On the boundedness of Bernoulli processes*

Witold Bednorz and Rafał Latała

Abstract

We present a positive solution to the so-called Bernoulli Conjecture concerning the characterization of sample boundedness of Bernoulli processes. We also discuss some applications and related open problems.

AMS 2010 Subject Classification. Primary: 60E15, 60G17; Secondary: 60G50, 60B12, 42A20
Key words and phrases: Bernoulli process, chaining methods, regularity of stochastic processes, random Fourier series, VC-classes.

1 Introduction and Notation

One of the fundamental issues of probability theory is the investigation of suprema of stochastic processes. Besides various practical motivations it is closely related to such important theoretical problems as boundedness and continuity of sample paths of stochastic processes, convergence of orthogonal series, random series and stochastic integrals, estimates of norms of random vectors and random matrices, limit theorems for random vectors and empirical processes, combinatorial matching theorems and many others.

In particular in many situations one needs to find lower and upper bounds for the quantity $\mathbb{E}\sup_{t\in T} X_t$, where $(X_t)_{t\in T}$ is a stochastic process. For a large class of processes (including Gaussian and Bernoulli processes) finiteness of this quantity is equivalent to the sample boundedness, i.e. to the condition $\mathbb{P}(\sup_{t\in T} X_t < \infty) = 1$. To avoid measurability problems one may either assume that T is countable or define $\mathbb{E}\sup_{t\in T} X_t := \sup_F \mathbb{E}\sup_{t\in F} X_t$, where the supremum is taken over all finite sets $F \subset T$. The modern approach to this problem is based on chaining techniques, already present in the work of A. N. Kolmogorov and successfully developed over the last 40 years (see the monographs [22] and [25]).

*Research supported by the NCN grant DEC-2012/05/B/ST1/00412

The most important case of centered Gaussian processes $(G_t)_{t\in T}$ is well understood. In this case the boundedness of the process is related to the geometry of the metric space (T,d), where $d(t,s) := (\mathbb{E}(G_t - G_s)^2)^{1/2}$. In the landmark paper [3], R. Dudley obtained an upper bound for $g(T) := \mathbb{E}\sup_{t\in T} G_t$ in terms of entropy numbers. Dudley's bound may be reversed for stationary processes [5], but not in general. In 1974 X. Fernique [5] showed that for any probability measure μ on the metric space (T,d),

$$g(T) \le L \sup_{t\in T} \int_0^\infty \log^{1/2}\Big(\frac{1}{\mu(B(t,x))}\Big)dx,$$

where L here and in the sequel denotes an universal constant and $B(t,x)$ is the ball in T centered at t with radius x. This can easily be shown to improve Dudley's estimate. In the seminal paper [14] M. Talagrand showed that Fernique's bound may be reversed, i.e. for any centered Gaussian process G_t there exists a probability measure μ (called a majorizing measure) on T such that

$$\sup_{t\in T} \int_0^\infty \log^{1/2}\Big(\frac{1}{\mu(B(t,x))}\Big)dx \le Lg(T).$$

In general finding a majorizing measure in a concrete situation is a highly nontrivial task. In [21] M. Talagrand proposed a more combinatorial approach to this problem and showed that constructing a majorizing measure is equivalent to finding a suitable sequence of *admissible* partitions of the set T. An increasing sequence $(\mathcal{A}_n)_{n\ge 0}$ of partitions of the set T is called admissible if $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \le N_n := 2^{2^n}$. The Fernique-Talagrand estimate may then be expressed as

$$\frac{1}{L}\gamma_2(T,d) \le g(T) \le L\gamma_2(T,d), \tag{1}$$

where

$$\gamma_2(T,d) := \inf \sup_{t\in T} \sum_{n=0}^{\infty} 2^{n/2}\Delta(A_n(t)),$$

and where the infimum runs over all admissible sequences of partitions. Here $A_n(t)$ is the unique set in $\mathcal{A}_n$ which contains t and $\Delta(A)$ denotes the diameter of the set A.

Any separable Gaussian process has a canonical Karhunen-Loève type representation $(\sum_{i=1}^\infty t_i g_i)_{t\in T}$, where $g_1, g_2, \ldots$ are i.i.d. standard normal Gaussian $\mathcal{N}(0,1)$ r.v's and T is a subset of ℓ^2. Another fundamental class of processes is obtained when in such a sum one replaces the Gaussian r.v's (g_i) by independent random signs. We detail this now.

Let I be a countable set and $(\varepsilon_i)_{i\in I}$ be a Bernoulli sequence i.e. a sequence of i.i.d. symmetric r.v's taking values ± 1. For $t \in \ell^2(I)$ the series $X_t := \sum_{i\in I} t_i\varepsilon_i$ converges a.s. and for $T \subset \ell^2(I)$ we may define a *Bernoulli process* $(X_t)_{t\in T}$ and try to estimate $b(T) := \mathbb{E}\sup_{t\in T} X_t$. There are two easy ways to bound $b(T)$. The first is a consequence of the uniform bound $|X_t| \le \|t\|_1 = \sum_{i\in I}|t_i|$, so that $b(T) \le \sup_{t\in T}\|t\|_1$. Another is based on the domination by the canonical Gaussian process $G_t := \sum_{i\in I} t_i g_i$. Indeed, assuming independence of (g_i) and (ε_i), Jensen's inequality implies

$$g(T) = \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i g_i = \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i\varepsilon_i |g_i| \ge \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i\varepsilon_i \mathbb{E}|g_i| = \sqrt{\frac{2}{\pi}}b(T). \tag{2}$$

Obviously also if $T \subset T_1 + T_2 = \{t^1 + t^2\colon\ t^l \in T_l\}$ then $b(T) \le b(T_1) + b(T_2)$, hence

$$\begin{aligned} b(T) &\le \inf\Big\{\sup_{t\in T_1}\|t\|_1 + \sqrt{\frac{\pi}{2}}g(T_2)\colon\ T \subset T_1 + T_2\Big\} \\ &\le \inf\Big\{\sup_{t\in T_1}\|t\|_1 + L\gamma_2(T_2)\colon\ T \subset T_1 + T_2\Big\}, \end{aligned}$$

where $\gamma_2(T) = \gamma_2(T, d_2)$ and d_2 is the ℓ^2-distance. It was open for about 25 years (under the name of Bernoulli conjecture) whether the above estimate may be reversed (see e.g. Problem 12 in [12] or Chapter 4 in [22]). Our main result, announced in [2], provides an affirmative answer.

Theorem 1.1. *For any set $T \subset \ell^2(I)$ with $b(T) < \infty$ we may find a decomposition $T \subset T_1 + T_2$ with $\sup_{t\in T_1}\sum_{i\in I}|t_i| \le Lb(T)$ and $g(T_2) \le Lb(T)$.*

Of course part of the difficulty is that the decomposition is neither unique nor canonical. Let us briefly describe some crucial ideas behind the proof, which uses a number of tools developed over the years by M. Talagrand. First of all we must review the proof of the lower bound of (1) in the modern approach, as in e.g. [22]. Every idea of this proof is used to its fullest in our approach.

As was nicely explained in [16] two fundamental facts behind this proof are Gaussian concentration and the Sudakov minoration principle. Gaussian concentration asserts that the fluctuations of the supremum of a Gaussian process are at worse like those of a single Gaussian r.v. with standard deviation about the diameter of the space (T, d) (irrelevant of the average value of this supremum). The Sudakov minoration says that the supremum of m Gaussian r.v's with distances at least a of each other is at least of the order $a\sqrt{\log m}$. These two principles can then be combined to obtain a "growth condition" as follows. If the space (T, d) contains m pieces H_l, which are at mutual distances at least a, and if each of these pieces is of diameter at most a small fraction of a, then the expected value of the supremum of the process over the whole index set T is larger by about $a\sqrt{\log m}$ than the minimum over l of the expected value of supremum of the process on the set H_l. This brings the idea to measure the "size" $F(A)$ of a subsets A of T by the expected value of the supremum of the process over A. One is then led to perform constructions in the abstract metric space (T, d) using only the value of the "functional" $F(A)$ over the subsets A of T. (The concept of functionals and related "growth conditions" was introduced and developed by M. Talagrand [20, 22] to simplify proofs and give a unified approach to various majorizing measure type results.) The basic ingredient to the proof is then a "decomposition lemma", which is a simple consequence of the growth condition through a "greedy" construction. Roughly speaking this decomposition lemma asserts that there exists a universal constant r with the property that any subset A of T can be partitioned into at most m pieces such that each piece either has the diameter at most $\Delta(A)/r$, or else it satisfies the condition that its every subset B of diameter at most $\Delta(A)/r^2$ satisfies $F(B) \leq F(A) - c\Delta(A)\sqrt{\log m}$ for some positive universal constant c. (The reader observes that the condition on B is *not* that its diameter is at most $\Delta(A)/r$ but the much more stringent requirement that its diameter is at most $\Delta(A)/r^2$. It is exactly this point which makes the proof delicate.) In words, every piece is either small, or it has the property that the value of the functional on its very small sub-pieces is quite smaller than on the whole of A. The admissible sequence of partitions we look for is then obtained by a recursive use of the decomposition lemma. Each set A belonging to $\mathcal{A}_n$ is partitioned in at most $N_n = 2^{2^n}$ sets to produce the partition $\mathcal{A}_{n+1}$. It is not obvious, but true, that the resulting sequence of partitions has the required properties. (Proving this is the tricky part of the whole proof.)

When working with Bernoulli processes (and many others) the situation is more complicated than in the Gaussian case and one needs to use a family of distances interpolating between the ℓ^2 and the ℓ^1 distances. Such distances were introduced by M. Talagrand in [17], [18], [19] and will be of constant use. An important concept in our proof is reducing the decomposition of the set T to constructing a suitable admissible sequence of partitions. Theorem 3.1 below is a refinement of previous results of M. Talagrand in the same direction, [18, 19, 22]. In some sense this type of result amounts to organize chaining in an efficient way. Indeed in [25] M. Talagrand used such a result to settle the long standing problem of convergence of random Fourier series in a very general case.

How, then, should one construct the required partitions?

M. Talagrand extended to Bernoulli processes both Gaussian concentration and the Sudakov minoration in [15] and [17] (see Theorems 2.5 and 2.7 below). The Sudakov minoration result provides a lower bound on the expected value of the supremum of variables X_{t_l} when the various points t_l are far from each other in the ℓ^2 sense, but it requires a control in the supremum norm of the elements t_l. (The overall idea is simply that by the central limit theorem a sum $\sum_i \varepsilon_i t_{l,i}$ looks more like a Gaussian r.v. if all the coefficients are small.) In order to apply this minoration to increasingly larger families, one needs to reduce the supremum norm. To do this M. Talagrand introduced in [17] the fundamental idea of "chopping maps". These replace the process of interest by a process where the control in the supremum norm is better, but which is related to the original process through an equally crucial comparison theorem (Theorem 2.2 below). This is essentially done by replacing each term $t_i\varepsilon_i$ by a sum $\sum_j \varphi_j(t_i)\varepsilon_{i,j}$ for new independent Bernoulli r.v's and certain functions φ_j, where we control uniformly $\sup|\varphi_j(t_i)|$, and where $|t_i| = \sum_j |\varphi_j(t_i)|$. In some sense in this procedure we "add more Bernoulli r.v's" to the process.

On the base of these tools M. Talagrand was able to prove in [19] a weaker form of Bernoulli conjecture with ℓ^p-diameter bound on the set T_1, $p > 1$ instead of ℓ^1-diameter. Although such a bound is not optimal, it was sufficient to obtain deep results about Rademacher cotype constants of operators on $C(K)$ spaces.

The main difficulty in using chopping maps optimally is that there are two ℓ^2-distances involved, the distance associated to the process before it is chopped, and the possibly much smaller distance associated to the process after it is chopped. This makes it very difficult not to loose information during the construction. For example, if we try to mimic the construction in the Gaussian case, and if at a given stage of the construction we have a set A with the property that on every subset of very small diameter the process is significantly smaller than on the whole of A, it is far from clear what this implies after applying a chopping map since sets of small diameter for the "smaller distance" need not be of small diameter for the larger distance. Maps other than chopping maps were used in [10], where the Bernoulli conjecture was verified for a very special class of subsets of ℓ^2. Proposition 2.10 below is a modification of the key new fact proved in that paper. It is the cornerstone of our paper. While Talagrand's chopping maps amount somehow to introduce new Bernoulli r.v's, a major new ingredient is that we find convenient at times to *remove* some of these variables (which can only decrease the size of the process). In the situation of Proposition 2.10 we consider a subset J of I and the process $X'_t = \sum_{i\in J} t_i\varepsilon_i$; that is, we remove the Bernoulli r.v's which are not indexed by J. We then have two ℓ^2-distances on the index set: a small one $\sqrt{\sum_{i\in J}(t_i - s_i)^2}$ and a large one $\sqrt{\sum_{i\in I}(t_i - s_i)^2}$. Roughly speaking the content of Proposition 2.10 is that if the index set has a small diameter with respect to the *smaller* distance we may decompose it into not too many sets which either have a small diameter with respect to the *larger* original distance or else have the property that the size of the process over the whole piece has decreased significantly when one drops the Bernoulli r.v's which are not indexed by J. The quantitative version of the result involves of course the ubiquitous term $\sqrt{\log m}$, where m is the number of pieces permitted.

