{-1}$. Hence, since $X_{i,0} = 0$,

$$\mathbb{E}\left\|\sum_{i=1}^{N}\sum_{j=0}^{N_i \wedge 1} X_{i,j}\right\|^q = \mathbb{E}\left\|\sum_{i=1}^{N} \delta_i X_i\right\|^q$$

where $\delta_i, i \le N$, are independent, and independent from (X_i), variables with $\mathbb{P}\{\delta_i = 0\} = 1 - \mathbb{P}\{\delta_i = 1\} = e^{-1}$. Again by Jensen's inequality (conditionally on the X_i's),

$$\mathbb{E}\left\|\sum_{i=1}^{N} \delta_i X_i\right\|^q \ge e^{-q}\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\|^q .$$

Summarizing, we have obtained that

$$\sum_{i=1}^{N} \|x_i\|^q \leq C\mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\|^q \leq Ce^q\mathbb{E}\left\|\sum_{i=1}^{N} \widetilde{N}_i x_i\right\| .$$

Now this inequality clearly cannot hold for all finite sequences (x_i) in a Banach space B which contains ℓ_∞^n 's uniformly since it does not hold for the canonical basis of ℓ_∞ . Therefore, by Theorem 9.16, B is of finite cotype and we are then in a position to appy Proposition 9.14 since $\mathbb{E}\widetilde{N}_i^p < \infty$ for all p . The proof is complete.

We now come back after this digression to the main application of Proposition 9.19. We namely investigate the relationship between the type condition and the iid SLLN of Marcinkiewicz-Zygmund. If X is a Radon random variable with values in a Banach space B , (X_i) denotes below a sequence of independent copies of X and, as usual, $S_n = X_1 + \cdots + X_n$, $n \geq 1$. Let $1 \leq p < 2$. In Theorem 7.9, we have seen that $S_n/n^{1/p} \to 0$ almost surely if and only if $\mathbb{E}\|X\|^p < \infty$ and $S_n/n^{1/p} \to 0$ in probability. Moreover, as already discussed thereafter, in type p spaces, $S_n/n^{1/p} \to 0$ in probability when $\mathbb{E}\|X\|^p < \infty$ and $\mathbb{E}X = 0$. We now show that this result is characteristic of the type p property.

Theorem 9.21. Let $1 \leq p < 2$ and let B be a Banach space. The following are equivalent:

(i) B is of type p ;

(ii) for every Radon random variable X with values in B , $S_n/n^{1/p} \to 0$ almost surely if and only if $\mathbb{E}\|X\|^p < \infty$ and $\mathbb{E}X = 0$.

Note of course that since every Banach space is of type 1 we recover Corollary 7.10.

Proof. Let us briefly recall the argument leading to (i) $\Rightarrow$ (ii) already discussed after Theorem 7.9. Under the type assumption and moment conditions $\mathbb{E}\|X\|^p < \infty$ and $\mathbb{E}X = 0$, the sequence $(S_n/n^{1/p})$ was shown to be arbitrarily close to a finite dimensional sequence, and thus tight. Since for every linear functional f , $f(S_n/n^{1/p}) \to 0$ in probability, it follows that $S_n/n^{1/p} \to 0$ in probability and we can conclude the proof of (i) $\Rightarrow$ (ii) by Theorem 7.9. Conversely, let us first show that when $\mathbb{E}\|X\|^p < \infty$, $\mathbb{E}X = 0$ and $S_n/n^{1/p} \to 0$ almost surely (or only in probability), then the sequence $(S_n/n^{1/p})$ is bounded in $L_1(B)$. Since X is centered, by a symmetrization argument it is enough to treat the case of a symmetric variable X . Then, by Lemma 7.2, we already know that

$$\sup_n \frac{1}{n^{1/p}}\mathbb{E}\left\|\sum_{i=1}^{n} X_i I_{\{\|X_i\| \leq n^{1/p}\}}\right\| < \infty.$$

However, it is easily seen by integration by parts that, under $\mathbb{E}\|X\|^p < \infty$,

$$\frac{1}{n^{1/p}}\mathbb{E}\left\|\sum_{i=1}^{n} X_i I_{\{\|X_i\|>n^{1/p}\}}\right\| \leq n^{1-1/p}\mathbb{E}(\|X\| I_{\{\|X\|>n^{1/p}\}})$$

is uniformly bounded in n . The claim follows. (One can also invoke for this result the version of Corollary 10.2 below for $n^{1/p}$.) By the closed graph theorem, there exists therefore a constant C such that for all centered random variables X with values in B

$$\sup_n \frac{1}{n^{1/p}}\mathbb{E}\|S_n\| \leq C(\mathbb{E}\|X\|^p)^{1/p}.$$

We conclude that B is of type p by Proposition 9.19.

The preceding theorem has an analog for the stable type. Let us briefly state this result and sketch its proof.

Theorem 9.22. Let $1 \leq p < 2$ and let B be a Banach space. The following are equivalent:

(i) B is of stable type p ;

(ii) for every symmetric Radon random variable X with values in B, $S_n/n^{1/p} \to 0$ in probability if and only if $\lim_{t\to\infty} t^p\mathbb{P}\{\|X\| > t\} = 0$.

Proof. We have noticed next to Theorem 7.9 that (ii) holds true in finite dimensional spaces. The implication (i) $\Rightarrow$ (ii) is then simply based on Proposition 9.13 and a finite dimensional argument as in Theorem 9.21. That $\lim_{t\to\infty} t^p\mathbb{P}\{\|X\| > t\} = 0$ when $S_n/n^{1/p} \to 0$ in probability is a simple consequence of Lévy's inequality (2.7) and Lemma 2.6; indeed, for each $\varepsilon > 0$ and all n large enough,

$$\frac{1}{4} \geq \mathbb{P}\{\|S_n\| > \varepsilon n^{1/p}\} \geq \frac{1}{2}\mathbb{P}\{\max_{i\leq n}\|X_i\| > \varepsilon n^{1/p}\} \geq \frac{1}{4}n\mathbb{P}\{\|X\| > \varepsilon n^{1/p}\}.$$

The implication (ii) $\Rightarrow$ (i) is obtained as in the last theorem via (iii) of Proposition 9.12 and some care in the closed graph argument.

Some more applications of the type condition in case of the law of the iterated logarithm have been described in Chapter 8 (Corollary 8.8) and we need not recall them here. We now would like to investigate some consequences related to the next chapter on the central limit theorem. They deal with pregaussian variables.

A Radon random variable X with values in a Banach space B such that for every f in B', $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$, is said to be *pregaussian* if there exists a Gaussian Radon variable X in B with the same covariance structure as X, i.e. for all f, g in B', $\mathbb{E}f(X)g(X) = \mathbb{E}f(G)g(G)$ (or just $\mathbb{E}f^2(X) = \mathbb{E}f^2(G)$). Since the distribution of a Gaussian variable is entirely determined by the covariance structure, we denote with some abuse by $G(X)$ a Gaussian variable with the same covariance structure as the pregaussian variable X. The concept of pregaussian variables and their integrability properties are closely related to type 2 and cotype 2. The following easy lemma is useful in this study.

Lemma 9.23. Let X be a pregaussian Radon random variable with values in a Banach space B and with associated Gaussian $G(X)$. Let Y be a Radon random variable in B such that for all f in B', $\mathbb{E}f^2(Y) \leq \mathbb{E}f^2(X)$ (and $\mathbb{E}f(Y) = 0$). Then Y is pregaussian and, for all $p > 0$, $\mathbb{E}\|G(Y)\|^p \leq 2\mathbb{E}\|G(X)\|^p$.

Proof. Since Y is Radon, we may assume that it is constructed on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathcal{A}$ countably generated. In particular, $L_2(\Omega, \mathcal{A}, \mathbb{P})$ is separable and we can find a countable orthonormal basis $(h_i)_{i\geq 1}$ of $L_2(\Omega, \mathcal{A}, \mathbb{P})$. Since Y is Radon (cf. Section 2.1), $\mathbb{E}(h_i Y)$ defines an element of B for every i. For every n, let now $G_n = \sum_{i=1}^{n} g_i \mathbb{E}(h_i Y)$ where (g_i) is an orthogaussian sequence. By Bessel's inequality, for every n and f in B',

$$\mathbb{E}f^2(G_n) \leq \mathbb{E}f^2(Y) \leq \mathbb{E}f^2(X) = \mathbb{E}f^2(G(X)).$$

Using (3.11), the Gaussian sequence (G_n) is seen to be tight in B. Since $\mathbb{E}f^2(G_n) \to \mathbb{E}f^2(Y)$, the limit is unique and (G_n) converges almost surely (Itô-Nisio) and in $L_2(B)$ to a Gaussian Radon random variable $G(Y)$ in B with the same covariance as Y. Since $\mathbb{P}\{\|G(Y)\| > t\} \leq 2\mathbb{P}\{\|G(X)\| > t\}$ for all $t > 0$ also follows from (3.11), the proof is complete.

Note that the factor 2 in Lemma 9.23 and its proof is actually not needed as follows from the deeper result described after (3.11). As a consequence of Lemma 9.23, note also that the sum of two pregaussian variables is also pregaussian. Further, if X is pregaussian and if A is a Borel set such that $XI_{\{X\in A\}}$ has still mean zero, then $XI_{\{X\in A\}}$ is also pregaussian.

To introduce to what follows, let us briefly characterize pregaussian variables in the sequence spaces ℓ_p, $1 \leq p < \infty$. Let $X = (X_k)_{k\geq 1}$ be weakly centered and square integrable with values in ℓ_p. Let $G = (G_k)_{k\geq 1}$ be a Gaussian sequence of real variables with covariance structure determined by $\mathbb{E}G_k G_\ell = \mathbb{E}X_k X_\ell$ for all k, ℓ. G is the natural candidate for $G(X)$. Note that $\mathbb{E}|G_k|^p = c_p(\mathbb{E}|G_k|^2)^{p/2} = c_p(\mathbb{E}|X_k|^2)^{p/2}$ where $c_p = \mathbb{E}|g|^p$, g standard normal. It follows that if

$$\sum_{k=1}^{\infty} (\mathbb{E}|X_k|^2)^{p/2} < \infty, \tag{9.16}$$

by a simple approximation, G is seen to define a Gaussian random variable with values in ℓ_p with the same covariance structure as X and such that

$$\mathbb{E}\|G\|^p = \sum_{k=1}^{\infty} \mathbb{E}|G_k|^p = \sum_{k=1}^{\infty} (\mathbb{E}|X_k|^2)^{p/2} < \infty.$$

Therefore $X = (X_k)_{k\geq 1}$ in $\ell_p, 1 \leq p < \infty$, is pregaussian if and only if (9.16) holds. More generally, it can be shown that X with values in $L_p = L_p(S, \Sigma, \mu)$, $1 \leq p < \infty$, where μ is σ-finite, is pregaussian if and only if (it has weak mean zero and)

$$\int_S (\mathbb{E}|X(s)|^2)^{p/2} d\mu(s) < \infty.$$

The next two propositions are the main observations on pregaussian random variables and their relations to type 2 and cotype 2.

Proposition 9.24. Let B be a Banach space. Then B is of type 2 if and only if every mean zero Radon random variable X with values in B such that $\mathbb{E}\|X\|^2 < \infty$ is pregaussian and we have the inequality

$$\mathbb{E}\|G(X)\|^2 \leq C\mathbb{E}\|X\|^2$$

for some constant C depending only on the type 2 constant of B. The equivalence is still true when $\mathbb{E}\|X\|^2 < \infty$ is replaced by X bounded.

Proposition 9.25. Let B be a Banach space. Then B is of cotype 2 if and only if each pregaussian (Radon) random variable X with values in B satisfies $\mathbb{E}\|X\|^2 < \infty$ and we have the inequality

$$\mathbb{E}\|X\|^2 \leq C\mathbb{E}\|G(X)\|^2$$

for some constant C depending only on the cotype 2 constant of B. The equivalence is still true when $\mathbb{E}\|X\|^2 < \infty$ is replaced by $\mathbb{E}\|X\|^p < \infty$ for some $0 < p \leq 2$.

Proof of Proposition 9.24. We know from Proposition 9.11 that if B is of type 2, for some $C > 0$ and all finite sequences (x_i) in B,

$$\mathbb{E}\left\|\sum_i g_i x_i\right\|^2 \leq C\sum_i \|x_i\|^2.$$

Let now X with $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. There exists an increasing sequence $(\mathcal{A}_N)$ of finite σ-algebras such that if $X^N = \mathbb{E}^{\mathcal{A}_N} X$, $X^N \to X$ almost surely and in $L_2(B)$. Since $\mathcal{A}_N$ is finite, X^N can be written as a finite sum $\sum_i x_i I_{A_i}$ with the A_i disjoint. Then

$$G(X^N) = \sum_i g_i (\mathbb{P}(A_i))^{1/2} x_i \textit{ and } \mathbb{E}\|X^N\|^2 = \sum_i \|x_i\|^2 \mathbb{P}(A_i)$$

so that $\mathbb{E}\|G(X^N)\|^2 \le C\mathbb{E}\|X^N\|^2 \le C\mathbb{E}\|X\|^2$ which thus holds for every N. Since the type 2 inequality also holds for quotient norms (with the same constant), the sequence $(G(X^N))$ of Gaussian random variables is seen to be tight. Since further $\mathbb{E}f^2(G(X^N)) = \mathbb{E}f^2(X^N) \to \mathbb{E}f^2(X)$ for every f in $B', (G(X^N))$ necessarily converges weakly to some Gaussian variable $G(X)$ with the same covariance structure as X. By Skorokhod's theorem, cf. Section 2.1, one obtains that $\mathbb{E}\|G(X)\|^2 \le C\mathbb{E}\|X\|^2$. This establishes the first part of Proposition 9.24.

Conversely, it is sufficient to show that if every centered variable X such that $\|X\|_\infty \le 1$ is pregaussian, then B is type 2. Let (x_i) be a sequence in B such that $\sum_i \|x_i\|^2 = 1$. Consider X with distribution $X = \pm x_i/\|x_i\|$ with probability $\|x_i\|^2/2$. By hypothesis X is pregaussian and $G(X)$ must be $\sum_i g_i x_i$ which therefore defines a convergent series. So is then $\sum_i \varepsilon_i x_i$ and the proof is complete.

Proof of Proposition 9.25. Assume first B is of cotype 2. Then for some constant C and all finite sequences (x_i) in B

$$\sum_i \|x_i\|^2 \le C\mathbb{E}\|\sum_i g_i x_i\|^2. \tag{9.17}$$

(This is a priori *weaker* than the cotype 2 inequality with Rademachers – see (9.10) and Remark 9.26 below). Given X pregaussian with values in B, let $Y = \varepsilon X I_{\{\|X\| \le t\}}$ where $t > 0$ and ε is a Rademacher random variable independent of X. By Lemma 9.23, Y is again pregaussian and $\mathbb{E}\|G(Y)\|^2 \le 2\mathbb{E}\|G(X)\|^2$. Arguing then exactly as in the first part of the proof of Proposition 9.24 by finite dimensional approximation, one obtains that

$$\mathbb{E}\|Y\|^2 \le C\mathbb{E}\|G(Y)\|^2 \le 2C\mathbb{E}\|G(X)\|^2 .$$

Since $t > 0$ is arbitrary, it follows that $\mathbb{E}\|X\|^2 < \infty$ and the inequality of the proposition holds.

Turning to the converse, let us first show that, given $0 < p \le 2$, if every pregaussian variable X in B satisfies $\mathbb{E}\|X\|^p < \infty$, then (9.17) holds. We actually show that if $\sum_i g_i x_i$ converges almost surely, then

$\sum_i \|x_i\|^2 < \infty$. We assume $p < 2$, which is the most difficult case, and set $r = 2/2-p$. Let α_i be positive numbers such that $\sum_i \alpha_i^r = 1$ and define X by:

$$X = \pm \alpha_i^{(1-r)/p} x_i \text{ with probability } \alpha_i^r/2.$$

It is easily seen that $G(X)$ is precisely $\sum_i g_i x_i$ and therefore, by hypothesis,

$$\mathbb{E}\|X\|^p = \sum_i \alpha_i \|x_i\|^p < \infty.$$

Since this holds of each such sequence (α_i), by duality it must be that $\sum_i \|x_i\|^2 < \infty$ which is the announced claim.

To conclude the proof, let us show how (9.17) implies that B is of cotype 2. Recall first we proved before that (9.17) implies

$$\mathbb{E}\|X\|^2 \leq 2C\mathbb{E}\|G(X)\|^2 \tag{9.18}$$

for every pregaussian variable X in B. Let (x_i) be a finite sequence in B. For every $t > 0$,

$$\begin{aligned} \Big\|\sum_i g_i x_i\Big\|_2 &\leq \Big\|\sum_i g_i I_{\{|g_i| \leq t\}} x_i\Big\|_2 + \Big\|\sum_i g_i I_{\{|g_i| > t\}} x_i\Big\|_2 \\ &\leq t\Big\|\sum_i \varepsilon_i x_i\Big\|_2 + \Big\|\sum_i g_i I_{\{|g_i| > t\}} x_i\Big\|_2 \end{aligned}$$

where we have used the contraction principle. If we now apply (9.18) to $X = \sum_i g_i I_{\{|g_i|>t\}} x_i$, we see that

$$\mathbb{E}\Big\|\sum_i g_i I_{\{|g_i|>t\}} x_i\Big\|^2 \leq 2C\mathbb{E}(|g|^2 I_{\{|g|>t\}})\mathbb{E}\Big\|\sum_i g_i x_i\Big\|^2 .$$

Choose now $t > 0$ be small enough in order for $\mathbb{E}(|g|^2 I_{\{|g|>t\}})$ to be less than $(8\mathrm{C})^{-1}$. Together with (9.17), we then obtain that

$$\sum_i \|x_i\|^2 \leq C\mathbb{E}\Big\|\sum_i g_i x_i\Big\|^2 \leq 4t^2 C\mathbb{E}\Big\|\sum_i \varepsilon_i x_i\Big\|^2 .$$

Proposition 9.25 is established.

Remark 9.26. The preceding proof indicates in particular that a Banach space B is of cotype 2 *if and only if* for all finite sequences (x_i) in B

$$\sum_i \|x_i\|^2 \le C\mathbb{E}\left\|\sum_i g_i x_i\right\|^2 .$$

This can also be obtained by the conjunction of Proposition 9.14 and Theorem 9.16. The preceding direct proof in this case is however simpler since it does not use the deep Theorem 9.16.

After pregaussian random variables, we discuss some results on spectral measures of stable distributions in spaces of stable type. If B is a Banach space of stable type p, $1 \le p < 2$, and if (x_i) is a sequence in B such that $\sum_i \|x_i\|^p < \infty$, then the series $\sum_i \theta_i x_i$ converges almost surely and defines a p-stable Radon random variable in B. In other words, if m is the finite discrete measure

$$m = \sum_i \frac{\|x_i\|^p}{2}(\delta_{-x_i/\|x_i\|} + \delta_{+x_i/\|x_i\|}),$$

m is the spectral measure of a p-stable random vector X in B. Moreover

$$\|X\|_{p,\infty} \le C|m|^{1/p} = C\left(\int \|x\|^p dm(x)\right)^{1/p}$$

for some constant C depending only on the stable type p property of B. Recall from Chapter 5 that m symmetrically distributed on the unit sphere of B is unique. Recall further (cf. Corollary 5.5) that if X is a p-stable Radon random variable with values in a Banach space B, there exists a spectral measure m of X such that $\int \|x\|^p dm(x) < \infty$. The parameter $\sigma_p(X)$ of X is defined as $\sigma_p(X) = (\int \|x\|^p dm(x))^{1/p}$ (which is unique among all possible spectral measures) and we always have (cf. (5.13))

$$K_p^{-1}\sigma_p(X) \le \|X\|_{p,\infty}.$$

This inequality is two-sided when $0 < p < 1$ and what we have just seen is that when $1 \le p < 2$ and $\sum_i \|x_i\|^p < \infty$ in a Banach space of stable type $p, X = \sum_i \theta_i x_i$ converges almost surely and $\|X\|_{p,\infty} \le C\sigma_p(X)$. This property actually extends to general measures on type p-stable Banach spaces.

Theorem 9.27. Let $1 \le p < 2$. A Banach space B is of stable type p if and only if every positive finite Radon measure m on B such that $\int \|x\|^p dm(x) < \infty$ is the spectral measure of a p-stable Radon random vector X in B. Furthermore, if this is the case, there exists a constant C such that

$$\|X\|_{p,\infty} \le C\left(\int \|x\|^p dm(x)\right)^{1/p} .$$

Proof. The choice of a discrete measure m as above proves the "if" part of the statement. Suppose now that B is of stable type p. Let m_1 denote the image of the measure $\|x\|^p dm(x)$ by the map $x \to x/\|x\|$. Let further Y_j be independent random variables distributed like $m_1/|m_1|$. The natural candidate for X is given by the series representation

$$c_p |m_1|^{1/p} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j Y_j$$

(cf. Corollary 5.5). In order to show that this series converges, note the following: since B is of stable type p, by Corollary 9.7 together with Proposition 9.12 (i), B is of Rademacher type p_1 for some $p_1 > p$. Therefore

$$\mathbb{E} \left\| \sum_{j=1}^{\infty} j^{-1/p} \varepsilon_j Y_j \right\| \leq C \left(\sum_{j=1}^{\infty} j^{-p_1/p} \right)^{1/p_1} < \infty$$

from which the required convergence easily follows. X thus defines a p-stable random variable with spectral measure m_1 and therefore also m. The inequality of the theorem follows from the same argument (and from some of the elementary material in Chapter 5).

As yet another application, let us briefly mention an alternate approach to p-stable random variables in Banach spaces of stable type p. This approach goes through stochastic integrals and follows a classical construction of S. Kakutani. Let (S, Σ, m) be any measure space. Define a p-stable *random measure* M based on (S, Σ, m) in the following way: $(M(A))_{A \in \Sigma}$ is a collection of real random variables such that for every A, $M(A)$ is p-stable with parameter $m(A)^{1/p}$ and whenever (A_i) are disjoint, the sequence $(M(A_i))$ is independent. For a step function φ of the form $\varphi = \sum_i \alpha_i I_{A_i}, \alpha_i \in \mathbb{R}, A_i \in \Sigma$ disjoint, the stochastic integral $\int \varphi dM$ is well-defined as

$$\int \varphi dM = \sum_i \alpha_i M(A_i).$$

It is a p-stable random variable with parameter $\|\varphi\|_p = (\int |\varphi|^p dm)^{1/p}$. Therefore, by a density argument, it is easy to define the stochastic integral $\int \varphi dM$ for any φ in $L_p(S, \Sigma, m)$.

The question now raises of the possibility of this construction for functions φ taking their values in a Banach space B. As suspected, the class of Banach spaces B of type p-stable is the one in which the preceding stochastic integral $\int \varphi dM$ can be defined when $\int \|\varphi\|^p dm < \infty$. In case φ is the identity map on B, $\int \varphi dM$ is a p-stable random variable in B with spectral measure m so that we recover the conclusion of Theorem 9.27.

Proposition 9.28. Let $1 \le p < 2$. A Banach space B is of stable type p if and only if for any measure space (S, Σ, m) and any p-stable random measure M based on (S, Σ, m) the stochastic integral $\int \varphi dM$ with $\int \|\varphi\|^p dm < \infty$ defines a p-stable Radon random variable with values in B and

$$\left\| \int \varphi dM \right\|_{p,\infty} \le C \left(\int \|\varphi\|^p dm \right)^{1/p} .$$

Proof. Sufficiency is embedded in Theorem 9.27 with the choice for φ of the identity on (B, m). Conversely, if φ is a step function $\sum_i x_i I_{A_i}$ with x_i in B and A_i mutually disjoint,

$$\int \varphi dM = \sum_i x_i M(A_i)$$

which is equal in distribution to $\sum_i m(A_i)^{1/p} \theta_i x_i$. Now, if B is of stable type p,

$$\left\| \int \varphi dM \right\|_{p,\infty} \le C \left(\sum_i m(A_i) \|x_i\|^p \right)^{1/p} = C \left(\int \|\varphi\|^p dm \right)^{1/p} .$$

Hence the map $\varphi \to \int \varphi dM$ can be extended to a bounded operator from $L_p(m; B)$ into $L_{p,\infty}(B)$, and a fortiori into $L_0(B)$. This concludes the proof of Proposition 9.28.

We conclude this chapter with a result on almost sure boundedness and convergence of sums of independent random variables in spaces which do not contain subspaces isomorphic to the space c_0 of all scalar sequences convergent to 0. This result is not directly related to type and cotype since these are local properties in the sense that they only depend on the collection of finite dimensional subspaces of the given space whereas the property discussed here involves *infinite* dimensional subspaces. It is however natural and convenient for further purposes to record this result at this stage.

Let (X_i) be a sequence of independent *real* symmetric random variables such that the sequence (S_n) of the partial sums is almost surely bounded, i.e. $\mathbb{P}\{\sup_n |S_n| < \infty\} = 1$. Then, it is well-known that (S_n) is almost surely convergent. This is actually contained in one of the steps of the proof of Theorem 2.4 but let us emphasize here the argument. Let (ε_i) be a Rademacher sequence independent of (X_i) and recall the partial integration notations $\mathbb{P}_\varepsilon, \mathbb{P}_X, \mathbb{E}_\varepsilon, \mathbb{E}_X$. By symmetry and the assumption, for every $\delta > 0$, there exists a finite number M such that for all n

$$\mathbb{P}\left\{ \left| \sum_{i=1}^n \varepsilon_i X_i \right| \le M \right\} \ge 1 - \delta^2 .$$

Let n be fixed for a moment. By Fubini's theorem, if

$$A = \left\{ \omega ; \mathbb{P}_\varepsilon \left\{ \left| \sum_{i=1}^{n} \varepsilon_i X_i(\omega) \right| \leq M \right\} \geq 1 - \delta \right\} ,$$

then $\mathbb{P}_X(A) \geq 1 - \delta$. Now, if $\delta \leq 1/8$, by Lemma 4.2 and (4.3), for every ω in A,

$$\sum_{i=1}^{n} X_i(\omega)^2 = \mathbb{E}_\varepsilon \left| \sum_{i=1}^{n} \varepsilon_i X_i(\omega) \right|^2 \leq 2\sqrt{2} M .$$

Hence, for all n, $\mathbb{P}\{\sum_{i=1}^{n} X_i(\omega)^2 \leq 2\sqrt{2}M\} \geq 1 - \delta$. It follows that $\sum_i X_i^2 < \infty$ almost surely. Thus, by Fubini's theorem, $\sum_i \varepsilon_i X_i$ converges almost surely and the claim follows.

It can easily be shown that this argument extends, for example, to Hilbert space valued random variables. Actually, this property is satisfied for random variables taking their values in, and only in, Banach spaces which do not contain subspaces isomorphic to c_0. This is the content of the next theorem. A sequence (Y_n) of random variables is almost surely bounded if $\mathbb{P}\{\sup_n \|Y_n\| < \infty\} = 1$.

Theorem 9.29. Let B be a Banach space. The following are equivalent:

(i) B does not contain subspaces isomorphic to c_0;

(ii) for every sequence (X_i) of independent symmetric Radon random variables in B, the almost sure boundedness of the sequence (S_n) of the partial sums implies its convergence;

(iii) for every sequence (x_i) in B, if $(\sum_{i=1}^{n} \varepsilon_i x_i)$ is almost surely bounded, $\sum_i \varepsilon_i x_i$ converges almost surely;

(iv) for every sequence (x_i) in B, if $(\sum_{i=1}^{n} \varepsilon_i x_i)$ is almost surely bounded, $x_i \to 0$.

Proof. The implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) are obvious. Let us show that (iv) $\Rightarrow$ (ii). Let (X_i) in B with $\sup_n \|S_n\| < \infty$ with probability one and let (ε_i) be a Rademacher sequence independent of (X_i). By symmetry and Fubini's theorem, for almost all ω on the probability space supporting the X_i's, $\sup_n \|\sum_{i=1}^{n} \varepsilon_i X_i(\omega)\| < \infty$ almost surely. Hence, by (iv), $X_i(\omega) \to 0$. Similarly, if we take blocks for any strictly increasing sequence (n_k) of integers, $\sum_{n_k < i \leq n_{k+1}} X_i \to 0$ almost surely when $k \to \infty$. Hence $S_{n_{k+1}} - S_{n_k} \to 0$ in probability and by Lévy's inequality (2.6), (S_n) is a Cauchy sequence in probability, and thus convergent. Therefore (ii) holds by Theorem 2.4.

The main point in this proof is the equivalence between (i) and (iv). That (iv) $\Rightarrow$ (i) is clear by the choice of $x_i = e_i$, the canonical basis of c_0. Let us then show the converse implication and let us proceed by contradiction. Let (x_i) be a sequence in B such that $\inf_i \|x_i\| > 0$ and $\mathbb{P}\{\sup_n \|\sum_{i=1}^n \varepsilon_i x_i\| < \infty\} = 1$. We may assume that the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is the canonical product space $\{-1,+1\}^{\mathbb{N}}$ equipped with its natural σ-algebra and product measure. It is easy to see that for every $C \in \mathcal{A}$,

$$\lim_{i\to\infty} \mathbb{P}(C \cap \{\varepsilon_i = -1\}) = \lim_{i\to\infty} \mathbb{P}(C \cap \{\varepsilon_i = +1\}) = \frac{1}{2}\mathbb{P}(C).$$

Let us pick $M < \infty$ so that $\mathbb{P}(A) > 1/2$ where

$$A = \{\sup_n \|\sum_{i=1}^n \varepsilon_i x_i\| \le M\}.$$

By the previous observation, we can define inductively an increasing sequence of integers (n_i) such that for every sequence of signs (a_i) and every k,

$$\mathbb{P}(A \cap \{\varepsilon_{n_1} = a_1\} \cap \cdots \cap \{\varepsilon_{n_k} = a_k\}) > 2^{-k-1}. \tag{9.19}$$

Put $\varepsilon_i' = \varepsilon_i$ if i is one of the n_j's, $\varepsilon_i' = -\varepsilon_i$ if not. The sequences (ε_i) and (ε_i') are equidistributed. Therefore, if

$$A' = \left\{\sup_n \left\|\sum_{i=1}^n \varepsilon_i' x_i\right\| \le M\right\},$$

(9.19) also holds for A' with respect to (ε_i'). Since $\varepsilon_{n_i} = \varepsilon_{n_i}'$ and $\mathbb{P}\{\varepsilon_{n_i} = a_i, i \le k\} = 2^{-k}$, it follows by intersection that there is an ω in $A \cap A'$ such that $\varepsilon_{n_i} = a_i$ for all $i = 1, \ldots, k$. Thus

$$\left\|\sum_{i=1}^k a_i x_{n_i}\right\| = \left\|\frac{1}{2}\left(\sum_{j=1}^{n_k} \varepsilon_j(\omega) x_j + \sum_{j=1}^{n_k} \varepsilon_j'(\omega) x_j\right)\right\| \le M.$$

Since the integer k and the signs $a_1, a_2, \ldots, a_k$ have been fixed arbitrary, this inequality implies that the series $\sum_i x_{n_i}$ is weakly unconditionally convergent, that is $\sum_i |f(x_{n_i})| < \infty$ for every f in B', while $\inf_i \|x_{n_i}\| > 0$. The conclusion is then obtained from the following classical result on basic sequences in Banach spaces.

Lemma 9.30. Let (y_i) be a sequence in a Banach space B such that for every f in B', $\sum_i |f(y_i)| < \infty$, and such that $\inf_i \|y_i\| > 0$. Then, there exists a subsequence (y_{i_k}) of (y_i) which is equivalent to the canonical basis of c_0 in the sense that, for some $C > 0$ and all finite sequences (α_k) of real numbers,

$$C^{-1} \max_k |\alpha_k| \le \left\|\sum_k \alpha_k y_{i_k}\right\| \le C \max_k |\alpha_k|.$$

Proof. As a consequence of the hypotheses, we know in particular that $\inf_i \|y_i\| > 0$ while $y_i \to 0$ weakly. It is then a well-known and important result (cf. e.g. [Li-T1], p.5) that one can extract a subsequence (y_{n_i}) which is basic in the sense that every element in the span of (y_{n_i}) can be written uniquely as $y = \sum_i \alpha_i y_{n_i}$ for some sequence of scalars (α_i). Then, necessarily $\alpha_i \to 0$ since $\inf_i \|y_{n_i}\| > 0$, and, by the closed graph theorem we already have the lower inequality in the statement of the lemma. Since $\sum_i |f(y_i)| < \infty$ for all f in B', by another application of the closed graph theorem, for some C and all f in B', $\sum_i |f(y_i)| \le C\|f\|$. The conclusion is then obvious: for all finite sequences (α_k) of scalars,

$$\left\| \sum_k \alpha_k y_{i_k} \right\| = \sup_{\|f\|\le 1} \left| \sum_k \alpha_k f(y_{i_k}) \right| \le \max_k |\alpha_k| \sup_{\|f\|\le 1} \sum_k |f(y_{i_k})| \le C \max_k |\alpha_k| .$$

This proves the lemma which thus concludes the proof of Theorem 9.29.

Remark 9.31. As a consequence of Theorem 9.29, and more precisely its proof, if (X_i) is a sequence of independent symmetric Radon random variables with values in a Banach space B such that $\sup_n \|S_n\| < \infty$ almost surely but (S_n) does *not* converge, there exist an ω and a subsequence $(i_k) = (i_k(\omega))$ such that $(X_{i_k}(\omega))$ is equivalent to the canonical basis of c_0. Indeed, the various assertions of the theorem are obviously equivalent to say that if $\sup_n \|S_n\| < \infty$, then $X_i \to 0$ almost surely. By Fubini's theorem, there exists an ω on the space supporting the X_i 's such that $\sup_n \| \sum_{i=1}^n \varepsilon_i X_i(\omega)\| < \infty$ but $\inf_i \|X_i(\omega)\| > 0$. The remark then follows from the proof of the implication (i) $\Rightarrow$ (iv).

Notes and references

This chapter reproduces, although not up to the original, parts of the excellent notes [Pi16] by G. Pisier where the interested reader can find more Banach space theory oriented results and in particular quantitative dimensional results. A complete exposition of type and cotype and their relations to local theory of Banach spaces is the book [Mi-S] by V.D. Milman and G. Schechtman. (A more recent "volumic" description of local theory is the book [Pi18] by G. Pisier.) We refer to these works for accurate references. For a more operator theoretical point of view, see [Pie], [Pi15], [TJ2]. The Lecture Notes [Sch3] by L. Schwartz surveys much of the connections between Probability and Geometry in Banach spaces until 1980. See also the exposition [Wo2] by W.A. Woyczyński.

Dvoretzky's theorem was established in 1961 [Dv]. The new proof by V.D. Milman [Mi1] using isoperimetric methods and amplified later on in the paper [F-L-M] considerably influenced the developments of the local theory of Banach spaces. A detailed account on applications of isoperimetric inequalities and concentration of measure phenomena to Geometry of Banach spaces may be found in [Mi-S]. The "concentration dimension" of a Gaussian variable was introduced by G. Pisier [Pi16] (see also [Pi18]) in a Gaussian version of Dvoretzky's theorem and the first proof of Theorem 9.2 is taken from [Pi16]. The second proof is due to Y. Gordon [Gor1], [Gor2]. The Dvoretzky-Rogers lemma appeared in [D-R]; various simple proofs are given in the modern literature, e.g. [F-L-M], [Mi-S], [Pi16], [TJ2]. The fundamental Theorems 9.6 and 9.16 are due to B. Maurey and G. Pisier [Mau-Pi], with an important contribution by J.-L. Krivine [Kr]. The proof of Theorem 9.6 through stable distributions and their representation is due to G. Pisier [Pi12] and was motivated by the results of W.B. Johnson and G. Schechtman on embedding ℓ_p^m into ℓ_1^n [J-S1]. Embeddings via stable variables had already been used in [B-DC-K].

The notions of type and cotype of Banach spaces were explicitely introduced by B. Maurey in the Maurey-Schwartz Seminar 1972/73 (see also [Mau1]) and independently by J. Hoffmann-Jørgensen [HJ1] (cf. [Pi15]). The basic Theorem 9.10 is due to S. Kwapień [Kw1]. Proposition 9.12 (iii) was known since the paper [M-P2] in which Lemma 5.8 is established. Proposition 9.13 comes from [Ro1] (improved in this form in [Led4] and [Pi16]). Comparison of averages with symmetric random variables in Banach spaces not containing ℓ_∞^n's is described in [Mau-Pi]. The proof of Proposition 9.14 is however due to S. Kwapień [Pi16]. We learned its optimality as well as its dual version (Proposition 9.15) from G. Schechtman and J. Zinn (personal communication). Operators of stable type and their possible different definitions (in the context of Probability in Banach spaces) are examined in [P-R1].

The relation of the strong law of large numbers (SLLN) with geometric convexity conditions goes back to the origin of Probability in Banach spaces. In 1962, A. Beck [Be] showed that the SLLN of Corollary 9.18 holds if and only if B is B-convex; a Banach space B is called B-convex if for some $\varepsilon > 0$ and some integer n, for all sequences $(x_i)_{i\leq n}$ in the unit ball of B, one can find a choice of signs $(\varepsilon_i)_{i\leq n}, \varepsilon_i = \pm 1$, with $\|\sum_{i=1}^n \varepsilon_i x_i\| \leq (1-\varepsilon)n$. This property was identified to B not containing ℓ_1^n's in [Gi] and then completely elucitated with the concept of type by G. Pisier [Pi1]. Theorem 9.17 is due to J. Hoffmann-Jørgensen and G. Pisier [HJ1], [Pi1], [HJ-P]. Note the prior contribution [Wo1] in smooth normed spaces. Proposition 9.19 was observed in [Pi3] while Proposition 9.20 is taken from the paper [A-G-M-Z]. More on "Poissonization" may be found there as well as in, e.g., [A-A-G] and [Ar-G2]. A. de Acosta established Theorem 9.21 in [Ac6], partly motivated by the results of M.B. Marcus and W.A. Woyczyński [M-W] on weak laws of large numbers and stable type (Theorem 9.22, but [M-W] goes beyond this statement). More SLLN's are discussed in [Wo2]. See also [Wo3].

Lemma 9.23 is part of the folklore on pregaussian covariances. (9.16) goes back to [Va1]. Propositions 9.24 and 9.25 have been noticed by many authors, e.g. [Pi3], [Ja2], [C-T2], [A-G] (attributed to X. Fernique), [Ar-G2] (through the central limit theorem), etc. Theorem 9.27 has been deduced by several authors from more general statements on Lévy measures and their integrability properties (cf. [Ar-G2], [Li]). The proof through the representation is borrowed from [Pi16] as also the approach through stochastic integrals (see also [M-P3], [Ro2]). Note that while L_p-spaces, $1 \leq p < 2$, are not of stable type, one can still describe spectral measures of p-stable random variables in L_p. The proof goes again through the representation together with arguments similar to those used in (5.19) (that this study actually extends). One can show for example in this way that if m is a (say) probability measure on the unit sphere of ℓ_p and if $Y = (Y_k)$ has law m, then m is the spectral measure of a p-stable random variable in $\ell_p, 1 \leq p < 2$, if and only if

$$\sum_k \mathbb{E}\left(|Y_k|^p \left(1 + \log^+ \frac{|Y_k|}{\|Y_k\|_p}\right)\right) < \infty.$$

This has been known by S. Kwapień and G. Pisier for some time and presented in [C-R-W] and [G-Z1].

That S_n converges almost surely when $\sup_n |S_n| < \infty$ for real random variables is classically deduced from Kolmogorov's converse inequality (and the three series theorem). Theorem 9.29 on Banach spaces not containing c_0 is due to J. Hoffmann-Jørgensen [H-J2] and S. Kwapień [Kw2] (for the main implication (i) $\Rightarrow$ (iv)). Lemma 9.30 goes back to [B-P] (cf. [Li-T1]).

Chapter 10. The central limit theorem

Chapter 10. The central limit theorem

The study of strong limit theorems for sums of independent random variables like the strong law of large numbers or the law of the iterated logarithm in the preceding chapters showed that in Banach spaces these can only be reasonably understood when the corresponding weak property, that is tightness or convergence in probability, is realized. It was shown indeed that under natural moment conditions, the strong statements actually reduce to the corresponding weak ones. On the line or in finite dimensional spaces, these moment conditions usually automatically ensure the weak limiting property. As we pointed out, this is no more the case in general Banach spaces.

There is some point, therefore, to attempt to investigate one typical tightness question in Banach space. One such example is provided by the central limit theorem (in short CLT). The CLT is of course one of the main topic in Probability theory. Here also, its study will indicate the typical problems and difficulties to achieve tightness in Banach spaces. We only investigate here the very classical CLT for sums of independent and identically distributed random variables with normalization $\sqrt{n}$. This framework is actually rich enough already to analyze the main questions. In the first section of this chapter, we present some general facts on the CLT. In the second one, we make use of the type and cotype conditions to extend to certain classes of Banach spaces the classical characterization of the CLT. In the last paragraph, we describe a small ball criterion, which might be of independent interest, as well as an almost sure randomized CLT of possible application in Statistics. Let us mention that this study of the classical CLT will be further developed in the empirical process framework in Chapter 14.

In the whole chapter we deal with Radon random variables, and actually for more convenience, with Borel random variables with values in a *separable* Banach space. Some results actually extend, with only minor modifications, to our usual more general setting, like for example the results of Section 10.3. We leave this to the interested reader. Let thus B denote a separable Banach space. If X is a (Borel) random variable with values in B , we denote by (X_i) a sequence of independent copies of X , and set, as usual, $S_n = X_1 + \cdots + X_n$, $n \geq 1$.

10.1. Some general facts about the central limit theorem

We start with the fundamental definition of central limit property. Let X be a (Borel) random variable with values in a separable Banach space B . X is said to satisfy the *central limit theorem* (CLT) in B if the sequence $(S_n/\sqrt{n})$ converges weakly in B .

Once this definition has been given, one of the main questions is of course to decide when a random variable X satisfies the CLT, and if possible in terms only of the distribution of X. It is well-known that on the line a random variable X satisfies the CLT if and only if $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, and if X satisfies the CLT, the sequence $(S_n/\sqrt{n})$ converges weakly to a normal distribution with mean zero and variance $\mathbb{E}X^2$. The sufficienty part can be established by various methods, for example Lévy's method of characteristic functions or Lindeberg's truncation approach.

We would like to outline here the necessity of $\mathbb{E}X^2 < \infty$ which is particularly clear using the methods developed so far in this book (and somewhat annoying by the usual ones). Let us show more precisely that if the sequence $(S_n/\sqrt{n})$ is stochastically bounded (of course necessary for X to satisfy the CLT), then $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$. Once $\mathbb{E}X^2 < \infty$ has been shown, the centering will be obvious from the strong law of large numbers. Further, replacing X by $X - X'$ where X' is an independent copy of X, we can assume without loss of generality that X is symmetric. For every $n \geq 1$ and $i = 1, \ldots, n$, let

$$u_i = u_i(n) = \frac{X_i}{\sqrt{n}} I_{\{|X_i| \leq \sqrt{n}\}} .$$

By the contraction principle (Lemma 6.5), for any $t > 0$,

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} u_i\right| > t\right\} \leq 2\mathbb{P}\{|S_n/\sqrt{n}| > t\} .$$

By hypothesis, choose $t = t_0$ independent of n such that the right side of this inequality is less than $1/72$. By Proposition 6.8, we get that

$$\mathbb{E}\left|\sum_{i=1}^{n} u_i\right|^2 \leq 18(1 + t_0^2)$$

uniformly therefore in n. Hence, by orthogonality and identical distribution,

$$\mathbb{E}\left(|X|^2 I_{\{|X| \leq \sqrt{n}\}}\right) \leq 18(1 + t_0^2) \tag{10.1}$$

and thus the result when n tends to infinity.

The sufficiency of the conditions $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$ for a random variable X to satisfy the CLT clearly extends to the case where X takes values in a finite dimensional space. It is not too difficult to see that this extends also to Hilbert space. Concerning necessity, we note that the preceding argument allows to conclude that $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$ for any random variable X satisfying the CLT in a Banach space of *cotype 2*. Indeed, the orthogonality property leading to (10.1) is just the cotype 2 inequality. There are however spaces in which the CLT does not necessarily imply that $\mathbb{E}\|X\|^2 < \infty$. Rather than to give an example at this stage, we refer to the forthcoming Proposition 10.8 in which we will actually realize that $\mathbb{E}\|X\|^2 < \infty$ is necessary for X to satisfy the CLT (in and) *only* in cotype 2 spaces.

If X satisfies the CLT in B however, for any linear functional f in B', the scalar random variable $f(X)$ satisfies the CLT with limiting Gaussian with variance $\mathbb{E}f^2(X) < \infty$. Hence, the sequence $(S_n/\sqrt{n})$ actually converges weakly to a Gaussian random variable $G = G(X)$ with the same covariance structure as X. In other words and in the terminology of Chapter 9, a random variable X satisfying the CLT is necessarily *pregaussian*. By Proposition 9.25, we can then recover in particular that a random variable X with values in a Banach space B of cotype 2 satisfying the CLT is such that $\mathbb{E}\|X\|^2 < \infty$.

We mentioned above that in general a random variable satisfying the CLT does not necessarily have a strong second moment. What can then be said on the integrability properties of the norm of X when X satisfy the CLT in an arbitrary Banach space? The next lemma describes the best possible result in this direction. It shows that if the strong second moment is not always available, nevertheless a close property holds. The gap however induces conceptually a rather deep difference.

Lemma 10.1. Let X be a random variable in B satisfying the CLT. Then X has mean zero and

$$\lim_{t\to\infty} t^2 \mathbb{P}\{\|X\| > t\} = 0 .$$

In particular, $\mathbb{E}\|X\|^p < \infty$ for every $0 < p < 2$.

Proof. The mean zero property follows from the second result together with the law of large numbers. Replacing X by $X - X'$ where X' is an independent copy of X and with some trivial desymmetrization argument, we need only consider the case of a symmetric random variable X. Let $0 < \varepsilon \le 1$. For any $t > 0$, if $G = G(X)$ denotes the limiting Gaussian distribution of the sequence $(S_n/\sqrt{n})$,

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \frac{\|S_n\|}{\sqrt{n}} > t \right\} \le \mathbb{P}\{\|G\| \ge t\} .$$

Since G is Gaussian, $\mathbb{E}\|G\|^2 < \infty$ and we can therefore find $t_0 = t_0(\varepsilon)$ (≥ 3) large enough so that $\mathbb{P}\{\|G\| \ge t_0\} \le \varepsilon t_0^{-2}$. (We are thus actually only using that $\lim_{t\to\infty} t^2 \mathbb{P}\{\|G\| > t\} = 0$, the property we would like to establish for X.) Hence, there exists $n_0 = n_0(\varepsilon)$ such that for all $n \ge n_0$

$$\mathbb{P}\left\{ \frac{\|S_n\|}{\sqrt{n}} > t_0 \right\} \le 2\varepsilon t_0^{-2} .$$

By Lévy's inequality (2.7) for symmetric random variables,

$$\mathbb{P}\{\max_{i\le n}\|X_i\| > t_0\sqrt{n}\} \le 4\varepsilon t_0^{-2} \ (\le \frac{1}{2}) .$$

By Lemma 2.6,

$$n\mathbb{P}\{\|X\| > t_0\sqrt{n}\} \le 8\varepsilon t_0^{-2}$$

which therefore holds for all $n \ge n_0$. Let now t be such that $t_0\sqrt{n} \le t < t_0\sqrt{n+1}$ for some $n \ge n_0$. Then

$$t^2\mathbb{P}\{\|X\| > t\} \le t_0^2(n+1)\mathbb{P}\{\|X\| > t_0\sqrt{n}\} \le 16\varepsilon ,$$

that is,

$$\limsup_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} \le 16\varepsilon .$$

Since $\varepsilon > 0$ is arbitrarily small, this proves the lemma.

It is useful to observe that the preceding argument can also be applied to $S_k/\sqrt{k}$ instead of X for *each fixed* k with bounds *uniform in* k. Indeed, if $Y_1,\dots,Y_m$ denote independent copies of $S_k/\sqrt{k}$, then $\sum_{i=1}^{m} Y_i/\sqrt{m}$ has the same distribution as $S_{mk}/\sqrt{mk}$ and the argument remains the same. In this way, we can state the following corollary to the proof of Lemma 10.1.

Corollary 10.2. If X satisfies the CLT, then

$$\lim_{t\to\infty} t^2 \sup_n \mathbb{P}\left\{\frac{\|S_n\|}{\sqrt{n}} > t\right\} = 0 .$$

In particular, for any $0 < p < 2$,

$$\sup_n \mathbb{E}\left\|\frac{S_n}{\sqrt{n}}\right\|^p < \infty .$$

It will be convenient for the sequel to retain a simple quantitive version of this result that immediately folows from the proof of Lemma 10.1 and the preceding observation; namely, if, for some $\varepsilon > 0$,

$$\text{(10.2)} \qquad \sup_n \mathbb{P}\left\{\frac{\|S_n\|}{\sqrt{n}} > \varepsilon\right\} \le \frac{1}{8}, \quad \text{then} \quad \sup_n \mathbb{E}\left\|\frac{S_n}{\sqrt{n}}\right\| \le 20\varepsilon .$$

Let us mention further that the preceding argument leading to Corollary 10.2 extends to more general normalization sequences (a_n) like e.g., $a_n = n^{1/p}$, $0 < p < 2$, or $a_n = (2nLLn)^{1/2}$ since the only property really used is that $a_{mk} \leq C a_m a_k$ for some constant C. Finally, as an alternate approach to the proof of Corollary 10.2, one can use Hoffmann-Jørgensen's inequalities. Combine to this aim Proposition 6.8, Lemma 7.2 and Lemma 10.1.

In the following, we adopt the notation

$$CLT(X) = \sup_n \mathbb{E} \left\| \frac{S_n}{\sqrt{n}} \right\| .$$

By Corollary 10.2, $CLT(X) < \infty$ when X satisfies the CLT. It will be seen below that $CLT(\cdot)$ defines a norm on the linear space of all random variables satisfying the CLT.

At this point, we would like to open a parenthesis and mention a few words about what can be called the bounded form of the CLT. We can say indeed that a random variable X with values in B satisfies the *bounded* CLT if the sequence $(S_n/\sqrt{n})$ is bounded in probability, that is, if for each $\varepsilon > 0$ there is a positive finite number A such that

$$\sup_n \mathbb{P} \left\{ \frac{\|S_n\|}{\sqrt{n}} > A \right\} < \varepsilon .$$

The proofs of Lemma 10.1 and Corollary 10.2 of course carry over to conclude that if X satisfies the bounded CLT, then $\mathbb{E}X = 0$ and

$$\sup_{t>0} \sup_n t^2 \mathbb{P} \left\{ \frac{\|S_n\|}{\sqrt{n}} > t \right\} < \infty .$$

In particular, we also have that $CLT(X) < \infty$ as already indicated by (10.2). As we have seen to start with, on the line, the bounded CLT also implies that $\mathbb{E}X^2 < \infty$, and similarly $\mathbb{E}\|X\|^2 < \infty$ in cotype 2 spaces. In the scalar case therefore, the bounded and true CLT are equivalent, and, as will follow from subsequent results, this equivalence actually extends to Hilbert spaces and even cotype 2 spaces. It is not difficult however to see that this equivalence already fails in L_p-spaces when $p > 2$. Rather than to detail an example at this stage, we refer to Theorem 10.10 below where a characterization of the CLT and bounded CLT in L_p-spaces will clearly indicate the difference between these two properties. It is an open problem to characterize those Banach spaces in which bounded and true CLT are equivalent.

The bounded CLT does not in general imply that X is pregaussian. Since however $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$ for every f in B', there exists by the finite dimensional CLT a Gaussian process $G = (G_f)_{f \in B'_1}$ indexed by the unit ball B'_1 of B' with the same covariance structure as X. Moreover, G is almost surely bounded since by the finite dimensional CLT and convergence of moments, (using Skorokhod's theorem on weak convergence),

$$\mathbb{E} \sup_{f \in B_1'} |G_f| \leq CLT(X) < \infty .$$

In particular, by Sudakov's minoration (Theorem 3.18), the family $\{f(X); f \in B_1'\}$ is relatively compact in L_2. However, this Gaussian process G need not define in general a Radon probability distribution on B. Let us consider indeed the following easy but meaningful example. Denote by $(e_k)_{k \geq 1}$ the canonical basis of c_0 and consider the random variable X with values in c_0 defined by

$$X = \left(\frac{\varepsilon_k e_k}{(2 \log(k+1))^{1/2}} \right) \tag{10.3}$$

where (ε_k) is a Rademacher sequence. Let us sketch that X satisfies the bounded CLT. By independence and identical distribution, for every $A > 0$,

$$\mathbb{P} \left\{ \frac{\|S_n\|}{\sqrt{n}} > A \right\} = 1 - \prod_k \left(1 - \mathbb{P} \left\{ \left| \sum_{i=1}^n \frac{\varepsilon_i}{\sqrt{n}} \right| > A(2 \log(k+1))^{1/2} \right\} \right) .$$

From the subgaussian inequality (4.1),

$$\mathbb{P} \left\{ \left| \sum_{i=1}^n \frac{\varepsilon_i}{\sqrt{n}} \right| > A(2 \log(k+1))^{1/2} \right\} \leq 2 \exp(-A^2 \log(k+1)) .$$

Therefore, if A is large enough independently of n,

$$\mathbb{P} \left\{ \frac{\|S_n\|}{\sqrt{n}} > A \right\} \leq 4 \sum_k \exp(-A^2 \log(k+1))$$

from which it clearly follows that X satisfies the bounded CLT. However, X does not satisfy the CLT in c_0 since X is not pregaussian. Indeed, the natural Gaussian structure with the same covariance as X should be given by

$$G = \left(\frac{g_k e_k}{(2 \log(k+1))^{1/2}} \right)$$

where (g_k) is an orthogaussian sequence. But we know that

$$\limsup_{k \to \infty} \frac{|g_k|}{(2 \log(k+1))^{1/2}} = 1 \quad \text{almost surely.}$$

Therefore G is a bounded Gaussian process but does not define a Gaussian random variable with values in c_0.

The choice of c_0 in this example is not casual as shown by the next result.

Proposition 10.3. Let B be the separable Banach space. In order that every random variable X with values in B that satisfies the bounded CLT is pregaussian, it is necessary and sufficient that B does not contain an isomorphic copy of c_0.

Proof. Example (10.3) provides necessity. Assume B does not contain c_0 and let X be defined on $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying the bounded CLT in B. Since $\mathbb{E}\|X\| < \infty$, there exists a sequence $(\mathcal{A}_N)$ of finite sub-σ-algebras of $\mathcal{A}$ such that if $X^N = \mathbb{E}^{\mathcal{A}_N} X$, the sequence (X^N) converges almost surely and in $L_1(B)$ to X. By Jensen's inequality, $CLT(X^N) \leq CLT(X)$ for every N. Set $Y^N = X^N - X^{N-1}$, $N \geq 1$ ($X^0 = 0$). Since finite valued, for each N, Y^N is pregaussian. Denote by (G_N) independent Gaussian random variables in B such that G_N has the same covariance structure as Y^N. As is easily seen, for every f in B' and every N,

$$\mathbb{E}f^2\left(\sum_{i=1}^{N} G_i\right) = \mathbb{E}f^2(X^N).$$

Hence, by the finite dimensional CLT (and convergence of moments), for every N,

$$\mathbb{E}\left\|\sum_{i=1}^{N} G_i\right\| \leq CLT(X^N) \leq CLT(X).$$

Thus the sequence $\left(\sum_{i=1}^{N} G_i\right)$ is almost surely bounded. Since B does not contain c_0, it converges almost surely (Theorem 9.29) to a Gaussian random variable G which satisfies $\mathbb{E}f^2(G) = \mathbb{E}f^2(X)$ for every f in B'. Hence X is pregaussian.

After this short digression on the bounded CLT we come back to the general study of the CLT. We first recall from Chapter 2 some of the criteria which might be used in order to establish weak convergence of the sequence $(S_n/\sqrt{n})$. From the finite dimensional CLT and Theorem 2.1, a random variable X with values in a separable Banach space B satisfies the CLT if and only if for each $\varepsilon > 0$ one can find a compact set K in B such that

$$\mathbb{P}\left\{\frac{S_n}{\sqrt{n}} \in K\right\} \geq 1 - \varepsilon \quad \text{for every } n$$

(or only every n large enough). Alternatively, and in terms of finite dimensional approximation, a random variable X with values in B such that $\mathbb{E}f(X) = 0$ and $\mathbb{E}f^2(X) < \infty$ for every f in B' (i.e. $f(X)$ satisfies the scalar CLT for every f) satisfies the CLT if and only if for every $\varepsilon > 0$ there is a finite dimensional subspace F of B such that

$$\mathbb{P}\left\{ \| T(\frac{S_n}{\sqrt{n}}) \| > \varepsilon \right\} < \varepsilon \quad \text{for every } n$$

(or only n large enough) where $T = T_F$ denotes the quotient map $B \to B/F$ (cf. (2.4)). By (10.2), equivalently,

$$CLT(T(X)) < \varepsilon \,. \tag{10.4}$$

Note that such a property is realized as soon as there exists, for every $\varepsilon > 0$, a step mean zero random variable Y such that $CLT(X - Y) < \varepsilon$. (The converse actually also holds cf. [Pi3].) Note further from these considerations that X satisfies the CLT as soon as for each $\varepsilon > 0$ there is a random variable Y satisfying the CLT such that $CLT(X - Y) < \varepsilon$; in particular, the linear space of all random variables satisfying the CLT equipped with the norm $CLT(\cdot)$ defines a Banach space.

Before turning to the next section, let us continue with these easy observations and mention some comments about symmetrization. By centering and Jensen's inequality on (10.4) for example, clearly, X satisfies the CLT if and only if $X - X'$ does where X' is an independent copy of X. When trying to establish that a random variable X satisfies the CLT, it will thus basically be enough to deal with a symmetric X. In the same spirit, we also see from Lemma 6.3 that X satisfies the CLT if and only if εX does where ε denotes a Rademacher random variable independent of X. This property is actually one example of a general randomization argument which might be worthwhile to detail at this point. If $0 < p,\ q < \infty$, denote by $L_{p,q}$ the space of all real random variable ξ such that

$$\|\xi\|_{p,q} = \left(q \int_0^\infty (t^p \mathbb{P}\{|\xi| > t\})^{q/p} \frac{dt}{t} \right)^{1/q} < \infty \,.$$

$L_{p,p}$ is just L_p by the usual integration by parts formula and $L_{p,q_1} \subset L_{p,q_2}$ if $q_1 \leq q_2$.

Proposition 10.4. Let X be a random variable with values in B such that $CLT(X) < \infty$ (in particular $\mathbb{E}X = 0$). Let further ξ be a non-zero real random variable in $L_{2,1}$ independent of X. Then

$$\frac{1}{2}\mathbb{E}|\xi| \; CLT(X) \leq CLT(\xi X) \leq 2\|\xi\|_{2,1} \; CLT(X) \,.$$

In particular ξX satisfies the CLT (and $\mathbb{E}X = 0$) if and only if X does.

Proof. The second assertion follows from (10.4) and the inequalities applied to $T(X)$ for quotient maps T. We first prove the right hand side inequality. Assume to begin with that ξ is the indicator function I_A of some set A. Then clearly, by independence and identical distribution, for every n,

$$\mathbb{E}\|\sum_{i=1}^{n}\xi_i X_i\| = \mathbb{E}\|\sum_{i=1}^{S_n(\xi)} X_i\| \leq \mathbb{E}((S_n(\xi))^{1/2})\ CLT(X)$$
$$\leq \sqrt{n}\sqrt{\mathbb{P}(A)}\ CLT(X)$$

where ξ_i are independent copies of ξ and $S_n(\xi) = \xi_1 + \cdots + \xi_n$. Hence, in this case,

$$CLT(\xi X) \leq \sqrt{\mathbb{P}(A)}\ CLT(X)\,.$$

Now classical extremal properties of indicators in the spaces $L_{p,1}$ yield the conclusion. Supposing first $\xi \geq 0$, let, for each $\varepsilon > 0$,

$$\xi^{(\varepsilon)} = \sum_{k=1}^{\infty}\varepsilon k I_{\{\varepsilon(k-1)<\xi\leq\varepsilon k\}} = \sum_{k=1}^{\infty}\varepsilon I_{\{\xi>\varepsilon(k-1)\}}\,.$$

By the triangle inequality and the preceding,

$$CLT(\xi^{(\varepsilon)}X) \leq \sum_{k=1}^{\infty}\varepsilon(\mathbb{P}\{\xi > \varepsilon(k-1)\})^{1/2} CLT(X)$$
$$\leq \|\xi\|_{2,1} CLT(X)\,.$$

Letting ε tend to 0 yield the result in this case. The general one follows by writing $\xi = \xi^+ - \xi^-$.

To establish the reverse inequality, note first that by centering and Lemma 6.3,

$$CLT(\varepsilon\xi X) \leq 2CLT(\xi X)\,,$$

where ε is a Rademacher variable independent of X and ξ. Use then the contraction principle conditionally on X and ξ in the form of Lemma 4.5. The conclusion follows.

Proposition 10.4 in particular applies when ξ is a standard normal variable. Therefore, the normalized sums $S_n/\sqrt{n}$ can be regarded as conditionally Gaussian and several Gaussian tools and techniques can be used. This will be one of the arguments in the empirical process approach to the CLT developed in

Chapter 14. Note further that this Gaussian randomization is similar to the one put forward in the series representation of stable random variables. This relation to stables is not fortuitous and the difficulty in proving tightness in the CLT resembles in some sense to the difficulty in showing existence of a p-stable random vector with a given spectral measure. Let us mention also that while for general sums of independent random variables, Gaussian randomization is heavier than Rademacher randomization (cf. (4.9)), the crucial point in Proposition 10.4 is that we are dealing with independent *identically distributed* random variables. The condition ξ in $L_{2,1}$ has been shown to be best possible in general [L-T1], although L_2 is (necessary and) sufficient in various classes of spaces; this L_2-multiplication property is perhaps related to some geometrical properties of the underlying Banach space. Note finally that the argument of the proof of Proposition 10.4 is not really limited to the normalization $\sqrt{n}$ of the CLT and that similar statements can be obtained in case for example of the iid laws of large numbers with normalization $n^{1/p}$, $0 < p < 2$, and in case of the law of the iterated logarithm.

10.2. Some central limit theorems in certain Banach spaces

In this paragraph, we try to find conditions on the distribution only of a Borel random variable X with values in a separable Banach space B in order it satisfies the CLT. As we know, this is a difficult question in general spaces and at the present time it has a clear cut answer only for special classes of Banach spaces. In this section, we present a sample of these results. We start with some examples and negative facts in order to set up the framework of the study.

Let us first mention that a Gaussian random variable clearly satisfies the CLT. A random variable X with values in a finite dimensional Banach space B satisfies the CLT if and only if $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. As will be shown below, this equivalence extends to infinite dimensional Hilbert spaces, but actually only to them! In general, very bad situations can occur and strong assumptions on the distribution of a random variable X have no reason to ensure that X satisfies the CLT. For example, the random variable in c_0 defined by (10.3) is symmetric and *almost surely bounded* but does not satisfy the CLT. It fails the CLT since it is not pregaussian, but if we go back to Example 7.11, we have a bounded symmetric pregaussian (elementary verification) variable in c_0 which does not satisfy the bounded CLT, hence the CLT. (In [Ma-Pl], there is even an example of a bounded symmetric pregaussian random variable in c_0 satisfying the bounded CLT but failing the CLT.) Even the fact that these examples are constructed in c_0, a space with "bad" geometrical properties (of no non-trivial type or cotype) is not restrictive. There exist indeed spaces of type

$2-\varepsilon$ and cotype $2+\varepsilon$ for every $\varepsilon > 0$ in which one can find bounded pregaussian random variables failing the CLT [Led3].

In spite of these negative examples, some positive results can be obtained. In particular, and as indicated by the last mentioned example, spaces of type 2 and/or cotype 2 play a special role. This is also made clear by Propositions 9.24 and 9.25 connecting type 2 and cotype 2 with pregaussian random variables and some inequalities involving those. The first theorem extends to type 2 spaces the sufficient conditions on the line for the CLT.

Theorem 10.5. Let X be a mean zero random variable such that $\mathbb{E}\|X\|^2 < \infty$ with values in a separable Banach space B of type 2. Then X satisfies the CLT. Conversely, if in a (separable) Banach space B, every random variable X such that $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$ satisfies the CLT, then B must be of type 2.

Proof. The definition of type 2 in the form of Proposition 9.11 immediately implies that for X such that $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$,

$$CLT(X) \leq C(\mathbb{E}\|X\|^2)^{1/2}. \tag{10.5}$$

If, given $\varepsilon > 0$, we choose a mean zero random variable Y with finite range such that $\mathbb{E}\|X-Y\|^2 < \varepsilon^2/C^2$, the preceding inequality applied to $X-Y$ yields $CLT(X-Y) < \varepsilon$ and thus X satisfies the CLT ((10.4)). Conversely, if, in B, each mean zero random variable X with $\mathbb{E}\|X\|^2 < \infty$ satisfies the CLT, such a random variable is necessarily pregaussian. The second part of the theorem therefore simply follows from the corresponding one in Proposition 9.24. Alternatively, one can invoke a closed graph argument to obtain (10.5) and apply then Proposition 9.19.

We would like to mention at this stage for further purposes that Theorem 10.5 easily extends to operators of type 2. Let us state, for later reference, the following corollary to the proof of Theorem 10.5.

Corollary 10.6. Let $u : E \to F$ be an operator of type 2 between two separable Banach spaces E and F. Let X be a random variable with values in E such that $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. Then the random variable $u(X)$ satisfies the CLT in F.

The next statement is the dual result of Theorem 10.5 for cotype 2 spaces.

Theorem 10.7. Let X be a pregaussian random variable with values in a separable cotype 2 Banach space B. Then X satisfies the CLT. Conversely, if in a (separable) Banach space B, any pregaussian random variable satisfies the CLT, then B must be of cotype 2.

Proof. By Proposition 9.25, in a cotype 2 space,

$$\mathbb{E}\|X\|^2 \leq C\mathbb{E}\|G(X)\|^2 < \infty$$

for any pregaussian random variable X with associated Gaussian variable $G(X)$. Now, for each n, $S_n/\sqrt{n}$ is pregaussian and associated to $G(X)$ too. Hence,

$$CLT(X) \leq (C\mathbb{E}\|G(X)\|^2)^{1/2} .$$

Let now $X^N = \mathbb{E}^{\mathcal{A}_N} X$ where $(\mathcal{A}_N)$ is a sequence of finite σ-algebras generating the σ-algebra of X. Then (X^N) converges almost surely and in $L_2(B)$ to X. For each N, $X - X^N$ is still pregaussian and since

$$\mathbb{E}f^2(X - X^N) \leq 2\mathbb{E}f^2(X) = 2\mathbb{E}f^2(G(X))$$

for every f in B', it follows from (3.11) that the sequence $(G(X-X^N))$ is tight (since $G(X)$ is). It can only converge to 0 since, for every f, $\mathbb{E}f^2(X - X^N) \to 0$. By the Gaussian integrability properties (Corollary 3.2), for each $\varepsilon > 0$ one can find an N such that $\mathbb{E}\|G(X - X^N)\|^2 < \varepsilon^2/C$. Thus $CLT(X - X^N) < \varepsilon$ and X satisfies the CLT by the tightness criterion described in Section 10.1. Conversely, recall that a random variable X satisfying the CLT is such that $\mathbb{E}\|X\|^p < \infty$, $p < 2$ (Lemma 10.1). We need then simply recall one assertion of Proposition 9.25. Theorem 10.7 is thus established.

In cotype 2 spaces, random variables satisfying the CLT have a strong second moment (because they are pregaussian, or cf. Section 10.1). This actually only happens in cotype 2 spaces as shown by the next statement. This result complements Theorem 10.7.

Proposition 10.8. If in a (separable) Banach space B every random variable X satisfying the CLT is in $L_2(B)$, then B is of cotype 2.

Proof. Assume B is not of cotype 2. Since (cf. Remark 9.26) the cotype 2 definitions with either Gaussian or Rademacher averages are equivalent, there exists a sequence (x_j) in B such that $\sum_j g_j x_j$

converges almost surely but for which $\sum_j \|x_j\|^2 = \infty$. On some suitable probability space consider

$$X = \sum_{j=1}^{\infty} 2^{j/2} I_{A_j} g_j x_j$$

where A_j are disjoints sets with $\mathbb{P}(A_j) = 2^{-j}$ and independent from the orthogaussian sequence (g_j). Then $\mathbb{E}\|X\|^2 = \sum_j \|x_j\|^2 = \infty$. Let us show however that X satisfies the CLT which leads therefore to a contradiction and proves the proposition. Let (g_j^i) be independent copies of (g_j) and (A_j^i) independent copies of (A_j), all of them assumed to be independent. For every n, let $N(n)$ be the smallest integer such that $2^{N(n)} \geq 2^5 n$. Let

$$\Omega_0 = \bigcup_{j=N(n)+1}^{\infty} \bigcup_{i=1}^{n} A_j^i .$$

Ω_0 only depends on the A_j^i 's and $\mathbb{P}(\Omega_0) \leq 2^{-5}$. Moreover, on the complement of Ω_0,

$$S_n = \sum_{j=1}^{N(n)} 2^{j/2} \left(\sum_{i=1}^{n} I_{A_j^i} g_j^i \right) x_j .$$

We now use the law of large numbers to show that, conditionally on (A_j^i), the Gaussian variables $\sum_{i=1}^{n} I_{A_j^i} g_j^i$ have a variance close to $n 2^{-j}$. For every $t > 0$,

$$\begin{aligned} \mathbb{P}\left\{ \bigcup_{j=1}^{N(n)} \left(\frac{1}{n} \sum_{i=1}^{n} I_{A_j^i} > (t+1) 2^{-j} \right) \right\} &\leq \sum_{j=1}^{N(n)} \mathbb{P}\left\{ \sum_{i=1}^{n} (I_{A_j^i} - \mathbb{P}(A_j^i)) > tn2^{-j} \right\} \\ &\leq \frac{1}{t^2 n} \sum_{j=1}^{N(n)} 2^j \leq \frac{2^7}{t^2} . \end{aligned}$$

Let us take $t = 2^6$. Hence, for every n, there is a set of probability bigger than $1 - 2^{-4}$ such that conditionally on the A_j^i 's, $S_n / \sqrt{n}$ has the same distribution as $\sum_{j=1}^{N(n)} \eta_j x_j$ where η_j are independent normal random variables with variances less than $2^6 + 1 \leq 2^8$. We can then write, for every n and $s > 0$,

$$\begin{aligned} \mathbb{P}\left\{ \frac{\|S_n\|}{\sqrt{n}} > s \right\} &\leq 2^{-4} + \mathbb{P}\left\{ \| \sum_{j=1}^{N(n)} \eta_j x_j \| > s \right\} \\ &\leq 2^{-4} + \frac{2^4}{s} \mathbb{E} \| \sum_{j=1}^{\infty} g_j x_j \| \end{aligned}$$

where we have used the contraction principle in the last step. If we now choose $s = 2^8\mathbb{E}\|\sum_{j=1}^{\infty} g_j x_j\|$ (for example), we deduce from (10.2) that

$$CLT(X) \le 20 \cdot 2^8 \mathbb{E}\|\sum_{j=1}^{\infty} g_j x_j\| .$$

It is now easy to conclude that X satisfies the CLT. The same inequality when a finite number of x_j are 0 allows indeed to organize a finite dimensional approximation of X in the $CLT(\cdot)$-norm. X therefore satisfies the CLT and the proof is thereby completed.

In the spirit of the preceding proof and as a concrete example, let us note in passing that a convergent Rademacher series $X = \sum_i \varepsilon_i x_i$ satisfies the CLT if and only if the corresponding Gaussian series $\sum_i g_i x_i$ converges. Necessity is obvious since $\sum_i g_i x_i$ is a Gaussian variable with the same covariance as X. Concerning sufficiency, if (ε_{ij}) and (g_{ij}) are respectively doubly-indexed Rademacher and orthogaussian sequences, and starting with a finite sequence (x_i), by (4.8), for every n,

$$\begin{aligned}
\mathbb{E}\|\frac{S_n}{\sqrt{n}}\| &= \mathbb{E}\|\sum_i (\sum_{j=1}^{n} \frac{\varepsilon_{ij}}{\sqrt{n}}) x_i\| \\
&\le (\frac{\pi}{2})^{1/2} \mathbb{E}\|\sum (\sum_{j=1}^{n} \frac{g_{ij}}{\sqrt{n}}) x_i\| \\
&= (\frac{\pi}{2})^{1/2} \mathbb{E}\|\sum_i g_i x_i\|
\end{aligned}$$

where the last step follows from the Gaussian rotational invariance. By approximation, X is then easily seen to satisfy the CLT when $\sum_i g_i x_i$ converges almost surely.

If we recall (Theorem 9.10) that a Banach space of type 2 and cotype 2 is isomorphic to a Hilbert space, the conjunction of Theorem 10.5 and Proposition 10.8 yields an isomorphic characterization of Hilbert space by the CLT.

Corollary 10.9. A (separable) Banach space B is isomorphic to a Hilbert space if and only if for every random variable X with values in B the conditions $\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$ are necessary and sufficient for X to satisfy the CLT.