Even after this principle has been clarified, it is still a very non-trivial technical problem to define an appropriate family of "functionals" to measure the "size" of the pieces of our partition. These functionals at time "add" new Bernoulli r.v's and at time "remove" some. Of course the difficulty is to find an exact balance between these two operations to ensure that no essential information is lost. Our functionals, defined in Section 4, depend on four parameters J, u, k, j. The parameter $j \in \mathbb{Z}$ indicates "how much chopping we have performed". The other three parameters keep track of which Bernoulli r.v's we still use in the functional. A new feature of this construction is that our functionals depend not only on which stage of the construction we are at, but also on which piece we are trying to partition. At each step we use a "decomposition lemma", which we give in Corollary 5.3, somewhat similar in spirit to that of the Gaussian case. Another new feature is that this lemma is not obtained only through a growth condition. To prove it we also apply in an essential way Proposition 2.10 mentioned above. In contrast with the Gaussian case, the decomposition lemma now produces three distinct types of pieces. Two of the types of pieces behave as in the Gaussian case. The new type of piece has the property that its size (as measured by the proper functional) has decreased compared to the set we partitioned after ignoring a suitable subset of the Bernoulli r.v's.

Our proof also uses in an essential way the technique of "counters" introduced by M. Talagrand to keep suitably track of the "past" of the construction, c.f. [22, Chapter 5].

Theorem 1.1 yields another striking characterization of boundedness for Bernoulli processes. For a random variable X and $p > 0$ we set $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$.

Corollary 1.2. *Suppose that $(X_t)_{t\in T}$ is a Bernoulli process with $b(T) < \infty$. Then there exist $t^1, t^2, \ldots \in \ell^2$ such that $T - T \subset \overline{\mathrm{conv}}\{t^n\colon\ n \geq 1\}$ and $\|X_{t^n}\|_{\log(n+2)} \leq Lb(T)$ for all $n \geq 1$.*

The converse statement easily follows from the union bound and Chebyshev's inequality. Indeed, suppose that $T - T \subset \overline{\mathrm{conv}}\{t^n\colon\ n \geq 1\}$ and $\|X_{t^n}\|_{\log(n+2)} \leq M$. Then for $u \geq 1$,

$$\mathbb{P}\Big(\sup_{s\in T-T} X_s \geq uM\Big) \leq \mathbb{P}\Big(\sup_{n\geq 1} X_{t^n} \geq uM\Big) \leq \sum_{n\geq 1} \mathbb{P}(X_{t^n} \geq u\|X_{t^n}\|_{\log(n+2)}) \leq \sum_{n\geq 1} u^{-\log(n+2)}$$

and integration by parts easily yields $\mathbb{E}\sup_{s\in T-T} X_t \leq LM$. Moreover for any $t_0 \in T$,

$$b(T) = \mathbb{E}\sup_{t\in T}(X_t - X_{t_0}) = \mathbb{E}\sup_{t\in T}(X_{t-t_0}) \leq \mathbb{E}\sup_{s\in T-T} X_s \leq LM.$$

One of the motivations to state the Bernoulli Conjecture was a question of X. Fernique about vector-valued random Fourier series (which we solve in Theorem 8.1 below). Another interesting application of Theorem 1.1 is a Levy-Ottaviani type maximal inequality for VC-classes (Theorem 8.2).

To put Theorem 1.1 in a proper perspective, we will briefly explain that is it just the first step towards a much more ambitious program outlined in Talagrand's book [25]. One way to describe (1) in words is that "chaining explains the size of Gaussian processes". The best chaining bound one can obtain for the supremum of a Gaussian process is of the correct order. Now, the bound $\sum_i t_i\varepsilon_i \le \sum_i |t_i|$ on a Bernoulli process is of a different nature, in the sense that it makes no use of cancellation between the various terms. In some sense, Theorem 1.1 can be reformulated as "chaining explains the part of boundedness which is due to cancellation". That is, chaining explains the boundedness of the part T_2 of the process, while the boundedness of the T_1 part owes nothing to cancellation. It is argued in [25] that the phenomenon that "chaining explains the part of boundedness due to cancellation" could be true in many more situations (empirical processes, infinitely divisible processes). Here we just briefly discuss the case of empirical processes.

Let $(X_i)_{i\le N}$ be i.i.d. r.v's with values in a measurable space $(S,\mathcal{S})$ and $\mathcal{F}$ be a class of measurable functions on S. It is a fundamental problem, strictly related to the investigation of uniform laws of large numbers, uniform central limit theorems and various applications in asymptotic statistics c.f. [4, 26], to relate the quantity

$$\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i\le N}(f(X_i)-\mathbb{E}f(X_i)) \tag{3}$$

with the geometry of the class $\mathcal{F}$. A first situation is when one already controls

$$\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i\le N}|f(X_i)|,$$

a situation where there is no cancellation. A second situation is when one can bound the quantity (3) using chaining. Since then one has to use Bernstein's inequality (36), this requires not only a control of the size of $\mathcal{F}$ with respect to the ℓ^2 norm, but also with respect to the ℓ^∞ norm. M. Talagrand then conjectures that the general situation is a mixture of these two cases. The precise technical statement is given in Conjecture 9.2 below.

A discretized version of this problem concerning the"selector processes" based on the i.i.d. sequence $(\delta_i)_{i\in I}$ will also be discussed in Section 9.

In a somewhat different direction, we would like to mention a very beautiful generalization of the Bernoulli Conjecture formulated by S. Kwapień (private communication).

Problem 1.3. *Let $(F,\|\ \|)$ be a normed space and (u_i) be a sequence of vectors in F such that the series $\sum_{i\ge1}u_i\varepsilon_i$ converges a.s. Does there exist* *a universal constant* L *and a decomposition* $u_i = v_i + w_i$ *such that*

$$\sup_{\eta_i=\pm 1}\Big\|\sum_{i\geq 1} v_i\eta_i\Big\| \leq L\mathbb{E}\Big\|\sum_{i\geq 1} u_i\varepsilon_i\Big\| \quad \textit{and} \quad \mathbb{E}\Big\|\sum_{i\geq 1} w_i g_i\Big\| \leq L\mathbb{E}\Big\|\sum_{i\geq 1} u_i\varepsilon_i\Big\|?$$

Theorem 1.1 shows that the answer is positive for $F = \ell^\infty$, in general however we may only assume that F is a subspace of ℓ^∞. The difficulty here is that our proof gives very little additional information about the decomposition given by Theorem 1.1, in particular there is no reason for sets T_1 and T_2 to be contained in the linear space spanned by the index set T.

The paper is organized as follows. In Section 2 we gather general results about Bernoulli processes. The main new ingredient there is Proposition 2.10. In Section 3 we show how to reduce finding a required decomposition of the index set to constructing a suitable admissible sequence of partitions. In Section 4 on the base of chopping maps we define functionals and in Section 5 we show that they satisfy a Talagrand-type decomposition condition stated in Corollary 5.3. In Section 6 we inductively construct a required admissible sequence of partitions and conclude proofs of the main results stated above in Section 7. In Section 8 we present two applications of our main result and in the last Section 9 we discuss in more details the situation of "selector processes".

Acknowledgments. We would like to thank professors Stanisław Kwapień and Michel Talagrand for constant encouragement to work on the problem. Upon seeing our original proof, Michel Talagrand was able to simplify a number of technical details, and we are grateful to him for allowing us to freely use some of his arguments as well as for multiple comments that improved the presentation of the paper.

Notation. By $(\varepsilon_i)_i$, $(\varepsilon_{i,j})_{i,j}$ and $(\varepsilon_{i,j,k})_{i,j,k}$ we denote independent Bernoulli sequences. We use letter L to denote positive universal constants that may change from line to line, and L_i for positive universal constants that are the same at each occurrence.

By $\Delta_{\ell^2(I)}(T)$ (or $\Delta_2(T)$ if the set I is clear from the context) we denote the diameter with respect to the ℓ^2-distance of the set $T \subset \ell^2(I)$.

2 Estimates for Bernoulli processes

In the first part of this section we gather several well known estimates for suprema of Bernoulli processes and discuss some of their consequences that play a crucial role in the proof of the main result.

We start with the following simple bound on the diameter of the index set.

Lemma 2.1. *For any $T \subset \ell^2(I)$ we have $\Delta_2(T) \leq 4b(T)$.*

Proof. Let $X_t := \sum_i t_i\varepsilon_i$ for $t \in T$. Take any $t, s \in T$, then

$$b(T) \geq \mathbb{E}\max\{X_t, X_s\} = \mathbb{E}\max\{X_t - X_s, 0\} = \frac{1}{2}\mathbb{E}|X_t - X_s| \geq \frac{1}{4}\|t-s\|_2.$$

□

Obviously by Jensen's inequality we have

$$\mathbb{E}\sup_{t\in T}\sum_{i\in J} t_i\varepsilon_i \leq \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i\varepsilon_i \quad \text{for } J \subset I. \tag{4}$$

Much deeper is the following Talagrand's comparison theorem for Bernoulli processes (cf. Theorem 2.1 in [17] or the proof of Theorem 4.12 in [12]).

Theorem 2.2. *Suppose that $\varphi_i \colon \mathbb{R} \to \mathbb{R}$, $i \in I$ are contractions (i.e. $|\varphi_i(x) - \varphi_i(y)| \leq |x - y|$) and $\varphi_i(0) = 0$ for all $i \in I$. Then for any $T \subset \ell^2(I)$,*

$$\mathbb{E}\sup_{t\in T}\sum_{i\in I} \varphi_i(t_i)\varepsilon_i \leq \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i\varepsilon_i.$$

Remark 2.3. Since

$$\mathbb{E}\sup_{t\in T}\sum_{i\in I} \varphi_i(t_i)\varepsilon_i = \mathbb{E}\sup_{t\in T}\sum_{i\in I} (\varphi_i(t_i) - \varphi_i(0))\varepsilon_i$$

we may replace the assumption that $\varphi_i(0) = 0$ with $(\varphi_i(0)) \in \ell^2(I)$ (which for contractions is equivalent to $(\varphi_i(t_i)) \in \ell^2(I)$ for some/all $t \in \ell^2(I)$).

A typical application of Theorem 2.2 is the following.

Corollary 2.4. *Suppose that $(f_{i,j})$ and (g_i) are functions on $\mathbb{R}$ such that for all $i \in I$, $x, y \in \mathbb{R}$,*

$$\sum_{j\in J} |f_{i,j}(x) - f_{i,j}(y)| \leq |g_i(x) - g_i(y)|.$$

Let T be a set such that $(g_i(t_i)) \in \ell^2(I)$ and $(f_{i,j}(t_i)) \in \ell^2(I \times J)$ for all $t \in T$. Then

$$\mathbb{E}\sup_{t\in T}\sum_{i\in I, j\in J} f_{i,j}(t_i)\varepsilon_{i,j} \leq \mathbb{E}\sup_{t\in T}\sum_{i\in I} g_i(t_i)\varepsilon_i.$$

Proof. Without loss of generality we may assume that the sequences $(\varepsilon_{i,j})$ and (ε_i) are independent. It is enough to observe that

$$\mathbb{E}\sup_{t\in T}\sum_{i\in I, j\in J} f_{i,j}(t_i)\varepsilon_{i,j} = \mathbb{E}\sup_{t\in T}\sum_{i\in I}\Big(\sum_{j\in J} f_{i,j}(t_i)\varepsilon_{i,j}\Big)\varepsilon_i$$

and that for any values of $\varepsilon_{i,j} \in \{\pm 1\}$ and $x, y \in \mathbb{R}$,

$$\Big| \sum_{j \in J} f_{i,j}(x)\varepsilon_{i,j} - \sum_{j \in J} f_{i,j}(y)\varepsilon_{i,j} \Big| \leq |g_i(x) - g_i(y)|.$$

The assertion follows by applying conditionally Theorem 2.2. □

Next we state the concentration property of Bernoulli processes (cf. [15] or [11, Corollary 4.10]).

Theorem 2.5. *Let $(a_t)_{t \in T}$ be a sequence of real numbers indexed by a set $T \subset \ell^2(I)$ and $S := \sup_{t \in T}(a_t + \sum_{i \in I} t_i \varepsilon_i)$ be such that $|S| < \infty$ a.s. Then*

$$\mathbb{P}(|S - \mathrm{Med}(S)| \geq u) \leq 4 \exp\Big(-\frac{u^2}{16\sigma^2} \Big) \quad \text{for } u > 0,$$

where $\sigma := \sup_{t \in T} \|t\|_2$. In particular $\mathbb{E}|S| < \infty$, $|\mathbb{E}S - \mathrm{Med}(S)| \leq L\sigma$ and

$$\mathbb{P}(|S - \mathbb{E}(S)| \geq u) \leq 2 \exp\Big(-\frac{u^2}{L_1 \sigma^2} \Big) \quad \text{for } u > 0.$$

Theorem 2.5 easily implies the following fact [10, Corollary 1].

Proposition 2.6. *Let $(Y_t^k)_{t \in T}$, $1 \leq k \leq m$ be i.i.d. Bernoulli processes and $\sigma := \sup_{t \in T} \|Y_t^1\|_2$. Then for any process $(Z_t)_{t \in T}$ independent of $(Y_t^k \colon t \in T, k \leq m)$ we have*

$$\mathbb{E} \max_{1 \leq k \leq m} \sup_{t \in T} (Z_t + Y_t^k) \leq \mathbb{E} \sup_{t \in T} (Z_t + Y_t^1) + L_2 \sigma \sqrt{\log m}.$$

Another important property of Bernoulli processes is a Sudakov-type minoration formulated and proved by M. Talagrand (cf. [17] or [22, Theorem 4.2.4]).

Theorem 2.7. *Suppose that vectors $t_1, \ldots, t_m \in \ell^2(I)$ and numbers $a, b > 0$ satisfy*

$$\forall_{l \neq l'} \ \|t_l - t_{l'}\|_2 \geq a \quad \text{and} \quad \forall_l \ \|t_l\|_\infty \leq b. \tag{5}$$

Then

$$\mathbb{E} \sup_{l \leq m} \sum_{i \in I} t_{l,i} \varepsilon_i \geq \frac{1}{L_3} \min\Big\{ a\sqrt{\log m}, \frac{a^2}{b} \Big\}.$$

The next proposition, also due to M. Talagrand, combines concentration and minoration properties for Bernoulli processes [22, Proposition 4.2.2]. It exactly parallels the Gaussian case.

Proposition 2.8. *Consider vectors $t_1, \ldots, t_m \in \ell^2(I)$ and numbers $a, b > 0$ such that (5) holds. Then for any $\sigma > 0$ and any sets $H_l \subset B_{\ell^2(I)}(t_l, \sigma)$,*

$$b\Big(\bigcup_{l \leq m} H_l \Big) \geq \frac{1}{L_4} \min\Big\{ a\sqrt{\log m}, \frac{a^2}{b} \Big\} - L_5 \sigma \sqrt{\log m} + \min_{l \leq m} b(H_l).$$

Proposition 2.8 together with a simple greedy algorithm yields the following decomposition result for Bernoulli processes. This again parallels the Gaussian case.

Corollary 2.9. *Suppose that* $\|t\|_\infty \leq b$ *for all* $t \in T$ *and* $b\sqrt{\log m} \leq \sigma$. *Then there exists sets* $C_1, \ldots, C_{m-1} \subset T$ *such that* $\Delta_{\ell^2(I)}(C_i) \leq L_6\sigma$ *and for each nonempty set* $D \subset T \setminus \bigcup_{k \leq m-1} C_k$ *with* $\Delta_{\ell^2(I)}(D) \leq \sigma$,

$$b(D) \leq b(T) - \sigma\sqrt{\log m}.$$

Proof. Let $L_6 = \max\{2, 2L_4(L_5+2)\}$ and $a = \frac{1}{2}L_6\sigma$. Then

$$\min\left\{a\sqrt{\log m}, \frac{a^2}{b}\right\} = a\sqrt{\log m} \geq L_4(L_5+2)\sigma\sqrt{\log m}.$$

If $T \subset \bigcup_{i \leq m-1} B(t_i, a)$ for some $t_1, \ldots, t_{m-1} \in T$ there is nothing to prove, otherwise we choose inductively vectors $t_1, t_2, \ldots, t_{m-1}$. To this end we set $T_1 := T$ and $T_k := T \setminus \bigcup_{l<k} B(t_l, a)$ for $k > 1$ and choose $t_k \in T_k$ in such a way that

$$b(T_k \cap B(t_k, \sigma)) \geq \sup_{t \in T_k} b(T_k \cap B(t, \sigma)) - \sigma\sqrt{\log m}.$$

Let $C_k := T \cap B(t_k, a)$ for $k \leq m-1$. Then obviously $\Delta_{\ell^2(I)}(C_k) \leq L_6\sigma$. Take any $D \subset T_m = T \setminus \bigcup_{k<m} C_k$ with $\Delta_{\ell^2(I)}(D) \leq \sigma$ and choose any $t_m \in D$ so that $D \subset B(t_m, \sigma) \cap T_m$. By construction the condition (5) holds. Let $H_l := B(t_l, \sigma) \cap T_l$, for $l < m$ and $H_m := D$. Then by the choice of t_l it follows that

$$\min_{1 \leq l \leq m} b(H_l) \geq b(D) - \sigma\sqrt{\log m}.$$

So by Proposition 2.8

$$\begin{aligned} b(T) \geq b\Big(\bigcup_{l \leq m} H_l\Big) &\geq \frac{1}{L_4}\min\left\{a\sqrt{\log m}, \frac{a^2}{b}\right\} + b(D) - (L_5+1)\sigma\sqrt{\log m} \\ &\geq b(D) + \sigma\sqrt{\log m}. \end{aligned}$$

□

The last result of this section is a modification of Proposition 1 from [10] and contains the crucial idea of "removing" some of Bernoulli r.v's. Before we state it let us introduce a bit of notation. For $\emptyset \neq J \subset I$, $t \in \ell^2(I)$, $T \subset \ell^2(I)$ we define $t_J := (t_i)_{i \in J} \in \ell^2(J)$,

$$b_J(T) := \mathbb{E}\sup_{t \in T}\sum_{i \in J}\varepsilon_i t_i,$$

$$d_J(t,s) := \|t_J - s_J\|_2, \quad t, s \in \ell^2(I)$$

and

$$B_J(t,a) := \{s \in \ell^2(I) \colon d_J(s,t) \leq a\}, \ a \geq 0.$$

Proposition 2.10. *Consider a positive integer m, numbers $b, c, \sigma > 0$ and $\lambda \geq 1$ that satisfy $b\sqrt{\log m} \leq \lambda\sigma$ and $T \subset \ell^2(I)$ such that*

$$\forall_{t,s\in T}\ d_J(t,s) \leq c, \quad \|t-s\|_\infty \leq b. \tag{6}$$

Then there exist $t_1, \ldots, t_m \in T$ such that either $T \subset \bigcup_{l\leq m} B_I(t_l, \sigma)$ or

$$b_J\Big(T \setminus \bigcup_{l\leq m} B_I(t_l,\sigma)\Big) \leq b_I(T) - \Big(\frac{1}{4\lambda L_3}\sigma - L_7 c\Big)\sqrt{\log m}. \tag{7}$$

Proposition 2.10 and Corollary 2.9 present two ways to decompose the index set of a Bernoulli process. Combination of both statements will yield the main decomposition result, Corollary 5.3. Observe that we use in Proposition 2.10 two distances d_J and d_I. What is fundamental here is that we assume that the diameter of the set T is small only with respect to the smaller distance d_J and we show that it may be covered by a certain number of balls with respect to the larger distance d_I and a remaining set with a small value of b_J. The proof is based on concentration and minorization properties of Bernoulli processes, but they are combined in a different way than in Proposition 2.8.

Proof. If $T \subset \bigcup_{l\leq m} B_I(t_l,\sigma)$ for some $t_1, \ldots, t_m \in T$ or $m = 1$ there is nothing to prove, so we will assume that this is not the case. We may also choose the universal constant L_7 in such a way that $L_3 L_7 \geq 1$, so it is enough to consider the case $\sigma \geq 2c$ (since otherwise $\frac{1}{4\lambda L_3}\sigma - L_7 c < 0$).

Since $b_J(T) = b_J(T - t)$ for any $t \in \ell^2(I)$, we may and will assume that $0 \in T$, so that

$$\|t_J\|_2 \leq c, \quad \|t\|_\infty \leq b \leq \frac{\lambda\sigma}{\sqrt{\log m}} \qquad \text{for } t \in T.$$

We need to show that

$$\alpha < b_I(T) - \Big(\frac{1}{4\lambda L_3}\sigma - L_7 c\Big)\sqrt{\log m}, \tag{8}$$

where

$$\alpha := \inf_{t_1,\ldots,t_m\in T} b_J\Big(T \setminus \bigcup_{l\leq m} B_I(t_l,\sigma)\Big).$$

Let $\varepsilon_i^{(k)}$, $i \in J$, $k = 1, \ldots, m$ be independent Bernoulli r.v's, independent of $(\varepsilon_i)_{i\in I}$. Let

$$Y_t^{(k)} := \sum_{i\in J} t_i \varepsilon_i^{(k)}, \quad Z_t := \sum_{i\in I\setminus J} t_i \varepsilon_i.$$

Then for any k,

$$b(T) = \mathbb{E}\sup_{t\in T}(Z_t + Y_t^{(k)}),$$

and therefore Proposition 2.6 yields

$$\mathbb{E}\max_{1\le k\le m}\sup_{t\in T}(Z_t+Y_t^{(k)})\le b(T)+L_2c\sqrt{\log m}. \tag{9}$$

We set $T_1=T$ and define a random point $t_1\in T_1$ that depends only on $(\varepsilon_i^{(1)})_{i\in J}$ such that

$$Y_{t_1}^{(1)}>\sup_{t\in T_1}Y_t^{(1)}-c\sqrt{\log m}.$$

We continue this construction and inductively define random points $t_k\in T$, $k\le m$ that depend only on $(\varepsilon_i^{(l)})_{l\le k,i\in J}$. If $t_1,\ldots,t_{k-1}$ are already defined we set

$$T_k:=T\setminus\bigcup_{l\le k-1}B_I(t_l,\sigma)$$

and we choose a random point $t_k\in T_k$ such that

$$Y_{t_k}^{(k)}>\sup_{t\in T_k}Y_t^{(k)}-c\sqrt{\log m}.$$

The process $(Y_t^{(k)})$ is independent of the set T_k and for $k\le m$,

$$Y_{t_k}^{(k)}+c\sqrt{\log m}>\sup_{t\in T_k}Y_t^{(k)}\quad\text{and}\quad\mathbb{E}\sup_{t\in T_k}Y_t^{(k)}\ge\alpha.$$

We have

$$\begin{aligned}\mathbb{E}\max_{1\le k\le m}\sup_{t\in T}(Z_t+Y_t^{(k)})&\ge\mathbb{E}\Big(\max_{1\le k\le m}Z_{t_k}+\min_{1\le k\le m}Y_{t_k}^{(k)}\Big)\\&\ge\mathbb{E}\max_{1\le k\le m}Z_{t_k}+\alpha-c\sqrt{\log m}+\mathbb{E}\min_{1\le k\le m}\Big(\sup_{t\in T_k}Y_t^{(k)}-\alpha\Big)\\&\ge\mathbb{E}\max_{1\le k\le m}Z_{t_k}+\alpha-c\sqrt{\log m}+\mathbb{E}\min_{1\le k\le m}\Big(\sup_{t\in T_k}Y_t^{(k)}-\mathbb{E}\sup_{t\in T_k}Y_t^{(k)}\Big).\end{aligned}\tag{10}$$

Observe that for $1\le l<k\le m$,

$$d_{I\setminus J}(t_k,t_l)\ge d_I(t_k,t_l)-d_J(t_k,t_l)\ge\sigma-c\ge\frac{1}{2}\sigma,$$

and hence Theorem 2.7 with $a=\sigma/2$ (and using independence of Z_t and of the random points (t_k)) implies

$$\mathbb{E}\max_{1\le k\le m}Z_{t_k}\ge\frac{1}{4\lambda L_3}\sigma\sqrt{\log m}. \tag{11}$$

Since $(Y_t^{(k)})$ is independent of the set T_k, Theorem 2.5 gives that for $u>0$,

$$\mathbb{P}\Big(\sup_{t\in T_k}Y_t^{(k)}-\mathbb{E}\sup_{t\in T_k}Y_t^{(k)}\le-u\Big)\le2\exp\Big(-\frac{u^2}{L_1c^2}\Big),$$

so that

$$\mathbb{P}\Big(\min_{k\le m}\Big(\sup_{t\in T_k} Y_t^{(k)} - \mathbb{E}\sup_{t\in T_k} Y_t^{(k)}\Big)\le -u\Big)\le \min\Big\{1, 2m\exp\Big(-\frac{u^2}{L_1c^2}\Big)\Big\},$$

and integration by parts yields

$$\mathbb{E}\min_{k\le m}\Big(\sup_{t\in T_k} Y_t^{(k)} - \mathbb{E}\sup_{t\in T_k} Y_t^{(k)}\Big)\ge -Lc\sqrt{\log m}. \tag{12}$$

Estimates (9)-(12) imply (8) and complete the proof. □

3 Partitions

One of the main difficulties of the proof of Theorem 1.1 is that there is no canonical way to decompose the index set of Bernoulli processes. M. Talagrand connected finding this decomposition with constructing a suitable sequence of partitions (cf. Theorem 2.6.3 in [22]). Theorem 3.1 and its proof are based on Talagrand' ideas. The main new ingredient here is the introduction of the sets $I_n(A)$ – they will enable us to "remove" some of Bernoulli r.v's from the process during the inductive partition construction and efficiently use Proposition 2.10.