The preceding results are perhaps the most satisfactory ones on the CLT in Banach spaces although they actually concern rather small classes of spaces. While the case of cotype 2 spaces might be consider as completely understood, this is not exactly true for type 2 spaces. Indeed, if Theorem 10.4 indicates that

$\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$ are sufficient for X to satisfy the CLT in a type 2 space, the integrability condition $\mathbb{E}\|X\|^2 < \infty$ need not conversely be necessary (it is only in cotype 2 spaces, therefore Hilbert). As we have seen, in general, when X satisfies the CLT, one only knows that (Lemma 10.1),

$$\lim_{t\to\infty} t^2 \mathbb{P}\{\|X\| > t\} = 0 .$$

It is therefore of some interest to try to understand the spaces in which the best possible necessary conditions, i.e., X pregaussian and $\lim_{t\to\infty} t^2 \mathbb{P}\{\|X\| > t\} = 0$, are also sufficient for a random variable X to ¿satisfy the CLT. One convenient way to investigate this class of spaces is an inequality, similar in some sense to the type and cotype inequalities, but which combines moment assumptions and the pregaussian character. More precisely, let us say that a separable Banach space B satisfies the inequality $\mathrm{Ros}(p)$, $1 \le p < \infty$, if there is a constant C such that for any finite sequence (X_i) of independent pregausssian random variables with values in B with associated Gaussian variables $(G(X_i))$ (which may be assumed to be independent)

$$\mathbb{E}\|\sum_i X_i\|^p \le C\left(\sum_i \mathbb{E}\|X_i\|^p + \mathbb{E}\|\sum_i G(X_i)\|^p\right) . \tag{10.6}$$

This inequality is the vector valued version of an inequality discovered by H. P. Rosenthal (hence the appellation) on the line. That (10.6) holds on the line for any p is easily deduced from, for example, Proposition 6.8 and the observation that

$$\mathbb{E}|\sum_i G(X_i)|^p = C_p(\sum_i \mathbb{E}G(X_i)^2)^{p/2} = C_p(\sum_i \mathbb{E}X_i^2)^{p/2} = C_p(\mathbb{E}|\sum_i X_i|^2)^{p/2} .$$

The same argument together with Proposition 9.25 shows that cotype 2 spaces also satisfy $\mathrm{Ros}(p)$ for every p, $1 \le p < \infty$. It can be shown actually that the spaces of cotype 2 are the only ones with this property (cf. [Led4]).

The main interest of the inequality $\mathrm{Ros}(p)$ in connection with the CLT lies in the following observation.

Theorem 10.10. Let B be a separable Banach space satisfying $\mathrm{Ros}(p)$ for some $p > 2$. Then, a random variable X with values in B satisfies the CLT if and only if it is pregaussian and $\lim_{t\to\infty} t^2 \mathbb{P}\{\|X\| > t\} = 0$.

Proof. The necessity has been discussed in Section 10.1 and holds, as we know, in any space. Turning to sufficiency, our aim is to show that for any symmetric pregaussian X with values in B

$$CLT(X) \le C(\|X\|_{2,\infty} + \mathbb{E}\|G(X)\|) \tag{10.7}$$

for some C. This property easily implies the conclusion. Indeed, given X symmetric, pregaussian and such that $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\|>t\}=0$ and given $\varepsilon>0$, let us first choose t large enough in order that, if $Y=XI_{\{\|X\|\le t\}}$ (which is still pregaussian by Lemma 9.23),

$$\|X-Y\|_{2,\infty}+\mathbb{E}\|G(X-Y)\|<\frac{\varepsilon}{2C}\,.$$

This can be obtained since $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\|>t\}=0$ and $\lim_{t\to\infty}\mathbb{E}f^2(XI_{\{\|X\|>t\}})=0$ for every f in B'. Consider then $(\mathcal{A}_N)$ a sequence of finite σ-algebras generating the σ-algebra of X and set $Y^N=\mathbb{E}^{\mathcal{A}_N}Y$. Then $Y^N\to Y$ almost surely and in $L_2(B)$, and, as usual, $G(Y-Y^N)\to 0$ in $L_2(B)$. Applying (10.7) to $X-Y$ and $Y-Y^N$ for N large enough yields $CLT(X-Y^N)<\varepsilon$ and therefore X satisfies the CLT from this finite dimensional approximation. If X is not symmetric, replace it for example by εX where ε is a Rademacher random variable independent of X.

It therefore suffices indeed to establish (10.7). Assume by homogeneity that $\|X\|_{2,\infty}\le 1$. For each n, we have that,

$$\frac{1}{\sqrt{n}}\mathbb{E}\|\sum_{i=1}^n X_i\|\le\mathbb{E}\|\sum_{i=1}^n u_i\|+\sqrt{n}\mathbb{E}(\|X\|I_{\{\|X\|>\sqrt{n}\}})$$

where $u_i=u_i(n)=n^{-1/2}X_iI_{\{\|X_i\|\le\sqrt{n}\}}$, $i=1,\ldots,n$. Since $\|X\|_{2,\infty}\le 1$, by integration by parts it is easily seen that

$$\sup_n\sqrt{n}\mathbb{E}(\|X\|I_{\{\|X\|>\sqrt{n}\}})\le 2\,.$$

Applying the inequality $\operatorname{Ros}(p)$, $p>2$, to the u_i's, we get that

$$\mathbb{E}\|\sum_{i=1}^n u_i\|^p\le C(n\mathbb{E}\|u_1\|^p+2\mathbb{E}\|G(X)\|^p)$$

since the u_i's are pregaussian and $\mathbb{E}\|G(u_1)\|^p\le 2\mathbb{E}\|G(X)\|^p$ (Lemma 9.23). But now, since $p>2$ and $\|X\|_{2,\infty}\le 1$,

$$\begin{aligned}n\mathbb{E}\|u_1\|^p&\le n^{1-p/2}\int_0^{\sqrt{n}}\mathbb{P}\{\|X\|>t\}\,dt^p\\&\le n^{1-p/2}\int_0^{\sqrt{n}}\frac{dt^p}{t^2}=\frac{p}{p-2}\,.\end{aligned}$$

(10.7) thus follows (recall Gaussian random vectors have all their moments equivalent). Theorem 10.10 is established.

Type 2 spaces of course satisfy $\mathrm{Ros}(2)$ but the important property in Theorem 10.10 is $\mathrm{Ros}(p)$ for $p > 2$. We already noticed that cotype 2 spaces verify $\mathrm{Ros}(p)$ for all p; in particular L_p-spaces with $1 \le p \le 2$. When $2 < p < \infty$, L_p satisfies $\mathrm{Ros}(p)$ for the corresponding p. This follows from the real inequality together with Fubini's theorem. Indeed, if (X_i) are independent pregaussian random variables in $L_p = L_p(S, \Sigma, \mu)$ where (S, Σ, μ) is σ-finite,

$$\begin{aligned}
\mathbb{E}\|\sum_i X_i\|^p &= \int_S \mathbb{E}|\sum_i X_i(s)|^p d\mu(s) \\
&\le \int_S C\left(\sum_i \mathbb{E}|X_i(s)|^p + \mathbb{E}|\sum_i G(X_i)(s)|^p\right) d\mu(s) \\
&= C\left(\sum_i \mathbb{E}\|X_i\|^p + \mathbb{E}\|\sum_i G(X_i)\|^p\right) .
\end{aligned}$$

When $2 < p < \infty$, L_p is of type 2 and enters the setting of Theorem 10.5 but since it satisfies $\mathrm{Ros}(p)$ and $p > 2$ we can also apply the more precise Theorem 10.10. Together with the characterization (9.16) of pregaussian structures, the CLT is therefore completely understood in the L_p-spaces, $1 \le p < \infty$.

One might wonder from the preceding for some more examples of spaces satisfying $\mathrm{Ros}(p)$ for some $p > 2$, especially among the class of type 2 spaces. It can be shown that Banach lattices which are r-convex and s-concave for some $r > 2$ and $s < \infty$ (cf. [Li-T2]) belong to this class. However, already $\ell_2(\ell_r)$, $r > 2$, which is of type 2, verifies $\mathrm{Ros}(p)$ for *no* $p > 2$. Actually, this is true of $\ell_2(B)$ as soon as B is a Banach space which is *not* of cotype 2. To see this, note that if B is not of cotype 2, for every $\varepsilon > 0$, one can find a pregaussian random variable Y in B such that

$$\|Y\| = 1 \quad \text{almost surely and} \quad \mathbb{E}\|G(Y)\|^2 < \varepsilon .$$

Consider then independent copies Y_i of Y and set $X_i = Y_i e_i$ in $\ell_2(B)$ where (e_i) is the canonical basis. Assume then that $\ell_2(B)$ satisfies $\mathrm{Ros}(p)$ for some $p > 2$ and apply this inequality to the sample $(X_i)_{i \le N}$. Since Gaussian moments are all equivalent, we should have that, for some constant C and all N,

$$\mathbb{E}|\sum_{i=1}^N X_i\|^p \le C\left(\sum_{i=1}^N \mathbb{E}\|X_i\|^p + (\mathbb{E}\|\sum_{i=1}^N G(X_i)\|^2)^{p/2}\right) .$$

That is, since $G(X_i) = G(Y_i)e_i$,

$$\begin{aligned}
N^{p/2} &\le C(N + (N\mathbb{E}\|G(Y)\|^2)^{p/2}) \\
&\le C(N + (\varepsilon N)^{p/2}) .
\end{aligned}$$

Hence, if ε is small enough and N tends to infinity, this leads to a contradiction.

Thus finding spaces satisfying $\mathrm{Ros}(p)$ for $p > 2$ seems a difficult task, and so the CLT under the best possible necessary conditions. One interesting problem in this context would be to know whether Theorem 10.10 has some converse; that is, if in a Banach space B, the conditions X pregaussian and $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} = 0$ are sufficient for X to satisfy the CLT, does B satisfy $\mathrm{Ros}(p)$ for some $p > 2$? This could be in analogy with theorems on the laws of large numbers and type of Banach space (cf. Corollary 9.18).

As a remark we would like to briefly come back at this stage to the bounded CLT and show that the bounded CLT and true CLT already differ in L_p for $p > 2$. It is clear from the proof of Theorem 10.10 (cf. (10.7)) that a pregaussian variable in a $\mathrm{Ros}(p)$-space, $p > 2$, such that $\|X\|_{2,\infty} < \infty$ satisfies the bounded CLT. To prove our claim, it is therefore simply enough to construct a pregaussian random variable X in, for example, ℓ_p, $2 < p < \infty$, such that

$$0 < \limsup_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} < \infty . \tag{10.8}$$

To this aim, let N be an integer valued random variable such that $\mathbb{P}\{N = i\} = ci^{-1-2/p}$, $i \geq 1$. Let (ε_i) be a Rademacher sequence independent of N and consider X in ℓ_p given by

$$X = \sum_{i=1}^{\infty} \varepsilon_i I_{\{N^2 \leq i < N^2+N\}} e_i$$

where (e_i) is the canonical basis of ℓ_p. Then $\|X\| = N^{1/p}$ and (10.8) clearly holds. We are left with showing that X is pregaussian and we use (9.16). That is, it suffices to show that

$$\sum_{i=1}^{\infty} (\mathbb{P}\{N \leq i < N^2 + N\})^{p/2} < \infty ;$$

but this is clear since by definition of N, $\mathbb{P}\{N \leq i < N^2 + N\}$ is of the order of $i^{-1/2-1/p}$ when $i \to \infty$, and $p > 2$.

In conclusion to this section, we present some remarks on the relation between the CLT and the LIL (for simplicity we understand here by LIL only the compact law of the iterated logarithm) in Banach spaces. On the line and in finite dimensional spaces, CLT and LIL are of course equivalent, that is, a random variable X satisfies the CLT if and only if it satisfies the LIL, since they are both characterized by the moment conditions

$\mathbb{E}X = 0$ and $\mathbb{E}\|X\|^2 < \infty$. However, the conjunction of Corollary 10.9 and Corollary 8.8 indicates that this equivalence already fails in infinite dimensional Hilbert space where one can find a random variable satisfying the LIL but failing the CLT. This observation together with Dvoretzky's theorem (Theorem 9.1) actually shows that the implication LIL $\Rightarrow$ CLT only holds in finite dimensional spaces. Indeed, if B is an infinite dimensional Banach space in which every random variable satisfying the LIL also satisfies the CLT, by a closed graph argument, for some constant C and every X in B,

$$CLT(X) \leq C\Lambda(X)$$

where we recall that $\Lambda(X) = \limsup_{n\to\infty} \|S_n\|/a_n$ (non random). By Theorem 9.1, the same inequality would hold for all step random variables with values in a Hilbert space, and hence, by approximation, for all random variables. But this is impossible as we have seen. Hence we can state

Theorem 10.11. Let B be a separable Banach space in which every random variable satisfying the LIL also satisfies the CLT. Then B is finite dimensional.

Concerning the implication CLT $\Rightarrow$ LIL, a general statement is available. Indeed, if X satisfies the CLT, then trivially $S_n/a_n \to 0$ in probability, and since X is pregaussian the unit ball of the reproducing kernel Hilbert space associated to X is compact. The characterization of the LIL in Banach spaces (Theorem 8.6) then yields the following theorem.

Theorem 10.12. Let X be a random variable with values in a separable Banach space B satisfying the CLT. Then X satisfies the LIL if and only if $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$.

The moment condition $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$ is of course necessary in this statement since it is not comparable to the tail behavior $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} = 0$ necessary for the CLT. Despite this general satisfactory result, the question of the implication CLT $\Rightarrow$ LIL is not solved for all that. Theorem 10.12 indicates that the spaces in which random variables satisfying the CLT also satisfy the LIL are exactly those in which the CLT implies the integrability property $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$. This is of course the case for cotype 2 spaces but the characterization of the CLT in L_p-spaces shows that L_p with $p > 2$ does not satisfy this property. An argument similar to the one used for Theorem 10.11, but this time with Theorem 9.16 instead of Dvoretzky's theorem, shows then that the spaces satisfying CLT $\Rightarrow$ LIL are necessarily of cotype $2+\varepsilon$ for every $\varepsilon > 0$. But a final characterization has still to be obtained.

10.3. A small ball criterion for the central limit theorem

In this last paragraph, we develop a criterion for the CLT which, while certainly somewhat difficult to verify in practice, involves in its elaboration several interesting arguments and ideas developed throughout this book. The result therefore presents some interest from a theoretical point of view. The idea of its proof can be used further for an almost sure randomized version of the CLT.

Recall that we deal in all this chapter with a separable Banach space B. We have noticed, prior to Theorem 3.3, that for a Gaussian Radon random variable G, each ball centered at the origin has a positive mass for the distribution of G. It therefore follows that if X is a Borel random variable satisfying the CLT in B, for every $\varepsilon > 0$,

$$\liminf_{n\to\infty} \mathbb{P}\left\{ \frac{\|S_n\|}{\sqrt{n}} < \varepsilon \right\} > 0 . \tag{10.9}$$

It turns out that, conversely, if a Gaussian cylindrical measure charges each ball centered at the origin, then it is Radon. Surprisingly, this converse extends to the CLT. Namely, If (10.9) holds for every $\varepsilon > 0$, and if the necessary tail condition $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} = 0$ holds, then X satisfies the CLT. This is theoretical small ball criterion for the CLT. It can thus be stated as follows.

Theorem 10.13. Let X be a random variable with values in a separable Banach space B. Then X satisfies the CLT if and only if the following two properties are satisfied:

(i) $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} = 0$;

(ii) for each $\varepsilon > 0$, $\alpha(\varepsilon) = \liminf_{n\to\infty} \mathbb{P}\{\|S_n/\sqrt{n}\| < \varepsilon\} > 0$.

Before turning to the proof of this result, we would like to mention a few facts, one of which will be of help in the proof. As we have seen, (i) and (ii) are necessary, and best possible. Indeed, the tail condition (i) cannot be suppressed in general, i.e. (ii) does not necessarily imply (i) (cf. [L-T3]). However, and we would like to detail this point, (ii), and for *one* $\varepsilon > 0$ only, already implies the bounded CLT, that is the sequence ¿ $(S_n/\sqrt{n})$ is bounded in probability (and therefore (ii) implies $\sup_{t>0} t^2\mathbb{P}\{\|X\| > t\} < \infty$). This claim is based on the inequality of Proposition 6.4. Replacing X by $X - X'$ where X' is an independent copy of X, it is enough to deal with the symmetrical case. Let $Y_1, \dots, Y_m$ be independent copies of $S_n/\sqrt{n}$. Since $\sum_{i=1}^m Y_i/\sqrt{m}$ has the same distribution as $S_{mn}/\sqrt{mn}$, by Proposition 6.4, for n large enough,

$$\frac{\alpha(\varepsilon)}{2} \le \mathbb{P}\left\{ \|\sum_{i=1}^m Y_i\| < \varepsilon\sqrt{m} \right\} \le \frac{3}{2}\left(1 + \sum_{i=1}^m \mathbb{P}\{\|Y_i\| > \varepsilon\sqrt{m}\}\right)^{-1/2}$$

and thus, for all n large enough and every m,

$$m\mathbb{P}\left\{\frac{\|S_n\|}{\sqrt{n}} > \varepsilon\sqrt{m}\right\} \leq \frac{9}{\alpha(\varepsilon)^2}\,.$$

As announced therefore, $(S_n/\sqrt{n})$ is bounded in probability. In particular, $CLT(X) < \infty$ and while X is not necessarily pregaussian, at least there is a bounded Gaussian process with the same covariance structure as X and the family $\{f(X); f \in B', \|f\| \leq 1\}$ is totally bounded in L_2. All that was noticed in Section 10.1 when we discussed the bounded CLT.

Let us note that the preceding argument based on Proposition 6.4 shows similarly that a random variable X satisfies the CLT if there is a compact set K in B such that

$$\liminf_{n\to\infty} \mathbb{P}\left\{\frac{S_n}{\sqrt{n}} \in K\right\} > 0\,.$$

This improved version of the usual tightness criterion may be useful to understand the intermediate level of Theorem 10.13.

Proof of Theorem 10.13. Replacing as usual X by $X - X'$ we may and do assume that X is symmetric. The sequence (X_i) of independent copies of X has therefore the same distribution as $(\varepsilon_i X_i)$ where (ε_i) is a Rademacher sequence independent of (X_i). Recall we denote by $\mathbb{P}_\varepsilon$, $\mathbb{E}_\varepsilon$ (resp. $\mathbb{P}_X$, $\mathbb{E}_X$) conditional probability and expectation with respect to the sequence (X_i) (resp. (ε_i)). We show, and this is enough, that there is a numerical constant C such that, given $\delta > 0$, one can find a finite dimensional subspace F of B with quotient map $T = T_F : B \to B/F$ such that

$$\limsup_{n\to\infty} \mathbb{P}_X\{\mathbb{E}_\varepsilon\|\sum_{i=1}^n \varepsilon_i \frac{T(X_i)}{\sqrt{n}}\| > C\delta\} \leq \delta\,. \tag{10.10}$$

The proof is based on the isoperimetric inequality for product measures of Theorem 1.4 which was one of the main tools in the study of strong limit theorems and which proves to be also of some interest in weak statements like here the CLT. We use here the full statement of Theorem 1.4 and not only (1.13). The main step in this proof will be to show that if

$$A = \{x = (x_i)_{i\leq n} \in B^n; \mathbb{E}_\varepsilon\|\sum_{i=1}^n \varepsilon_i \frac{T(x_i)}{\sqrt{n}}\| \leq 2\delta\}\,,$$

for each $\delta > 0$ there exists a finite dimensional subspace F such that if $T = T_F$,

$$\mathbb{P}\{(X_i)_{i\leq n} \in A\} \geq \theta(\delta) > 0 \tag{10.11}$$

for all n large enough where $\theta(\delta) > 0$ depends only on $\delta > 0$. Let us show how to conclude when (10.11) is satisfied. For integers q, k, recall $H(A, q, k)$. If $(X_i)_{i \le n} \in H(A, k, q)$, there exist $j \le k$ and $x^1, \ldots, x^q$ in A such that

$$\{1, \ldots, n\} = \{i_1, \ldots, i_j\} \cup I$$

where $I = \bigcup_{\ell=1}^{q} \{i \le n : X_i = x_i^\ell\}$. By monotonicity of Rademacher averages (cf. Remark 6.18),

$$\begin{aligned} \mathbb{E}_\varepsilon \| \sum_{i=1}^{n} \varepsilon_i \frac{T(X_i)}{\sqrt{n}} \| &\le k \max_{i \le n} \frac{\|X_i\|}{\sqrt{n}} + \mathbb{E}_\varepsilon \| \sum_{i \in I} \varepsilon_i \frac{T(X_i)}{\sqrt{n}} \| \\ &\le k \max_{i \le n} \frac{\|X_i\|}{\sqrt{n}} + \sum_{\ell=1}^{q} \mathbb{E}_\varepsilon \| \sum_{i=1}^{n} \varepsilon_i \frac{T(x_i^\ell)}{\sqrt{n}} \| \\ &\le k \max_{i \le n} \frac{\|X_i\|}{\sqrt{n}} + 2q\delta \, . \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{P}_X \{ \mathbb{E}_\varepsilon \| \sum_{i=1}^{n} \varepsilon_i \frac{T(X_i)}{\sqrt{n}} \| > (2q+1)\delta \} \le \mathbb{P}^* \{ (X_i)_{i \le n} \notin H(A, q, k) \} \\ + \mathbb{P} \{ k \max_{i \le n} \frac{\|X_i\|}{\sqrt{n}} > \delta \} \, . \end{aligned}$$

Now, under (10.11), Theorem 1.4 tells us that for some numerical constant K, which we might choose to be an integer for convenience,

$$\mathbb{P}^* \{ (X_i)_{i \le n} \in H(A, q, k) \} \le \left[K \left(\frac{\log(1/\theta(\delta))}{k} + \frac{1}{q} \right) \right]^k \, .$$

Let us choose q to be $2K$, and take then $k = k(\delta)$ large enough depending on $\delta > 0$ only in order for the preceding probability to be less than $\delta/2$. Since, by (i),

$$\lim_{n \to \infty} \mathbb{P} \{ \max_{i \le n} \|X_i\| > \delta \sqrt{n} / k(\delta) \} = 0 \, ,$$

(10.10) is satisfied and the CLT for X will hold.

We have therefore to establish (10.11). We write, for every n,

$$\begin{aligned} \mathbb{P}_X \{ \mathbb{E}_\varepsilon \| \sum_{i=1}^{n} \varepsilon_i \frac{T(X_i)}{\sqrt{n}} \| > 2\delta \} \le \mathbb{P} \{ \| \sum_{i=1}^{n} \varepsilon_i \frac{X_i >}{\sqrt{n}} \| < \delta \} \\ + \mathbb{P} \{ \mathbb{E}_\varepsilon \| \sum_{i=1}^{n} \varepsilon_i \frac{T(X_i)}{\sqrt{n}} \| - \| \sum_{i=1}^{n} \varepsilon_i \frac{T(X_i)}{\sqrt{n}} \| > \delta \} \end{aligned}$$

since $\|T\| \leq 1$. By (ii), (10.11) will hold as soon as

$$\limsup_{n\to\infty} \mathbb{P}\{\mathbb{E}_\varepsilon \|\sum_{i=1}^n \varepsilon_i \frac{T(X_i)}{\sqrt{n}}\| - \|\sum_{i=1}^n \varepsilon_i \frac{T(X_i)}{\sqrt{n}}\| > \delta\} < \alpha(\delta)$$

for some appropriate choice of T . Setting, for every n ,

$$u_i = u_i(n) = \frac{X_i}{\sqrt{n}} I_{\{\|X_i\| \leq c(\delta)\sqrt{n}\}}\,, \quad i = 1, \ldots, n$$

where $c(\delta) > 0$ is to be specified, it is actually enough by (i) to check that

$$\limsup_{n\to\infty} \mathbb{P}\{\mathbb{E}_\varepsilon \|\sum_{i=1}^n \varepsilon_i T(u_i)\| - \|\sum_{i=1}^n \varepsilon_i T(u_i)\| > \delta\} < \alpha(\delta)\,. \tag{10.12}$$

To this aim, we use the concentration properties of Rademacher averages (Theorem 4.7). We have however to switch to expectations instead of medians, but this is easy. Conditionally on (X_i) , denote by M a median of $\|\sum_{i=1}^n \varepsilon_i T(u_i)\|$ and let $\sigma = \sup_{\|f\|\leq 1} \left(\sum_{i=1}^n f^2(T(u_i))\right)^{1/2}$ where the supremum runs here over the unit ball of the dual space of B/F for an F to be chosen. By Theorem 4.7, for every $t > 0$,

$$\mathbb{P}_\varepsilon\{|\|\sum_{i=1}^n \varepsilon_i T(u_i)\| - M| > t\} \leq 4\exp(-\frac{t^2}{8\sigma^2}) \leq 32\frac{\sigma^2}{t^2}\,.$$

In particular, integrating by parts,

$$|\mathbb{E}_\varepsilon \|\sum_{i=1}^n \varepsilon_i T(u_i)\| - M| \leq 12\sigma\,.$$

Hence, if $\sigma \leq \delta/24$,

$$\mathbb{P}_\varepsilon\{\mathbb{E}_\varepsilon \|\sum_{i-1}^n \varepsilon_i T(u_i)\| - \|\sum_{i=1}^n \varepsilon_i T(u_i)\| > \delta\} \leq 128\frac{\sigma^2}{\delta^2}\,.$$

Thus, integrating with respect to $\mathbb{P}_X$,

$$\mathbb{P}\{\mathbb{E}_\varepsilon \|\sum_{i=1}^n \varepsilon_i T(u_i)\| - \|\sum_{i=1}^n \varepsilon_i T(u_i)\| > \delta\} \leq \mathbb{P}\{\sigma > \frac{\delta}{24}\} + \frac{128}{\delta^2}\mathbb{E}\sigma^2 \leq \frac{10^3}{\delta^2}\mathbb{E}\sigma^2\,. \tag{10.13}$$

To prove (10.12), we have thus just simply to show that $\mathbb{E}\sigma^2$ can be made arbitrarily small independently of n for a well chosen large enough subspace F of B . To this aim, recall from the discussion prior to this proof that, under (ii), $CLT(X) < \infty$ and that this implies that $\{f(X); f \in B', \|f\| \leq 1\}$ is relatively compact in L_2. We use Lemma 6.6 (6.5) (and the contraction principle) to see that, for all n,

$$\begin{aligned} \mathbb{E}\sigma^2 &= \mathbb{E} \sup_{\|f\| \leq 1} \sum_{i=1}^n f^2(T(u_i)) \\ &\leq \sup_{\|f\| \leq 1} \mathbb{E} f^2(T(X)) + 8c(\delta) CLT(X) . \end{aligned}$$

Choose then $c(\delta) > 0$ to be less than $\delta^2 \alpha(\delta)/16 \cdot 10^3 CLT(X)$, and choose also $T: B \to B/F$ associated to some large enough finite dimensional subspace F of B in order that $\sup_{\|f\| \leq 1} \mathbb{E} f^2(T(X)) \leq \delta^2 \alpha(\delta)/2 \cdot 10^3$. According to (10.13), we see then that (10.12) is satisfied, which was to prove. Theorem 10.13 is therefore established in this way.

We conclude this chapter with an almost sure randomized version of the CLT. The interest in such a result lies in its proof itself, which is similar in nature to the preceding one, and in possible statistical applications. It might be worthwhile to recall Proposition 10.4 before the statement. As there also, if X is a Banach space valued random variable and ξ a real random variable independent of X, and if (X_i) (resp. (ξ_i)) are independent copies of X (resp. ξ), the sequences (X_i) and (ξ_i) are understood to be independent (constructed on different probability spaces).

Theorem 10.14. Let X be a mean zero random variable with values in a separable Banach space B and let ξ be a real random variable in $L_{2,1}$ independent of X such that $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$. The following are equivalent:

(i) $\mathbb{E}\|X\|^2 < \infty$ and X satisfies the CLT;

(ii) for almost every ω on the probability space supporting the X_i's, the sequence $\left(\sum_{i=1}^n \xi_i X_i(\omega)/\sqrt{n}\right)$ converges in distribution.

In either case, the limit of $\left(\sum_{i=1}^n \xi_i X_i(\omega)/\sqrt{n}\right)$ does not depend on ω and is distributed like $G(X)$, the Gaussian distribution with the same covariance structure as X.

Proof. It is plain that under (ii) the product ξX satisfies the CLT. Hence, since X has mean zero and $\xi \not\equiv 0$, X also satisfies the CLT by Proposition 10.4. To show that $\mathbb{E}\|X\|^2 < \infty$, replacing ξ by $\xi - \xi'$ where ξ' is an independent copy of ξ, we may assume ξ to be symmetric. For almost every ω, $\left(\sum_{i=1}^n \xi_i X_i(\omega)/\sqrt{n}\right)$ is bounded in probability. Hence, by Lévy's inequality (2.7), the same holds for the sequence $(|\xi_n|\|X_n(\omega)\|/\sqrt{n})$. Since ξ is non-zero, it follows that for almost all ω

$$\sup_n \frac{\|X_n(\omega)\|}{\sqrt{n}} < \infty .$$

Therefore, by independence and the Borel-Cantelli lemma, $\mathbb{E}\|X\|^2 < \infty$. This proves the implication (ii) $\Rightarrow$ (i).

The main tool in the proof of the converse implication (i) $\Rightarrow$ (ii) is the following lemma. This lemma may be considered as some vector valued extension of a strong law of large numbers for squares.

Lemma 10.15. In the setting of Theorem 10.14, if $\mathbb{E}\|X\|^2 < \infty$, for some numerical constant K , almost surely,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}_\xi \|\sum_{i=1}^n \xi_i X_i\| \le K \limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}\|\sum_{i=1}^n \xi_i X_i\|$$

where, as usual, $\mathbb{E}_\xi$ denotes partial integration with respect to the sequence (ξ_i) .

Proof. Set $M = \limsup_{n\to\infty} \mathbb{E}\|\sum_{i=1}^n \xi_i X_i/\sqrt{n}\|$, assumed to be finite. By Lemma 6.3, since $\mathbb{E}\xi = 0$, replacing ξ by $\varepsilon\xi$, we may and do assume ξ to be symmetric. By the Borel-Cantelli lemma, and monotonicity of the averages, it suffices to show that for some K , and all $\varepsilon > 0$,

$$\sum_n \mathbb{P}_X\{\mathbb{E}_\xi\|\sum_{i=1}^{2^n} \xi_i X_i\| > K(M+\varepsilon)2^{n/2}\} < \infty ,$$

or further, by definition of M , that

$$\sum_n \mathbb{P}_X\{\mathbb{E}_\xi\|\sum_{i=1}^{2^n} \xi_i X_i\| > K\mathbb{E}\|\sum_{i=1}^{2^n} \xi_i X_i\| + \varepsilon 2^{n/2}\} < \infty .$$

To show this, we make use of the isoperimetric approach based on Theorem 1.4 developed in Section 6.3. Since $\mathbb{E}\|X\|^2 < \infty$, by Lemmas 7.6 and 7.8, there exists a sequence (k_n) of integers such that $\sum_n 2^{-k_n} < \infty$ and

$$\sum_n \mathbb{P}\{\sum_{i=1}^{k_n} \|X_i\|^* > \varepsilon 2^{n/2}\} < \infty$$

where $(\|X_i\|^*)$ is the non-increasing rearrangement of $(\|X_i\|)_{i\le 2^n}$. We now make use of Remark 6.18 which applies similarly to the averages in the symmetric sequence (ξ_i) . Note that $\mathbb{E}|\xi| \le 1$. Hence by (6.23) adapted to (ξ_i) with $q = 2K_0$ and $k = k_n$ ($\geq q$ for n large enough),

$$\mathbb{P}_X\{\mathbb{E}_\xi\|\sum_{i=1}^{2^n}\xi_i X_i\| > 2q\mathbb{E}\|\sum_{i=1}^{2^n}\xi_i X_i\| + \varepsilon 2^{n/2}\}$$

$$\leq 2^{-k_n} + \mathbb{P}\{\sum_{i=1}^{k_n}\|X_i\|^* > \varepsilon 2^{n/2}\} .$$

Letting $K = 2q = 4K_0$, the proof of Lemma 10.15 is complete.

We now conclude the proof of the theorem. Since X satisfies the CLT, for every $k \geq 1$, there exists a finite dimensional subspace F_k of B such that if $T_k = T_{F_k} : B \to B/F_k$

$$CLT(T_k(X)) \leq \frac{1}{2Kk\|\xi\|_{2,1}} .$$

Recall from Proposition 10.4 that $CLT(\xi X) \leq 2\|\xi\|_{2,1} CLT(X)$. If we now apply Lemma 10.15 to $T_k(X)$ for each k, there exists Ω_k with $\mathbb{P}(\Omega_k) = 1$ such that for all ω in Ω_k

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\mathbb{E}_\xi\|\sum_{i=1}^{n}\xi_i T_k(X_i(\omega))\| \leq \frac{1}{k} .$$

Let also Ω_0 the set of full probability obtained when Lemma 10.15 is applied to X itself. Let $\Omega^0 = \bigcap_{k\geq 0}\Omega_k$; $\mathbb{P}(\Omega^0) = 1$. Let now $\omega \in \Omega^0$. For each $\varepsilon > 0$, there exists a finite dimensional subspace F of B such that if $T = T_F$,

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\mathbb{E}_\xi\|\sum_{i=1}^{n}\xi_i T(X_i(\omega))\| < \varepsilon^2 .$$

Hence, if $n \geq n_0(\varepsilon)$,

$$\mathbb{P}_\xi\{\|T(\sum_{i=1}^{n}\xi_i X_i(\omega)/\sqrt{n})\| > \varepsilon\} \leq \varepsilon .$$

It follows that the sequence $(\sum_{i=1}^{n}\xi_i X_i(\omega)/\sqrt{n})$ is tight (it is bounded in probability si nce $\omega \in \Omega_0$). We conclude the proof by identifying the limit. Using basically that $\sum_{i=1}^{n} f^2(X_i)/n \to \mathbb{E}f^2(X)$ almost surely, it is not difficulty to see, by the Lindeberg CLT (cf. e.g. [Ar-G2]) for example, that for every f in B' there exists a set Ω_f of probability one such that for all $\omega \in \Omega_f$, $(\sum_{i=1}^{n}\xi_i f(X_i(\omega))/\sqrt{n})$ converges in distribution to a normal variable with variance $\mathbb{E}f^2(X)$. The proof is then easily completed by considering a weakly dense countable subset in B'. Theorem 10.14 is established.

Notes and references

This chapter only concentrate on the classical central limit theorem for sums of identically distributed random variables under the normalization $\sqrt{n}$. We would like to refer to the book of A. Araujo and E. Giné [Ar-G2] for a more complete account on the general CLT for real and Banach space valued random variables, as well as for precise and detailed historical and recent references. See also the nice paper [A-A-G]. We also mention the recent book by V. Paulauskas and A. Rachkauskas [P-R2] on rates of convergence in the vector valued CLT, a topic not covered here. We note further that some more results on the CLT, using empirical processes methods, will be presented in Chapter 14.

Starting with Donsker's invariance principle [Do], the study of the CLT for Banach space valued random variables was initiated by E. Mourier [Mo] and R. Fortet [F-M1], [F-M2], S. R. S. Varadhan [Var], R. Dudley and V. Strassen [D-S], L. Le Cam [LC]. The proof we present of the necessity of $\mathbb{E}X^2 < \infty$ on the line and similarly of $\mathbb{E}\|X\|^2 < \infty$ in cotype 2 spaces is due to N. C. Jain [Ja1] who also showed necessity of $\|X\|_{2,\infty} < \infty$ in any Banach space [Ja1], [Ja2]. The improved Lemma 10.1 was then noticed independently in [A-A-G] and [P-Z]. Corollary 10.2 was observed in [Pi3]. Proposition 10.3 is due to G. Pisier and J. Zinn [P-Z]. The randomization property of Proposition 10.4 has been known for some time independently by X. Fernique and G. Pisier and put forward in the paper [G-Z2]. Our proof is Pisier's and the best possibility of $L_{2,1}$ was shown in [L-T1].

The extension to Hilbert spaces of the classical CLT was obtained by S.R.S. Varadhan [Var]. A further extension in some smooth spaces, anticipating type 2, is described in [F-M1]. Examples of bounded random variables in $C[0,1]$ failing the CLT were provided in [D-S], and in L_p, $1 \le p < 2$, by R. Dudley (cf. [Kue2]). A decisive step was accomplished by J. Hoffmann-Jørgensen and G. Pisier [HJ-P] with Theorem 10.5 and N. C. Jain [Ja2] with Theorem 10.7 and Proposition 10.8. This proposition is due independently to D. Aldous [Ald]. See also [Pi3], [HJ3], [HJ4]. Rosenthal's inequality appeared in [Ros]. Its interest in the study of the CLT in L_p-spaces was put forward in [G-M-Z] and further by J. Zinn in [Zi2] where Theorem 10.10 is explained. An attempt of a systematic study of Rosenthal's inequality for vector valued random variables is undertaken in [Led4]. The characterization of the CLT in L_p (and more general Banach lattices) goes back to [P-Z] (cf. also [G-Z1]). That the best possible necessary conditions for the CLT are not sufficient in $\ell_2(B)$ when B is not of cotype 2 is due to J. Zinn [Zi2], [G-Z1]. Some further CLTs in $\ell_p(\ell_q)$ are investigated in this last paper [G-Z1]. The example on the CLT and bounded CLT in ℓ_p, $2 < p < \infty$, is taken from [P-Z].

Theorem 10.11 is due to G. Pisier and J. Zinn [P-Z] thanks to several early results on the CLT and the LIL in L_p-spaces, $2 \leq p < \infty$. G. Pisier [Pi3] established Theorem 10.12 assuming a strong moment and with a proof containing the essential step of the general case. The final result is due independently to V. Goodman, J. Kuelbs, J. Zinn [G-K-Z] and B. Heinkel [He2]. The comments on the implication CLT $\Rightarrow$ LIL are taken from [Pi3].

The small ball criterion (Theorem 10.13) was obtained in [L-T3]. On the line, the result was noticed in [J-O]. We learned how to use Kanter's inequality (Proposition 6.4) in this study from X. Fernique. The proof of Theorem 10.13 with the isoperimetric approach is new. Theorem 10.14 is due to J. Zinn and the authors and also appeared in [L-T3] (with a different proof). Its proof, and in particular Lemma 10.15, has been used recently in bootstrapping of empirical measures by E. Giné and J. Zinn [G-Z4].

Chapter 11. Regularity of random processes

Chapter 11. Regularity of random processes

In Chapter 9 we described how certain conditions on Banach spaces can ensure the existence and tightness of some probability measures. For example, if (x_i) is a sequence in a type 2 Banach space B such that $\sum_i \|x_i\|^2 < \infty$, then the series $\sum_i g_i x_i$ is almost surely convergent and defines a Gaussian Radon random variable with values in B. These conditions were further used in Chapters 9 and 10 to establish tightness properties of sums of independent random variables, especially in the context of central limit theorems.

In this chapter, another approach to existence and tightness of certain measures is taken in the framework of random functions and processes. Given a random process $X = (X_t)_{t \in T}$ indexed by some set T, we investigate sufficient conditions for almost sure boundedness or continuity of the sample paths of X in terms of the "geometry" (in the metrical sense) of T. By geometry, we mean some metric entropy or majorizing measure condition which estimates the size of T in function of some parameters related to X. The setting of this study has its roots in a celebrated theorem of Kolmogorov which gives sufficient conditions for the continuity of processes X indexed by a compact subset of $\mathbb{R}$ in terms of a Lipschitz condition on the *increments* $X_s - X_t$ of the processes. Under this type of incremental conditions on the processes, this result was extended to processes indexed by regular subsets T of $\mathbb{R}^N$ and then further to abstract index sets T. In this chapter, we present several results in this general abstract setting. The first section deals with the metric entropy condition. The results, which naturally extend the more classical ones, are rather easy to prove and to use but nevertheless can be shown to be sharp in many respects. The second paragraph investigates majorizing measure conditions which are more precise than entropy conditions as they take more into account the local geometry of the index set. That majorizing measures are a key notion will be shown in the next chapter on Gaussian processes. These sufficient entropy or majorizing measure conditions for sample boundedness or continuity are utilized in the proofs in a rather similar manner: the main idea is indeed based on the rather classical covering technique and chaining argument already contained in Kolmogorov's theorem. In Section 11.3, we present important examples of applications to Gaussian, Rademacher and chaos processes.

Common to this chapter is the datum of a random process $X = (X_t)_{t \in T}$, that is a collection (X_t) of random variables, indexed by some parameter set T which we assume to be a metric or pseudo-metric space. By pseudo-metric recall we mean that T is equipped with a distance d which does not necessarily separate points ($d(s,t) = 0$ does not always imply $s = t$). Our main objective is thus to find sufficient conditions in order for X to be almost surely bounded or continuous, or to possess a version with these properties. We usually work with processes $X = (X_t)_{t\in T}$ which are in L_p, $1 \le p < \infty$, or, more generaly, in some Orlicz space L_ψ, i.e., $\|X_t\|_\psi < \infty$ for all t. Concerning almost sure boundedness, and according to what was described in Chapter 2, we therefore simply understand supremum like $\sup_{t\in T} X_t$, $\sup_{t\in T} |X_t|$, $\sup_{s,t\in T} |X_s - X_t|, \ldots,$ as lattice supremum in L_ψ; for example,

$$\mathbb{E} \sup_{s,t\in T} |X_s - X_t| = \sup\{\mathbb{E} \sup_{s,t\in F} |X_s - X_t| ;\ F \text{ finite in } T\} .$$

We avoid in this way the usual measurability questions and moreover reduce the estimates we will establish to the case of a *finite* parameter set T. We could of course also use separable versions which we do anyway in the study of sample continuity.

One more word before turning to the object of this chapter. In the theorems below, we usually bound the quantity $\mathbb{E} \sup_{s,t\in T} |X_s - X_t|$ (for T finite for simplicity). Of course

$$\mathbb{E} \sup_{s,t\in T} |X_s - X_t| = \mathbb{E} \sup_{s,t\in T} (X_s - X_t)$$

which is also equal to $2\mathbb{E} \sup_{t\in T} X_t$ if the process is *symmetric* (i.e., the distribution in $\mathbb{R}^T$ of $-X$ and X are the same, for example Gaussian processes). We also have that, for every t_0 in T,

$$\mathbb{E} \sup_{t\in T} X_t \le \mathbb{E} \sup_{t\in T} |X_t| \le \mathbb{E}|X_{t_0}| + \mathbb{E} \sup_{s,t\in T} |X_s - X_t|$$

and, when X is symmetric,

$$\mathbb{E} \sup_{t\in T} X_t \le \mathbb{E} \sup_{t\in T} |X_t| \le \mathbb{E}|X_{t_0}| + 2\mathbb{E} \sup_{t\in T} X_t .$$

These inequalities are used freely below. The example of T being reduced to one point shows that the estimates we establish do not hold in general for $\mathbb{E} \sup_{t\in T} |X_t|$ rather than $\mathbb{E} \sup_{t\in T} X_t$ or $\mathbb{E} \sup_{s,t\in T} |X_s - X_t|$. The supremum notations will also often be shortened in $\sup_T$ or $\sup_t$, $\sup_{s,t}$, etc.

Recall that a Young function ψ is a convex increasing function on $\mathbb{R}_+$ with $\lim_{t\to\infty} \psi(t) = \infty$ and $\psi(0) = 0$. The Orlicz space $L_\psi = L_\psi(\Omega, \mathcal{A}, \mathbb{P})$ associated to ψ is defined as the space of all real random variables Z on $(\Omega, \mathcal{A}, \mathbb{P})$ such that $\mathbb{E}\psi(|Z|/c) < \infty$ for some $c > 0$. Recall furthermore it is a Banach space for the norm

$$\|Z\|_\psi = \inf\{c > 0 ;\ \mathbb{E}\psi(|Z|/c) \le 1\} .$$

The general question we study in the first two sections of this chapter is the following: given a Young function ψ and a random process $X = (X_t)_{t\in T}$ indexed by (T, d) and in L_ψ (i.e. $\|X_t\|_\psi < \infty$ for all t) satisfying the Lipschitz conditions in L_ψ

$$\|X_s - X_t\|_\psi \le d(s,t) \qquad \text{for all } s, t \in T\,, \tag{11.1}$$

find then estimates of $\sup_t X_t$ and sufficient conditions for sample boundedness or continuity of X in terms of "the geometry of $(T, d; \psi)$". By this we mean the size of T measured in terms of d and ψ. Note that we could take as pseudo-metric d the one given by the process itself $d(s,t) = \|X_s - X_t\|_\psi$ so that T may be measured in terms of X. The main idea will be to convey through the incremental conditions (11.1) boundedness and continuity of X in the function space L_ψ to the corresponding almost sure properties. This will be accomplished in a chaining argument.

Our main geometric measures of $(T, d; \psi)$ are the metric entropy condition and the majorizing measure condition. The first section develops the results under the concept of entropy which we already encountered in some of the preceding chapters in necessity results.

Let us note before turning to these results that the study of the continuity (actually uniform continuity) will always follow rather easily from the various bounds established for boundedness that thus appears as the main case of this investigation. This situation is rather classical and we already met it for example in Chapters 7-9 in the study of limit theorems. Let us also mention that the fact that we are working with pseudo-metrics rather than metrics is not really important since we can always identify two points s and t in T such that $d(s,t) = 0$; under (11.1), $X_s = X_t$ almost surely. Further, all the conditions we will use, entropy or majorizing measures, imply that (T, d) is totally bounded. For simplicity, one can therefore reduce everything if one wishes it to the case of (T, d) metric compact.

11.1. Regularity of random processes under metric entropy conditions

Let (T, d) be a pseudo-metric space. For each $\varepsilon > 0$, denote by $N(T, d; \varepsilon)$ the smallest number of open balls of radius $\varepsilon > 0$ in the pseudo-metric d which form a covering of T. (Recall we could work equivalently with closed balls.) T is totally bounded for d if and only if $N(T, d; \varepsilon) < \infty$ for every $\varepsilon > 0$, a property which will always be satisfied under all the conditions we will deal with. Denote further by $D = D(T)$ the diameter of (T, d) i.e.,

$$D = \sup\{d(s,t);\ s, t \in T\}$$

(finite or infinite).

The following theorem is the main regularity result under metric entropy condition for processes with increments satisfying (11.1). It only concerns at this point boundedness but continuity will be achieved similarly later on. ψ^{-1} denotes the inverse function of ψ .

Theorem 11.1. Let $X = (X_t)_{t\in T}$ be a random process in L_ψ such that for all s,t in T

$$\|X_s - X_t\|_\psi \le d(s,t) .$$

Then, if

$$\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon < \infty ,$$

X is almost surely bounded and we actually have

$$\mathbb{E} \sup_{s,t\in T} |X_s - X_t| \le 8 \int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon .$$

It is clear that the convergence of the entropy integral is understood when $\varepsilon \to 0$. The numerical constant 8 is without any special meaning.

We will actually prove a somewhat better result whose proof is not more difficult.

Theorem 11.2. Let ψ be a Young function and let $X = (X_t)_{t\in T}$ be a random process in $L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$ such that for all measurable set A in Ω and all s,t in (T,d) ,

$$\int_A |X_s - X_t| d\mathbb{P} \le d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) .$$

Then, for all A ,

$$\int_A \sup_{s,t\in T} |X_s - X_t| d\mathbb{P} \le 8\mathbb{P}(A) \int_0^D \psi^{-1}(\mathbb{P}(A)^{-1}N(T,d;\varepsilon))d\varepsilon .$$

Note that this statement does not really concern ψ but rather the function $u\psi^{-1}(1/u)$, $0 < u \le 1$, and should perhaps be preferably stated in this way.

Before proving Theorem 11.2, let us explain the advantages of its formulation and how it includes Theorem 11.1. For the latter, simply note that when (11.1) holds, by convexity and Jensen's inequality

$$\begin{aligned}\int_A |X_s - X_t| d\mathbb{P} &= d(s,t)\mathbb{P}(A) \int_A \psi^{-1} \circ \psi\left(\frac{|X_s - X_t|}{d(s,t)}\right) \frac{d\mathbb{P}}{\mathbb{P}(A)} \\ &\le d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\mathbb{E}\psi\left(\frac{|X_s - X_t|}{d(s,t)}\right)\right) \\ &\le d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) .\end{aligned}$$

Conversely it should be noted that if Z is a positive random variable such that for every measurable set A

$$\int_A Z d\mathbb{P} \leq \mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right),$$

then, letting $A = \{Z > u\}$ and using Chebyshev's inequality we have

$$\mathbb{P}\{Z > u\} \leq \frac{1}{u}\int_{\{Z>u\}} Z d\mathbb{P} \leq \frac{1}{u}\mathbb{P}\{Z > u\}\psi^{-1}\left(\frac{1}{\mathbb{P}\{Z > u\}}\right)$$

so that for every $u > 0$,

$$\mathbb{P}\{Z > u\} \leq \frac{1}{\psi(u)}.$$

For Young functions of exponential type the latter is equivalent to say that $\|Z\|_\psi < \infty$. However, for power type functions, it is less restrictive and this is why Theorem 11.2 is more general in its hypotheses. It includes for example conditions in weak L_p-spaces. Let $1 < p < \infty$ and assume indeed we have a random process X such that for all s, t in T

$$\|X_s - X_t\|_{p,\infty} \leq d(s,t). \tag{11.2}$$

Then, as is easily seen by integration by parts,

$$\begin{aligned}\int_A |X_s - X_t| d\mathbb{P} &\leq q d(s,t)\mathbb{P}(A)^{1/q} \\ &\leq q d(s,t)\mathbb{P}(A)\left(\frac{1}{\mathbb{P}(A)}\right)^{1/p}\end{aligned}$$

where $q = p/p-1$ is the conjugate of p.

This for the advantages in the hypotheses. Concerning the conclusion, the formulation in Theorem 11.2 allows to obtain directly some quite sharp integrability and tail estimates of the sup-norm of the process X much in the spirit of those described in the first part of the book. If, for example, ψ is such that for some constant $C = C_\psi$ and all $x, y \geq 1$

$$\psi^{-1}(xy) \leq C\psi^{-1}(x)\psi^{-1}(y)$$

(which is the case for example when $\psi(x) = x^p$, $1 \leq p < \infty$), then the conclusion of Theorem 11.2 is that for all measurable set A

$$\int_A \sup_{s,t} |X_s - X_t| d\mathbb{P} \leq 8CE\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right)$$

where

$$E = E(T, d; \psi) = \int_0^D \psi^{-1}(N(T, d; \varepsilon))d\varepsilon \,,$$

assumed of course to be finite. Then, by Chebyshev's inequality as before, for every $u > 0$,

$$\mathbb{P}\{\sup_{s,t} |X_s - X_t| > u\} \le \left(\psi^{-1}\left(\frac{u}{8CE}\right)\right)^{-1} . \tag{11.3}$$

This applies in particular to the preceding setting of (11.2) for $\psi(x) = x^p$, $1 < p < \infty$, in which case we get that

$$\| \sup_{s,t} |X_s - X_t| \|_{p,\infty} \le 8qCE \,.$$

Hence, under the finiteness of the entropy integral, we conclude to a degree of integrability for the supremum which is exactly the same as the one we started with on the individuals $(X_s - X_t)$. In case we have $\|X_s - X_t\|_p \le d(s,t)$ instead of (11.2) we end up with a small gap in this order of idea. This can however easily be repaired. Indeed, if Z is a positive random variable such that $\|Z\|_q = 1$ and if we set $\mathbb{Q}(A) = \int_A Z^q d\mathbb{P}$, then, from the assumption, for every measurable A and all s, t,

$$\int_A |X_s - X_t| d\mathbb{Q} \le d(s,t)\mathbb{Q}(A)^{1/q} \,.$$

Theorem 11.2 with respect to $\mathbb{Q}$ yields

$$\int \sup_{s,t} |X_s - X_t| d\mathbb{Q} \le 8CE$$

from which, by uniformity in $\mathbb{Q}$, it follows that

$$\| \sup_{s,t} |X_s - X_t| \|_p \le 8CE \,.$$

If ψ is an exponential function $\psi_q(x) = \exp(x^q) - 1$, then (11.3) immediately implies that

$$\| \sup_{s,t} |X_s - X_t| \|_{\psi_q} \le C_q E$$

for some constant C_q depending only on q. In this case actually, the general estimate (11.3) can be improved into a *deviation* inequality. Assume ψ satisfies now

$$\psi^{-1}(xy) \le C(\psi^{-1}(x) + \psi^{-1}(y))$$

for all $x, y \geq 1$. The functions ψ_q satisfy this inequality. In this situation, Theorem 11.2 indicates that for every measurable set A ,

$$\int_A \sup_{s,t} |X_s - X_t| d\mathbb{P} \leq 8C\mathbb{P}(A)\left(D\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + E\right) .$$

This easily implies for every $u > 0$

$$\mathbb{P}\{\sup_{s,t} |X_s - X_t| > 8C(E+u)\} \leq \left(\psi\left(\frac{u}{D}\right)\right)^{-1} . \tag{11.4}$$

This is an inequality of the type we obtained for bounded Gaussian or Rademacher processes in Chapters 3 and 4 with *two* parameters, E , the entropy integral, which measures some information on the sup-norm in L_p , $0 \leq p < \infty$, and the diameter D which may be assimilated to weak moments. For the purpose of comparison, note that, obviously,

$$E = \int_0^D \psi^{-1}(N(T, d; \varepsilon))d\varepsilon \geq \psi^{-1}(1)D$$

($\psi^{-1}(1) > 0$ by convexity) and E is much bigger than D in general.

We now prove Theorem 11.2.

Proof of Theorem 11.2. It is enough to prove the inequality of the statement with T finite. Let ℓ_0 be the largest integer (in $\mathbb{Z}$) such that $2^{-\ell} \geq D$; let also ℓ_1 be the smallest integer ℓ such that the open balls $B(t, 2^{-\ell})$ of center t and radius $2^{-\ell}$ in the metric d are reduced to exactly one point. For each $\ell_0 \leq \ell \leq \ell_1$, let $T_\ell \subset T$ of cardinality $N(T, d; 2^{-\ell})$ such that the balls $\{B(t, 2^{-\ell});\ t \in T_\ell\}$ form a covering of T . By induction, define maps $h_\ell : T_\ell \to T_{\ell-1}$, $\ell_0 < \ell \leq \ell_1$, such that $t \in B(h_\ell(t), 2^{-\ell+1})$. Set then $k_\ell : T \to T_\ell$, $k_\ell = h_{\ell+1} \circ \cdots \circ h_{\ell_1}$, $\ell_0 \leq \ell \leq \ell_1$ ($k_{\ell_1} =$ identity). We can then write the fundamental *chaining* identity which is at the basis of most of the results on boundedness and continuity of processes under conditions on increments. Since $2^{-\ell_0} \geq D$, T_{ℓ_0} is reduced to one point, call it t_0 . Then, for every t in T ,

$$X_t - X_{t_0} = \sum_{\ell=\ell_0+1}^{\ell_1} (X_{k_\ell(t)} - X_{k_{\ell-1}(t)}) .$$

It follows that

$$\sup_{s,t \in T} |X_s - X_t| \leq 2 \sum_{\ell=\ell_0+1}^{\ell_1} \sup_{t \in T} |X_{k_\ell(t)} - X_{k_{\ell-1}(t)}| .$$

Now, for fixed ℓ, observe that

$$\mathrm{Card}\{(X_{k_\ell(t)} - X_{k_{\ell-1}(t)});\ t \in T\} \le N(T,d;2^{-\ell}).$$

Further, by construction, $d(k_\ell(t), k_{\ell-1}(t)) \le 2^{-\ell+1}$ for every t. Hence the hypothesis indicates that for every t, $\ell_0 < \ell \le \ell_1$ and A measurable

$$\int_A |X_{k_\ell(t)} - X_{k_{\ell-1}(t)}| d\mathbb{P} \le 2^{-\ell+1}\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right).$$

The conclusion will then easily follow from the following elementary lemma.

Lemma 11.3. Let $(Z_i)_{i\le N}$ be positive random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that for all $i \le N$ and all measurable set A in Ω,

$$\int_A Z_i d\mathbb{P} \le \mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right).$$

Then, for every measurable set A,

$$\int_A \max_{i\le N} Z_i d\mathbb{P} \le \mathbb{P}(A)\psi^{-1}\left(\frac{N}{\mathbb{P}(A)}\right).$$

Proof. Let $(A_i)_{i\le N}$ be a (measurable) partition of A such that $Z_i = \max_{j\le N} Z_j$ on A_i. Then

$$\begin{aligned}\int_A \max_{i\le N} Z_i d\mathbb{P} = \sum_{i=1}^N \int_{A_i} Z_i d\mathbb{P} &\le \sum_{i=1}^N \mathbb{P}(A_i)\psi^{-1}\left(\frac{1}{\mathbb{P}(A_i)}\right)\\ &\le \mathbb{P}(A)\psi^{-1}\left(\frac{N}{\mathbb{P}(A)}\right)\end{aligned}$$

where we have used in the last step that ψ^{-1} is concave and that $\sum_{i=1}^N \mathbb{P}(A_i) = \mathbb{P}(A)$. The lemma is proved.

Note that the lemma applies when $\max_{i\le N} \|Z_i\|_\psi \le 1$. This lemma of course describes one of the key points of the entropy approach through the intervention of the cardinality N.

We can now conclude the proof of Theorem 11.2. Together with the chaining inequality, the lemma implies that for any measurable set A,

$$\begin{aligned}\int_A \sup_{s,t} |X_s - X_t| d\mathbb{P} &\le 2\sum_{\ell=\ell_0+1}^{\ell_1} \int_A \sup_t |X_{k_\ell(t)} - X_{k_{\ell-1}(t)}| d\mathbb{P}\\ &\le 4\mathbb{P}(A)\sum_{\ell>\ell_0} 2^{-\ell}\psi^{-1}(\mathbb{P}(A)^{-1}N(T,d;2^{-\ell})).\end{aligned}$$

The conclusion follows from a simple comparison between series and integral:

$$\sum_{\ell>\ell_0} 2^{-\ell}\psi^{-1}(\mathbb{P}(A)^{-1}N(T,d;2^{-\ell})) \le 2\sum_{\ell>\ell_0}\int_{2^{-\ell-1}}^{2^{-\ell}} \psi^{-1}(\mathbb{P}(A)^{-1}N(T,d;\varepsilon))d\varepsilon$$
$$\le 2\int_0^D \psi^{-1}(\mathbb{P}(A)^{-1}N(T,d;\varepsilon))d\varepsilon$$

by definition of ℓ_0. (Note that similar simple comparison between series and integral will be used frequently throughout this chapter and the next ones, usually without any further comments.) Theorem 11.2 is therefore established.

Remark 11.4. What was thus used in this proof is the existence, for every ℓ, of a finite subset T_ℓ of T such that, for every t, there exists $t_\ell \in T_\ell$ with $d(t,t_\ell) \le 2^{-\ell}$, and with the property

$$\sum_\ell 2^{-\ell}\psi^{-1}(\operatorname{Card} T_\ell) < \infty .$$

It should be noted further that the preceding proof also applies to the random functions $X = (X_t)_{t\in T}$ with the following property: for every ℓ ($\ge \ell_0$), every (t,t_ℓ) as before and every measurable set A,

$$\int_A |X_t - X_{t_\ell}|d\mathbb{P} \le 2^{-\ell}\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + M_\ell\mathbb{P}(A) \tag{11.5}$$

where (M_ℓ) is a sequence of positive numbers satisfying $\sum_\ell M_\ell < \infty$. [This property is in particular realized when, for some C, for all measurable sets A and all s,t in T,

$$\int_A |X_s - X_t|d\mathbb{P} \le d(s,t)\mathbb{P}(A)(\psi^{-1}(\frac{1}{\mathbb{P}(A)}) + C) .]$$

The final bound of course involves then this quantity in the form

$$\int_A \sup_{s,t\in T}|X_s - X_t|d\mathbb{P} \le 8\mathbb{P}(A)\int_0^D \psi^{-1}(\mathbb{P}(A)^{-1}N(T,d;\varepsilon))d\varepsilon + 2\mathbb{P}(A)\sum_{\ell\ge\ell_0} M_\ell$$

where we recall that ℓ_0 is the largest integer ℓ such that $2^{-\ell} \ge D$. The proof is straightforward. This simple observation can be useful as will be illustrated in Section 11.3.

Remark 11.5. We collect here some further easy observations which will be useful next. First note that in Theorems 11.1 and 11.2 (and Remark 11.4 too) we might have as well equipped T with the pseudo-metric $d(s,t) = \|X_s - X_t\|_\psi$ induced by X itself. Further, since the hypotheses and conclusions of these theorems actually only involve the increments $X_s - X_t$ in absolute values, and since the only property used on them is that they satisfy the triangle inequality, the preceding results may be trivially extended to the setting of random distances; that is, random processes

$(D(s,t))_{s,t\in T}$ on $T \times T$ such that for all s,t,u in T, $D(s,s) = 0 \leq D(s,t) = D(t,s) \leq D(s,u) + D(u,t)$ with probability one. This includes $D(s,t) = |X_s - X_t|^\alpha$, $0 < \alpha \leq 1$, and $D(s,t) = \min(1, |X_s - X_t|)$ for example. Actually we could also include in this way random process with values in a Banach space by setting $D(s,t) = \|X_s - X_t\|$ (or some of the preceding variations). By random process with values in a Banach space, we simply mean here a collection $X = (X_t)_{t\in T}$ of random variables with values in a Banach space. An example will be discussed in Section 11.3. More extensions of this type can be obtained; we leave them to the interested reader. The one we mentioned will be useful to include some classical statements.

We now present the continuity theorem in the preceding framework.

Theorem 11.6. Let ψ be a Young function and let $X = (X_t)_{t\in T}$ be a random process in $L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$ such that for all measurable set A in Ω and all s,t in (T,d),

$$\int_A |X_s - X_t| d\mathbb{P} \leq d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right).$$

Then, if

$$\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon < \infty ,$$

X admits a version $\widetilde{X}$ with almost all sample paths bounded and (uniformly) continuous on (T,d). Moreover, $\widetilde{X}$ satisfies the following property: for each $\varepsilon > 0$, there exists $\eta > 0$, depending only on ε and the finiteness of the entropy integral but not on the process X, itself such that

$$\mathbb{E} \sup_{d(s,t)<\eta} |\widetilde{X}_s - \widetilde{X}_t| < \varepsilon .$$

Proof. We take the notations of the proof of Theorem 11.2. The main point is to show that, when T is finite, for every $\eta > 0$ and $\ell_0 \leq \ell \leq \ell_1$,

$$\mathbb{E} \sup_{d(s,t)<\eta} |X_s - X_t| \leq \eta\psi^{-1}(N(T,d;2^{-\ell})^2) + 8\sum_{m>\ell} 2^{-m}\psi^{-1}(N(T,d;2^{-m})) . \tag{11.6}$$

Let η and ℓ be fixed. The proof of Theorem 11.2 indicates that the chaining identity

$$X_t - X_{k_\ell(t)} = \sum_{m=\ell+1}^{\ell_1} (X_{k_m(t)} - X_{k_{m-1}(t)})$$

implies that

$$\mathbb{E}\sup_{t\in T}|X_t - X_{k_\ell(t)}| \leq 2\sum_{m>\ell} 2^{-m}\psi^{-1}(N(T,d;2^{-m}))\,.$$

Let now

$$U = \{(x,y)\in T_\ell\times T_\ell\,;\ \exists u,v \ \text{ in } T \ \text{ such that } d(u,v)<\eta \ \text{ and } k_\ell(u)=x\,,\ k_\ell(v)=y\}\,.$$

If $(x,y)\in U$, we *fix* $u_{x,y}$, $v_{x,y}$ such that $k_\ell(u_{x,y})=x$, $k_\ell(v_{x,y})=y$ and $d(u_{x,y},v_{x,y})<\eta$. By Lemma 11.3,

$$\begin{aligned}\mathbb{E}\sup_{(x,y)\in U}|X_{u_{x,y}} - X_{v_{x,y}}| &\leq \eta\psi^{-1}(\text{ Card } U)\\ &\leq \eta\psi^{-1}(N(T,d;2^{-\ell})^2)\,.\end{aligned}$$

Let now s,t be arbitrary in T satisfying $d(s,t)<\eta$. Set $x=k_\ell(s)$, $y=k_\ell(t)$. Clearly $(x,y)\in U$. We can then write by the triangle inequality

$$\begin{aligned}|X_s - X_t| &\leq |X_s - X_{k_\ell(s)}| + |X_{k_\ell(s)} - X_{u_{x,y}}| + |X_{u_{x,y}} - X_{v_{x,y}}|\\ &\quad + |X_{v_{x,y}} - X_{k_\ell(t)}| + |X_{k_\ell(t)} - X_t|\\ &\leq \sup_{(x,y)\in U}|X_{u_{x,y}} - X_{v_{x,y}}| + 4\sup_{r\in T}|X_r - X_{k_\ell(r)}|\end{aligned}$$

where we have used that $k_\ell(u_{x,y}) = k_\ell(s) = x$ and similarly for y. We then have clearly (11.6).

We can now conclude the proof of the theorem. We have obtained by (11.6) that, under the finiteness of the entropy integral, for each $\varepsilon>0$ there exists $\eta>0$ depending only on $\varepsilon>0$ and T,d,ψ such that, for every finite and thus also countable subset S of T,

$$\mathbb{E}\sup_{\substack{s,t\in S\\ d(s,t)<\eta}}|X_s - X_t| < \varepsilon\,.$$

Since (T,d) is totally bounded, there exists S countable and dense in T. Setthen $\widetilde{X}_t = X_t$ if $t\in S$ and $\widetilde{X}_t = \lim X_s$ where this limit, in probability or L_1, is taken for $s\to t$, $s\in S$. Then $(\widetilde{X}_t)_{t\in T}$ is cleary a version of X which satisfies all the required properties. To see in particular that $(\widetilde{X}_t)_{t\in T}$ has uniformly continuous sample paths on (T,d), let, for each n, $\eta_n>0$ be such that

$$\mathbb{E}\sup_{d(s,t)<\eta_n}|\widetilde{X}_s - \widetilde{X}_t| < 4^{-n}\,.$$

Then, if $A_n = \{ \sup_{d(s,t)<\eta_n} |\widetilde{X}_s - \widetilde{X}_t| > 2^{-n} \}$, $\sum_n \mathbb{P}(A_n) < \infty$ and the claim ¿follows from the Borel-Cantelli lemma. The proof of Theorem 11.6 is complete.

It is plain that Remarks 11.4 and 11.5 also apply in the context of Theorem 11.6. Further, the dependence of $\eta > 0$ we carefully describe in Theorem 11.6 has some easy consequences to tightness results. Assume (T, d) is a compact metric space and denote by $C(T)$ the Banach space of continuous functions on T equipped with the sup-norm. By Prokhorov's criterion and the Arzela-Ascoli characterization of compact sets in $C(T)$ (cf. e.g. [Bi]), it is easily seen that a family $\mathcal{X}$ of random variables $X = (X_t)_{t\in T}$ is relatively compact in the weak topology of probability distributions on $C(T)$ as soon as, for some t , $\{X_t;\ X \in \mathcal{X}\}$ is relatively compact as real random variables and, for each $\varepsilon > 0$, there is $\eta > 0$ such that for every X in $\mathcal{X}$

$$\mathbb{E} \sup_{d(s,t)<\eta} |X_s - X_t| < \varepsilon \, .$$

But this last condition is exactly what is provided by Theorem 11.6 under the entropy condition. We thus have the following consequence which will be of interest in the study of the central limit theorem in Chapter 14.

Corollary 11.7. Let (T, d) be compact and let ψ be a Young function. Assume that

$$\int_0^D \psi^{-1}(N(T, d; \varepsilon))d\varepsilon < \infty \, .$$

Let $\mathcal{X}$ be a family of separable random processes $X = (X_t)_{t\in T}$ in $L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$ such that for all s, t in T and A measurable

$$\int_A |X_s - X_t| d\mathbb{P} \le d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) .$$

Then each element of $\mathcal{X}$ defines a tight probability distribution on $C(T)$ and $\mathcal{X}$ is weakly relatively compact if and only if, for some $t \in T$, $\{X_t; X \in \mathcal{X}\}$ is weakly relatively compact (as measures on $\mathbb{R}$).

As a first application, we would like in particular to indicate how the preceding results contain the continuity theorem of Kolmogorov. We state it in its classical and usual form although its various sharpening are deduced similarly.

Corollary 11.8. Let $X = (X_t)_{t\in[0,1]}$ be a separable random process indexed on $[0,1]$ such that for some $\alpha > 0$ and $p > 1$ and all s, t in $[0,1]$,

$$\mathbb{E}|X_s - X_t|^\alpha \le |s - t|^p \, .$$

Then X has almost surely bounded and continuous sample paths.

Proof. We apply Theorem 11.6 together with Remark 11.5. We distinguish between two cases. If $\alpha \geq p$, then

$$\|X_s - X_t\|_\alpha \leq d(s,t)$$

where d is the metric $d(s,t) = |s-t|^{p/\alpha}$ ($p/\alpha \leq 1$). As is obvious, $N([0,1], d; \varepsilon)$ is of the order $\varepsilon^{-\alpha/p}$ and since $p > 1$ the corresponding entropy integral with $\psi(x) = x^\alpha$ in Theorem 11.6 is finite and the conclusion follows in this case. When $\alpha \leq p$, then, for all s, t in $[0,1]$

$$\||X_s - X_t|^\gamma\|_p \leq d(s,t)$$

where $\gamma = \alpha/p \leq 1$ and here $d(s,t) = |s-t|$. Apply then again Theorem 11.6 with this time Remark 11.5; indeed, since $\gamma \leq 1$, $|X_s - X_t|^\gamma$ defines a random distance and here $N([0,1], d; \varepsilon) \sim \varepsilon^{-1}$. The proof is complete.

While the preceding evaluations seem quite easy, they however appear to be sharp in various instances. The case of Gaussian processes treated in Section 11.3 and the next chapter is a first example. Other examples concerning processes indexed by regular subsets of $\mathbb{R}^N$ were treated in the literature (see e.g. [Ha], [H-K], [Pi9], [Ib], [Ta12], etc.). They basically indicate that, if *every* random process satisfying a certain Lipschitz condition is almost surely bounded or continuous, then a corresponding integral is convergent. This is the natural formulation of the necessity results. They do not concern one single process but rather the whole family of processes satisfying the same incremental condition. Only in the Gaussian case the necessary conditions can concern one single process due to the comparison theorems. We shall come back to this in Chapter 12.

11.2. Regularity of random processes under majorizing measure conditions

One main feature (and weakness) of the entropy condition is that it gives some "weight" to *each* piece of T. This does not present any inconvenient if T is in some sense homogeneous; we will see how this is the case in Chapter 13 and how metric entropy is best possible (necessary) for some processes in such a homogeneous setting (cf. also the closing comments of Section 11.1). In general however, one has rather to think at some geometrical measure of T which takes into account the possible lack of homogeneity of T. One way to handle this is the concept of majorizing measure.

Given a pseudo-metric space (T, d) and a Young function ψ as before, say that a probability measure m on T is a *majorizing measure* for $(T, d; \psi)$, if

$$\gamma = \gamma_m(T, d; \psi) = \sup_{t \in T} \int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon < \infty \tag{11.7}$$

where $B(t, \varepsilon)$ is the open ball in the d-metric of center t and radius $\varepsilon > 0$. (Again we could use essentially equivalently closed balls.) We thus call (11.7) a majorizing measure condition as opposed to the entropy condition studied before. This definition clearly gives a way to take more into account local properties of the geometry of T. Our aim in this section will be to show how one can control random processes satisfying Lipschitz conditions under a majorizing measure condition as we did before with entropy and actually in a more efficient way.

We start with some remarks for a better comprehension of condition (11.7) and for comparison with the results of Section 11.1. If s and s' are two points in T with $d(s, s') = 2\eta > 0$, the open balls $B(s, \eta)$ and $B(s', \eta)$ are disjoint. Thus, if m is a probability measure on (T, d), one of these two balls, say $B(s, \eta)$, has a measure less than or equal to $1/2$. Therefore

$$\sup_{t \in T} \int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon \geq \int_0^\eta \psi^{-1}\left(\frac{1}{m(B(s,\varepsilon))}\right) d\varepsilon \geq \eta\psi^{-1}(2),$$

so that, as for entropy, if $D = D(T)$ is the diameter of (T, d),

$$\gamma_m(T, d; \psi) \geq \frac{1}{2}\psi^{-1}(2)D. \tag{11.8}$$

Also, if m is a majorizing measure satisfying (11.7), then (T, d) is totally bounded and actually

$$\sup_{\varepsilon > 0} \varepsilon\psi^{-1}(N(T, d; \varepsilon)) \leq 2\gamma_m(T, d; \psi). \tag{11.9}$$

The proof of this easy fact is already instructive on the way to use majorizing measures. Let $N(T, d; \varepsilon) \geq N$. There exist $t_1, \ldots, t_N$ such that $d(t_i, t_j) \geq \varepsilon$ for all $i \neq j$. By definition of $\gamma_m(T, d; \psi) = \gamma$, for each $i \leq N$,

$$\frac{\varepsilon}{2}\psi^{-1}\left(\frac{1}{m(B(t_i, \varepsilon/2))}\right) \leq \gamma,$$

that is $m(B(t_i, \varepsilon/2)) \geq [\psi(2\gamma/\varepsilon)]^{-1}$. Since the balls $B(t_i, \varepsilon/2)$, $i \leq N$, are disjoint and m is a probability, it follows that $\psi(2\gamma/\varepsilon) \geq N$ which is the result (11.9).

More important now is to observe that entropy conditions are stronger than majorizing measure conditions. That, is there is a probability measure m on T such that

$$\sup_{t\in T}\int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon \leq K\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon \tag{11.10}$$

where K is some numerical constant. This can be established in great generality (cf. [Ta12]) but we actually only prove it here in a special case. The general study of bounds on stochastic processes using majorizing measures runs indeed in many technicalities in which we decided not to enter here for the simplicity of the exposition. We will thus restrict this study to the case of a special class of Young functions ψ for which things become simpler. This restriction however does not hide the main idea and interest of the majorizing measure technique. As already mentioned, this study can be conducted in a rather large generality and we refer to [Ta12] where this program is performed.

For the rest of this paragraph, we hence assume that ψ is a Young function such that for some constant C and all $x,y\geq 1$,

$$\psi^{-1}(xy)\leq C(\psi^{-1}(x)+\psi^{-1}(y)) \quad \text{and} \quad \int_0 \psi^{-1}\left(\frac{1}{x}\right)dx<\infty\,. \tag{11.11}$$

This class covers the main examples we have in mind, namely the exponential Young functions $\psi_q(x)=\exp(x^q)-1$, $1\leq q<\infty$ (and also $\psi_\infty(x)=\exp(\exp\, x)-e$). For simplicity in the notations, let us assume moreover that $C=1$ in (11.11) (which is actually easily seen not to be a restriction).