We recall that an increasing sequence $(\mathcal{A}_n)_{n\ge 0}$ of partitions of T is called *admissible* if $\mathcal{A}_0=\{T\}$ and $|\mathcal{A}_n|\le N_n:=2^{2^n}$. For such partitions and $t\in T$ by $A_n(t)$ we denote the unique set in $\mathcal{A}_n$ which contains t. To each set $A\in\mathcal{A}_n$ we will associate a point $\pi_n(A)$ and an integer $j_n(A)$. To simplify the notation we set $j_n(t):=j_n(A_n(t))$ and $\pi_n(t):=\pi_n(A_n(t))$. The main new feature in the next theorem is the introduction of the sets $I_n(A)$.

Theorem 3.1. *Suppose that $M>0$, $r\ge 2$, $(\mathcal{A}_n)_{n\ge 0}$ is an admissible sequence of partitions of $T\subset \ell^2(I)$, and for each $A\in\mathcal{A}_n$ there exists an integer $j_n(A)$ and a point $\pi_n(A)\in T$ satisfying the following assumptions:*
i) $\|t-s\|_2\le\sqrt{M}r^{-j_0(T)}$ for $t,s\in T$,
ii) if $n\ge 1$, $\mathcal{A}_n\ni A\subset A'\in\mathcal{A}_{n-1}$ then either
a) $j_n(A)=j_{n-1}(A')$ and $\pi_n(A)=\pi_{n-1}(A')$
or
b) $j_n(A)>j_{n-1}(A')$, $\pi_n(A)\in A'$ and

$$\sum_{i\in I_n(A)}\min\{(t_i-\pi_n(A)_i)^2, r^{-2j_n(A)}\}\le M2^nr^{-2j_n(A)} \text{ for all } t\in A,$$

where for any $t\in A$,

$$I_n(A)=I_n(t):=\big\{i\in I\colon\ |\pi_{k+1}(t)_i-\pi_k(t)_i|\le r^{-j_k(t)} \text{ for } 0\le k\le n-1\big\}.$$

Then there exist sets T_1,T_2 such that $T\subset T_1+T_2$ and

$$\sup_{t^1\in T_1}\|t^1\|_1\le LM\sup_{t\in T}\sum_{n=0}^{\infty}2^nr^{-j_n(t)} \quad \text{and} \quad \gamma_2(T_2)\le L\sqrt{M}\sup_{t\in T}\sum_{n=0}^{\infty}2^nr^{-j_n(t)}. \tag{13}$$

Remark. Note that if $t, s \in A \in \mathcal{A}_n$ then for $0 \le k \le n$, $A_k(t) = A_k(s)$ and as a consequence $j_k(t) = j_k(s)$, $\pi_k(t) = \pi_k(s)$ and $I_n(t) = I_n(s)$. Therefore the definition of $I_n(A)$ does not depend on the choice of $t \in A$.

Proof. Obviously we may assume that $\sup_{t\in T} \sum_{n\ge 0} 2^n r^{-j_n(t)} < \infty$, which in particular implies that $\lim_{n\to\infty} j_n(t) = \infty$. Define

$$m(t,i) := \inf\big\{n \ge 0\colon\ |\pi_{n+1}(t)_i - \pi_n(t)_i| > r^{-j_n(t)}\big\}, \quad t \in T, i \in I,$$

so that $I_n(t) = \{i\colon\ m(t,i) \ge n\}$ for $n \ge 0$.

Observe that

$$|\pi_{n+1}(t)_i - \pi_n(t)_i| \le r^{-j_n(t)} I_{\{j_{n+1}(t) > j_n(t)\}} \quad \text{for } 0 \le n < m(t,i). \tag{14}$$

Since $j_n(t)$ is nondecreasing sequence of integers, for i such that $m(t,i) = \infty$ the limit $\pi_\infty(t)_i := \lim_{n\to\infty} \pi_n(t)_i$ exists. Therefore we may define $\pi(t)$ by the formula

$$\pi(t)_i := \pi_{m(t,i)}(t)_i, \quad t \in T, i \in I.$$

We set

$$T_1 := \{t - \pi(t)\colon\ t \in T\} \quad \text{and} \quad T_2 := \{\pi(t)\colon\ t \in T\},$$

so that obviously $T \subset T_1 + T_2$.

To estimate $\|t - \pi(t)\|_1$ we define

$$\tau(t,i) := \inf\big\{n \ge 0\colon\ |\pi_n(t)_i - t_i| > \frac{1}{2} r^{-j_n(t)}\big\}, \quad t \in T, i \in I$$

and

$$J_n(t) := \{i \in I\colon\ \tau(t,i) = n\}.$$

Observe that $\tau(t,i) \le m(t,i) + 1$ and if $\tau(t,i) = \infty$ then $\pi(t)_i = \pi_\infty(t)_i = t_i$. Therefore we have

$$\|t - \pi(t)\|_1 = \sum_{n=0}^{\infty} \sum_{i\in J_n(t)} |t_i - \pi_{m(t,i)}(t)_i|.$$

From (14) we get

$$|\pi_0(t)_i - \pi_{m(t,i)}(t)_i| \le \sum_{n=0}^{m(t,i)-1} |\pi_{n+1}(t)_i - \pi_n(t)_i| \le \sum_{j=j_0(t)}^{\infty} r^{-j} \le 2r^{-j_0(t)},$$

and moreover for $i \in J_0(t)$, it holds that $|t_i - \pi_0(t)_i| \ge \frac{1}{2} r^{-j_0(t)}$. Thus

$$\sum_{i\in J_0(t)} |t_i - \pi_{m(t,i)}(t)_i| \le 5 \sum_{i\in J_0(t)} |t_i - \pi_0(t)_i| \le 10 r^{j_0(t)} \sum_{i\in I} |t_i - \pi_0(t)_i|^2$$
$$\le 10 M r^{-j_0(t)},$$

where the last estimate follows by the assumption i).

If $i \in J_n(t)$, $n \geq 1$ then $m(t,i) \geq n-1$ and

$$
\begin{aligned}
|t_i - \pi_{m(t,i)}(t)_i| &\leq |t_i - \pi_{n-1}(t)_i| + \sum_{k=n-1}^{m(t,i)-1} |\pi_{k+1}(t)_i - \pi_k(t)_i| \\
&\leq \frac{1}{2} r^{-j_{n-1}(t)} + \sum_{k=n-1}^{\infty} r^{-j_k(t)} I_{\{j_{k+1}(t) > j_k(t)\}} \\
&\leq \frac{1}{2} r^{-j_{n-1}(t)} + \sum_{l=j_{n-1}(t)}^{\infty} r^{-l} \leq 3r^{-j_{n-1}(t)}.
\end{aligned}
$$

Hence

$$
\|t - \pi(t)\|_1 \leq 10Mr^{-j_0(t)} + 3\sum_{n=1}^{\infty} r^{-j_{n-1}(t)} |J_n(t)|.
$$

To estimate $|J_n(t)|$ for $n \geq 1$ we may assume that $j_n(t) > j_{n-1}(t)$, since otherwise assumption ii)a) yields $\pi_n(t) = \pi_{n-1}(t)$ and $|J_n(t)| = 0$. For $i \in J_n(t)$ we have either $i \in I_n(t)$ or $m(t,i) = n-1$. Since $|\pi_n(t)_i - t_i| > \frac{1}{2} r^{-j_n(t)}$ for $i \in J_n(t)$ we get by the assumption ii)b)

$$
\frac{1}{4} r^{-2j_n(t)} |J_n(t) \cap I_n(t)| \leq \sum_{i \in I_n(t)} \min\{|t_i - \pi_n(t)_i|^2, r^{-2j_n(t)}\} \leq M2^n r^{-2j_n(t)}.
$$

If $m(t,i) = n-1$ then $|\pi_n(t)_i - \pi_{n-1}(t)_i| > r^{-j_{n-1}(t)}$. Let $n' := \inf\{k \leq n-1\colon j_k(t) = j_{n-1}(t)\}$. Then, since $\pi_n(t) \in A_{n-1}(t) \subset A_{n'}(t)$, $j_{n-1}(t) = j_{n'}(t) > j_{n'-1}(t)$ and $\pi_{n-1}(t) = \pi_{n'}(t)$, the assumption ii)b) used this time for n' yields

$$
\begin{aligned}
&r^{-2j_{n-1}(t)} |\{i\colon\ m(t,i) = n-1\}| \\
&\qquad \leq \sum_{i \in I_{n'}(t)} \min\{|\pi_n(t)_i - \pi_{n-1}(t)_i|^2, r^{-2j_{n-1}(t)}\} \\
&\qquad \leq M2^{n-1} r^{-2j_{n-1}(t)}.
\end{aligned}
$$

Thus

$$
|J_n(t)| \leq |J_n(t) \cap I_n(t)| + |\{i\colon\ m(t,i) = n-1\}| \leq 9M2^{n-1}
$$

and

$$
\|t - \pi(t)\|_1 \leq 10Mr^{-j_0(t)} + 27M\sum_{n=1}^{\infty} 2^{n-1} r^{-j_{n-1}(t)} \leq 37M \sup_{t \in T} \sum_{n=0}^{\infty} 2^n r^{-j_n(t)}.
$$

To bound $\gamma_2(T_2)$ we will construct sets $U_n \subset \ell^2(I)$ such that $|U_0| = 1$, $|U_n| \leq N_n$ for $n \geq 1$ and use [22, Theorem 1.3.5] to get

$$
\gamma_2(T_2) \leq L \sup_{t \in T} \sum_{n=0}^{\infty} 2^{n/2} \mathrm{dist}(\pi(t), U_n). \tag{15}
$$

To this end we define

$$U_n := \{\pi_{m(t,i)\wedge n}(t)\colon\ t \in T\},$$

where $\pi_{m(t,i)\wedge n}(t) = (\pi_{m(t,i)\wedge n}(t)_i)_{i\in I}$. Observe that for $s \in A_n(t)$, $\pi_k(s) = \pi_k(t)$ for $k \le n$ and $\{i\colon\ m(t,i) \ge n\} = \{i\colon\ m(s,i) \ge n\}$ so that $m(t,i)\wedge n = m(s,i) \wedge n$. Hence $|U_n| \le |\mathcal{A}_n| \le N_n$ for $n \ge 1$ and $U_0 = \{\pi_0(T)\}$.

To estimate $\mathrm{dist}(\pi(t), U_n)$, first notice that

$$\mathrm{dist}(\pi(t), U_n) \le \|\pi(t) - \pi_{m(t,i)\wedge n}(t)\|_2 \le \sum_{l=n}^{\infty} \|(\pi_{l+1}(t) - \pi_l(t))1_{\{m(t,i)\ge l+1\}}\|_2.$$

The condition $m(t,i) \ge l+1$ implies $|\pi_{l+1}(t)_i - \pi_l(t)_i| \le r^{-j_l(t)}$. If $j_{l+1}(t) = j_l(t)$ then $\pi_{l+1}(t) = \pi_l(t)$, otherwise $\pi_{l+1}(t) \in A_l(t)$ and by the assumption ii)b)

$$\begin{aligned}\|(\pi_{l+1}(t) - \pi_l(t))1_{\{m(t,i)\ge l+1\}}\|_2^2 &\le \sum_{i\in I_{l+1}(t)} \min\{|\pi_{l+1}(t)_i - \pi_l(t)_i|^2, r^{-2j_l(t)}\} \\ &\le M2^l r^{-2j_l(t)}.\end{aligned}$$

Therefore

$$\mathrm{dist}(\pi(t), U_n) \le \sum_{l=n}^{\infty} \sqrt{M} 2^{l/2} r^{-j_l(t)}$$

and

$$\sum_{n=0}^{\infty} 2^{n/2} \mathrm{dist}(\pi(t), U_n) \le \sqrt{M} \sum_{l=0}^{\infty} 2^{l/2} r^{-j_l(t)} \sum_{n=0}^{l} 2^{n/2} \le L\sqrt{M} \sum_{l=0}^{\infty} 2^l r^{-j_l(t)}.$$

Hence the estimate for $\gamma_2(T_2)$ follows by (15). □

4 Chopping maps

In this section on the base of the so-called chopping maps we define functionals that will play a key role in the proof of Theorem 1.1. Chopping maps were introduced by M. Talagrand in [17], he used them to prove a weak form of the Bernoulli Conjecture ([19] and [22, Chapter 4]).