Let us now show how in this case we may prove (11.10). Let, as usual, ℓ_0 be the largest integer with $2^{-\ell_0}\geq D$ where D is the diameter of (T,d). For every $\ell\geq\ell_0$, let $T_\ell\subset T$ denote the set of the centers of a minimal family such that the balls $B(t,2^{-\ell})$, $t\in T_\ell$, cover T. By definition, Card $T_\ell=N(T,d;2^{-\ell})$. Consider then the probability measure m on T given by

$$m=\sum_{\ell>\ell_0}2^{-\ell+\ell_0}N(T,d;2^{-\ell})^{-1}\sum_{t\in T_\ell}\delta_t$$

where δ_t is Dirac measure at t. Clearly, for every t and $\ell>\ell_0$,
$m(B(t,2^{-\ell}))\geq 2^{-\ell+\ell_0}N(T,d;2^{-\ell})^{-1}$. Hence

$$\begin{aligned}\int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon &\leq \sum_{\ell>\ell_0}2^{-\ell}\psi^{-1}\left(\frac{1}{m(B(t,2^{-\ell}))}\right)\\ &\leq \sum_{\ell>\ell_0}2^{-\ell}\psi^{-1}(2^{\ell-\ell_0}N(T,d;2^{-\ell}))\,.\end{aligned}$$

Using (11.11) this is then estimated by

$$\begin{aligned}\sum_{\ell>\ell_0} 2^{-\ell}\psi^{-1}(2^{\ell-\ell_0}) + \sum_{\ell>\ell_0} 2^{-\ell}\psi^{-1}(N(T,d;2^{-\ell}))\\ &\leq 2^{-\ell_0+1}\int_0^{1/2}\psi(x^{-1})dx + 2\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon\\ &\leq 4D\int_0^{1/2}\psi(x^{-1})dx + 2\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon\\ &\leq 2\left(1+2(\psi^{-1}(1))^{-1}\int_0^{1/2}\psi(x^{-1})dx\right)\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon\end{aligned}$$

and the announced claim follows. Note that the constant however depends on ψ.

Let us note that the preceding proof actually shows that when $\int_0^D \psi^{-1}(N(T,d;\varepsilon))d\varepsilon < \infty$, then m is a majorizing measure which satisfies (in addition to (11.10))

$$\lim_{\eta\to 0}\sup_{t\in T}\int_0^\eta \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon = 0\,. \tag{11.12}$$

This condition is the one which enters to obtain continuity properties of stochastic processes as opposed to (11.7) which deals with boundedness. We therefore sometimes speak of bounded, resp. continuous, majorizing measure conditions. This is another advantage of majorizing measures upon entropy, to be able to give weaker sufficient conditions for sample boundedness than for sample continuity.

Let us now enter the heart of the matter and show how majorizing measures are used to control processes. Our approach will be to associate to each majorizing measure a (non-unique) ultrametric distance δ finer than the original distance d and for which there still exists a majorizing measure. This ultrametric structure is at the basis of the understanding of majorizing measures and will appear as the key notion in the study of necessity in the next section. Alternatively, it can be seen as a way to *discretize* majorizing measures. This then allows to use chaining arguments exactly as with entropy conditions. This program is accomplished in Proposition 11.10 and Corollary 11.12 below. We however start with a simple lemma which allows to conveniently reduce to a finite index set T.

Lemma 11.9. Let (T,d) be a pseudo-metric space with diameter $D(T)$. Let m be a probability measure on (T,d) and recall we set

$$\gamma_m(T,d;\psi) = \sup_{t\in T}\int_0^{D(T)} \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon\,.$$

Then, if A is a finite (or compact) subset of T, there is a probability measure μ on (A,d) such that

$$\begin{aligned}\gamma_\mu(A,d;\psi) &= \sup_{t\in A}\int_0^{D(A)} \psi^{-1}\left(\frac{1}{\mu(B_A(t,\varepsilon))}\right) d\varepsilon \\ &\le 2\sup_{t\in A}\int_0^{D(A)/2} \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon\end{aligned}$$

where $D(A)$ is the diameter of (A,d) and $B_A(t,\varepsilon)$ the ball *in* A of center t and radius $\varepsilon > 0$. In particular, $\gamma_\mu(A,d;\psi) \le 2\gamma_m(T,d;\psi)$.

Proof. For t in T, take $\varphi(t)$ in A with

$$d(t,\varphi(t)) = d(t,A) = \inf\{d(t,y)\,;\ y\in A\}\,.$$

Set $\mu = \varphi(m)$, so μ is supported by A. Fix x in A. For t in T, we have $d(t,\varphi(t)) = d(t,A) \le d(t,x)$ and thus $d(x,\varphi(t)) \le 2d(x,t)$. It follows that $\varphi(B(x,\varepsilon)) \subset B_A(x,2\varepsilon)$ and thus $\mu(B_A(x,2\varepsilon)) \ge m(B(x,\varepsilon))$. The proof is easily completed.

Recall a metric space (U,δ) is ultrametric if δ satisfied the improved triangle inequality

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

The main feature of ultrametric spaces is that two balls of the same radius are either disjoint or equal. The next proposition deals with the nice functions ψ satisfying (11.11); we recall however that, at the expense of (severe) complications, a similar study can be driven in the general setting (cf. [Ta12]).

Proposition 11.10. Let ψ satisfying (11.11) and let (T,d) be a finite metric space with diameter D. Let m be a probability measure on (T,d) and recall we set

$$\gamma_m(T,d;\psi) = \sup_{t\in T}\int_0^{D} \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon\,.$$

There exist an ultrametric distance δ on T such that $d(s,t)\le\delta(s,t)$ for all s,t in T and a probability measure μ on (T,δ) such that

$$\gamma_\mu(T,\delta;\psi) \le K_\psi \gamma_m(T,d;\psi)$$

where K_ψ is a constant depending only on ψ.

Proof. Let ℓ_0 be the largest integer ℓ such that $4^{-\ell} \geq D$ and ℓ_1 be the smallest one such that the balls $B(t, 4^{-\ell})$ of center t and radius $4^{-\ell}$ in the metric d are reduced to exactly one point. Assume $T = (t_i)$. Set $T_{\ell_1,i} = \{t_i\}$ for every i. For every $\ell = \ell_1 - 1, \ldots, \ell_0$, we construct by induction on $i \geq 1$, points $x_{\ell,i}$ and subsets $T_{\ell,i}$ of T as follows: setting $T_{\ell,0} = \emptyset$,

$$m(B(x_{\ell,i}, 4^{-\ell})) = \max\{m(B(x, 4^{-\ell}))\,;\ x \notin \bigcup_{j<i} T_{\ell,j}\}\,,$$

$$T_{\ell,i} = \bigcup\{T_{\ell+1,k}; T_{\ell+1,k} \cap B(x_{\ell,i}, 4^{-\ell+1}) \neq \emptyset\,,\ \forall j < i\,,\ T_{\ell+1,k} \not\subset T_{\ell,j}\}\,.$$

Define then $\delta(s,t) = 4^{-\ell+2}$ where ℓ is the largest integer such that s and t belong to the same $T_{\ell,i}$ for some i. δ is clearly an ultrametric distance. By decreasing induction on ℓ, it is easily verified that the diameter of each set $T_{\ell,i}$ is less than $4^{-\ell+2}$. It then clearly follows by definition of δ that $d(s,t) \leq \delta(s,t)$ for all s, t in T.

For each ℓ, $(T_{\ell,i})$ forms the family of the δ-balls of radius $4^{-\ell+2}$. By construction, the balls $B(x_{\ell,i}, 4^{-\ell})$ when i varies are disjoint so that $\sum_i m(B(x_{\ell,i}, 4^{-\ell})) \leq 1$. Let $t_{\ell,i}$ be a fixed point in $T_{\ell,i}$. Consider

$$\mu' = \sum_{\ell=\ell_0}^{\ell_1} 4^{-\ell+\ell_0+1} \sum_i m(B(x_{\ell,i}, 4^{-\ell}))\delta_{t_{\ell,i}}$$

where δ_t is Dirac measure at t. There is a probability measure $\mu \geq \mu'$. If $t \in T_{\ell,i}$, note that, by construction,

$$\mu(T_{\ell,i}) \geq 4^{-\ell+\ell_0-1} m(B(t, 4^{-\ell}))\,. \tag{11.13}$$

We evaluate $\gamma_\mu(T, \delta; \psi)$ using (11.13) and the properties of ψ. Let t be fixed, $t \in T_{\ell,i}$. By (11.13) and (11.11),

$$\sum_{\ell=\ell_0}^{\ell_1} 4^{-\ell}\psi^{-1}\left(\frac{1}{\mu(T_{\ell,i})}\right) \leq \sum_{\ell \geq \ell_0} 4^{-\ell}\psi^{-1}(4^{\ell-\ell_0+1}) + \sum_{\ell \geq \ell_0} 4^{-\ell}\psi^{-1}\left(\frac{1}{m(B(t, 4^{-\ell}))}\right)\,.$$

By definition of ℓ_0 (and (11.8)), this easily implies the conclusion. The proof is complete.

Remark 11.11. If (T, δ) is ultrametric and ψ satisfies (11.11), we may observe the following from the preceding proof: for every ℓ, let $\mathcal{B}_\ell$ be the family of the balls B of radius $2^{-\ell}$ (or $4^{-\ell}$ to agree with the proof of Proposition 11.10); then, perhaps more important than the probability measure μ we constructed (although it is actually equivalent), is the datum of a family of weights $\alpha(B, \ell) \geq 0$, $B \in \mathcal{B}_\ell$, such that

$\sum_{B \in \mathcal{B}_\ell} \alpha(B,\ell) \leq 1$ (the measures $m(B(x_{\ell,i},4^{-\ell}))$ in the preceding proof). Further, Proposition 11.10 may be expressed in the following "discretized" formulation. Denote by ℓ_0 the largest integer ℓ such that $2^{-\ell} \geq D$; then, there exists, for all $\ell \geq \ell_0$, finite sets T_ℓ in T and maps $\pi_\ell : T \to T_\ell$ such that $\pi_{\ell-1} \circ \pi_\ell = \pi_{\ell-1}$ and $d(t,\pi_\ell(t)) \leq 2^{-\ell}$ for every t and ℓ, and a *discrete* probability measure μ on $\{T_\ell; \ell \geq \ell_0\}$ satisfying

$$\sup_{t\in T} \sum_{\ell \geq \ell_0} 2^{-\ell} \psi^{-1}\left(\frac{1}{\mu(\{\pi_\ell(t)\})}\right) \leq K_\psi \gamma_m(T,d;\psi)$$

where K_ψ only depends on ψ. Indeed, if δ is the ultrametric structure obtained in Proposition 11.10, for every $\ell \geq \ell_0$ denote by $\mathcal{B}_\ell$ the family of the δ-balls of radius $2^{-\ell}$. For every $t \in T$, there is a unique element B of $\mathcal{B}_\ell$ with $t \in B$. Let then $\pi_\ell(t)$ be one fixed point of B and let $\mu_\ell(\{\pi_\ell(t)\}) = \mu(B)$. The probability measure $\sum_{\ell \geq \ell_0} 2^{-\ell+\ell_0+1} \mu_\ell$ fulfills the conditions of the claim.

Provided with the preceding results, we now present sufficient conditions in terms of majorizing measures for a random process to be almost surely bounded or continuous. The results are the analogs (actually improvements), in the setting of functions ψ satisfying (11.11), of Theorems 11.1, 11.2 and 11.6 dealing with entropy. We first establish the main result for a *general* Young function ψ, in the case of an ultrametric index set, and then deduce the general case from Proposition 11.10 for a function ψ satisfying (11.11). As an alternate approach, one may use the preceding discretization (Remark 11.11) that, however, does not really clarify the steps in which the property (11.11) of ψ is used. We refer to [Ta12] for more details on this point and a more general study for arbitrary Young functions ψ.

Proposition 11.12. Let ψ be an arbitrary Young function and let $X = (X_t)_{t\in T}$ be a random process in $L_1 = L_1(\Omega,\mathcal{A},\mathbb{P})$ indexed by a finite ultrametric space (T,δ) such that, for all s,t in T and all measurable sets A in Ω,

$$\int_A |X_s - X_t| d\mathbb{P} \leq \delta(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right).$$

Then, for any probability measure μ on (T,δ) and any measurable set A,

$$\int_A \sup_{s,t\in T} |X_s - X_t| d\mathbb{P} \leq K\mathbb{P}(A) \sup_{t\in T} \int_0^D \psi^{-1}\left(\frac{1}{\mathbb{P}(A)\mu(B(t,\varepsilon))}\right) d\varepsilon$$

where K is numerical and D is the diameter of (T,δ).

Proof. Set $T = \{t_i\}$ and let A in $\mathcal{A}$. Let (A_i) be a measurable partition of A such that, on A_i,

$$\sup_{t\in T} |X_t - X_x| = |X_{t_i} - X_x|$$

where x is one (arbitrary) fixed point of T. Thus,

$$\int_A \sup_{t \in T} |X_t - X_x| d\mathbb{P} = \sum_i \int_{A_i} |X_{t_i} - X_x| \, d\mathbb{P} .$$

Let ℓ_0 be the largest integer ℓ such that $2^{-\ell} \geq D$ and, for $\ell \geq \ell_0$, denote by $\mathcal{B}_\ell$ the family of the balls of radius $2^{-\ell}$. For every B in $\mathcal{B}_\ell$, we fix $x(B) \in B$, and take $x(T) = x$. Further, we let $\pi_\ell(t) = x(B(t, 2^{-\ell}))$, $t \in T$, $\ell \geq \ell_0$. The usual chaining identity then yields, for every t in T,

$$|X_t - X_x| \leq \sum_{\ell > \ell_0} |X_{\pi_\ell(t)} - X_{\pi_{\ell-1}(t)}| .$$

For every B in $\mathcal{B}_\ell$, set $A_B = \cup\{A_i ; t_i \in B\}$. Thus, we can write

$$\begin{aligned}
\int_A \sup_{t \in T} |X_t - X_x| d\mathbb{P} &\leq \sum_i \sum_{\ell > \ell_0} \int_{A_i} |X_{\pi_\ell(t_i)} - X_{\pi_{\ell-1}(t_i)}| d\mathbb{P} \\
&= \sum_{\ell > \ell_0} \sum_{B \in \mathcal{B}_\ell} \sum_{t_i \in B} \int_{A_i} |X_{\pi_\ell(t_i)} - X_{\pi_{\ell-1}(t_i)}| d\mathbb{P} \\
&= \sum_{\ell > \ell_0} \sum_{B \in \mathcal{B}_\ell} \int_{A_B} |X_{x(B)} - X_{x(\overline{B})}| d\mathbb{P}
\end{aligned}$$

where $B \subset \overline{B}$ and $\overline{B} \in \mathcal{B}_{\ell-1}$. Hence, from the hypothesis, since $\delta(x(B), x(\overline{B})) \leq 2^{-\ell+1}$,

$$\int_A \sup_{t \in T} |X_t - X_x| d\mathbb{P} \leq \sum_{\ell > \ell_0} \sum_{B \in \mathcal{B}_\ell} 2^{-\ell+1} \mathbb{P}(A_B) \psi^{-1} \left(\frac{1}{\mathbb{P}(A_B)} \right) .$$

Let now μ be a probability measure on (T, δ) and set

$$M = \sup_{t \in T} \int_0^D \psi^{-1} \left(\frac{1}{\mathbb{P}(A) \mu(B(t, \varepsilon))} \right) d\varepsilon$$

so that, for every t in T,

$$\sum_{\ell \geq \ell_0} 2^{-\ell} \psi^{-1} \left(\frac{1}{\mathbb{P}(A) \mu(B(t, 2^{-\ell}))} \right) \leq 2M .$$

Integrating with respect to μ yields

$$\sum_{\ell \geq \ell_0} 2^{-\ell} \sum_{B \in \mathcal{B}_\ell} \mu(B) \psi^{-1} \left(\frac{1}{\mathbb{P}(A) \mu(B)} \right) \leq 2M ,$$

while integrating with respect to the measure ν on T such that $\nu(\{t_i\}) = \mathbb{P}(A_i)$ yields

$$\sum_{\ell\geq\ell_0} 2^{-\ell} \sum_{B\in\mathcal{B}_\ell} \mathbb{P}(A_B)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)\mu(B)}\right) \leq 2\mathbb{P}(A)M\,.$$

We next observe the following. If $\mathbb{P}(A_B) \leq \mathbb{P}(A)\mu(B)$,

$$\mathbb{P}(A_B)\psi^{-1}\left(\frac{1}{\mathbb{P}(A_B)}\right) \leq \mathbb{P}(A)\mu(B)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)\mu(B)}\right)$$

since $\psi(u) \leq u\psi'(u)$ which shows that the function $u\psi^{-1}(1/u)$ is increasing. If $\mathbb{P}(A_B) \geq \mathbb{P}(A)\mu(B)$, we have simply that

$$\mathbb{P}(A_B)\psi^{-1}\left(\frac{1}{\mathbb{P}(A_B)}\right) \leq \mathbb{P}(A_B)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)\mu(B)}\right)\,.$$

Assembling this observation with what we obtained previously yields

$$\int_A \sup_{t\in T} |X_t - X_x| d\mathbb{P} \leq 8\mathbb{P}(A)M$$

from which the conclusion follows. Proposition 11.12 is established.

Together with Lemma 11.9 and Proposition 11.10, the preceding basic result yields the following general theorem for functions ψ satisfying (11.11).

Theorem 11.13. Let ψ be a Young function satisfying (11.11) and let $X = (X_t)_{t\in T}$ be a random process in $L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$ indexed by the pseudo-metric space (T, d) such that, for all s, t in T and all measurable sets A in Ω ,

$$\int_A |X_s - X_t| d\mathbb{P} \leq d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right)\,.$$

Then, for any probability measure m on (T, d) and any measurable set A ,

$$\int_A \sup_{s,t\in T} |X_s - X_t| d\mathbb{P} \leq K_\psi \mathbb{P}(A)\left(D\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + \sup_{t\in T}\int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon\right)$$

where $D = D(T)$ is the diameter of (T, d) and K_ψ only depends on ψ . In particular,

$$\mathbb{E}\sup_{s,t\in T} |X_s - X_t| \leq K_\psi \sup_{t\in T}\int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon\,.$$

Let us mention that the various comments next to Theorem 11.2 about integrability and tail behavior of the supremum of the processes under study can be repeated similarly from Theorem 11.13; simply replace the entropy integral by the corresponding majorizing measure integral. In particular, as an analog of (11.4), we have in the setting of Theorem 11.13 that for every $u > 0$

$$\mathbb{P}\left\{\sup_{s,t\in T}|X_s - X_t| > K_\psi(\gamma + u)\right\} \le \left(\psi\left(\frac{u}{D}\right)\right)^{-1} \tag{11.14}$$

where

$$\gamma = \gamma_m(T, d; \psi) = \sup_{t\in T}\int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon\,.$$

The next result concerns continuity of random processes under majorizing measure conditions. It is the analog of Theorem 11.6 and the proof actually simply needs to adapt appropriately the proof of Theorem 11.6. As announced, the majorizing measure condition has to be strengthened into (11.12).

Theorem 11.14. Let ψ be a Young function satisfying (11.11) and let $X = (X_t)_{t\in T}$ be a random process in $L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$ such that for all s, t in (T, d) and all measurable sets A in Ω,

$$\int_A |X_s - X_t| d\mathbb{P} \le d(s,t)\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right).$$

Assume there is a probability measure m on (T, d) such that

$$\lim_{\eta\to 0}\sup_{t\in T}\int_0^\eta \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon = 0\,.$$

Then X admits a version $\widetilde{X}$ with almost all sample paths (uniformly) continuous on (T, d). Moreover, $\widetilde{X}$ satisfies the following property: for each $\varepsilon > 0$, there exists $\eta > 0$ depending only on the preceding limit, i.e. only on T, d, ψ, m but not on X, such that

$$\mathbb{E}\sup_{d(s,t)<\eta}|\widetilde{X}_s - \widetilde{X}_t| < \varepsilon\,.$$

Proof. We simply sketch the steps of the proof of Theorem 11.6 in the majorizing measure setting. For each $\eta > 0$, set

$$\gamma(\eta) = \sup_{t\in T}\int_0^\eta \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon$$

so that $\lim_{\eta\to 0}\gamma(\eta) = 0$. Fix $\eta > 0$. If A is a finite subset of T, we know from Lemma 11.9, or more precisely its proof, that there is a probability measure μ on (A, d) such that

$$\sup_{t\in T}\int_0^\eta \psi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right) d\varepsilon \le 2\gamma(\eta/2) \le 2\gamma(\eta)\,.$$

This observation allows to assume that T is finite in what follows. Let ℓ be the largest integer such that $2^{-\ell} \geq D$. If (T,d) is ultrametric, the proof of Theorem 11.13 and its notations yield

$$\mathbb{E} \sup_{t \in T} |X_t - X_{\pi_\ell(t)}| \leq K\gamma(\eta)$$

for some (numerical) K. We can then simply repeat in this case the argument leading to (11.6) in the proof of Theorem 11.6 to get that, for every $\tau > 0$,

$$\mathbb{E} \sup_{d(s,t)<\tau} |X_s - X_t| \leq \tau\psi^{-1}(N(T,d;2^{-\ell})^2) + K\gamma(\eta). \tag{11.15}$$

Proposition 11.10, adapted to the present case, i.e. with η replacing the diameter, allows then to extend this property to the case of a general index set T, K depending then on ψ. Since (T,d) is totally bounded under the majorizing measure condition, the proof of Theorem 11.14 is completed exactly as the one of Theorem 11.6.

Note that the precise dependence of η on ε in the last assertion of Theorem 11.14 can be made explicit from (11.15). This will not be required in the sequel so that we only gave the statement that will be sufficient in our applications. This will be in particular the case for the majorizing measure versions of Corollary 11.7 which we need not state since completely similar. Note further, that a deviation inequality of the type (11.14) may be obtained for supremum over $d(s,t) < \eta$ from (11.15). We leave the details to the interested reader.

Various remarks developed in Section 11.1 in the context of entropy apply similarly in the setting of majorizing measure conditions. This is the case for example with Remark 11.5 which we need not repeat here; we however use it freely below. This is also the case for Remark 11.4 which might be worthwile to detail in this context. Here is its analog.

Remark 11.15. In the setting of Theorem 11.13, assume the process $X = (X_t)_{t\in T}$ satisfies the following weaker assumption: for every $t \in T$ and integer ℓ, and every measurable set A and s in the ball of center t and radius $2^{-\ell}$,

$$\int_A |X_s - X_t| d\mathbb{P} \leq 2^{-\ell}\mathbb{P}(A)\psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + M_\ell(t)\mathbb{P}(A) \tag{11.16}$$

where $(M_\ell(t))$ is a sequence of positive numbers such that

$$\sup_{t\in T} \sum_\ell M_\ell(t) < \infty.$$

Then, the conclusion of Theorem 11.13 holds similarly, i.e. under the majorizing measure condition, X is almost surely bounded. The quantitative bounds of course involve then the preceding quantity. In case of Theorem 11.14 dealing with continuity, the condition on $(M_\ell(t))$ has to be strengthened into

$$\lim_{\ell_0\to\infty} \sup_{t\in T} \sum_{\ell\geq\ell_0} M_\ell(t) = 0\,.$$

As for Remark 11.4, this extension to processes satisfying (11.16) follows directly from the proof of Theorem 11.13. We will see in the next paragraph how these simple observations can be rather useful in various applications.

11.3. Examples of applications

In the last paragraph of this chapter, we present some (rather important) examples for which the preceding results can be applied. They concern Gaussian and Rademacher processes and their corresponding chaos processes. In particular, the sufficient conditions we describe in order a Gaussian process be sample bounded or continuous may be considered as the first part of the study of the regularity of Gaussian processes. The second part devoted to necessity is the object of the next chapter.

Before entering these examples, let us briefly indicate an elementary but convenient remark. We deal here with the Young functions $\psi_q(x) = \exp(x^q) - 1$, $1 \leq q < \infty$. We have that $\psi_q^{-1}(x) = (\log(1+x))^{1/q}$. The point of this observation is that we can deal equivalently, in the entropy or majorizing measure conditions, with the functions $(\log x)^{1/q}$, $x \geq 1$. For the entropy condition, it is clear that

$$\int_0^\infty (\log N(T,d;\varepsilon))^{1/q} d\varepsilon = \int_0^D (\log N(T,d;\varepsilon))^{1/q} d\varepsilon \geq (\log 2)^{1/q} D$$

and thus

$$\int_0^D \psi_q^{-1}(N(T,d;\varepsilon)) d\varepsilon \leq 3 \int_0^\infty (\log N(T,d;\varepsilon))^{1/q} d\varepsilon\,. \tag{11.17}$$

The reverse inequality (with constant 1) is obvious. Note that we can write indifferently, with the function $(\log x)^{1/q}$, the integral up to D or ∞ since $N(T,d;\varepsilon) = 1$ if $\varepsilon \geq D$. Similarly, for the majorizing measure condition, we have that for every probability measure m on T

$$\sup_{t\in T} \int_0^D \psi_q^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right) d\varepsilon \leq 4 \sup_{t\in T} \int_0^\infty \left(\log \frac{1}{m(B(t,\varepsilon))}\right)^{1/q} d\varepsilon \tag{11.18}$$

and trivially also a reverse inequality. A similar property of course also holds for continuous majorizing measure conditions. We can further deal with $\psi_\infty(x) = \exp(\exp\, x) - e$ and replace $\psi_\infty^{-1}(x)$ by $\log(1+\log x)$, or the more commonly used $\log^+ \log x$ (provided the diameter is taken in account in the inequalities). Accordingly, we use freely below either $\psi_q^{-1}(x)$ or $(\log x)^{1/q}$ depending on the context and/or historical references; *actually*, $(\log x)^{1/q}$ will be used most often.

Let $X = (X_t)_{t \in T}$ be a Gaussian process. Recall from Chapter 3 that by this we mean that the distribution of any finite dimensional random vector $(X_{t_1}, \dots, X_{t_N})$, $t_i \in T$, is Gaussian. The distribution of X is therefore completely determined by its covariance structure $\mathbb{E} X_s X_t$, $s, t \in T$. The question of course raises to know under which condition(s) on its covariance structure, the Gaussian process X is (or admits a version which is) almost surely bounded or continuous.

Set

$$d_X(s,t) = \|X_s - X_t\|_2 , \qquad s,t \in T .$$

The knowledge of the covariance structure implies a complete knowledge of this L_2-metric d_X, and conversely, at least if $\mathbb{E} X_t^2$, $t \in T$, is known. d_X is therefore a natural pseudo-metric on the index set T associated to the Gaussian process X. According to the previous sections, we might try to know how the "geometry" of (T, d_X) describes boundedness or continuity properties of the sample paths of X. In this order of ideas, the results of the preceding sections provide a rather precise description of the situation (they will actually be shown to be best possible in the next chapter). To start with however, let us first mention that the comparison theorems of Section 3.3 can also be efficient in this study. Indeed, if X and Y are two Gaussian processes such that $d_Y(s,t) \le d_X(s,t)$ for all s,t, and if X has nice regularity properties, boundedness or continuity of the sample paths, then by the results of Section 3.3, these can be "transferred" to Y. This is clear for boundedness by the integrability properties of supremum of Gaussian processes. For continuity, we can use the following lemma.

Lemma 11.16. Let $Y = (Y_t)_{t \in T}$ be a Gaussian process and d be a metric on T such that $d_Y(s,t) \le d(s,t)$. Then, for every $\eta > 0$,

$$\mathbb{E} \sup_{d(s,t) < \eta} |Y_s - Y_t| \le K \big(\sup_{t \in T} \mathbb{E} \sup_{d(s,t) < \eta} |Y_s - Y_t| + \eta (\log N(T,d;\eta))^{1/2} \big)$$

where K is numerical.

Proof. Fix $\eta > 0$ and let $N = N(T,d;\eta)$ (assumed to be finite and larger than 2). Let $U = (u_1,\dots,u_N)$ in T be such that the d-balls of radius η and center u_i cover T. Clearly

$$\sup_{d(s,t)<\eta} |Y_s - Y_t| \le 2\max_{u\in U}\Big(\sup_{d(t,u)<\eta} |Y_t - Y_u|\Big) + \max_{\substack{u,v\in U\\ d(u,v)<3\eta}} |Y_u - Y_v| \,.$$

By (3.6) and the fact that $d_Y(t,u) \le d(t,u)$, we have

$$\mathbb{E}\max_{u\in U}\Big(\sup_{d(t,u)<\eta} |Y_t - Y_u|\Big) \le 2\max_{u\in U}\mathbb{E}\sup_{d(t,u)<\eta} |Y_t - Y_u| + 3\eta(\log N)^{1/2} \,.$$

Similarly, by (3.13),

$$\mathbb{E}\max_{\substack{u,v\in U\\ d(u,v)<3\eta}} |Y_u - Y_v| \le 3\eta(\log N^2)^{1/2}$$

and the lemma is proved.

The preceding claim then follows immediately from this lemma when $d = d_X$ using Corollary 3.19. Note that the Gaussian comparison properties are only used in this approach through Corollary 3.19, that is Sudakov's minoration.

Let therefore X be a Gaussian process with associated pseudo-metric d_X. The integrability properties of Gaussian variables indicate that, for all s,t in T,

$$\|X_s - X_t\|_{\psi_2} \le 2d_X(s,t) \,.$$

We are thus *immediately* in the setting of processes satisfying a Lipschitz condition as studied in the previous sections. The next two statements are then direct consequences of the results obtained there. The first one which deals with entropy is known as Dudley's theorem. The numerical constant has no reason to be sharp.

Theorem 11.17. Let $X = (X_t)_{t\in T}$ be a Gaussian process. then

$$\mathbb{E}\sup_{t\in T} X_t \le 24\int_0^\infty (\log N(T,d_X;\varepsilon))^{1/2} d\varepsilon \,.$$

Further, if this entropy integral is convergent, X has a version with almost all sample paths (uniformly) continuous on (T,d_X).

Theorem 11.18. Let $X = (X_t)_{t\in T}$ be a Gaussian process. Then, for some numerical constant K and any probability measure m on (T, d_X),

$$\mathbb{E}\sup_{t\in T} X_t \le K \sup_{t\in T} \int_0^\infty \left(\log \frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon .$$

Further, if m satisfies

$$\lim_{\eta\to 0} \sup_{t\in T} \int_0^\eta \left(\log \frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon = 0 ,$$

X admits a version with almost all sample paths (uniformly) continuous on (T, d_X).

Note that if we are asked for continuity properties of a Gaussian process X with respect to another metric d for which T is compact, we need simply assume in addition that d_X is continuous on (T, d), in other words that X is continuous in L_2 (or in probability). Actually, if (T, d) is metric compact, a Gaussian process $X = (X_t)_{t\in T}$ is continuous on (T, d) if and only if it is continuous on (T, d_X) and d_X is continuous on (T, d). Sufficiency is obvious. If X is d-continuous, so is d_X. For $\eta > 0$, let $A_\eta = \{(s,t) \in T \times T\,;\ d_X(s,t) \le \eta\}$. This is a closed set in $T \times T$ and $\bigcap_{\eta>0} A_\eta = A_0$. Fix $\varepsilon > 0$. By compactness, there is $\eta > 0$ and a finite set $A' \subset A_0$ such that whenever $(s,t) \in A_\eta$, there exists $(s', t') \in A'$ with $d(s, s')$, $d(t, t') \le \varepsilon$. We have by the triangle inequality

$$|X_s - X_t| \le |X_s - X_{s'}| + |X_{s'} - X_{t'}| + |X_{t'} - X_t| .$$

Since $(s', t') \in A_0$, $X_{s'} = X_{t'}$ with probability one. It follows that

$$\mathbb{E} \sup_{d_X(s,t)\le\eta} |X_s - X_t| \le 2\mathbb{E} \sup_{d(s,t)\le\varepsilon} |X_s - X_t| .$$

By the integrability properties of Gaussian random vectors, the right side of this inequality goes to 0 with ε, and thus the left side with η. It follows that X is d_X-uniformly continuous.

Recall that Theorem 11.18 is more general than Theorem 11.17 ((11.10)). It is remarkable that these two theorems drawn from the rather general results of the previous sections, which apply to large classes of processes, are sharp in this Gaussian setting. As will be discussed in Chapter 12, Theorem 11.17 may indeed be compared to Sudakov's minoration (Theorem 3.18) and, actually, the existence of a majorizing measure in Theorem 11.18 will be shown to be *necessary* for X to be bounded or continuous.

Closely related to Gaussian processes are Rademacher processes. Following Chapter 4, we say that a process $X = (X_t)_{t\in T}$ is a Rademacher process if there exists a sequence $(x_i(t))$ of functions on T such that for every t, $X_t = \sum_i \varepsilon_i x_i(t)$ assumed to converge almost surely (i.e., $\sum_i x_i(t)^2 < \infty$). Recall that (ε_i) denotes a Rademacher sequence, i.e. a sequence of independent random variables taking the values ± 1 with probability $1/2$. The basic observation is that according to the subgaussian inequality (4.1), as in the Gaussian case,

$$\|X_s - X_t\|_{\psi_2} \le 5\|X_s - X_t\|_2$$

for all s, t in T. The preceding Theorems 11.17 and 11.18 therefore also apply to Rademacher processes. In particular, for any probability measure m on T equipped with the pseudo-metric $d(s,t) = (\sum_i |x_i(s) - x_i(t)|^2)^{1/2}$, we have

$$\mathbb{E}\sup_t \sum_i \varepsilon_i x_i(t) \le K \sup_t \int_0^\infty \left(\log \frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon \tag{11.19}$$

for some numerical constant K.

The Gaussian results actually apply to the general class of the so-called subgaussian processes. A centered process $X = (X_t)_{t\in T}$ is said to be *subgaussian* with respect to a metric or pseudo-metric d on T if, for all s, t in T and every λ in $\mathbb{R}$,

$$\mathbb{E}\exp \lambda(X_s - X_t) \le \exp\left(\frac{\lambda^2}{2} d(s,t)^2\right). \tag{11.20}$$

Gaussian and Rademacher processes are subgaussian with respect to (a multiple of) their associated L_2-metric. If X is subgaussian with respect to d, by Chebyshev's inequality, for every $\lambda, u > 0$,

$$\mathbb{P}\{|X_s - X_t| > u\} \le 2\exp\left(-\lambda u + \frac{\lambda^2}{2} d(s,t)^2\right).$$

Minimizing over λ ($\lambda = u/d(s,t)^2$) yields

$$\mathbb{P}\{|X_s - X_t| > u\} \le 2\exp(-u^2/2d(s,t)^2)$$

for all $u > 0$. Hence, for all s, t in T,

$$\|X_s - X_t\|_{\psi_2} \le 5d(s,t).$$

This is the property we use on subgaussian processes. Actually, elementary computations on the basis of a series expansion of the exponential function shows that if Z is a real mean zero random variable such that $\|Z\|_{\psi_2} \le 1$, then, for all $\lambda \in \mathbb{R}$,

$$\mathbb{E} \exp \lambda Z \le \exp C^2 \lambda^2$$

where C is numerical. Therefore, changing d by some multiple of it shows that the subgaussian definition (11.20) is equivalent to say that $\|X_s - X_t\|_{\psi_2} \le d(s,t)$ for all s, t, or also $\mathbb{P}\{|X_s - X_t| > u\} \le C \exp(-u^2/Cd(s,t)^2)$ for some constant C and all $u > 0$. We use this freely below.

As announced, Theorem 11.17 and 11.18 apply similarly to subgaussian processes. What we will actually use in applications concerning subgaussian processes (in Chapter 14) is the majorizing measure version of Corollary 11.7 for families of subgaussian processes. Let us record at this stage the following statement for further reference.

Proposition 11.19. Assume there is a probability measure m on (T, d) such that

$$\lim_{\eta \to 0} \sup_{t \in T} \int_0^\eta \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/2} d\varepsilon = 0 .$$

Then, for each $\varepsilon > 0$, there exists $\eta > 0$ such that for every (separable) process $X = (X_t)_{t \in T}$ which is subgaussian with respect to d,

$$\mathbb{E} \sup_{d(s,t) < \eta} |X_s - X_t| < \varepsilon .$$

Turning back to Rademacher processes $X = (X_t)_{t \in T}$, $X_t = \sum_i \varepsilon_i x_i(t)$, we have that $\|X_s - X_t\|_2 = (\sum_i |x_i(s) - x_i(t)|^2)^{1/2}$. In Section 4.1, we learned estimates of $\|X_s - X_t\|_{\psi_q}$, $2 < q \le \infty$, for other metrics than this ℓ_2-metric, namely $\ell_{p,\infty}$-metrics, $\|(x_i(s) - x_i(t))\|_{p,\infty}$ where p is the conjugate of q. These results yield then further entropy or majorizing measure bounds of Rademacher processes in terms of ψ_q and these metrics. In particular, we have the following statement. Since $\|(x_i(s) - x_i(t))\|_{p,\infty}$ need not be distances in general, the proof is actually carried over with the true metrics $\|X_s - X_t\|_{\psi_q}$.

Lemma 11.20. Let $X = (X_t)_{t \in T}$ be a Rademacher process, $X_t = \sum_i \varepsilon_i x_i(t)$, $t \in T$. For $1 \le p < 2$, let $d_{p,\infty}(s,t) = \|(x_i(s) - x_i(t))\|_{p,\infty}$, s, t in T. Then, for any probability measure m on $(T, d_{p,\infty})$,

$$\mathbb{E} \sup_t \sum_i \varepsilon_i x_i(t) \le K_p \sup_{t \in T} \int_0^\infty \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/q} d\varepsilon$$

where K_p only depends on p and $q = p/p-1$ is the conjugate of $p > 1$; when $p = 1$,

$$\mathbb{E} \sup_t \sum_i \varepsilon_i x_i(t) \leq K \sup_{t \in T} \int_0^\infty \log \left(1 + \log \frac{1}{m(B(t,\varepsilon))} \right) d\varepsilon .$$

After these classical and important examples, we now investigate some more specialized ones. They concern Gaussian processes with vector values and Gaussian chaos. We say Gaussian but actually these applications are exactly the same for Rademacher processes on the basis of the corresponding results in Chapter 4 and the previous discussion. We leave it to the interested reader to translate the results to the Rademacher case.

One of the main interests of these applications is the use of Remarks 11.4 and 11.15 concerning processes satisfying (11.5) or (11.16). In order to put the results in a clearer perspective, we decided to present the first application to vector valued Gaussian processes using the tool of entropy and Remark 11.4, and the second one using majorizing measures. It is indeed fruitful to first analyze the questions in terms of entropy. Of course, Theorem 11.21 below also holds under the corresponding majorizing measure conditions.

We do not seek the greatest generality in the definition of processes with vector values. Assume simply we are given a separable Banach space B and a family $X = (X_t)_{t \in T}$ of Borel random variables X_t with values in B indexed by T. X is Gaussian if each finite sample $(X_{t_1}, \ldots, X_{t_N})$, $t_i \in T$, is Gaussian in B^N. We may then ask similarly for almost sure boundedness or continuity properties of the sample paths of $X = (X_t)_{t \in T}$ in B. As a first simple observation, set, for all s,t in T,

$$d_X(s,t) = \|X_s - X_t\|_2 = (\mathbb{E}\|X_s - X_t\|^2)^{1/2} .$$

From (3.5), we have that

$$\|X_s - X_t\|_{\psi_2} \leq 8 d_X(s,t) .$$

We can then make use of Remark 11.5 to realize that if

$$\int_0^\infty (\log N(T, d_X; \varepsilon))^{1/2} d\varepsilon < \infty , \tag{11.21}$$

then X has a version with almost all sample paths bounded and continuous on (T, d_X).

The metric d_X in (3.5) is however too strong. This inequality (3.5) is indeed a consequence of the precise deviation inequalities for norms of Gaussian random vectors in the form of Lemma 3.1. These involve, besides what can be called the "strong" parameter d_X , a "weak" parameter. Let us set indeed,

$$\sigma_X(s,t) = \sigma(X_s - X_t) = \sup_{\|f\|\le 1} (\mathbb{E} f^2(X_s - X_t))^{1/2} ,$$

$s, t \in T$. Then we know from Lemma 3.1 that, for all s, t in T , and for all $u > 0$,

$$\mathbb{P}\{\|X_s - X_t\| > 2d_X(s,t) + u\sigma_X(s,t)\} \le \exp(-u^2/2) .$$

This is (basically) equivalent to say that

$$\|(\|X_s - X_t\| - 2d_X(s,t))^+\|_{\psi_2} \le 2\sigma_X(s,t)$$

from which it follows that for every measurable set A

$$\int_A \|X_s - X_t\| d\mathbb{P} \le 2\sigma_X(s,t)\mathbb{P}(A)\psi_2^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + 2d_X(s,t)\mathbb{P}(A) . \tag{11.22}$$

We are therefore in a position to make use of the general setting developed in the previous sections, and in particular of Remarks 11.4 and 11.15. In this way, we obtain the following result; as announced, it is described in the setting of entropy for a somewhat clearer picture of the argument but also holds under the corresponding majorizing measures conditions. It improves upon (11.21).

Theorem 11.21. Let $X = (X_t)_{t\in T}$ be a Gaussian process with values in a separable Banach space B . Recall the weak and strong distances σ_X and d_X on T introduced above. Then, if

$$\int_0^\infty (\log N(T, \sigma_X; \varepsilon))^{1/2} d\varepsilon < \infty \qquad \text{and} \qquad \int_0^\infty \log^+ \log N(T, d_X; \varepsilon) d\varepsilon < \infty ,$$

X has a version with almost surely bounded and continuous paths on (T, d_X) .

Proof. Let us first show that there exists a sequence (a_ℓ) of positive numbers such that $\sum_\ell a_\ell < \infty$ and

$$\sum_\ell 2^{-\ell} (\log N(T, d_X; a_\ell))^{1/2} < \infty .$$

Set $b_k = (\log N(T, d_X; 2^{-k}))^{1/2}$ for every k. Since $\int_0^\infty \log^+ \log N(T, d_X; \varepsilon) d\varepsilon < \infty$, we have $\sum_k 2^{-k} \log^+ b_k < \infty$. Define $\ell_k = [\log_2^+(2^k b_k)] + 1$ where $[\cdot]$ is integer part and $\log_2$ the logarithm of base 2. We let then a_ℓ to be 2^{-k} for all ℓ with $\ell_k < \ell \leq \ell_{k+1}$. Clearly

$$\sum_\ell a_\ell \leq \sum_k 2^{-k} \ell_{k+1} < \infty .$$

Further

$$\sum_\ell 2^{-\ell} (\log N(T, d_X; a_\ell))^{1/2} \leq \sum_k 2^{-\ell_k + 1} b_k$$

which is finite too by definition of ℓ_k.

According to this, let now, for each ℓ, A_ℓ be minimal in T such that the d_X-balls with centers in A_ℓ and radius a_ℓ cover T. If D is such a ball, let further $C_\ell(D)$ be minimal in D such that the σ_X-balls with centers in $C_\ell(D)$ and radius $2^{-\ell}$ cover D. Set then, for every ℓ, $T_\ell = \bigcup_D C_\ell(D)$. We have

$$\log \ \mathrm{Card}\, T_\ell \leq \log N(T, \sigma_X; 2^{-\ell}) + \log N(T, d_X; a_\ell) .$$

Further, for every $t \in T$ and every ℓ, there exists, by construction, t_ℓ in T_ℓ such that $\sigma_X(t, t_\ell) \leq 2^{-\ell}$ and $d_X(t, t_\ell) \leq 2a_\ell$. If we now use (11.22) and recall that $\sum_\ell a_\ell < \infty$, we see that we are exactly in the setting of Remark 11.4. The conclusion therefore follows since this remark applies similarly to continuity as we noticed it.

Our last application deals with chaos processes. Gaussian and Rademacher chaos were introduced in Chapters 3 and 4 respectively where their integrability and tail behavior properties were investigated. As in those chapters, we restrict here to chaos of order 2. We deal moreover only with real valued chaos; we indicate at the end how the application can be amplified to more general cases. As before finally, we only deal with the Gaussian case; with the corresponding results in Chapter 4, the theorem we will obtain applies similarly to Rademacher chaos.

Recall (g_i) denotes an orthogaussian sequence. $X = (X_t)_{t \in T}$ is a Gaussian chaos process of order 2 if there is a sequence $(x_{ij}(t))$ of (real) functions on T such that, for all t, $X_t = \sum_{i,j} g_i g_j x_{ij}(t)$ where the sum is almost surely convergent. Following Section 3.2, we introduce two distances d_1 and d_2 on T by setting $d_1(s,t) = \|X_s - X_t\|_2$ and

$$d_2(s,t) = \sup_{|h| \leq 1} \Big|\sum_{i,j} h_i h_j (x_{ij}(s) - x_{ij}(t))\Big|$$

for s,t in T. With respect to Section 3.2, we do not consider the third parameter since we have seen there that, for real sequences, the associated decoupled chaos is equivalent to X (at least if the diagonal terms are zero). It follows from Lemma 3.8 and the comments introducing it that there is a numerical constant K such that, for all s,t, $d_2(s,t) \le K d_1(s,t)$ and

$$\mathbb{P}\{|X_s - X_t| > u d_1(s,t) + u^2 d_2(s,t)\} \le K \exp(-u^2/K) \tag{11.23}$$

for every $u > 0$. In particular (for some possibly different K),

$$\mathbb{P}\{|X_s - X_t| > u d_1(s,t)\} \le K \exp(-u/K).$$

We could then apply the results of the preceding sections and show boundedness and continuity of X in terms of the only distance d_1 with respect to the Young function $\psi_1(x) = \exp(x) - 1$. However, as in the previous application, the incremental estimates (11.23) involving the two distances d_1 and d_2 are more precise and lead to sharper conditions. (11.23) is used in the context of Remarks 11.4 and 11.15. To this aim, note that it implies that if $a \ge d_2(s,t)$, for all $u > 0$,

$$\mathbb{P}\{|X_s - X_t| > \frac{1}{a} d_1^2(s,t) + 2au\} \le K \exp(-u/K)$$

(since $a^{-1} d_1^2(s,t) + 2au \ge \sqrt{u} d_1(s,t) + u d_2(s,t)$). Hence, for some (possibly different) numerical constant K,

$$\|(|X_s - X_t| - \frac{1}{a} d_1^2(s,t))^+\|\psi_1 \le K a.$$

Therefore, if $a \ge d_2(s,t)$ and A is measurable in Ω,

$$\int_A |X_s - X_t| d\mathbb{P} \le K a \mathbb{P}(A) \psi_1^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + \frac{1}{a} d_1^2(s,t) \mathbb{P}(A). \tag{11.24}$$

These relations put us in the right situation in order to apply Theorems 11.13 and 11.14 together with Remark 11.15.

We can now state our result on almost sure boundedness and continuity of Gaussian chaos processes.

Theorem 11.22. Let $X_t = \sum_{i,j} g_i g_j x_{ij}(t)$, $t \in T$, be a Gaussian chaos processes of order 2 as just described with associated metrics d_1 and d_2. Assume there exist probability measures m_1 and m_2 on T such that respectively

$$\lim_{\eta \to 0} \sup_{t \in T} \int_0^\eta \left(\log \frac{1}{m_1(B_1(t,\varepsilon))}\right)^{1/2} d\varepsilon = 0$$

and

$$\lim_{\eta\to 0}\sup_{t\in T}\int_0^\eta \log\frac{1}{m_2(B_2(t,\varepsilon))}d\varepsilon = 0$$

where $B_i(t,\varepsilon)$ is the d_i-ball of center t and radius $\varepsilon > 0$ $(i = 1,2)$. Then $X = (X_t)_{t\in T}$ admits a version with almost all sample paths bounded and continuous on (T, d_1).

Proof. We only show that if T is finite and if M is a number such that

$$\sup_{t\in T}\int_0^\infty \left(\log\frac{1}{m_i(B_i(t,\varepsilon))}\right)^{i/2} d\varepsilon \le M\,, \quad i = 1,2,$$

then

$$\mathbb{E}\sup_{s,t\in T}|X_s - X_t| \le KM$$

for some numerical K. From this and the material discussed in the preceding section, it is not difficult to deduce the full conclusion of the statement.

Let thus T be finite and M be as before. According to Proposition 11.10, we may and do assume that d_1 and d_2 are ultrametric. For every t and ℓ set

$$\gamma(t,\ell) = 2^{-\ell}\left(\log\frac{1}{m_1(B_1(t,2^{-\ell}))}\right)^{1/2}$$

so that $\sum_\ell \gamma(t,\ell) \le KM$. Denote by $k(t,\ell)$ the largest integer k such that $2^{2j}\gamma(t,\ell+j) \le 1$ for all $j \le k$. We may observe that

$$\sum_\ell 2^{-2k(t,\ell)} \le 8KM\,. \tag{11.25}$$

To show this, let $L_n = \{\ell; \ell + k(t,\ell) = n\}$. We note that if $\ell \ne \ell'$ are elements of L_n, then $k(t,\ell) \ne k(t,\ell')$ from which it follows that $\sum_{\ell\in L_n} 2^{-2k(t,\ell)} \le 2^{-2k(t,\ell_0)+1}$ where $k(t,\ell_0) = \min\{k(t,\ell);\ \ell \in L_n\}$. But then, by definition of $k(t,\ell_0)$, $2^{-2k(t,\ell_0)} \le 4\gamma(t,\ell_0 + k(t,\ell_0) + 1) = 4\gamma(t, n+1)$. (11.25) clearly follows and implies in particular that

$$\sum_\ell \gamma(t,\ell + k(t,\ell)) \le 8KM\,. \tag{11.26}$$

As another property of the integers $k(t,\ell)$, note, as is easily checked, that $\ell + k(t,\ell)$ is increasing in ℓ. For every t and ℓ, consider the subset $H(t, 2^{-2\ell})$ of T consisting of the balls C of radius $2^{-\ell-k(t,\ell)-1}$

included in $B_1(t, 2^{-\ell-k(t,\ell)})$ such that $k(s,\ell) = k(t,\ell)$ for all s in C. From the definition of $k(t,\ell)$, the latter property only depends on C, and not on s in C. Since $t + k(t,\ell)$ is increasing in ℓ, $H(t, 2^{-2\ell-2}) \subset H(t, 2^{-2\ell})$. Further the subsets $H(t, 2^{-2\ell})$ when $t \in T$ form the family of the balls of radius $2^{-2\ell}$ for the (ultrametric) distance d' given by

$$d'(s,t) = 2^{-2\ell} \quad \text{where} \quad \ell = \sup\{j \,:\, \exists u \text{ such that } s,t \in H(u, 2^{-2j})\}\,.$$

Recall now from the assumptions that, for all t,

$$\sum_{\ell} 2^{-2\ell} \log \frac{1}{m_2(B_2(t, 2^{-2\ell}))} \le KM\,. \tag{11.27}$$

Let $B(t, 2^{-2\ell})$ be the ball of center t and radius $2^{-2\ell}$ for the distance $d = \max(d', d_2)$. Such a ball is the intersection of a d'-ball of radius $2^{-2\ell}$ and a d_2-ball of radius $2^{-2\ell}$. To this ball we can associate the weight

$$2^{-k(t,\ell)+k_0-1} m_1(B_1(t, 2^{-\ell-k(t,\ell)})) m_2(B_2(t, 2^{-2\ell}))$$

where k_0 is the smallest possible value for $k(t,\ell)$, $t \in T$, $\ell \in \mathbb{Z}$. We obtain in this way a family of weights as described in Remark 11.11. One can then construct a probability measure m on (T,d) such that, by (11.26) and (11.27), for all t in T,

$$\sum_{\ell} 2^{-2\ell} \log \frac{1}{m(B(t, 2^{-2\ell}))} \le KM$$

for some numerical K. We are then in a position to conclude. Let s and t be such that $d(s,t) \le 2^{-2\ell}$. Then, by construction, $d_1(s,t) \le 2^{-\ell-k(t,\ell)}$ and $d_2(s,t) \le 2^{-2\ell}$. Hence, from (11.24), for every measurable set A,

$$\int_A |X_s - X_t| d\mathbb{P} \le K 2^{-2\ell} \mathbb{P}(A) \psi^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + 2^{-2k(t,\ell)}\,.$$

We therefore see that we are exactly in the situation described by Remark 11.15 and (11.16) since (11.25) holds. We thus conclude the proof of Theorem 11.22 in this way.

Let us mention to conclude this chapter that the previous theorem might be extended to vector valued chaos processes, that is processes $X_t = \sum_{i,j} g_i g_j x_{ij}(t)$ where the functions $x_{ij}(t)$ take their values in a Banach spaces. According to the study of Section 3.2, three distances would then be involved with different entropy or majorizing measure conditions on each of them ($d+1$ distances for chaos of order d !). We do not pursue in this direction.

Notes and references

Various references have presented during the past years the theory of random processes on abstract index sets and their regularity properties as developed in this chapter. Exposition on Gaussian processes have been given in particular in the course [Fer4] by X. Fernique, and also by R. Dudley [Du2], V. N. Sudakov [Su4] and N. C. Jain and M. B. Marcus [J-M3]. General sufficient conditions for non-Gaussian processes satisfying incremental Orlicz conditions to be almost surely bounded or continuous are presented in the notes [Pi13] by G. Pisier and emphasized in [Fer7] and [We]. We refer to these authors for more accurate references and in particular to [Du2] for a careful historical description of the early developments of the theory of Gaussian (and non-Gaussian too) processes. Our exposition is based on [Pi13] and the recent work [Ta12].

The study of random processes under metric entropy conditions actually started with the Gaussian results of Section 11.3. The notion of ε-entropy goes back to Kolmogorov. The landmark paper [Du1] by R. Dudley, where Theorem 11.17 is established, introduced this fundamental abstract framework in the field. Credit for the introduction of ε-entropy applied to regularity of processes goes to V. Strassen (see [Du1]) and V. N. Sudakov [Su1]. It was only slowly realized after that the Gaussian structure of this result only relies on the appropriate integrability properties of the increments $X_s - X_t$ and that the technique could be extended to large classes of non-Gaussian processes. On the basis of Kolmogorov's continuity theorem (which already contains the fundamental chaining argument) and this observation, several authors investigated sufficient conditions for boundedness and continuity of processes whose increments $X_s - X_t$ are nicely controlled. Among the various articles, let us mention (see also [Du2]) [De], [Bou], [J-M1] (on subgaussian processes, see also [J-M3]), [Ha], [H-K], [N-N], [Ib] and [Kô] and [Pi9] on the important case of increments in L_p. The general Theorem 11.1 is due to G. Pisier [Pi13] (on the basis of [Pi9] thanks to an observation of X. Fernique). Its refined version Theorem 11.2 is equivalent to the (perhaps somewhat unorthodox) formulation of [Fer7]. The tail behaviors deduced from this statement were precisely analyzed in some cases in [Ale1] in the context of empirical processes (see also [Fer7], [We]). The uniform continuity and compactness results (Theorem 11.6 and Corollary 11.7) simplify prior proofs by X. Fernique [Fer11]. Kolmogorov's theorem (Corollary 11.8) may be found e.g. in [Slu], [Ne1], [Bi].

We refer to the survey paper [He3] for a history of majorizing measures. In [G-R-R], A. M. Garsia, E. Rodemich and H. Rumsey establish a real variable lemma using integral bounds involving majorizing measures to be applied to regularity of random processes. This lemma was further used and refined in [Gar] and [G-R] and usually provides interesting moduli of continuity. Let us note that our approach to majorizing measures is not completely similar to this real variable lemma. Our concerns go more to integrability and tail behaviors rather than moduli of continuity. More precisely, the technique of [G-R-R], refined by C. Preston [Pr1], [Pr2] and B. Heinkel [He1], [He3], allows for example to show the following *non-random* property. Let f be real (continuous) on some metric space (T,d) and let ψ be a Young function. Given a probability measure m on (T,d), denote by $\|\widetilde{f}\|_{L_\psi(m\times m)}$ the Orlicz norm with respect to ψ of

$$\widetilde{f}(s,t)=\frac{|f(s)-f(t)|}{d(s,t)}I_{\{d(s,t)\neq 0\}}$$

in the product space $(T\times T\,,\ m\times m)$. Then one can show (cf. [He3]) that for all s,t in T,

$$|f(s)-f(t)|\leq 20\|\widetilde{f}\|_{L_\psi(m\times m)}\sup_{u\in T}\int_0^{d(s,t)/2}\psi^{-1}\left(\frac{1}{m(B(u,\varepsilon))^2}\right)d\varepsilon\,.$$

Note the square of $m(B(u,\varepsilon))$ in the majorizing measure integral. (While this square is irrelevant when ψ has exponential growth, this is not the case in general, and a main concern of the paper [Ta12] is to remove this square when it is not needed.) In concrete situations, the evaluation of the entropy integral (usually for Lebesgue measure on some compact subset of $\mathbb{R}^N$) yields therefore various moduli of continuity and actually allows to study bounds on $\sup\limits_{s,t}|X_s-X_t|/d(s,t)^\alpha$, $\alpha>0$. These arguments have been proved useful in stochastic calculus by several authors (see e.g. [S-V], [Yo], [B-Y], [DM] etc., and in particular the recent connections between regularity of (stationary) Gaussian processes and regularity of local times of lévy processes put forward by M. T. Barlow [Bar1], [Bar2], [B-H]). On the basis of the seminal result of [G-R-R], C. Preston [Pr1], [Pr2] developed the concept of majorizing measures and basically obtained, in [Pr1], Theorem 11.18. However, in his main statement, C. Preston unnecessarily restricts his hypotheses. He was apparently not aware of the power of the present formulation which was put forward by X. Fernique who completely established Theorem 11.18 [Fer4]. X. Fernique developed in the mean time a somewhat different point of view based on duality of Orlicz norms (cf. [Fer3], [Fer4]). Our exposition is taken from [Ta12] to which we actually refer for a more complete exposition involving general Young functions ψ. The ultrametric structure and discretization procedure (Proposition 11.10 and Corollary 11.12), implicit in [Ta5], are described in [A-G-O-Z], [Ta7], and [Ta12].

As mentioned, Theorem 11.17 is due to R. Dudley [Du1], [Du2] after early observations and contributions by V. Strassen and V. N. Sudakov while Theorem 11.18 is due to X. Fernique [Fer4]. Various comments on regularity of Gaussian processes are taken from these references as well as from [M-S1], [M-S2] (where Slepian's lemma is introduced in this study), [J-M3], [Ta5], [Ad]. Lemma 11.16 is taken from [Fer8] (to show a result of [M-S2]). Let us mention here a volumic approach to regularity of Gaussian processes in the paper [Mi-P] where local theory of Banach spaces is used to prove a conjecture of [Du1]. Subgaussian processes have been emphasized in their relation to the central limit theorem in [J-M1], [He1], [J-M3] (cf. Chapter 14). Lemma 11.20 comes from [M-P2] (see also [M-P3]) and will be crucial in Chapter 13. The new technique of using simultaneously different majorizing measure conditions for different metrics in Theorems 11.21 and 11.22 is due to the second author. It becomes very natural in the new presentation of majorizing measures introduced in [Ta18]; see also [Ta19]. Theorem 11.22 on regularity of Gaussian and Rademacher chaos processes improves observations of [Bo6].

Chapter 12. Regularity of Gaussian and stable processes

Chapter 12. Regularity of Gaussian and stable processes

In the preceding chapter, we presented *sufficient* metric entropy and majorizing measure conditions for sample boundedness and continuity of random processes satisfying incremental conditions. In particular, these results were applied to Gaussian processes in Section 11.3. The main concern of this chapter is *necessity.* We will see indeed, as one of the main results, that the sufficient majorizing measure condition in order for a Gaussian process to be almost surely bounded or continuous is actually also necessary. This characterization thus provides a complete understanding of the regularity properties of Gaussian paths. The arguments of proof rely heavily on the basic ultrametric structure which lies behind a majorizing measure condition.

This characterization is performed in the first section which is completed by some equivalent formulations of the main result. Let us mention at this point that the study of necessity for non-Gaussian processes in this framework involves, rather than one given process, the whole family of processes satisfying incremental conditions with respect to the same Young function ψ and metric d. In the Gaussian case, Slepian's lemma and the comparison theorems (Section 3.3), which appear as a cornerstone in this study of necessity, confound those two situations and things become simpler. We refer the interested reader to [Ta12] for a study of necessity for general processes in this setting of incremental conditions. A noticeable exception is however the case of stable processes. The series representation and conditional use of Gaussian techniques allow here to describe necessary conditions as for Gaussian processes. This extension to p-stable processes, $1 \leq p < 2$, is the subject of Section 12.2; as we will see, it is sufficiency which appears to be more difficult in the stable case. This chapter is completed with applications to subgaussian processes and some remarkable type properties of the injection map $\mathrm{Lip}(T) \to C(T)$ when there is a Gaussian or stable majorizing measure on (T,d). The difficult subject of Rademacher processes is discussed through some conjectures in the very last part.

12.1. Regularity of Gaussian processes

Recall that a random process $X = (X_t)_{t\in T}$ is said to be Gaussian if each finite linear combination $\sum_i \alpha_i X_{t_i}$, $\alpha_i \in \mathbb{R}$, $t_i \in T$, is a real Gaussian variable. The distribution of the Gaussian process X is therefore completely determined by its covariance structure $\mathbb{E}X_sX_t$, $s,t \in T$. As we know, to study the regularity properties of X, it is fruitful to analyze the geometry of T for the induced L_2-pseudo-metric $d_X(s,t) = \|X_s - X_t\|_2$, $s,t \in T$.

We have seen in Theorem 11.18 that for *any* probability measure m on (T, d_X),

$$\mathbb{E} \sup_{t \in T} X_t \le K \sup_{t \in T} \int_0^\infty \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/2} d\varepsilon \tag{12.1}$$

where $B(t,\varepsilon)$ is the open ball of center t and radius $\varepsilon > 0$ in the pseudo-metric d_X and X has almost surely bounded sample paths if, for some probability measure m, the majorizing measure integral on the right of (12.1) is finite. There is a similar result about continuity. Recall further that $\mathbb{E} \sup_{t \in T} X_t$ is simply understood here as

$$\mathbb{E} \sup_{t \in T} X_t = \sup\{\mathbb{E} \sup_{t \in F} X_t \,;\ F \ \text{ finite in } T\} .$$

K is further some numerical constant which may vary from line to line in what follows. By (11.10), (12.1) contains the familiar entropy bound (Theorem 11.17)

$$\mathbb{E} \sup_{t \in T} X_t \le K \int_0^\infty (\log N(T, d_X; \varepsilon))^{1/2} d\varepsilon . \tag{12.2}$$

Now, we have seen in Theorem 3.18 a lower bound in terms of entropy numbers which indicates that

$$\sup_{\varepsilon > 0} \varepsilon (\log N(T, d_X; \varepsilon))^{1/2} \le K \mathbb{E} \sup_{t \in T} X_t . \tag{12.3}$$

These two bounds (12.2) and (12.3) appear to be rather close from each other. There is however a small gap which may be put forward by the example of an independent Gaussian sequence (Y_n) such that $\|Y_n\|_2 = (\log(n+1))^{-1/2}$ for all n. This sequence, which defines an almost surely bounded process by (3.7), shows that boundedness of Gaussian processes cannot be characterized by the metric entropy integral in (12.2). The reason for this failure is due to the possible lack of homogeneity of T for the metric d_X. We will see in the next chapter that in, a *homogeneous* setting, the metric entropy condition *does characterize* almost sure boundedness and continuity of Gaussian processes.

As we know, majorizing measures provide a way to take into account the possible lack of homogeneity of (T, d_X). Further, by (11.9) and (11.10), the majorizing measure integral of (12.1) lies in between the entropic bounds (12.2) and (12.3). As a main result, we will now show that the minoration (12.3) can be improved into

$$\sup_{t \in T} \int_0^\infty \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/2} d\varepsilon \le K \mathbb{E} \sup_{t \in T} X_t \tag{12.4}$$

for *some* probability measure m on T. Hence, together with (12.1), existence of a majorizing measure for the function ψ_2^{-1}, or $(\log(1/x))^{1/2}$ (cf. (11.18)), completely characterizes boundedness of Gaussian processes X in terms only of their associated L_2-metric d_X. There is a similar result for continuity.

The proof of this result, to which we now turn, requires a rather involved study of majorizing measures. This will be accomplished in an abstract setting which we now describe.

The majorizing measure conditions which will be included in this study concern in particular the ones associated to the Young functions $\psi_q(x) = \exp(x^q) - 1$. To unify the exposition, let us thus consider a strictly decreasing function h on $(0,1]$ with $h(1) = 0$ and $\lim\limits_{x \to 0} h(x) = \infty$. We assume that for all x, y in $(0,1]$

$$h(xy) \leq h(x) + h(y)\,. \tag{12.5}$$

[This condition may be weakened into $h(xy) \leq h(x) + ch(y)$ for some positive c and the reader can verified that the subsequent arguments go through in this case, the numerical constant of Theorem 12.5 depending then on h.] As announced, the main examples we have in mind with this function h are the examples of $h_q(x) = (\log(1/x))^{1/q}$, $1 \leq q < \infty$, and also $h_\infty(x) = \log(1 + \log(1/x))$.

Let us mention that we never attempt to find sharp numerical constants, but always use crude, but simple, bounds.

If (T,d) is a metric space, recall we denote by $D = D(T)$ its diameter, i.e.

$$D = D(T) = \sup\{d(s,t)\,;\ s,t \in T\}\,.$$

Given a probability measure m on the metric space (T,d), let

$$\gamma_m(T) = \gamma_m(T,d) = \sup_{t \in T} \int_0^\infty h(m(B,(t,\varepsilon))d\varepsilon$$

where $B(t,\varepsilon)$ is the open ball of center t and radius $\varepsilon > 0$ in (T,d). Here, and throughout this study, when no ambiguity arises, we adopt the convention that $B(t,\varepsilon)$ denotes the ball for the distance on the space which contains t. We also let

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

where the infimum is taken over all probability measures m on T. For a subspace A of T, $\gamma(A) = \gamma(A,d)$ refers to the quantity associated to the metric space (A,d), i.e. $\gamma(A) = \inf \gamma_m(A)$ where the infimum is taken over the probability measures supported by A.

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

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

The nice feature of ultrametric spaces is that two balls of the same radius are either identical or disjoint.

From now on, and until further notice, we assume that all the metric spaces are *finite*. Given a metric space (T,d) and an ultrametric space (U,δ), say that a map φ from U onto T is a contraction if $d(\varphi(u),\varphi(v)) \le \delta(u,v)$ for u,v in U. Define the functional $\alpha(T) = \alpha(T,d)$ by

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

Although $\gamma(T)$ comes first in mind as a way to measure the size of T, the quantity $\alpha(T)$ is easier to manipulate and yields stronger results. We first collect some simple facts.

Lemma 12.1. The following hold under the preceding notations.

(i) $\gamma(T) \le \alpha(T)$.

(ii) If $A \subset T$, then $\gamma(A) \le 2\gamma(T)$.

(iii) If U is ultrametric and $A \subset U$, then $\gamma(A) \le \gamma(U)$.

(iv) If $A \subset T$, then $\alpha(A) \le \alpha(T)$.

(v) $\alpha(T) = \inf\{\gamma(U); U$ is ultrametric, $D(U) \le D(T)$ and T is the image of U by a contraction $\}$ and this infimum is attained.

(vi) $D(T) \le 2[h(1/2)]^{-1}\gamma(T)$.

Proof. (i) Let φ be a contraction from U onto T, μ a probability measure on U and $m = \varphi(\mu)$. For u in U, $\varepsilon > 0$, we have, since φ is a contraction, that $\varphi^{-1}(B(\varphi(u),\varepsilon)) \supset B(u,\varepsilon)$, so $m(B(\varphi(u),\varepsilon)) \ge \mu(B(u,\varepsilon))$. Since h is decreasing and φ onto, we get $\gamma_m(T) \le \gamma_\mu(U)$, so $\gamma(T) \le \gamma(U)$ since μ is arbitrary; therefore $\gamma(T) \le \alpha(T)$ since U and φ are arbitrary. (ii) This has already been shown in Lemma 11.9 but let us briefly recall the argument. For t in T, take $a(t)$ in A with $d(t,a(t)) = d(t,A)$.

Let m be a probability measure on T and let $\mu = a(m)$ so that μ is supported by A. Fix x in A. For t in T, we have $d(t, A) \leq d(t, x)$, so $d(t, a(t)) \leq d(t, x)$ and $d(x, a(t)) \leq 2d(x, t)$. Since $\mu = a(m)$, it follows that $\mu(B(x, 2\varepsilon)) \geq m(B(x, \varepsilon))$. Hence, by a change of variables, $\gamma_\mu(A) \leq 2\gamma_m(T)$ which gives the results. (iii) With the notations of the proof of (ii), the ultrametricity gives $d(x, a(t)) \leq \max(d(x,t), d(t, a(t)) \leq d(x,t)$ and thus $\gamma(A) \leq \gamma(U)$ in this case. (iv) Let U be ultrametric and let φ be a contraction from U onto T. By (iii), we get that $\alpha(A) \leq \alpha(\varphi^{-1}(A)) \leq \gamma(U)$ and thus $\alpha(A) \leq \alpha(T)$. (v) If (U, δ) is ultrametric and φ is a contraction from U onto T, consider the distance δ_1 on U given by $\delta_1(u,v) = \min(\delta(u,v), D(T))$. Then (U, δ_1) is ultrametric and φ is still a contraction from (U, δ_1) onto T. By the argument of (i), $\gamma(U, \delta_1) \leq \gamma(U, \delta)$. The last assertion follows by a standard compactness argument. (vi) Take two points s and t in T and let $\eta = d(s,t)$. The balls $B(s, \eta/2)$ and $B(t, \eta/2)$ are disjoint so that if m is a probability measure on T, one of these balls, say the first, has a measure less than $1/2$. Therefore

$$\gamma_m(T) \geq \int_0^{\eta/2} h(m(B(s,\varepsilon)))d\varepsilon \geq \frac{\eta}{2} h(\frac{1}{2})$$

from which the result follows since m, s, t are arbitrary and $h(1/2) > 0$ by (12.5). The proof of the lemma is complete.

The next lemma is one of the key tools of this investigation. It exhibits a behavior of α that resembles a strong form of subadditivy.

Lemma 12.2. Let T be a finite metric space with diameter $D = D(T)$. Suppose that we have a finite covering $A_1, \ldots, A_n$ of T. Then, for every positive numbers $a_1, \ldots, a_n$ with $\sum\limits_{i=1}^{n} a_i \leq 1$,

$$\alpha(T) \leq \max_{i \leq n} [\alpha(A_i) + D(T) h(a_i)] .$$

Proof. From Lemma 12.1 (v), for every $i = 1, \ldots, n$, there exists an ultrametric space (U_i, δ_i) of diameter less than D, a contraction φ_i from U_i onto A_i and a probability measure μ_i on U_i such that $\alpha(A_i) = \gamma_{\mu_i}(U_i)$ (or arbitrarily close). Let U be the disjoint sum of the spaces $(U_i)_{i \leq n}$. Define the distance δ on U by $\delta(u,v) = \delta_i(u,v)$ whenever u, v belong to the same U_i, and $\delta(u,v) = D$ otherwise. Then (U, δ) is ultrametric and the map φ from U onto T given by $\varphi(u) = \varphi_i(u)$ for u in U_i is a contraction. Consider the positive measure μ' on U given by $\mu' = \sum\limits_{i=1}^{n} a_i \mu_i$. Since $|\mu'| \leq 1$, there is a probability μ on

U with $\mu \geq \mu'$. Take then u in U and let i be such that $u \in U_i$. By (12.5),

$$\begin{aligned} h(\mu(B(u,\varepsilon))) &\leq h(\mu'(B(u,\varepsilon))) \\ &\leq h(a_i \mu_i(B(u,\varepsilon))) \\ &\leq h(\mu_i(B(u,\varepsilon))) + h(a_i)\,. \end{aligned}$$

It follows that

$$\begin{aligned} \int_0^\infty h(\mu(B(u,\varepsilon)))d\varepsilon &= \int_0^D h(\mu(B(u,\varepsilon)))d\varepsilon \\ &\leq \int_0^\infty h(\mu_i(B(u,\varepsilon)))d\varepsilon + Dh(a_i) \\ &\leq \alpha(A_i) + Dh(a_i)\,. \end{aligned}$$

Therefore

$$\alpha(U) \leq \gamma_\mu(U) \leq \max_{i \leq n}[\alpha(A_i) + Dh(a_i)]$$

from which the conclusion follows.

The next lemma is the basic step in the subsequent construction. If (T,d) is a finite metric space, let, for every integer k (in $\mathbb{Z}$),

$$\beta_k(T) = \alpha(T) - \sup_{x \in T} \alpha(B(x, 6^{-k}))\,.$$

Lemma 12.3. Let (T,d) be a finite metric space of diameter less than 6^{-k}. We are necessarily in one of the following two cases:

(i) either there exists a subset S of T of diameter less than 6^{-k-1} satisfying

$$\beta_{k+2}(S) \geq 2(\alpha(T) - \alpha(S))\,; \tag{12.6}$$

(ii) or there exists balls $(B_i)_{1 \leq i \leq N}$ of radius 6^{-k-2} with centers at mutual distance larger than $3 \cdot 6^{-k-2}$ such that

$$\text{for all } i\,,\ \alpha(B_i) \geq \alpha(T) - 6^{-k+1} h\left(\frac{1}{1+N}\right)\,. \tag{12.7}$$

Proof. Suppose that (i) does not hold. By induction, we construct points x_i, $i \geq 1$, in T in the following manner: $\alpha(B(x_1, 6^{-k-2}))$ is maximal, and, if $x_1, \ldots, x_{i-1}$ have been constructed, we take x_i such that

$$\alpha(B(x_i, 6^{-k-2})) = \max\{\alpha(B(x, 6^{-k-2}))\,,\ \forall j < i\,, \qquad d(x, x_j) \geq 3 \cdot 6^{-k-2}\}\,.$$

for every i, set then

$$S_i = B(x_i, 3\cdot 6^{-k-2})\setminus \bigcup_{j<i} B(x_j, 3\cdot 6^{-k-2})\,.$$

Since (i) does not hold, necessarily, for any i, $\beta_{k+2}(S_i) \le 2(\alpha(T)-\alpha(S_i))$. By construction $\beta_{k+2}(S_i) = \alpha(S_i) - \alpha(B(x_i, 6^{-k-2}))$. Thus, for all i,

$$\begin{aligned} \alpha(B(x_i, 6^{-k-2})) &= \alpha(S_i) - \beta_{k+2}(S_i) \\ &\ge \alpha(S_i) - 2(\alpha(T)-\alpha(S_i)) \\ &\ge \alpha(T) - 3(\alpha(T)-\alpha(S_i))\,. \end{aligned} \tag{12.8}$$

The union of the S_i's covers T. They can be assumed to be ordered in such a way that the sequence $(\alpha(S_i))$ is decreasing. If we let $a_i = (i+1)^{-2}$, we have $\sum_{i\ge 1} a_i \le 1$. Therefore, by Lemma 12.2, there exists $i_0 \ge 1$ such that

$$\alpha(S_{i_0}) \ge \alpha(T) - 6^{-k} h((i_0+1)^{-2})\,. \tag{12.9}$$

By (12.5), $h((i_0+1)^{-2}) \le 2h((i_0+1)^{-1})$. Hence, if $I = \{1,\ldots,i_0\}$, since $(\alpha(S_i))$ is decreasing, we see from (12.9) that for all i in I,

$$\alpha(S_i) \ge \alpha(T) - 2\cdot 6^{-k} h\left(\frac{1}{1+\ \mathrm{Card} I}\right)\,. \tag{12.10}$$

Combining (12.10) with (12.8) yields that for all i in I,

$$\alpha(B(x_i, 6^{-k-2})) \ge \alpha(T) - 6^{-k+1} h\left(\frac{1}{1+\ \mathrm{Card} I}\right)\,.$$

Letting $B_i = B(x_i, 6^{-k-2})$ for i in I and $N = \ \mathrm{Card} I$ shows that we are in case (ii) of the statement and that (12.7) is satisfied. Lemma 12.3 is established.

We now perform the main construction. Given a metric space T, we exhaust it with the alternative of Lemma 12.3 and construct in this way subsets (actually balls) which are well separated and whose α-functionals are big enough and carry enough information on $\alpha(T)$ itself. Iterative use of this proposition gives raise to a "tree" and an ultrametric structure. For two subsets A, B of a metric space (T,d), let $d(A,B) = \inf\{d(a,b)\,; a\in A\,,\ b \in B\}$.

Proposition 12.4. Let (T,d) be a finite metric space of diameter less than 6^{-k}. There exists an integer $\ell \geq k$ and subsets $(B_i)_{1\leq i\leq N}$ of T of diameter less than $6^{-\ell-1}$ such that $d(B_i,B_j)\geq 6^{-\ell-2}$ for $i\neq j$, the diameter of $\bigcup_{i\leq N} B_i$ is $\leq 6^{-\ell+1}$ and

$$\text{for all } i\,,\ \alpha(B_i)\geq \alpha(T)-2\cdot 6^{-k+1}h\left(\frac{1}{1+N}\right)\,. \tag{12.11}$$

Proof. By induction, we construct a decreasing sequence (T_m) of subsets of T such that $T_0=T$, $D(T_m)\leq 6^{-k-m}$ and

$$\beta_{k+m+1}(T_m)\geq 2(\alpha(T_{m-1})-\alpha(T_m)) \tag{12.12}$$

for all $m\geq 1$. The construction stops (since T is finite) for some $m=n$ and in T_n we are necessarily in case (ii) of Lemma 12.3 (since if not we would be able to continue the exhaustion). We first note that, for all $m\leq n$,

$$\alpha(T)-\alpha(T_m)\leq \beta_{k+m+1}(T_m)\,. \tag{12.13}$$

Indeed, this is clearly the case for $m=1$ (since we even have in this case that $\beta_{k+2}(T_1)\geq 2(\alpha(T)-\alpha(T_1))$). Assume then that (12.13) is satisfied for m and let us show it for $m+1$. We have by (12.12) that

$$\begin{aligned}\beta_{k-m+2}(T_{m+1}) &\geq 2(\alpha(T_m)-\alpha(T_{m+1}))\\ &\geq \alpha(T_m)-\alpha(T_{m+1})+\beta_{k+m+1}(T_m)\end{aligned}$$

since $\alpha(T_m)-\alpha(T_{m+1})\geq \beta_{k+m+1}(T_m)$ by definition of the of the functional β. Thus, by the induction hypothesis,

$$\beta_{k+m+2}(T_{m+1})\geq \alpha(T)-\alpha(T_{m+1})$$

and (12.13) indeed holds.

Set now $\ell=k+n$. Since in T_n we are in case (ii) of Lemma 12.3, we can find balls $(B'_i)_{1\leq i\leq N}$ (ot T_n) of radius $6^{-\ell-2}$ with centers at mutual distance $\geq 3\cdot 6^{-\ell-2}$ such that, for all i,

$$\alpha(B'_i\cap T_n)\geq \alpha(T_n)-6^{-k+1}h\left(\frac{1}{1+N}\right)\,. \tag{12.14}$$

Since $\alpha(B'_i \cap T_n) \le \alpha(T_n) - \beta_{\ell+1}(T)$, combining (12.13) and (12.14) yields that

$$\alpha(B'_i \cap T_n) \ge \alpha(T) - 2 \cdot 6^{-k+1} h\left(\frac{1}{1+N}\right)$$

and the proof of Proposition 12.4 is complete with $B_i = B'_i \cap T_n$.

Let U be an ultrametric space. For x in U, $k \in \mathbb{Z}$, let $N_k(x)$ be the number of disjoint balls of radius 6^{-k-1} which are contained in $B(x, 6^{-k})$. Define

$$\xi_x(U) = \sum_{k \in \mathbb{Z}} 6^{-k} h(1/N_k(x)),$$
$$\xi(U) = \inf_{x \in U} \xi_x(U).$$

We note that if $D(U) \le 6^{-k_0}$ and $B(x, 6^{-k_1}) = \{x\}$ for all x, we have

$$\xi_x(U) = \sum_{k_0 \le k < k_1} 6^{-k} h(1/N_k(x)).$$

We can now state and prove the main conclusion of the preceding construction.

Theorem 12.5. There is a numerical constant K with the following property: for each function h satisfying (12.5) and each finite metric space (T, d), there exist an ultrametric space (U, δ) and a map $\phi : U \to T$ such that the following conditions hold:

$$\alpha(T) \le K\xi(U);$$

$$\text{for all } u, v \text{ in } U,\ \delta(u,v) \le d(\phi(u), \phi(v)) \le 6^3 \delta(u,v).$$

Proof. Let k_0 be the largest integer (in $\mathbb{Z}$) with $6^{-k_0} \ge D(T)$. Consider two points u, v of T with $d(u,v) = D(T)$. The space $U = (\{u,v\}, d)$ is ultrametric and the canonical injection ϕ from U in T satisfies $6^{-k_0-1} \le d(\phi(u), \phi(v)) \le 6^{-k_0}$. The balls $B(u, 6^{-k_0-1})$, $B(v, 6^{-k_0-1})$ are disjoint; so we have

$$\xi(U) \ge 6^{-k_0-1} h(1/2) \ge 6^{-1} h(1/2) D(T).$$

We intend to prove the theorem with $K = 4 \cdot 6^3$. By the preceding, the result holds unless $\alpha(T) \ge 4 \cdot 6^2 h(1/2) D(T)$. It thus remains to prove the theorem in that case only. By induction over $k \ge k_0$, we construct a family $\mathcal{B}$ of subsets A of T in the following way. The construction starts with $A = T$

and each step is performed by an application of Proposition 12.4 to each element of $\mathcal{B}$ obtained at the step before. That is, if $A \in \mathcal{B}$ with diameter $\leq 6^{-k}$, there exist integers $k(A) \geq k$ and $N(A) \geq 1$, subsets $(B_i(A))_{1 \leq i \leq N(A)}$ of A of diameter $\leq 6^{-k(A)-1}$ and such that $d(B_i(A), B_j(A)) \geq 6^{-k(A)-2}$, and the diameter of $\bigcup_{i \leq N(A)} B_i(A)$ is $\leq 6^{-k(A)+1}$, with the following property: for all i,

$$\alpha(B_i(A)) \geq \alpha(A) - 2 \cdot 6^{-k(A)+1} h\left(\frac{1}{1+N(A)}\right). \tag{12.15}$$

The construction stops when each element A of $\mathcal{B}$ is reduced to exactly one point and we denote by U the collection of points of T obtained in this way. For u, v in U, there exists A in $\mathcal{B}$ such that for two different $B_i(A)$, $B_j(A)$, $u \in B_i(A)$, $v \in B_j(A)$. We then set $\delta(u,v) = 6^{-k(A)-2}$. δ is ultrametric on U. Further, if ϕ is the canonical injection map from U into T, we have by construction that $\delta(u,v) \leq d(\phi(u), \phi(v)) \leq 6^3 \delta(u,v)$. Fix x in U. Denote by $(A_\ell)_{\ell \geq 1}$ the decreasing sequence of elements of $\mathcal{B}$ that contain x; $A_1 = T$. By (12.15), for every $\ell \geq 1$,

$$\alpha(A_{\ell+1}) \geq \alpha(A_\ell) - 2 \cdot 6^{-k(A_\ell)+1} h\left(\frac{1}{1+N(A_\ell)}\right).$$

Since $\alpha(\{x\}) = 0$, summation of those inequalities yields

$$\alpha(T) \leq 12 \sum_{\ell \geq 1} 6^{-k(A_\ell)} h\left(\frac{1}{1+N(A_\ell)}\right). \tag{12.16}$$

For $k \geq k_0 + 3$, let $B(x, 6^{-k})$ be the δ-ball of U with center x and radius 6^{-k}. By definition of δ, if $k = k(A_\ell) + 2$ for some ℓ, then $N_k(x) = N(A_\ell)$, while if there is no such ℓ, $N_k(x) = 1$. Hence, from (12.16), and (12.5) since $1 + N(A_\ell) \leq 2N(A_\ell)$, we get that

$$\begin{aligned}
\alpha(T) &\leq 12\left(h(1/2) \sum_{k > k_0} 6^{-k} + 6^2 \xi_x(U)\right) \\
&\leq 12(6h(1/2)D(T) + 6^2 \xi_x(U)).
\end{aligned}$$

Since we are in the case $D(T) \leq \alpha(T)/4 \cdot 6^2 h(1/2)$, it follows that $\alpha(T) \leq 4 \cdot 6^3 \xi_x(U)$. Since x is arbitrary in U, we have $\alpha(T) \leq 4 \cdot 6^3 \xi(U)$ which is the announced claim.

Provided with the abstract Theorem 12.5, we can now prove existence of majorizing measures for bounded Gaussian processes (at least, to start with, indexed by a finite set). *In the rest of this section,* $h(x) = (\log(1/x))^{1/2}$.

Theorem 12.6. Let $X = (X_t)_{t\in T}$ be a Gaussian process indexed by a finite set T, and provide T with the canonical distance $d_X(s,t) = \|X_s - X_t\|_2$. Then

$$\alpha(T, d_X) \leq K\mathbb{E}\sup_{t\in T} X_t$$

where K is a numerical constant.

The fact here that d_X does not possibly separate all points of T is no problem: simply identify s and t such that $d_X(s,t) = 0$ and the new index set $\widetilde{T}$ obtained in this way is such that $\alpha(\widetilde{T}, d_X) = \alpha(T, d_X)$ and $\mathbb{E}\sup_{t\in\widetilde{T}} X_t = \mathbb{E}\sup_{t\in T} X_t$.

Let U, ϕ be as given by the application of Theorem 12.5 to the space (T, d_X) (or $(\widetilde{T}, d_X)$). It is thus enough to show that $\xi(U) \leq K\mathbb{E}\sup_{u\in U} X_{\phi(u)}$. We note that for u, v in U, $d_X(\phi(u), \phi(v)) \geq \delta(u,v)$ so that the theorem is a consequence of the following result that we single out for future reference. It is at this point, *actually the unique place in this study*, that the Gaussian structure through the comparison theorems based on Slepian's lemma plays its key role.

Proposition 12.7. Let (U, δ) be a finite ultrametric space. Then, for each Gaussian process $X = (X_u)_{u\in U}$ such that $d_X(u,v) \geq \delta(u,v)$ whenever $u, v \in U$, we have $\xi(U) \leq K\mathbb{E}\sup_{u\in U} X_u$ where K is numerical.

Proof. Let k_0 be the largest integer such that $6^{-k_0} \geq D(U)$. For $k > k_0$, let $\mathcal{B}_k$ be the collection of the balls of radius 6^{-k}. Let $\mathcal{B} = \bigcup_{k>k_0} \mathcal{B}_k$. Consider an independent family $(g_B)_{B\in>B}$ of standard normal variables. For u in U, $k > k_0$, we write simply $g_{u,k} = g_{B(u,6^{-k})}$. We let further $Z_u = \sum_{k>k_0} 6^{-k} g_{u,k}$. Let u, v in U and let ℓ be the largest such that $\delta(u,v) \leq 6^{-\ell}$. Then $B(u, 6^{-k}) = B(v, 6^{-k})$ for $k \leq \ell$, so $Z_u - Z_v = \sum_{k>\ell} 6^{-k}(g_{u,k} - g_{v,k})$. It follows that

$$\|Z_u - Z_v\|_2 \leq \sqrt{2}\sum_{k>\ell} 6^{-k} \leq 2\cdot 6^{-\ell-1} \leq 2\delta(u,v) \leq 2d_X(u,v).$$

Corollary 3.14 shows that it is enough to establish that $\xi(U) \leq A\mathbb{E}\sup_{u\in U} Z_u$ for some constant A. According to (3.14), we take A such that $(\log N)^{1/2} \leq A\mathbb{E}\max_{i\leq N} g_i$ for all N where (g_i) is an orthogaussian sequence. By induction over n, we establish the following statement:

(H_n) If U has diameter $\leq 6^{-k_0}$ and if, for each x in U, $B(x, 6^{-k}) = \{x\}$ with $k - k_0 \leq n$, then $\xi(U) \leq A\mathbb{E}\sup_{u\in U} Z_u$.

For $n = 0$, U contains only one point so that $\xi(U) = 0$ and (H_0) holds. Let us assume that (H_n) holds and let us prove (H_{n+1}). We enumerate $\mathcal{B}_{k_0+1}$ as $\{B_1, \ldots, B_q\}$. For $i \leq q$, let $\Omega_i = \{\forall p \leq q\,,\ p \neq i\,,\ g_{B_p} < g_{B_i}\}$. For u in U, define

$$Z'_u = \sum_{k > k_0+1} 6^{-k} g_{u,k} = Z_u - 6^{-k_0-1} g_{u,k_0+1}\,.$$

For $i \leq q$ consider a measurable map τ_i from Ω to B_i that satisfies $Z'_{\tau_i} = \sup_{u \in B_i} Z'_u$. Define now a measurable map τ from Ω to U by $\tau(\omega) = \tau_i(\omega)$ for ω in Ω_i. We have

$$\begin{aligned}
\mathbb{E} \sup_{u \in U} Z_u \geq \mathbb{E} Z_\tau &= \sum_{i \leq q} \mathbb{E}(I_{\Omega_i} Z_{\tau_i}) \\
&\geq \sum_{i \leq q} \mathbb{E}(I_{\Omega_i}(6^{-k_0-1} g_{B_i} + Z'_{\tau_i})) \\
&= 6^{-k_0-1} \sum_{i \leq q} \mathbb{E}(I_{\Omega_i} g_{B_i}) + \sum_{i \leq q} \mathbb{E}(I_{\Omega_i} Z'_{\tau_i})\,.
\end{aligned}$$

Now

$$\sum_{i \leq q} \mathbb{E}(I_{\Omega_i} g_{B_i}) = \mathbb{E} \max_{i \leq q} g_{B_i} \geq A^{-1} (\log q)^{1/2}\,.$$

Further, the independence of the variables $(g_B)_{B \in \mathcal{B}}$ shows that I_{Ω_i} and Z'_{τ_i} are independent, and thus

$$\mathbb{E}(I_{\Omega_i} Z'_{\tau_i}) = \mathbb{P}(\Omega_i) \mathbb{E} Z'_{\tau_i} = \frac{1}{q} \mathbb{E} Z'_{\tau_i}\,.$$

By the induction hypothesis, for every i,

$$A \mathbb{E} Z'_{\tau_i} = A \mathbb{E} \sup_{u \in B_i} Z'_u \geq \xi(B_i)\,.$$

Since the definition of ξ makes it clear that for each i

$$\xi(B_i) + 6^{-k_0-1} (\log q)^{1/2} \geq \xi(U)\,,$$

the proof is complete.

Theorem 12.6 proves the existence of a majorizing measure for Gaussian processes when the index set is finite since $\gamma(T) \leq \alpha(T)$ (Lemma 12.1). We now deduce from this finite case the existence of a majorizing measure for almost surely bounded general Gaussian processes. The use of the functional α actually yields a seemingly *stronger*, but equivalent (see Remark 12.11), statement with, in this form, some interesting consequences to be developed next.

Anticipating on the next section, let us mention that the following two theorems, as well as their consequences, on necessary conditions for boundedness and continuity of sample paths of Gaussian processes actually hold similarly for other processes *once* a statement for T finite analogous to Theorem 12.6 can be established for them. This is the procedure which will indeed be followed for stable processes in the next section. We therefore write the *proofs* below with a general function h.

Metric spaces are no longer always finite.

Theorem 12.8. Consider a bounded Gaussian process $X = (X_t)_{t \in T}$. Then, there exists a probability measure m on (T, d_X) such that

$$\sup_{t \in T} \int_0^\infty \left(\log \frac{1}{\sup\{m(\{s\}); d_X(s,t) < \varepsilon\}} \right)^{1/2} d\varepsilon \le K \mathbb{E} \sup_{t \in T} X_t$$

where K is a numerical constant.

Proof. Theorem 12.6 shows that for each finite subset F of T,

$$\alpha(F) \le K \mathbb{E} \sup_{t \in F} X_t \le K \mathbb{E} \sup_{t \in T} X_t \, .$$

It is hence enough to show that if $\alpha = \sup\{\alpha(F) ; \ F \subset T , \ F \text{ finite}\}$, there is a probability measure m on T such that the left hand side of the inequality of the theorem is less than $K\alpha$. Denote by k_0 the largest integer with $2^{-k_0} \ge D(T)$. Since X is almost surely bounded, $N(T, d_X; \varepsilon) < \infty$ for every $\varepsilon > 0$ (Theorem 3.18). For $\ell \ge k_0$, let T_ℓ be a finite subset of T such that each point of T is within a distance $\le 2^{-\ell}$ of a point of T_ℓ. Consider a map a_ℓ from T to T_ℓ such that $d_X(t, a_\ell(t)) \le 2^{-\ell}$. For each k, we know that $\alpha(T_\ell) \le \alpha$. So there exist an ultrametric space (U_ℓ, δ_ℓ), a contraction φ_ℓ from U_ℓ on T_ℓ, and a probability μ_ℓ on U_ℓ such that $\gamma_{\mu_\ell}(U_\ell) \le \alpha$. This implies that, for every u in U_ℓ and every ℓ,

$$\sum_{k > k_0} 2^{-k} h(m_\ell(B(u, 2^{-k}))) \le 2\alpha \, . \tag{12.17}$$

To each ball B in an ultrametric space U, we associate a point $v(B)$ of B. Let $\mathcal{B}_k$ be the family of balls of radius 2^{-k} of U_ℓ. Denote by m_k^ℓ the probability measure on T that, for each B in $\mathcal{B}_k$, assigns mass $\mu_\ell(B)$ to the point $a_k(\varphi_\ell(v(B)))$. We note that m_k^ℓ is supported by T_k. Fix t in T. Choose t' in T_ℓ

with $d_X(t,t') \le 2^{-\ell}$. Take u in U_ℓ such that $\varphi_\ell(u) = t'$. For each k, we have $\delta_\ell(u, v(B(u,2^{-k}))) \le 2^{-k}$ so that $d_X(t', \varphi_\ell(v(B(u,2^{-k})))) \le 2^{-k}$ and

$$d_X(t, a_k(\varphi_\ell(v(B(u,2^{-k}))))) \le 2^{-k+1} + 2^{-\ell}.$$

We set $t_k^\ell = a_k(\varphi_\ell(v(B(u,2^{-k}))))$ so $d_X(t, t_k^\ell) \le 2^{-k+1} + 2^{-\ell}$, and $m_k^\ell(\{t_k^\ell\}) \ge m_\ell(B(u,2^{-k}))$. It follows from (12.17) that we have

$$\sum_{k>k_0} 2^{-k} h(m_k^\ell(\{t_k^\ell\})) \le 2\alpha.$$

Let $\mathcal{U}$ be a ultrafilter in $\mathbb{N}$. Since t_k^ℓ belongs to the finite set T_k, the limit $t_k = \lim_{\ell\to\mathcal{U}} t_k^\ell$ exists, and $d_X(t,t_k) \le 2^{-k+1}$. Since m_k^ℓ is supported by the finite set T_k, the limit $m_k = \lim_{\ell\to\mathcal{U}} m_k^\ell$ exists and thus, for each t in U,

$$\sum_{k>k_0} 2^{-k} h(m_k(\{t_k\})) \le 2\alpha.$$

Let $m = \sum_{k>k_0} 2^{k_0-k} m_k$, so m is a probability on T. We note that, by (12.5),

$$\begin{aligned} h(m(\{t_k\})) &\le h(2^{k_0-k} m_k(\{t_k\})) \\ &\le h(2^{k_0-k}) + h(m_k(\{t_k\})). \end{aligned}$$

It follows that

$$\begin{aligned} \sum_{k>k_0} 2^{-k} h(m(\{t_k\})) &\le \sum_{k>k_0} 2^{-k} h(2^{k_0-k}) + 2\alpha \\ &\le KD(T) + 2\alpha \end{aligned}$$

where we have used (12.5) and $2^{-k_0} \ge D(T)$. Recall from Lemma 12.1 that $D(T) \le K\alpha$. The conclusion then follows since for $\varepsilon \ge 2^{-k+1}$, $\sup\{m(\{s\})\,;\ d_X(s,t) < \varepsilon\} \ge m(\{t_k\})$.

We now present the necessary majorizing measure condition for almost sure continuity of Gaussian processes.

Theorem 12.9. Consider a Gaussian process $X = (X_t)_{t\in T}$ that is almost surely bounded and continuous (or that admits a version which is almost surely bounded and continuous) on (T, d_X). Then, there exists a probability measure m on (T, d_X) such that

$$\lim_{\eta\to 0} \sup_{t\in T} \int_0^\eta \left(\log \frac{1}{\sup\{m(\{s\})\,;\ d_X(s,t) < \varepsilon\}} \right)^{1/2} d\varepsilon = 0.$$

Proof. For $n \geq 1$, let $a_n = 2^{-n} D$ where $D = D(T)$. Consider a family $B_{n,1}, \ldots, B_{n,p(n)}$ of d_X-balls of radius a_n that covers T where $p(n) = N(T, d_X; a_n)$. So, for $i \leq p(n)$, if we denote by $t(n,i)$ the center of $B_{n,i}$, we have

$$\mathbb{E} \sup_{t \in B_{n,i}} X_t = \mathbb{E} \sup_{t \in B_{n,i}} (X_t - X_{t(n,i)}) \leq \mathbb{E} \sup_{t \in B_{n,i}} |X_t - X_{t(n,i)}| \leq \beta(a_n)$$

where we have set $\beta(\eta) = \sup_{t \in T} \mathbb{E} \sup_{d_X(s,t) < \eta} |X_s - X_t|$. Denote by $d_{n,i}$ the diameter of $B_{n,i}$ (that can be smaller than $2a_n$). Theorem 12.8 shows that there is a probability measure $m_{n,i}$ on $B_{n,i}$ such that for each t in $B_{n,i}$ we have

$$\int_0^{d_{n,i}} h(\sup\{m_{n,i}(\{s\}); d_X(s,t) < \varepsilon\}) d\varepsilon \leq K\beta(a_n). \tag{12.18}$$

Let $m'_{n,i} = \frac{1}{2}(\delta_{t(n,i)} + m_{n,i})$ and let

$$m' = \sum_{n \geq 1} \sum_{i \leq p(n)} n^{-2} p(n)^{-1} m'_{n,i},$$

so $|m'| \leq 1$. There is a probability m on T such that $m \geq m'$. Fix t in T, $0 < \eta \leq D$. Let n be the smallest integer with $a_n \leq \eta$, so $\eta \leq 2a_n$. We note that if $t \in B_{n,i}$,

$$\sup\{m(\{s\}); d_X(s,t) < \varepsilon\} \geq \frac{1}{2n^2 p(n)} \sup\{m'_{n,i}(\{s\}); d_X(s,t) < \varepsilon\}.$$

Also, for $\varepsilon \geq \min(d_{n,i}, a_n)$,

$$\sup\{m(\{s\}); d_X(s,t) < \varepsilon\} \geq \frac{1}{2n^2 p(n)}.$$

From (12.18) therefore, if $t \in B_{n,i}$,

$$\begin{aligned}
\int_0^\eta h(\sup\{m(\{s\}); d_X(s,t) < \varepsilon\}) d\varepsilon &\leq \int_0^{d_{n,i}} h(\sup\{m_{n,i}(\{s\}); d_X(s,t) < \varepsilon\}) d\varepsilon + \eta h\left(\frac{1}{2n^2 p(n)}\right) \\
&\leq K\beta(a_n) + \eta h\left(\frac{1}{p(n)}\right) h\left(\frac{1}{2n^2}\right) \\
&\leq K\beta(\eta) + \eta h(N(T, d_X; \frac{\eta}{2})^{-1}) + \eta h(\frac{1}{2}(\log 2)^2 (\log \frac{D}{\eta})^{-2})
\end{aligned}$$

since $a_n \leq \eta \leq 2a_n$. Now, if the Gaussian process X is continuous on (T, d_X), $\lim_{\eta \to 0} \beta(\eta) = 0$ by the integrability properties of Gaussian random vectors. Further, if $h(x) = (\log(1/x))^{1/2}$, we see from Corollary 3.19 that

$$\lim_{\eta \to 0} \eta h(N(T, d_X; \eta)^{-1}) = 0. \tag{12.19}$$

These observations conclude the proof of Theorem 12.9.

The two preceding theorems therefore describe necessary majorizing measure conditions for a Gaussian process to have bounded or continuous sample paths that, together with Theorem 11.18, thus provide a complete description of regularity properties of Gaussian processes. In the last part of this section, we describe a consequence of this result in terms of a convex hull representation of Gaussian processes.

Theorem 12.10. Let $X = (X_t)_{t \in T}$ be a bounded Gaussian process. Let $\sigma = \sup_{t \in T} (\mathbb{E}|X_t|^2)^{1/2}$ and $M = \mathbb{E} \sup_{t \in T} |X_t|$. Then there exists a Gaussian sequence $(Y_n)_{n \geq 1}$ with $\|Y_n\|_2 \leq KM(\log n + M^2/\sigma^2)^{-1/2}$ such that, for each t in T, we can write

$$X_t = \sum_{n \geq 1} \alpha_n(t) Y_n$$

where $\alpha_n(t) \geq 0$, $\sum_{n \geq 1} \alpha_n(t) \leq 1$ and the series converges almost surely and in L_2. Moreover, each Y_n is a linear combination of at most two variables of the type X_t. If (and only if) X is continuous, (Y_n) can be chosen such that $\lim_{n \to \infty} (\log n)^{1/2} \|Y_n\|_2 = 0$.

Before the proof, let us mention that if (Y_n) is as in the theorem, by the Borel-Cantelli lemma and Gaussian tail, it defines an almost surely bounded sequence. We even have that, for some numerical constant K_1,

$$\mathbb{P}\{\sup_{n \geq 1} |Y_n| > K_1(M + u\sigma)\} \leq K_1 \exp(-u^2)$$

for all $u > 0$. Indeed, if K is such that $\|Y_n\|_2 \leq KM(\log n + M^2/\sigma^2)^{-1/2}$ in Theorem 12.10, we have

$$\begin{aligned} \mathbb{P}\{\sup_{n \geq 1} |Y_n| > 2K(M + u\sigma)\} &\leq 2 \sum_{n \geq 1} \exp(-4K^2(M + u\sigma)^2 / 2\|Y_n\|_2^2) \\ &\leq 2 \exp(-u^2) \sum_{n \geq 1} n^{-2}. \end{aligned}$$

Since $\sup_{t \in T} |X_t| \leq \sup_{n \geq 1} |Y_n|$, we note in particular that the majorizing measure theorem contains, at least qualitatively, the tail behavior of isoperimetric nature of norms of Gaussian random vectors (cf. Section 3.1). We shall partially come back to this at the end of the section. The representation of Theorem 12.10 also implies, with a little more effort, that X is continuous when $\lim_{n \to \infty} (\log n)^{1/2} \|Y_n\|_2 = 0$.

Proof. We only show the assertion concerning boundedness, the continuous case being obtained with some easy modifications on the basis of Theorem 12.9. Let k_0 be the largest integer with $2^{-k_0} \geq D(T)$. It follows from Theorem 12.8 that there is a probability measure m on T such that for each t in T, we have

$$\sum_{k\geq k_0} 2^{-k} h(\sup\{m(\{s\}); d_X(s,t) < 2^{-k}\}) \leq KM\,. \tag{12.20}$$

For t in T, $k \geq k_0$, we pick t_k such that $d_X(t,t_k) < 2^{-k}$ and

$$m(\{t_k\}) = \sup\{m(\{s\}); d_X(s,t) < 2^{-k}\}\,.$$

We can assume that t_{k_0} does not depend on t. From (12.20), we see in particular that $2^{-k}h(m(\{t_k\})) \leq KM$. Thus, for each t in T, t_k belongs to the finite set

$$A_k = \{s \in T\,;\ m(\{s\}) \geq h^{-1}(2^k KM)\}\,.$$

Let $b_k = 2^{-k+k_0-1}h^{-1}(M/D)$. Using the fact that $D \leq 2\sigma \leq (2\pi)^{1/2}M$, it follows from (12.5) and (12.20) that

$$\sum_{k\geq k_0} 2^{-k} h(b_k m(\{t_k\})) \leq K_1 M \tag{12.21}$$

for some constant K_1. For each t in T, $k \geq k_0$, we define

$$a_{t,k} = 2^{-k}(h(b_k m(\{t_k\})) + h(b_{k+1} m(\{t_{k+1}\})))\,.$$

Form (12.21), $\sum\limits_{k\geq k_0} a_{t,k} \leq 3K_1M$. Define then, for $k \geq k_0$,

$$z_{t,k} = 6K_1 M a_{t,k}^{-1}(X_{t_{k+1}} - X_{t_k})\,.$$

Let Z_k be the set of all $z_{t,k}$ for t in T. Since t_k, t_{k+1} belong to the finite set A_{k+1}, Z_k is finite. Let Z be the union of the sets Z_k for $k > k_0$. Fix $\varepsilon > 0$. We note that

$$\|X_{t_{k+1}} - X_{t_k}\|_2 \leq d_X(t,t_{k+1}) + d_X(t,t_k) \leq 3 \cdot 2^{-k-1}$$

so, if $\|z_{t,k}\|_2 \geq \varepsilon$, we have $a_{t,k} \leq 9 \cdot 2^{-k} K_1 M/\varepsilon$ and thus

$$h(b_k m(\{t_k\})) + h(b_{k+1} m(\{t_{k+1}\})) \leq 9K_1 M/\varepsilon\,.$$

This implies that

$$m(\{t_k\}), m(\{t_{k+1}\}) \geq 2^{k-k_0+1}\left(h^{-1}\left(\frac{M}{D}\right)\right)^{-1} h^{-1}\left(\frac{9K_1M}{\varepsilon}\right).$$

Since m is a probability, this shows that there are at most

$$2^{-k+k_0-1}h^{-1}\left(\frac{M}{D}\right)\left(h^{-1}\left(\frac{9K_1M}{\varepsilon}\right)\right)^{-1}$$

possible choices for either t_k or t_{k+1} when $\|z_{t,k}\|_2 \geq \varepsilon$. Therefore

$$\operatorname{Card}\{z \in Z_k; \|z\|_2 \geq \varepsilon\} \leq 2^{-2k+2k_0-2}\left[h^{-1}\left(\frac{M}{D}\right)\left(h^{-1}\left(\frac{9K_1M}{\varepsilon}\right)\right)^{-1}\right]^2$$

and

$$\operatorname{Card}\{z \in Z; \|z\|_2 \geq \varepsilon\} \leq \left[h^{-1}\left(\frac{M}{D}\right)\left(h^{-1}\left(\frac{9K_1M}{\varepsilon}\right)\right)^{-1}\right]^2.$$

We can thus index Z as a sequence $(Y_n)_{n\geq 1}$ such that $\|Y_n\|_2$ does not increase. For each n,

$$n \leq \operatorname{Card}\{z \in Z; \|z\|_2 \geq \|Y_n\|_2\} \leq \left[h^{-1}\left(\frac{M}{D}\right)\left(h^{-1}\left(\frac{9K_1M}{\|Y_n\|_2}\right)\right)^{-1}\right]^2$$

so that

$$\|Y_n\|_2 \leq 9K_1M\left(\frac{M}{D} + h\left(\frac{1}{\sqrt{n}}\right)\right)^{-1}. \tag{12.22}$$

Since $D \leq 2\sigma$, for $h(x) = (\log(1/x))^{1/2}$, (12.22) implies that, for all $n \geq 1$, $\|Y_n\|_2 \leq 18K_1M(\log n + M^2/\sigma^2)^{-1/2}$. In particular, by the Borel-Cantelli lemma, the Gaussian sequence (Y_n) is bounded almost surely. For each t in T, we have

$$X_t - X_{t_{k_0}} = \sum_{k\geq k_0}(X_{t_{k+1}} - X_{t_k}) = \sum_{k\geq k_0}(6K_1M)^{-1}a_{t,k}z_{t,k}.$$

Since $\sum_{k\geq k_0} a_{t,k} \leq 3K_1M$, this implies that $X_t - X_{t_{k_0}} = \sum_{n\geq 1}\alpha_n(t)Y_n$ where $\sum_{n\geq 1}\alpha_n(t) \leq 1/2$ and where the series converges almost surely since (Y_n) is bounded. Since $\|X_{t_{k_0}}\|_2 \leq \sigma$, the proof of Theorem 12.10 is complete.

Remark 12.11. This remark is in order to show that several of the techniques developed in the proofs of the preceding theorems go beyond the Gaussian case and apply to a rather general setting. As we noted, the Gaussian structure and Slepian's lemma were actually only basic in the proof of Proposition 12.7. For simplicity, let us consider the family of Young functions ψ_q , $1 \leq q \leq \infty$ (and associated functions $h_q(x) = (\log(1/x))^{1/q}$) although some more general functions may be imagined. If (T, d) is a metric space and m is probability measure on (T, d) , set, for $1 \leq q \leq \infty$,

$$\gamma_m^{(q)}(T) = \gamma_m^{(q)}(T, d) = \sup_{t \in T} \int_0^\infty h_q(m(B(t, \varepsilon)))d\varepsilon$$

and

$$\gamma^{(q)}(T) = \gamma^{(q)}(T, d) = \inf \gamma_m^{(q)}(T, d)$$

where the infimum is taken over all probability measures m on T . (It might be useful to recall (11.18) at this point and the easy comparison between ψ_q^{-1} and h_q). K_q denotes below a constant depending only on q and not necessarily the same at each occurence. Our first observation, which we deduce from the proof of Theorem 12.8, is that there is a probability m on (T, d) such that

$$\sup_{t \in T} \int_0^\infty h_q(\sup\{m(\{s\}); d(s, t) < \varepsilon\})d\varepsilon \leq K_q \gamma^{(q)}(T, d) .$$

If T is finite, Proposition 11.10 indeed indicates that $\alpha(T, d) \leq K_q \gamma^{(q)}(T, d)$ (where α is defined with $h = h_q$). The proof of Theorem 12.8, which is given with a general function h , then implies the result. There is a similar result about continuous majorizing measures. From this observation, let us mention further the following one. Consider a stochastic process $X = (X_t)_{t \in T}$ continuous in L_1 (say) on (T, d) , more precisely such that $\|X_s - X_t\|_1 \leq d(s, t)$ for all s, t in T . Assume again that $\gamma^{(q)}(T, d) < \infty$. Then, from the preceding, the proof of Theorem 12.10 shows that there is a sequence (Y_n) of random variables such that, for every n ,

$$\|Y_n\|_1 \leq K_q \gamma^{(q)}(T, d) \left(\frac{\gamma^{(q)}(T, d)}{D(T)} + h_q\left(\frac{1}{\sqrt{n}}\right)\right)^{-1}$$

and such that, for every t , $X_t = \sum_n \alpha_n(t) Y_n$ where $\alpha_n(t) \geq 0$, $\sum_n \alpha_n(t) \leq 1$ and the series converges alsmost surely and in L_1 . The main point in the Gaussian case is of course that $\gamma^{(2)}(T, d_X) \leq K \mathbb{E} \sup_{t \in T} X_t$.

It is interesting to interpret Theorem 12.10 (and similar comments may be given in the context of the observations of Remark 12.11) as a result about subsets of Hilbert space associated to bounded Gaussian processes. Let $X = (X_t)_{t\in T}$ be a bounded Gaussian process on $(\Omega, \mathcal{A}, \mathbb{P})$. Let $H = L_2(\Omega, \mathcal{A}, \mathbb{P})$ and identify T with the subset of H consisting of the family $(X_t)_{t\in T}$. Then, Theorem 12.10 proves the existence of a sequence (y_n) in H such that, for some $M > 0$ and all n, $|y_n| \le M(\log(n+1))^{-1/2}$, and $T \subset \operatorname{Conv}(y_n)$. Let us rewrite this observation as a perhaps more geometrical statement about the finite dimensional Hilbert space. Consider H of dimension N and denote by σ normalized measure on its unit ball. For a subset T of H, consider

$$V(T) = \int \sup_{y\in T} |\langle x, y\rangle| d\sigma(x) .$$

This quantity has been studied in geometry under the name of mixed volume and plays an important role in the local theory of Banach spaces. Fix an orthonormal basis $(e_i)_{i\le N}$ of H. For t in H, let $X_t = \sum_{i=1}^{N} g_i\langle t, e_i\rangle$ where (g_i) is a standard Gaussian sequence. Since the distribution of $(g_i)_{i\le N}$ is rotation invariant,

$$\ell(T) = \mathbb{E}\sup_{t\in T}|X_t| = \mathbb{E}\left(\sum_{i=1}^{N} g_i^2\right)^{1/2} \int \sup_{y\in T}|\langle x, y\rangle| d\sigma(x)$$

and thus $K^{-1}N^{1/2}V(T) \le \ell(T) \le KN^{1/2}V(T)$ for some numerical K. Define now

$$C(T) = \inf\{a > 0\,;\ \exists (y_n) \text{ in } H\,,\ |y_n| \le a(\log(n+1))^{-1/2}\,,\ T \subset \operatorname{Conv}(y_n)\} .$$

Then Theorem 12.10 can be reformulated as

$$K^{-1}N^{-1/2}C(T) \le V(T) \le KN^{-1/2}C(T) . \tag{12.23}$$

As a closing remark, note that if we go back to the tail estimate (11.14) for the Young function $\psi = \psi_2$ associated to Gaussian processes, we see that the processes techniques and existence of majorizing measures (12.4) yield deviation inequalities with the two parameters similar to those obtained from isoperimetric considerations in Chapter 3. Such inequalities can also be deduced as we have seen from Theorem 12.10 and elementary considerations on Gaussian sequences.

12.2. Necessary conditions for boundedness and continuity of stable processes

Let $X = (X_t)_{t\in T}$ be a p-stable process, $0 < p < 2$. Recall that by this we mean that every finite linear combination $\sum_i \alpha_i X_{t_i}$, $\alpha_i \in \mathbb{R}$, $t_i \in T$, is a p-stable real random variable. As in the Gaussian case, we may consider the associated pseudo-metric d_X on T given by

$$d_X(s,t) = \sigma(X_s - X_t)\,, \qquad s,t \in T ,$$

where $\sigma(X_s - X_t)$ is the parameter of the real stable random variable $X_s - X_t$ (cf. Chapter 5).

Contrary to the Gaussian case, the pseudo-metric d_X does not entirely determine the distribution and therefore the regularity properties of a p-stable process X when $0 < p < 2$. The distribution of X is however determined by its spectral measure (Theorem 5.2) and, as we already noted in Chapter 5 through the representation, the existence and finiteness of a spectral measure for a p-stable process with $0 < p < 1$ already ensures its almost surely boundedness and continuity. This is no more the case for $1 \le p < 2$ on which we concentrate here. If a characterization in term of d_X is therefore hopeless, nevertheless a best possible necessary majorizing measure condition for (T, d_X), similar to the one of the Gaussian case, for almost sure boundedness and continuity of p-stable processes, $1 \le p < 2$, exists. It is the purpose of this paragraph to describe this result.

As we mentioned, the difference with the Gaussian setting is that this necessary condition is far from being sufficient. Sufficient conditions for sample boundedness or continuity of p-stable processes may be obtained from Theorem 11.2 at some weak level; for example, if $1 < p < 2$, since

$$\|X_s - X_t\|_{p,\infty} \le C_p d_X(s,t)\,, \qquad s, t \in T\,,$$

the finiteness of the entropy integral

$$\int_0^D (N(T, d_X; \varepsilon))^{1/p}\, d\varepsilon$$

implies that X has a version with almost all paths bounded and continuous on (T, d_X). The difference however between necessity described below in Theorem 12.12 and this sufficient condition is huge. Trying to characterize regularity properties of p-stable processes, $1 \le p < 2$, seems to be a difficult question, still under study. The paper [Ta13] reflects for example some of the main problems.

For $1 \le p \le 2$, let q be the conjugate of p. For $p > 1$, $0 < x \le 1$, recall we set $h_q(x) = (\log(1/x))^{1/q}$; further $h_\infty(x) = \log(1 + \log(1/x))$. These functions satisfy (12.5) and (12.6) and enter the setting of the abstract analysis of the preceding section. The main result of this section is the extension to p-stable processes, $1 \le p < 2$, of Theorems 12.8 and 12.9.

Theorem 12.12. Let $1 \le p < 2$. Let $X = (X_t)_{t \in T}$ be a sample bounded p-stable process. Then there is a probability measure m on (T, d_X) such that

$$\sup_{t \in T} \int_0^\infty h_q(\sup\{m(\{s\}); d_X(s,t) < \varepsilon\}) d\varepsilon \le K_p \| \sup_{s,t \in T} |X_s - X_t| \|_{p,\infty}$$

where K_p only depends on p. (In particular, $\gamma_m^{(q)}(T,d_X) \le K_p \| \sup_{s,t\in T} |X_s - X_t| \|_{p,\infty}$.) If, moreover, X has (or admits a version with) almost surely continuous sample paths on (T,d_X), there exists a probability measure m on (T,d_X) such that

$$\lim_{\eta\to 0} \sup_{t\in T} \int_0^\eta h_q(\sup\{m(\{s\}); d_X(s,t) < \varepsilon\}) d\varepsilon = 0 .$$

As we noted (Remark 12.11), provided with such a result, Theorem 12.10 has an extension to the stable case. This observation puts into light the gap between this necessity result and sufficient conditions for stable processes to be bounded or continuous simply because if (θ_n) is a standard p-stable sequence with $1 < p < 2$ (say), $(\theta_n/(\log n)^{1/q})$ is not almost surely bounded. For some special class of stationary p-stable process, strongly or harmonizable stationary process, we will see however in Chapter 13 that the necessary conditions of Theorem 12.12 are also sufficient.

We only give the proof of Theorem 12.12 for $p > 1$. The proof when $p = 1$ follows the same pattern with however some further (delicate) arguments inherent to this case. We restrict for clarity to $p > 1$ refering to [Ta8] for the complete result. In the following therefore $1 < p < 2$. To put Theorem 12.12 in perspective, let us go back to Chapter 5, Section 5.3. There we saw how for a p-stable process X (Theorem 5.10),

$$\sup_{\varepsilon>0} \varepsilon(\log N(T,d_X;\varepsilon))^{1/q} \le K_p \| \sup_{t\in T} |X_t| \|_{p,\infty} .$$

(By (11.9), Theorem 12.12 improves upon this result.) The main idea of the proof of this Sudakov type minoration was to realize the stable process $X = (X_t)_{t\in T}$ as conditionally Gaussian by the series representation of stable variables. We need not go back to all the details here and refer to the proof of Theorem 5.10 for this point. It was described there how $(X_t)_{t\in T}$ has the same distribution as $(X_t^\omega)_{t\in T}$ defined on $\Omega\times\Omega'$ where, for each ω in Ω, $(X_t^\omega)_{t\in T}$ is Gaussian. If d_ω is the canonical metric associated to the Gaussian process $(X_t^\omega)_{t\in T}$, the main tool was comparison between the random distances d_ω and d_X given by (cf. (5.20))

$$\mathbb{P}\{\omega; d_\omega(s,t) \le \varepsilon d_X(s,t)\} \le \exp(-c_\alpha \varepsilon^{-\alpha}) \tag{12.24}$$

for all s,t in T, $\varepsilon > 0$, where $1/\alpha = 1/p - 1/2$ and c_α only depends on α. From (12.24), the idea was to "transfer" Gaussian minoration inequalities of Sudakov's type on the random distances d_ω into similar ones for d_X. The idea of the proof of Theorem 12.12 is similar, trying to transfer the stronger minoration results of Section 12.1. It turns out unfortunately that it does not seem possible to use mixtures of majorizing measures. Instead, we are going to use the machinery of ultrametric structures of the last section to reduce the proof to the simpler, yet non-trivial, following property.

Theorem 12.13. Let $1 < p < 2$ and let $X = (X_t)_{t\in T}$ be a p-stable process indexed by a finite set T. Provide T with the canonical distance d_X. Then

$$\alpha(T, d_X) \le K_p \| \sup_{s,t\in T} |X_s - X_t| \|_{p,\infty}$$

where K_p only depends on p.

This result is the analog of the Gaussian Theorem 12.6 and the functional α is the one used in Section 12.1 with $h = h_q$. We noted there that, with similar proofs than in the Gaussian case, Theorem 12.12 follows from Theorem 12.13 (for continuity, note that (12.19) holds with $h = h_q$ by (5.21)). We therefore concentrate now on the proof of the latter. Recall that if (U, δ) is ultrametric, we denote by $\mathcal{B}_k$ the family of balls of U of radius 6^{-k} ($k \in \mathbb{Z}$). Let k_0 be the largest such that $\mathrm{Card}\mathcal{B}_{k_0} = 1$. Let us set for simplicity $\Delta = 6^{-k_0}$. For x in U, we denote by $N_k(x) = N(x,k)$ the number of disjoint balls of $\mathcal{B}_{k+1}$ which are contained in $B(x, 6^{-k})$. We set

$$\xi_x^{(q)}(U) = \sum_{k\ge k_0} 6^{-k} (\log N(x,k))^{1/q} ,$$
$$\xi^{(q)}(U) = \inf_{x\in U} \xi_x^{(q)}(U) .$$

By Theorem 12.5 applied to (T, d_X) and $h = h_q$, we see that in order to establish Theorem 12.13, it is enough to prove the analog of Proposition 12.7 in this stable case. That is, if (U, δ) is a finite ultrametric space and $X = (X_u)_{u\in U}$ a p-stable process with $d_X \ge \delta$, then we have

$$\xi^{(q)}(U) \le K_p \| \sup_{u,v\in U} |X_u - X_v| \|_{p,\infty} . \tag{12.25}$$

Since the arguments of the proof of (12.25) and Theorem 12.13 have already proved their usefulness in some other contexts, we will try to detail (at least the first steps of) this study in a possible general context. Let us set

$$M = \| \sup_{u,v\in U} |X_u - X_v| \|_{p,\infty} .$$

X is conditionally Gaussian and we denote by $X^\omega = (X_u^\omega)_{u\in U}$, for every ω in Ω , the conditional Gaussian processes (see the proof of Theorem 5.10). By Fubini's theorem and definition of M , there exists a set $\Omega_1 \subset \Omega$ with $\mathbb{P}(\Omega_1) \geq 1/2$ and such that for ω in Ω_1 we have

$$\mathbb{P}\{\sup_{u,v\in U} |X_u^\omega - X_v^\omega| > 4M\} \leq \frac{1}{2}\,.$$

By the integrability properties of norms of Gaussian processes (Corollary 3.2) and Theorem 12.6 we know the following: for ω in Ω_1 , there exists a probability μ_ω on (U, d_ω) where $d_\omega(u,v) = \|X_u^\omega - X_v^\omega\|_2$, such that

$$\sup_{x\in U} \int_0^\infty h_2(\mu_\omega(y\in U\,;\ d_\omega(x,y)<\varepsilon))d\varepsilon \leq KM\,. \tag{12.26}$$

where K is a numerical constant.

The function $h_2(t) = (\log(1/t))^{1/2}$ is convex for $t \leq e^{-1/2}$ but not for $0 < t \leq 1$. For that reason, it will be more convenient to use the function $h_2'(t) = h_2(t/3)$ (the choice of 3 is rather arbitrary) which is convex on $(0,1]$. To prove (12.25), and therefore Theorem 12.13, we will exhibit a subset Ω_2 of Ω with $\mathbb{P}(\Omega_2) \geq 3/4$ (for example), and constants K_1, K_2, K_3 depending on p on ly, such that for ω in Ω_2 and each probability measure μ on U , we have

$$\xi^{(q)}(U) \leq K_1 \sup_{x\in U} \int_0^{K_2\Delta} h_2'(\mu(y\in U\,;\ d_\omega(x,y)<\varepsilon))d\varepsilon + K_3\Delta\,. \tag{12.27}$$

We observe that $h_2'(t) \leq h_2(t) + (\log 3)^{1/2}$; so combining with (12.26) since $\mathbb{P}(\Omega_1 \cap \Omega_2) > 0$, we get that $\xi^{(q)}(U) \leq K_1(KM + K_2(\log 3)^{1/2}\Delta) + K_3\Delta$. It is easily seen that $\Delta \leq K_4 M$ so that (12.25) holds.

We thus establish now (12.27). The philosophy of the approach is that a large value of $\xi^{(q)}(U)$ means that (U,δ) is big in an appropriate sense. Since (12.24) means that $d_\omega(u,v)$ is, most of the time, not much smaller than $d_X(u,v) \geq \delta(u,v)$, we can expect that U will be big with respect to d_ω for most values of ω . The construction is made rather delicate however by the following feature: while (12.24) tells us precisely how $d_\omega(u,v)$ behaves compared to $d_X(u,v)$, if we take another couple (u',v') , we have no information about the joint behavior of $(d_\omega(u,v)\,,\ d_\omega(u',v'))$.

As announced, we will try to develop the proof of (12.25) in a possible general framework. Let us therefore assume that we have a family (d_ω) of random distances on (U,δ) such that for some strictly increasing function θ on $\mathbb{R}_+$ with $\lim_{\varepsilon\to 0} \theta(\varepsilon) = 0$, and all u, v in U and $\varepsilon > 0$,

$$\mathbb{P}\{\omega\,;\ d_\omega(u,v) \leq \varepsilon\delta(u,v)\} \leq \theta(\varepsilon)\,. \tag{12.28}$$

By (12.24), the stable case corresponds to $\theta(\varepsilon) = \exp(-c_\alpha \varepsilon^{-\alpha})$ where $1/\alpha = 1/p - 1/2$.

To show (12.27), it will be enough to show that for some probability measure λ on U

$$\begin{aligned} &K_5^{-1} \int_U d\lambda(x) \xi_x^{(q)}(U) \\ &\le K_1 \int_U d\lambda(x) \int_0^{K_2\Delta} h_2'(\mu(y \in U\,;\ d_\omega(x,y) < \varepsilon)) d\varepsilon + K_3 \Delta\,. \end{aligned} \tag{12.29}$$

We choose as a convenient probability measure λ the following: λ is homogeneous in the sense that the mass of any ball of radius 6^{-k} is devided evenly among all the balls of radius 6^{-k-1} that it contains. Equivalently, for any x in U, $\lambda(\{x\}) = \left(\prod_{k \ge k_0} N(x,k)\right)^{-1}$. Let now $k \ge k_0$ be fixed. Let $B_1 \ne B_2$ in $\mathcal{B}_k$. Let further $b, c > 0$ and define

$$A(B_1, B_2, b, c) = \{\omega \in \Omega\,;\ \lambda \otimes \lambda((x,y) \in B_1 \times B_2\,;\ d_\omega(x,y) \le b6^{-k}) \ge c\lambda(B_1)\lambda(B_2)\}\,.$$

Under (12.28), one checks immediately by Fubini's theorem and Chebyshev's inequality that

$$\mathbb{P}(A(B_1, B_2, b, c)) \le \frac{\theta(b)}{c}\,. \tag{12.30}$$

For $B \in \mathcal{B}_k$ and $i \le k$, there is a unique $B' \in \mathcal{B}_i$ that contains B. We denote by $N(B,i)$ the number of elements of $\mathcal{B}_{i+1}$ contained in B', so $N(B,i) = N(x,i)$ whenever x belongs to B. In particular, $N(B,k)$ is the number of elements of $\mathcal{B}_{k+1}$ that are contained in B. Also, if $i \le k' \le k$ and $B \in \mathcal{B}_k$, $B' \in \mathcal{B}_{k'}$, $B \subset B'$, we have $N(B,i) = N(B',i)$. We denote by $\mathcal{B}_k'$ the subset of $\mathcal{B}_k$ that consists of the balls B in $\mathcal{B}_k$ for which

$$N(B,k) > 2 \prod_{i<k} N(B,i)\,.$$

Let B in $\mathcal{B}_k'$, so that in particular $N(B,k) > 1$. There exist therefore two balls B_1, B_2 in $\mathcal{B}_{k+1}$, $B_1, B_2 \subset B$, $B_1 \ne B_2$. We consider the event

$$C(B, B_1, B_2) = A(B_1, B_2, b, (2N(B,k))^{-2})$$

where $b = b(B,k)$ is chosen as

$$b = \theta^{-1}((2N(B,k))^{-6})\,. \tag{12.31}$$

It follows from (12.30) that $\mathbb{P}(C(B,B_1,B_2)) \le (2N(B,k))^{-4}$.

For D in $\mathcal{B}_k$, we consider the event

$$\overline{A}(D) = \bigcup C(B,B_1,B_2)$$

where the union is taken over all choices of $j \ge k$, B in $\mathcal{B}'_j$, $B \subset D$, B_1, B_2 in $\mathcal{B}_{j+1}$, $B_1 \ne B_2$, $B_1, B_2 \subset B$. (If no such choice is possible, we set $\overline{A}(D) = \emptyset$.)

Lemma 12.14. Under the previous notations,

$$\mathbb{P}(\overline{A}(D)) \le (2\prod_{i<k} N(D,i))^{-2} .$$

Proof. The proof goes by decreasing induction over k . If k is large enough that D has only one point, then $\overline{A}(D) = \emptyset$ and the lemma holds. Assume now that we have proved the lemma for $k+1$ and let D in $\mathcal{B}_k$. Assume first that $D \in \mathcal{B}'_k$, so $N(D,k) \ge 2$. Let

$$\begin{aligned} \overline{A}_1 &= \bigcup\{\overline{A}(D')\,;\ D' \in \mathcal{B}_{k+1}\ ,\ D' \subset D\}\ , \\ \overline{A}_2 &= \bigcup\{C(D,B_1,B_2)\,;\ B_1, B_2 \in \mathcal{B}_{k+1}\ ,\ B_1 \ne B_2\ ,\ B_1, B_2 \subset D\}\ . \end{aligned}$$

We have $\overline{A}(D) = \overline{A}_1 \cup \overline{A}_2$ and by the induction hypothesis, since $N(D,k) \ge 2$,

$$\mathbb{P}(\overline{A}_1) \le N(D,k)(2\prod_{i\le k} N(D,i))^{-2} \le \frac{1}{2}(2\prod_{i<k} N(D,i))^{-2} .$$

Using that $\mathbb{P}(C(D,B_1,B_2)) \le (2N(B,k))^{-4}$,

$$\begin{aligned} \mathbb{P}(\overline{A}_2) &\le \frac{1}{2}N(D,k)^2 \cdot (2N(D,k))^{-4} \\ &\le \frac{1}{2}N(D,k)^{-2} \le \frac{1}{2}(2\prod_{i<k} N(D,i))^{-2} . \end{aligned}$$

The result therefore follows in this case. If $D \notin \mathcal{B}'_k$, with the same notation as before, $\overline{A}(D) = \overline{A}_1$. Thus, by the induction hypothesis,

$$\mathbb{P}(\overline{A}(D)) \le N(D,k)(2\prod_{i\le k} N(D,i))^{-2} \le (2\prod_{i<k} N(D,i))^{-2} .$$

This completes the proof of Lemma 12.14.

Let now $\Omega_2 = \Omega \backslash \overline{A}(U)$, so $\mathbb{P}(\Omega_2) \geq 3/4$. Let us fix furthermore ω in Ω_2 . For B in $\mathcal{B}'_k$, we set

$$a(B,k) = 6^{-k-2} b(B,k)$$

where $b(B,k)$ is given by (12.31). For x in B , set

$$H_x = \{z \in U\, ;\ d_\omega(x,z) \leq a(B,k)\}\,.$$

We then have that, for all y in U ,

$$\lambda(x \in B\, ;\ y \in H_x) \leq \frac{3\lambda(B)}{2N(B,k)}\,. \tag{12.32}$$

Indeed, suppose otherwise and let y in U such that (12.32) does not hold. For D in $\mathcal{B}_{k+1}$, $D \subset B$, we have $\lambda(D) = \lambda(B)/N(B,k)$. It follows that there are at least two balls B_1, B_2 of $\mathcal{B}_{k+1}$, $B_1, B_2 \subset B$ such that, for $\ell = 1, 2$,

$$\lambda(x \in B_\ell\, ;\ y \in H_x) \geq \frac{\lambda(B_\ell)}{2N(B,k)}\,.$$

For x_1, x_2 in H_y , $d_\omega(x_1,x_2) \leq 2a(B,k)$; so we have

$$\lambda \otimes \lambda((x_1,x_2) \in B_1 \times B_2\, ;\ d_\omega(x_1,x_2) \leq 2a(B,k)) \geq \lambda(B_1)\lambda(B_2)/4N(B,k)^2\,.$$

This, however, contradicts the fact that $\omega \notin C(B,B_1,B_2)$ (by definition of Ω_2) and thus shows (12.32).

Let μ be a probability measure on U . Since the function h'_2 is convex, we have

$$I(B) = \frac{1}{\lambda(B)} \int_B h'_2(\mu(H_x))d\lambda(x) \geq h'_2\Big(\int g\, d\mu\Big)$$

where $g(y) = \lambda(x \in B\, ;\ y \in H_x)/\lambda(B)$. It follows from (12.32) that $0 \leq g \leq 3/2N(B,k)$, so that, by definition of h'_2 , $I(B) \geq (\log(2N(B,k))^{1/2}$. In particular therefore

$$a(B,k)I(B) \geq 6^{-k-2} b(B,k)(\log(2N(B,k))^{1/2}\,. \tag{12.33}$$

For x in U , let us enumerate as $k_1(x) < \cdots < k_{\ell(x)}(x)$ the indexes k such that $B(x, 6^{-k}) \in \mathcal{B}'_k$. Note that $k_1(x) = k_0$. For $\ell \leq \ell(x)$, let

$$c(x,\ell) = a(B(x, 6^{-k_\ell(x)}),\ k_\ell(x))\,.$$

We have

$$\int_U \sum_{\ell\le\ell(x)} c(x,\ell)h_2'(\mu(y\in U\,;\ d_\omega(x,y)<c(x,\ell)))d\lambda(x)$$

(12.34)

$$=\sum\int_B a(B,k)h_2'(\mu(y\in U\,;\ h_\omega(x,y)\le a(B,k)))d\lambda(x)$$

where the summation is taken over each value of k and each B in $\mathcal{B}_k'$. By (12.33), we see that the latter quantity (12.34) dominates

$$\sum 6^{-k-2}b(B,k)(\log(2N(B,k)))^{1/2}\lambda(B) \tag{12.35}$$

where the summation is over the same range. Observe that for $\ell<\ell(x)$, $c(x,\ell+1)\le c(x,\ell)/8$ (by definition of $\mathcal{B}_k'$ and since θ is increasing). Also, $c(x,1)\le\theta^{-1}(1/2)\Delta$. It follows that, for every x,

$$\sum_{\ell\le\ell(x)} c(x,\ell)h_2'(\mu(y\in U\,;\ d_\omega(x,y)<c(x,\ell)))$$
$$\le 2\int_0^{\theta^{-1}(1/2)\Delta} h_2'(\mu(y\in U\,;\ d_\omega(x,y)<\varepsilon))d\varepsilon\,.$$

Let us summarize in a statement what we have obtained so far in this general approach. This statement could possibly be of some use in a related context.

Proposition 12.15. Let (U,δ) be an ultrametric space and (d_ω) a family of random distances on (U,δ) such that for some increasing function θ on $\mathbb{R}_+$ such that $\lim_{\varepsilon\to0}\theta(\varepsilon)=0$ and all u,v in U and $\varepsilon>0$,

$$\mathbb{P}\{\omega\,;\ d_\omega(u,v)\le\varepsilon\delta(u,v)\}\le\theta(\varepsilon)\,.$$

Then, there exists Ω_0 with $\mathbb{P}(\Omega_0)\ge3/4$ such that for all ω in Ω_0 and every probability measure μ on U, in the previous notations,

$$\int_U d\lambda(x)\int_0^{\theta^{-1}(1/2)\Delta} h_2'(\mu(y\in U\,;\ d_\omega(x,y)<\varepsilon))d\varepsilon$$
$$\ge 2^{-7}\sum 6^{-k}\theta^{-1}((2N(B,k))^{-6})(\log(2N(B,k)))^{1/2}\lambda(B)$$

where the summation is taken over each value of k and each B in $\mathcal{B}_k'$.

From this general result, we can now complete the proof of Theorem 12.13 with the choice of θ given by (12.24).

Proof of Theorem 12.13. By (12.24), we take $\theta(\varepsilon) = \exp(-c_\alpha \varepsilon^{-\alpha})$. Then, from (12.31),

$$b(B,k) = \left(\frac{6}{c_\alpha}\log(2N(B,k))\right)^{-1/\alpha} .$$

Since $1/2 - 1/\alpha = 1/q$, the right side of the inequality of Proposition 12.15 is simply $2^{-7}(6/c_\alpha)^{-1/\alpha}\int_U \eta_x(U)d\lambda(x)$ where

$$\eta_x(U) = \sum_{\ell \le \ell(x)} 6^{-k_\ell(x)}(\log(2N(x,k_\ell(x)))^{1/q} .$$

Therefore, in order to establish (12.29) (with $K_2 = (c_\alpha/\log 2)^{1/\alpha}$), we need simply find the appropriate lower bound for $\eta_x(U)$. For x in U , $s < \ell(x)$ and

$$k_0 = k_1(x) < \cdots < k_s(x) \le i < k_{s+1}(x) ,$$

we have

$$N(x,i) \le 2^{2^{i-k_0}} N(x,k_s(x))^{2^{i-k_s(x)}} \ldots N(x,k_1(x))^{2^{i-k_1(x)}}$$

as is shown by immediate induction over i . Therefore,

$$(\log N(x,i))^{1/q} \le (2^{i-k_0}\log 2)^{1/q} + \sum_{\ell=1}^{s}(2^{i-k_\ell(x)}\log N(x,k_\ell(x)))^{1/q} .$$

A simple computation then shows that there are constants K_6, K_7 such that $\xi_x^{(q)}(U) \le K_6\eta_x(U) + K_7\Delta$. This shows (12.29) and concludes the proof of Theorem 12.13.

12.3. Applications and conjectures on Rademacher processes

The first application deals with subgaussian processes. Recall from Section 11.3 ((11.20)) that a centered process $Y = (Y_t)_{t\in T}$ is subgaussian with respect to a metric or pseudometric d if for all s,t in T and λ in $\mathbb{R}$,

$$\mathbb{E}\exp\lambda(Y_s - Y_t) \le \exp\left(\frac{\lambda^2}{2}d^2(s,t)\right) .$$

As a consequence of the general study of Chapter 11, we noted there that the sufficient conditions in order for a Gaussian process to have bounded or continuous sample paths also apply for subgaussian processes. That is, if m is a probability measure on (T,d), then

$$\mathbb{E}\sup_{t\in T} Y_t \le K \sup_{t\in T}\int_0^\infty \left(\log\frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon$$

where K is numerical. But now, as a consequence of the main result of Section 12.1 and in particular Theorem 12.8, we see that if d is the pseudo-metric of a Gaussian process $X=(X_t)_{t\in T}$, then

$$\mathbb{E}\sup_{t\in T} Y_t \le K\mathbb{E}\sup_{t\in T} X_t$$

where K is a numerical constant. We have thus the following statement. Its second part is proved similarly.

Theorem 12.16. Let $X=(X_t)_{t\in T}$ be a Gaussian process and $Y=(Y_t)_{t\in T}$ be subgaussian with respect to the canonical distance d_X associated to X. Then, $\mathbb{E}\sup_{t\in T} Y_t \le K\mathbb{E}\sup_{t\in T} X_t$ for some numerical constant K. In particular, Y is almost surely bounded if X is. Further, if X is continuous, Y has a version with almost all sample paths continuous.

The second application concerns some remarkable type properties of the canonical injection map j : $\mathrm{Lip}(T)\to C(T)$ first considered by J. Zinn. Let (T,d) be any compact metric space with diameter $D=D(T)$. Denote by $C(T)$ the space of continuous functions on T with the supnorm $\|\cdot\|_\infty$ and by $\mathrm{Lip}(T)$ the space of Lipschitz functions f on (T,d) provided with the norm

$$\|f\|_{\mathrm{Lip}} = D^{-1}\|f\|_\infty + \sup_{s\ne t}\frac{|f(s)-f(t)|}{d(s,t)}.$$

Consider the canonical injection map $j:\ \mathrm{Lip}(T)\to C(T)$. For every $1\le p\le 2$, let (θ_i) be a p-stable standard sequence (if $p=2$, $(\theta_i)=(g_i)$, the orthogaussian sequence). Denote by $T_p(j)$ the smallest constant C such that for every finite sequence (x_i) in $\mathrm{Lip}(T)$,

$$\|\,\|\sum_i \theta_i j(x_i)\|_\infty\|_{p,\infty} \le C\left(\sum_i \|x_i\|^p_{\mathrm{Lip}}\right)^{1/p}.$$

Denote further by $T'_p(j)$, $1\le p\le 2$, the smallest constant C such that

$$\mathbb{E}\|\sum_i \varepsilon_i j(x_i)\|_\infty \le C\begin{cases} (\sum_i \|x_i\|^2_{\mathrm{Lip}}) & \text{if } p=2\\[2ex] \|(\|x_i\|_{\mathrm{Lip}})\|_{p,\infty} & \text{if } 1\le p<2.\end{cases}$$

The introduction of those two type constants is motivated by the two possible definitions of p-stable operators as described at the end of Section 9.2. From this section actually, it can be seen that $T_2(j)$ and $T_2'(j)$ are equivalent, that is, for some numerical constant K, $K^{-1}T_2(j) \leq T_2'(j) \leq KT_2(j)$. The second inequality is simply that Gaussian averages "dominate" the corresponding Rademacher ones, while the first inequality is obtained by partial integration and moment equivalences of Gaussian averages. For the same reason as the latter, for $1 \leq p < 2$, $T_p(j) \leq K_p T_p'(j)$ (cf. one portion of the equivalence (iii) in Proposition 9.12). In general however, the type constants T_p and T_p' of general operators between Banach spaces are not equivalent when $1 \leq p < 2$.

What we will discover however is that, for this particular operator j, $T_p(j)$ and $T_p'(j)$ are equivalent for every $1 \leq p \leq 2$, and their finiteness equivalent to the existence of a majorizing measure condition on (T,d) for $h_q(x) = (\log(1/x))^{1/q}$ where q is the conjugate of p ($h_\infty(x) = \log(1+\log(1/x))$ if $p=1$). Recall that if m is a probability measure m on (T,d), we let

$$\gamma_m^{(q)}(T,d) = \sup_{t\in T}\int_0^\infty h_q(m(B(t,\varepsilon)))d\varepsilon\,,$$

and set

$$\gamma^{(q)}(T,d) = \inf \gamma_m^{(q)}(T,d)$$

where the infimum runs over all probability measures m on T. We then have the following theorem.

Theorem 12.17. Let $1 \leq p \leq 2$. There is a constant K_p depending only on p such that

$$K_p^{-1}T_p'(j) \leq \gamma^{(q)}(T,d) \leq K_p T_p(j)\,.$$

Proof. We first prove the left hand side inequality for $p=2$. Let (x_i) be a finite sequence in $\mathrm{Lip}(T)$ such that, by homogeneity, $\sum_i \|x_i\|_{\mathrm{Lip}}^2 = 1$. Let $X = (X_t)_{t\in T}$ where $X_t = \sum_i \varepsilon_i x_i(t)$, $t \in T$, and set $d_X(s,t) = \|X_s - X_t\|_2 = (\sum_i |x_i(s) - x_i(t)|^2)^{1/2}$. Then, since X is subgaussian with respect to d_X (cf. (11.19)), for some numerical constant K,

$$\begin{aligned}\mathbb{E}\|\sum_i \varepsilon_i j(x_i)\|_\infty &= \mathbb{E}\sup_{t\in T}|\sum_i \varepsilon_i x_i(t)| \\ &\leq \sup_{t\in T}\left(\sum_i x_i(t)^2\right)^{1/2} + K\gamma^{(2)}(T,d_X)\,.\end{aligned}$$

Now, by definition of $\|\cdot\|_{\text{Lip}}$, for every s,t in T, $(\sum\limits_i x_i(t)^2)^{1/2} \le D$ and $d_X(s,t) \le d(s,t)$. Thus

$$\begin{aligned}\mathbb{E}\|\sum_i \varepsilon_i j(x_i)\|_\infty &\le D + K\gamma^{(2)}(T,d)\\ &\le K'\gamma^{(2)}(T,d)\end{aligned}$$

where we have used Lemma 12.1 (vi) in the last step. The result thus follows by homogeneity. Using Lemma 11.20, the proof is entirely similar for $1 \le p < 2$.

Let us now establish that $\gamma^{(q)}(T,d) \le K_p T_p(j)$. It is easy to see that if A is a subset of T and j' the canonical injection from $\operatorname{Lip}(A)$ into $C(A)$, then $T_p(j') \le T_p(j)$. On the other hand, we have shown in the proof of Theorem 12.8 that

$$\gamma^{(q)}(T) \le K_p \sup\{\alpha(A)\,;\ A \subset T\,,\ A \ \text{ finite}\}$$

(where, for simpliticy, we do not specify q in α). Hence, it is enough to show that, when T is finite, $\alpha(T) \le K_p T_p(j)$. Here, K_p denotes some constant depending only on p and not necessarily the same in each occurence. We claim that it is enough to show that when (T,d) is finite and ultrametric, there is a finite family (x_i) in $\operatorname{Lip}(T)$ with $(\sum\limits_i \|x_i\|^p_{\text{Lip}})^{1/p} \le 128$ (for example!) and such that if, for t in T, $X_t = \sum\limits_i \theta_i x_i(t)$, then the canonical distance d_X of the p-stable process $X = (X_t)_{t\in T}$ satisfies $d_X \ge d$. Indeed, we would then have $128 T_p(j) \ge \|\sup\limits_{t\in T}|X_t|\|_{p,\infty}$, and, by the conjunction of Theorem 12.5 and Proposition 12.7 for $p=2$, or (12.25) for $1 \le p < 2$, the result. (We use here the fact that (12.25) also holds for $p=1$, cf. [Ta8].)

Let k_0 be the largest such that $4^{-k_0} \ge D$. For $k \ge k_0$, denote by $\mathcal{B}_k$ the family of balls of T of radius 4^{-k}. Since T is finite, there exists k_1 such that $B(x, 4^{-k_1}) = \{x\}$ for every x in T. Let $\mathcal{B} = \bigcup\limits_{k_0 \le k \le k_1} \mathcal{B}_k$. For $\varepsilon = (\varepsilon_B)_{B\in\mathcal{B}} \in \mathcal{E} = \{0,1\}^{\mathcal{B}}$, we define

$$\varphi_\varepsilon = \sum_{k=k_0+1}^{k_1} \sum_{B\in\mathcal{B}_k} 4^{-k}\varepsilon_B I_B\,.$$

We note that $\|\varphi_\varepsilon\|_\infty \le \sum\limits_{k>k_0} 4^{-k} \le 2D$. Consider now s,t in T and let ℓ be the largest such that $d(s,t) \le 4^{-\ell}$. If $B \in \mathcal{B}_m$ for some $m \le \ell$, we have $I_B(s) = I_B(t)$. It follows that

$$|\varphi_\varepsilon(s) - \varphi_\varepsilon(t)| \le \sum_{k>\ell} 4^{-k} \le 2d(s,t)\,.$$

This shows that $\|\varphi_\varepsilon\|_{\mathrm{Lip}} \le 4$. The definition of ℓ shows that the two balls $B_1 = B(s, 4^{-\ell-1})$, $B_2 = B(t, 4^{-\ell-1})$ are different. Since they belong to $\mathcal{B}_{\ell+1}$, we have that

$$\begin{aligned}|\varphi_\varepsilon(s) - \varphi_\varepsilon(t)| &\ge 4^{-\ell-1}|\varepsilon_{B_1}(s) - \varepsilon_{B_2}(t)| - \sum_{k>\ell+1} 4^{-k} \\ &\ge 4^{-\ell-1}|\varepsilon_{B_1}(s) - \varepsilon_{B_2}(t)| - \frac{1}{2}4^{-\ell-1}.\end{aligned}$$

Since $|\varepsilon_{B_1}(s) - \varepsilon_{B_2}(t)|$ is zero or one, we have

$$|\varphi_\varepsilon(s) - \varphi_\varepsilon(t)| \ge \frac{1}{2}4^{-\ell-1}|\varepsilon_{B_1}(s) - \varepsilon_{B_2}(t)|.$$

Set $N = 2^{\,\mathrm{Card}\mathcal{B}}$. It follows from the preceding inequality that

$$\left(\frac{1}{N}\sum_{\varepsilon\in\mathcal{E}} |\varphi_\varepsilon(s) - \varphi_\varepsilon(t)|^p\right)^{1/p} \ge \frac{1}{2^{1/p}} \cdot \frac{1}{2}4^{-\ell-1} \ge \frac{1}{32}d(s,t).$$

Set then $x_\varepsilon = 32N^{-1/p}\varphi_\varepsilon$, $\varepsilon \in \mathcal{E}$. The family $(x_\varepsilon)_{\varepsilon\in\mathcal{E}}$ is a finite family of elements of $\mathrm{Lip}(T)$ such that $\left(\sum_\varepsilon \|x_\varepsilon\|^p_{\mathrm{Lip}}\right)^{1/p} \le 128$. Further, if $(\theta_\varepsilon)_{\varepsilon\in\mathcal{E}}$ is an independent family of standard p-stable random variables and if $X_t = \sum_{\varepsilon\in\mathcal{E}} \theta_\varepsilon x_\varepsilon(t)$, $t \in T$, then we have just shown that $d_X(s,t) \ge d(s,t)$, for all s,t in T. As announced, this concludes the proof of Theorem 12.17.