For $u < v$ we define the non-increasing function $\varphi_{u,v}$ by the formula

$$\varphi_{u,v}(x) := \min\{v, \max\{x, u\}\} - \min\{v, \max\{0, u\}\}.$$

In other words $\varphi_{u,v}$ is the unique continuous function, which is constant on half lines $(-\infty, u]$ and $[v, \infty)$, has slope 1 on the interval $[u, v]$ and takes value

0 at 0. Observe that $|\varphi_{u,v}(x)| \leq v-u$, $|\varphi_{u,v}(x)-\varphi_{u,v}(y)| \leq \min\{|x-y|, v-u\}$ and

$$\varphi_{u_0,u_k}(x) = \sum_{l=1}^{k} \varphi_{u_{l-1},u_l}(x) \quad \text{for } u_0 < u_1 < \ldots < u_k. \tag{16}$$

Functions $\varphi_{u_i,u_{i+1}}$ are called by M. Talagrand *chopping maps*, since the interval $[u_0, u_k]$ is "chopped" into smaller intervals $[u_i, u_{i+1}]$ laying side to side, $\varphi_{u_i,u_{i+1}}$ changes only on intervals $[u_i, u_{i+1}]$ and property (16) holds.

Lemma 4.1. *For any $u_0 < u_1 < \ldots < u_k$ and $x, y \in \mathbb{R}$ we have*

$$\sum_{l=1}^{k} |\varphi_{u_{l-1},u_l}(x) - \varphi_{u_{l-1},u_l}(y)| = |\varphi_{u_0,u_k}(x) - \varphi_{u_0,u_k}(y)| \leq |x-y|. \tag{17}$$

In particular

$$\sum_{l=1}^{k} |\varphi_{u_{l-1},u_l}(x)| \leq |x| \quad \textit{and} \quad \sum_{l=1}^{k} \varphi_{u_{l-1},u_l}(x)^2 \leq x^2. \tag{18}$$

Proof. W.l.o.g. we may assume that $x > y$. Then $\varphi_{u,v}(x) \geq \varphi_{u,v}(y)$ for any u, v and (17) follows by (16). The "in particular" part easily follows taking $y = 0$. □

Let $G_i = \{u_{i,0} < u_{i,1} < \ldots < u_{i,k_i}\}$, $i \in I$ be finite subsets of $\mathbb{R}$ and $\mathcal{G} = (G_i)_{i \in I}$. For $t \in \ell^2(I)$ we define "chopped" Bernoulli processes

$$X_t(G_i, i) := \sum_{l=1}^{k_i} \varphi_{u_{i,l-1},u_{i,l}}(t_i)\varepsilon_{i,l}$$

and

$$X_t(\mathcal{G}) := \sum_{i \in I} X_t(G_i, i) = \sum_{i \in I} \sum_{l=1}^{k_i} \varphi_{u_{i,l-1},u_{i,l}}(t_i)\varepsilon_{i,l}.$$

Note that for $t \in \ell^2(I)$ by (18) we get

$$\sum_{i \in I} \sum_{l=1}^{k_i} |\varphi_{u_{i,l-1},u_{i,l}}(t_i)|^2 \leq \sum_{i \in I} t_i^2 < \infty$$

and $X_t(\mathcal{G})$ is well defined. We also consider the canonical distance $d_{\mathcal{G}}$ associated to the process $X_t(\mathcal{G})$ given by

$$d_{\mathcal{G}}(s,t)^2 := \mathbb{E}|X_t(\mathcal{G}) - X_s(\mathcal{G})|^2 = \sum_{i \in I} \sum_{l=1}^{k_i} |\varphi_{u_{i,l-1},u_{i,l}}(t_i) - \varphi_{u_{i,l-1},u_{i,l}}(s_i)|^2.$$

Proposition 4.2. *i) For any family of finite sets* $\mathcal{G} = (G_i)_{i\in I}$ *and* $T \subset \ell^2(I)$ *we have*

$$\mathbb{E}\sup_{t\in T} X_t(\mathcal{G}) \leq b(T) = \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i\varepsilon_i.$$

ii) If $\mathcal{G} = (G_i)_{i\in I}$ *and* $\mathcal{G}' = (G'_i)_{i\in I}$ *are two families of finite subsets of* $\mathbb{R}$ *such that for all* $i \in I$*,*

$$G_i \subset G'_i,\ \max_i G_i = \max_i G'_i \text{ and } \min_i G_i = \min_i G'_i \tag{19}$$

then for any $T \subset \ell^2(I)$*,*

$$\mathbb{E}\sup_{t\in T} X_t(\mathcal{G}') \leq \mathbb{E}\sup_{t\in T} X_t(\mathcal{G}).$$

Proof. Part i) follows easily by Corollary 2.4 and (17).

To show part ii) let $G_i = \{u_{i,0} < u_{i,1} < \ldots < u_{i,k_i}\}$ and $[u_{i,l-1}, u_{i,l}] \cap G'_i = \{s_{i,l,0} < s_{i,l,1} < \ldots < s_{i,l,k_{i,l}}\}$. Then

$$\mathbb{E}\sup_{t\in T} X_t(\mathcal{G}') = \mathbb{E}\sup_{t\in T}\sum_{i\in I}\sum_{l=1}^{k_i}\sum_{j=1}^{k_{i,l}} \varphi_{s_{i,l,j-1},s_{i,l,j}}(t_i)\varepsilon_{i,l,j}$$

and the assertion follows by Corollary 2.4 and (17). □

Inequality (17) yields

$$d_{\mathcal{G}}(s,t) \leq \|s-t\|_2 \quad \text{for } s,t \in \ell^2(I). \tag{20}$$

The next proposition shows how to compare $d_{\mathcal{G}}$ with $d_{\mathcal{G}'}$.

Proposition 4.3. *Let* $\mathcal{G} = (G_i)_{i\in I}$ *and* $\mathcal{G}' = (G'_i)_{i\in I}$ *be two families of finite subsets of* $\mathbb{R}$ *such that* $G_i \subset G'_i$ *and* $G_i = \{u_{i,0} < u_{i,1} < \ldots < u_{i,k_i}\}$ *for all* $i \in I$*.*
i) If $\max_i G_i = \max_i G'_i$ *and* $\min_i G_i = \min_i G'_i$ *then* $d_{\mathcal{G}'} \leq d_{\mathcal{G}}$*.*
ii) If $|G'_i \cap (u_{i,l-1}, u_{i,l}]| \leq q$ *for all* $i \in I$*,* $1 \leq l \leq k_i$ *then* $d_{\mathcal{G}} \leq \sqrt{q}\, d_{\mathcal{G}'}$*.*

Proof. Part i) follows by (17) and the inequality $\sum_l |a_l|^2 \leq (\sum_l |a_l|)^2$. To show ii) we also use (17) and the bound $(\sum_{l=1}^k |a_l|)^2 \leq k\sum_{l=1}^k |a_l|^2$. □

We are now ready to define functionals and related distances. Let $r \geq 4$ be an integer to be chosen later. For $x \in \mathbb{R}$ and $k \in \mathbb{Z}$ we set

$$G(x,k) := \{pr^{-k} :\ p \in \mathbb{Z}\} \cap [x - 4r^{-k}, x + 4r^{-k}).$$

In other words if $p_k(x) = \lceil r^k x\rceil \in \mathbb{Z}$, i.e. $(p_k(x)-1)r^{-k} < x \leq p_k(x)r^{-k}$ then

$$G(x,k) = \{pr^{-k} \colon p_k(x) - 4 \leq p \leq p_k(x) + 3\}.$$

For an integer $j \geq k$ we set

$$G(x,k,j) := \{pr^{-j}:\ (p_k(x)-4)r^{-k} \leq pr^{-j} \leq (p_k(x)+3)r^{-k}\}$$
$$= \{pr^{-j}:\ w_{k,j}(x) \leq p \leq v_{k,j}\},$$

where $w_{k,j}(x) := (p_k(x)-4)r^{j-k}$ and $v_{k,j}(x) := (p_k(x)+3)r^{j-k}$. Then $G(x,k,k) = G(x,k)$ and

$$j' \geq j \geq k \ \Rightarrow G(x,k,j) \subset G(x,k,j'),\ \min G(x,k,j) = \min G(x,k,j') \text{ and } \max G(x,k,j) = \max G(x,k,j'). \tag{21}$$

For $u \in \ell^2(I)$, integers $j \geq k$ and $J \subset I$ we define the process $X_t(J,u,k,j)$ by

$$X_t(J,u,k,j) := X_t((G(u_i,k,j))_{i\in J}) = \sum_{i\in J} \sum_{p=w_{k,j}(u_i)+1}^{v_{k,j}(u_i)} \varphi_{(p-1)r^{-j},pr^{-j}}(t_i)\varepsilon_{i,p}.$$

For $T \subset \ell^2(I)$ we set

$$F(T,J,u,k,j) := \mathbb{E}\sup_{t\in T} X_t(J,u,k,j).$$

Increasing the parameter j corresponds to "adding" new Bernoulli r.v's, while increasing the parameter k results in "removing" some of Bernoulli r.v's from the process $X_t(J,u,k,j)$.

Let us denote by $d(J,u,k,j)$ the canonical distance associated to the process $(X_t(J,u,k,j))$, i.e.

$$d(J,u,k,j)(t,s) := \Big(\mathbb{E}(X_t(J,u,k,j) - X_s(J,u,k,j))^2\Big)^{1/2}$$

and let $\Delta(T,J,u,k,j)$ denote the diameter of the set $T \subset \ell^2(I)$ with respect to $d(J,u,k,j)$.

Proposition 4.2i) and (20) easily yield the following.

Proposition 4.4. *For any $J \subset I$, $u \in \ell^2(I)$, integers $j \geq k$ and $T \subset \ell^2(I)$ we have*

$$F(T,J,u,k,j) \leq b(T)$$

and

$$\Delta(T,J,u,k,j) \leq \Delta_{\ell^2(I)}(T).$$

We also have the following comparison of distinct functionals and related distances.

Proposition 4.5. *If $J' \subset J \subset I$, integers $j \geq k$ and $j' \geq k'$ satisfy $j' \geq j$ and $k' \geq k$ then for any $u \in \ell^2(I)$ and $T \subset \ell^2(I)$ we have*

$$F(T, J', u, k', j') \leq F(T, J, u, k, j)$$

and

$$\Delta(T, J', u, k', j') \leq \Delta(T, J, u, k, j).$$

Proof. The monotonicity of $F(T, J, u, k, j)$ with respect to the set J and the variable k easily follows by the definition of $X_t(J, u, k, j)$ and (4). The monotonicity with respect to j is a consequence of Proposition 4.2 ii) and (21).

Monotonicity of distances $d(T, J, u, k, j)$ with respect to J and k is quite obvious, and with respect to j follows by Proposition 4.3i). □

We conclude this section with a lemma that gives a lower bound for the constructed distances.

Lemma 4.6. *For $s, t, u \in \ell^2(I)$, $J \subset I$ and $j \geq k$,*

$$d(J, u, k, j)(t, s)^2 \geq \frac{1}{2} \sum_{i \in J} \min\{|s_i - t_i|^2, r^{-2j}\} I_{\{|s_i - u_i| \leq 2r^{-k}\}}.$$

Proof. It is easy to reduce to the case when $|s_i - u_i| \leq 2r^{-k}$ and $|s_i - t_i| \leq r^{-j}$ for all $i \in J$. Then for any $i \in J$, $\min G(u_i, k, j) \leq s_i, t_i \leq \max G(u_i, k, j)$ and for at most two integers p, $\varphi_{(p-1)r^{-j}, pr^{-j}}(t_i) \neq \varphi_{(p-1)r^{-j}, pr^{-j}}(s_i)$. The estimate follows by (16), since $(a + b)^2 \leq 2a^2 + 2b^2$. □

5 Decomposition Lemmas

In this section we derive several decomposition results for our functionals $F(T, J, u, k, j)$. First two propositions are based on results of Section 2. We combine them to get Corollary 5.3 on which we will base our inductive construction of suitable partitions.

The first proposition immediately follows from Corollary 2.9.

Proposition 5.1. *Let $T \subset \ell^2(I)$, $u \in \ell^2(I)$, $J \subset I$ and $j \geq k$. If $r^{-j}\sqrt{\log m} \leq \sigma$ then there exist sets $C_1, \ldots, C_{m-1} \subset T$ such that*

$$\Delta(C_l, J, u, k, j) \leq L_6 \sigma \quad 1 \leq l \leq m - 1$$

and for any $\emptyset \neq D \subset T \setminus \bigcup_{l<m} C_l$ with $\Delta(D, J, u, k, j) \leq \sigma$, it holds

$$F(D, J, u, k, j) \leq F(T, J, u, k, j) - \sigma\sqrt{\log m}.$$

The next result is a consequence of Proposition 2.10.

Proposition 5.2. *Let $u, u' \in \ell^2(I)$, $J \subset I$, $j \geq k$ and $J' \subset J$ be such that $|u_i - u'_i| \leq 2r^{-k}$ for all $i \in J'$. Let T be a subset of $\ell^2(I)$ with $\Delta(T, J, u, k, j+2) \leq c$. If $r^{-j-1}\sqrt{\log m} \leq \sigma$ and $L_8 c \leq \sigma$ then there exist sets $A_1, \ldots, A_m \subset T$ such that*

$$\Delta(A_l, J, u, k, j+1) \leq \sigma \quad \text{for } 1 \leq l \leq m$$

and either $T \subset \bigcup_{l \leq m} A_l$ or

$$F\Big(T \setminus \bigcup_{l=1}^{m} A_l, J', u', j+2, j+2\Big) \leq F(T, J, u, k, j+1) - \frac{1}{L_9}\sigma\sqrt{\log m}. \quad (22)$$

Proof. Let $\mathcal{G} = (G_i)_{i \in J}$, $\mathcal{G}' = (G'_i)_{i \in J}$, where

$$G_i = G(u_i, k, j+1), \ i \in J$$

and

$$G'_i = \begin{cases} G_i & \text{for } i \in J \setminus J', \\ G_i \cup G(u'_i, j+2, j+2) & \text{for } i \in J'. \end{cases}$$

Since $r \geq 4$ and $j \geq k$ we have

$$G(u'_i, j+2, j+2) \subset [u'_i - 4r^{-j-2}, u'_i + 4r^{-j-2}) \subset (u'_i - r^{-k}, u'_i + r^{-k}).$$

Moreover $|u_i - u'_i| \leq 2r^{-k}$ for $i \in J'$, and therefore the sets G_i and G'_i satisfy the condition (19) and Proposition 4.2ii) yields

$$\mathbb{E} \sup_{t \in T} X_t(\mathcal{G}') \leq \mathbb{E} \sup_{t \in T} X_t(\mathcal{G}) = F(T, J, u, k, j+1).$$

Since $|G(u'_i, j+2, j+2)| = 8$, Proposition 4.3ii) with $q = 9$ yields $d_{\mathcal{G}} \leq 3 d_{\mathcal{G}'}$.

For $i \in J'$ we have $|u_i - u'_i| \leq 2r^{-k}$, so that

$$|pr^{-j-2} - u'_i| \leq 4r^{-j-2} \ \Rightarrow |pr^{-j-2} - u_i| \leq 2r^{-k} + 4r^{-j-2} \leq 3r^{-k}$$

and therefore $G(u'_i, j+2, j+2) \subset G(u_i, k, j+2)$. Thus

$$\Delta(T, J', u', j+2, j+2) \leq \Delta(T, J, u, k, j+2) \leq c.$$

We apply Proposition 2.10 with $b = r^{-j-1}, \lambda = 6$ and σ^*, I^*, J^*, T^* instead of σ, I, J and T, where $\sigma^* := \sigma/6$,

$$I^* := \{(i, u) \colon \ i \in J, \ u \in G'_i \setminus \{\min G_i\}\},$$

$$J^* := \{(i, u) \colon \ i \in J', \ u \in G(u'_i, j+2, j+2) \setminus \{\min G(u'_i, j+2, j+2)\}\}$$

and for $A \subset T$,

$$A^* := \{(\varphi_{u-,u}(t_i))_{(i,u)} \colon \ t \in A, \ (i, u) \in I^*\},$$

where for $(i,u)\in I^*$, $u-$ denotes the largest element of G'_i smaller than u. Observe that with the notation of Proposition 2.10 we have for $A\subset T$

$$b_{I^*}(A^*)=\mathbb{E}\sup_{t\in A}X_t(\mathcal{G}') \quad \text{and} \quad b_{J^*}(A^*)=F(A,J',u',j+2,j+2).$$

It is not hard to check that all the assumptions of the proposition are satisfied. Hence there exist sets $A_1,\dots,A_m\subset T$ such that $A_l^*\subset B_{I^*}(t_l^*,\sigma^*)$ for some $t_l^*\in T^*$ and

$$\begin{aligned}F\Big(T\setminus\bigcup_{l=1}^{m}A_l,J',u',j+2,j+2\Big)&\leq \mathbb{E}\sup_{t\in T}X_t(\mathcal{G}')-\Big(\frac{1}{144L_3}\sigma-L_7c\Big)\sqrt{\log m}\\ &\leq F(T,J,u,k,j+1)-\Big(\frac{1}{144L_3}\sigma-L_7c\Big)\sqrt{\log m}.\end{aligned}$$

Hence condition (22) holds if we take $L_8=288L_3L_7$ and $L_9=288L_3$. We conclude by observing that the condition $A_l^*\subset B_{I^*}(t_l^*,\sigma^*)$ implies that for $s,t\in A_l$, we have $d_{\mathcal{G}}(s,t)\leq 3d_{\mathcal{G}'}(s,t)\leq 6\sigma^*=\sigma$, and hence $\Delta(A_l,J,u,k,j+1)\leq\sigma$, $1\leq l\leq m$. □

We finish this section with a crucial corollary which states that our functionals satisfy a Talagrand-type decomposition condition. Talagrand's constructions of admissible partitions for various classes of stochastic processes were based on conditions of similar nature, which roughly state that each set may be partitioned into a number of pieces with either a small diameter or a small value of a suitable functional on subsets of small diameter.

In our case each set may be decomposed into pieces of three types. Pieces of type (C3) have small diameters and pieces of type (C1) have small value of a functional on subsets with sufficiently small diameters, in both cases we do not change values of parameters k, J and u. Pieces satisfying conditions (C2) are of different type – they have both small diameters and small value of functionals, however we increase the parameter k and allow changes in parameters u and J.

Corollary 5.3. *There exists a positive integer r_0 with the following property. Consider $T\subset\ell^2(I)$, $J\subset I$, $u\in\ell^2(I)$, $u'\in T$, $c\geq 0$ and integers $j\geq k$, $n\geq 1$, $r\geq r_0$ and set*

$$J':=\{i\in J:\ |u_i-u'_i|\leq 2r^{-k}\}.$$

Then we can find $p\leq N_n$ and a partition $(A_l)_{l\leq p}$ of T such that each set A_l satisfies one of the following properties:

$$\begin{aligned}&\textit{for any } D\subset A_l \textit{ with } \Delta(D,J,u,k,j+2)\leq\frac{1}{L_{10}}2^{n/2}r^{-j-1}\\ &F(D,J,u,k,j+2)\leq F(T,J,u,k,j+2)-\frac{1}{L_{11}}2^nr^{-j-1}\end{aligned}\qquad\text{(C1)}$$

or

$$\Delta(A_l, J', u', j+2, j+2) \le \Delta(A_l, J, u, k, j+2) \le 2^{n/2} r^{-j-1}, \qquad \text{(C2a)}$$

$$\begin{aligned} F(A_l, J', u', j+2, j+2) &\le F(T, J, u, k, j+1) - \frac{1}{L_{12}} 2^n r^{-j-1} \\ &\le F(T, J, u, k, j) - \frac{1}{L_{12}} 2^n r^{-j-1} \qquad \text{(C2b)} \end{aligned}$$

or

$$\Delta(A_l, J, u, k, j+1) \le 2^{n/2} r^{-j-1}. \qquad \text{(C3)}$$

Proof. Let $m := \sqrt{N_n}$ so that $\sqrt{\log m} = 2^{(n-1)/2}\sqrt{\log 2}$. Without loss of generality we may also assume $L_8 \ge 1$ (where L_8 is the absolute constant given by Proposition 5.2).

We first apply Proposition 5.1 with $j+2$ and $\sigma = \frac{1}{L_6 L_8} 2^{n/2} r^{-j-1}$. Observe that $r^{-j-2}\sqrt{\log m} \le r^{-j-2} 2^{(n-1)/2} \le \sigma$ if $r_0 \ge L_6 L_8$. This way we obtain the decomposition $T = \bigcup_{l \le m-1} C_l \cup A_1$, where $\Delta(C_l, J, u, k, j+2) \le c := \frac{1}{L_8} 2^{n/2} r^{-j-1}$ and A_1 satisfies the condition (C1) with $L_{10} := L_6 L_8$, $L_{11} := (2/\log(2))^{1/2} L_6 L_8$.

Now for $l \le m-1$ we apply Proposition 5.2 with $T = C_l$, $\sigma = 2^{n/2} r^{-j-1}$ and we decompose C_l into at most $m+1$ sets that satisfy either (C2b) with $L_{12} := (2/\log(2))^{1/2} L_9$ or (C3). Since $G(u_i', j+2, j+2) \subset G(u_i, k, j+2)$ for $i \in J'$ and $L_8 \ge 1$ we get $\Delta(C_l, J', u', j+2, j+2) \le \Delta(C_l, J, u, k, j+2) \le c \le 2^{n/2} r^{-j-1}$ and (C2a) follows.

This way we decompose the set T into at most $1 + (m-1)(m+1) = N_n$ sets A_l satisfying one of the conditions (C1)-(C3). □

6 Partition construction

To prove Theorem 1.1 with the use of Theorem 3.1 we need to construct a suitable admissible sequence of partitions $(\mathcal{A}_n)_{n \ge 0}$ of the index set T. In this section we present such a construction.

We use the following notation. For $A \in \mathcal{A}_n$, $n \ge 1$ by A' we will denote the unique set in $\mathcal{A}_{n-1}$ such that $A \subset A'$. For $t \in T$ and $n \ge 0$, $A_n(t)$ is the unique element of $\mathcal{A}_n$ which contains t. Moreover if to each set $A \in \mathcal{A}_n$ is assigned a certain quantity (which may be a point, a number or a set) $\alpha_n(A)$, then to shorten the notation we write $\alpha_n(t)$ for $\alpha_n(A_n(t))$.

The following simple lemma will be very useful. It was proven in [25], we rewrite its proof for the sake of completeness.

Lemma 6.1 ([25, Lemma 2.6.3])**.** *Let $\alpha > 1$ and $(a_n)_{n \ge 0}$ be a sequence of positive numbers such that $\sup_n a_n < \infty$. Define*

$$V := \{m \ge 0 : \ a_n < a_m \alpha^{|n-m|} \text{ for all } n \ge 0, n \ne m\}.$$

Then

$$\sum_{n\geq 0} a_n \leq \frac{2\alpha}{\alpha-1} \sum_{m\in V} a_m.$$

Proof. We define a partial order on $\mathbb{N}$ by $n \prec m$ if and only if $a_m \geq a_n\alpha^{|n-m|}$. Then V is just the set of maximal elements of $\prec$, i.e. if $m \in V$, $m \prec m'$ then $m' = m$. Moreover, since a_n is bounded there cannot exist an infinite sequence of integers increasing with respect to $\prec$. Therefore for each $n \in \mathbb{N}$ there exists $m \in V$ such that $n \prec m$. Thus

$$\sum_{n\geq 0} a_n \leq \sum_{m\in V} a_m \sum_{n\geq 0} \alpha^{-|n-m|} \leq \frac{2\alpha}{\alpha-1} \sum_{m\in V} a_m.$$

□

We are now ready to describe the partition construction. It is based on the iterative application of Corollary 5.3. Unfortunately we will need to control several parameters. The integers $k_n \leq j_n$, the points $u_n \in T$ and the sets $J_n \subset I$ are related to the functionals studied in the previous sections. The parameter $p_n = 0$ means that we will use Corollary 5.3 to decompose the set and $p_n > 0$ means that we will wait $2\kappa - p_n$ steps before doing it.

Let us first summarize the main dependencies between these quantities. The first condition gives initial values of parameters

$$p_0(T) = 0,\ j_0(T) = k_0(T) = j_0,\ J_0(T) = I. \tag{P1}$$

The next requirement is a mild regularity condition (in all conditions below we assume that $A \in \mathcal{A}_n$ for some $n \geq 1$)

$$j_{n-1}(A') \leq j_n(A) \leq j_{n-1}(A') + 2, \quad k_{n-1}(A') \leq k_n(A). \tag{P2}$$

Observe that we do not bound the difference $k_n(A) - k_{n-1}(A')$ from above. Now we state a crucial estimate for the diameter of the set A:

$$p_n(A) = 0 \ \Rightarrow\ \Delta(A, J_n(A), u_n(A), k_n(A), j_n(A)) \leq 2^{n/2} r^{-j_n(A)}, \tag{P3}$$

and its version for a positive value of the counter $p_n(A)$:

$$p_n(A) > 0 \ \Rightarrow\ \Delta(A, J_n(A), u_n(A), k_n(A), j_n(A)) \leq 2^{(n-p_n(A))/2} r^{-j_n(A)+1}. \tag{P4}$$

We require that "parameters k, J, u do not change unless $p_n(A) = 1$"

$$p_n(A) \neq 1 \ \Rightarrow\ u_n(A) = u_{n-1}(A'),\ k_n(A) = k_{n-1}(A'), J_n(A) = J_{n-1}(A'). \tag{P5}$$

Next condition describes how parameters change if $p_n(A) = 1$:

$$p_n(A) = 1 \Rightarrow u_n(A) \in A',\ j_n(A) = j_{n-1}(A') + 2 \text{ and } J_n(A) = \{i \in J_{n-1}(A') \colon\ |u_n(A)_i - u_{n-1}(A')_i| \leq 2r^{-k_{n-1}(A')}\}. \tag{P6}$$

For $p_{n-1}(A') \neq 0$ parameter j_n does not change

$$p_{n-1}(A') \neq 0 \;\Rightarrow\; j_n(A) = j_{n-1}(A'). \tag{P7}$$

Last two conditions describe the behavior of the counter p_n

$$p_n(A) > 0 \;\Rightarrow\; p_n(A) = p_{n-1}(A') + 1, \tag{P8}$$

and

$$p_n(A) = 0 \Rightarrow\; p_{n-1}(A') \in \{0, 2\kappa - 1\},\; j_n(A) \leq j_{n-1}(A') + 1. \tag{P9}$$

Proposition 6.2. *Suppose that $r = 2^\kappa$, where κ is a sufficiently large positive integer and $T \subset \ell^2(I)$ satisfies $\Delta_2(T) \leq r^{-j_0}$. Then there exists an admissible sequence of partitions $(\mathcal{A}_n)_{n\geq 0}$ of T, points $u_n(A) \in T$, sets $J_n(A) \subset I$ and integers $k_n(A) \leq j_n(A)$, $0 \leq p_n(A) \leq 2\kappa - 1$, $A \in \mathcal{A}_n$ which satisfy conditions (P1)-(P9). Moreover for all $t \in T$,*

$$\sum_{n=0}^{\infty} 2^n r^{-j_n(t)} \leq K(r)(r^{-j_0(T)} + b(T)), \tag{23}$$

where $K(r)$ is a constant that depends only on r.