We conclude this chapter with some observations and conjectures on Rademacher processes. Recall first that the various results on Gaussian processes described in Section 12.1 can be formulated in the language of subsets of Hilbert space (as, for example, next to Theorem 12.10). For example, if $X = (X_t)_{t\in T}$ is a Gaussian process given by $X_t = \sum_i g_i x_i(t)$, where $(x_i(t))$ is a sequence of functions on T, we may identify T with the subset of ℓ_2 consisting of the elements $(x_i(t))$, $t \in T$. Boundedness of the Gaussian process X is then characterized by the majorizing measure condition $\gamma^{(2)}(T)$ for the Hilbertian distance $|\cdot|$ on $T \subset \ell_2$.

By Rademacher process $X = (X_t)_{t\in T}$ indexed by a set T, we mean that for some sequence $(x_i(t))$ of functions on T, $X_t = \sum_i \varepsilon_i x_i(t)$, assumed to be almost surely convergent (i.e. $\sum_i x_i(t)^2 < \infty$). As in Chapter 4, let $r(T) = \mathbb{E}\sup_{t\in T}|X_t|$. A Rademacher process is subgaussian. If $d_X(s,t) = \|X_s - X_t\|_2$, s,t in T, we thus know that $r(T) < \infty$ whenever there is a probability measure m on T such that $\gamma_m^{(2)}(T, d_X) < \infty$. In order to present our observations, it is convenient (and fruitfull too) to identify as before T with a subset of ℓ_2. By (11.19), we can write that

$$r(T) \le |T| + K\gamma^{(2)}(T)$$

where $|T| = \sup_{t \in T} |t|$, and $\gamma^{(2)}(T)$ is as before the majorizing measure condition on $T \subset \ell_2$ with respect to the canonical metric $|\cdot|$ on ℓ_2 and K some numerical constant.

From this result, one might wonder, as for Gaussian processes, for some possible necessary majorizing measure condition in order for a Rademacher process to be almost surely bounded (or continuous). Denote by B_1 the unit ball of $\ell_1 \subset \ell_2$. Since $r(T) \le \sup_{t \in T} \sum_i |x_i(t)|$, a natural *conjecture* would be that, for some numerical constant K ,

$$T \subset K r(T) B_1 + A$$

where A is a subset of ℓ_2 such that $\gamma^{(2)}(A) \le K r(T)$. This result would completely characterize boundedness of Rademacher process. Theorem 4.15 supports this conjecture. By the convex hull representation of subsets A for which $\gamma^{(2)}(A)$ is controlled (Theorem 12.10), one may look equivalently for a subset A of ℓ_2 such that $A \subset \operatorname{Conv}(y^n)$ where (y^n) is a sequence in ℓ_2 such that, for all n , $|y^n| \le K r(T)(\log(n+1))^{-1/2}$.

Let us call M *-decomposable* a Rademacher process $X = (X_t)_{t \in T}$, or rather T identified with a subset of ℓ_2 , such that, for some M and some A in ℓ_2 ,

$$T \subset M B_1 + A \qquad \text{and} \qquad \gamma^{(2)}(A) \le M . \tag{12.36}$$

In the last part of this chapter, we would like to briefly describe some examples of M-decomposable Rademacher processes which go in the direction of the preceding conjecture.

As a first example, let us consider the case of a subset T of ℓ_2 for which there is a probability measure m such that $\gamma_m^{(q)}(T, \|\cdot\|_{p,\infty}) < \infty$, where $1 \le p < 2$ and q is the conjugate of p . As a consequence of Lemma 11.20 and Remark 12.11, there exists a sequence (z^n) in ℓ_2 such that, if $M = K_p \gamma^{(q)}(T, \|\cdot\|_{p,\infty})$ where K_p only depends on p , $\|z^n\|_{p,\infty} \le M(\log(n+1))^{-1/q}$ ($M \log(1 + \log n)$ if $q = \infty$) and $T \subset \operatorname{Conv}(z^n)$. We show that there exists then a sequence (y^n) in ℓ_2 with $|y^n| \le K'_p M(\log(n+1))^{-1/2}$ for all n , such that,

$$T \subset K'_p M B_1 + \operatorname{Conv}(y^n)$$

where K'_p only depends on p . That is, T is $K'_p M$-decomposable. To check this assertion, let us restrict for simplicity to the case $q < \infty$, $q = \infty$ being similar. For every n ,

$$\|z^n\|_{p,\infty} = \sup_{i \ge 1} i^{1/p} z_i^{n*} \le M(\log(n+1))^{-1/q}$$

where (z_i^{n*}) denotes the non-increasing rearrangement of $(|z_i^n|)$. Let $i_0 = i_0(n)$ be the largest integer $i \geq 1$ such that $i \leq (M/q)^q \log(n+1)$ so that

$$\sum_{i=1}^{i_0} z_i^{n*} \leq M . \tag{12.37}$$

Let then, for each n, y^n be the element of $\mathbb{R}^{\mathbb{N}}$ defined by $y_i^n = z_i^n$ if $z_i^n \notin \{z_1^{n*}, \ldots, z_{i_0}^{n*}\}$, $y_i^n = 0$ otherwise. Clearly

$$|y^n|^2 = \sum_{i>i_0} (z_i^{n*})^2 \leq K_p'' M^2 (\log(n+1))^{-1}$$

and the claims follows by (12.37).

Hence, under a majorizing measure condition of the type $\gamma^{(q)}(T, \|\cdot\|_{p,\infty}) < \infty$, T is decomposable. The next proposition describes a more general result based on Lemma 4.9. It deals with a natural class of bounded Rademacher processes which, according to the conjecture, could possibly describe all Rademacher processes.

Proposition 12.18. Let (a^n) be a sequence in ℓ_2 such that for some M

$$\sum_n \mathbb{P}\{|\sum_i \varepsilon_i a_i^n| > M\} \leq \frac{1}{8} .$$

Let $T \subset \operatorname{Conv}(a^n)$. Then T is KM-decomposable where K is numerical. More precisely, one can find a sequence (y^n) in ℓ_2 such that, for all n, $|y^n| \leq KM(\log(n+1))^{-1/2}$ with the property that $T \subset KMB_1 + \operatorname{Conv}(y^n)$.

Proof. By Lemma 4.2, $(\sum_i (a_i^n)^2)^{1/2} \leq 2\sqrt{2}M$ for all n. Denote by $K_1 \geq 1$ the numerical constant of Lemma 4.9 and set $M_1 = 2\sqrt{2}K_1 M$. By this lemma, there exist, for each n, u^n and v^n in ℓ_2 such that $a^n = u^n + v^n$ and

$$\sum_i |u_i^n| \leq M_1 \quad \text{and} \quad \exp\left(-\frac{K_1 M_1^2}{|v^n|^2}\right) \leq 2\mathbb{P}\{|\sum_i \varepsilon_i a_i^n| > M_1\} .$$

If, for $\varepsilon > 0$, $N(\varepsilon) = \operatorname{Card}\{n\,;\ |v^n| \geq \varepsilon\}$, it follows from the second of these inequalities and the hypothesis that $N(\varepsilon) \geq \frac{1}{4}\exp(K_1 M_1^2/\varepsilon^2)$. We can therefore rearrange (v^n) as a sequence (y^n) such that $|y^n|$ does not increase. In particular, for every N,

$$N \leq \operatorname{Card}\{n\,;\ |v^n| \geq |y^N|\} \leq \frac{1}{4}\exp\left(\frac{K_1 M_1^2}{|y^N|^2}\right) .$$

Hence, for all N, $|y^N| \le K_1 M_1^2/\log(4N)$. The conclusion immediately follows since $\mathrm{Conv}(v^n) = \mathrm{Conv}(y^n)$.

Notes and References

The main results of this chapter are taken from the two papers [Ta5] and [Ta8] where existence of majorizing measures for bounded Gaussian and p-stable processes, $1 \le p < 2$, has been established. This chapter basically compiles, with some omissions, these articles. References on Gaussian processes (in the spirit of what is developed in this chapter) presenting the results at their stage of developments are the article by R. Dudley [Du2], the well-known notes by X. Fernique [Fer4] and the paper by N. C. Jain and M. B. Marcus [J-M3]. A recent account on both continuity and extrema is the set of notes by R. J. Adler [Ad].

The finiteness of Dudley's entropy condition for almost surely bounded stationary Gaussian processes is due to X. Fernique [Fer4]. The result corresponds to Theorems 12.8 and 12.9 in an homogeneous setting and will be detailed in the next chapter. (That the entropy integral is not necessary in general was known since the examples of [Du1].) X. Fernique also established in [Fer5] the a posteriori important case of existence of a majorizing measure in case of an ultrametric index set. He conjectured back in 1974 the validity of the general case. This result has thus been obtained in [Ta5]. Let us mention also a "volumic" approach to regularity of Gaussian processes, actually control of subsets of Hilbert space, in the papers [Du1] and [Mi-P] (see also [Pi18]).

One limitation of Theorem 12.5 is that it can help to understand only processes which are essentially described by a single distance. This limit to the case of Gaussian or p-stable processes. In the recent work [Ta18], this theorem is extended to a case where the single distance is replaced by a family of distances. The ideas of this result, when restricted to the case of one single distance, yield the new proof of Theorem 12.5 presented here, that is different from the original proof of [Ta5], and that is somewhat simpler and more constructive. Another contribution of [Ta18] is a new method to replace, in the proof of Proposition 12.7, the use of Slepian's lemma by the use of Sudakov minoration and of concentration of measure (expressed by Borell inequality in the Gaussian case). The use of these tools allow in [Ta18] to extend Theorem 12.12 (properly reformulated) to a large class of infinitely divisible processes.

Theorem 12.12 comes from [Ta8]. It improves upon the homogoneous case as well as the Sudakov type minorations obtained previously by M. B. Marcus and G. Pisier [M-P2] (cf. Chapter 5). The somewhat more general study for the proof has already been used in empirical processes in [L-T4].

Theorem 12.16 was known for a long time to follow from the majorizing measure conjecture. Its proof is actually rather indirect and it would be desirable to find a direct argument. As indicated by G. Pisier [Pi6] and described by the results of [Ta12], its validity actually implies conversely the existence of majorizing measures. This would yield a new approach to the results of Section 12.1. The type 2 property of the canonical injection map $j: \mathrm{Lip}(T) \to C(T)$ has been put into light by J. Zinn [Zi1] to connect the Jain-Marcus central limit theorem for Lipschitz processes [J-M1] to the general type 2 theory of the CLT (cf. Chapter 14). The majorizing measure version of Zinn's map for $p = 2$ was noticed in [He1] and in [Ju] for $1 \leq p < 2$ (cf. also [M-P2], [M-P3]). Theorem 12.17 was obtained in [Ta5].

Consider $1 \leq \alpha < \infty$, and an independent identically distributed sequence (θ_i), where the law of θ_i has a density $a_\alpha e^{-|x|^\alpha}$ with respect to Lebesgue's measure (a_α is the normalizing factor). Building on the ideas of [Ta18], in [Ta19] Theorem 12.8 is (properly reformulated) extended to the processes $(X_t)_{t \in T}$, where $X_t = \sum\limits_i \theta_i x_i(t)$, and where $(x_i(t))$ is a sequence of functions on T such that $\sum\limits_i x_i^2(t) < \infty$. The still open case of Rademacher processes would correspond to the case " $\alpha = \infty$ ".

Chapter 13. Random Fourier series

Chapter 13. Random Fourier series

In Chapter 11, we evaluated random processes indexed by an arbitrary index set T. We now take advantage of some homogeneity properties of T and investigate in this setting, using the general conclusions of Chapters 11 and 12, the more concrete processes of random Fourier series. The tools developed so far lead indeed to a definitive treatment of those objects with applications to Harmonic Analysis. Our main reference for this chapter is the work by M. B. Marcus and G. Pisier [M-P1], [M-P2] to which we refer for an historical background and accurate references and priorities.

In a first paragraph, we briefly indicate how majorizing measure and entropy conditions coincide in an homogeneous setting. We can therefore deal next only with the simpler minded entropy conditions. Using necessity of majorizing measure (entropy) condition for boundedness and continuity of (stationary) Gaussian processes, we investigate and characterize in this way, in the second section, almost sure boundedness and continuity of large classes of random Fourier series. The case of stable random Fourier series and strongly stationary processes is studied next with the conclusions of Section 12.2. We conclude the chapter with some results and comments on random Fourier series with vector valued coefficients.

Let us note that we usualy deal in this chapter with complex Banach spaces and use, often without further notice, the trivial extensions to the complex case of various results (like e.g. the contraction principle).

13.1. Stationarity and entropy

In this paragraph, we show that on translation invariant metrics, majorizing measure and entropy conditions are the same. We shall adopt (throughout this chapter) the following homogeneous setting. Let G be a locally compact Abelian group with unit element 0. Let $\lambda(\cdot) = |\cdot|$ be the normalized translation invariant (Haar) measure on G. Consider furthermore a metric or pseudo-metric d on G which is translation invariant in the sense that $d(u+s, u+t) = d(s,t)$ for all u, s, t in G. Let finally T be a compact subset of nonempty interior of G.

Recall that $N(T, d; \varepsilon)$ denotes the minimal number of open balls (with centers in T) of radius $\varepsilon > 0$ in the pseudo-metric d necessary to cover T. More generally, for subsets A, B of G, denote by $N(A, B)$ the minimal number of translates of B (by elements of A) necessary to cover A, i.e.

$$N(A,B) = \inf\{N \geq 1\,;\ \exists t_1, \ldots, t_N \in A\,,\ A \subset \bigcup_{i=1}^{N} (t_i + B)\}\,.$$

Here $t + B = \{t + s\,;\ s \in B\}$. We let similarly for subsets A, B of G, $A + B = \{s + t\,;\ s \in A\,,\ t \in B\}$ and define in the same way $A - B$, etc. In particular, we set $T' = T + T$ and $T'' = T + T' = T + T + T$.

The following lemma is an elementary statement on the preceding covering numbers which will be useful in various parts of this chapter. We set $B(0, \varepsilon) = \{t \in G\,;\ d(t, 0) < \varepsilon\}$.

Lemma 13.1. Under the previous notations, we have:

(i) if $B = T' \cap B(0, \varepsilon)$, $\varepsilon > 0$, then $N(T, d; \varepsilon) = N(T, B)$;

(ii) $N(T, d; 2\varepsilon) \leq N(T, B(0, \varepsilon) - B(0, \varepsilon))$;

(iii) if $A \subset G$, $N(T, A) \geq |T|/|A|$;

(iv) if $A \subset T'$, $N(T, A - A) \leq |T''|/|A|$.

Proof. (i) follows immediately from the definitions. (ii) If, for t in T, there exists t_i in T such that $t \in t_i + B(0, \varepsilon) - B(0, \varepsilon)$, this means that $t = t_i + u - v$ where $u, v \in B(0, \varepsilon)$. Hence, by translation invariance,

$$d(t, t_i) = d(u - v, 0) \leq d(u, 0) + d(0, v) < 2\varepsilon$$

so that $N(T, d; 2\varepsilon) \leq N(T, B(0, \varepsilon) - B(0, \varepsilon))$. (iii) It is enough to consider the case $N(T, A) < \infty$. Then, if $T \subset \bigcup_{i=1}^{N} (t_i + A)$,

$$|T| \leq \sum_{i=1}^{N} |t_i + A| = N|A|$$

which gives the result. (iv) Assume $|A| > 0$. Let $\{t_1, \ldots, t_M\}$ be maximal in T under the conditions $(t_i + A) \cap (t_j + A) = \emptyset$, $\forall i \neq j$. If $t \in T$, by maximality, $(t + A) \cap (t_i + A) \neq \emptyset$ for some $i = 1, \ldots, M$. Hence, $t + u = t_i + v$ for some u, v in A. Therefore $t \in t_i + A - A$ and $T \subset \bigcup_{i=1}^{M} (t_i + A - A)$. This implies that $M \geq N(T, A - A)$. Now $(t_i + A)_{i \leq M}$ are disjoint in $T'' = T + T'$ and thus

$$|T''| \geq \sum_{i=1}^{M} |t_i + A| = M|A| \geq N(T, A - A)|A|.$$

The proof of Lemma 13.1 is complete.

Provided with this lemma we can compare majorizing measure and entropy conditions in translation invariant situations. The idea is simply that if a majorizing measure exists then Haar measure is also a majorizing measure from which the conclusion then follows from the preceding lemma.

Proposition 13.2. Let ψ be a Young function. In the preceding notations, let m be a probability measure on $T \subset G$ and denote by $D = D(T)$ the d-diameter of T. Then

$$\frac{1}{2}\int_0^D \psi^{-1}\left(\frac{|T|}{|T''|}N(T,d;\varepsilon)\right)d\varepsilon \le \sup_{t\in T}\int_0^D \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\varepsilon\,.$$

Proof. Let us denote by M the right hand side of the inequality that we have to establish. Since $\psi^{-1}(1/x)$ is convex, by Jensen's inequality,

$$\begin{aligned} M &\ge \int_0^D \frac{1}{|T|}\int_T \psi^{-1}\left(\frac{1}{m(B(t,\varepsilon))}\right)d\lambda(t)d\varepsilon \\ &\ge \int_0^D \psi^{-1}\left(\frac{|T|}{\int_T m(B(t,\varepsilon))d\lambda(t)}\right)d\varepsilon\,. \end{aligned}$$

Now, by Fubini's theorem and translation invariance,

$$\int_T m(B(t,\varepsilon))d\lambda(t) = \int_T |T\cap B(s,\varepsilon)|dm(s) \le |T'\cap B(0,\varepsilon)|\,.$$

Hence

$$M \ge \int_0^D \psi^{-1}\left(\frac{|T|}{|T'\cap B(0,\varepsilon)|}\right)d\varepsilon\,.$$

To conclude, by (ii) and (iv) of Lemma 13.1, for every $\varepsilon > 0$,

$$N(T,d;2\varepsilon) \le \frac{|T''|}{|T'\cap B(0,\varepsilon)|}$$

from which Proposition 13.2 follows.

Note that, conversely, if $\lambda_{T''}$ denotes restricted normalized Haar measure on $T'' = T+T+T$, then, for any $\eta > 0$,

$$\sup_{t\in T}\int_0^\eta \psi^{-1}\left(\frac{1}{\lambda_{T''}(B(t,\varepsilon))}\right)d\varepsilon \le \int_0^\eta \psi^{-1}\left(\frac{|T''|}{|T|}N(T,d;\varepsilon)\right)d\varepsilon\,. \tag{13.1}$$

(Note the complete equivalence when $T = G$ is compact.) To show (13.1), observe that, if $t \in T$, $t + T' \cap B(0,\varepsilon) \subset T'' \cap B(t,\varepsilon)$. Hence, by translation invariance and (i) and (iii) of Lemma 13.1,

$$\lambda_{T''}(B(t,\varepsilon)) \ge \frac{1}{|T''|}\lambda(T'\cap B(0,\varepsilon)) \ge \frac{|T|}{|T''|}N(T,d;\varepsilon)\,.$$

Proposition 13.2 allows to state in terms of entropy the characterization of almost sure boundedness and continuity of stationary Gaussian processes. Before stating the result, it is convenient to introduce some notations concerning entropy integrals. Let (T,d) be a pseudo-metric space with diameter D. For $1 \le q < \infty$, we let

$$E^{(q)}(T,d) = \int_0^\infty (\log N(T,d;\varepsilon))^{1/q} d\varepsilon .$$

As usual, the integral is taken from 0 to ∞ but of course stops at D. When $q = \infty$, we let

$$E^{(\infty)}(T,d) = \int_0^\infty \log(1 + \log N(T,d;\varepsilon)) d\varepsilon .$$

A process $X = (X_t)_{t \in G}$ indexed by G is stationary if, for every u in G, $(X_{u+t})_{t\in G}$ has the same distribution as X. If X is a stationary Gaussian process, its corresponding L_2-metric $d_X(s,t) = \|X_s - X_t\|_2$ is translation invariant. As a corollary to the majorizing measure characterization of boundedness and continuity of Gaussian processes described in Chapter 12, and Proposition 13.2, we can state:

Theorem 13.3. Let G be a locally compact Abelian group and let T be a compact metrizable subset of G of nonempty interior. Let $X = (X_t)_{t\in G}$ be a stationary Gaussian process indexed by G. Then X has a version with almost all bounded and continuous sample paths on $T \subset G$ if and only if d_X is continuous on $G \times G$ and $E^{(2)}(T, d_X) < \infty$. Moreover, there is a numerical constant K such that

$$K^{-1}(E^{(2)}(T,d_X) - L(T)^{1/2} D(T)) \le \mathbb{E} \sup_{t\in T} X_t \le K E^{(2)}(T, d_X)$$

where $D(T)$ is the diameter of (T, d_X) and $L(T) = \log(|T''|/|T|)$. Note that if $T = G$ is compact, $L(T) = 0$.

Sufficiency in this theorem is simply Dudley's majoration (Theorem 11.17), together with the continuity of d_X. Necessity and the left side inequality follow from Theorem 12.9 together with Proposition 13.2. (It might be useful to recall at this point the simple comparison (11.17).)

Note the following dichotomy contained in the statement of Theorem 13.3: stationary Gaussian processes are either almost surely continuous or almost surely unbounded (according to the finiteness or not of the entropy integral).

The choice of the compact set T does not affect the qualitative conclusion of Theorem 13.3. Indeed, let actually $(X_t)_{t\in G}$ be any stationary process. If T_1 and T_2 are two compact subsets of G with nonempty interiors, then $(X_t)_{t\in T_1}$ has a version with continuous sample paths if and only if $(X_t)_{t\in T_2}$ does. This is obvious by stationarity since each of the sets T_1 and T_2 can be covered by finitely many translates of the other. Consequently, if G is the union of a countable family of compact sets, then $(X_t)_{t\in T}$ has a version with continuous sample paths if and only if the entire process $(X_t)_{t\in G}$ does. This applies in the most important cases such as $G = \mathbb{R}^n$. Boundedness (lattice-boundedness) of the stationary Gaussian process of Theorem 13.3 over *all* G holds if and only if

$$\int_0^\infty (\log N(G, d_X; \varepsilon))^{1/2} d\varepsilon < \infty .$$

This can be shown by either repeating the arguments of the proof of Theorem 13.3 or by using the Bohr compactification of G and the fact that under the preceding entropy condition, X has a version which is an almost periodic function on G. The proof of this fact indicate further that, under this condition, a stationary Gaussian process $X = (X_t)_{t\in G}$ indexed by G admits an expansion as a series

$$X_t = \sum_n g_n \operatorname{Re}(a_n\gamma_n(t)) + \sum_n g_n' \operatorname{Im}(a_n\gamma_n(t)) , \quad t \in G ,$$

where (a_n) is a sequence of complex numbers in ℓ_2, (γ_n) a sequence of continuous characters of G and (g_n), (g_n') independent standard Gaussian sequences. We refer to [M-P1, p. 134-138] for more details on these comments.

Theorem 13.3 is one of the main ingredients in the study of random Fourier series. Before turning to this topic in the subsequent section, it is convenient for comparison and possible estimates in concrete situations to mention an equivalent of the entropy integral $E^{(q)}(T,d)$ for a translation invariant metric d. Assume for simplicity that T is a symmetric subset of G and let $\sigma(t) = d(t,0)$, $t \in T'$. Consider the non-decreasing rearrangement $\overline{\sigma}$ of σ on $[0, |T'|]$ (with respect to T'); more precisely, for $0 < \delta \le |T'|$,

$$\overline{\sigma}(\delta) = \sup\{\varepsilon > 0 ;\ |T' \cap B(0,\varepsilon)| < \delta\} .$$

For $1 < p \le 2$, let

$$I^{(p)}(T,d) = \int_0^{|T'|} \frac{\overline{\sigma}(\varepsilon)}{\varepsilon\left(\log \frac{4|T'''|}{\varepsilon}\right)^{1/p}} d\varepsilon$$

where $T''' = T + T + T + T$. By Lemma 13.1 and elementary arguments it can be shown that, if q is the conjugate of p, $E^{(q)}(T,d)$ and $I^{(p)}(T,d)$ are essentially of the same order. Namely, for some constant $K_p > 0$ only depending on p,

$$K_p^{-1}(D + I^{(p)}(T,d)) \leq E^{(q)}(T,d) \leq K_p(D + I^{(p)}(T,d))$$

where D is the diameter of (T,d) . A similar result holds for $p = 1$ (and $q = \infty$).

13.2. Random Fourier series

In this paragraph, we take advantage of Theorem 13.3 to study a class of random Fourier series and develop some applications. The interested reader will find in the book [M-P1] a general and historical introduction to random Fourier series. We more or less only concentrate here on one typical situation. In particular, we only consider Abelian groups. We refer to [M-P1] for the non-Abelian case as well as for further aspects.

Throughout this section, we fix the following notations. G is a locally compact Abelian group with identity element 0 and Haar measure $\lambda(\cdot) = |\cdot|$. We denote by V a compact neighborhood of 0 (by translation invariance, the results would apply to all compact sets with nonempty interior). Let Γ be the dual group of all characters γ of G (i.e. γ is a continuous complex valued function on G such that $|\gamma(t)| = 1$, $\gamma(s)\gamma(t) = \gamma(s+t)$ for all s, t in G). Fix a countable subset A of Γ . Since the main interest will be the study of random Fourier series with spectrum in A , and since A generates a closed separable subgroup of Γ , we can assume without restricting the generality that Γ itself is separable. In particular, all the compact subsets of G are *metrizable*. As a concrete example of this setting, one may consider the compact group $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ (identified with $[0, 2\pi]$) with $\mathbb{Z}$ as dual group.

$V \subset G$ being as before, we agree to denote by $\|\cdot\| = \sup_{t \in V} |\cdot|$ the sup -norm on the Banach space $C(V)$ of all continuous *complex* valued functions on V (or on G when $V = G$).

Let $(a_\gamma)_{\gamma \in A}$ be complex numbers such that $\sum_{\gamma \in A} |a_\gamma|^2 < \infty$. (We sometimes omit $\gamma \in A$ in the summation symbol only indicated then by $\sum_\gamma$ or just $\sum$.) Let also $(g_\gamma)_{\gamma \in A}$ be a standard Gaussian sequence. Following the comments at the end of Theorem 13.3, let us first consider the Gaussian process $X = (X_t)_{t \in G}$ given by

$$X_t = \sum_{\gamma \in A} a_\gamma g_\gamma \gamma(t) , \qquad t \in G . \tag{13.2}$$

The question of course arises of the almost sure continuity or uniform convergence of the series X on V. Since (g_γ) is a symmetric sequence, these two properties are equivalent: if X is continuous (or admits a continuous version) on V, then by Itô-Nisio's theorem (Theorem 2.1), X converges uniformly (for any ordering of A). The process defined by (13.2) is a (complex valued) Gaussian process indexed by G with associated L_2 metric

$$d_X(s,t) = \left(\sum_{\gamma \in A} |a_\gamma|^2 |\gamma(s) - \gamma(t)|^2 \right)^{1/2} , \; s,t \in G ,$$

which is translation invariant. Since X is complex valued, to enter the setting of Theorem 13.3, we use the following. If we let

$$X'_t = \sum_{\gamma \in A} g_\gamma \operatorname{Re}(a_\gamma \gamma(t)) + \sum_{\gamma \in A} g'_\gamma \operatorname{Im}(a_\gamma \gamma(t)) , \; t \in G ,$$

where (g'_γ) is an independent copy of (g_γ), $X' = (X'_t)_{t \in G}$ is a real stationary Gaussian process such that $d_{X'} = d_X$. The series X and X' converge uniformly almost surely simultaneously. As a consequence of Theorem 13.3, X' admits a version with almost all sample paths continuous on V and therefore converges uniformly almost surely on V if and only if $E^{(2)}(V, d_{X'}) = E^{(2)}(V, d_X) < \infty$, and thus the same holds for X. That is, we have the following statement.

Theorem 13.4. Let X be the random Fourier series of (13.2). The following are equivalent:

(i) for some (all) ordering of A, the series X converges uniformly almost surely on V;

(ii) $\sup\limits_{\substack{F \subset A \\ F \text{ finite}}} \mathbb{E} \| \sum_{\gamma \in F} a_\gamma g_\gamma \gamma \|^2 < \infty$;

(iii) $E^{(2)}(V, d_X) < \infty$.

Further, for some numerical constant K,

$$\begin{aligned} & K^{-1} \left(E^{(2)}(V, d_X) - L(V)^{1/2} \left(\sum_{\gamma \in A} |a_\gamma|^2 \right)^{1/2} \right) \\ (13.3) \qquad & \leq (\mathbb{E}\|X\|^2)^{1/2} \leq K \left(E^{(2)}(V, d_X) + \left(\sum_{\gamma \in A} |a_\gamma|^2 \right)^{1/2} \right) \end{aligned}$$

where we recall that $L(V) = \log(|V''|/|V|)$ where $V'' = V + V + V$ and $\|X\| = \sup\limits_{t \in V} |X_t|$.

Proof. If X converges uniformly in some ordering, then (ii) is satisfied by the integrability properties of Gaussian random vectors (Corollary 3.2) and conditional expectation. Let F be finite in A and denote by X^F the finite sum $\sum_{\gamma \in F} a_\gamma g_\gamma \gamma$. Then, as a consequence of Theorem 13.3 (considering as indicated before the natural associated real stationary Gaussian series), we have that (13.3) holds for X^F, with thus a numerical constant K independent of F finite in A. Then (iii) holds under (ii) by increasing F to A. Finally, and in the same way, if $E^{(2)}(V, d_X) < \infty$, for any ordering of A, X converges in L_2 with respect to the uniform norm on $C(V)$, and thus almost surely by Itô-Nisio's theorem: to see this, simply use a Cauchy argument in inequality (13.3) and dominated convergence in the entropy integral.

Note that we recover from this statement the equivalence between almost surely boundedness and continuity of Gaussian random Fourier series.

As the main question of this study, we shall be interested in similar results and estimates when the Gaussian sequence $(g_\gamma)_{\gamma \in A}$ is replaced by a Rademacher sequence $(\varepsilon_\gamma)_{\gamma \in A}$ or, more generally, by some *symmetric* sequence $(\xi_\gamma)_{\gamma \in A}$ of real random variables. While we are dealing with Fourier series with complex coefficients, for simplicity however we only consider real probabilistic structures. The complex case can actually easily be deduced. Recall that, by the symmetry assumption, the sequence (ξ_γ) can be replaced by $(\varepsilon_\gamma \xi_\gamma)$ where (ε_γ) is an independent Rademacher sequence. This is the setting we will adopt. By standard symmetrization procedure, the results apply similarly to a sequence (ξ_γ) of independent mean zero random variables (cf. Lemma 6.3).

Assume therefore we are given a sequence of complex numbers $(a_\gamma)_{\gamma \in A}$ and a sequence $(\xi_\gamma)_{\gamma \in A}$ of real random variables satisfying

$$\sum_{\gamma \in A} |a_\gamma|^2 \mathbb{E}|\xi_\gamma|^2 < \infty .$$

Consider the process $Y = (Y_t)_{t \in G}$ defined by

$$Y_t = \sum_{\gamma \in A} a_\gamma \varepsilon_\gamma \xi_\gamma \gamma(t) , \quad t \in G , \tag{13.4}$$

where, as before, (ε_γ) is a Rademacher sequence independent of (ξ_γ). We associate to this process the L_2 pseudo-metric

$$d_Y(s,t) = \left(\sum_{\gamma \in A} |a_\gamma|^2 \mathbb{E}|\xi_\gamma|^2 |\gamma(s) - \gamma(t)|^2 \right)^{1/2} , \quad s, t \in G ,$$

which is translation invariant. If V is as usual a compact neighborhood of 0, we shall be interested, as previously in the Gaussian case, in the almost sure uniform convergence on V of the random Fourier series Y of (13.4). The main objective is to try to obtain bounds similar to the ones of the Gaussian random Fourier series ((13.2)). The basic idea consists in first writing the best possible entropy estimates conditionally on the sequence (ξ_γ) (that is, along the Rademacher sequence) and then integrating and making use of the translation invariant properties. The first step is simply the content of the following lemma which is the entropic (and complex) version of (11.19) and Lemma 11.20.

Lemma 13.5. Let $x_i(t)$ be (complex) functions on a set T such that $\sum_i |x_i(t)|^2 < \infty$ for every t in T. Let $1 \le p \le 2$ and equip T with the pseudo-metric (functional) d_p defined by

$$d_2(s,t) = \left(\sum_i |x_i(s) - x_i(t)|^2 \right)^{1/2}$$

if $p = 2$ and

$$d_p(s,t) = \|(x_i(s) - x_i(t))\|_{p,\infty}$$

if $1 \le p < 2$. Then, for some K_p depending only on p,

$$\left(\mathbb{E} \sup_{t \in T} \left| \sum_i \varepsilon_i x_i(t) \right|^2 \right)^{1/2} \le K_p (D + E^{(q)}(T, d_p))$$

where $D = \sup_{t \in T} \left(\sum_i |x_i(t)|^2 \right)^{1/2}$ or $\sup_{t \in T} \|(x_i(t))\|_{p,\infty}$ according as $p = 2$ or $p < 2$ and where $q = p/p-1$ is the conjugate of p. Further, if $E^{(q)}(T, d_p) < \infty$, $\left(\sum_i \varepsilon_i x_i(t) \right)_{t \in T}$ has a version with continuous paths on (T, d_p).

As announced, this lemma applied conditionally together with the translation invariance property allows to evaluate random Fourier series of the type (13.4). The main results in this direction is the following theorem. Recall that for $V \subset G$ we set $L(V) = \log(|V''|/|V|)$ where $V'' = V + V + V$.

Theorem 13.6. If $E^{(2)}(V, d_Y) < \infty$, the random Fourier series Y of (13.4) converges uniformly on V with probability one. Further, for some numerical constant K,

$$(\mathbb{E}\|Y\|^2)^{1/2} \le K \left((1 + L(V)^{1/2}) \left(\sum_{\gamma \in A} |a_\gamma|^2 \mathbb{E}|\xi_\gamma|^2 \right)^{1/2} + E^{(2)}(V, d_Y) \right).$$

Proof. There is no loss of generality to assume by homogeneity that (for example)

$$\left(\sum_\gamma |a_\gamma|^2 \mathbb{E}|\xi_\gamma|^2\right)^{1/2} \leq \frac{1}{8}\,.$$

Denote by Ω_0 the set, of probability one, of all ω 's for which $\sum_\gamma |a_\gamma|^2 |\xi_\gamma(\omega)|^2 < \infty$. For ω in Ω_0, introduce the translation invariant pseudo-metric

$$d_\omega(s,t) = \left(\sum_\gamma |a_\gamma|^2 |\xi_\gamma(\omega)|^2 |\gamma(s) - \gamma(t)|^2\right)^{1/2}, \quad s,t \in G\,.$$

Clearly, for every s,t, $\mathbb{E} d_\omega^2(s,t) = d_Y^2(s,t)$ and

$$d_\omega(s,t) \leq D(\omega) = 2\left(\sum_\gamma |a_\gamma|^2 |\xi_\gamma(\omega)|^2\right)^{1/2} < \infty\,.$$

We may as well assume that $D(\omega) > 0$ for all ω in Ω_0. Conditionally on the set Ω_0, which only depends on the sequence (ξ_γ), we now apply Lemma 13.5 (for $p = 2$). We get that, $\mathbb{E}_\varepsilon$ denoting as usual partial integration with respect to the Rademacher sequence (ε_γ),

$$\begin{aligned} &\left(\mathbb{E}_\varepsilon \sup_{t \in V} \Big|\sum_\gamma a_\gamma \varepsilon_\gamma \xi_\gamma(\omega)\gamma(t)\Big|^2\right)^{1/2} \\ &\leq K\left(D(\omega) + \int_0^{D(\omega)} (\log N(V, d_\omega; \varepsilon))^{1/2} d\varepsilon\right). \end{aligned} \tag{13.5}$$

The idea is now to integrate this inequality with respect to the sequence (ξ_γ) and use the translation invariance properties to conclude. To this aim, let us set, for every integer $n \geq 1$,

$$W_n = \{t \in V'\,;\ d_Y(t,0) \leq 2^{-n}\}$$

where $V' = V + V$. For all ω in Ω_0, define a sequence $(b_n(\omega))_{n \geq 0}$ of positive numbers by letting $b_0(\omega) = D(\omega)$ and, for $n \geq 1$,

$$b_n(\omega) = \max\left(2^{-n},\ 4\int_{W_n} d_\omega(t,0)\frac{dt}{|W_n|}\right) \quad (2^{-n} \text{ if } |W_n| = 0)\,.$$

Observe that, for every $n \geq 1$,

$$(\mathbb{E} b_n^2)^{1/2} \leq 2^{-n} + 4 \left(\int_{W_n} \mathbb{E} d_\omega^2(t,0) \frac{dt}{|W_n|} \right)^{1/2}$$
$$\leq 2^{-n} + 4 \left(\int_{W_n} d_Y^2(t,0) \frac{dt}{|W_n|} \right)^{1/2} \leq 5 \cdot 2^{-n} ,$$

so that in particular $b_n(\cdot) \to 0$ almost surely. Let us denote by Ω_1 the set of full probability consisting of Ω_0 and of the set on which the sequence (b_n) converges to 0. For ω in Ω_1, and $n \geq 1$, set

$$B(n,\omega) = \{ t \in W_n \, ; \; d_\omega(t,0) \leq b_n(\omega)/2 \} .$$

Clearly, by Fubini's theorem and definition of $b_n(\omega)$,

$$|B(n,\omega)| \geq \frac{|W_n|}{2} . \tag{13.6}$$

We are now ready to integrate in ω inequality (13.5). To this aim, we make use of the following elementary lemma.

Lemma 13.7. Let $f : (0,D] \to \mathbb{R}_+$ be decreasing. Let also (b_n) be a sequence of positive numbers with $b_0 = D$ and $b_n \to 0$. Then

$$\int_0^D f(x)\, dx \leq \sum_{n=0}^\infty b_n f(b_{n+1}) .$$

By this lemma, for ω in Ω_1, we have

$$\int_0^{D(\omega)} (\log N(V, d_\omega; \varepsilon))^{1/2} d\varepsilon \leq \sum_{n=0}^\infty b_n(\omega) (\log N(V, d_\omega; b_{n+1}(\omega)))^{1/2} .$$

By Lemma 13.1 and (13.6), for ω in Ω_1 and $n \geq 0$, we can write

$$\begin{aligned} N(V, d_\omega; b_{n+1}(\omega)) &\leq N(V, B(n+1,\omega) - B(n+1,\omega)) \\ &\leq \frac{|V''|}{|B(n+1,\omega)|} \\ &\leq 2 \frac{|V''|}{|W_{n+1}|} \\ &\leq 2 \frac{|V''|}{|V|} N(V, W_{n+1}) \\ &\leq 2 \frac{|V''|}{|V|} N(V, d_Y; 2^{-n-1}) . \end{aligned}$$

Hence (13.5) reads as

$$\left(\mathbb{E}_\varepsilon \sup_{t\in V}\left|\sum_\gamma a_\gamma\varepsilon_\gamma\xi_\gamma(\omega)\gamma(t)\right|^2\right)^{1/2}$$
$$\leq K\left(D(\omega)+\sum_{n=0}^{\infty} b_n(\omega)\left(\log\left(2\frac{|V''|}{|V|}N(V,d_Y;2^{-n-1})\right)\right)^{1/2}\right)$$

for every ω in Ω_1. Since $(\mathbb{E}b_n^2)^{1/2}\leq 5\cdot 2^{-n}$, we have simply to integrate with respect to ω and we then get, by the triangle inequality,

$$(\mathbb{E}\sup_{t\in V}|Y_t|^2)^{1/2}\leq K\left(\frac{1}{4}+5\sum_{n=0}^{\infty}2^{-n}\left(\log\left(2\frac{|V''|}{|V|}N(V,d_Y;2^{-n-1})\right)\right)^{1/2}\right)$$

which yields the inequality of the theorem. A Cauchy argument together with dominated convergence in the entropy integral and Itô-Nisio's theorem proves then the almost sure uniform convergence of Y on V. The proof of Theorem 13.6 is complete.

Theorem 13.6 expresses a bound for the random Fourier series Y very similar to the one described previously for Gaussian Fourier series. Actually, the comparison between Theorem 13.6 and the lower bound in (13.3) indicate that, for some numerical constant K,

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\xi_\gamma\gamma\|^2\right)^{1/2}\leq K((1+L(V)^{1/2})\left(\mathbb{E}\|\sum_\gamma a_\gamma(\mathbb{E}|\xi_\gamma|^2)^{1/2}g_\gamma\gamma\|^2\right)^{1/2}. \tag{13.7}$$

(What we get actually is that

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\xi_\gamma\gamma\|^2\right)^{1/2}\leq K\left((1+L(V)^{1/2})\left(\sum_\gamma|a_\gamma|^2\mathbb{E}|\xi_\gamma|^2\right)^{1/2}\right.$$
$$\left.+\left(\mathbb{E}\|\sum_\gamma a_\gamma(\mathbb{E}|\xi_\gamma|^2)^{1/2}g_\gamma\gamma\|^2\right)^{1/2}\right)$$

but we will not use this improved formulation in the sequel. Recall that $L(V)\geq 0$.) In particular, we see by the contraction principle that

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\xi_\gamma\gamma\|^2\right)^{1/2}\leq K((1+L(V)^{1/2})\sup_\gamma(\mathbb{E}|\xi_\gamma|^2)^{1/2}\left(\mathbb{E}\|\sum_\gamma a_\gamma g_\gamma\gamma\|^2\right)^{1/2}. \tag{13.8}$$

One might of course wonder at this stage when an inequality like (13.8) can be reversed. By the contraction principle in the form of Lemma 4.5,

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\xi_\gamma\gamma\|^2\right)^{1/2} \geq \inf_\gamma \mathbb{E}|\xi_\gamma| \left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\gamma\|^2\right)^{1/2} .$$

If we try to exploit this information in order to reverse inequality (13.8), we are faced with the question of knowing whether the Rademacher series $\sum_\gamma a_\gamma\varepsilon_\gamma\gamma$ dominates in the sup-norm the coresponding Gaussian series $\sum_\gamma a_\gamma g_\gamma\gamma$. We know ((4.8)) that the converse is always satisfied, i.e.

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\gamma\|^2\right)^{1/2} \leq K\left(\mathbb{E}\|\sum_\gamma a_\gamma g_\gamma\gamma\|^2\right)^{1/2} \tag{13.9}$$

(with $K = (\pi/2)^{1/2}$) but the other way does not hold in general, unless we deal with a Banach space of finite cotype (cf. (4.9) and Proposition 9.14). Although $C(V)$ is of no finite cotype, the more general estimates (13.7) or (13.8) that we have obtained actually show that the inequality we are looking for is satisfied here due to the particular structure of the vector valued coefficients $a_\gamma\gamma$. The proof of this result is similar to the argument used in the proof of Proposition 9.25 showing that the Rademacher and Gaussian cotype 2 definitions are equivalent. Further, this property is indeed related to some cotype 2 Banach space of remarkable interest which we discuss next. The following proposition is the announced result of the equivalence of Gaussian and Rademacher random Fourier series.

Proposition 13.8. For some numerical constant K,

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma g_\gamma\gamma\|^2\right)^{1/2} \leq K((1+L(V)^{1/2})\left(\mathbb{E}\|\sum_\gamma a_\gamma\varepsilon_\gamma\gamma\|^2\right)^{1/2} .$$

(Recall that by (13.9) the reserve inequality is satisfied with a numerical constant independent of V.) Further, $\sum_\gamma a_\gamma g_\gamma\gamma$ and $\sum_\gamma a_\gamma\varepsilon_\gamma\gamma$ converge uniformly almost surely simultaneously.

Proof. The second assertion follows from the inequalities and a Cauchy argument and the integrability properties of both Gaussian and Rademacher series. We can thus assume we deal with finite sums. Let $c > 0$ to be specified. By the triangle inequality,

$$\left(\mathbb{E}\|\sum_\gamma a_\gamma g_\gamma\gamma\|^2\right)^{1/2} \leq \left(\mathbb{E}\|\sum_\gamma a_\gamma g_\gamma I_{\{|g_\gamma|\leq c\}}\gamma\|^2\right)^{1/2} + \left(\mathbb{E}\|\sum_\gamma a_\gamma g_\gamma I_{\{|g_\gamma|>c\}}\gamma\|^2\right)^{1/2} .$$

By the contraction principle, the first term on the right of this inequality is smaller than

$$c\left(\mathbb{E}\|\sum_{\gamma} a_\gamma \varepsilon_\gamma \gamma\|^2\right)^{1/2} .$$

To the second, we apply (13.7) to see that it bounded by

$$K(1+L(V)^{1/2})(\mathbb{E}|g|^2 I_{\{|g|>c\}})^{1/2}\left(\mathbb{E}\|\sum_{\gamma} a_\gamma g_\gamma \gamma\|^2\right)^{1/2}$$

where g is a standard normal variable. Let us then choose $c>0$ in order that

$$K(1+L(V)^{1/2})(\mathbb{E}|g|^2 I_{\{|g|>c\}})^{1/2} \leq \frac{1}{2} .$$

This can be achieved with c of the order of $1+L(V)^{1/2}$ from which the proposition then follows. (Actually a smaller function of $L(V)$ would suffice but we need not be concerned with this here.)

As a corollary of Theorem 13.7 and Proposition 13.8, we can now summarize the results on random Fourier series $Y=(Y_t)_{t\in G}$ of the type

$$Y_t = \sum_{\gamma\in A} a_\gamma \xi_\gamma \gamma(t) , \qquad t \in G ,$$

where $\sum_\gamma |a_\gamma|^2 < \infty$ and (ξ_γ) is a symmetric sequence of real random variables such that $\sup_\gamma \mathbb{E}|\xi_\gamma|^2 < \infty$ and $\inf_\gamma \mathbb{E}|\xi_\gamma| > 0$. Recall that by the symmetry assumption, (ξ_γ) has the same distribution as $(\varepsilon_\gamma \xi_\gamma)$ where (ε_γ) is an independent Rademacher sequence.

Corollary 13.9. Let V be as usual a compact symmetric neighborhood of the unit of G. Let $Y=(Y_t)_{t\in G}$ be as just defined with associated metric

$$d(s,t) = \left(\sum_\gamma |a_\gamma|^2 |\gamma(s)-\gamma(t)|^2\right)^{1/2} , \quad s,t\in G .$$

Then Y converges uniformly on V almost surely if and only if the entropy integral $E^{(2)}(V,d)$ is finite. Furthermore, for some numerical constant K,

$$\begin{aligned}
&(K(1+L(V)^{1/2}))^{-1}\inf_\gamma \mathbb{E}|\xi_\gamma| \left(\left(\sum_\gamma |a_\gamma|^2\right)^{1/2} + E^{(2)}(V,d)\right) \\
&\quad \leq (\mathbb{E}\|Y\|^2)^{1/2} \leq K(1+L(V)^{1/2})\sup_\gamma (\mathbb{E}|\xi_\gamma|^2)^{1/2}\left(\left(\sum_\gamma |a_\gamma|^2\right)^{1/2} + E^{(2)}(V,d)\right) .
\end{aligned}$$

Note in particular that Y converges uniformly almost surely if and only if the associated Gaussian Fourier series $\sum_\gamma a_\gamma g_\gamma \gamma$ does (a special case of which discussed previously being of course constitued by the choice for (ξ_γ) of a Rademacher sequence). Let us mention further that several of the comments following Theorem 13.3 apply similarly in the context of Corollary 13.9 and that the equivalences of Theorem 13.4 hold in the same way. In particular, boundedness and continuity are equivalent.

The following is a consequence of Corollary 13.9.

Corollary 13.10. Let Y be as in Corollary 13.9 with $\mathbb{E}|\xi_\gamma|^2 = 1$ for all γ in A. Then $(Y_t)_{t\in V}$ satisfies the central limit theorem in $C(V)$ if and only if $E^{(2)}(V,d) < \infty$.

Proof. $(\sum_\gamma a_\gamma g_\gamma \gamma(t))_{t\in V}$ is a Gaussian process with the same covariance structure as $(Y_t)_{t\in V}$. If $(Y_t)_{t\in V}$ satisfies the CLT it is necessarily pregaussian so that $E^{(2)}(V,d) < \infty$ by (13.3). To prove sufficiency, consider independent copies of $(Y_t)_{t\in V}$ associated to independent copies (ξ_γ^i) of the sequence (ξ_γ). Since, for all γ,

$$\frac{1}{\sqrt{n}}\left(\mathbb{E}|\sum_{i=1}^n \xi_\gamma^i|^2\right)^{1/2} = (\mathbb{E}|\xi_\gamma|^2)^{1/2} = 1\,,$$

the right hand side of the inequality of Corollary 13.9 together with an approximation argument shows that $(Y_t)_{t\in V}$ satisfies the CLT in $C(V)$ (cf. (10.4)).

Turning back to Proposition 13.8 and its proof, let us now explicit a remarkable Banach algebra of cotype 2 whose cotype 2 property actually basically amounts to inequality (13.8) and the conclusion of Corollary 13.9. Let us assume here for simplicity that G is (Abelian and) *compact.* Denote by Γ its discrete dual group. Introduce then the space $C_{\text{a.s.}} = C_{\text{a.s.}}(G)$ of all sequences of complex numbers $a = (a_\gamma)_{\gamma\in\Gamma}$ in ℓ_2 such that the Gaussian Fourier series $\sum_{\gamma\in\Gamma} a_\gamma g_\gamma \gamma(t)$, $t \in G$, converges uniformly on G almost surely. By what was described so far, we know that we get the same definition when the orthogaussian sequence (g_γ) is replaced by a Rademacher sequence (ε_γ), or even some more general symmetric sequence which enters the framework of Corollary 13.9. Alternatively, $C_{\text{a.s.}}$ is characterized by $E^{(2)}(G,d) < \infty$ where $d(s,t) = \left(\sum_\gamma |a_\gamma|^2 |\gamma(s) - \gamma(t)|^2\right)^{1/2}$, $s,t \in G$, which thus provides a *metric* description of $C_{\text{a.s.}}$ as opposed to the preceding *probabilistic* definition. Equip now the space $C_{\text{a.s.}}$ with the norm

$$[\![a]\!] = \left(\mathbb{E}\|\sum_{\gamma\in\Gamma} a_\gamma g_\gamma \gamma\|^2\right)^{1/2}$$

for which it becomes a Banach space. By (13.9) and Proposition 13.8 an equivalent norm is obtained when (g_γ) is replaced by (ε_γ). In the same way, moment equivalences of both Gaussian and Rademacher series allow to consider L_p-norms for $1 \le p \ne 2 < \infty$. Convenient also is to observe that

$$\frac{1}{4}([\![\operatorname{Re}(a)]\!] + [\![\operatorname{Im}(a)]\!]) \le [\![a]\!] \le 2([\![\operatorname{Re}(a)]\!] + [\![\operatorname{Im}(a)]\!]) \tag{13.10}$$

where $\operatorname{Re}(a) = (\operatorname{Re}(a_\gamma))_\gamma$, $\operatorname{Im}(a) = (\operatorname{Im}(a_\gamma))_\gamma$. The right hand side is clear by the triangle inequality. The left hand side follows from the contraction principle. Indeed, if (b_γ) and (α_γ) are complex numbers such that $|\alpha_\gamma| = 1$ for all γ, then

$$\left(\mathbb{E}\|\sum_\gamma \alpha_\gamma b_\gamma g_\gamma \gamma\|^2\right)^{1/2} \le 2\left(\mathbb{E}\|\sum_\gamma b_\gamma g_\gamma \gamma\|^2\right)^{1/2} .$$

Replacing α_γ by $\overline{\alpha}_\gamma$ and b_γ by $\alpha_\gamma b_\gamma$, we get that, since $|\alpha_\gamma| = 1$,

$$\left(\mathbb{E}\|\sum_\gamma b_\gamma g_\gamma \gamma\|^2\right)^{1/2} \le 2\left(\mathbb{E}\|\sum_\gamma \alpha_\gamma b_\gamma g_\gamma \gamma\|^2\right)^{1/2} .$$

For $\alpha_\gamma = a_\gamma/|a_\gamma|$ and $b_\gamma = \operatorname{Re}(a_\gamma), \operatorname{Im}(a_\gamma)$, (13.10) easily follows using one more time the contraction principle.

It is remarkable that the space $C_{\text{a.s.}}$ which arises from a sup-norm has nice cotype properties. This is the content of the following proposition which, as announced, basically amounts to inequality (13.8).

Proposition 13.11. The space $C_{\text{a.s.}}$ is of cotype 2.

Proof. We need to show that there is some constant C such that if $a^1, \ldots, a^N$ are elements of $C_{\text{a.s.}}$, then

$$\sum_{i=1}^N [\![a^i]\!]^2 \le C\mathbb{E}[\![\sum_{i=1}^N \varepsilon_i a^i]\!]^2 .$$

By observation (13.10), we need only prove this for real elements $a^1, \ldots, a^N$. Consider then the element a of $C_{\text{a.s.}}$ defined for every γ in Γ by $a_\gamma = \left(\sum_{i=1}^N |a_\gamma^i|^2\right)^{1/2}$. By Jensen's inequality and (4.3), it is clear that

$$\mathbb{E}[\![\sum_{i=1}^N \varepsilon_i a^i]\!]^2 \ge \frac{1}{2}[\![a]\!]^2$$

so that we have only to show that $\sum_{i=1}^{N} [\![a^i]\!]^2 \leq C[\![a]\!]^2$. We simply deduce this from (13.8). Independently of the basic orthogaussian sequence (g_γ) , let $A^1, \ldots, A^N$ be disjoint sets with equal probability $1/N$. Let, for γ in Γ ,

$$\xi_\gamma = \sqrt{N} \sum_{i=1}^{N} \frac{a_\gamma^i}{|a_\gamma|} I_{A^i} .$$

Clearly

$$\sum_{i=1}^{N} [\![a^i]\!]^2 = \mathbb{E} \| \sum_\gamma a_\gamma \xi_\gamma g_\gamma \gamma \|^2 .$$

Since $\mathbb{E}|\xi_\gamma|^2 = 1$ for every γ , the conclusion follows from (13.8). Proposition 13.11 is therefore established.

It is interesting to present some remarkable subspaces of the Banach space $C_{\text{a.s.}} = C_{\text{a.s.}}(G)$. Let G be as before a compact Abelian group. A subset Λ of the dual group Γ of G is called a *Sidon* set if there is a constant C such that for every finite sequence (α_γ) of complex numbers

$$\sum_{\gamma \in \Lambda} |\alpha_\gamma| \leq C \| \sum_{\gamma \in \Lambda} \alpha_\gamma \gamma \| .$$

Since the reverse inequality (with $C = 1$) is clearly satisfied, $\{\gamma ;\ \gamma \in \Lambda\}$ generates, when Λ is a Sidon set, a subspace of $C(G)$ isomorphic to ℓ_1 . A typical example is provided by a lacunary sequence (in the sense of Hadamard) in $\mathbb{Z}$, the dual group of the torus group. As we have seen in Chapter 4, Section 4.1, the Rademacher sequence in the Cantor group is another example of Sidon set. If $a = (a_\gamma)_{\gamma \in \Gamma}$ is a sequence of complex numbers vanishing outside some Sidon set Λ of Γ , then the norm $[\![a]\!]$ is equivalent to the ℓ_1 norm $\sum_\gamma |a_\gamma|$. ℓ_1 is of cotype 2 and it is remarkable that the norm $[\![\cdot]\!]$ preserves this property. On the other hand, $C_{\text{a.s.}}$ is of no type $p > 1$ (since this is not the case for ℓ_1).

The consideration of the space $C_{\text{a.s.}}$ gives raise to another interesting observation. Let G be compact Abelian. Any function f in $L_2(G)$ admits a Fourier expansion $\sum_{\gamma \in \Gamma} \widehat{f}(\gamma)\gamma$ which converges to f in $L_2(G)$. We denote by $[\![\widehat{f}]\!]$ the norm in $C_{\text{a.s.}}$ of the Fourier transform $\widehat{f} = (\widehat{f}(\gamma))_{\gamma \in \Gamma}$ of f . We first note the following. Let F be a complex valued function on $\mathbb{C}$ such that

$$|F(x) - F(y)| \leq |x - y| \quad \text{and} \quad |F(x)| \leq 1 \text{ for all } x, y \in \mathbb{C} .$$

Let f be a function in $C(G)$ such that $\widehat{f}$ belongs to $C_{\text{a.s.}}$. Then, $h = F \circ f$ belongs to $C_{\text{a.s.}}$ and for some numerical constant $K_1 \geq 1$,

$$[\![\widehat{h}]\!] \leq K_1 (\|f\| + [\![\widehat{f}]\!]) \tag{13.11}$$

where we recall that $\|\cdot\|$ is the sup-norm (on G). This property is an easy consequence of the comparison theorems for Gaussian processes. For any t in G, set $f_t(x) = f(t+x)$. If (X_t) is the Gaussian process $X_t = \sum_\gamma \widehat{f}(\gamma) g_\gamma \gamma(t)$, for any s, t in G,

$$\|X_s - X_t\|_2 = \|f_s - f_t\|_2$$

where the L_2 norm on the left is understood with respect to the Gaussian sequence (g_γ) and the one on the right with respect to Haar measure on G. For (13.11), since F is 1-Lipschitz, we have that

$$\|h_s - h_t\|_2 \leq \|f_s - f_t\|_2 \quad \text{and} \quad \|h\|_2 \leq 1$$

and the inequality then follows from (for example) Corollary 3.14 (in some complex formulation).

Let B be the space of the functions f in $C(G)$ whose Fourier transform $\widehat{f}$ belongs to $C_{\text{a.s.}} = C_{\text{a.s.}}(G)$. Equip B with the norm

$$|\!|\!| f |\!|\!| = K_1(3\|f\| + [\![\widehat{f}]\!]),$$

where K_1 is the numerical constant which appears in (13.11), for which it becomes a Banach space. The trigonometric polynomials are dense in B (cf. Theorem 3.4). As a result, B is actually a Banach algebra for the pointwise product such that $|\!|\!| fh |\!|\!| \leq |\!|\!| f |\!|\!| \, |\!|\!| h |\!|\!|$ for all f, h in B. This can be proved directly or as a consequence of (13.11). Let F be defined as

$$F(x) = \begin{cases} x^2/2 & \text{if } |x| \leq 1, \\ x^2/2|x|^2 & \text{if} |x| > 1. \end{cases}$$

Then, if f, h are in B with $\|f\|$, $\|h\| \leq 1$, since $4fh = (f+h)^2 - (f-h)^2$, as a corollary of (13.11) applied with this F,

$$[\![\widehat{fh}]\!] \leq 2K_1(2 + [\![\widehat{f}]\!] + [\![\widehat{h}]\!]).$$

The inequality $|\!|\!| fh |\!|\!| \leq |\!|\!| f |\!|\!| \, |\!|\!| h |\!|\!|$ follows from definition of $|\!|\!| \cdot |\!|\!|$.

Let $A(G)$ be the algebra of the elements of $C(G)$ whose Fourier transform is absolutely summable. The preceding algebra B is an example of a (strongly homfflgeneous) Banach algebra with $A(G) \not\subseteq B \not\subseteq C(G)$ on which all Lipschitzian functions of order 1 operate. One might wonder for a *minimal* algebra with these properties. In this order of idea, the following algebra might be of some interest. Let $\widetilde{B}$ be the space of the functions f in $C(G)$ such that

$$\sup\left\{E\sup_{t\in G}\left|\sum_{i=1}^{n} a_i g_i f_t(x_i)\right| \; ; \; n\geq 1\, , \; x_1,\ldots,x_n\in G\, , \; \sum_{i=1}^{n} a_i^2\leq 1\right\}<\infty\, .$$

It is not difficult to see, as before, that this quantity (at the exception perhaps of a numerical factor) defines a norm on $\widetilde{B}$ for which $\widetilde{B}$ is a Banach algebra with $A(B)\not\subseteq\widetilde{B}\not\subseteq C(G)$ on which all 1-Lipschitzian functions operate. Further, $\widetilde{B}$ is smaller than B. To see it, let (Z_i) be independent random variables with values in G and common distribution λ (normalized Haar measure on G). Then, if $f\in\widetilde{B}$,

$$\sup_n \frac{1}{\sqrt{n}}\mathbb{E}\sup_{t\in G}|\sum_{i=1}^{n} g_i f_t(Z_i)|<\infty\, .$$

Therefore, by the central limit theorem in finite dimension, it follows that the Gaussian process with L_2-metric

$$(\mathbb{E}|f_s(Z_1)-f_t(Z_1)|^2)^{1/2}=\|f_s-f_t\|_2\, , \qquad s,t\in G\, ,$$

is almost surely bounded. But then (Theorem 13.4), f belongs to $C_{\text{a.s.}}$ which yields the claim.

A deeper analysis of the Banach algebra $\widetilde{B}$ is still to be done. Is it in particular the smallest on which the 1-Lipschitzian functions operate?

13.3. Stable random Fourier series and strongly stationary processes

In the preceding sections, we investigated stationary Gaussian processes and random Fourier series of the type $\sum_\gamma a_\gamma \xi_\gamma \gamma$ where (ξ_γ) is a symmetric sequence of real random variables satisfying basically $\sup_\gamma \mathbb{E}|\xi_\gamma|^2 < \infty$. We shall be interested here in possible extensions to the stable case and to random Fourier series as before when (ξ_γ) satisfies some weaker moment assumptions (a typical example of both topics being given by the choice, for (ξ_γ), of a standard p-stable sequence).

Throughout this section, G denotes a locally compact Abelian group with dual group Γ, identity 0 and Haar measure $\lambda(\cdot) = |\cdot|$. If $X = (X_t)_{t\in G}$ is a *stationary* Gaussian process continuous in L_2, by Bochner's theorem, there is a measure m in Γ which represents the covariance of X in the sense that for all finite sequences of real numbers (α_j) and (t_j) in G,

$$\mathbb{E}|\sum_j \alpha_j X_{t_j}|^2 = \int_\Gamma |\sum_j \alpha_j \gamma(t_j)|^2 \, dm(\gamma) .$$

Let $0 < p \leq 2$. Say that a p-stable process $X = (X_t)_{t\in G}$ indexed by G is *strongly stationary*, or *harmonizable*, if there is a finite positive Radon measure m concentrated on Γ such that, for all finite sequences (α_j) of real numbers and (t_j) of elements of G,

$$\mathbb{E}\exp i \sum_j \alpha_j X_{t_j} = \exp\left(-\frac{1}{2}\int_\gamma \left|\sum_j \alpha_j \gamma(t_j)\right|^p dm(\gamma)\right) . \tag{13.12}$$

Going back to the spectral representation of stable processes (Theorem 5.2), we thus assume in this definition of strongly stationary stable processes a special property of the spectral measure m, namely to be concentrated on the dual group Γ of G. This property is motivated by several facts, some of which will become clear in the subsequent developments. In particular, strongly stationary stable processes are stationary in the usual sense; however, and refering to [M-P2], contrary to the Gaussian case not all stationary stable processes are strongly stationary.

It is worthwhile mentioning an example. Let $\widetilde{\theta}$ be a complex stable random variable such that, as a variable in $\mathbb{R}^2$, $\widetilde{\theta}$ has for spectral measure uniform distribution on the unit circle. That is, if $\widetilde{\theta} = \theta_1 + i\theta_2 = (\theta_1, \theta_2)$ and $\alpha = \alpha_1 + i\alpha_2 = (\alpha_1, \alpha_2) \in \mathbb{C}$,

$$\mathbb{E}\exp i \ \mathrm{Re}(\overline{\alpha}\widetilde{\theta}) = \mathbb{E}\exp i(\alpha_1\theta_1 + \alpha_2\theta_2) = \exp\left(-\frac{1}{2}{c'_p}^p |\alpha|^p\right) \tag{13.13}$$

where $c'_p = \left(\int_0^{2\pi} |\cos x|^p dx/2\pi\right)^{1/p}$. (Only in the Gaussian case, $p = 2$, θ_1 and θ_2 are necessarily independent.) This definition is one 2-dimensional extension of the real stable variables for which spectral measures are concentrated on $\{-1,+1\}$. Let A be a countable subset of Γ. Let further $(\widetilde{\theta}_\gamma)_{\gamma\in A}$ be a sequence of independent variables distributed as $\widetilde{\theta}$ and $(a_\gamma)_{\gamma\in A}$ be complex numbers with $\sum\limits_\gamma |a_\gamma|^p < \infty$. Then

$$X_t = \sum_{\gamma\in A} a_\gamma \widetilde{\theta}_\gamma \gamma(t)\,, \qquad t \in G\,,$$

defines a *complex* strongly stationary p-stable process. Its real and imaginary parts are real strongly stationary p-stable processes in the sense of definition (13.12). The spectral measure is discrete in this case. Strongly stationary stable processes therefore include random Fourier series with stable coefficients. Since, up to a constant depending on p only, $\mathrm{Re}\widetilde{\theta}$ and $\mathrm{Im}\widetilde{\theta}$ are standard p-stable real variables, this study can be shown to include similarly random Fourier series of the type $\sum\limits_\gamma a_\gamma\theta_\gamma\gamma$ where (θ_γ) is a standard p-stable sequence (real). We shall come back to this from a somewhat simpler point of view later.

In the first part of this section, we extend to strongly stationary p-stable processes, $1 \le p < 2$, the Gaussian characterization of Theorem 13.3. Necessity will follow easily from Section 12.2 and Proposition 13.2. Sufficiency uses the series representation of stable processes (Corollary 5.3) together with ideas developed in the preceding section on random Fourier series (conditional estimates and integration in presense of translation invariant metrics). As usual, we exclude from such a study the case $0 < p < 1$ since the only property that m is finite already ensures that a p-stable process, $0 < p < 1$, has a version with bounded and continuous paths. It is thus assumed henceforth that $1 \le p \le 2$, and actually also that $p < 2$ since the case $p = 2$ has already been studied.

Let $X = (X_t)_{t\in G}$ be a strongly stationary p-stable process with $1 \le p < 2$ and with spectral measure m in (13.12). For s,t in G, denote by $d_X(s,t)$ the parameter of the real p-stable variable $X_s - X_t$, that is

$$d_X(s,t) = \left(\int_\Gamma |\gamma(s) - \gamma(t)|^p dm(\gamma)\right)^{1/p}.$$

d_X defines a translation invariant pseudo-metric. Let V be a fixed compact neighborhood of the unit element 0 of G. We shall always assume that V is metrizable. d_X is then continuous on $V \times V$ (by dominated convergence) (and thus also on $G \times G$). We know from Theorem 12.12 a general necessary condition for boundedness of p-stable processes. Together with Proposition 13.2, it yields that if X has a version with almost all sample paths bounded on V, then $E^{(q)}(V, d_X) < \infty$ where $q = p/p-1$ is the conjugate of p. Furthermore, for some constant K_p depending only on p,

$$\|\sup_{t\in V} |X_t|\|_{p,\infty} \geq K_p^{-1}(E^{(q)}(V, d_X) - L(V)^{1/q} D(V)) \tag{13.14}$$

where $D(V)$ is the diameter of (V, d_X) and $L(V) = \log(|V''|/|V|)$ $(V'' = V + V + V)$. (When $q = \infty$ we agree that $L(V)^{1/q} = \log(1 + \log(|V''|/|V|))$.) In the following, $\|\sup_{t\in V} |X_t|\|_{p,\infty}$ will be denoted for simplicity by $\|X\|_{p,\infty}$.

This settles the necessity portion of this study of almost sure boundedness and continuity of strongly stationary p-stable processes. We now turn to sufficiency and show, as a main result, that the preceding necessary entropy condition is also sufficient. As announced, the proof makes use of various arguments developed in Section 13.2. We only prove the result for $1 < p < 2$. The case $p = 1$ can basically be obtained similarly with however some more care. We refer to [Ta14] for the case $p = 1$.

Theorem 13.12. Let $X = (X_t)_{t\in G}$ be a strongly stationary p-stable process, $1 \leq p < 2$. Let $q = p/p-1$. Then X has a version with almost all sample paths bounded and continuous on $V \subset G$ if and only if $E^{(q)}(V, d_X) < \infty$. Moreover, there is a constant K_p depending on p only such that

$$\begin{aligned} &K_p^{-1}(|m|^{1/p} + E^{(q)}(V, d_X) - L(V)^{1/q} D(V)) \\ &\qquad \leq \|X\|_{p,\infty} \leq K_p((1 + L(V)^{1/q})|m|^{1/p} + E^{(q)}(V, d_X)) \end{aligned}$$

where m is the spectral measure of X and $D(V)$ the diameter of (V, d_X).

Proof: Necessity and the left hand side inequality of the theorem have been discussed above. Recall the series representation of Corollary 5.3. Let Y_j be independent random variables distributed as $m/|m|$. Let further w_j be independent complex random variables all of them uniformly distributed on the unit circle of $\mathbb{R}^2$. Assume the sequences (ε_j), (Γ_j), (w_j), (Y_j) to be independent. Then, by Corollary 5.3 and (13.12), X has the same distribution as

$$c_p (c_p')^{-1} |m|^{1/p} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j \ \mathrm{Re}(w_j Y_j(t)), \qquad t \in G,$$

where c_p' appears in (13.13). For simplicity in the notation, we denote again by X this representation. Under $E^{(q)}(V, d_X) < \infty$, we will show, *exactly* as in the proof of Theorem 13.6, that this series has a version with almost all sample paths bounded and continuous, satisfying moreover the inequality of the theorem. By homogeneity, let us assume that $|m| = 1$. Conditionally on (Γ_j),

(w_j) and (Y_j), we apply Lemma 13.5. For every ω of the probability space supporting these sequences, denote by $d_\omega(s,t)$ the translation invariant functional

$$d_\omega(s,t) = \|(\Gamma_j^{-1/p}(\omega)\ \mathrm{Re}(w_j(\omega)(Y_j(\omega,s) - Y_j(\omega,t))))\|_{p,\infty} .$$

Since the Y_j 's take their values in Γ, the dual group of G, it is easy to verify that d_ω is, for almost all ω, continuous on $V \times V$. Indeed, if τ is a metric for which V is compact, we deduce from (5.8) and Corollary 5.9 that

$$\mathbb{E} \sup_{\substack{\tau(s,t)<\varepsilon \\ s,t\in V}} d_\omega(s,t) \le K_p \mathbb{E} \sup_{\substack{\tau(s,t)<\varepsilon \\ s,t\in V}} |Y_1(s) - Y_1(t)|$$

for all $\varepsilon > 0$. (K_p denotes here and below some constant only depending on $1 < p < 2$.) The claim thus follows. Further, since $\|\ \mathrm{Re}(w_1(Y_1(s) - Y_1(t)))\|_p = c_p' d_X(s,t)$, we have from exactly the same tools that

$$\mathbb{E} d_\omega(s,t) \le K_p d_X(s,t) \tag{13.15}$$

for all s,t. Once this has been observed, the rest of the proof is entirely similar to the proof of Theorem 13.6. By Lemma 13.5, letting $D(\omega) = 2\|(\Gamma_j^{-1/p}(\omega))\|_{p,\infty}$,

$$\begin{aligned} &\mathbb{E}_\varepsilon \sup_{t\in V} \left| \sum_{j=1}^{\infty} \Gamma_j^{-1/p}(\omega)\varepsilon_j\ \mathrm{Re}(w_j(\omega)Y_j(\omega,t)) \right| \\ &\le K_p \left(D(\omega) + \int_0^{D(\omega)} (\log N(V,d_\omega;\varepsilon))^{1/q} d\varepsilon \right) \end{aligned} \tag{13.16}$$

and, if this entropy integral is finite, $\left(\sum\limits_{j=1}^{\infty} \Gamma_j^{-1/p}(\omega)\varepsilon_j\ \mathrm{Re}(w_j(\omega)Y_j(\omega,t)) \right)_{t\in V}$ has a version with almost all (with respect to (ε_j)) sample paths continuous (since d_ω is continuous on $V \times V$). W_n and $b_n(\omega)$ being as in the proof of Theorem 13.6, we have from (13.15) that $\mathbb{E} b_n \le K_p 2^{-n}$. Furthermore, from Lemma 13.7 and exactly in the same way,

$$\int_0^{D(\omega)} (\log N(V,d_\omega;\varepsilon))^{1/q} d\varepsilon \le \sum_{n=0}^{\infty} b_n(\omega) \left(\log\left(2\frac{|V''|}{|V|} N(V,d_X;2^{-n-1}) \right) \right)^{1/q} .$$

Integrating with respect to ω, it follows that, when $E^{(q)}(V,d_X)<\infty$, for almost all ω the entropy integral on the left is finite and thus, by the preceding, $\left(\sum_{j=1}^{\infty}\Gamma_j^{-1/p}(\omega)\varepsilon_j\ \mathrm{Re}(w_j(\omega)Y_j(\omega,t))\right)_{t\in T}$ has a version with continuous sample paths with respect to (ε_j). Therefore, by Fubini's theorem and the representation, X has a version with continuous paths on V. Furthermore, integrating (13.16) together with the fact that $\mathbb{E}b_n\le K_p2^{-n}$ yields

$$\mathbb{E}\|X\|\le K_p(1+L(V)^{1/q}+E^{(q)}(V,d_X)).$$

Since $\mathbb{E}\|X\|$ is equivalent to $\|X\|_{p,\infty}$ (Proposition 5.6), the proof of Theorem 13.12 is thus complete (recall we have assumed by homogeneity that $|m|=1$).

Motivated by the previous result, we further investigate in the second part of this section random Fourier series with the objective of enlarging the conclusions of Theorem 13.6 or Corollary 13.9. In particular, we would like to study the case of a sequence (ξ_γ) there not necessarily in L_2. One typical example is a stable random Fourier series $\sum_\gamma a_\gamma\theta_\gamma\gamma$ where (θ_γ) is a standard p-stable sequence. We have seen that this example can be shown to enter the previous setting. We now present some natural extensions in the context of random Fourier series.

As in Section 13.2, G is a locally compact Abelian group with unit 0 and dual group Γ, V a fixed compact symmetric neighborhood of 0 and A a countable subset of Γ. Let $1\le p<2$. Let $(a_\gamma)_{\gamma\in A}$ be a sequence of complex numbers such that $\sum_\gamma|a_\gamma|^p<\infty$ and let $(\xi_\gamma)_{\gamma\in A}$ be a sequence of *independent and symmetric* real random variables. We are interested in the almost sure uniform convergence of the random Fourier series $Y=(Y_t)_{t\in G}$ where

$$Y_t=\sum_{\gamma\in A}a_\gamma\xi_\gamma\gamma(t)\,,\qquad t\in G\,, \tag{13.17}$$

in terms of the translation invariant pseudo-metric

$$d(s,t)=\left(\sum_{\gamma\in A}|a_\gamma|^p|\gamma(s)-\gamma(t)|^p\right)^{1/p},\qquad s,t\in G\,.$$

The technique of proof of Theorems 13.6 and 13.12 enables to extend the results of Section 13.2 to random variables ξ_γ which do not have finite second moments.

Theorem 13.13. Assume that $\sup\limits_\gamma \|\xi_\gamma\|_{p,\infty} < \infty$. Then, if $E^{(q)}(V,d) < \infty$ where $q = p/p-1$ and d is defined above, the random Fourier series Y of (13.17) converges uniformly on V with probability one. Further, for some constant K_p depending on p only,

$$\|Y\|_{p,\infty} \leq K_p \sup_\gamma \|\xi_\gamma\|_{p,\infty} \left((1 + L(V)^{1/q}) \left(\sum_\gamma |a_\gamma|^p \right)^{1/p} + E^{(q)}(V,d) \right)$$

(where $\|Y\|_{p,\infty} = \| \sup\limits_{t \in V} |Y_t| \|_{p,\infty}$).

Proof. It is entirely similar (actually somewhat simpler) to the proofs of Theorems 13.6 and 13.12 so that we only mention a few observations. By independence and symmetry, Y can be replaced by

$$\sum_{\gamma \in A} a_\gamma \varepsilon_\gamma \xi_\gamma \gamma(t) , \qquad t \in G ,$$

where (ε_γ) is a Rademacher sequence independent of (ξ_γ). We then use Lemma 13.5 conditionally on (ξ_γ) with respect to the metric

$$d_\omega(s,t) = \|(|a_\gamma \xi_\gamma(\omega)(\gamma(s) - \gamma(t))|)_{\gamma \in A}\|_{p,\infty} .$$

Since the ξ_γ are *independent*, we can integrate with respect to ω and use Lemma 5.8 and the hypothesis $\sup\limits_\gamma \|\xi_\gamma\|_{p,\infty} < \infty$. The proof is completed similarly.

Note that if the sequence (ξ_γ) is only a *symmetric* sequence, the preceding theorem holds similarly *but* with $\sup\limits_\gamma \|\xi_\gamma\|_p$ instead of $L_{p,\infty}$ moments. The argument is similar but, to integrate d_ω, since the ξ_γ need not be independent, we simply use that

$$d_\omega(s,t) \leq \left(\sum_\gamma |a_\gamma|^p |\xi_\gamma(\omega)|^p |\gamma(s) - \gamma(t)|^p \right)^{1/p} .$$

If the tails of the random variables ξ_γ are close to the tail of a standard p-stable variable, $1 \leq p < 2$, then the entropy condition $E^{(q)}(V,d) < \infty$ is also necessary for the random Fourier series Y of (13.17) to be almost surely bounded. More precisely, assume that for some $u_0 > 0$ and $\delta > 0$,

$$\mathbb{P}\{|\xi_\gamma| > u\} \geq \delta u^{-p}$$

for all $u > u_0$ and all γ. Hence, by (5.2), if (θ_γ) is a standard p-stable sequence,

$$\inf_\gamma \mathbb{P}\{|\xi_\gamma| > u\} / \mathbb{P}\{|\theta_\gamma| > u\}$$

is bounded below for u sufficiently large. Therefore, if $\sum_\gamma a_\gamma \xi_\gamma \gamma$ converges uniformly almost surely, the same holds for $\sum_\gamma a_\gamma \theta_\gamma \gamma$ by Lemma 4.6. But now we deal with a stable process and the complex version of (13.14) yields the announced claim. This approach can of course also be used to deduce Theorem 13.13 from Theorem 13.12.

Finally, the various comments developed in the Gaussian setting next to Theorem 13.3 also apply in this stable case. Similarly for the equivalences of Theorem 13.4 in the context of stable Fourier series $\sum_\gamma a_\gamma \theta_\gamma \gamma$.

13.4. Vector valued random Fourier series

In the last part of this chapter, we present some applications of the previous results to vector valued stationary processes and random Fourier series. The results are still fragmentary and only concern so far Gaussian variables.

Let B be a separable Banach space with dual space B'. Recall that by a process $X = (X_t)_{t \in T}$ with values in B we simply mean a family such that, for each t, X_t is a Borel random variable with values in B. X is Gaussian if for every $t_1, \dots t_N$ in T, $(X_{t_1}, \dots, X_{t_N})$ is Gaussian (in B^N). As in the preceding sections, let G be a locally compact Abelian group with identity 0 and dual group of characters Γ. Let us fix also a compact metrizable neighborhood V of 0. A process $X = (X_t)_{t \in G}$ indexed by G with values in B is said to be stationary if, as in the real case, for every u in G, $(X_{u+t})_{t \in G}$ has the same distribution (on B^G) as X. Since B is separable, this is equivalent to say that for every f in B', the real process $(f(X_t))_{t \in G}$ is stationary.

Almost sure boundedness and continuity of vector valued stationary Gaussian processes may be characterized rather easily through the corresponding properties along linear functionals. This is the content of the following statement. While this result is related to tensorization of Gaussian measures studied in Section 3.3, it does not seem possible to deduce it from the comparison theorems based on Slepian's lemma. Instead we use majorizing measures and the deep results of Chapter 12. Recall that $L(V) = \log(|V''|/|V|)$, $V'' = V + V + V$.

Theorem 13.14. Let $X = (X_t)_{t \in G}$ be a stationary Gaussian process with values in B. Then, for some numerical constant K,

$$\frac{1}{2}\Big(\mathbb{E}\|X_0\| + \sup_{\|f\| \le 1} \mathbb{E} \sup_{t \in V} |f(X_t)|\Big) \le \mathbb{E} \sup_{t \in V} \|X_t\| \le K(1 + L(V)^{1/2})\Big(\mathbb{E}\|X_0\| + \sup_{\|f\| \le 1} \mathbb{E} \sup_{t \in V} |f(X_t)|\Big).$$

Further, X has a version with almost all continuous paths on V if and only if $\|X_s - X_t\|_2$ is continuous on $V \times V$ and

$$\lim_{\eta \to 0} \sup_{\|f\| \leq 1} \mathbb{E} \sup_{t \in \eta V} |f(X_t)| = 0 .$$

Note that, by the results of Section 13.1, X is continuous on V if and only if

$$\lim_{\eta \to 0} \sup_{\|f\| \leq 1} \int_0^\eta (\log N(V, d_{f(X)}; \varepsilon))^{1/2} d\varepsilon = 0$$

(and $\|X_s - X_t\|_2$ is continuous on $V \times V$) where $d_{f(X)}(s,t) = \|f(X_s) - f(X_t)\|_2$, $f \in B'$, $s,t \in G$.

Proof. We only show boundedness and the inequalities of the theorem. Continuity follows similarly together with the preceding observation. Let B'_1 be the unit ball of B'. We consider X as a process indexed by $B'_1 \times G$. By Theorem 12.9, the real Gaussian process $X_0 = (f(X_0))_{f \in B'_1}$ indexed by B'_1 has a majorizing measure; that is, there exists a probability measure m on (B'_1, d_{X_0}) such that

$$\sup_{f \in B'_1} \int_0^\infty \left(\log \frac{1}{m(B(f,\varepsilon))} \right)^{1/2} d\varepsilon \leq K \mathbb{E} \sup_{f \in B'_1} f(X_0) = K \mathbb{E} \|X_0\| \tag{13.18}$$

where, as in Chapter 12, $B(f,\varepsilon)$ is the ball of radius ε with respect to the metric on the space which contains its center f, that is, $d_{X_0}(f,g) = \|f(X_0) - g(X_0)\|_2$, $f,g \in B'_1$. (We use further this convention about balls in metric spaces below.) We intend to use (13.1) so let $\lambda_{V''}$ be restricted normalized Haar measure on $V'' \subset G$. If we bound X considered as a (real) process on $B'_1 \times V$ with the majorizing measure integral for $m \times \lambda_{V''}$ (on $B'_1 \times V'' \supset B'_1 \times V$), we get from Theorem 11.18 and Lemma 11.9 (which applies similarly for the function $(\log(1/x))^{1/2}$) that,

$$\mathbb{E} \sup_{(f,t) \in B'_1 \times V} f(X_t) \leq K \sup_{(f,t) \in B'_1 \times V} \int_0^\infty \left(\log \frac{1}{m \times \lambda_{V''}(B((f,t),\varepsilon))} \right)^{1/2} d\varepsilon \tag{13.19}$$

where $B((f,t),\varepsilon)$ is the ball for the L_2 pseudo-metric of X on $B'_1 \times G$, i.e.

$$d((f,t),(g,s)) = \|f(X_t) - g(X_s)\|_2 ,$$

$f,g \in B'$, $s,t \in G$. To control the integral on the right of (13.19), note the following. By the triangle inequality and stationarity,

$$d((f,t),(g,s)) \leq d_{X_0}(f,g) + d_{f(X)}(s,t)$$

where we recall that $d_{f(X)}(s,t) = \|f(X_s) - f(X_t)\|_2$. It follows that, for all (f,t) in $B_1' \times V$ and all $\varepsilon > 0$,

$$m \times \lambda_{V''}(B((f,t),2\varepsilon)) \geq m(B(f,\varepsilon))\lambda_{V''}(B_{d_{f(X)}}(t,\varepsilon))$$

where $B_{d_{f(X)}}(t,\varepsilon)$ is the ball with respect to the metric $d_{f(X)}$. Therefore,

$$\mathbb{E} \sup_{B_1' \times V} f(X_t) \leq 2K \left(\sup_{B_1'} \int_0^\infty \left(\log \frac{1}{m(B(f,\varepsilon))} \right)^{1/2} d\varepsilon + \sup_{B_1' \times V} \int_0^\infty \left(\log \frac{1}{\lambda_{V''}(B_{d_{f(X)}}(t,\varepsilon))} \right)^{1/2} d\varepsilon \right) .$$

The first term on the right of this inequality is controlled by (13.18). We use (13.1) (which applies similarly with the function $(\log(1/x))^{1/2}$) to see that the second term is smaller than or equal to

$$\sup_{f \in B_1'} \int_0^\infty \left(\log \left(\frac{|V''|}{|V|} N(V, d_{f(X)}; \varepsilon) \right) \right)^{1/2} d\varepsilon .$$

Summarizing, we get from Theorem 13.3 that for some numerical constant K ,

$$\mathbb{E} \sup_{t \in V} \|X_t\| \leq K(\mathbb{E}\|X_0\| + \sup_{f \in B_1'} \mathbb{E} \sup_{t \in V} |f(X_t)| + L(V)^{1/2} \sup_{f \in B_1'} (\mathbb{E} f^2(X_0))^{1/2}) .$$

This inequality is stronger than the upper bound of the theorem. The minoration inequality is obvious. The proof is, therefore, complete.

One interesting application of Theorem 13.14 concerns Gaussian random Fourier series with vector valued coefficients. Let A be a fixed countable subset of the dual group of characters Γ of G . Let $(g_\gamma)_{\gamma \in A}$ be an orthogaussian sequence and $(x_\gamma)_{\gamma \in A}$ be a sequence of elements of a Banach space B . We assume that B is a complex Banach space. Suppose that the series $\sum_\gamma g_\gamma x_\gamma$ is convergent. Define then (using the contraction principle) the Gaussian random Fourier series $X = (X_t)_{t \in G}$ by

$$X_t = \sum_{\gamma \in A} g_\gamma x_\gamma \gamma(t) , \qquad t \in G . \tag{13.20}$$

As in the scalar case, one might wonder for the almost sure uniform convergence of the series (13.20) (in the sup-norm $\sup_{t \in V} \| \cdot \|$) or, equivalently (by Itô-Nisio's theorem), the almost sure continuity of the process X on V . Theorem 13.14 implies the following.

Corollary 13.15. In the preceding notations, there is a numerical constant K such that

$$\begin{aligned}\frac{1}{2}\left(\mathbb{E}\left\|\sum_\gamma g_\gamma x_\gamma\right\| + \sup_{\|f\|\le 1}\mathbb{E}\sup_{t\in V}\left|\sum_\gamma f(x_\gamma)g_\gamma\gamma(t)\right|\right) \\ \le \mathbb{E}\sup_{t\in V}\left\|\sum_\gamma g_\gamma x_\gamma\gamma(t)\right\| \\ \le K(1+L(V)^{1/2})\left(\mathbb{E}\left\|\sum_\gamma g_\gamma x_\gamma\right\| + \sup_{\|f\|\le 1}\mathbb{E}\sup_{t\in V}\left|\sum_\gamma f(x_\gamma)g_\gamma\gamma(t)\right|\right)\,.\end{aligned}$$

Further, $\sum_\gamma g_\gamma x_\gamma\gamma$ converges uniformly almost surely if and only if

$$\lim \sup_{\|f\|\le 1}\mathbb{E}\sup_{t\in V}\left|\sum_{\gamma\in F^c} f(x_\gamma)g_\gamma\gamma(t)\right| = 0$$

where the limit is taken over the finite sets F increasing to A.

The scalar case investigation of Sections 13.2 and 13.3 invites to consider the same convergence question of vector valued random Fourier series of the type (13.20) when for example the Gaussian sequence (g_γ) is replaced by a Rademacher sequence (ε_γ) or a standard p-stable sequence (θ_γ), $1\le p<2$. These questions are not yet answered. By the equivalence of scalar Gaussian and Rademacher Fourier series (Proposition 13.8), it is plain from Corollary 13.15 that a Rademacher series $\sum_\gamma \varepsilon_\gamma x_\gamma\gamma$ is characterized as the corresponding Gaussian one provided $\sum_\gamma \varepsilon_\gamma x_\gamma$ and $\sum_\gamma g_\gamma x_\gamma$ converge simultaneously. We know that this holds for all sequences (x_γ) in B if and only if B is of finite cotype (Theorem 9.16) but in general $\sum_\gamma \varepsilon_\gamma x_\gamma$ is only dominated by $\sum_\gamma g_\gamma x_\gamma$ ((4.8)). We however conjecture that Corollary 13.15 and its inequality also hold when (g_γ) is replaced by (ε_γ) (note of course that the left hand side inequality is trivial). This conjecture is supported by the fact that it holds for Rademacher processes which are $Kr(T)$-decomposable in the terminology of Section 12.3; this is checked immediately reproducing the argument of the proof of Theorem 13.14. The case of a p-stable standard sequence, $1\le p<2$, in Corollary 13.15 is also open.

Stationary real Gaussian (and strongly stationary p-stable) processes are either continuous or unbounded. This follows from the characterizations we described and extends to the classes of random Fourier series studied there. To conclude, we analyze this dichotomy for general random Fourier series with vector valued coefficients.

Let A be a countable subset of Γ. Let further $(x_\gamma)_{\gamma\in A}$ be a sequence in a complex Banach space B and let $(\xi_\gamma)_{\gamma\in A}$ be independent symmetric real random variables. Assuming that $\sum_\gamma \xi_\gamma x_\gamma$ is almost surely convergent for one (or, by symmetry, all) ordering of A, consider the random Fourier series $X = (X_t)_{t\in G}$ given by

$$X_t = \sum_{\gamma\in A} \xi_\gamma x_\gamma \gamma(t)\,, \qquad t \in G\,. \tag{13.21}$$

Let V be as usual a compact neighborhood of 0 in G. The random Fourier series X is said to be almost surely uniformly bounded if for some ordering of $A = \{\gamma_n; n \geq 1\}$ the partial sums $\sum_{n=1}^{N} \xi_{\gamma_n} x_{\gamma_n} \gamma_n$, $N \geq 1$, are almost surely uniformly bounded with respect to the norm $\sup_{t\in V} \|\cdot\|$. Since (ξ_γ) is a symmetric sequence, by Lévy's inequalities, the preceding is independent of the ordering of A and is equivalent to the boundedness of X as a process on V. Similarly, the random Fourier series X of (13.21) is almost surely uniformly convergent if for some (or all by Itô-Nisio's theorem) ordering $A = \{\gamma_n; n \geq 1\}$, the preceding partial sums converge uniformly almost surely. Equivalently, X defines an almost surely continuous process on V with values in B. The next theorem describes how these two properties are equivalent for scalar random Fourier series of the type (13.21) and similarly for B-valued coefficients if B does not contain an isomorphic copy of c_0. The proof is based on Theorem 9.29 which is identical for complex Banach spaces.

Theorem 13.16. For a Banach space B, the following are equivalent:

(i) B does not contain subspaces isomorphic to c_0.

(ii) Every almost surely uniformly bounded random Fourier series of the type (13.21) with coefficients in B is almost surely uniformly convergent.

Proof. If (ii) does not hold, there exists a series (13.21) such that, for some ordering $A = \{\gamma_n; n \geq 1\}$, $\sum_n \xi_{\gamma_n} x_{\gamma_n} \gamma_n$ is an almost surely bounded series which does not converge. Let us set for simplicity $\xi_n = \xi_{\gamma_n}$ and $x_n = x_{\gamma_n}$ for all n. By Remark 9.31, there exist ω and a sequence (n_k) such that $(\xi_{n_k}(\omega) x_{n_k} \gamma_{n_k})$ is equivalent in the norm $\sup_{t\in V} \|\cdot\|$ to the canonical basis of c_0. Note that $\|\xi_{n_k}(\omega) x_{n_k}\| = \sup_{t\in V} \|\xi_{n_k}(\omega) x_{n_k} \gamma_{n_k}(t)\|$ so that, in particular (since $\gamma_{n_k}(0) = 1$),

$$\inf_k \|\xi_{n_k}(\omega) x_{n_k}\| > 0\,.$$

In the same way, for every finite sequence (α_k) of complex numbers with $|\alpha_k| \leq 1$,

$$\Big\| \sum_k \alpha_k \xi_{n_k}(\omega) x_{n_k} \Big\| \leq C$$

for some constant C. We can then apply Lemma 9.30 to extract a further subsequence from $(\xi_{n_k}(\omega)x_{n_k})$ which will be equivalent to the canonical basis of c_0. This shows that (i) $\Rightarrow$ (ii). To prove the converse implication, we exhibit a random Fourier series of the type (13.21) (actually Gaussian) with coefficients in c_0 which is bounded but not uniformly convergent. Let G be the compact Cantor group $\{-1,+1\}^{\mathbb{N}}$ and set $V = G$. The characters on G consist of the Rademacher coordinate maps $\varepsilon_n(t)$. On some (different) probability space, let (g_n) be a standard Gaussian sequence. For every n, let further $I(n)$ denote the set of integers $\{2^n+1,\ldots,2^{n+1}\}$. Define then $X = (X_t)_{t\in G}$ by

$$X_t = \sum_n 2^{-n}\left(\sum_{i\in I(n)} g_i\varepsilon_i(t)\right) e_n\,, \qquad t\in G\,,$$

where (e_n) is the canonical basis of c_0. X is a (Gaussian) random Fourier series with values in c_0. It is almost surely bounded. To see it, note that

$$\begin{aligned}\sup_N\sup_{t\in G}\left\|\sum_{n=1}^N 2^{-n}\left(\sum_{i\in I(n)} g_i\varepsilon_i(t)\right)e_n\right\| &= \sup_N\sup_{t\in G}\sup_{n\le N} 2^{-n}\left|\sum_{i\in I(n)} g_i\varepsilon_i(t)\right| \\ &= \sup_n 2^{-n}\sum_{i\in I(n)}|g_i|\end{aligned}$$

(where we have used that (ε_n) generates ℓ_1 in $L_\infty(G)$). Now $\sup\limits_n 2^{-n}\sum\limits_{i\in I(n)}|g_i| < \infty$ almost surely. Indeed, $\sup\limits_n 2^{-n}\sum\limits_{i\in I(n)}\mathbb{E}|g_i| < \infty$, and, by Chebyshev's inequality,

$$\mathbb{P}\left\{\left|\sum_{i\in I(n)}|g_i| - \mathbb{E}|g_i|\right| > 2^n\right\} \le 2^{-n}\,.$$

Hence the claim by the Borel-Cantelli lemma. From exactly the same argument, X is not almost surely uniformly convergent. The proof of Theorem 13.16 is complete.

Notes and References

The main references to this chapter are the book [M-P1] by M. B. Marcus and G. Pisier and their paper [M-P2] (see also [M-P3]). Random Fourier series go back to Paley, Salem and Zygmund. Kahane's ideas [Ka1] significantly contributed to the neat achievements of [M-P1].

Theorem 13.3 is due to X. Fernique [Fer4] (with of course a direct entropic proof). (See also [J-M3] for an exposition of this result more in the setting of random Fourier series.) It is the translation invariant version of the results of Section 12.1 and the key point in the subsequent investigation of random Fourier series. The equivalence of boundedness and continuity of stationary Gaussian processes was known previously as Belaev's dichotomy [Bel] (see also [J-M3]) (a similar result for random Fourier series was proved by Billard, see [Ka1]). The basic Theorem 13.6 (in the case $G = \mathbb{R}$) is due to M. B. Marcus [Ma1], extended later in [M-P1]. The proof we present is somewhat different and simpler; it has been put forward in [Ta14]. It does not use non-decreasing rearrangements as presented in [M-P1] (see also [Fer6]). The equivalence between Gaussian and Rademacher random Fourier series was put forward in [Pi6] and [M-P1]. The remarkable Banach space $C_{\text{a.s.}}(G)$ and associated Banach algebra $C_{\text{a.s.}}(G) \cap C(G)$ have been investigated by G. Pisier [Pi6], [Pi7]. He further provided an harmonic analysis description of $C_{\text{a.s.}}$ as the predual of a space of Fourier multipliers. A Sidon set generates a subspace isomorphic to ℓ_1 in $C(G)$. As a remarkable result, it was shown conversely by J. Bourgain and V. D. Milman [B-M] that if a subset Λ of Γ is such that the subspace C_Λ of $C(G)$ of all functions whose Fourier transform is supported by Λ is of finite cotype (i.e. does not contain ℓ_∞^n 's uniformly), then Λ must be a Sidon set. (A prior contribution assuming C_Λ of cotype 2 is due to G. Pisier [Pi5]). For further conclusions on random Fourier series, in particular in non-Abelian groups, examples and quantitative estimates, we refer to [M-P1].

The results of Section 13.3 are taken from [M-P2] for the case $1 < p < 2$. The picture is completed in [Ta14] with the case $p = 1$ (and with a proof which inspired the proofs presented here). The complex probabilistic structures are carefully described in [M-P2]. Extensions to random Fourier series with infinitely divisible coefficients and ξ -radial processes are studied in [Ma4]. Further extensions to very general Random Fourier series and harmonic processes are obtained in [Ta17].

The study of stationary vector valued Gaussian process was initiated by X. Fernique to whom Theorem 13.14 is due [Fer10] (see also [Fer12]). He further extended this result in [Fer13]. Since the conclusion does not involve majorizing measures, one might wonder for a proof that does not use this tool. Theorem 13.14 was recently used in [I-M-M-T-Z] and [Fer14]. Theorem 13.16 is perhaps new.

Finally, related to the results of this chapter, note the following. Various central limit theorems (Corollary 13.10) for random Fourier series can be established [M-P1] with applications of the techniques to the empirical characteristic function [Ma2]. A law of the iterated logarithm for the empirical characteristic function can also be proved [Led1], [La]. Gaussian and Rademacher random Fourier quadratic forms are studied and characterized in [L-M] with the results of Section 13.2 and 13.4. In particular, it is shown there how random Fourier quadratic forms with either Rademacher or standard Gaussian sequences converge simultaneously.

Chapter 14. Empirical process methods in Probability in Banach spaces

Chapter 14. Empirical process methods in Probability in Banach spaces

The purpose of this chapter is to present applications of the random process techniques developed so far to infinite dimensional limit theorems, and in particular the central limit theorem (CLT). More precisely, we will be interested for example in the CLT in the space $C(T)$ of continuous functions on a compact metric space T. Since $C(T)$ is not well behaved with respect to the type or cotype 2 properties, we will have rather to seek for nice classes of random variables in $C(T)$ for which a central limit property can be established. This point of view leads to enlarge this framework and to investigate limit theorems for empirical measures or processes. Random geometric descriptions of the CLT may then be produced through this approach as well as complete description for nice classes of functions (indicator functions of some sets) on which the empirical processes are indexed. While these random geometric descriptions do not solve the central limit problem in infinite dimension (and are probably of little use in applications), they however clearly describe the main difficulties inherent to the problem from the empirical point of view.

We do not try to give here a complete account on empirical processes and their limiting properties but rather concentrate on some useful methods and ideas related to the material already discussed in this book. The examples of techniques we chose to present are borrowed from the work by R. Dudley [Du4], [Du5] and E. Giné and J. Zinn [G-Z2], [G-Z3], and we actually refer the interested reader to these authors for a complete exposition. The first section of this chapter presents various results on the CLT for subgaussian and Lipschitz processes in $C(T)$ under metric entropy or majorizing measure conditions. In the second section, we introduce the language of empirical processes and discuss the effect of pregaussianness in two cases: the first one concerns uniformly bounded classes while the second provides a random geometric description of Donsker classes, i.e. classes for which the CLT holds. Vapnik-Chervonenkis classes of sets form the matter of Section 14.3 where it is shown how these classes satisfy the classical limit properties uniformly over all probability measures, and are actually characterized in this way.

14.1. The central limit theorem for Lipschitz processes

Let (T,d) be a compact metric space and denote by $C(T)$ the separable Banach space of all continuous functions on T equipped with the sup-norm $\|\cdot\|_\infty$. A Borel random variable X with values in $C(T)$ may be denoted in the processes notation as $X=(X_t)_{t\in T}=(X(t))_{t\in T}$ and $(X(t))_{t\in T}$ has all its sample paths continuous on (T,d). If X is a random variable, we denote as usual by (X_i) a sequence of independent copies of X and let $S_n=X_1+\cdots+X_n$, $n\geq 1$.

A subset K of $C(T)$ is relatively compact if and only if it is bounded and uniformly equicontinuous (Arzela-Ascoli). Equivalently, this is the case if there exists t_0 in T and a finite number M such that $|x(t_0)| \le M$ for all x in K and for all $\varepsilon > 0$ there exists $\eta = \eta(\varepsilon) > 0$ such that $|x(s) - x(t)| < \varepsilon$ for all x in K and s,t in T with $d(s,t) < \eta$. Combining with Prokhorov's Theorem 2.1 and the finite dimensional CLT, it follows that a random variable $X = (X(t))_{t\in T}$ satisfies the CLT in $C(T)$ if and only if $\mathbb{E}X(t) = 0$ and $\mathbb{E}X(t)^2 < \infty$ for all t and, for each $\varepsilon > 0$, there is $\eta = \eta(\varepsilon) > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{d(s,t)<\eta} \left| \frac{S_n(s) - S_n(t)}{\sqrt{n}} \right| > \varepsilon \right\} < \varepsilon . \tag{14.1}$$

Since the space $C(T)$ has no non-trivial type or cotype, and does not satisfy any kind of Rosenthal's inequality (cf. Chapter 10), the results that we can expect on the CLT in $C(T)$ can only concern special classes of random variables. We concentrate on the classes of subgaussian and Lipschitz variables, the first of which naturally extends the class of Gaussian variables (which trivially satisfy the CLT).

Recall that a centered process $X = (X(t))_{t\in T}$ is said to be subgaussian with respect to a metric d on T if for all real numbers λ and all s,t in T,

$$\mathbb{E}\exp \lambda(X(s) - X(t)) \le \exp \frac{\lambda^2}{2} d(s,t)^2 .$$

Changing if necessary d into a multiple of it, we may require equivalently that $\|X(s) - X(t)\|_{\psi_2} \le d(s,t)$ for all s,t in T (or $\mathbb{P}\{|X(s) - X(t)| > ud(s,t)\} \le C\exp(-u^2/C)$ for all $u > 0$ and some constant C). We have seen in Section 11.3 that if (T,d) satisfies the majorizing measure condition

$$\lim_{\eta\to 0} \sup_{t\in T} \int_0^\eta \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/2} d\varepsilon = 0$$

for some probability measure m on T, then the subgaussian process X has a version with almost all sample paths continuous on (T,d). It therefore defines (actually its version which we denote in the same way) a Radon random variable in $C(T)$. Note that by the main result of Chapter 12 and existence of majorizing measures for bounded and continuous processes, the preceding condition is (essentially) equivalent to the existence of a Gaussian random variable G in $C(T)$ such that $\|G(s) - G(t)\|_2 \ge d(s,t)$ for all s,t in T. Now, under one of these (equivalent) assumptions, it is easily seen that the subgaussian process X also satisfies the CLT in $C(T)$. Indeed, by independence and identical distribution of the summands, $S_n/\sqrt{n}$ is seen, for every n, to be also subgaussian with respect to d. Then, from Proposition 11.19, we deduce that for every $\varepsilon > 0$, one can find $\eta > 0$ depending only on ε, T, d such that, uniformly in n,

$$\mathbb{E} \sup_{d(s,t)<\eta} \left| \frac{S_n(s) - S_n(t)}{\sqrt{n}} \right| < \varepsilon .$$

Hence, X satisfies the CLT by (14.1). We have therefore the following result.

Theorem 14.1. Let X be a Borel random variable in $C(T)$ which is subgaussian with respect to d. Assume there is a probability measure m on (T, d) such that

$$\lim_{\eta \to 0} \sup_{t \in T} \int_0^\eta \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/2} d\varepsilon = 0 .$$

Then X satisfies the CLT.

We turn to the second class of random variables in $C(T)$ we will study here which are the Lipschitz random variables. They will be shown to be conditionally subgaussian and will therefore satisfy the CLT under conditions similar to the ones used for subgaussian variables. One first and main result is the following theorem.

Theorem 14.2. Let X be a Borel random variable in $C(T)$ such that $\mathbb{E}X(t) = 0$ and $\mathbb{E}X(t)^2 < \infty$ for all t in T. Assume there is a positive random variable M in L_2 such that for all ω and all s, t in T,

$$|X(\omega, s) - X(\omega, t)| \leq M(\omega) d(s,t) .$$

Then, if (T, d) satisfies the majorizing measure condition

$$\lim_{\eta \to 0} \sup_{t \in T} \int_0^\eta \left(\log \frac{1}{m(B(t,\varepsilon))} \right)^{1/2} d\varepsilon = 0$$

for some probability measure m on (T, d), X satisfies the CLT in $C(T)$.

Recall that we may assume equivalently (Theorem 12.9) that d is the L_2-pseudo-metric of a Gaussian random variable in $C(T)$. We would like to mention that the exposition of the proof of Theorem 14.2 we give is slightly more complicated than it should be. It should actually be similar to the proof of Theorem 14.5 below. We chose this exposition in order to include in the same pattern Theorem 14.3.

Proof. Let X, and (X_i), be defined on $(\Omega, \mathcal{A}, \mathbb{P})$. By Proposition 10.4, or a simple symmetrization argument, we may and do assume that X is symmetrically distributed. Thus, (X_i) has the same distribution as $(\varepsilon_i X_i)$ where (ε_i) is a Rademacher sequence constructed on some different probability space. There is further a sequence (M_i) of independent copies of M such that $|X_i(\omega,s) - X_i(\omega,t)| \le M_i(\omega)d(s,t)$ for all i, all ω, and all s,t in T. By the subgaussian inequality (4.1), for every ω, every integer n and every $u > 0$, and all s,t in T,

$$\begin{aligned}\mathbb{P}_\varepsilon\left\{\frac{1}{\sqrt{n}}\left|\sum_{i=1}^n \varepsilon_i(X_i(\omega,s) - X_i(\omega,t))\right| > u\right\} &\le 2\exp\left(-\frac{u^2}{\frac{2}{n}\sum_{i=1}^n |X_i(\omega,s)-X_i(\omega,t)|^2}\right)\\ &\le 2\exp\left(-\frac{u^2}{\frac{2}{n}d(s,t)^2\sum_{i=1}^n M_i(\omega)^2}\right),\end{aligned}$$

where $\mathbb{P}_\varepsilon$ is, as usual, integration with respect to (ε_i). Let then $a > 0$ to be specified and set for every integer n and every t in T,

$$Y^n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i X_i(t) I_{\left\{\sum_{j=1}^n M_j^2 \le a^2 n\right\}} .$$

From the preceding, it clearly follows that for all s,t in T, all n and $u > 0$,

$$\mathbb{P}\{|Y^n(s) - Y^n(t)| > a\,ud(s,t)\} \le 2\exp(-u^2/2) .$$

That is to say, for some numerical constant K, the processes $((Ka)^{-1}Y^n(t))_{t\in T}$ are subgaussian with respect to d. Therefore, under the majorizing measure condition of the theorem, we know from Proposition 11.19 that for all $\delta > 0$ there exists $\eta > 0$ depending only on δ, T, d, m such that, uniformly in n,

$$\mathbb{E}\sup_{d(s,t)<\eta}|Y^n(s) - Y^n(t)| < a\delta . \tag{14.2}$$

It is now easy to conclude the proof of Theorem 14.2. Fix $\varepsilon > 0$ and let $a = a(\varepsilon) > 0$ be such that $a^2 \ge 2\mathbb{E}M^2/\varepsilon$. Hence $\mathbb{P}\left\{\sum_{j=1}^n M_j^2 > a^2 n\right\} \le \varepsilon/2$ for all n. For all $\eta > 0$, we can write

$$\begin{aligned}\mathbb{P}\left\{\sup_{d(s,t)<\eta}\left|\frac{S_n(s)-S_n(t)}{\sqrt{n}}\right| > \varepsilon\right\} &\le \frac{\varepsilon}{2} + \mathbb{P}\{\sup_{d(s,t)<\eta}|Y^n(s) - Y^n(t)| > \varepsilon\}\\ &\le \frac{\varepsilon}{2} + \frac{1}{\varepsilon}\mathbb{E}\sup_{d(s,t)<\eta}|Y^n(s) - Y^n(t)| .\end{aligned}$$

If we then choose $\eta = \eta(\varepsilon) > 0$ small enough in order for (14.2) to be satisfied with $\delta = \varepsilon^2/2a$, we find that X satisfies (14.1) and therefore the CLT. The proof of Theorem 14.2 is complete.

Note that if the continuous majorizing measure condition in Theorem 14.2 is weakened into the corresponding bounded one, then we can only conclude in general to the bounded CLT for the Lipschitz variable X. That the continuous majorizing measure condition is necessary is made clear by the example of the random variable $X = (\varepsilon_n/(\log(n+1))^{1/2})$ on $C(\mathbb{N} \cup \{\infty\})$ which is Lipschitzian with respect to the distance of the bounded, but not continuous, Gaussian sequence $(g_n/(\log(n+1))^{1/2})$.

Although $C(T)$ is not of type 2, it is interesting to mention that Theorem 14.2 on Lipschitz random variables can be related to the general results on the CLT in type 2 spaces of Chapter 10. Actually, it rather concerns *operators* of type 2 and more precisely the canonical injection map $j : \mathrm{Lip}(T) \to C(T)$ investigated in Section 12.3. Recall we denote by $\mathrm{Lip}(T)$ the space of Lipschitz functions x on T equipped with the norm

$$\|x\|_{\mathrm{Lip}} = D^{-1}\|x\|_\infty + \sup_{s \neq t} \frac{|x(s) - x(t)|}{d(s,t)}$$

where $D = D(T)$ is the diameter of (T,d). We have seen in Theorem 12.17 that if there is a (bounded) majorizing measure on (T,d) for the function $(\log 1/u)^{1/2}$, then j is an operator of type 2 and that its type 2 constant $T_2(j)$ satisfies $T_2(j) \leq K\gamma^{(2)}(T,d)$ for some numerical constant K. Let now X be Lipschitzian with respect to d as in Theorem 14.2. Then $\mathbb{E}\|X\|^2_{\mathrm{Lip}} < \infty$ and since j is type 2, one might wish to use the CLT result for operators of type 2 (Corollary 10.6). There is however a small problem here since $\mathrm{Lip}(T)$ need *not* be separable and X not a Radon random variable in this space. This can be turned around in several ways. For example, from Proposition 9.11 for operators, we already have that, for every n,

$$\frac{1}{\sqrt{n}}\mathbb{E}\left\|\sum_{i=1}^n j(X_i)\right\|_\infty \leq 2T_2(j)(\mathbb{E}\|X\|^2_{\mathrm{Lip}})^{1/2}. \tag{14.3}$$

In particular, X already satisfies the bounded CLT in $C(T)$. Now, if there is a probability measure m on (T,d) such that,

$$\lim_{\eta\to 0}\sup_{t\in T}\int_0^\eta \left(\log\frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon = 0, \tag{14.4}$$

it is not difficult to see that the proof of Theorem 12.17 can be modified to show that for every $\varepsilon > 0$ there exists a finite dimensional subspace F of $C(T)$ such that if T_F is the quotient map $C(T) \to C(T)/F$, then

$T_2(T_F \circ j) < \varepsilon$. Applying (14.3) to $T_F \circ j$ then easily yields the CLT. In this last step however, this approach basically amounts to the original proof of Theorem 14.2. As an alternate, but also somewhat cumbersome argument, one can show that under (14.4) there exists a distance d' on T such that $d(s,t)/d'(s,t) \to 0$ when $d(s,t) \to 0$ and for which still $\gamma^{(2)}(T,d') < \infty$. Since the balls for the norm in $\mathrm{Lip}(T,d)$ are compact in $\mathrm{Lip}(T,d')$, the Lipschitz random variable X of Theorem 14.2 takes its values in some separable subspace of $\mathrm{Lip}(T,d')$. Corollary 10.6 can then be applied.

Since a random variable X satisfying the CLT in a Banach space B does not necessarily verify $\mathbb{E}\|X\|^2 < \infty$ but rather $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} = 0$ (cf. Lemma 10.1), it was conjectured for some time that the hypothesis M in L_2 in Theorem 14.1 could possibly be weakened into M in $L_{2,\infty}$, i.e. $\sup_{t>0} t^2\mathbb{P}\{M > t\} < \infty$. The next result shows how this is indeed the case. It is assumed explicitely that X is pregaussian since this does not follow anymore from the Lipschitz assumption when M is not in L_2. The proof relies on inequality (6.30) and Lemma 5.8.

Theorem 14.3. Let X be a pregaussian random variable in $C(T)$. Assume there is a positive random variable M in $L_{2,\infty}$ such that for all ω and all s,t in T

$$|X(\omega,s) - X(\omega,t)| \le M(\omega)d(s,t)\,.$$

Then, if there is a probability measure m on (T,d) such that

$$\lim_{\eta\to 0}\sup_{t\in T}\int_0^\eta \left(\log\frac{1}{m(B(t,\varepsilon))}\right)^{1/2} d\varepsilon = 0\,,$$

X satisfies the CLT in $C(T)$.

Proof. We first need transform the (necessary) pregaussian property into a majorizing measure condition. There exists a Gaussian variable in $C(T)$ with L_2-metric $d_X(s,t) = \|X_s - X_t\|_2$. By the comments next to Theorem 11.18, this Gaussian process is also continuous with respect to d_X and thus, by Theorem 12.9, there is a probability measure m' on (T,d_X) which satisfies the same majorizing measure condition as m on (T,d). We would like to have this property for the *maximum* of those two distances d and d_X. Clearly, $\mu = m \times m'$ on $T \times T$ equipped with the metric $\widetilde{d}((s,t),(s',t')) = \max(d(s,s'), d_X(t,t'))$ satisfies

$$\lim_{\eta\to 0}\sup_{T\times T}\int_0^\eta \left(\log\frac{1}{\mu(B((s,t),\varepsilon))}\right)^{1/2} d\varepsilon = 0\,.$$

We now simply project on the diagonal. For each couple (s,t) in $T \times T$, one can find (by a compactness argument) a point $\varphi(s,t)$ in T such that $\widetilde{d}((s,t),(\varphi(s,t),\varphi(s,t)) \leq 2\widetilde{d}((s,t),(u,u))$ for all u in T. Then, if $d(s,u)$ and $d_X(t,u)$ are both $< \varepsilon$, it follows by definition of $\varphi(s,t)$ that $d(\varphi(s,t),u)$ and $d_X(\varphi(s,t),u)$ are $< 3\varepsilon$. Hence $B((u,u),\varepsilon) \subset \varphi^{-1}(B(u,3\varepsilon))$ where $B(u,3\varepsilon)$ is the ball in T of center u and radius 3ε for the metric $\max(d,d_X)$. Therefore, letting $\widetilde{m} = \varphi(\mu)$, $\widetilde{m}(B(u,3\varepsilon)) \geq \mu(B((u,u),\varepsilon))$ and thus

$$\lim_{\eta\to\infty} \sup_{t\in T} \int_0^\eta \left(\log \frac{1}{\widetilde{m}(B(t,\varepsilon))}\right)^{1/2} d\varepsilon = 0\,.$$

It follows from this discussion that, replacing d by $\max(d,d_X)$, we may and do assume in the following that $d_X \leq d$.

We can now turn to the proof of Theorem 14.2. Instead of refering to (6.30), it is simpler, since we will only be concerned with real variables, to state and prove again the inequality we will need.

Lemma 14.4. Let (Z_i) be a finite sequence of independent real symmetric random variables in L_2. Then, for all $u > 0$,

$$\mathbb{P}\left\{\left|\sum_i Z_i\right| > u\left(4\|(Z_i)\|_{2,\infty} + \left(\sum_i \mathbb{E}Z_i^2\right)^{1/2}\right)\right\} \leq 4\exp(-u^2/6)\,.$$

Proof. Set $\sigma = \left(\sum\limits_i \mathbb{E}Z_i^2\right)^{1/2}$. For any random $A > 0$, we can write by the triangle inequality and definition of $\|(Z_i)\|_{2,\infty}$ that

$$\begin{aligned}\left|\sum_i Z_i\right| &\leq \left|\sum_i Z_i I_{\{|Z_i|\leq A\}}\right| + \sum_i |Z_i| I_{\{|Z_i|>A\}} \\ &\leq \left|\sum_i Z_i I_{\{|Z_i|\leq A\}}\right| + \frac{2}{A}\|(Z_i)\|_{2,\infty}^2\,.\end{aligned}$$

Therefore, if we let $A = u^{-1}\|(Z_i)\|_{2,\infty}$,

$$\mathbb{P}\left\{\left|\sum_i Z_i\right| > u(4\|(Z_i)\|_{2,\infty} + \sigma)\right\} \leq \mathbb{P}\left\{\left|\sum_i Z_i I_{\{|Z_i|\leq A\}}\right| > u(2Au+\sigma)\right\}\,.$$

Let us now observe that on the set $\{|Z_i| \leq A\}$

$$\frac{|Z_i|}{2Au+\sigma} \leq \frac{|Z_i|}{\sigma} \wedge \frac{1}{2u}\,.$$

Hence, by symmetry of the variables Z_i and the contraction principle in the form of (4.7) (applied conditionally on the Z_i 's),

$$\mathbb{P}\left\{\left|\sum_i Z_i\right| > u(4\|(Z_i)\|_{2,\infty} + \sigma)\right\} \le 2\mathbb{P}\left\{\left|\sum_i \varepsilon_i \left(\frac{|Z_i|}{\sigma} \wedge \frac{1}{2u}\right)\right| > u\right\}$$

where (ε_i) is an independent Rademacher sequence. We need now simply apply Kolmogorov's inequality (Lemma 1.6) to get the result. Lemma 14.4 is proved.

Provided with this lemma, the proof of Theorem 14.3 is very much like the proof of Theorem 14.2, substituing the inequality of Lemma 14.4 to the subgaussian inequality. Since M is in $L_{2,\infty}$, by Lemma 5.8, one can find, for each $\varepsilon > 0$, $a = a(\varepsilon) > 0$ such that, for every n,

$$\mathbb{P}\{\|(M_j)_{j\le n}\|_{2,\infty} > a\sqrt{n}\} \le \frac{\varepsilon}{2}.$$

Let, for all n and t in T,

$$Y^n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i X_i(t) I_{\{\|(M_j)_{j\le n}\|_{2,\infty} \le a\sqrt{n}\}}.$$

Lemma 14.4 implies that for all s,t in T and all $u > 0$,

$$\mathbb{P}\{|Y^n(s) - Y^n(t)| > (4a+1)ud(s,t)\} \le 4\exp(-u^2/6)$$

since $|X_i(s) - X_i(t)| \le M_i d(s,t)$ and $\|X_i(s) - X_i(t)\|_2 \le d(s,t)$ for all i and s,t in T. From this result, the proof of Theorem 14.3 is completed exactly as the proof of Theorem 14.2 by the subgaussian results (Proposition 11.19).

We conclude this section with an analogous study of some spectral measures of p-stable random vectors in $C(T)$. We already know the clos e relationships between the Gaussian CLT and the question of existence of a stable random vector with given spectral measure. The next example is another instance of this observation.

Given a positive finite Radon measure ν on $C(T)$, we would like to determine conditions under which ν is the spectral measure of some p-stable random variable in $C(T)$ with $1 \le p < 2$ (recall the case $p < 1$ is trivial, cf. Chapter 5). Since this seems a difficult task in general, we consider, as for the CLT, the particular case corresponding to Lipschitz processes. Assume for simplicity (and without any loss of generality) that ν is a probability measure so that it is the distribution of a random variable Y in $C(T)$.

Theorem 14.5. Let $1 \leq p < 2$ and $q = p/p-1$. Let $Y = (Y(t))_{t \in T}$ be a random variable in $C(T)$ such that $\mathbb{E}|Y(t)|^p < \infty$ for all t and such that for all ω and all s, t in T,

$$|Y(\omega, s) - Y(\omega, t)| \leq M(\omega) d(s, t)$$

for some positive random variable M in L_p. Assume there is a probability measure m on (T, d) such that

$$\lim_{\eta \to 0} \sup_{t \in T} \int_0^\eta \left(\log \frac{1}{m(B(t, \varepsilon))} \right)^{1/q} d\varepsilon = 0$$

(if $q < \infty$; if $q = \infty$, use the function $\log^+ \log$). Then, the distribution ν of Y is the spectral measure of a p-stable random variable with values in $C(T)$.

Proof. It is similar to the proof of Theorem 14.2. For notational convenience, we restrict ourselves to the case $q < \infty$. Recall the series representation of stable random vectors and processes (Corollary 5.3). Let (Y_j) be independent copies of Y, (ε_j) be a Rademacher sequence and assume as usual that (Γ_j), (ε_j), (Y_j) are independent. For each t, since $\mathbb{E}|Y(t)|^p < \infty$, the series $\sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j Y_j(t)$ is almost surely convergent (and defines a p-stable real random variable). It will be enough to show that for each $\varepsilon > 0$ one can find $\eta > 0$ such that

$$\mathbb{E} \sup_{d(s,t) < \eta} \left| \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j (Y_j(s) - Y_j(t)) \right| < \varepsilon . \tag{14.5}$$

The series $\sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j Y_j$ is then seen to be convergent almost surely and in L_1 in $C(T)$ (Itô-Nisio theorem). By Corollary 5.5, $c_p^{-1} \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j Y_j$ therefore defines there a p-stable random variable with spectral measure ν.

To establish (14.5), we first note that by independence, the contraction principle and (5.8),

$$\begin{aligned} \mathbb{E} \sup_{d(s,t) < \eta} \left| \sum_{j=1}^{\infty} \Gamma_j^{-1/p} \varepsilon_j (Y_j(s) - Y_j(t)) \right| &\leq \mathbb{E} \sup_{j \geq 1} \left(\frac{j}{\Gamma_j} \right)^{1/p} \mathbb{E} \sup_{d(s,t) < \eta} \left| \sum_{j=1}^{\infty} j^{-1/p} \varepsilon_j (Y_j(s) - Y_j(t)) \right| \\ &\leq K_p \mathbb{E} \sup_{d(s,t) < \eta} \left| \sum_{j=1}^{\infty} j^{-1/p} \varepsilon_j (Y_j(s) - Y_j(t)) \right| \end{aligned}$$

where K_p only depends on p. Using Lemma 1.7, for every ω on the space supporting (Y_j), and every s,t in T and $u>0$,

$$\mathbb{P}_\varepsilon\left\{\left|\sum_{j=1}^{\infty} j^{-1/p}\varepsilon_j(Y_j(\omega,s)-Y_j(\omega,t))\right|>u\right\}\leq 2\exp\left(-\frac{u^q}{C_q d(s,t)^q\|(j^{-1/p}M_j(\omega))\|_{p,\infty}^q}\right)$$

where (M_j) is a sequence of independent copies of M and where we have used that $|Y_j(\omega,s)-Y_j(\omega,t)|\leq M_j(\omega)d(s,t)$. Under the majorizing measure condition of the statement, we deduce from Theorem 11.14 that for each $\varepsilon>0$, one can find $\eta>0$ such that, uniformly in ω,

$$\mathbb{E}_\varepsilon \sup_{d(s,t)<\eta}\left|\sum_{j=1}^{\infty} j^{-1/p}\varepsilon_j(Y_j(\omega,s)-Y_j(\omega,t))\right|\leq \varepsilon\|(j^{-1/p}M_j(\omega))\|_{p,\infty}.$$

Integrating with respect to ω using Corollary 5.9 implies (14.5) and, as announced, the conclusion.

14.2. Empirical processes and random geometry

In this section, we examine the CLT through yet another angle, namely by empirical process methods. We actually only present a short overview of these empiral techniques with, in particular, a random geometric characterization of classes for which the central limit property holds. We refer to [Du5] and [G-Z3] for some of the basics of the theory as well as for a more detailed investigation.

We first introduce the empirical process language. Let $(S,\mathcal{S})$ be a measurable space. If P is a probability on $(S,\mathcal{S})$, (X_i) will denote here, unless otherwise indicated, a sequence of independent random variables defined on some probability space $(\Omega,\mathcal{A},\mathbb{P})$ with values in S and with common law P. We will also use randomizing sequences like Rademacher or standard Gaussian sequences (ε_i) or (g_i) and denote accordingly by $\mathbb{P}_\varepsilon$, $\mathbb{E}_\varepsilon$, $\mathbb{P}_g$, $\mathbb{E}_g$ partial integration with respect to (ε_i) or (g_i). The *empirical measures* P_n associated to P are defined as the random measures on S given by

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

(recall the X_i's have common law P).

In this section (and the next one), $L_p=L_p(P)$, $0\leq p\leq\infty$, is understood to be $L_p(S,\mathcal{S},P;\mathbb{R})$ (we write L_p or $L_p(P)$ depending on the context and the necessity of specifying the underlying probability P). $\|f\|_p$ denotes the L_p-norm ($1\leq p\leq\infty$) of the measurable function f on S, $d_p(f,g)=\|f-g\|_p$ its associated metric. If f is in $L_1 = L_1(P)$, we denote further $P(f) = E(f) = \int f dP$. We need also consider the random spaces $L_p(P_n)$, $1 \leq p < \infty$, with their norms

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

where f is a function on S, and denote by $d_{n,p}$ the associated random distances.

By class of functions on S, we will always mean here a family $\mathcal{F}$ of (real) measurable functions f on $(S, \mathcal{S})$ such that $\|f(x)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|f(x)| < \infty$ for all x in S. (For any family $(a(f))_{f\in\mathcal{F}}$ of numbers indexed by a class $\mathcal{F}$, we set, with some abuse, $\|a(f)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|a(f)|$.) Given P on $(S, \mathcal{S})$, the (centered) *empirical processes* based on P and indexed by a class $\mathcal{F} \subset L_1(P)$ are defined as

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

As always in this book, we do not enter the various and possibly intricated measurability questions that the study of empirical processes raises. In order not to hide the main ideas we intend to emphasize here, we shall assume all classes $\mathcal{F}$ to be *countable*. We could instead require a separability assumption on the processes $((P_n - P)(f))_{f\in\mathcal{F}}$.

Since we are assuming that $\|f(x)\|_{\mathcal{F}} < \infty$ for every x in S, the maps $f \to f(X_i)$, $i \in \mathbb{N}$, define random elements in the space $\ell_\infty(\mathcal{F})$ of bounded functions $\mathcal{F} \to \mathbb{R}$ equipped with the sup-norm $\|\cdot\|_{\mathcal{F}}$. In this study of empirical processes, we are therefore dealing with random variables taking their values in the non-separable (unless $\mathcal{F}$ is finite) Banach space $\ell_\infty(\mathcal{F})$ entering, since we are assuming $\mathcal{F}$ countable, our general setting of infinite dimensional random variables (cf. Section 2.3). Many results presented throughout this book therefore apply in this empirical setting.

Limit properties are of course the main topic in the study of empirical processes as a way to approximate a given law P by empirical data P_n. We have the following definitions. A class $\mathcal{F}$ as before is said to be a *Glivenko-Cantelli class for* P, or P satisfies the strong law of large numbers uniformly on $\mathcal{F}$, if, with probability one,

$$\lim_{n\to\infty}\|P_n(f) - P(f)\|_{\mathcal{F}} = 0.$$

This definition extends the classical result due to Glivenko and Cantelli according to which the class $\mathcal{F}$ of the indicator functions of the intervals $[0, t]$, $0 \leq t \leq 1$, is a Glivenko-Cantelli class for every probability P on

$[0,1]$. Since weak convergence is involved, the definition of the central limit property in this non-separable framework requires some more care. Write for convenience $\nu_n = \sqrt{n}(P_n - P)$, $n \in \mathbb{N}$. Then a class $\mathcal{F}$ of functions on S is said to be a *Donsker class for* P , or P satisfies the central limit theorem uniformly on $\mathcal{F}$, if there is a Gaussian Radon probability measure γ_P on $\ell_\infty(\mathcal{F})$ such that for every real bounded continuous function φ on $\ell_\infty(\mathcal{F})$,

$$\lim_{n\to\infty} \int^* \varphi(\nu_n) d\mathbb{P} = \int \varphi d\gamma_P \, .$$

The use of the upper integral takes into account the measurability questions. By the finite dimensional CLT, the probability measure γ_P is the law of a Gaussian process G_P indexed by $\mathcal{F}$ with covariance given by

$$\mathbb{E} G_P(f) G_P(g) = P(fg) - P(f)P(g) \, , \qquad f, g \in \mathcal{F} \, .$$

Further, γ_P being Radon on $\ell_\infty(\mathcal{F})$ is equivalent to say that G_P admits a version with almost all sample paths bounded and continuous on $\mathcal{F}$ with respect to the metric $\|(f - P(f)) - (g - P(g))\|_2$, $f, g \in \mathcal{F}$ (cf. [G-Z3]). If this property is realized the class $\mathcal{F}$ is said to be P*-pregaussian* so that a P-Donsker class is of course P-pregaussian. As before, these definitions extends the classical Kolmogorov-Smirnov-Donsker theorem for the class $\mathcal{F}$ of the indicator functions of the intervals $[0,t]$, $0 \le t \le 1$; the Gaussian process G_P appears as a generalization of the Brownian bridge (with P Lebesgue measure on $[0,1]$). We note for further purposes that if G_P is continuous in the previous sense and $\|P(f)\|_{\mathcal{F}} < \infty$ there exists a Gaussian process W_P with L_2-metric given by $\mathbb{E}|W_P(f) - W_P(g)|^2 = \|f - g\|_2^2 = d_2(f,g)^2$, f, g in $\mathcal{F}$ (the analog of the Brownian motion), which is almost surely continuous on $(\mathcal{F}, d_2)$. We may simply take for example $W_P(f) = G_P(f) + \theta P(f)$ where θ is a standard normal variable independent of G_P . To conclude finally this set of definitions, we should introduce Strassen classes satisfying the law of the iterated logarithm. Since we will basically only be concerned with the CLT here, we leave this to the interested reader (cf. e.g. [K-D], [Du5], [D-P]).

As for the CLT in the space of continuous functions (cf. the preceding section), a class $\mathcal{F}$ is a P-Donsker class if and only if the processes ν_n satisfy a Prokhorov type asymptotic equicontinuity condition. We refer to [Du4], [Du5], [G-Z2], [G-Z3] for complete description and proof of the following statement which extends (14.1) to this empirical framework. It is already expressed in its randomized version (cf. Proposition 10.4) which will be useful in the sequel. For every $\eta > 0$, we let $\mathcal{F}_\eta = \{f - g \, ; \; f, g \in \mathcal{F} \, , \; d_2(f,g) < \eta\}$.

Theorem 14.6. Let $\mathcal{F}$ be a class of functions on $(S, \mathcal{S}, P)$ such that $\|P(f)\|_{\mathcal{F}} < \infty$. Then $\mathcal{F}$ is a Donsker class for P if and only if $(\mathcal{F}, d_2)$ is totally bounded and for every $\varepsilon > 0$ there exists $\eta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_\eta} > \varepsilon \right\} < \varepsilon .$$

The equivalence holds similarly if the Rademacher sequence (ε_i) is replaced by an orthogaussian sequence (g_i). From the integrability properties in the CLT, in the form for example of Corollary 10.2 and (10.2), note that if $\mathcal{F}$ is a P-Donsker class we also have that

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\|_{\mathcal{F}_\eta} = 0 \tag{14.6}$$

and similarly with (g_i) in place of (ε_i).

Provided with these definitions and observations, we now turn to the two results on Donsker classes we intend to present. The first one describes the effect of pregaussianness on the equicontinuity condition of Theorem 14.6 for uniformly bounded classes of functions. It combines Sudakov's minoration with real exponential bounds. For every $\varepsilon > 0$ and integer n, $\mathcal{F}_{\varepsilon,n}$ denotes $\mathcal{F}_\eta$ for $\eta = (\varepsilon/\sqrt{n})^{1/2}$.

Theorem 14.7. Let $\mathcal{F}$ be a uniformly bounded class of functions on $(S, \mathcal{S}, P)$. Then, $\mathcal{F}$ is a P-Donsker class if and only if it is P-pregaussian and, for some (or all) $\varepsilon > 0$,

$$\left\| \sum_{i=1}^{n} \varepsilon_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_{\varepsilon,n}} \to 0 \ \text{ in probability} .$$

Proof. Assume without loss of generality that $\|f\|_\infty \le 1$ for all f in $\mathcal{F}$. Only sufficiency requires a proof. Let $\varepsilon > 0$ be fixed. Since $\mathcal{F}$ is P-pregaussian, we know that W_P is a Gaussian process which has a continuous version on $(\mathcal{F}, d_2)$. Therefore, by Sudakov's minoration (Corollary 3.19), $\lim_{n\to\infty} c_n(\varepsilon) = 0$ where

$$c_n(\varepsilon) = \left(\frac{\varepsilon}{\sqrt{n}}\right)^{1/2} \left(\log N\left(\mathcal{F}, d_2; \frac{1}{2}\left(\frac{\varepsilon}{\sqrt{n}}\right)^{1/2}\right)\right)^{1/2} .$$

By definition of the entropy numbers, there exists $\mathcal{G} = \mathcal{G}(\varepsilon, n)$ maximal in $\mathcal{F}$ with respect to the relations $d_2(f,g) \ge (\varepsilon/\sqrt{n})^{1/2}$ such that

$$\operatorname{Card}\mathcal{G} \le \exp(c_n(\varepsilon)^2 \sqrt{n}/\varepsilon) . \tag{14.7}$$

By maximality, for every f in $\mathcal{F}$ there exists g in $\mathcal{G}$ satisfying $d_2(f,g) < (\varepsilon/\sqrt{n})^{1/2}$. Therefore, for every $\eta > 0$, and n sufficiently large depending on η , we can write for all $\delta > 0$,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathcal{F}_\eta} > 3\delta\right\} \le 2\mathbb{P}\left\{\left\|\sum_{i=1}^{n}\varepsilon_i > f(X_i)/\sqrt{n}\right\|_{\mathcal{F}_{\varepsilon,n}} > \delta\right\} + \mathbb{P}\left\{\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathcal{G}_\eta} > \delta\right\}.$$

So, by hypothesis, it is enough to show that for all $\delta > 0$,

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\mathbb{P}\left\{\left\|\sum_{i=1}^{n}\varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathcal{G}_\eta} > \delta\right\} = 0. \tag{14.8}$$

Set now, for every n ,

$$A(\varepsilon,n) = \{\forall f \ne g \ \text{ in } \mathcal{G} = \mathcal{G}(\varepsilon,n)\,,\ d_{n,2}(f,g) \le 2d_2(f,g)\}$$

where we recall that $d_{n,2}(f,g)$ are the random distances $\left(\sum_{i=1}^{n}(f-g)^2(X_i)/n\right)^{1/2}$. Let $h = f-g$, $f \ne g$ in $\mathcal{G}$; then $\|h\|_\infty \le 2$ since $\mathcal{F}$ is uniformly bounded by 1 and $\|h\|_2 \ge (\varepsilon/\sqrt{n})^{1/2}$ by definition of $\mathcal{G}$. By Lemma 1.6, for all n large enough,

$$\begin{aligned}\mathbb{P}\{\|h\|_{n,2} > 2\|h\|_2\} &\le \mathbb{P}\left\{\sum_{i=1}^{n}(h^2(X_i) - \mathbb{E}h^2(X_i)) > 3n\|h\|_2^2\right\}\\ &\le \exp(-n\|h\|_2^2/50)\\ &\le \exp(-\varepsilon\sqrt{n}/50)\,.\end{aligned}$$

Hence, by (14.7),

$$\limsup_{n\to\infty}\mathbb{P}(A(\varepsilon,n)^c) \le \limsup_{n\to\infty}(\ \mathrm{Card}\mathcal{G}(\varepsilon,n))^2\exp(-\varepsilon\sqrt{n}/50) = 0\,. \tag{14.9}$$

For each n and ω in $A(\varepsilon,n)$, consider the Gaussian process

$$Z_{\omega,n}(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i f(X_i(\omega))\,, \qquad f \in \mathcal{G}\,.$$

Since $\omega \in A(\varepsilon, n)$, clearly $\mathbb{E}_g|Z_{\omega,n}(f) - Z_{\omega,n}(f')|^2 \leq 4d_2(f, f')$. Now W_P has d_2 as associated L_2 -metric and possesses a continuous version on $(\mathcal{F}, d_2)$. It then clearly follows, from Lemma 11.16 for example, that

$$\lim_{\eta \to 0} \mathbb{E}_g \|Z_{\omega,n}(f)\|_{\mathcal{G}_\eta} = 0$$

which therefore holds for all n and ω in $A(\varepsilon, n)$. Standard comparison of Rademacher averages to Gaussian averages combined with (14.9) then implies (14.8) and thus the conclusion. Theorem 14.7 is established.

The second result of this section investigates further the influence of pregaussianness in the study of Donsker classes $\mathcal{F}$ (no more necessarily uniformly bounded). While we only used Sudakov's minoration before, we now take advantage of existence of majorizing measures (Chapter 12). The result we present indicates rather precisely how the pregaussian property actually controls a whole "portion" of $\mathcal{F}$. In the remaining part, no cancellation (one of the main features of the study of sums of independent random variables) occurs. For clarity, we first give a quantitative rather than a qualitative statement. If P is a probability on $(S, \mathcal{S})$ and $\mathcal{F}$ a class of functions in $L_2 = L_2(P)$, recall the Gaussian process $W_P = (W_P(f))_{f \in \mathcal{F}}$. For classes of functions $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2$, we write $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ to signify that each f in $\mathcal{F}$ can be written as $f_1 + f_2$ where $f_1 \in \mathcal{F}_1$, $f_2 \in \mathcal{F}_2$.

Theorem 14.8. There is a numerical constant K with the following property: for every P -pregaussian class $\mathcal{F}$ such that $\|f\|_{\mathcal{F}} \in L_1 = L_1(P)$ and for every n , there exist classes $\mathcal{F}_1^n, \mathcal{F}_2^n$ in $L_2 = L_2(P)$ such that $\mathcal{F} \subset \mathcal{F}_1^n + \mathcal{F}_2^n$ and

(i) $$\mathbb{E}\left\|\sum_{i=1}^n |f(X_i)|/\sqrt{n}\right\|_{\mathcal{F}_1^n} \leq K\left(\mathbb{E}\|W_P(f)\|_{\mathcal{F}} + \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathcal{F}}\right),$$

(ii) $$\mathbb{E}\left\|\sum_{i=1}^n g_i f(X_i)/\sqrt{n}\right\|_{\mathcal{F}_2^n} \leq K\mathbb{E}\|W_P(f)\|_{\mathcal{F}}.$$

Proof. We may and do assume that $\mathcal{F}$ is a finite class. By Theorem 12.6, there exists an ultrametric distance $\delta \geq d_2$ on $\mathcal{F}$ and a probability measure m on $(\mathcal{F}, \delta)$ such that

(14.10) $$\sup_{f \in \mathcal{F}} \int_0^\infty \left(\log \frac{1}{m(B(f,\varepsilon))}\right)^{1/2} d\varepsilon \leq K\mathbb{E}\|W_P(f)\|_{\mathcal{F}}$$

where $B(f, \varepsilon)$ are the balls for the metric δ . K is further some numerical constant, possibly changing from line to line below, and eventually yielding the constant of the statement. We use (14.10) as in Proposition

11.10 and Remark 11.11. Denote by ℓ_0 the largest ℓ for which $2^{-\ell} \geq D$ where D is the d_2-diameter of $\mathcal{F}$. For every $\ell \geq \ell_0$, let $\mathcal{B}_\ell$ be the family of δ-balls of radius $2^{-\ell}$. For every f in $\mathcal{F}$, there is a unique element B of $\mathcal{B}_\ell$ with $f \in B$. Let then $\pi_\ell(f)$ be one fixed point of B and let $\mu_\ell(\{\pi_\ell(f)\}) = m(B)$. Let further $\mu = \sum_{\ell \geq \ell_0} 2^{-\ell+\ell_0+1}\mu_\ell$ which defines a probability measure. We note that $d_2(f, \pi_\ell(f)) \leq 2^{-\ell}$ for all f and ℓ and that $\pi_{\ell-1} \circ \pi_\ell = \pi_{\ell-1}$. From (14.10),

$$\sup_{f \in \mathcal{F}} \sum_{\ell \geq \ell_0} 2^{-\ell} \left(\log \frac{2^{\ell-\ell_0}}{\mu(\{\pi_\ell(f)\})} \right)^{1/2} \leq K \mathbb{E} \|W_P(f)\|_{\mathcal{F}} \tag{14.11}$$

(where we have used the definition of ℓ_0). Let now n be fixed (so that we do not specify it every time we should). For every f in $\mathcal{F}$ and $\ell > \ell_0$, set

$$a(f, \ell) = \sqrt{n} 2^{-\ell} \left(\log \frac{2^{\ell-\ell_0}}{\mu(\{\pi_\ell(f)\})} \right)^{-1/2} .$$

Given f in $\mathcal{F}$ and $x \in S$, let

$$\ell(x, f) = \sup_\ell \{\forall j \leq \ell\, ,\ |\pi_j(f)(x) - \pi_{j-1}(f)(x)| \leq a(f, j)\} .$$

Define then f_2 by $f_2(x) = \pi_{\ell(x,f)}(f)(x)$ and $f_1 = f - f_2$ and let $\mathcal{F}_1 = \mathcal{F}_1^n = \{f_1 \,;\ f \in \mathcal{F}\}$, $\mathcal{F}_2 = \mathcal{F}_2^n = \{f_2 \,;\ f \in \mathcal{F}\}$, with the obvious abuse in notation. The classes $\mathcal{F}_1^n$ and $\mathcal{F}_2^n$ are the classes of the expected decomposition and we thus would like to show that they satisfy (i) and (ii) respectively.

We start with (ii). Set $\mathcal{F}_2 - \mathcal{F}_2 = \{f_2 - f_2' \,;\ f_2, f_2' \in \mathcal{F}_2\}$ and $u = \mathbb{E}\|W_P(f)\|_{\mathcal{F}}$. We work with $\mathcal{F}_2 - \mathcal{F}_2$ rather than $\mathcal{F}_2$ since the process bounds of Chapter 11 are usually stated in this way. By definition of u, this will make no difference. We evaluate, for every $t > 0$ (or only $t \geq t_0$ large enough), the probability

$$\mathbb{P} \left\{ \mathbb{E}_g \left\| \sum_{i=1}^n g_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_2 - \mathcal{F}_2} > tu \right\} .$$

In a first step, let us show that this probability is less than $\mathbb{P}(A(t)^c)$ where, for K_2 to be specified later,

$$\begin{aligned} A(t) = \{&\forall \ell \geq \ell_0\, ,\ \forall f \in \mathcal{F}\, , \\ &\|(\pi_\ell(f) - \pi_{\ell-1}(f)) I_{\{|\pi_\ell(f) - \pi_{\ell-1}(f)| \leq a(f,\ell)\}}\|_{n,2} \leq K_2^{-1} t 2^{-\ell}\} \end{aligned}$$

(recall the random norms and distances $\|\cdot\|_{n,2}$, $d_{n,2}$). Let f, f' in $\mathcal{F}$ and denote by j the largest ℓ with $\pi_\ell(f) = \pi_\ell(f')$. Then $\ell(x, f) \geq j$ if and only if $\ell(x, f') \geq j$. That is, we can write for every x in S that

$$\begin{aligned} f_2(x) - f_2'(x) = {} &(\pi_{\ell(x,f)}(f)(x) - \pi_j(f)(x)) I_{\{\ell(\cdot, f) \geq j\}}(x) \\ &- (\pi_{\ell(x,f')}(f')(x) - \pi_j(f')(x)) I_{\{\ell(\cdot, f') \geq j\}}(x)) . \end{aligned}$$

It follows that

$$\begin{aligned} d_{n,2}(f_2, f_2') &= \|f_2 - f_2'\|_{n,2} \\ &\le \sum_{\ell \ge j} \|(\pi_\ell(f) - \pi_{\ell-1}(f)) I_{\{|\pi_\ell(f) - \pi_{\ell-1}(f)| \le a(f,\ell)\}}\|_{n,2} \\ &\quad + \sum_{\ell \ge j} \|(\pi_\ell(f') - \pi_{\ell-1}(f')) I_{\{|\pi_\ell(f') - \pi_{\ell-1}(f')| \le a(f',\ell)\}}\|_{n,2} \end{aligned}$$

and thus, on the set $A(t)$, for all f, f',

$$d_{n,2}(f_2, f_2') \le K_2^{-1} t 2^{-j+2} \le 8 K_2^{-1} t \delta(f, f')$$

(by definition of δ and j). From this property, it follows from the majorizing measure bound of Theorem 11.18 and (14.10) that, for all $t > 0$,

$$\mathbb{P}\left\{ \mathbb{E}_g \left\| \sum_{i=1}^n g_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_2 - \mathcal{F}_2} > tu \right\} \le \mathbb{P}(A(t)^c) \tag{14.12}$$

for K_2 well chosen from (14.10) and the constant of Theorem 11.18. We need therefore evaluate $\mathbb{P}(A(t)^c)$. To this aim, we use exponential inequalities in the form, for example, of Lemma 1.6. Note that $\|\pi_\ell(f) - \pi_{\ell-1}(f)\|_2 \le 3 \cdot 2^{-\ell}$. Recentering, we deduce from Lemma 1.6 that, for all f in $\mathcal{F}$, and $\ell > \ell_0$,

$$\mathbb{P}\{\|(\pi_\ell(f) - \pi_{\ell-1}(f)) I_{\{|\pi_\ell(f) - \pi_{\ell-1}(f)| \le a(f,\ell)\}}\|_{n,2} > K_2^{-1} t 2^{-\ell}\} \le \exp(-tn 2^{-2\ell} a(f,\ell)^{-2})$$

for all $t \ge t_1$ large enough (independent of n, f and ℓ). By definition of $a(f,\ell)$, this probability is estimated by

$$\exp\left(-t \log\left(\frac{2^{\ell-\ell_0}}{\mu(\{\pi_\ell(f)\})}\right)\right).$$

If $c \ge 2$, $\exp(-t \log c) \le (ct^2)^{-1}$ as soon as $t \ge t_2$ where t_2 is numerical. Therefore, if $t \ge t_0 = \max(t_1, t_2)$, we have obtained that

$$\mathbb{P}(A(t)^c) \le \frac{1}{t^2} \sum_{\ell > \ell_0} \sum_{\{\pi_\ell(f)\}} 2^{\ell-\ell_0} \mu(\{\pi_\ell(f)\}) \le \frac{1}{t^2}$$

where we have used that $\pi_{\ell-1} \circ \pi_\ell = \pi_{\ell-1}$ and that μ is a probability. By (14.12), integration by parts then yields

$$\mathbb{E} \left\| \sum_{i=1}^n g_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_2 - \mathcal{F}_2} \le \left(t_0 + \frac{1}{t_0}\right) u$$

from which (ii) immediately follows since, for any f in $\mathcal{F}$, $\|f_2\|_2 \le \|f\|_2 + 2^{-\ell_0} \le 5u$.

The main observation to establish (i) is the following: since $|\pi_{\ell(x,f)+1}(f)(x) - \pi_{\ell(x,f)}(f)(x)| > a(f,\ell+1)$, for every f_1 in $\mathcal{F}_1$,

$$\|f_1\|_1 = E|f| \le \sum_{\ell \ge \ell_0} E(|f - \pi_\ell(f)| I_{\{|\pi_{\ell+1}(f) - \pi_\ell(f)| > a(f,\ell+1)\}}) .$$

By Cauchy-Schwartz and Chebyshev's inequalities,

$$\begin{aligned}\|f_1\|_1 &\le \sum_{\ell \ge \ell_0} a(f,\ell+1)^{-1} \|f - \pi_\ell(f)\|_2 \|\pi_{\ell+1}(f) - \pi_\ell(f)\|_2 \\ &\le \sum_{\ell \ge \ell_0} a(f,\ell+1)^{-1} 3 \cdot 2^{-2\ell-1} \\ &\le K_1 \frac{u}{\sqrt{n}}\end{aligned}$$

for some numerical constant K_1, where the last inequality is (14.11) (recall that $u = \mathbb{E}\|W_P(f)\|_{\mathcal{F}}$). It is then easy to conclude. Set

$$v = \mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}} .$$

Since $\mathcal{F}_1 \subset \mathcal{F} - \mathcal{F}_2$, we already know from (ii) that

$$\mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_1} \le v + Ku .$$

From the comparison properties for Rademacher averages (Theorem 4.12),

$$\mathbb{E}\left\| \sum_{i=1}^n \varepsilon_i |f(X_i)|/\sqrt{n} \right\|_{\mathcal{F}_1} \le 2(v + Ku) ,$$

and further, by Lemma 6.3,

$$\mathbb{E}\left\| \sum_{i=1}^n (|f(X_i)| - \mathbb{E}|f(X_i)|)/\sqrt{n} \right\|_{\mathcal{F}_1} \le 4(v + Ku) .$$

Since $\|f_1\|_1 \le K_1 u n^{-1/2}$ for all f_1 in $\mathcal{F}_1$, (i) immediately follows. The proof of Theorem 14.8 is therefore complete.