Proof. Define $F_n(A) := F(A, J_n(A), u_n(A), k_n(A), j_n(A))$. We will additionally require the following two conditions, which will help us to prove (23): first

$$p_n(A) = 1 \;\Rightarrow\; F_n(A) \leq F_{n-1}(A') - \frac{1}{L_{12}} 2^{n-1} r^{-j_n(A)+1}, \tag{P10}$$

and second,
if $n \geq 2$, $p_n(A) = p_{n-1}(A') = 0$ and $j_n(A) = j_{n-1}(A')$ then for any $D \subset A$ with $\Delta(D, J_n(A), u_n(A), k_n(A), j_n(A) + 2) \leq \frac{1}{L_{10}} 2^{(n-1)/2} r^{-j_n(A)-1}$ we have

$$\begin{aligned} &F(D, J_n(A), u_n(A), k_n(A), j_n(A) + 2) \\ &\leq F(A', J_n(A), u_n(A), k_n(A), j_n(A) + 2) - \frac{1}{L_{11}} 2^{n-1} r^{-j_n(A)-1} \\ &\leq F(A', J_{n-1}(A'), u_{n-1}(A'), k_{n-1}(A'), j_{n-1}(A')) - \frac{1}{L_{11}} 2^{n-1} r^{-j_n(A)-1}. \end{aligned} \tag{P11}$$

We assume that κ is large enough so that $r \geq \max\{r_0, 4L_{10}^2\}$, where r_0 is given by Corollary 5.3.

We start the construction with $\mathcal{A}_0 = \mathcal{A}_1 = \{T\}$, $k_1(T) = j_1(T) = k_0(T) = j_0(T) = j_0$, $p_1(T) = p_0(T) = 0$, $J_1(T) = J_0(T) = I$ and $u_1(T) = u_0(T) = t_0$, where t_0 is a point in T. Since

$$\Delta(T, J_n(A), u_n(A), k_n(A), j_n(A)) \leq \Delta_2(T) \leq r^{-j_0}$$

conditions (P1)-(P11) are satisfied for $n \le 1$.

Assume now that $\mathcal{A}_n$, $n \ge 1$ is already constructed and fix set $B \in \mathcal{A}_n$. We will split this set into at most N_n sets in $\mathcal{A}_{n+1}$ this way $|\mathcal{A}_{n+1}| \le N_n|\mathcal{A}_n| \le N_n^2 = N_{n+1}$ as required.

If $1 \le p_n(B) \le 2\kappa - 2$ we do not split B. That is, we decide that $B \in \mathcal{A}_{n+1}$ and we set $p_{n+1}(B) := p_n(B) + 1$, $k_{n+1}(B) := k_n(B)$, $j_{n+1}(B) := j_n(B)$, $J_{n+1}(B) := J_n(B)$ and $u_{n+1}(B) := u_n(B)$. It is easy to see that all required conditions holds for B and $n+1$.

If $p_n(B) = 2\kappa - 1$ we do not split B either, but this time we set $p_{n+1}(B) := 0$, $k_{n+1}(B) := k_n(B)$, $j_{n+1}(B) := j_n(B)$, $J_{n+1}(B) := J_n(B)$ and $u_{n+1}(B) := u_n(B)$. The condition (P3) for $A = B$ and $n+1$ follows by (P4) for $A = B$.

Finally assume that $p_n(B) = 0$ then we will split B using Corollary 5.3 with $T = B$, $u = u_n(B)$, u' any point in B, $J = J_n(B)$, $k = k_n(B)$ and $j = j_n(B)$. We obtain a partition $B = \bigcup_{l \le m} A_l$, $m \le N_n$ and each of the sets A_l satisfies one of the conditions (C1)-(C3). Let $A = A_l$ be one of these sets.

If A satisfies (C1) we set $p_{n+1}(A) := 0$, $j_{n+1}(A) := j_n(B)$, $k_{n+1}(A) := k_n(B)$, $J_{n+1}(A) := J_n(B)$ and $u_{n+1}(A) := u_n(B)$. The first inequality in (P11) for A and $n+1$ follows now by (C1) and the second one by Proposition 4.5.

If A satisfies (C2a)-(C2b) we define $p_{n+1}(A) := 1$, $j_{n+1}(A) := k_{n+1}(A) = j_n(B) + 2$, $u_{n+1}(A) := u'$ and

$$J_{n+1}(A) := J' = \{i \in J_n(B)\colon\ |u_n(B)_i - u'_i| \le 2r^{-k_n(B)}\}.$$

Property (P4) for A and $n+1$ follows by (C2a) and property (P10) by (C2b).

Finally if A satisfies (C3) we define $p_{n+1}(A) := 0$, $j_{n+1}(A) = j_n(B) + 1$, $k_{n+1}(A) = k_n(B)$, $J_{n+1}(A) := J_n(B)$ and $u_{n+1}(A) = u_n(B)$. Condition (P3) for A and $n+1$ now follows by (C3).

This way we constructed an admissible partition that satisfies (P1)-(P11). To finish the proof we need to show (23).

Observe that $F_n(A) \le F_{n-1}(A')$: for $p_n(A) = 1$ this obviously follows from (P10), while for $p_n(A) \ne 1$, we have $u_{n-1}(A') = u_n(A)$, $J_{n-1}(A') = J_n(A)$, $j_{n-1}(A') \le j_n(A)$ and $k_{n-1}(A') = k_n(A)$ and we may use Proposition 4.5.

Fix $t \in T$ and define $a_n = a_n(t) := 2^n r^{-j_n(t)}$. If $p_n(t) = 0$ and $n \ge 2$ then either $j_{n-1}(t) < j_n(t)$ and $a_{n-1} > a_n$ or $j_{n-1}(t) = j_n(t)$, $p_{n-1}(t) = 0$, which by (P11) gives $a_n \le 2L_{11} r F_{n-1}(t) \le 2L_{11} r b(T)$ or $p_{n-1}(t) = 2\kappa - 1$, which yields $p_{n-2\kappa}(t) = 0$, $j_{n-2\kappa}(t) = j_n(t) - 2$ and $a_{n-2\kappa} = a_n$. If $p_n(t) > 0$ then taking $n' := \inf\{m \ge n\colon\ p_m(t) = 0\}$ we get $j_{n'}(t) = j_n(t)$, $p_{n'}(t) = 0$ and $a_n < a_{n'}$. This shows that

$$\sup_n a_n \le \max\{a_0, a_1, 2L_{11} r b(T)\} \le K(r)(r^{-j_0} + b(T)) < \infty.$$

Let

$$V_0 := \{n \geq 0\colon\ a_m < 2^{|m-n|} a_n \text{ for all } m \geq 0,\ m \neq n\}.$$

If $n \in V_0$ then $a_{n+1} = 2^{n+1} r^{-j_{n+1}(t)} < 2a_n = 2^{n+1} r^{-j_n(t)}$, so that

$$V_0 \subset V_1 := \{n \geq 0\colon j_n(t) < j_{n+1}(t)\}.$$

By Lemma 6.1 with $\alpha = 2$ we have

$$\sum_{n \geq 0} a_n \leq 4 \sum_{n \in V_0} a_n \leq 4 \sum_{n \in V_1} a_n.$$

Let us enumerate the elements of V_1 as $1 \leq n_0 < n_1 < n_2 < \ldots$ and set

$$V_2 := \{n_q\colon\ a_{n_m} < 2^{|m-q|} a_{n_q} \text{ for all } m \geq 0,\ m \neq q\}.$$

Lemma 6.1 applied once again implies

$$\sum_{n \geq 0} a_n \leq 4 \sum_{n \in V_1} a_n \leq 16 \sum_{n \in V_2} a_n.$$

Fix $n = n_q \in V_2$. If $j_{n-1}(t) < j_n(t)$ then $n - 1 = n_{q-1}$ and (since $r \geq 4$)

$$a_{n_{q-1}} = a_{n-1} \geq \frac{r}{2} a_n \geq 2a_n,$$

which contradicts the definition of V_2. Hence $j_{n-1}(t) = j_n(t) < j_{n+1}(t)$. We have the following 4 possibilities.

1. $j_{n+1}(t) = j_n(t) + 2$, then $p_{n+1}(t) = 1$ and by (P10) applied with $A = A_{n+1}(t)$,

$$a_n = r2^n r^{-j_{n+1}(t)+1} \leq L_{12} r (F_n(t) - F_{n+1}(t)).$$

2. $j_{n+1}(t) = j_n(t) + 1$ and $j_{n_{q+1}+1}(t) = j_{n_{q+1}}(t) + 2$, then $p_{n_{q+1}+1}(t) = 1$, $j_{n_{q+1}+1}(t) = j_n(t) + 3$ and by (P10) applied with $A = A_{n_{q+1}+1}(t)$,

$$a_n \leq \frac{1}{4} r^3 a_{n_{q+1}+1} \leq \frac{1}{2} L_{12} r^2 (F_{n_{q+1}}(t) - F_{n_{q+1}+1}(t)).$$

3. $p_{n-1}(t) = 2\kappa - 1$, then $p_{n-2\kappa+1}(t) = 1$, $j_{n-2\kappa}(t) < j_{n-2\kappa+1}(t) = j_n(t)$, so $n - 2\kappa = n_{q-1}$ and by (P10) applied with $A = A_{n_{q-1}+1}(t)$,

$$\begin{aligned} a_n = 2^{2\kappa-1} a_{n_{q-1}+1} &\leq 2^{2\kappa} L_{12} r^{-1} (F_{n_{q-1}}(t) - F_{n_{q-1}+1}(t)) \\ &= L_{12} r (F_{n_{q-1}}(t) - F_{n_{q-1}+1}(t)). \end{aligned}$$

4. $p_{n-1}(t) = 0$, $j_{n+1}(t) = j_n(t) + 1$ and $j_{n_{q+1}+1}(t) = j_{n_{q+1}}(t) + 1$. Then $p_{n_{q+1}+1}(t) = 0$, moreover by the definition of V_2,

$$2^{n_{q+1}} r^{-j_n(t)-1} = a_{n_{q+1}} < 2a_{n_q} = 2^{n+1} r^{-j_n(t)},$$

which yields $n_{q+1} - n \leq \kappa$. In particular this implies $p_m(t) = 0$ for all $n \leq m \leq n_{q+1} + 1$. Hence $k_{n_{q+1}+1}(t) = k_n(t)$, $j_{n_{q+1}+1}(t) = j_n(t) + 2$, $u_{n_{q+1}+1}(t) = u_n(t)$ and $J_{n_{q+1}+1}(t) = J_n(t)$. Therefore (P3) used for $n = n_{q+1} + 1$ and $A = A_{n_{q+1}+1}(t)$ implies

$$\begin{aligned}\Delta(A_{n_{q+1}+1}(t), J_n(t), u_n(t), k_n(t), j_n(t) + 2) &\leq 2^{(n_{q+1}+1)/2} r^{-j_n(t)-2} \\ &\leq \frac{1}{L_{10}} 2^{(n-1)/2} r^{-j_n(t)-1},\end{aligned}$$

where the last estimate follows since $n_{q+1} - n \leq \kappa$ and $r = 2^\kappa \geq (2L_{10})^2$. Then either $q = 0$ or $n \geq 2$ and then we may apply (P11) to $D = A_{n_{q+1}+1}$, $A = A_n(t)$ and get

$$a_n \leq 2L_{11} r (F_{n-1}(t) - F_{n_{q+1}+1}(t)).$$

This shows that for $n = n_q \in V_2$, either $q = 0$ or $a_n \leq K(r)(F_{n_{q-1}}(t) - F_{n_{q+2}}(t))$. By monotonicity of the map $l \mapsto F_{n_l}(t)$ this gives (with a value of $K(r)$ which may change at each occurrence)

$$\sum_{n \geq 0} a_n \leq 16 \sum_{n \in V_2} a_n \leq 16 a_{n_0} + K(r) F_0(T) \leq K(r)(r^{-j_0} + b(T)).$$

□

7 Proofs of the Main Results

We are now ready to present proofs of the main Theorem 1.1 and Corollary 1.2. By Theorem 3.1 in order to decompose the index set T it is enough to find a suitable admissible sequence of partitions. To construct such a sequence we defined in Section 4 a family of functionals and showed with the help of results gathered in Section 2 that they satisfy the Talagrand-type decomposition condition presented in Corollary 5.3. Then iterative application of the lattest result enabled us to construct inductively a sequence of partitions. To conclude we need to verify that our sequence satifies conditions from Theorem 3.1. In particular we need to construct (on the base of points $u_n(A)$) points $\pi_n(A)$ and show that sets $I_n(A)$ defined in Theorem 3.1 are related to sets $J_n(A)$ from Proposition 6.2. .