It is worthwhile mentioning that the proof of Theorem 14.8 actually yields more than its statement. We have shown indeed that, with high probability, the class $\mathcal{F}_2^n$ equipped with the random distances $d_{n,2}$ is a Lipschitz image of $(\mathcal{F}, \delta)$ which is controlled by the pregaussian hypothesis. The class $\mathcal{F}_1^n$ is controlled in $L_1(P_n)$. In this sense, Theorem 14.8 may appear as a random geometric description of the central limit property. If $\mathcal{F}$ is P-Donsker, then $\mathcal{F}$ is decomposed in two classes, the first for which the random distances $d_{n,2}$ are controlled by the (necessary) P-pregaussian property, the second being controlled in the $\|\cdot\|_{n,1}$ random norms for which no cancellation occurs. Conversely, such a decomposition clearly contains the Donsker property. Note tha the levels of truncation chosen in the proof of Theorem 14.8 for this decomposition correspond, when $\mathcal{F}$ is reduced to one point, to the classical level $\sqrt{n}$.

To draw a possible qualitative version of Theorem 14.8, let us state the following (see also [Ta4], [G-Z3]). Recall $\mathcal{F}_\eta = \{f - g;\ f, g \in \mathcal{F},\ d_2(f,g) < \eta\}$.

Corollary 14.9. Let $\mathcal{F}$ be P-pregaussian. Then, for all $\eta > 0$ and every integer n, one can find classes $\mathcal{F}_1^n(\eta)$, $\mathcal{F}_2^n(\eta)$ in $L_2(P)$ such that $\mathcal{F}_\eta \subset \mathcal{F}_1^n(\eta) + \mathcal{F}_2^n(\eta)$ and

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \mathbb{E} \left\| \sum_{i=1}^n g_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_2^n(\eta)} = 0$$

and

$$\limsup_{\eta\to 0} \limsup_{n\to\infty} \mathbb{E} \left\| \sum_{i=1}^n |f(X_i)|/\sqrt{n} \right\|_{\mathcal{F}_1^n(\eta)} \le K \limsup_{\eta\to 0} \limsup_{n\to\infty} \mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i f(X_i)/\sqrt{n} \right\|_{\mathcal{F}_\eta}$$

(where K is numerical). In particular, $\mathcal{F}$ is P-Donsker if and only if

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \mathbb{E} \left\| \sum_{i=1}^n |f(X_i)|/\sqrt{n} \right\|_{\mathcal{F}_1^n(\eta)} = 0\,.$$

14.3. Vapnik-Chervonenkis classes of sets

While the previous section dealt with random characterizations of Donsker classes, this paragraph is devoted to the study of nice classes of indicator functions for which the classical limit theorems can be established. These classes of sets are the so-called Vapnik-Chervonenkis classes which naturally extend the case of the intervals $[0,t]$, $0 \le t \le 1$, on $[0,1]$. As we will see moreover, the limit properties of empirical processes indexed by Vapnik-Chervonenkis classes actually hold *uniformly* over all probability distributions.

Let S be a set and $\mathcal{C}$ be a class of subsets of S. Let A be a subset of S of cardinality k. Say that $\mathcal{C}$ *shatters* A if each subset of A is the trace of an element of $\mathcal{C}$, i.e. $\mathrm{Card}(\mathcal{C} \cap A) = 2^k$ where $\mathcal{C} \cap A = \{C \cap A\,;\ C \in \mathcal{C}\}$. Say that $\mathcal{C}$ is a *Vapnik-Chervonenkis class* (VC class in short) if there exists an integer $k \geq 1$ such that no subset A of S of cardinality k is shattered by $\mathcal{C}$, i.e. for all A in S with $\mathrm{Card}A = k$, we have $\mathrm{Card}(\mathcal{C} \cap A) < 2^k$. Denote by $v(\mathcal{C})$ the smallest k with this property. The class $\mathcal{C} = \{[0,t]\,;\ 0 \leq t \leq 1\}$ in $[0,1]$ is a VC class with $v(\mathcal{C}) = 2$. The following result is the most striking fact about VC classes.

Proposition 14.10. Let $\mathcal{C}$ be a VC class in S and let $v = v(\mathcal{C})$. Then, for any finite subset A of S,

$$\mathrm{Card}(\mathcal{C} \cap A) \leq \ \mathrm{Card}\{B \subset A\,;\ \ \mathrm{Card}B < v\}\,.$$

In particular, if $\mathrm{Card}A = n$ and $n \geq v$,

$$\mathrm{Card}(\mathcal{C} \cap A) \leq \left(\frac{en}{v}\right)^v\,.$$

The second part of the proposition follows from the fact that $\mathrm{Card}\{B \subset A;\ \ \mathrm{Card}B < v\} = \sum_{j<v} \binom{n}{j}$ and an easy estimate of the latter. It indicates in particular that we pass from the a priori information that $\mathrm{Card}(\mathcal{C} \cap A) < 2^n$ to a *polynomial* growth of this cardinality. This section is devoted to the applications to empirical processes indexed by VC classes of this basic result.

We may note that if $B \subset A$ is such that $\mathrm{Card}(\mathcal{C} \cap A) = 2^{\,\mathrm{Card}B}$, then $\mathrm{Card}B < v$. Proposition 14.10 therefore follows from the more general following result (by letting $\mathcal{U} = \mathcal{C} \cap A$) whose proof uses rearrangement techniques.

Proposition 14.11. Let A be a finite set and $\mathcal{U}$ a class of subsets of A. Then,

$$\mathrm{Card}\mathcal{U} \leq \ \mathrm{Card}\{B \subset A\,;\ B \ \text{ is shattered by } \mathcal{U}\}\,.$$

Proof. The idea is to find a simple operation (symmetrization) that will make $\mathcal{U}$ more regular while at the same time not decreasing the number of sets shattered by $\mathcal{U}$. One then applies this operation until the set $\mathcal{U}$ is so regular that the result is obvious. Given x in A, we define $T_x(\mathcal{U}) = \{T_x(U)\,;\ U \in \mathcal{U}\}$ where for U in $\mathcal{U}$, $T_x(U) = U\backslash\{x\}$ if $x \in U$ and $U\backslash\{x\} \notin \mathcal{U}$ and $T_x(U) = U$ otherwise. The first observation is that

$$\mathrm{Card}T_x(\mathcal{U}) = \ \mathrm{Card}\mathcal{U}\,. \tag{14.13}$$

To show this, it suffices to establish that T_x is one-to-one on $\mathcal{U}$. Suppose that $T_x(U_1) = T_x(U_2)$ for U_1, U_2 in $\mathcal{U}$. Then, by definition of the operation T_x, $U_1 \backslash \{x\} = U_2 \backslash \{x\}$. Since $U_1 = U_2$ when $x \in T_x(U_1)$, let us assume that $x \notin T_x(U_1)$ and $U_1 \neq U_2$ and proceed to a contradiction. Suppose for example that $x \in U_1$, $x \notin U_2$. Then, since $U_1 \backslash \{x\} = U_2 \in \mathcal{U}$, $T_x(U_1) = U_1$, so $x \in T_x(U_1)$ which is a contradiction. This shows (14.13). Let us now establish that if $T_x(\mathcal{U})$ shatters B, then $\mathcal{U}$ shatters B. If $x \notin B$, $T_x(\mathcal{U})$ and $\mathcal{U}$ have the same trace on B. If $x \in B$, for $B' \subset B \backslash \{x\}$ there is $T \in T_x(\mathcal{U})$ such that $T \cap B = B' \cup \{x\}$. T is of the form $T_x(U)$ for some U in $\mathcal{U}$. Since $x \in T$, both U and $U \backslash \{x\}$ belong to $\mathcal{U}$, so $\mathcal{U}$ shatters B. We can now conclude the proof of the proposition. Let $w(\mathcal{U}) = \sum_{U \in \mathcal{U}} \mathrm{Card} U$. Let $\mathcal{U}'$ such that $\mathcal{U}'$ is obtained from $\mathcal{U}$ by applications of some transformations T_x, and such that $w(\mathcal{U}')$ is minimal. Then, for each U in $\mathcal{U}'$, x in U, we must have $U \backslash \{x\} \in \mathcal{U}'$ for otherwise $w(T_x(\mathcal{U}')) < w(\mathcal{U}')$. This means that $\mathcal{U}'$ is hereditary in the sense that if $B' \subset B \in \mathcal{U}'$, then $B' \in \mathcal{U}'$. In particular, $\mathcal{U}'$ shatters each set it contains so that the result of the proposition is obvious for $\mathcal{U}'$. Since by (14.13), $\mathrm{Card}\mathcal{U}' = \mathrm{Card}\mathcal{U}$ and $\mathcal{U}$ shatters more sets than $\mathcal{U}'$, the proof is complete.

Let now $(S, \mathcal{S})$ be a measurable space and consider a class $\mathcal{C} \subset \mathcal{S}$. If Q is a probability measure on $(S, \mathcal{S})$, we let, for A, B in $\mathcal{S}$,

$$d_Q(A, B) = (Q(A \triangle B))^{1/2} = \|I_A - I_B\|_2$$

(where $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$ and where the norm $\|\cdot\|_2$ is understood with respect to Q). Recall the entropy numbers $N(\mathcal{C}, d_Q; \varepsilon)$. The next theorem is a consequence of Proposition 14.10 and appears as the fundamental property in the study of limit theorems for empirical processes indexed by VC classes.

Theorem 14.12. Let $\mathcal{C} \subset \mathcal{S}$ be a VC class. There is a numerical constant K such that for all probability measures Q on $(S, \mathcal{S})$ and all $0 < \varepsilon \leq 1$,

$$\log N(\mathcal{C}, d_Q; \varepsilon) \leq K v(\mathcal{C}) \left(1 + \log \frac{1}{\varepsilon}\right) .$$

Proof. Let Q and ε be fixed. Let N be such that $N(\mathcal{C}, d_Q; \varepsilon) \geq N$. There exist $A_1, \ldots, A_N$ in $\mathcal{C}$ such that $d_Q(A_k, A_\ell) \geq \varepsilon$ for all $k \neq \ell$. If (X_i) are independent random variables distributed according to Q, $\mathbb{P}\{X_i \notin A_k \triangle A_\ell\} \leq 1 - \varepsilon^2$ and thus

$$\mathbb{P}\{(A_k \triangle A_\ell) \cap \{X_1, \ldots, X_M\} = \emptyset\} \leq (1 - \varepsilon^2)^M$$

for all $k \neq \ell$ and integer M. Therefore, if M is chosen such that $N^2(1-\varepsilon^2)^M < 1$,

$$\mathbb{P}\{\exists k \neq \ell\,,\ (A_k \triangle A_\ell) \cap \{X_1, \dots, X_M\} \neq \emptyset\} > 0$$

and there exist therefore points $x_1, \dots, x_M$ in S such that, for $k \neq \ell$, $(A_k \triangle A_\ell) \cap \{x_1, \dots, x_M\}) \neq \emptyset$. Proposition 14.10 then indicates that, necessarily,

$$N \leq \left(\frac{eM}{v}\right)^v$$

where $v = v(\mathcal{C})$, at least when $M \geq v$. Take $M = [2\varepsilon^{-2}\log N] + 1$ so that $N^2(1-\varepsilon^2)^M < 1$. Assume first that $\log N \geq v$ (≥ 1) so that $v \leq M \leq 4\varepsilon^{-2}\log N$. Then the inequality $N \leq (eM/v)^v$ yields

$$\begin{aligned}\log N &\leq v\left(\log\frac{4e}{\varepsilon^2} + \log\left(\frac{\log N}{v}\right)\right)\\ &\leq v\log\frac{4e}{\varepsilon^2} + \frac{1}{e}\log N\end{aligned}$$

where we have used that $\log x \leq x/e$, $x > 0$. It follows that

$$\log N \leq 2v\log\frac{4e}{\varepsilon^2}\,,$$

an inequality which is also satisfied when $\log N \leq v$ since $\varepsilon \leq 1$. Since $N \leq N(\mathcal{C}, d_Q; \varepsilon)$ is arbitrary, Theorem 14.12 is established.

To agree with the notations and language of the previous section, we identify a class $\mathcal{C} \subset \mathcal{S}$ with the class $\mathcal{F}$ of indicator functions I_C, $C \in \mathcal{C}$ and thus write $\|\cdot\|_{\mathcal{C}}$ for $\|\cdot\|_{\mathcal{F}}$. We also assume that we deal only with countable classes $\mathcal{C}$ in order to avoid the usual measurability questions. This condition may be replaced by some appropriate separability assumption. The next statement is one of the main results on VC classes. It expresses that VC classes are Donsker for all underlying probability distributions.

Theorem 14.13. Let $\mathcal{C} \subset \mathcal{S}$ be a VC class. Then $\mathcal{C}$ is a Donsker class for every probability measure P on $(S, \mathcal{S})$.

Proof. Let P be a probability measure on $(S, \mathcal{S})$. As a first simple consequence of Theorem 14.12, $\mathcal{C}$ is totally bounded in $L_2 = L_2(P)$. By Theorem 14.6, it therefore suffices to show that

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \mathbb{P}\left\{\sup_{d_P(C,D) < \eta} \left|\sum_{i=1}^n \varepsilon_i (I_C - I_D)(X_i)/\sqrt{n}\right| > \varepsilon\right\} = 0 \tag{14.14}$$

for all $\varepsilon > 0$ (where, of course, C, D belong to $\mathcal{C}$). Recall the empirical measures $P_n(\omega) = \sum_{i=1}^{n} \delta_{X_i(\omega)}/n$, $\omega \in \Omega$, $n \in \mathbb{N}$. For every ω and n , the random (in the Rademacher sequence (ε_i)) process $\left(\sum_{i=1}^{n} \varepsilon_i I_C(X_i(\omega))/\sqrt{n}\right)_{C \in \mathcal{C}}$ is subgaussian with respect to $d_{P_n(\omega)}$. Since Theorem 14.12 actually implies that

$$\lim_{\eta \to 0} \sup_{Q} \int_0^{\eta} (\log N(\mathcal{C}, d_Q; \varepsilon))^{1/2} d\varepsilon = 0 \,,$$

it is plain that Proposition 11.19 yields a conclusion which is *uniform* over the distances $d_{P_n(\omega)}$ in the sense that for every $\varepsilon > 0$, there exists $\eta = \eta(\varepsilon) > 0$ such that *for all* n *and* ω

$$\mathbb{E}_\varepsilon \sup_{d_{P_n(\omega)}(C,D) < \eta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i (I_C - I_D)(X_i(\omega)) \right| < \varepsilon^2 \tag{14.15}$$

where, as usual, $\mathbb{E}_\varepsilon$ is partial integration with respect to (ε_i) . From this result, the proof will be completed if the random distances $d_{P_n(\omega)}$ can be replaced by d_P , at least on a large enough set of ω 's. In the same way as we have (14.15), if $\mathcal{C}$ is VC,

$$\sup_n \frac{1}{\sqrt{n}} \mathbb{E} \left\| \sum_{i=1}^{n} \varepsilon_i I_C(X_i) \right\|_{\mathcal{C}} < \infty \,.$$

By Lemma 6.3 (actually the subsequent comment), we also have that

$$\sup_n \frac{1}{\sqrt{n}} \mathbb{E} \left\| \sum_{i=1}^{n} (I_C(X_i) - P(C)) \right\|_{\mathcal{C}} < \infty \,.$$

The same property holds for $\mathcal{C} \triangle \mathcal{C} = \{C \triangle C' \,;\ C, C' \in \mathcal{C}\}$ since it is also VC by Proposition 14.10. Hence, given $\varepsilon > 0$ and $\eta = \eta(\varepsilon) > 0$ so that (14.15) is satisfied, there exists n_0 such that $\mathbb{P}(A(n)) \leq \varepsilon$ for all $n \geq n_0$ where

$$A(n) = \{\|(P_n - P)(C \triangle C')\|_{\mathcal{C} \triangle \mathcal{C}} > \eta^2/4\} \,.$$

We can then easily conclude. For all $n \geq n_0$, using (14.15),

$$\begin{aligned}
&\mathbb{P} \left\{ \sup_{d_P(C,D) < \eta/2} \left| \sum_{i=1}^{n} \varepsilon_i (I_C - I_D)(X_i)/\sqrt{n} \right| > \varepsilon \right\} \\
&\leq \varepsilon + \mathbb{P} \left\{ A(n)^c \,;\ \sup_{d_{P_n}(C,D) < \eta} \left| \sum_{i=1}^{n} \varepsilon_i (I_C - I_D)(X_i)/\sqrt{n} \right| > \varepsilon \right\} \\
&\leq 2\varepsilon \,.
\end{aligned}$$

This gives (14.14) so that Theorem 14.13 is established.

The preceding proof based on the key Theorem 14.12 actually carries more information than the actual statement of Theorem 14.13. It indeed indicates a uniformity property over all probability measures P. With the same argument leading to (14.15) through Theorem 14.12, and as we actually used it in the proof, there is a numerical constant K such that if $\mathcal{C}$ is VC,

$$\sup_n \mathbb{E}\|\nu_n(C)\|_{\mathcal{C}} \le Kv(\mathcal{C})^{1/2} \tag{14.16}$$

for all probability measures P on $(S,\mathcal{S})$ where we recall that $\nu_n = \sqrt{n}(P_n - P)$. This property implies a uniform strong law of large numbers in the sense that for some sequence (a_n) of positive numbers decreasing to 0 and any P

$$\sup_n \frac{1}{a_n}\|(P_n - P)(C)\|_{\mathcal{C}} < \infty \qquad \text{almost surely}.$$

This may be obtained for example from Theorem 8.6 (or the "bounded" version of Theorem 10.12) which yield the best possible $a_n = (LLn/n)^{1/2}$. One may also invoke the SLLN result in the form of Theorem 7.9 for example. (Recall that since we are dealing with indicators, the corresponding random variables in $\ell_\infty(\mathcal{C})$ are uniformly bounded.)

It is remarkable that these uniform limit properties actually characterize VC classes. Following (14.16), suppose for example that we are given a class $\mathcal{C}$ in $\mathcal{S}$ such that for some finite M,

$$\sup_n \mathbb{E}\|\nu_n(C)\|_{\mathcal{C}} \le M \tag{14.17}$$

for all probability measures P on $(S,\mathcal{S})$. If we then recall the Gaussian processes G_P with covariance $P(C\cap D) - P(C)P(D)$, $C, D \in \mathcal{C}$, introduced in Section 14.2, we also have, by the finite dimensional CLT, that

$$\mathbb{E}\|G_P(C)\|_{\mathcal{C}} \le M \tag{14.18}$$

for all P. In particular, if $P = \frac{1}{k}\sum_{i=1}^{k}\delta_{x_i}$ where $x_1,\dots,x_k$ are points in S, G_P can be realized as

$$G_P(C) = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} g_i(\delta_{x_i}(C) - P(C)), \qquad C \in \mathcal{C},$$

where (g_i) is a standard normal sequence. Therefore, from (14.18),

$$\mathbb{E}\left\|\sum_{i=1}^{k} g_i \delta_{x_i}(C)\right\|_{\mathcal{C}} \leq (M+1)\sqrt{k}\,. \tag{14.19}$$

Suppose now that $\mathcal{C}$ is not a VC class. Then, for every k, there exists a subset $A = \{x_1, \ldots, x_k\}$ of S of cardinality k such that $\mathrm{Card}(\mathcal{C} \cap A) = 2^k$. Therefore, as is obvious, for all $\alpha = (\alpha_1, \ldots, \alpha_k)$ in $\mathbb{R}^k$

$$\sum_{i=1}^{k} |\alpha_i| \leq 2\left\|\sum_{i=1}^{k} \alpha_i \delta_{x_i}(C)\right\|_{\mathcal{C}}\,. \tag{14.20}$$

Integrating this inequality along the orthogaussian sequence (g_i) leads to a contradiction with (14.19) when k is large enough. Therefore $\mathcal{C}$ is necessarily a VC class under (14.18) and thus a fortiori under (14.17). (By an argument close in spirit to the closed graph theorem as in Proposition 14.14 below, it is actually enough to have that $\mathcal{C}$ is P-pregaussian for every P.)

The next proposition strengthens this conclusion, showing in particular that the conclusion is not restricted to the CLT.

Proposition 14.14. Let $\mathcal{C}$ be a (countable) class in $\mathcal{S}$. Assume there exists a decreasing sequence of positive numbers (a_n) tending to 0 such that for every probability P on $(S, \mathcal{S})$, the sequence $(\|(P_n - P)(C)\|_{\mathcal{C}}/a_n)$ is bounded in probability. Then $\mathcal{C}$ is a VC class.

Proof. We may assume that $\sup_n (a_n\sqrt{n})^{-1} < \infty$. By Hoffmann-Jørgensen's inequalities (Proposition 6.8),

$$\sup_n \frac{1}{a_n}\mathbb{E}\|(P_n - P)(C)\|_{\mathcal{C}} < \infty\,.$$

From Lemma 6.3, we see that, for all n,

$$\mathbb{E}\left\|\sum_{i=1}^{n} \varepsilon_i (I_C(X_i) - P(C))\right\|_{\mathcal{C}} \leq 2\mathbb{E}\left\|\sum_{i=1}^{n} (I_C(X_i) - P(C))\right\|_{\mathcal{C}}\,.$$

Therefore, if we set

$$\Phi(P) = \sup_n \frac{1}{na_n}\mathbb{E}\left\|\sum_{i=1}^{n} \varepsilon_i I_C(X_i)\right\|_{\mathcal{C}}\,,$$

we clearly get that

$$\Phi(P) \leq \sup_n \frac{1}{\sqrt{n}a_n} + 2\sup_n \frac{1}{a_n}\mathbb{E}\|(P_n - P)(C)\|_{\mathcal{C}}\,.$$

Hence $\Phi(P) < \infty$ for all P. We would like to show actually that there is some constant M such that

(14.21)
$$\Phi(P) \leq M \qquad \text{for all probabilities } P\,.$$

To this aim, let us first show that if P^0 and P^1 are two probability measures on $(S, \mathcal{S})$, and if $P = \alpha P^0 + (1-\alpha)P^1$, $0 \leq \alpha \leq 1$, then $\Phi(P) \geq \alpha\Phi(P^0)$. Let (X_i^0) (resp. (X_i^1)) be independent random variables with common distribution P^0 (resp. P^1). Let further (δ_i) be independent of everything that was introduced before and consisting of independent random variables with law $\mathbb{P}\{\delta_i = 0\} = 1 - \mathbb{P}\{\delta_i = 1\} = \alpha$. Then (X_i) has the same distribution as $(X_i^{\delta_i})$. In particular, by the contraction principle,

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i I_C(X_i)\right\|_{\mathcal{C}} \geq \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i I_C(X_i^0) I_{\{\delta_i = 0\}}\right\|_{\mathcal{C}}.$$

Jensen's inequality on (δ_i) yields then

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i I_C(X_i)\right\|_{\mathcal{C}} \geq \alpha\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i I_C(X_i^0)\right\|_{\mathcal{C}}$$

and thus the announced inequality $\Phi(P) \geq \alpha\Phi(P^0)$. This observation easily implies (14.21). Indeed, if it is not realized, one can find a sequence (P^k) of probabilities on $(S, \mathcal{S})$ such that $\Phi(P^k) \geq 4^k$ for all k. If we then let $P = \sum_{k=1}^{\infty} 2^{-k}P^k$, $\Phi(P) \geq 2^{-k}\Phi(P^k) \geq 2^k$ for all k, contradicting $\Phi(P) < \infty$.

We can now conclude the proof of Proposition 14.14. If $\mathcal{C}$ is not VC, there exists, for all k, $A = \{x_1, \ldots, x_k\}$ in S such that (14.20) holds. Take then $P = \frac{1}{k}\sum_{i=1}^k \delta_{x_i}$. Fix then n large enough so that $n > 4Mna_n$, which is possible since $a_n \to 0$. Consider $k > 2n^2$ in order that $\mathbb{P}(\Omega_0) \geq 1/2$ where

$$\Omega_0 = \{\forall i \neq j \leq n\,,\ X_i \neq X_j\}\,.$$

Then, since $\Phi(P) \leq M$, we can write by (14.20) that

$$\begin{aligned} Mna_n &\geq \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i I_C(X_i)\right\|_{\mathcal{C}} \\ &\geq \int_{\Omega_0}\left\|\sum_{i=1}^n \varepsilon_i I_C(X_i)\right\|_{\mathcal{C}} \geq \frac{n}{2}\mathbb{P}(\Omega_0) \geq \frac{n}{4} \end{aligned}$$

which is a contradiction. The proof of Proposition 14.14 is complete.

As in Section 14.1, the nice limit properties of empirical processes indexed by VC classes may be related to the type property of a certain operator between Banach spaces. Denote by $M(S,\mathcal{S})$ the Banach space of bounded measures μ on $(S,\mathcal{S})$ equipped with the norm $\|\mu\| = |\mu|(S)$. Let $\mathcal{C} \subset \mathcal{S}$; consider the operator $j : M(S,\mathcal{S}) \to \ell_\infty(\mathcal{C})$ defined by $j(\mu) = (\mu(C))_{C\in\mathcal{C}}$. We denote by $T_2(j)$ the type 2 constant of j , that is the smallest constant C such that for all finite sequences (μ_i) in $M(S,\mathcal{S})$,

$$\mathbb{E}\left\|\sum_i \varepsilon_i j(\mu_i)\right\|_{\mathcal{C}} \le C\left(\sum_i \|\mu_i\|^2\right)^{1/2}$$

(provided there exists one). We have the following result.

Therorem 14.15. In the preceding notations, $\mathcal{C}$ is a VC class if and only if j is an operator of type 2 . Moreover, for some numerical constant K ,

$$K^{-1}(v(\mathcal{C}))^{1/2} \le T_2(j) \le K(v(\mathcal{C}))^{1/2}.$$

Proof. We establish that $T_2(j) \le K(v(\mathcal{C}))^{1/2}$. Let (μ_i) be a finite sequence in $M(S,\mathcal{S})$. To prove the type 2 inequality, we may assume that the measures μ_i are positive and, by homogeneity, that $\sum_i \|\mu_i\|^2 = 1$. Set $Q = \sum_i \|\mu_i\|\mu_i$. Then Q is a probability measure on $(S,\mathcal{S})$ and we have clearly that

$$\left(\sum_i |\mu_i(C) - \mu_i(D)|^2\right)^{1/2} \le d_Q(C,D)$$

for all C,D . Therefore, the entropy version of inequality (11.19) together with Theorem 14.12 applied to the process $\left(\sum_i \varepsilon_i\mu_i(C)\right)_{C\in\mathcal{C}}$ where $\mathcal{C} \subset \mathcal{S}$ is VC yields

$$\begin{aligned}\mathbb{E}\left\|\sum_i \varepsilon_i j(\mu_i)\right\|_{\mathcal{C}} &\le 1 + K\int_0^\infty (\log N(\mathcal{C}, d_Q; \varepsilon))^{1/2} d\varepsilon \\ &\le 1 + K'(v(\mathcal{C}))^{1/2}\end{aligned}$$

where K, K' are numerical constants. Since $v(\mathcal{C}) \ge 1$, the right side inequality of the theorem follows. To prove the converse inequality, set $v = v(\mathcal{C})$. Since $T_2(j) \ge 1$, we can assume that $v \ge 2$. By definition of v , there exists $A = \{x_1, \ldots, x_{v-1}\}$ in S such that $\mathrm{Card}(\mathcal{C} \cap A) = 2^{v-1}$. Then

$$\mathbb{E}\left\|\sum_{i=1}^{v-1} \varepsilon_i j(\delta_{x_i})\right\|_{\mathcal{C}} \le T_2(j)(v-1)^{1/2},$$

and, by (14.20),

$$\frac{1}{2}(v-1) \le T_2(j)(v-1)^{1/2}\,.$$

Therefore $T_2(j) \ge (v-1)^{1/2}/2 \ge v^{1/2}/4$ which completes the proof of Theorem 14.15.

From this result together with the limit theorems for sums of independent random variables with values in Banach spaces with some type (or for *operators* with some type) (cf. Chapter 7, 8, 9, 10), one can essentially deduce again Theorem 14.13 and the consequent strong limit theorems for empirical processes indexed by VC classes. Moreover, one can note that $\mathcal{C}$ is actually a VC class as soon as the associated operator j has *some* type $p > 1$, simply because if this is realized, we are in a position to apply Proposition 14.14. Finally, property (14.20) makes it clear that the notion of VC class is related to ℓ_1^n spaces. the following statement is the Banach space theory formulation of the previous investigation.

Theorem 14.16. Let $x_1, \ldots, x_n$ be functions on some set T taking the values ± 1. Let $r(T) = \mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^{n}\varepsilon_i x_i(t)\right|$. There exists a numerical constant K such that for all k such that $k \le r(T)^2/Kn$, one can find $m_1 < m_2 < \cdots < m_k$ in $\{1,\ldots,n\}$ such that the set of values $\{x_{m_1}(t),\ldots,x_{m_k}(t)\}$, $t \in T$, is exactly $\{-1,+1\}^k$. In other words, the subsequence $x_{m_1},\ldots,x_{m_k}$ generates in $\ell_\infty(T)$ a subspace isometric to ℓ_1^k.

Proof. Let

$$M = \mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^{n}\varepsilon_i\left(\frac{1+x_i(t)}{2}\right)\right|$$

and consider the class $\mathcal{C}$ of subsets of $\{1,2,\ldots,n\}$ of the form $\{i \in \{1,\ldots,n\}\,;\ x_i(t) = 1\}$, $t \in T$. Theorem 14.15 applied to this class yields

$$M \le T_2(j)\sqrt{n} \le K(nv(\mathcal{C}))^{1/2}\,.$$

Therefore, by definition of $v(\mathcal{C})$, the conclusion of the theorem is fulfilled for all $k < M^2/K^2n$. Note that $r(T) \le 2M + \sqrt{n}$. Since we may assume $M \ge \sqrt{n}$ (otherwise there is nothing to prove), we see that when $k \le r(T)^2/K'n$ for some large enough K', we have that $k < M^2/K^2n$ in which case we already know the conclusion holds.

Notes and references

As announced, this chapter only presents a few examples of the empirical process methods and their applications. Our framework is essentially the one put forward in the work of E. Giné and J. Zinn [G-Z2], [G-Z3] itself initiated by R. Dudley [Du4]. Some general references on empirical processes are the notes and books [Ga], [Du5], [Pol], [G-Z3]. The interested reader will complete appropriately this chapter and this short discussion with those references and the papers cited there.

The central limit theorems for subgaussian and Lipschitz processes (Theorems 14.1 and 14.2) are due to N. C. Jain and M. B. Marcus [J-M1]. R. Dudley and V. Strassen [D-S] introduced entropy in this study and established Theorem 14.2 with M uniformly bounded. Another prior partial result to Theorem 14.2 may be found in [Gin] (where the technique of proof is actually close to the nowadays bracketing arguments - see below). These authors worked under the metric entropy condition; the majorizing measure version of these results was obtained in [He1]. In [Zi1], J. Zinn connects the Jain-Marcus CLT for Lipschitz processes with the type 2 property of the operator $j:\ \mathrm{Lip}(T) \to C(T)$. The analog of theorem 14.2 for random variables in c_0 is studied in [Pau], [He4], [A-G-O-Z], [Ma-Pl]. A stronger version of Theorem 14.3 is established in [A-G-O-Z] where it is shown that the conclusion actually holds for local Lipschitz processes, namely processes X in $C(T)$ such that for all t in T and $\varepsilon > 0$,

$$\| \sup_{d(s,t)<\varepsilon} |X_s - X_t| \, \|_{2,\infty} \leq \varepsilon \, .$$

The proof of this result relies on bracketing techniques in the context of empirical processes. (The arguments of proof are related to Theorem 14.8 and the truncations used there; a similar decomposition is developed from which the local Lipschitz conditions provides the control of the corresponding $L_1(P_n)$-portion.) Bracketing was initiated in [Du4]; further results are obtained in [Oss], [A-G-Z], [L-T4]. In those last two articles, analogous results for the LIL are discussed, improving upon [Led1]. The simple proof of the (weaker) Theorem 14.3 and the inequality of Lemma 14.4 are due to B. Heinkel [He7] (see also [He8]). Theorem 14.5 is taken from [M-P3].

The equicontinuity criterion for Donsker classes is due to R. Dudley [Du4]; its randomized version (as stated here as Theorem 14.6) is taken from [G-Z2]. E. Giné and J. Zinn made clever use of the Gaussian randomization and the Gaussian process techniques together with exponential bounds to achieve remarkable progress in the understanding of the Donsker property. Theorem 14.7 is theirs [G-Z2]. Theorem 14.8 and the random geometric description of Donsker classes have been obtained in [Ta4], motivated by the investigation and prior results in [G-Z2]. We refer to [G-Z3] for an alternate exposition and more details. Further statements in this spirit appear in [L-T4]. Donsker classes of sets are investigated in [G-Z2], [G-Z3], [Ta6] (see also [V-C2]).

Vapnik-Chervonenkis classes of sets were introduced in [V-C1] and were shown there to satisfy uniform laws of large numbers. CLT, LIL and invariance principle for VC classes have been established respectively in [Du4], [K-D], [D-P]. Our exposition of this section is based on the observations by G. Pisier [Pi14]. Proposition 14.10 was established, independently, in [V-C1], [Sa], [Sh]. The proof based on Proposition 14.11 seems to be due to P. Frankl [Fr]. We learned it from V. D. Milman. The main Theorem 14.12 on the uniform entropy control of VC classes has been observed by R. Dudley [Du4] to whom Theorem 14.13 is due. Th at VC classes are actually characterized by uniform limit properties of the empirical measures was noticed in a particular case (for the Donsker property) in [D-D] and completely understood via the map j and Theorem 14.15 by G. Pisier [Pi14]. Theorem 14.16 is also taken from [Pi14]. Universal Donsker classes of functions do not seem to have similar nice descriptions; for some results connected with type 2 map, cf. [Zi4]. Let us also mention to conclude the extension of the VC definition to classes of functions (VC graphs) with in particular a nice characterization of the Donsker property [Ale2] in the spirit of the best possible conditions for the CLT in Banach spaces (cf. Chapter 10). See [A-T] for the corresponding LIL result.

Chapter 15. Applications to Banach space theory

Chapter 15. Applications to Banach space theory

This last chapter emphasizes some applications of isoperimetric methods and process techniques of Probability in Banach spaces to local theory of Banach spaces. The applications we present are only a sample of some of the recent developments in local theory of Banach spaces (and we refer to the lists of references and seminars and proceedings for further main examples in the historical developments). They demonstrate the power of probabilistic ideas in this context. This chapter is organized along its subtitles of rather independent context. Several questions and conjectures are presented in addition, some with details as in Sections 15.2 and 15.6, the others in the last paragraph on miscellaneous problems.

15.1. Subspaces of small codimension

Before turning to the object of this first section, it is convenient to briefly present a covering lemma in the spirit of Lemma 9.5 which will be of help here. We denote by $B_2 = B_2^N$ the Euclidean unit ball of $\mathbb{R}^N$.

Lemma 15.1. There exists a subset H of $2B_2^N$ of cardinality at most 5^N such that $B_2^N \subset \mathrm{Conv}H$.

Proof. It is similar to the proof of Lemma 9.5. Let N be fixed. For $0 < \delta < 1$, let $\widetilde{H}$ be maximal in B_2 such that $|x-y| > \delta$ for all x, y in $\widetilde{H}$. Then the balls of radius $\delta/2$ with centers in $\widetilde{H}$ are disjoint and contained in $(1+\delta/2)B_2$. Comparing the volumes yields

$$\mathrm{Card}\widetilde{H}\left(\frac{\delta}{2}\right)^N \ \mathrm{vol}B_2 \le \left(1+\frac{\delta}{2}\right)^N \ \mathrm{vol}B_2$$

so that $\mathrm{Card}\widetilde{H} \le (1+2/\delta)^N$. By maximality of $\widetilde{H}$, it is easily seen that each x in B_2 can be written as

$$x = \sum_{k=0}^{\infty} \delta^k h_k$$

where $(h_k) \subset \widetilde{H}$. Take then $\delta = 1/2$, $H = 2\widetilde{H}$ and the lemma follows.

As a consequence of this lemma, note that $\widetilde{H} = \frac{1}{2}H$ of the preceding proof is such that $\widetilde{H} \subset B_2^N$, $\mathrm{Card}\widetilde{H} \le 5^N$ and

$$|x| \le 2\sup_{h\in\widetilde{H}} |\langle x, h\rangle|$$

for all x in $\mathbb{R}^N$.

Let T be a convex body of $\mathbb{R}^N$, that is T is a compact convex symmetric (about the origin) subset of $\mathbb{R}^N$ with nonempty interior. As usual, denote by (g_i) an orthonormal Gaussian sequence. We let (as in Chapter 3)

$$\ell(T) = \mathbb{E} \sup_{t \in T} \left| \sum_{i=1}^{N} g_i t_i \right| = \int_{\mathbb{R}^N} \sup_{t \in T} |\langle x, t \rangle| d\gamma_N(x)$$

where $t = (t_1, \ldots, t_N)$ in $\mathbb{R}^N$ and γ_N is the canonical Gaussian measure on $\mathbb{R}^N$. The result of this section describes a way of finding subspaces F of $\mathbb{R}^N$ whose intersection with T has a small diameter as soon as the codimension of F is large with respect to $\ell(T)$. This result is one of the main steps in the proof of the remarkable quotient of a subspace theorem of V.D. Milman [Mi2]. As for Dvoretzky's theorem, we present two proofs of the results, both of them based on Gaussian variables. The first one uses isoperimetry and Sudakov's minoration, the second one the Gaussian comparison theorems.

For g_{ij}, $1 \le i \le N$, $1 \le j \le k$, $k \le N$, a family of independent standard normal variables, we denote by G the random operator $\mathbb{R}^N \to \mathbb{R}^k$ with matrix (g_{ij}). If S is a subset of $\mathbb{R}^N$, $\mathrm{diam} S = \sup\{|x-y|\,;\ x, y \in S\}$. If F is a vector subspace of a vector space B, the *codimension* of F in B is the dimension of the quotient space B/F.

Theorem 15.2. Let T be a convex body in $\mathbb{R}^N$. There exists a numerical constant K such that for all $k \le N$

$$\mathbb{P}\left\{ \mathrm{diam}(T \cap \mathrm{Ker} G) \le K \frac{\ell(T)}{\sqrt{k}} \right\} \ge 1 - \exp(-k/K)\,.$$

In particular, there exists a subspace F of $\mathbb{R}^N$ of codimension k (obtained as $F = \mathrm{Ker} G(\omega)$ for some appropriate ω) such that

$$\mathrm{diam}(T \cap F) \le K \frac{\ell(T)}{\sqrt{k}}\,.$$

1^{st} **proof of Theorem 15.2.** We start with an elementary observation. For every x in $\mathbb{R}^N$ and all $u > 0$,

$$\mathbb{P}\{|G(x)| < u|x|\} \le \left(\frac{eu^2}{k}\right)^{k/2}. \tag{15.1}$$

We may assume by homogeneity that $|x| = 1$. Then $|G(x)|^2$ has the distribution of $\sum_{i=1}^{k} g_i^2$. For every $\lambda > 0$,

$$\mathbb{E}\exp\left(-\lambda \sum_{i=1}^{k} g_i^2\right) = (\mathbb{E}\exp(-\lambda g_1^2))^k = \left(\frac{1}{1+2\lambda}\right)^{k/2}.$$

Therefore,

$$\mathbb{P}\{|G(x)|^2 < u^2\} = \mathbb{P}\left\{\exp\left(-\lambda \sum_{i=1}^{k} g_i^2\right) > \exp(-\lambda u^2)\right\}$$
$$\leq \left(\frac{1}{1+2\lambda}\right)^{k/2} \exp(\lambda u^2)\,.$$

Choose then for example $\lambda = k/2u^2$ and (15.1) follows.

Then next lemma is the main step of this proof. It is based on the isoperimetric type inequalities for Gaussian measures.

Lemma 15.3. Under the previous notations, for every integer k and $u > 0$,

$$\mathbb{P}\{\sup_{x\in T} |G(x)| > 4\ell(T) + u\} \leq 5^k \exp\left(-\frac{u^2}{8R^2}\right)$$

where $R = \sup_{x\in T} |x|$.

Proof. By Lemma 15.1 (or rather its immediate consequence), there exists H in B_2^k of cardinality less than 5^k such that

$$|G(x)| \leq 2 \sup_{h\in H} |\langle G(x), h\rangle|$$

for all x in $\mathbb{R}^N$. It therefore suffices to show that for every $|h| \leq 1$,

$$\mathbb{P}\{\sup_{x\in T} |\langle G(x), h\rangle| > 2\ell(T) + \frac{u}{2}\} \leq \exp\left(-\frac{u^2}{8R^2}\right)\,. \tag{15.2}$$

By the Gaussian rotational invariance and definition of G, the process $(\langle G(x), h\rangle)_{x\in T}$ is distributed as $\left(|h| \sum_{i=1}^{N} g_i x_i\right)_{x\in T}$. Then (15.2) is a direct consequence of Lemma 3.1. (Of course, we need not be concerned here with sharp constants in the exponential function and the simpler inequality (3.2) for example can be used equivalently.) Lemma 15.3 is established.

We can now conclude this first proof of Theorem 15.2. By Sudakov's minoration (Theorem 3.18),

$$\sup_{\varepsilon>0} \varepsilon(\log N(T, \varepsilon B_2^N))^{1/2} \leq K_1 \ell(T)$$

for some numerical constant K_1. Therefore, there exists a subset S of T of cardinality less than $\exp(K_1^2 k)$ such that the Euclidean balls of radius $k^{-1/2}\ell(T)$ and centers in S cover T. Let $\overline{T} = 2T \cap (k^{-1/2}\ell(T) B_2^N)$ and define the random variables

$$A = \sup_{x\in\overline{T}} |G(x)|\,, \qquad B = \inf_{x\in S} \frac{|G(x)|}{|x|}\,.$$

By Lemma 15.3 applied to $\overline{T}$, for every $u > 0$,

$$\mathbb{P}\{A > (8+u)\ell(T)\} \le 5^k \exp\left(-\frac{ku^2}{8}\right),$$

so that, if $u = 6$ for example ($k \ge 1$),

$$\mathbb{P}\{A > 14\ell(T)\} \le \exp(-2k).$$

On the other hand, by (15.1),

$$\mathbb{P}\{B < u\} \le \exp(K_1^2 k)\left(\frac{eu^2}{k}\right)^{k/2}.$$

If we choose here $u = K_2^{-1}\sqrt{k}$ where $K_2 = \exp(K_1^2 + 3)$, it follows that

$$\mathbb{P}\{B < K_2^{-1}\sqrt{k}\} \le \exp(-2k).$$

Therefore, the set $\{A \le 14\ell(T), B \ge K_2^{-1}\sqrt{k}\}$ has a probability larger than $1 - 2\exp(-2k) \ge 1 - \exp(-k)$. On this set of ω's, take $x \in T \cap \mathrm{Ker}G(\omega)$. There exists y in S with $|x-y| \le k^{-1/2}\ell(T)$. Since $G(\omega, x) = 0$, we can write that

$$\begin{aligned} |x| \le |y| + \frac{\ell(T)}{\sqrt{k}} \le \frac{|G(\omega,y)|}{B(\omega)} + \frac{\ell(T)}{\sqrt{k}} &= \frac{|G(\omega, x-y)|}{B(\omega)} + \frac{\ell(T)}{\sqrt{k}} \\ &\le \frac{A(\omega)}{B(\omega)} + \frac{\ell(T)}{\sqrt{k}} \le (14K_2 + 1)\frac{\ell(T)}{\sqrt{k}}. \end{aligned}$$

Hence $\mathrm{diam}(T \cap \mathrm{Ker}G(\omega)) \le K\ell(T)/\sqrt{k}$ with $K = 2(14K_2 + 1)$. This completes the first proof of Theorem 15.2.

2^{nd} **Proof of Theorem 15.2.** It is based on Corollay 3.13. Let S be a closed subset of the unit sphere S_2^{N-1} of $\mathbb{R}^N$. Let (g_i), (g_i') be independent orthogaussian sequences. For x in S and y in the Euclidean unit sphere S_2^{k-1} of $\mathbb{R}^k$, define the two Gaussian processes

$$X_{x,y} = \langle G(x), y\rangle + g = \sum_{i=1}^{N}\sum_{j=1}^{k} g_{ij} x_i y_j + g$$

where g is a standard normal variable independent of (g_{ij}) and

$$Y_{x,y} = \sum_{i=1}^{N} g_i x_i + \sum_{j=1}^{k} g_j' y_j.$$

It is easy to see that for all x, x' in S and y, y' in S_2^{k-1},

$$\mathbb{E}(X_{x,y}X_{x',y'}) - \mathbb{E}(Y_{x,y}Y_{x',y'}) = (1 - \langle x, x' \rangle)(1 - \langle y, y' \rangle)$$

so that this difference is always non-negative and equal to 0 when $x = x'$. By Corollary 3.13, for every $\lambda > 0$,

$$\mathbb{P}\{\inf_{x \in S} \sup_{y \in S_2^{k-1}} Y_{x,y} > \lambda\} \leq \mathbb{P}\{\inf_{x \in S} \sup_{y \in S_2^{k-1}} X_{x,y} > \lambda\}. \tag{15.3}$$

By definition of $X_{x,y}$ and $Y_{x,y}$, it follows that the right hand side of this inequality is majorized by

$$\mathbb{P}\{\inf_{x \in S} |G(x)| + g > \lambda\} \leq \mathbb{P}\{\inf_{x \in S} |G(x)| > 0\} + \frac{1}{2}\exp(-\lambda^2/2)$$

while the left hand side is bounded below by

$$\begin{aligned}
\mathbb{P}\left\{\left(\sum_{j=1}^{k} g_j'^2\right)^{1/2} - \sup_{x \in S} \sum_{i=1}^{N} g_i x_i > \lambda\right\} &\geq \mathbb{P}\{Z_k > 2\lambda\} - \mathbb{P}\left\{\sup_{x \in S} \sum_{i=1}^{N} g_i x_i > \lambda\right\} \\
&\geq \mathbb{P}\{Z_k > 2\lambda\} - \frac{1}{\lambda}\mathbb{E}\sup_{x \in S} \sum_{i=1}^{N} g_i x_i
\end{aligned}$$

where we have set $Z_k = \left(\sum_{j=1}^{k} g_j'^2\right)^{1/2}$. We can write

$$\begin{aligned}
k = \mathbb{E}Z_k^2 &\leq \frac{k}{10^2} + \int_{\{Z_k^2 \geq k/10^2\}} Z_k^2 d\mathbb{P} \\
&\leq \frac{k}{10^2} + (\mathbb{E}Z_k^4)^{1/2}(\mathbb{P}\{Z_k^2 \geq k/10^2\})^{1/2};
\end{aligned}$$

since $\mathbb{E}Z_k^4 \leq (k+1)^2$ (elementary computations), we get that

$$\mathbb{P}\{Z_k^2 \geq k/10^2\} \geq \left[\left(1 - \frac{1}{10^2}\right)\frac{k}{k+1}\right] \geq \frac{1}{2}$$

at least if $k \geq 3$ something we may assume without any loss of generality (increasing if necessary K in the conclusion of the theorem). Hence, by the preceeding, if $\lambda = \sqrt{k}/20$, the left side of (15.3) is larger than

$$\frac{1}{2} - \frac{20}{\sqrt{k}}\mathbb{E}\sup_{x \in S} \sum_{i=1}^{N} g_i x_i .$$

Let then $S = r^{-1}T \cap S_2^{N-1}$ where $r > 0$. We have obtained from (15.3) (with $\lambda = \sqrt{k}/20$) that

$$\mathbb{P}\{\inf_{x\in S} |G(x)| > 0\} \geq \frac{1}{2} - \frac{20}{\sqrt{k}}\frac{\ell(T)}{r} - \frac{1}{2}\exp(-k/800)\,.$$

Hence, if we choose $r = K\ell(T)/\sqrt{k}$ for K numerical large enough, we see from the preceding inequality that $\mathbb{P}\{\inf_{x\in S}|G(x)| > 0\} > 0$. There exists therefore an ω such that $|G(\omega,x)| > 0$ for all x in S. That is, if we let $F = \operatorname{Ker} G(\omega)$ (which is of codimension k), $S \cap F = \emptyset$. By definition of $S = r^{-1}T \cap S_2^{N-1}$, this implies that, for every x in $T \cap F$, $|x| < r = K\ell(T)/\sqrt{k}$. The qualitative conclusion of the theorem already follows. To improve the preceding argument into the quantitative estimate, we n eed simply combine it with concentration properties. We indeed improve the minoration of the right hand side of (15.3) in the following way. Let m and M be respectively medians of $Z_k = (\sum_{j=1}^k g_j^2)^{1/2}$ and $\sup_{x\in S}\sum_{i=1}^N g_i x_i$. For every $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\{\inf_{x\in S}\sup_{y\in S_2^{k-1}} Y_{x,y} > \lambda\} &\geq \mathbb{P}\left\{Z_k - \sup_{x\in S}\sum_{i=1}^N g_i x_i > \lambda\right\}\\ &\geq \mathbb{P}\{Z_k \geq m-\lambda\} - \mathbb{P}\left\{\sup_{x\in S}\sum_{i=1}^N g_i x_i > m - 2\lambda\right\}\\ &\geq 1 - \mathbb{P}\{m - Z_k > \lambda\} - \mathbb{P}\left\{\sup_{x\in S}\sum_{i=1}^N g_i x_i > M + (m - M - 2\lambda)\right\}.\end{aligned}$$

By Lemma 3.1, if $m - M - 2\lambda > 0$, this is larger than

$$1 - \frac{1}{2}\exp(-\lambda^2/2) - \frac{1}{2}\exp[-(m-M-2\lambda)^2/2]\,.$$

Let us then choose $\lambda = (m-M)/3$ to get the lower bound

$$1 - \exp[-(m-M)^2/18]$$

and (15.3) thus reads as

$$\mathbb{P}\{\inf_{x\in S}|G(X)| > 0\} \geq 1 - \frac{3}{2}\exp\{-(m-M)^2/18]\,.$$

We have seen previously that $m \geq \sqrt{k}/10$. On the other hand, if S is as before $r^{-1}T \cap S_2^{N-1}$,

$$M \leq 2\mathbb{E}\sup_{x\in S}\sum_{i=1}^N g_i x_i \leq \frac{2}{r}\ell(T)$$

so that

$$\mathbb{P}\{\ \mathrm{diam}(T \cap\ \mathrm{Ker} G) \leq 2r\} \geq \mathbb{P}\{\inf_{x \in S} |G(x)| > 0\}$$
$$\geq 1 - \frac{3}{2} \exp\left[-\frac{1}{18} \left(\frac{\sqrt{k}}{10} - \frac{2\ell(T)}{r} \right)^2 \right] .$$

The conclusion of the theorem follows. Let us observe again that the simpler inequality (3.2) may be used instead of Lemma 3.1.

Note that this second proof may be rather simplified yielding moreover best constants in the statement of the theorem by the use of Theorem 3.16. Take $X_{x,y} = \langle G(x), y \rangle$ and

$$Y_{x,y} = \sum_{i=1}^{N} g_i x_i + |x| \sum_{j=1}^{k} g'_j y_j \, .$$

For all x, x' in S and y, y' in S_2^{k-1}, we have

$$\mathbb{E}|Y_{x,y} - Y_{x',y'}|^2 - \mathbb{E}|X_{x,y} - X_{x',y'}|^2 = |x|^2 + |x'|^2 - 2|x|\,|x'|\langle y, y' \rangle - 2\langle x, x' \rangle (1 - \langle y, y' \rangle)$$
$$\geq |x|^2 + |x'|^2 - 2|x|\,|x'|\langle y, y' \rangle - 2|x|\,|x'|(1 - \langle y, y' \rangle)$$

so that this difference is always ≥ 0 and equal to 0 if $x = x'$. By Theorem 3.16, this implies that

$$\mathbb{E} \inf_{x \in S} \sup_{y \in S_2^{k-1}} Y_{x,y} \leq \mathbb{E} \inf_{x \in S} \sup_{y \in S_2^{k-1}} X_{x,y}$$

and hence, by definition of $X_{x,y}$ and $Y_{x,y}$ and with $S = r^{-1} T \cap S_2^{N-1}$,

$$\mathbb{E} \inf_{x \in S} |G(x)| \geq a_k - \frac{1}{r} \ell(T)$$

where $a_k = \mathbb{E} Z_k = \mathbb{E}((\sum_{j=1}^{k} {g'_j}^{\,2})^{1/2})$. By the law of large numbers, a_k is of the order of $\sqrt{k}$. Write then that

$$\mathbb{P}\{\inf_{x \in S} |G(x)| = 0\} \leq \mathbb{P}\{\inf_{x \in S} |G(x)| - \mathbb{E} \inf_{x \in S} |G(x)| \leq \frac{1}{r} \ell(T) - a_k\}$$

which is majorized using concentration ((1.6) e.g.) by

$$\exp\left[\frac{1}{2} \left(a_k - \frac{\ell(T)}{r} \right)^2 \right] .$$

As before, this yields the conclusion of the theorem, with, as announced, improved numerical constants.

15.2. Conjectures on Sudakov's minoration for chaos

Denote by $\ell_2(\mathbb{N}\times\mathbb{N})$ the space of all sequences $t=(t_{ij})$ indexed by $\mathbb{N}\times\mathbb{N}$ such that $|t| = \left(\sum_{i,j} t_{ij}^2\right)^{1/2} < \infty$. Let T be a subset of $\ell_2(\mathbb{N}\times\mathbb{N})$. Let further (g_i) be an orthogaussian sequence (on $(\Omega,\mathcal{A},\mathbb{P})$) and consider the Gaussian chaos process $\left(\sum_{i,j} g_i g_j t_{ij}\right)_{t\in T}$. While we have studied in Chapter 3 the integrability properties and tail behavior of $\sup_{t\in T}|\sum_{i,j} g_i g_j t_{ij}|$, little seems to be known on the "metric geometric" conditions on T equivalent to this supremum to be almost surely finite (if there are any?). The sufficient conditions of Theorem 11.22 are too strong and by no way necessary. One approach could be to view the preceding chaos process, after decoupling, as a mixture of Gaussian processes, i.e. to study, given ω', $\left(\sum_i g_i\left(\sum_j g_j'(\omega')t_{ij}\right)\right)_{t\in T}$ where (g_j') is a standard Gaussian sequence constructed on some different probability space $(\Omega',\mathcal{A}',\mathbb{P}')$. Unfortunately, this approach seems to be doomed to failure: the random distances

$$d_{\omega'}(s,t) = \left(\sum_i \left|\sum_j g_j'(\omega')(s_{ij}-t_{ij})\right|^2\right)^{1/2}$$

do not have the property that was essential in Section 12.2, i.e. that $\mathbb{P}'\{\omega'; d_{\omega'}(s,t)<\varepsilon\}$ is very small for $\varepsilon>0$ small.

Let us reduce to decoupled chaos and set

$$\ell^{(2)}(T) = \mathbb{E}\sup_{t\in T}\left|\sum_{i,j} g_i g_j' t_{ij}\right| .$$

(With respect to the notation of Section 15.1, the 2 in $\ell^{(2)}(T)$ indicates that we are dealing with chaos of order 2.) By the study of Section 13.2 we know that, at least for symmetric (t_{ij}), this reduction to the decoupled setting is no loss of generality in the understanding of when T defines an almost surely bounded chaos process. A first step in this study would be an analog of Sudakov's minoration. As we discussed it in Section 11.3 on sufficiency, there are two natural distances to consider. The first one is the usual L_2-metric $|s-t|$ and the second one is the injective metric or norm given by

$$\|t\|_\vee = \sup_{|h|,|h'|\le 1}|\langle th,h'\rangle| = \sup_{|h|\le 1}|th|$$

where $th = \left(\sum_j h_j t_{ij}\right)_i$. Clearly $\|t\|_\vee \le |t|$.

We first investigate an instructive example. Denote, for n fixed, by $\ell_2(n \times n)$ the subset of $\ell_2(\mathbb{N} \times \mathbb{N})$ consisting of the elements (t_{ij}) for which $t_{ij} = 0$ if i or $j > n$. Let T be the unit ball of $\ell_2(n \times n)$ for the injective norm $\|\cdot\|_\vee$. Clearly

$$\sup_{t \in T} \left| \sum_{i,j=1}^{n} g_i g'_j t_{ij} \right| \le \left(\sum_{i=1}^{n} g_i^2 \right)^{1/2} \left(\sum_{j=1}^{n} {g'_j}^{\,2} \right)^{1/2}$$

so that $\ell^{(2)}(T) \le n$.

Proposition 15.4. Let T be as before, $T = \{t \in \ell_2(n \times n); \|t\|_\vee \le 1\}$. There is a numerical constant $c > 0$ (independent of n) such that

(i) $(\log N(T, |\cdot|; c\sqrt{n}))^{1/2} \ge cn$;

(ii) $(\log N(T, \|\cdot\|_\vee; \frac{1}{2}))^{1/2} \ge cn$.

Proof. (ii) is obvious by volume considerations. We give a simple probabilistic proof of (i). Let $(\varepsilon_{ij})_{1 \le i,\, j \le n}$ be a doubly-indexed Rademacher sequence. For $|h|$, $|h'| \le 1$, by the subgaussian inequality (4.1), for every $u > 0$,

$$\mathbb{P}\left\{ \left| \sum_{i,j} h_i h'_j \varepsilon_{ij} \right| > u \right\} \le 2 \exp\left(-\frac{u^2}{2} \right).$$

By Lemma 15.1, there exists $H \subset 2B_2^n$, $\mathrm{Card} H \le 5^n$ such that $B_2^n \subset \mathrm{Conv} H$. By the preceding, since $\mathrm{Card} H \le 5^n$,

$$\mathbb{P}\left\{ \forall h, h' \in H; \left| \sum_{i,j} h_i h'_j \varepsilon_{ij} \right| > 4\sqrt{n} \right\} \le \frac{1}{2}.$$

By definition of $\|\cdot\|_\vee$ and H, it follows that $\mathbb{P}\{\|(\varepsilon_{ij})\|_\vee \le 4\sqrt{n}\} \ge 1/2$. Let η_{ij}, $1 \le i, j \le n$, be a family of ± 1. Then $(|\varepsilon_{ij} - \eta_{ij}|^2)$ is a sequence of random variables taking the values 0 and 1 with probability $1/2$. Recentering, Lemma 1.6 implies that for some $c > 0$,

$$\mathbb{P}\left\{ \sum_{i,j} |\varepsilon_{ij} - \eta_{ij}|^2 \le \frac{n^2}{4} \right\} \le \exp(-cn^2).$$

That is $\mathbb{P}\{|(\varepsilon_{ij} - \eta_{ij})| \le n/2\} \le \exp(-cn^2)$. So one needs at least $\frac{1}{2}\exp(cn^2)$ balls of radius $n/2$ in the Euclidean metric $|\cdot|$ to cover $\{\|(\varepsilon_{ij})\|_\vee \le 4\sqrt{n}\} \subset 4\sqrt{n}T$. Therefore, one needs at least $\frac{1}{2}\exp(cn^2)$ balls of radius $\sqrt{n}/8$ to cover T. The proof of Proposition 15.4 is complete.

It follows from this proposition that, for $\varepsilon = c\sqrt{n}$,

$$\varepsilon(\log N(T, |\cdot|; \varepsilon))^{1/4} \geq c^{3/2} \geq c^{3/2}\ell^{(2)}(T),$$

and, for $\varepsilon = 1/2$,

$$\varepsilon(\log N(T, \|\cdot\|_{\vee}; \varepsilon))^{1/2} \geq \frac{c}{2}n \geq \frac{c}{2}\ell^{(2)}(T)$$

since, as we have seen, $\ell^{(2)}(T) \leq n$. It is natural to *conjecture* that these inequalities are best possible, i.e. that for any subset T in $\ell_2(\mathbb{N} \times \mathbb{N})$,

$$\sup_{\varepsilon>0} \varepsilon(\log N(T, |\cdot|; \varepsilon))^{1/4} \leq K\ell^{(2)}(T) \tag{15.4}$$

and

$$\sup_{\varepsilon>0} \varepsilon(\log N(T, \|\cdot\|_{\vee}; \varepsilon))^{1/2} \leq K\ell^{(2)}(T)$$

for some numerical K. Recently (15.4) has been proved in [Ta16] (relying on an appropriate version of Theorem 15.2), where it is also proved that for $\delta > 2$

$$\sup_{\varepsilon>0} \varepsilon(\log N(T, \|\cdot\|_{\vee}; \varepsilon))^{1/\delta} \leq K(\delta)\ell^{(2)}(T).$$

We would like now to present a simple result that is somewhat related to these questions.

Proposition 15.5. Let (g_i), (g'_j), $(\widetilde{g}_{ij})$ be independent standard Gaussian sequences. Then, if $(x_{ij})_{1\leq i,j\leq n}$ is a finite sequence in a Banach space,

$$\mathbb{E}\left\|\sum_{i,j=1}^{n} \widetilde{g}_{ij}x_{ij}\right\| \leq \left(\frac{\pi}{2}\right)^{1/2} \sqrt{n}\mathbb{E}\left\|\sum_{i,j=1}^{n} g_i g'_j x_{ij}\right\|. \tag{15.4}$$

Proof. We simply write that

$$\sqrt{n}\mathbb{E}\left\|\sum_{i,j=1}^{n} g_i g'_j x_{ij}\right\| = \mathbb{E}\left\|\sum_{i,j,k=1}^{n} \widetilde{g}_{ik} g'_j x_{ij}\right\|.$$

By symmetry, if (ε_i) is a Rademacher sequence independent of $(\widetilde{g}_{ik})$, (g'_j),

$$\mathbb{E}\left\|\sum_{i,j,k=1}^{n} \widetilde{g}_{ik} g'_j x_{ij}\right\| = \mathbb{E}\left\|\sum_{i,j,k=1}^{n} \varepsilon_k \widetilde{g}_{ik} \varepsilon_j g'_j x_{ij}\right\|.$$

Jensen's inequality with respect to partial integration in (ε_i) shows then that

$$\mathbb{E}\left\|\sum_{i,j,k=1}^{n}\widetilde{g}_{ik}g_j'x_{ij}\right\| \geq \mathbb{E}\left\|\sum_{i,j=1}^{n}\widetilde{g}_{ik}g_j'x_{ij}\right\| .$$

The contraction principle in (g_j') (cf. (4.8)) and symmetry then imply the result.

15.3. An inequality of J. Bourgain

In his remarkable recent work on $\Lambda(p)$-sets [Bour], J. Bourgain establishes the following inequality. Let (ξ_i) be independent random variables with common distribution

$$\mathbb{P}\{\xi_i = 1\} = 1 - \mathbb{P}\{\xi_i = 0\} = \delta\,, \qquad 0 < \delta < 1\,.$$

Let T be a subset of $\mathbb{R}_+^N$ and set

$$E = \int_0^\infty (\log N(T, \varepsilon B_2^N))^{1/2} d\varepsilon$$

(assumed to be finite). Set further $R = \sup_{t\in T}|t|$. An element t in $\mathbb{R}^N$ has coordinates $t = (t_1, \ldots, t_N)$.

Theorem 15.6. Under the previous notations, there is a numerical constant K such that for all $p \geq 1$ and $1 \leq m \leq N$,

$$\left\|\sup_{t\in T}\max_{\mathrm{Card}I\leq m}\sum_{i\in I}\xi_i t_i\right\|_p \leq KR\left[\sqrt{\delta m} + \sqrt{p}\left(\log\frac{1}{\delta}\right)^{-1/2}\right] + K\left(\log\frac{1}{\delta}\right)^{-1/2} E\,.$$

The inequality of the theorem is equivalent to say (Lemma 4.10) that for some constant K and all $u > 0$

$$\begin{aligned}(15.5)\qquad & \mathbb{P}\left\{\sup_{t\in T}\max_{\mathrm{Card}I\leq m}\sum_{i\in I}\xi_i t_i > u + K\left[R\sqrt{\delta m} + \left(\log\frac{1}{\delta}\right)^{-1/2} E\right]\right\} \\ & \qquad \leq K\exp\left(-\frac{u^2}{KR^2}\log\frac{1}{\delta}\right) .\end{aligned}$$

We will establish (15.5) using the entropic bound (11.4) for the *vector valued* process $\sum_{i\in I}\xi_i t_i$, $I \subset \{1, \ldots, N\}$, $\mathrm{Card}I \leq m$, $t \in T$, with respect to the norm

$$\max_{\mathrm{Card}I\leq m}\sum_{i\in I}\xi_i t_i\,.$$

To this aim, the next crucial lemma will indicate that the increments of this process satisfy the appropriate regularity condition.

As usual, if $t=(t_1,\dots,t_N)\in\mathbb{R}^N$, $\|(t_i)\|_{2,\infty}$ denotes the weak ℓ_2 norm of the sequence $(t_i)_{i\le N}$. Recall that $\|(t_i)\|_{2,\infty}\le|t|$.

Lemma 15.7. Let t in $\mathbb{R}_+^N$ with $\|(t_i)\|_{2,\infty}\le 1$ and let also $1\le m\le N$. There is a numerical constant K such that for all $u\ge K\sqrt{\delta m}$,

$$\mathbb{P}\left\{\max_{\mathrm{Card}\le m}\sum_{i\in I}\xi_i t_i>u\right\}\le K\exp\left(-\frac{u^2}{K}\log\frac{1}{\delta}\right).$$

Proof: Let $(Z_i)_{i\le N}$ be the non-increasing rearrangement of the sequence $(\xi_i t_i)_{i\le N}$ so that

$$\max_{\mathrm{Card}I\le m}\sum_{i\in I}\xi_i t_i=\sum_{i=1}^m Z_i .$$

Note that since $\|(t_i)\|_{2,\infty}\le 1$, and $\xi_i=0$ or 1, we also have that $\|(\xi_i t_i)\|_{2,\infty}\le 1$, that is $Z_i\le i^{-1/2}$, $i\ge 1$. By the binomial estimate (Lemma 2.5), for every i and $u>0$,

$$\mathbb{P}\{Z_i>u\}=\mathbb{P}\left\{\sum_{j=1}^N I_{\{\xi_j t_j>u\}}\ge i\right\}\le\left(\frac{e}{i}\sum_{j=1}^N\mathbb{P}\{\xi_j t_j>u\}\right)^i .$$

Since $\|(t_i)\|_{2,\infty}\le 1$,

$$\begin{aligned}\sum_{j=1}^N\mathbb{P}\{\xi_j t_j>u\}&\le\sum_{j=1}^N\mathbb{P}\{\xi_1>u\sqrt{j}\}\\&=\sum_{j=1}^N\mathbb{P}\{\xi_1>u^2 j\}\le\mathbb{E}\left(\frac{\xi_1}{u^2}\right)=\frac{\delta}{u^2}.\end{aligned}$$

Therefore, for all integers $i\le N$ and all $u>0$,

$$\mathbb{P}\{Z_i>u\}\le\left(\frac{e\delta}{u^2 i}\right)^i . \tag{15.6}$$

Without loss of generality, we can assume that $m=2^p$. Let $u>0$ be fixed. Let $j\ge 1$ be the largest such that $u^2\ge 2e10^4 2^j$ and take $j=0$ if $u^2<2e10^4$. (We do not attempt to deal with sharp numerical constants.) We observe that, if $j\ge 1$,

$$\sum_{i=1}^{2^j}Z_i\le\sum_{i=1}^{2^j}i^{-1/2}\le 2\cdot 2^{j/2}\le\frac{u}{2}.$$

We also observe that

$$\sum_{i>2^j} Z_i \leq \sum_{\ell=j}^{p} 2^\ell Z_{2^\ell}\,.$$

It follows from these two observations that, for all $j \geq 0$,

$$\mathbb{P}\left\{\sum_{i=1}^{m} Z_i > u\right\} \leq \mathbb{P}\left\{\sum_{\ell=j}^{p} 2^\ell Z_{2^\ell} > \frac{u}{2}\right\} \leq \sum_{\ell=j}^{p} \mathbb{P}\{2^\ell > Z_{2^\ell} > u_\ell\}$$

whenever (u_ℓ) are positive numbers such that $\sum_{\ell=j}^{p} u_\ell \leq u/2$. Let $v_\ell = 10^{-2} u 2^{-(\ell-j)/2}$, $w_\ell = 10(\delta 2^\ell)^{1/2}$, $j \leq \ell \leq p$, and set $u_\ell = \max(v_\ell, w_\ell)$. It is clear that $\sum_{\ell \geq j} v_\ell \leq u/4$ while, if $u \geq K\sqrt{\delta m}$ where $K \geq 10^3$,

$$\sum_{\ell \leq p} w_\ell \leq 3 \cdot 10(\delta 2^{p+1})^{1/2} \leq \frac{u}{4}\,.$$

Hence we have that $\sum_{\ell=j}^{p} u_\ell \leq u/2$. By the binomial estimates (15.6),

$$\mathbb{P}\left\{\sum_{i=1}^{m} Z_i > u\right\} \leq \sum_{\ell=j}^{p} \left(\frac{e\delta 2^\ell}{u_\ell^2}\right)^{2^\ell} \leq \sum_{\ell=j}^{p} \exp\left[-2^\ell \log\left(\frac{u_\ell^2}{e\delta 2^\ell}\right)\right]. \tag{15.7}$$

Note that for $2\delta \leq x \leq 4$,

$$\log\left(\frac{x}{\delta}\right) \geq \frac{x}{4}\log\frac{1}{\delta}\,.$$

Therefore, the first term $(\ell = j)$ of the sum on the right of (15.7) gives raise to an exponent of the form

$$2^j \log\left(\frac{u_j^2}{e\delta 2^j}\right) \geq 2^j \log\left(\frac{v_j^2}{e\delta 2^j}\right) = 2^j \log\left(\frac{u^2}{e10^4 \delta 2^j}\right) \geq \frac{1}{4}\,\frac{u^2}{e10^4}\log\frac{1}{\delta}$$

provided that

$$2\delta \leq \frac{u^2}{e10^4 2^j} \leq 4\,.$$

This clearly holds by definition of $j \geq 1$ and also if $j = 0$ since $u \geq K\sqrt{\delta m}$ ($K \geq 10^3$). Hence

$$\exp\left[-2^j \log\left(\frac{u_j^2}{e\delta 2^j}\right)\right] \leq \exp\left(-\frac{u^2}{K}\log\frac{1}{\delta}\right) \tag{15.8}$$

for $K \geq 4e10^4$.

We now would like to show that for every $\ell > j$

$$2^\ell \log\left(\frac{u_\ell^2}{e\delta 2^\ell}\right) \geq \frac{6}{5} 2^{\ell-1} \log\left(\frac{u_{\ell-1}^2}{e\delta 2^{\ell-1}}\right) , \tag{15.9}$$

that is,

$$\log\left(\frac{u_\ell^2}{e\delta 2^\ell}\right) \geq \frac{3}{5} \log\left(\frac{u_{\ell-1}^2}{e\delta 2^{\ell-1}}\right) .$$

If $u_{\ell-1} = w_{\ell-1}$, then $u_\ell = w_\ell$ and this inequality is trivially satisfied since by definition of w_ℓ , the interior of the logarithm is constant. If $u_{\ell-1} = v_{\ell-1}$, we note the following:

$$\frac{u_\ell^2}{e\delta 2^\ell} \geq \frac{v_\ell^2}{e\delta 2^\ell} = \frac{1}{4} \frac{v_{\ell-1}^2}{e\delta 2^{\ell-1}} = \frac{1}{4} \frac{u_{\ell-1}^2}{e\delta 2^{\ell-1}} \geq \frac{1}{4} \frac{w_{\ell-1}^2}{e\delta 2^{\ell-1}} \geq 2^3 .$$

Using that $\log(x/4) \geq (3\log x)/5$ when $x \geq 2^5$, (15.9) is thus satisfied.

We can now conclude the proof of Lemma 15.7. By (15.7), (15.8) and (15.9), we have that

$$\begin{aligned}
\mathbb{P}\left\{\sum_{i=1}^m Z_i > u\right\} &\leq \sum_{k\geq 0} \exp\left[-\left(\frac{6}{5}\right)^k \frac{u^2}{K} \log\frac{1}{\delta}\right] \\
&\leq \sum_{k\geq 1} \exp\left(-\frac{k}{3} \frac{u^2}{K} \log\frac{1}{\delta}\right) \\
&\leq 2\exp\left(-\frac{u^2}{3K} \log\frac{1}{\delta}\right)
\end{aligned}$$

at least if $u^2\log(1/\delta) \geq 3K$. The inequality of Lemma 15.7 follows with (for example) $K = 12e10^4$ since there is nothing to prove if $u^2\log(1/\delta) \leq 3K$.

Proof of Theorem 15.6. For s, t in $\mathbb{R}^N$, define the random distance

$$D(s,t) = \left(\log\frac{1}{\delta}\right)^{1/2} \max_{\mathrm{Card} I \leq m} \left|\sum_{i\in I} \xi_i(s_i - t_i)\right| .$$

Lemma 15.7 implies that for all s, t and $u \geq K(m\delta\log\frac{1}{\delta})^{1/2}$,

$$\mathbb{P}\{D(s,t) > u|s-t|\} \leq K\exp\left(-\frac{u^2}{K}\right) .$$

K denotes a numerical constant not necessarily the same each time it appears below. From the preceding, for all $u > 0$,

$$\mathbb{P}\left\{D(s,t) > \left(u + K\left(m\delta\log\frac{1}{\delta}\right)^{1/2}\right)|s-t|\right\} \leq K\exp\left(-\frac{u^2}{K}\right) .$$

Hence, for all measurable sets A, and all s, t,

$$\int_A D(s,t)d\mathbb{P} \leq K\mathbb{P}(A)|s-t| \left(\psi_2^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + \left(m\delta \log \frac{1}{\delta}\right)^{1/2} \right)$$

where we recall that $\psi_2(x) = \exp(x^2) - 1$. We are therefore in a position to apply Remark 11.4 to the random distance $D(s,t)$ (cf. Remark 11.5). It follows that, for all measurable sets A,

$$\int_A \sup_{s,t \in T} D(s,t)d\mathbb{P} \leq K\mathbb{P}(A) \left(D\psi_2^{-1}\left(\frac{1}{\mathbb{P}(A)}\right) + R\left(m\delta \log \frac{1}{\delta}\right)^{1/2} + E \right)$$

and hence ((11.4)), for all $u > 0$,

$$\mathbb{P}\left\{ \sup_{s,t \in T} D(s,t) > u + K\left(R\left(m\delta \log \frac{1}{\delta}\right)^{1/2} + E \right) \right\} \leq K \exp\left(-\frac{u^2}{KR^2}\right).$$

Since for every t, by Lemma 15.7,

$$\mathbb{P}\left\{ D(0,t) > u + KR\left(m\delta \log \frac{1}{\delta}\right)^{1/2} \right\} \leq K \exp\left(-\frac{u^2}{KR^2}\right)$$

for all $u > 0$, inequality (15.5) follows. The proof of Theorem 15.6 is thus complete.

Note of course that we can replace the entropy integral by sharper majorizing measure conditions.

15.4. Invertibility of submatrices

This section is devoted to the exposition of one simple but significant result of the work [B-T1] (see also [B-T2]) by J. Bourgain and L. Tzafriri. The proof of this result uses several probabilistic ideas and arguments already encountered throughout this book.

Let A be a real $N \times N$ matrix considered as a linear operator on $\mathbb{R}^N$. Denote by $\|A\| = \|A\|_{2\to 2}$ its norm as an operator $\ell_2^N \to \ell_2^N$. If σ is a subset of $\{1, \ldots, N\}$, denote by R_σ the restriction operator, that is the projection from $\mathbb{R}^N$ onto the span of the unit vectors $(e_i)_{i\in\sigma}$ where (e_i) is the canonical basis of $\mathbb{R}^N$. R_σ^t is the transpose operator. By restricted invertibility of A, we mean the existence of a subset σ of $\{1, \ldots, N\}$ with cardinality of the order of N such that $R_\sigma A R_\sigma^t$ is an isomorphism. If the diagonal of A is the identity matrix I, this will be achieved by simply constructing the set σ such that $\|R_\sigma (A - I) R_\sigma^t\| < 1/2$.

The following statement is the main result we present.

Theorem 15.8. There is a numerical constant K with the following property: for every $0<\delta<K^{-1}$ and $N \geq K/\delta$, whenever $A=(a_{ij})$ is an $N \times N$ matrix with $a_{ii}=0$ for all i there exists a subset σ of $\{1,\ldots,N\}$ such that $\mathrm{Card}\sigma \geq \delta N$ and

$$\|R_\sigma A R_\sigma^t\| \leq K\sqrt{\delta}\|A\|\,.$$

The following restricted invertibility statement is an immediate consequence of Theorem 15.8.

Corollary 15.9. There is a numerical constant K with the following property: for every $0<\varepsilon\leq 1$, $c \geq 1$ and $N \geq Kc\varepsilon^{-2}$, whenever A is an $N \times N$ matrix of norm $\|A\| \leq c$ and with only 1's on the diagonal, there exists a subset σ of $\{1,\ldots,N\}$ of cardinality $\mathrm{Card}\sigma \geq \varepsilon^2 N/Kc$ such that $R_\sigma A R_\sigma^t$ restricted to $R_\sigma \ell_2^N$ is invertible and its inverse satisfies

$$\|(R_\sigma A R_\sigma^t)^{-1}\| < 1+\varepsilon\,.$$

The following proposition, whose proof is based on a random choice argument, is the main step in the proof of Theorem 15.8.

Proposition 15.10. Let $0<\delta<1$ and $N \geq 8/\delta$. Then, for all $N \times N$ matrices $A=(a_{ij})$ with $a_{ii}=0$, there exists $\sigma \subset \{1,\ldots,N\}$ such that $\mathrm{Card}\sigma \geq \delta N/2$ and

$$\|R_\sigma A R_\sigma^t\|_{2\to 1} \leq 50\delta\|A\|\sqrt{N}$$

where $\|\cdot\|_{2\to 1}$ is the operator norm $\ell_2^M \to \ell_1^M$ ($M=$ $\mathrm{Card}\sigma$).

Before proving this proposition, let us show how to establish then Theorem 15.8 from this result. We need the following second step which clarifies the passage through the norm $\|\cdot\|_{2\to 1}$.

Proposition 15.11. Let $u:\ell_2^n \to \ell_1^m$. There exists a subset τ of $\{1,\ldots,m\}$ with $\mathrm{Card}\tau \geq m/2$ such that

$$\|R_\tau u\|_{2\to 2} \leq \left(\frac{\pi}{m}\right)^{1/2}\|u\|_{2\to 1}\,.$$

Proof. Consider $u^*:\ell_\infty^m \to \ell_2^n$, $\|u^*\|_{\infty\to 2}=\|u\|_{2\to 1}$. By the "little Grothendieck theorem" (cf. [Pi15]), if $\pi_2(u^*)$ is the 2-summing norm of u^*,

$$\pi_2(u^*) \leq (\pi/2)^{1/2}\|u^*\|_{\infty\to 2}\,.$$

Therefore, by Pietsch's factorization theorem (cf. e.g. [Pie], [Pi15]), there exists $(\lambda_i)_{i\le m}$, $\lambda_i > 0$, $\sum\limits_{i=1}^{m} \lambda_i = 1$, such that for all $x = (x_1, \ldots, x_m)$ in ℓ_∞^m ,

$$|u^*(x)| \le \left(\frac{\pi}{2}\right)^{1/2} \|u^*\|_{\infty\to 2} \left(\sum_{i=1}^{m} \lambda_i x_i^2\right)^{1/2} .$$

Let $\tau = \{i \le m\,;\ \lambda_i \le 2/m\}$. Then $\mathrm{Card}\tau \ge m/2$ and, from the preceding,

$$\|u^* R_\tau^t\|_{2\to 2} \le \left(\frac{\pi}{2}\right)^{1/2} \|u^*\|_{\infty\to 2}$$

from which Proposition 15.11 follows.

These two propositions clearly imply the theorem. If $N \ge 8/\delta$ and A is an $N\times N$ matrix with $a_{ii} = 0$, there exists, by Proposition 15.10, σ in $\{1, \ldots, N\}$ such that $\mathrm{Card}\sigma \ge \delta N/2$ and

$$\|R_\sigma A R_\sigma^t\|_{2\to 1} \le 50\delta \|A\|\sqrt{N} .$$

Apply then Proposition 15.11 to $u = R_\sigma A R_\sigma^t$. It follows that one can find $\tau \subset \sigma$ with $\mathrm{Card}\tau \ge \frac{1}{2}\,\mathrm{Card}\sigma \ge \delta N/4$ and

$$\|R_\tau A R_\tau^t\|_{2\to 2} \le \left(\frac{\pi}{\mathrm{Card}\sigma}\right)^{1/2} 50\delta\|A\|\sqrt{N} \le 50(2\pi\delta)^{1/2}\|A\| .$$

Theorem 15.8 immediately follows.

Proof of Proposition 15.10. Let $0 < \delta < 1$ and let ξ_i be independent with distribution $\mathbb{P}\{\xi_i = 1\} = 1 - \mathbb{P}\{\xi_i = 0\} = \delta$. Let $A(\omega)$ be the random operator $\mathbb{R}^N \to \mathbb{R}^N$ with matrix $(\xi_i(\omega)\xi_j(\omega)a_{ij})$. We use a decoupling argument in the following form. For $I \subset \{1, \ldots, N\}$, let

$$B_I = \mathbb{E}\left\| \sum_{i\notin I, j\in I} \xi_i\xi_j a_{ij} e_i \otimes e_j \right\|_{2\to 1} .$$

Since

$$\frac{1}{2^N}\sum_I \sum_{i\notin I, j\in I} a_{ij} = \frac{4}{2^N}\sum_{i,j} a_{ij}$$

(recall that $a_{ii} = 0$), we have that

$$\mathbb{E}\|A(\omega)\|_{2\to 1} \le \frac{4}{2^N}\sum_I B_I . \tag{15.10}$$

We therefore estimate B_I for each fixed $I \subset \{1, \dots, N\}$. We have that

$$\Big\| \sum_{i \notin I, j \in I} \xi_i \xi_j a_{ij} e_i \otimes e_j \Big\|_{2 \to 1} = \sup_{|h| \le 1} \sum_{i \notin I} \xi_i \Big| \sum_{j \in I} \xi_j h_j a_{ij} \Big| .$$

For every i and h, let $d_i(h) = \sum_{j \in I} \xi_j h_j a_{ij}$. Therefore,

$$B_I = \mathbb{E} \sup_{|h| \le 1} \sum_{i \notin I} \xi_i |d_i(h)| .$$

Let (ξ_i') be an independent copy of (ξ_i). Working conditionally on ξ_j, ξ_j', $j \in I$, we get by centering and Jensen's inequality that

$$\begin{aligned} B_I &\le \mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} (\xi_i - \mathbb{E}\xi_i) |d_i(h)| \Big| + \delta \sup_{|h| \le 1} \sum_{i \notin I} |d_i(h)| \\ &\le \mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} (\xi_i - \xi_i') |d_i(h)| \Big| + \delta \sup_{|h| \le 1} \sum_{i \notin I} |d_i(h)| . \end{aligned}$$

By Cauchy-Schwarz,

$$\sup_{|h| \le 1} \sum_{i \notin I} |d_i(h)| \le \|A\| \sqrt{N} .$$

On the other hand, if (ε_i) is a Rademacher sequence independent of (ξ_i), by symmetry

$$\mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} (\xi_i - \xi_i') |d_i(h)| \Big| \le 2\mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} \varepsilon_i \xi_i |d_i(h)| \Big| .$$

By the comparison properties of Rademacher averages (Theorem 4.12), we have further that

$$\mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} \varepsilon_i \xi_i |d_i(h)| \Big| \le 2\mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} \varepsilon_i \xi_i d_i(h) \Big| .$$

Summarizing, we have obtained so far that

$$B_I \le 4\mathbb{E} \sup_{|h| \le 1} \Big| \sum_{i \notin I} \varepsilon_i \xi_i d_i(h) \Big| + \delta \|A\| \sqrt{N} .$$

Going back to the definition of $d_i(h)$, we see that

$$\sup_{|h|\le 1}\left|\sum_{i\notin I}\varepsilon_i\xi_i d_i(h)\right|^2=\sum_{j\in I}\xi_j\left|\sum_{i\notin I}\varepsilon_i\xi_i a_{ij}\right|^2 .$$

If we integrate this expression with respect to the Rademacher sequence and then with respect to the variables ξ_i, $i\notin I$, we obtain

$$\delta\sum_{j\in I}\xi_j\left(\sum_{i\notin I}a_{ij}^2\right) .$$

Since $\sum_i a_{ij}^2\le\|A\|^2$ for every j, we finally get that

$$\begin{aligned}B_I&\le 4\left(\mathbb{E}\sup_{|h|\le 1}\left|\sum_{i\notin I}\varepsilon_i\xi_i d_i(h)\right|^2\right)^{1/2}+\delta\|A\|\sqrt{N}\\&\le 4(\delta^2\|A\|^2N)^{1/2}+\delta\|A\|\sqrt{N}\\&\le 5\delta\|A\|\sqrt{N}\,.\end{aligned}$$

This estimate holds for any $I\subset\{1,\dots,N\}$. Hence, by (15.10),

$$\mathbb{E}\|A(\omega)\|_{2\to1}\le 20\delta\|A\|\sqrt{N}\,.$$

In particular,

$$\mathbb{P}\{\|A(\omega)\|_{2\to1}\le 50\delta\|A\|\sqrt{N}\}>\frac{1}{2}\,.$$

Since clearly

$$\mathbb{P}\left\{\sum_{i=1}^N\xi_i\ge\frac{\delta N}{2}\right\}>\frac{1}{2}\,,$$

there exists ω in the intersection of these two events; $\sigma=\{i\le N\,;\ \xi_i(\omega)=1\}$ then fulfills the conclusion of Proposition 15.10. The proof is complete.

15.5. Embedding subspaces of L_p into ℓ_p^N

This section is devoted to the following problem of local theory of Banach spaces. Given an n-dimensional subspace E of $L_p=L_p([0,1],dt)$, $1\le p<\infty$, and $\eta>0$, what is the smallest $N=N(E,\eta)$ such that there is a subspace F of ℓ_p^N with $d(E,F) \leq 1+\eta$? Here $d(E,F)$ is the Banach-Mazur distance between Banach spaces defined as the infinum of $\|U\|\,\|U^{-1}\|$ over all isomorphisms $U: E \to F$ (more precisely, the logarithm of d is a distance); $d(E,F)=1$ if E and F are isometric. Partial results in case $E=\ell_2^n$ or $E=\ell_p^n$, $1 \leq p < 2$, follow from the general study on Dvoretzky's theorem and the stable type in Chapter 9.

The results we present here are taken from [B-L-M] and [Ta15]. They are at this point different when $p=1$ or $p>1$ so that we distinguish these cases in two separate statements. In the first one, we need the concept of K-convexity constant [Pi11], [Pi16]. If E is a (finite dimensional) Banach space, denote by $K(E)$ the norm of the natural projection from $L_2(E)=L_2(\Omega,\mathcal{A},\mathbb{P};E)$ onto the span of the functions $\sum_i \varepsilon_i x_i$ where (ε_i) is a Rademacher sequence on $(\Omega,\mathcal{A},\mathbb{P})$. Thus, if $f \in L_2(E)$, we have

$$\Big\|\sum_i \varepsilon_i \mathbb{E}(\varepsilon_i f)\Big\|_{L_2(E)} \leq K(E)\|f\|_{L_2(E)} .$$

It has been noticed in [TJ2] that then the same inequality holds when (ε_i) is replaced by a standard Gaussian sequence (g_i). We use this freely below. It is known further and due to G. Pisier [Pi8] that $K(E) \leq C(\log(n+1))^{1/2}$ if E is a subspace of dimension n of L_1 (C numerical).

We can now state the two main results.

Theorem 15.12. There are numerical constants η_0 and C such that for any n-dimensional subspace E of L_1 and any $0<\eta<\eta_0$, we have

$$N(E,\eta) \leq CK(E)^2\eta^{-2}n .$$

In particular

$$N(E,\eta) \leq C\eta^{-2}n\log(n+1) .$$

Theorem 15.13. Let $1<p<\infty$. There are numerical constants η_0 and C such that for any n-dimensional subspace E of L_p and any $0<\eta<\eta_0$, we have

$$N(E,\eta) \leq Cp\eta^{-2}n^{p/2}(\log n)^2\log(\eta^{-1}n)$$

if $p \geq 2$ and

$$N(E,\eta) \leq C\eta^{-2}n(\log n)^2 \max\left(\frac{1}{p-1},\ \log(\eta^{-1}n)\right)$$

if $1 < p \leq 2$.

Common to the proofs of these theorems is a basic random choice argument. We first develop it in general.

We agree to denote below by $\|\cdot\|_p = \|\cdot\|_{L_p(\mu)}$, $1 \leq p < \infty$, the norm of $L_p(S,\Sigma,\mu)$ where μ is a σ-finite measure on (S,Σ) (hopefully with no confusion when (S,Σ,μ) varies). We first observe that, given $\eta > 0$, a finite dimensional subspace E of $L_p = L_p([0,1],dt)$ is always at a distance (in the sense of Banach-Mazur) less than $1+\eta$ of a subspace of ℓ_p^M for some (however large) M. This can be seen in a number of ways, e.g. using that $\|f - E^k f\|_p \to 0$ where E^k denotes the conditional expectation with respect to the k-th dyadic subalgebra of $[0,1]$.

From this observation, we concentrate on subspaces E of dimension n of ℓ_p^M for some fixed M. To embed E into $\ell_p^{M_1}$ where M_1 is of the order $M/2$, we will, for each coordinate, flip a coin and disregard the coordinate if "head" comes up. More formally, consider a sequence $\varepsilon = (\varepsilon_i)$ of Rademacher random variables and consider the operator $U_\varepsilon : \ell_p^M \to \ell_p^M$ given by: if $x = (x_i)_{i \leq M} \in \ell_p^M$, $U_\varepsilon(x) = ((1+\varepsilon_i)^{1/p} x_i)_{i \leq M}$. We will give conditions under which U_ε is, with large probability, almost an isometry when restricted to E. We note that $U_\varepsilon(\ell_p^M)$ is isometric to $\ell_p^{M_1}$ where $M_1 = \operatorname{Card}\{i \leq M\,;\ \varepsilon_i = 1\}$ and, with probability $\geq 1/2$, we have $M_1 \leq M/2$. For x in ℓ_p^M, consider the random variable

$$Z_x = \|U_\varepsilon(x)\|_p^p - \|x\|_p^p .$$

We have

$$Z_x = \sum_{i=1}^M (1+\varepsilon_i)|x_i|^p - \sum_{i=1}^M |x_i|^p = \sum_{i=1}^M \varepsilon_i |x_i|^p .$$

Let E_1 denote the unit ball of E and set

$$A_E = \sup_{x \in E_1} \left| \sum_{i=1}^M \varepsilon_i |x_i|^p \right| .$$

The restriction R_ε of U_ε to E satisfies $\|R_\varepsilon\| \leq (1+A_E)^{1/p}$, $\|R_\varepsilon^{-1}\| \leq (1-A_E)^{1/p}$ so that, when $A_E \leq 1/2$,

$$d(E, U_\varepsilon(E)) \leq 1 + \frac{12}{p} A_E .$$

Since $\mathbb{P}\{A_E \leq 3\mathbb{E}A_E\} \geq 2/3$, we have therefore obtained:

Proposition 15.14. Let E be a subspace of ℓ_p^M, $1 \leq p < \infty$, of dimension n and let A_E be as before. If $\mathbb{E}A_E \leq 1/6$, there exists $M_1 \leq M/2$ and a subspace H of $\ell_p^{M_1}$ such that

$$d(E,H) \leq 1 + \frac{36}{p}\mathbb{E}A_E \,.$$

In order to apply successfully this result, one must therefore ensure that $\mathbb{E}A_E$ is small. The next lemma is one of the useful tools in this order of ideas. It is an immediate consequence of a result of D. Lewis [Lew].

Lemma 15.15. Let E be an n-dimensional subspace of ℓ_p^M, $1 \leq p < \infty$. Then, one can find a probability measure μ on $\{1,\ldots,M\}$ and a subspace F of $L_p(\mu)$ isometric to E which admits a basis $(\psi_j)_{j \leq n}$ orthogonal in $L_2(\mu)$ such that $\sum\limits_{j=1}^{n} \psi_j^2 = 1$ and $\|\psi_j\|_2 = n^{-1/2}$ for all $j \leq n$.

In this lemma, if we split each atom of μ of mass $a \geq 2/M$ in $[aM/2]+1$ equal pieces, we can assume that each atom of μ has mass $\leq 2/M$ and that μ is now supported by $\{1,\ldots,M'\}$ where M' is an integer less than $3M/2$. Also we can assume that $\lambda_i = \mu(\{i\}) > 0$ for all $i \leq M'$. *We always use Lemma 15.15 in this convenient form below.*

Let F be as before and F_1 be the unit ball of $F \subset L_p(\mu)$. Our main task in the proof of both theorems will be to show that

$$\Lambda_E = \mathbb{E} \sup_{x \in F_1} \left| \sum_{i=1}^{M'} \lambda_i \varepsilon_i |x(i)|^p \right| \tag{15.11}$$

can be nicely estimated (recall that $\lambda_i = \mu(\{i\})$). If we denote by $G \subset \ell_p^{M'}$ the image of F by the map

$$L_p(\mu) \to \ell_p^{M'}$$
$$x = (x(i))_{i \leq M'} \to (\lambda_i^{1/p} x(i))_{i \leq M'} \,,$$

then $\mathbb{E}A_G = \Lambda_E$ and G is isometric to F. Applying Proposition 15.14 to G in $\ell_p^{M'}$, we see that if $\Lambda_E \leq 1/6$, exists a subspace H of $\ell_p^{M_1}$ with $M_1 \leq M'/2 \leq 3M/4$ such that

$$d(E,H) = d(F,H) = d(G,H) \leq 1 + \frac{36}{p}\Lambda_E \,. \tag{15.12}$$

This is the procedure we will use to establish Theorems 15.12 and 15.13. When $p = 1$, we estimate Λ_E by the K-convexity constant $K(E)$ of E while, when $p > 1$, we use Dudley's entropy majorizing result (Theorem 11.17) to suitably bound Λ_E.

We turn to the proof of Theorem 15.12.

Proof of Theorem 15.12. The main point is the following proposition. In this proof, $p = 1$.

Proposition 15.16. Let Λ_E be as obtained in (15.11). Then

$$\Lambda_E \leq CK(E) \left(\frac{n}{M}\right)^{1/2}$$

for some numerical constant C.

Proof. By an application of the comparison properties for Rademacher averages (Theorem 4.12) and comparison with Gaussian averages (4.8),

$$\Lambda_E \leq 2\mathbb{E} \sup_{x \in F_1} \left| \sum_{i=1}^{M'} \lambda_i \varepsilon_i x(i) \right| \leq \sqrt{2\pi}\mathbb{E} \sup_{x \in F_1} \left| \sum_{i=1}^{M'} \lambda_i g_i x(i) \right| .$$

(We may of course also invoke the Gaussian comparison theorems; (Theorem 4.12) is however simpler.)

We denote by $\langle \cdot , \cdot \rangle$ the scalar product in $L_2(\mu)$. There exists f in $L_\infty(\Omega, \mathcal{A}, \mathbb{P}; F) = L_\infty(F)$ of norm 1 such that

$$\sup_{x \in F_1} \langle \sum_{j=1}^{n} g_j \psi_j \, , \; x \rangle = \langle \sum_{j=1}^{n} g_j \psi_j \, , \; f \rangle \, .$$

Thus

$$\mathbb{E} \sup_{x \in F_1} \langle \sum_{j=1}^{n} g_j \psi_j \, , \; x \rangle = \sum_{j=1}^{n} \langle \psi_j \, , \; \mathbb{E}(g_j f) \rangle \, .$$

Setting $y_j = \mathbb{E}(g_j f)$, we have by definition of $K(F) = K(E)$,

$$\left\| \sum_{j=1}^{n} g_j y_j \right\|_{L_2(F)} \leq K(E) \, .$$

Now

$$\left\| \sum_{j=1}^{n} g_j y_j \right\|_{L_2(F)}^2 = \mathbb{E} \int \left| \sum_{j=1}^{n} g_j y_j(t) \right|^2 d\mu(t) = \int \left(\sum_{j=1}^{n} y_j(t)^2 \right) d\mu(t) \, .$$

Since $\sum_{j=1}^{n} \psi_j^2 = 1$, we can write by Cauchy-Schwarz that

$$\sum_{j=1}^{n} \langle \psi_j, y_j \rangle = \int \left(\sum_{j=1}^{n} \psi_j(t) y_j(t) \right) d\mu(t)$$
$$\leq \int \left(\sum_{j=1}^{n} y_j(t)^2 \right)^{1/2} d\mu(t) \leq K(E).$$

Hence we have obtained that

$$\mathbb{E} \sup_{x \in F_1} \langle \sum_{j=1}^{n} g_j n^{1/2} \psi_j, \ x \rangle \leq n^{1/2} K(E). \tag{15.13}$$

Set $v_i = I_{\{i\}}$ so that $(\lambda_i^{-1/2} v_i)_{i \leq M'}$ is an orthonormal basis of $L_2(\mu)$ ($\lambda_i = \mu(\{i\})$). Since $(n^{1/2}\psi_j)_{j \leq n}$ is an orthonormal basis of $F \subset L_2(\mu)$, the rotational invariance of Gaussian distribution indicates that

$$\mathbb{E} \sup_{x \in F_1} \langle \sum_{j=1}^{n} g_j n^{1/2} \psi_j, \ x \rangle = \mathbb{E} \sup_{x \in F_1} \langle \sum_{i=1}^{M'} \lambda_i^{-1/2} g_i v_i, \ x \rangle. \tag{15.14}$$

Hence (15.13) reads as

$$\mathbb{E} \sup_{x \in F_1} \langle \sum_{i=1}^{M'} \lambda_i^{-1/2} g_i v_i, x \rangle \leq n^{1/2} K(E).$$

Finally, since $\lambda_i \leq 2/M$ for all i, by the contraction principle,

$$\mathbb{E} \sup_{x \in F_1} \left| \sum_{i=1}^{M'} \lambda_i g_i x(i) \right| = \mathbb{E} \sup_{x \in F_1} \langle \sum_{i=1}^{M'} g_i v_i, \ x \rangle$$
$$\leq \max_{i \leq M'} \lambda_i^{1/2} \mathbb{E} \sup_{x \in F_1} \langle \sum_{i=1}^{M'} \lambda_i^{-1/2} g_i v_i, \ x \rangle$$
$$\leq \left(\frac{2n}{M} \right)^{1/2} K(E).$$

Proposition 15.16 follows (with $C = 2\sqrt{\pi}$).

We establish Theorem 15.12 by a simple iteration of the preceding proposition. We show that, for $\eta \leq C^{-1}$, there exists a subspace H of ℓ_1^N such that $d(E, H) \leq 1 + C\eta$ and $N \leq CK(E)^2 \eta^{-2} n$ where C is numerical. Note that for any Banach spaces X, Y,

$K(Y) \leq d(X,Y)^2 K(X)$. As we have seen, given $\eta > 0$, there exists M and a subspace H^0 of ℓ_1^M with $d(E, H^0) \leq 1+\eta$. In particular, $K(H^0) \leq (1+\eta)^2 K(E)$. Let C_1 be the constant of Proposition 15.16 and set $C_2 = 288 C_1$ and assume that $0 < \eta \leq 10^{-2}$. We can assume further that $M \geq C_2^2 K(E)^2 \eta^{-2} n$ otherwise we simply take $H = H^0$. By (15.12), we construct by induction a sequence of integers $(M_j)_{j\geq 0}$, $M_0 = M$, satisfying $M_{j+1} \leq 3M_j/4$ and subspaces H^j, $j \geq 0$, of $\ell_1^{M_j}$ such that, for all $j \geq 0$,

$$d(H^j, H^{j+1}) \leq 1 + C_2 K(E) \left(\frac{n}{M_j}\right)^{1/2}$$

and we stop at j_0, the first for which $M_{j_0} \leq C_2^2 K(E)^2 \eta^{-2} n$. Indeed, suppose that $j < j_0$ and that $M_0, \dots M_j$, $H^0, \dots, H^j$ have been constructed. Note that

$$\begin{aligned} \prod_{\ell=0}^{j} \left(1 + C_2 K(E) \left(\frac{n}{M_\ell}\right)^{1/2}\right) &\leq \prod_{\ell=0}^{j} \left(1 + C_2 K(E) \left(\frac{3}{4}\right)^{(j-\ell)/2} \left(\frac{n}{M_j}\right)^{1/2}\right) \\ &\leq e^{10\eta} \leq 2 \end{aligned} \tag{15.15}$$

(since $\eta \leq 10^{-2}$). If follows in particular that

$$K(H^j) \leq (1+\eta)^2 e^{10\eta} K(E) \leq 8K(E),$$

and thus, by Proposition 15.16,

$$\Lambda_{H^j} \leq C_1 K(H^j) \left(\frac{n}{M_j}\right)^{1/2} \leq 8 C_1 K(E) \left(\frac{n}{M_j}\right)^{1/2} \leq \frac{1}{6}$$

so that, by (15.12), there exist $M_{j+1} \leq 3M_j/4$ and a subspace H^{j+1} of $\ell_1^{M_{j+1}}$ with

$$\begin{aligned} d(H^j, H^{j+1}) &\leq 1 + 36 C_1 K(H^j) \left(\frac{n}{M_j}\right)^{1/2} \\ &\leq 1 + C_2 K(E) \left(\frac{n}{M_j}\right)^{1/2}. \end{aligned}$$

This proves the induction procedure. The result then follows. We have that

$$d(H^0, H^{j_0}) \leq \prod_{j=0}^{j_0-1} \left(1 + C_2 K(E) \left(\frac{n}{M_j}\right)^{1/2}\right) \leq e^{10\eta}$$

(by (15.15)) and thus $d(E, H^{j_0}) \le (1+\eta)e^{10\eta}$. The proof of Theorem 15.12 is therefore complete.

We now turn to the case $p > 1$ and the proof of Theorem 15.13.

Proof of Theorem 15.13. We first need the following easy consequences of Lemma 15.15.

Lemma 15.17. Let F be as given by Lemma 15.15. If $x \in F$, then

$$\max_{i \le M'} |x(i)| \le n^{\max(1/p, 1/2)} \|x\|_p .$$

Further, for all $r \ge 1$,

$$\mathbb{E} \left\| \sum_{j=1}^{n} g_j \psi_j \right\|_r \le C r^{1/2}$$

where C is numerical.

Proof. If $x = \sum_{j=1}^{n} \alpha_j \psi_j$, $\|x\|_2 = n^{-1/2} \left(\sum_{j=1}^{n} \alpha_j^2 \right)^{1/2}$, so that, for all $i \le M'$,

$$|x(i)| \le \left(\sum_{j=1}^{n} \alpha_j^2 \right)^{1/2} \left(\sum_{j=1}^{n} \psi_j(i)^2 \right)^{1/2} = n^{1/2} \|x\|_2 . \tag{15.16}$$

This already gives the result for $p \ge 2$ since then $\|x\|_2 \le \|x\|_p$. When $p \le 2$,

$$\|x\|_2^2 \le \max_i |x(i)|^{2-p} \|x\|_p^p \le n^{1-p/2} \|x\|_2^{2-p} \|x\|_p^p$$

and thus $\|x\|_2 \le n^{(1/p)-(1/2)} \|x\|_p$. By (15.16), the first claim in Lemma 15.17 is established in the case $p \le 2$ as well. The second claim immediately follows from Jensen's inequality and the fact that $\sum_{j=1}^{n} \psi_j^2 = 1$.

Recall that $\lambda_i = \mu(\{i\}) > 0$, $\lambda_i \le 2/M$, $i \le M'$. Let $J = \{i \le M' ;\ \lambda_i > 1/M^2\}$. Then, by Lemma 15.18,

$$\mathbb{E} \sup_{x \in F_1} \left| \sum_{i \notin J} \lambda_i \varepsilon_i |x(i)|^p \right| \le \sum_{i \notin J} \lambda_i n^{\max(1, p/2)} \le \frac{3}{2M} n^{\max(1,p/2)} .$$

Hence, by the triangle inequality and the contraction principle,

$$\Lambda_E \le \frac{3}{2M} n^{\max(1,p/2)} + \left(\frac{2}{M} \right)^{1/2} \mathbb{E} \sup_{x \in F_1} \left| \sum_{i \in J} \lambda_i^{1/2} \varepsilon_i |x(i)|^p \right| . \tag{15.17}$$

Our aim is to study the process $\sum_{i\in J} \lambda_i^{1/2}\varepsilon_i|x(i)|^p$, $x \in F_1$, with associated pseudo-metric

$$\delta(x,y) = \left(\sum_{i\in J} \lambda_i(|x(i)|^p - |y(i)|^p)^2\right)^{1/2} ,$$

and use Dudley's entropy integral to bound it appropriately. We therefore need to evaluate various entropy numbers. J being fixed as before, for $x = (x(i))_{i\le M'}$ in $L_p(\mu)$, we set

$$\|x\|_J = \max_{i\in J} |x(i)|$$

and agree to denote by B_J the corresponding unit ball. In these notations, let us first observe the following. If $x, y \in F$, $\|x\|_p$, $\|y\|_p \le 1$, then

(15.18) $$\delta(x,y) \le 2pn^{(p-2)/4}\|x-y\|_J \qquad \text{if } p \ge 2$$

and

(15.19) $$\delta(x,y) \le 6\|x-y\|_J^{p/2} \qquad \text{if } p \le 2 .$$

For a proof observe that if $a, b \ge 0$,

$$a^p - b^p \le p(a^{p-1} + b^{p-1})|a-b| .$$

Thus

$$\delta(x,y)^2 \le 2p^2 \sum_{i\in J} \lambda_i \max(|x(i)|^{2p-2} ,\ |y(i)|^{2p-2})|x(i)-y(i)|^2 .$$

If $p \ge 2$, we proceed as follows. For all i , $|x(i)|^{2p-2} \le \max_j |x(j)|^{p-2}|x(i)|^p$ and, if $i \in J$, $|x(i)-y(i)|^2 \le \|x-y\|_J^2$. Hence, since $\|x\|_p$, $\|y\|_p \le 1$, using Lemma 15.17,

$$\delta(x,y)^2 \le 2p^2n^{(p-2)/2}\|x-y\|_J^2 \sum_{i\in J} \lambda_i(|x(i)|^p + |y(i)|^p)$$

and (15.18) is satisfied. When $p \le 2$, Hölder's inequality with $\alpha = p/(2p-2)$ and $\beta = p/(2-p)$ shows that

$$\begin{aligned}
&\sum_{i\in J} \lambda_i \max(|x(i)|^{2p-2}, |y(i)|^{2p-2})|x(i)-y(i)|^{2-p} \\
&\quad \le \left(\sum_{i\in J} \lambda_i \max(|x(i)|^p, |y(i)|^p)\right)^{(2p-2)/p} \left(\sum_{i\in J} \lambda_i |x(i)-y(i)|^p\right)^{(2-p)/p} \\
&\quad \le (\|x\|_p^p + \|y\|_p^p)^{(2p-2)/p}\|x-y\|_p^{2-p} \le 4 .
\end{aligned}$$

This yields (15.19).

The following proposition is the basic entropy estimate which will be required for the case $p \geq 2$. It is based on the dual version of Sudakov's minoration ((3.15)) which will also be used for $p \leq 2$. For every $r \geq 1$, denote by $B_{F,r}$ the unit ball of F considered as a subspace of $L_r(\mu)$ (in particular $B_{F,p} = F_1$).

Proposition 15.18. Let $p \geq 2$ and let $F \subset L_p(\mu)$ be as given by Lemma 15.15. Then, for every $u > 0$,

$$\log N(B_{F,p}, uB_J) \leq \frac{C}{u^2} n \log M$$

where C is numerical.

Proof. Let $r > 1$ and r' be the conjugate of r. By (3.15), for some constant C and all $u > 0$,

$$u(\log N(B_{F,2}, uB_{F,r}))^{1/2} \leq C\mathbb{E} \sup_{x \in B_{F,r'}} \left| \sum_{i=1}^{M'} \lambda_i^{-1/2} g_i x(i) \right| .$$

As in (15.14), since $(n^{1/2}\psi_j)_{j \leq n}$ is an orthonormal basis of F, the rotational invariance of Gaussian measures implies that the expectation on the right of the preceding inequality is equal to

$$\mathbb{E} \sup_{x \in B_{F,r'}} \langle \sum_{j=1}^{n} g_j n^{1/2} \psi_j, x \rangle = n^{1/2} \mathbb{E} \left\| \sum_{j=1}^{n} g_j \psi_j \right\|_r .$$

Therefore, the second assertion in Lemma 15.17 shows that, for some numerical constant C and all $r > 1$ and $u > 0$

$$\log N(B_{F,2}, uB_{F,r}) \leq \frac{C}{u^2} rn . \tag{15.20}$$

It is then easy to conclude the proof of Proposition 15.18. Since $p \geq 2$, $B_{F,p} \subset B_{F,2}$. One the other hand, if $x \in B_{F,r}$ for some r,

$$\sum_{i=1}^{M'} \lambda_i |x(i)|^r \leq 1$$

and thus $|x(i)|^r \leq \lambda_i^{-1} \leq M^2$ if $i \in J$. Therefore, $B_{F,r} \subset eB_J$ if $r = \log M^2$. Hence, for this choice of r,

$$N(B_{F,p}, uB_J) \leq N(B_{F,2}, \frac{u}{e} B_{F,r})$$

and the conclusion follows from (15.20).

Let us now show how to deduce the portion $p \geq 2$ of Theorem 15.13. By (15.18),

$$\int_0^\infty (\log N(F_1, \delta; u))^{1/2} du \leq 2pn^{(p-2)/4} \int_0^{2\sqrt{n}} (\log N(B_{F,p}, uB_J))^{1/2} du .$$

Then note that since $n^{-1/2} B_{F,p} \subset B_J$ (Lemma 15.17), by simple volume considerations,

$$N(B_{F,p}, uB_J) \leq N(B_{F,p}, un^{-1/2} B_{F,p}) \leq \left(1 + \frac{2\sqrt{n}}{u}\right)^n .$$

Together with Proposition 15.18, we get that

$$\begin{aligned}\int_0^{2\sqrt{n}} (\log N(B_{F,p}, uB_J))^{1/2} du \leq & \int_0^1 \left(n \log\left(1 + \frac{2\sqrt{n}}{u}\right)\right)^{1/2} du \\ & + \int_1^{2\sqrt{n}} (Cn \log M)^{1/2} \frac{du}{u} .\end{aligned}$$

It follows that for some numerical constant C,

$$\int_0^\infty (\log N(F_1, \delta; u))^{1/2} du \leq [Cp^2 n^{p/2} (\log n)^2 \log M]^{1/2} .$$

Therefore, by Dudley's majoration theorem (Theorem 11.17) and by (15.17), if $n \leq M$,

$$\Lambda_E \leq \left[Cp^2 \frac{n^{p/2}}{M} (\log n)^2 \log M\right]^{1/2} .$$

An iteration on the basis of (15.12) similar to the one performed in the proof of Theorem 15.12 (but simpler) then yields the conclusion in case $p \geq 2$.

We turn next to the case $1 < p \leq 2$. It relies on the following analog of Proposition 15.18 that will basically be obtained by duality (after an interpolation argument).

Proposition 15.19. Let $1 < p \leq 2$ and let $F \subset L_p(\mu)$ be as given by Lemma 15.15. Then, for every $0 < u \leq 2n^{1/p}$,

$$\log N(B_{F,p}, uB_J) \leq \frac{C}{u^p} n \max\left(\frac{1}{p-1}, \log M\right)$$

where C is numerical.

Proof. Let $q = p/p-1$ be the conjugate of p. Let $v \le 2n^{(2-p)/2p}$. Let further $h \ge q$ and define θ, $0 \le \theta \le 1$, by

$$\frac{1}{q} = \frac{1-\theta}{2} + \frac{\theta}{h}.$$

For x, y in $B_{F,2}$, by Hölder's inequality,

$$\|x-y\|_q \le \|x-y\|_2^{1-\theta}\|x-y\|_h^{\theta} \le 2^{1-\theta}\|x-y\|_h^{\theta}$$

so that, if $\|x-y\|_h \le (v/2)^{1/\theta}$, then $\|x-y\|_q \le v$. Hence, (15.20) for $r = h$ yields

$$\log N(B_{F,2}, vB_{F,q}) \le \log N\left(B_{F,2}, \left(\frac{v}{2}\right)^{1/\theta} B_{F,h}\right) \le Chn\left(\frac{v}{2}\right)^{-2/\theta}.$$

Here and below, C denotes some numerical constant, not necessarily the same each time it appears. Since

$$\frac{1}{\theta} = \frac{p}{2-p} - \frac{2p}{(2-p)h},$$

it follows that (recall $v \le 2n^{(2-p)/2p}$),

$$\log N(B_{F,2}, vB_{F,q}) \le Chn\, n^{2/h}\left(\frac{v}{2}\right)^{-2p/(2-p)}.$$

Let us then choose

$$h = h(n) = \max(q, \log n)$$

so that

$$\log N(B_{F,2}, vB_{F,q}) \le Chn\left(\frac{v}{2}\right)^{-2p/(2-p)}.$$

The proof of (3.16) of the equivalence of Sudakov and dual Sudakov inequalities indicates that we have in the same way

$$\log N(B_{F,p}, vB_{F,2}) \le Chn\left(\frac{v}{2}\right)^{-2p/(2-p)} \tag{15.21}$$

where the constant C may be chosen independent of p.

As in the proof of Proposition 15.18, if $r = \log M^2$, $B_{F,r} \subset eB_J$. Thus, using furthermore the properties of entropy numbers, for $u, v > 0$,

$$\log N(B_{F,p}, uB_J) \le \log N(B_{F,p}, vB_{F,2}) + \log N\left(B_{F,2}, \frac{u}{ev}B_{F,r}\right).$$

Let us choose

$$v = 2^{p/2} \left(\frac{u}{e}\right)^{(2-p)/2}$$

(so that $v \le 2n^{(2-p)/2p}$ since $u \le 2n^{1/p}$). By (15.21) applied to the first term on the right of the preceding inequality and (15.20) to the second (for $r = \log M^2$), we get that, for some numerical constant C (recall that $1 < p \le 2$),

$$\log N(B_{F,p}, uB_J) \le Cu^{-p} n(h + \log M) .$$

Since we may assume that $n \le M$, by definition of $h = h(n)$, the proof of Proposition 15.19 is complete.

The proof of Theorem 15.13 for $1 < p \le 2$ is then completed exactly as before for $p \ge 2$. The preceding proposition together with Dudley's theorem ensures similarly that

$$\Lambda_E \le \left[C \frac{n}{M} (\log n)^2 \max \left(\frac{1}{p-1}, \log M \right) \right]^{1/2} .$$

An iteration on the basis of (15.12) concludes the proof in the same way.

15.6. Majorizing measures on ellipsoids

Consider a sequence (y^n) is a Hilbert space H such that $|y^n| \le (\log(n+1))^{-1/2}$ for all n and set $T = \{y^n ;\, n \ge 1\}$. We know (cf. Section 12.3) that the canonical Gaussian process $X_t = \sum_i g_i t_i$, $t = (t_i) \in T \subset H$, is almost surely bounded. The same holds of course if X is indexed by $\mathrm{Conv} T$. In particular, by Theorem 12.8, there is a majorizing measure on $\mathrm{Conv} T$. However, no explicit construction is known. The difficulty is that this majorizing measure should depend on the relative position of the y^n 's, not just on their lengths. More generally, we can ask: given a set A in H with a majorizing measures, construct a majorizing measure on $\mathrm{Conv} A$. Of a similar nature: given sets A_i , $i \le n$, with majorizing measure, construct a majorizing measure on $\frac{1}{n} \sum_{i=1}^{n} A_i$.

As an example of construction, we construct in this section explicit majorizing measures on ellipsoids. The construction is based on the evaluation of the entropy integral for products of balls.

Let (a_i) be a sequence of positive numbers with $\sum_i a_i^2 \le 1$. In $H = \ell_2$, consider the ellipsoid

$$\mathcal{E} = \left\{ x \in H ;\ \sum_i \frac{x_i^2}{a_i^2} \le 1 \right\} .$$

Let (g_i) be an orthogaussian sequence. Since

$$\mathbb{E}\sup_{x\in\mathcal{E}}\left|\sum_i g_i x_i\right|^2 = \mathbb{E}\left(\sum_i g_i^2 a_i^2\right) = \sum_i a_i^2 \le 1\,,$$

we know by Theorem 12.8 that there is a majorizing measure (for the function $(\log(1/x))^{1/2}$) on $\mathcal{E}$ with respect to the ℓ_2 metric (that can actually be made into a continuous majorizing measure). One may wonder for an explicit description of such a majorizing measure. This is the purpose of this section.

What will actually be explicit is the following. Let I_k , $k \ge 0$, be the disjoint sets of integers given by

$$I_k = \{i\,;\ 2^{-k-1} < a_i \le 2^{-k}\}\,.$$

Note that since $\sum\limits_i a_i^2 \le 1$, we have that $\sum\limits_k 2^{-2k}\ \mathrm{Card} I_k \le 4$. Consider then the ellipsoid

$$\mathcal{E}' = \left\{x\in H\,;\ \sum_k 2^{2k}\sum_{i\in I_k} x_i^2 \le 1\right\}\,.$$

We will exhibit a probability measure m on $\sqrt{3}\mathcal{E}'$ such that, for all x in $\mathcal{E}$,

$$\int_0^\infty \left(\log\frac{1}{m(B(x,\varepsilon))}\right)^{1/2} d\varepsilon \le C \tag{15.22}$$

where C is numerical and $B(x,\varepsilon)$ is the (Hilbertian) ball of center x and radius $\varepsilon > 0$ in $\sqrt{3}\mathcal{E}'$. Since $\mathcal{E}\subset\mathcal{E}'$, we can then use Lemma 11.9 to obtain a majorizing measure on $\mathcal{E}$ (however less explicit). It can be shown further to be a continuous majorizing measure. We leave this to this interested reader.

Let U be the set of all sequences of integers $\overline{n} = (n_k)_{k\ge0}$ such that $n_k \le k$ for all k and $\sum\limits_k 2^{-n_k}\ le3$. For such a sequence $\overline{n} = (n_k)$, let $\Pi(\overline{n})$ be the product of balls

$$\Pi(\overline{n}) = \left\{x\in H\,;\ \forall k\,,\ \sum_{i\in I_k} 2^{2k}x_i^2 \le 2^{-n_k}\right\}\,.$$

The family $\{\Pi(\overline{n})\,;\ \overline{n}\in U\}$ covers $\mathcal{E}'$. Indeed, given x in $\mathcal{E}'$ and k an integer, let m_k be the integer such that

$$2^{-m_k-1} < \sum_{i\in I_k} 2^{2k}x_i^2 \le 2^{-m_k}\,.$$

Then $x \in \Pi(\overline{n})$ with $\overline{n} = (n_k)$, $n_k = \min(k, m_k)$. Note conversely that $\Pi(\overline{n}) \subset \sqrt{3}\mathcal{E}'$ for every $\overline{n}$ in U. The main step in this construction will be to show that the products of balls $\Pi(\overline{n})$ satisfy the entropy conditions uniformly over all $\overline{n}$ in U; that is, for some constant C and all $\overline{n}$ in U,

$$\sum_p 2^{-p} (\log N(\Pi(\overline{n}), \ 2^{-p} B_2))^{1/2} \le C \tag{15.23}$$

(where B_2 is the unit ball of ℓ_2).

Let therefore $\overline{n} = (n_k)$ be a fixed element of U. Set for all k, $\ell_k = n_k + 2k$ and consider thus the product of balls

$$\Pi = \Pi(\overline{n}) = \left\{ x \in H \, ; \, \forall k \, , \ \sum_{i \in I_k} x_i^2 \le 2^{-\ell_k} \right\} .$$

Note that since $\sum_k 2^{-2k} \operatorname{Card} I_k \le 4$ and $\sum_k 2^{-n_k} \le 3$, by Cauchy-Schwartz and definition of ℓ_k,

$$\sum_k (2^{-\ell_k} \operatorname{Card} I_k)^{1/2} \le 2\sqrt{3} \, . \tag{15.24}$$

In addition, we may assume that the sequence $(2^{\ell_k} \operatorname{Card} I_k)$ is *increasing.* This property will allow to easily select the balls with small entropy in the product Π.

For any integer p, let $k(p)$ be the smallest such that

$$\sum_{k \ge k(p)} 2^{-\ell_k} \le 2^{-2p-2} \, .$$

Then

$$N(\Pi, \ 2^{-p} B_2) \le N(\Pi_p, \ 2^{-p-1} B_2)$$

where Π_p is the product $\Pi_{k<k(p)} B(k)$ of the finite dimensional Euclidean balls

$$B(k) = \left\{ (x_i)_{i \in I_k} \, ; \ \sum_{i \in I_k} x_i^2 \le 2^{-\ell_k} \right\} .$$

ΞWe agree further to denote in the same way by B_2 the unit ball of ℓ_2 and of the finite dimensional space ℓ_2^N for the corresponding dimension (clear from the context). Let us now recall that, by volume considerations, for every $\varepsilon > 0$,

$$N(B(k), \ \varepsilon B_2) \le \left(1 + \frac{2 \cdot 2^{-\ell_k/2}}{\varepsilon} \right)^{\operatorname{Card} I_k} . \tag{15.25}$$

It is easily seen further that if $(\varepsilon_k)_{k<p}$ are positive numbers with $\sum\limits_{k<k(p)} \varepsilon_k^2 \le 2^{-2p-2}$, then

$$N(\Pi_p\,,\ 2^{-p-1}B_2) \le \prod_{k<k(p)} N(B(k)\,,\ \varepsilon_k B_2)\,. \tag{15.26}$$

Let us choose in this inequality

$$\varepsilon_k^2 = 2^{-\ell_k-4(p-j)-1}$$

where j is such that $k(j-1) \le k < k(j)$. We have that

$$\begin{aligned}
\sum_{k<k(p)} \varepsilon_k^2 &= \sum_{j\le p} 2^{-4(p-j)-1} \sum_{k(j-1)\le k<k(j)} 2^{-\ell_k} \\
&\le \sum_{j\le p} 2^{-4(p-j)-1} \sum_{k\ge k(j-1)} 2^{-\ell_k} \\
&\le \sum_{j\le p} 2^{2j-4p-1} \le 2^{-2p-2}
\end{aligned}$$

by definition of $k(j-1)$. We are thus in a position to apply (15.26). Together with (15.25), it yields that

$$\begin{aligned}
\log N(\Pi, 2^{-p}B_2) &\le \sum_{k<k(p)} \log N(B(k)\,,\ \varepsilon_k B_2) \\
&\le \sum_{j\le p} \left(\sum_{k(j-1)\le k<k(j)} \mathrm{Card} I_k \right) \log(2^{2(p-j)+2})\,.
\end{aligned}$$

It follows that for some numerical constant C

$$\sum_p 2^{-p}(\log N(\Pi, 2^{-p}B_2))^{1/2} \le C \sum_j 2^{-j} \left(\sum_{k(j-1)\le k<k(j)} \mathrm{Card} I_k \right)^{1/2}. \tag{15.27}$$

For every j, set

$$u_j = \sum_{k(j-1)\le k<k(j)} (2^{-\ell_k}\ \mathrm{Card} I_k)^{1/2}$$

so that $\sum\limits_j u_j \le 2\sqrt{3}$ by (15.24). Since $(2^{\ell_k}\ \mathrm{Card} I_k)$ is increasing,

$$\sum_{k(j-1)\le k<k(j)} \mathrm{Card} I_k \le u_j (2^{\ell_{k(j)-1}}\ \mathrm{Card} I_{k(j)-1})^{1/2}\,. \tag{15.28}$$

By definition of $k(j)$, $\sum\limits_{k>k(j)-1} 2^{-\ell_k} > 2^{-2j-2}$. Hence

$$\sum_{k(j)-1<k<k(j+1)} 2^{-\ell_k} > 2^{-2j-2} - 2^{-(2j+1)-2} = 2^{-2j-3} .$$

Therefore

$$\begin{aligned} 2^{-2j}(2^{\ell_{k(j)-1}} \ \mathrm{Card} I_{k(j)-1})^{1/2} &\le 2^3 \left(\sum_{k(j)-1<k<k(j+1)} 2^{-\ell_k} \right) (2^{\ell_{k(j)-1}} \ \mathrm{Card} I_{k(j)-1})^{1/2} \\ &\le 2^3 \sum_{k(j)-1<k<k(j+1)} (2^{-\ell_k} \ \mathrm{Card} I_k)^{1/2} \\ &\le 2^3 (u_j + u_{j+1}) \end{aligned}$$

where we have used again that $(2^{\ell_k} \ \mathrm{Card} I_k)$ is increasing. Therefore, by (15.28),

$$\begin{aligned} \sum_j 2^{-j} \left(\sum_{k(j-1)\le k<k(j)} \mathrm{Card} I_k \right)^{1/2} &\le \sum_j 2^{-2j} u_j (2^{\ell_{k(j)-1}} \ \mathrm{Card} I_{k(j)-1})^{1/2} \\ &\le 2^{3/2} \sum_j (u_j(u_j + u_{j+1}))^{1/2} . \end{aligned}$$

Since $\sum\limits_j u_j \le 2\sqrt{3}$ by (15.24), the announced claim (15.23) follows from Cauchy-Schwarz inequality and (15.27).

We now make use of this result (15.23) to exhibit a concrete majorizing measure satisfying (15.22). Recall we denote by U the set of all sequences of integers $\overline{n} = (n_k)_{k\ge 0}$, $n_k \ge 0$, such that $n_k \le k$ for all k and $\sum\limits_k 2^{-n_k} \le 3$. For every $j \ge 0$, let φ_j be the restriction map from $\mathbb{N}^{\mathbb{N}}$ onto $\mathbb{N}^{j+1}$, i.e. $\varphi_j((n_k)) = (n_k)_{k\le j}$. Set $U_j = \varphi_j(U)$ and note that $\mathrm{Card} U_j \le (j+1)!$. As for the elements of U, for every $\overline{n} = (n_k)_{k\le j}$ in U_j, we denote by $\Pi(\overline{n})$ the (finite dimensional) product of balls

$$\Pi(\overline{n}) = \left\{ x \in H \, ; \ \forall k \le j , \ \sum_{i\in I_k} 2^{2k} x_i^2 \le 2^{-n_k} , \ \forall k > j , \ \forall i \in I_k , \ x_i = 0 \right\} .$$

For every j and $\overline{n}$ in U_j, let then $(x(j,\overline{n}))$ be a family of points in $\Pi(\overline{n})$ of cardinality $N(\Pi(\overline{n}), 2^{-j}B_2)$, such that the Euclidean balls of center in $(x(j,\overline{n}))$ and radius 2^{-j} cover $\Pi(\overline{n})$. Consider then

$$m = \sum_{j\ge 0} \sum_{\overline{n}\in U_j} (2^{j+1}(j+1)! N(\Pi(\overline{n}), 2^{-j}B_2))^{-1} \delta_{x(j,\overline{n})}$$

where δ_x is point mass at x. Since $\Pi(\overline{n}) \subset \sqrt{3}\mathcal{E}'$ for every $\overline{n}$ in U, m is a measure on $\sqrt{3}\mathcal{E}'$ such that, by construction, $|m| \leq 1$. Now, if x is in $\mathcal{E} \subset \mathcal{E}'$, there exists $\overline{n}$ in U with $x \in \Pi(\overline{n})$. For every $j \geq 0$, one can find $x(j)$ in $\Pi(\varphi_j(\overline{n}))$ with $|x - x(j)| \leq 2^{-j}$. Then, by definition of m,

$$\begin{aligned} m(B(x, 2^{-j+1})) &\geq m(B(x(j), 2^{-j})) \\ &\geq (2^{j+1}(j+1)! N(\Pi(\varphi_j(\overline{n})), 2^{-j}B_2))^{-1} . \end{aligned}$$

Note that, for every $\varepsilon > 0$,

$$N(\Pi(\varphi_j(\overline{n})), \varepsilon B_2) \leq N(\Pi(\overline{n}), \varepsilon B_2) .$$

Hence, by (15.23) (and $\mathcal{E}' \subset B_2$), the announced result (15.22) is clearly established. This concludes the explicit construction of a majorizing measure on ellipsoids we wanted to present.

15.7. Cotype of the canonical injection $\ell_\infty^N \to L_{2,1}$

In this section, we develop a consequence of Sudakov's minoration for Rademacher averages (Theorem 4.15) to the cotype properties of the injection $\ell_\infty^N \to L_{2,1}$, a result due to S. J. Montgomery-Smith [MS1].

Let S be a compact metric space and $C(S)$ be the Banach space of continuous functions on S equipped with the supnorm $\|\cdot\|_\infty$. Consider an operator T from $C(S)$ to a Banach space B. We denote by $C_2(T)$ the Rademacher cotype 2 constant of T, i.e. $C_2(T)$ is the smallest number C such that

$$\left(\sum_i \|T(x_i)\|^2\right)^{1/2} \leq C\mathbb{E}\left\|\sum_i \varepsilon_i x_i\right\|_\infty$$

for all finite sequences (x_i) in $C(S)$. In particular, we have that

$$\left(\sum_i \|T(x_i)\|^2\right)^{1/2} \leq C \sup_{s \in S} \sum_i |x_i(s)| . \tag{15.29}$$

This is expressed by saying that T is $(2,1)$-summing (cf. [Pie]). Hence, if $T : C(S) \to B$ is of cotype 2, it is $(2,1)$-summing (i.e. satisfies (15.29)) with a $(2,1)$-summing constant less than $C_2(T)$.

We recall that for a measurable function x on a probability space (S, Σ, μ), we can define a quasi-norm by

$$\|x\|_{2,1} = \int_0^\infty (\mu(|x| > t))^{1/2} dt .$$

This quasi-norm defines the Lorentz space $L_{2,1}(\mu)$. It is known (and easy to see, cf. [S-W]) that $\|\cdot\|_{2,1}$ is equivalent to a norm and, for simplicity, we will use below that $\|\cdot\|_{2,1}$ as defined above behaves like a norm.

The canonical injection $C(S) \to L_{2,1}(\mu)$ (or $L_\infty(\mu) \to L_{2,1}(\mu)$), for any μ, is $(2,1)$-summing. Indeed, by Cauchy-Schwartz,

$$\begin{aligned}\|x\|_{2,1}^2 &= \left(\int_0^{\|x\|_\infty} (\mu(|x| > t))^{1/2} dt\right)^2 \\ &\leq \|x\|_\infty \int_0^\infty \mu(|x| > t) dt = \|x\|_\infty \int |x| d\mu .\end{aligned}$$

Hence, if (x_i) is a finite sequence in $C(S)$,

$$\sum_i \|x_i\|_{2,1}^2 \leq \sum_i \|x_i\|_\infty \int |x_i| d\mu \leq \sup_s \left(\sum_i |x_i(s)|\right)^2 .$$

[G. Pisier has shown conversely [Pi17] that an operator $T : C(S) \to B$ that is $(2,1)$-summing (i.e. satisfying (15.29)) factors through $L_{2,1}(\mu)$, i.e. there exists a probability measure μ on S such that for all x in $C(S)$,

$$\|T(x)\| \leq KC\|x\|_{2,1} ,$$

where K is a numerical constant.]

It is a natural question whether the canonical injection $C(S) \to L_{2,1}(\mu)$ is also of cotype 2. The answer to this question turns out to be no [Ta10]. We nonetheless have the following rather remarkable result of [MS1].

Theorem 15.20. Consider a probability space (S, Σ, μ) where Σ is assumed to be finite with N atoms. Then the cotype 2 constant $C_2(j)$ of the canonical injection $j : L_\infty(\mu) \to L_{2,1}(\mu)$ satisfies

$$C_2(j) \leq K \log(1 + \log N)$$

where K is numerical.

When μ is uniform measure on N points, it can be shown that $C_2(j) \geq K^{-1}(\log(1 + \log N))^{1/2}$. We conjecture that in Theorem 15.20 the term $\log(1+\log N)$ can be replaced by $(\log(1+\log N))^{1/2}$. From the subsequent proof, this would be the case if the conjecture on Rademacher processes presented at the end of Chapter 12 were true.

Proof. Our proof of Theorem 15.20 relies on Sudakov's minoration for Rademacher processes (Theorem 4.15) that can be reformulated as follows.

Lemma 15.21. Consider a finite sequence (x_i) in $L_\infty(S, \Sigma, \mu)$ where Σ is finite such that $\mathbb{E}\|\sum_i \varepsilon_i x_i\|_\infty \leq 1$. Then, for all k, one can find a subalgebra Σ' of Σ that has at most 2^{2^k} atoms and Σ'-measurable functions x'_i on S such that, for all i, $x_i = x'_i + y_i + z_i$ where

$$\mathbb{E}\left\|\sum_i \varepsilon_i x'_i\right\|_\infty \leq 1\,,$$

$$\left\|\sum_i |y_i|\right\|_\infty = \sup \sum_i |y_i(s)| \leq K_1$$

and

$$\left\|\sum_i z_i^2\right\|_\infty = \sup_s \sum_i z_i(s)^2 \leq K_1 2^{-k/2}$$

for some numerical constant K_1.

Proof. There is no loss of generality to assume that S is finite and that Σ is the algebra of all its subsets. Assume $(x_i) = (x_i)_{i \leq M}$. By Theorem 4.15, the subset $\{(x_i(s))_{i \leq M}\,;\ s \in S\}$ of $\mathbb{R}^M$ can be covered by 2^{2^k} translates of the ball $K_1(B_1 + 2^{-k/2}B_2)$ where $B_1 = B_1^M$ and $B_2 = B_2^M$ are respectively the ℓ_1 and ℓ_2 unit balls of $\mathbb{R}^M$. Thus, we can find a partition of S in at most 2^{2^k} sets A such that, given A, there exists $s_A \in S$ with

$$\forall s \in A\,,\ (x_i(s) - x_i(s_A))_{i \leq M} \in K_1(B_1 + 2^{-k/2}B_2)\,.$$

For s in A, set then $x'_i(s) = x_i(s_A)$. The conclusion is then obvious.

As a consequence of this lemma, note that by the triangle inequality

$$\left(\sum_i \|x_i\|_{2,1}^2\right)^{1/2} \leq \left(\sum_i \|x'_i\|_{2,1}^2\right)^{1/2} + \left(\sum_i \|y_i\|_{2,1}^2\right)^{1/2} + \left(\sum_i \|z_i\|_{2,1}^2\right)^{1/2}. \tag{15.30}$$

Since $\|\sum_i |y_i|\|_\infty \leq K_1$, the fact that j is $(2,1)$-summing already indicates that

$$\left(\sum_i \|y_i\|_{2,1}^2\right)^{1/2} \leq K_1\,.$$

The main objective will be to establish the following fact.

Lemma 15.22. If Σ has N atoms, then

$$\left(\sum_i \|x_i\|_{2,1}^2\right)^{1/2} \leq K_2(\log(N+1))^{1/2} \left\|\sum_i x_i^2\right\|_\infty^{1/2}$$

for all finite sequences (x_i) of functions on (S,Σ,μ) where K_2 is numerical.

Once this has been established, it follows together with Lemma 15.21 and (15.30) that if $\mathbb{E}\|\sum_i \varepsilon_i x_i\|_\infty \leq 1$, for all k,

$$\left(\sum_i \|x_i\|_{2,1}^2\right)^{1/2} \leq \left(\sum_i \|x_i'\|_{2,1}^2\right)^{1/2} + K_1 + K_1 K_2 2^{-k/2}(\log(N+1))^{1/2} \tag{15.31}$$

where (x_i') are functions which are measurable with respect to a subalgebra Σ' of Σ with at most 2^{2^k} atoms and satisfy $\mathbb{E}\|\sum_i \varepsilon_i x_i'\|_\infty \leq 1$. If Σ has less than $2^{2^{k+1}}$ atoms, (15.31) reads as

$$\left(\sum_i \|x_i\|_{2,1}^2\right)^{1/2} \leq \left(\sum_i \|x_i'\|_{2,1}^2\right)^{1/2} + K_3$$

for some numerical K_3. Iterative use of this property shows that if (x_i) is a finite sequence in $L_\infty(S,\Sigma,\mu)$ where Σ has less than 2^{2^k} atoms with $\mathbb{E}\|\sum_i \varepsilon_i x_i\| \leq 1$, then, for some numerical constant K,

$$\left(\sum_i \|x_i\|_{2,1}^2\right)^{1/2} \leq K(k+1).$$

This shows indeed Theorem 15.20.

We are thus left with the proof of Lemma 15.22.

Proof of Lemma 15.22. It relies on two observations. First, if x is in $L_{2,1}(\mu)$, and each atom of μ has mass $\geq a > 0$, then

$$\|x\|_{2,1} \leq 2(1+\log(a^{-1/2}))^{1/2}\|x\|_2. \tag{15.32}$$

Indeed, assuming $\|x\|_2 = 1$, we have $\|x\|_\infty \leq a^{1/2}$. Hence,

$$\begin{aligned}\|x\|_{2,1} &\leq 1 + \int_1^{a^{-1/2}} (t\mu(|x|>t))^{1/2}\frac{dt}{\sqrt{t}} \\ &\leq 1 + \left(\int_0^\infty t\mu(|x|>t)dt\right)^{1/2}\left(\int_1^{a^{-1/2}}\frac{dt}{t}\right)^{1/2}\end{aligned}$$

and (15.32) follows.

The second observation is as follows. Let ν be the probability measure on (S,Σ) that assigns mass $1/N$ to each atom of Σ. Let $\theta = \frac{1}{2}(\mu+\nu)$. Since $\mu(|x|>t) \leq 2\theta(|x|>t)$, we have that $\|x\|_{L_{2,1}(\mu)} \leq \sqrt{2}\|x\|_{L_{2,1}(\theta)}$. θ assigns mass $\geq 1/2N$ to each atom of Σ. Thus, by (15.32),

$$\|x\|_{L_{2,1}(\theta)} \leq 2(1+\log(\sqrt{2N}))^{1/2}\|x\|_{L_2(\theta)} .$$

When $\sum_i x_i(s)^2 \leq 1$ for all s, $\sum_i \|x_i\|^2_{L_2(\theta)} \leq 1$ from which the conclusion then clearly follows. The lemma is established.

15.8. Miscellaneous problems

In this last section, we present various problems on (or related to) the topics developed in this book. Some of them have already been explained in their context so that we only briefly recall them here.

Problem 1. Does inequality (1.1) (Chapter 1) of A. Ehrhard [Eh1]

$$\Phi^{-1}(\gamma_N(\lambda A+(1-\lambda)B)) \geq \lambda\Phi^{-1}(\gamma_N(A))+(1-\lambda)\Phi^{-1}(\gamma_N(B))$$

hold for all *Borel* sets A,B of $\mathbb{R}^N$ (and not only convex ones)?

Problem 2. Consider a Gaussian measure μ on a separable Banach space of unit ball B_1. Consider the function on $\mathbb{R}_+$, $F(t)=\mu(tB_1)$. It follows from (1.2) that $\log F$ is concave; thus it has right and left derivatives at every point. It is shown in [By] that these derivatives are equal so that F is differentiable. (The proof of [Ta2] is erroneous). One may wonder how regular F is. Is it thus differentiable? Actually, in all the examples we know the function F has an analytic extension in a sector $|\arg z| < \theta$ (think for example of Wiener measure on $C[0,1]$). This is a fascinating fact, since a priori one would think that the regularity of F should be related to the speed at which $F(t)$ goes to zero when $t\to 0$.

Problem 3. Consider a locally convex Hausdorff topological vector space E and μ a Radon measure on E equipped with its Borel σ-algebra that is Gaussian in the sense that the law of every continuous linear functional is Gaussian under μ. It is known that there is a compact metrizable set K with $\mu(K)>0$. But does there exist a compact *convex* set K for which $\mu(K)>0$? An equivalent formulation is the following.

Consider the canonical Gaussian measure γ_N on $\mathbb{R}^N$ and a compact set $B \subset \mathbb{R}^N$ such that $\gamma_N(B) > 0$. Let E be the linear space generated by B, i.e. $E = \bigcup_n B_n$ where

$$B_n = \underbrace{[-n,n]B + [-n,n]B + \cdots + [-n,n]B}_{n \text{ times}}$$

(and $[-n,n]B = \{\lambda x\,;\ |\lambda| \le n\,,\ x \in B\}$, $D + D = \{x + y\,;\ x, y \in D\}$). Is it true that for some compact *convex* set A with $\gamma_N(A) > 0$, we have $A \subset E$? It is not difficult to see that this problem is equivalent to the following. Does there exist n with the following property: whenever $B \subset \mathbb{R}^N$ is such that $\gamma_N(B) \ge 1/2$, then B_n contains a convex set A for which $\gamma_N(A) \ge 1/2$. It does not seem to be known whether $n = 3$ works.

Problem 4. Following Corollary 8.8 (Chapter 8), are type 2 spaces the only spaces B in which the conditions $\mathbb{E}(\|X\|^2/LL\|X\|) < \infty$ and $\mathbb{E}f(X) = 0$, $\mathbb{E}f^2(X) < \infty$ for all f in B' are also sufficient for X to satisfy the bounded (say) LIL? It can be seen from Proposition 9.19 that a Banach space with this property is necessarily of type $2-\varepsilon$ for all $\varepsilon > 0$ (see [Pi3]).

Problem 5. Theorem 8.10 completely describes the cluster set $C(S_n/a_n)$ of the LIL sequence in Banach space. Using in particular this result, Theorem 8.11 provides a complete picture of the limits in the LIL when $S_n/a_n \to 0$ in probability. One might wonder what these limits become when (S_n/a_n) is only bounded in probability, that is when X only satisfies the bounded LIL. What is in particular $\Lambda(X) = \limsup_{n\to\infty} \|S_n\|/a_n$? Example 8.12 suggests that this investigation could be difficult.

Problem 6. Try to characterize Banach spaces in which every random variable satisfying the CLT also satisfies the LIL. We have seen in Chapter 10, after Theorem 10.12, that cotype 2 spaces have this property, while conversely if a Banach space B satisfies this implication, B is necessarily of cotype $2+\varepsilon$ for every $\varepsilon > 0$ [Pi3]. Are cotype 2 spaces the only ones?

Problem 7. Theorem 10.10 indicates that in a Banach space B satisfying the inequality $\mathrm{Ros}(p)$ for some $p > 2$, a random variable X satisfies the CLT if and only if it is pregaussian and $\lim_{t\to\infty} t^2\mathbb{P}\{\|X\| > t\} = 0$. Is it true that, if, in a Banach space B, these best possible necessary conditions for the CLT are also sufficient, then B is of $\mathrm{Ros}(p)$ for some $p > 2$? This could be in analogy with the law of large numbers and the type of a Banach space (Corollary 9.18).

Problem 8. More generally on Chapter 10 (and Chapter 14), try to understand when an infinite dimensional random variable satisfies the CLT. This is the course related to one of the main questions of

Probability in Banach spaces: how to achieve efficiently tightness of a sum of independent random variables in terms for example of the individual summands?

Problem 9. Try to characterize almost sure boundedness and continuity of Gaussian chaos processes of order $d \geq 2$. See Section 11.3 and Section 15.2 in this chapter for more details and some (very) partial results. As conjectured in Section 15.2, an analog of Sudakov's minoration would be a first step in this investigation. Recently, some positive results in this direction have been obtained by the second author [Ta16].

Problem 10. Recall the problem described in Section 15.6 of this chapter of the explicit construction of a majorizing measure on $\mathrm{Conv}T$ when there is one on T.

Problem 11. Almost sure boundedness and continuity of Gaussian processes are now understood via the tool of majorizing measures (Section 12.1). Try now to understand boundedness and continuity of p-stable processes when $1 \leq p < 2$. In particular, since the necessary majorizing measure conditions of Section 12.2 are no more sufficient when $p < 2$, what are the additional conditions to investigate? From the series representation of stable processes, this question is closely related to Problem 8. The paper [Ta13] describes some of the difficulties in such an investigation.

Problem 12. Is it possible to characterize boundedness (and continuity) of Rademacher processes as conjectured in Section 12.3?

Problem 13. Is there a minimal Banach algebra $\overline{B}$ with $A(G) \not\subseteq \overline{B} \not\subseteq C(G)$ on which all Lipschitzian functions of order 1 operate? What is the contribution to this question of the algebra $\widetilde{B}$ discussed at the end of Section 13.2. Concerning further this algebra $\widetilde{B}$, try to describe it from an Harmonic Analysis point of view as was done for $C_{\text{a.s.}}$ by G. Pisier [Pi6].

Problem 14. In the random Fourier series notations of Chapter 13, is it true that

$$\mathbb{E}\sup\left\|\sum_\gamma \xi_\gamma x_\gamma \gamma(t)\right\| \leq K\left(\mathbb{E}\left\|\sum_\gamma \xi_\gamma x_\gamma\right\| + \sup_{\|f\|\leq 1}\mathbb{E}\sup_{t\in V}\left|\sum_\gamma \langle f, x_\gamma\rangle \xi_\gamma \gamma(t)\right|\right)$$

for every (finite) sequence (x_γ) in a Banach space B when (ξ_γ) is either a Rademacher sequence (ε_γ) or a standard p-stable sequence (θ_γ), $1 < p < 2$ (and also $p = 1$ but for moments less than 1)? The constant K may depend on V as in the Gaussian case (Corollary 13.15).

Notes and references

Theorem 15.2 originates in the work of V. D. Milman on almost Euclidean subspace of a quotient (cf. [Mi2], [Mi-S], [Pi18]). V. D. Milman used indeed a (weaker) version of this result to establish the remarkable fact that if B is a Banach space of dimension N, there is a subspace of a quotient of B of dimension $[c(\varepsilon)N]$ which is $(1+\varepsilon)$-isomorphic to a Hilbert space. A. Pajor and N. Tomczak-Jaegermann [P-TJ] improved Milman's estimate and established Theorem 15.2 using the isoperimetric inequality on the sphere and Sudakov's minoration. The first proof presented here is the Gaussian version of their argument and is taken from [Pi18]. The second proof is due to Y. Gordon [Gor4] with quantative improvements kindly communicated to us by the author.

Proposition 15.5 was shown to us by G. Pisier. Section 15.3 presents a different proof of the sharp inequality that J. Bourgain [Bour] uses in his deep investigation on $\Lambda(p)$-sets.

Theorem 15.8 is taken from the work by J. Bourgain and L. Tzafriri [B-T1] (see also [B-T2] for more recent information). The simplification in the proof of Proposition 15.10 was noticed by J. Bourgain at the light of some arguments used in [Ta15].

Embedding subspaces of L_p into ℓ_p^N was considered in special cases in [F-L-M], [J-S1], [Pi12], [Sch3], [Sch4] (among others). In particular, a breakthrough was made in [Sch4] by G. Schechtman who used empirical distributions to obtain various early general results. Schechtman's method was refined and combined with deep facts from Banach space theory by J. Bourgain, J. Lindenstrauss and V. D. Milman [B-L-M]. In [Ta15], a simple random choice argument is introduced that simplifies the probabilistic part of the proofs of [B-L-M]. The crucial Lemma 15.15 is taken from the work by D. Lewis [Lew]. Theorem 15.12 was obtained in this way in [Ta15]. It is not known if the K-convexity constant $K(E)$ is necessary. The entropy computations of the proof of Theorem 15.13 are taken from [B-L-M].

The proof of the existence of a majorizing measure on ellipsoids was the first step of the second author on the way of his general solution of majorizing measures (Chapter 12). A refinement of this result (with a simplified proof) can be found in [Ta20], where it is shown to imply several very sharp discrepancy results.

The results of Section 15.7 are due to S. J. Montgomery-Smith [MS1] (that contains further developments). The proofs are somewhat simpler than those of [MS1] and taken from [MS-T].