Proof of Theorem 1.1. By homogeneity we may assume that $b(T) = \frac{1}{4}$, then $\Delta_2(T) \leq 1$ by Lemma 2.1. We apply Proposition 6.2 with $j_0 = 0$ and get an admissible sequence of partitions $(\mathcal{A}_n)_{n \geq 0}$, numbers $p_n(A), k_n(A), j_n(A)$ and points $u_n(A)$. First we inductively define points $\pi_n(A)$. We set $\pi_0(T) = u_0(T)$ and for $A \in \mathcal{A}_n$, $n \geq 1$ we define $\pi_n(A) = \pi_{n-1}(A')$ if $j_n(A) = j_{n-1}(A')$, $\pi_n(A) = u_n(A)$ if $p_n(A) = 1$ and choose for $\pi_n(A)$ an arbitrary point in A if $p_n(A) = 0$ and $j_n(A) > j_{n-1}(A')$.

As in Theorem 3.1 we set

$$I_n(t) := \big\{i \in I\colon\ |\pi_{q+1}(t)_i - \pi_q(t)_i| \le r^{-j_q(t)} \text{ for } 0 \le q \le n-1\big\}.$$

First we show that

$$|\pi_{n+1}(t)_i - u_n(t)_i| \le 2r^{-k_n(t)} \quad \text{ for } i \in I_{n+1}(t). \tag{24}$$

To this aim we define $J' = \{0\} \cup \{n \ge 1\colon\ p_n(t) = 1\}$. Then $\pi_n(t) = u_n(t)$ for $n \in J'$. Fix n and let n' be the largest element of J' such that $n' \le n$. Then by (P5) $u_n(t) = u_{n'}(t) = \pi_{n'}(t)$ and $k_n(t) = k_{n'}(t)$. Therefore for $i \in I_{n+1}(t)$,

$$\begin{aligned}|\pi_{n+1}(t)_i - u_n(t)_i| = |\pi_{n+1}(t)_i - \pi_{n'}(t)_i| &\le \sum_{q=n'}^{n} |\pi_{q+1}(t)_i - \pi_q(t)_i| \\ &\le \sum_{j \ge j_{n'}(t)} r^{-j} \le 2r^{-j_{n'}(t)} \le 2r^{-k_{n'}(t)} = 2r^{-k_n(t)}.\end{aligned}$$

Now we inductively show that $I_n(t) \subset J_n(t)$. For $n = 0$ both sets equals I. If $p_{n+1}(t) \ne 1$ then $I_{n+1}(t) \subset I_n(t) \subset J_n(t) = J_{n+1}(t)$ and if $p_{n+1}(t) = 1$ then $\pi_{n+1}(t) = u_{n+1}(t)$ so by (24), $|u_{n+1}(t)_i - u_n(t)_i| \le 2r^{-k_n(t)}$ for $i \in I_{n+1}(t)$ hence by (P6) and the induction assumption $I_{n+1}(t) \subset J_{n+1}(t)$.

Finally assume that $A \in \mathcal{A}_n$, $j_n(A) > j_{n-1}(A')$ and $t \in A$. Then $p_{n-1}(A') = 0$, $t, \pi_n(A) \in A'$, $I_n(A) \subset J_n(A) \subset J_{n-1}(A')$ and $|\pi_n(A)_i - u_{n-1}(A')_i| \le 2r^{-k_{n-1}(A')}$ for $i \in I_n(A)$. Hence Lemma 4.6 (applied with $J = I_n(A)$, $u = u_{n-1}(A')$, $s = \pi_n(A)$, $j = j_{n-1}(A')$ and $k = k_{n-1}(A')$), (P3) and (P2) yield

$$\begin{aligned}&\sum_{i \in I_n(A)} \min\{(t_i - \pi_n(A)_i)^2, r^{-2j_n(A)}\} \\ &\qquad \le \sum_{i \in I_n(A)} \min\{(t_i - \pi_n(A)_i)^2, r^{-2j_{n-1}(A')}\} \\ &\qquad \le 2\Delta(A', J_{n-1}(A'), u_{n-1}(A'), k_{n-1}(A'), j_{n-1}(A'))^2 \\ &\qquad \le 2^n r^{-2j_{n-1}(A')} \le r^4 2^n r^{-2j_n(A)}.\end{aligned}$$

Therefore all assumptions of Theorem 3.1 are satisfied with $M = r^4$ and Theorem 1.1 follows by (13) and (23) (since $r^{-j_0} = 1 = 4b(T)$).

□

Proof of Corollary 1.2. By Theorem 1.1 we know that $T \subset T_1 + T_2$ with $\sup_{t \in T_1} \|t\|_1 \le Lb(T)$ and $g(T_2) \le Lb(T)$. Then

$$T - T \subset (T_1 - T_1) + (T_2 - T_2) \subset \operatorname{conv}\{2(T_1 - T_1), 2(T_2 - T_2)\}.$$

Obviously $T_1 - T_1 \subset L\overline{\mathrm{conv}}\{e_i\colon\ i \in I\}$, where $(e_i)_{i\in I}$ is the canonical basis of $\ell^2(I)$. The majorizing measure theorem for Gaussian processes implies (cf. [22, Theorem 2.1.8]) that we can find vectors $(s^n)_{n\geq 1}$ in ℓ^2 such that $T_2 - T_2 \subset \overline{\mathrm{conv}}\{s^n\colon\ n \geq 1\}$ and $\sqrt{\log(n+1)}\|s_n\|_2 \leq Lg(T_2) \leq Lb(T)$. To finish the proof it is enough to notice that $\|X_{e_i}\|_p = \|\varepsilon_i\|_p = 1$ for any $p > 0$ and that by Khinthine's inequality $\|X_t\|_p \leq L\sqrt{p}\|t\|_2$ for $p \geq 1$. □

8 Selected Applications

The Bernoulli Conjecture was motivated by the following question of X. Fernique concerning random Fourier series. Let G be a compact Abelian group and $(F, \|\ \|)$ be a complex Banach space. Consider (finitely many) vectors $v_i \in F$ and characters χ_i on G. X. Fernique [6] showed that

$$\mathbb{E}\sup_{h\in G}\Big\|\sum_i v_i g_i \chi_i(h)\Big\| \leq L\Big(\mathbb{E}\Big\|\sum_i v_i g_i\Big\| + \sup_{\|x^*\|\leq 1}\mathbb{E}\sup_{h\in G}\Big|\sum_i x^*(v_i) g_i \chi_i(h)\Big|\Big)$$

and asked whether similar bound holds if one replaces Gaussian r.v's by random signs. Theorem 1.1 yields an affirmative answer.

Theorem 8.1. *For any compact Abelian group G any finite collection of vectors v_i in a complex Banach space $(F, \|\ \|)$ and characters χ_i on G we have*

$$\mathbb{E}\sup_{h\in G}\Big\|\sum_i v_i \varepsilon_i \chi_i(h)\Big\| \leq L\Big(\mathbb{E}\Big\|\sum_i v_i \varepsilon_i\Big\| + \sup_{\|x^*\|\leq 1}\mathbb{E}\sup_{h\in G}\Big|\sum_i x^*(v_i) \varepsilon_i \chi_i(h)\Big|\Big).$$

Remark. Since $\chi_i(e) = 1$, where e is the neutral element of G we have

$$\max\Big\{\mathbb{E}\Big\|\sum_i v_i \varepsilon_i\Big\|, \sup_{\|x^*\|\leq 1}\mathbb{E}\sup_{h\in G}\Big|\sum_i x^*(v_i)\varepsilon_i\chi_i(h)\Big|\Big\} \leq \mathbb{E}\sup_{h\in G}\Big\|\sum_i v_i\varepsilon_i\chi_i(h)\Big\|.$$

Therefore Theorem 8.1 gives a two-sided bound on $\mathbb{E}\sup_{h\in G}\|\sum_i v_i\varepsilon_i\chi_i(h)\|$.

Proof of Theorem 8.1. We need to show that for any bounded set $T \subset \mathbb{C}^n$, $n < \infty$,

$$\mathbb{E}\sup_{h\in G, t\in T}\Big|\sum_{i=1}^n t_i\varepsilon_i\chi_i(h)\Big| \leq L\Big(\mathbb{E}\sup_{t\in T}\Big|\sum_{i=1}^n t_i\varepsilon_i\Big| + \sup_{t\in T}\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^n t_i\varepsilon_i\chi_i(h)\Big|\Big). \tag{25}$$

Let $M := \mathbb{E}\sup_{t\in T}|\sum_{i=1}^n t_i\varepsilon_i|$. Theorem 1.1 implies that we can find a decomposition $T \subset T_1 + T_2$, with $\sup_{t^1\in T_1}\|t^1\|_1 \leq LM$ and

$$\mathbb{E}\sup_{t^2\in T_2}\Big|\sum_{i=1}^n t_i^2 g_i\Big| \leq LM. \tag{26}$$

Obviously

$$\mathbb{E}\sup_{h\in G,t\in T}\Big|\sum_{i=1}^{n}t_i\varepsilon_i\chi_i(h)\Big|$$
$$\leq \mathbb{E}\sup_{h\in G,t^1\in T_1}\Big|\sum_{i=1}^{n}t_i^1\varepsilon_i\chi_i(h)\Big|+\mathbb{E}\sup_{h\in G,t^2\in T_2}\Big|\sum_{i=1}^{n}t_i^2\varepsilon_i\chi_i(h)\Big|. \quad (27)$$

Since $|\sum_{i=1}^{n}t_i^1\varepsilon_i\chi_i(h)|\leq\sum_{i=1}^{n}|t_i^1||\chi_i(h)|=\|t^1\|_1$ we get

$$\mathbb{E}\sup_{h\in G,t^1\in T_1}\Big|\sum_{i=1}^{n}t_i^1\varepsilon_i\chi_i(h)\Big|\leq\sup_{t\in T^1}\|t^1\|_1\leq LM. \quad (28)$$

Estimate (2) and Fernique's theorem imply

$$\mathbb{E}\sup_{h\in G,t^2\in T_2}\Big|\sum_{i=1}^{n}t_i^2\varepsilon_i\chi_i(h)\Big|\leq\sqrt{\frac{\pi}{2}}\mathbb{E}\sup_{h\in G,t^2\in T_2}\Big|\sum_{i=1}^{n}t_i^2g_i\chi_i(h)\Big|$$
$$\leq L\Big(\mathbb{E}\sup_{t^2\in T_2}\Big|\sum_{i=1}^{n}t_i^2g_i\Big|+\sup_{t^2\in T_2}\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i^2g_i\chi_i(h)\Big|\Big). \quad (29)$$

The Marcus-Pisier estimate [13] yields for any $t^2\in T_2$,

$$\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i^2g_i\chi_i(h)\Big|\leq L\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i^2\varepsilon_i\chi_i(h)\Big|. \quad (30)$$

Since we may assume that $T_2\subset T-T_1$ we get

$$\sup_{t^2\in T_2}\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i^2\varepsilon_i\chi_i(h)\Big|$$
$$\leq\sup_{t\in T}\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i\varepsilon_i\chi_i(h)\Big|+\sup_{t^1\in T_1}\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i^1\varepsilon_i\chi_i(h)\Big|$$
$$\leq\sup_{t\in T}\mathbb{E}\sup_{h\in G}\Big|\sum_{i=1}^{n}t_i\varepsilon_i\chi_i(h)\Big|+LM. \quad (31)$$

Estimate (25) follows by (26)-(31). □

Another consequence of Theorem 1.1 is a Levy-Ottaviani type maximal inequality for VC-classes (see [9] for details). Recall that a class $\mathcal{C}$ of subsets of I is called a Vapnik-Chervonenkis class (or in short a VC-class) of order at most d if for any set $A\subset I$ of cardinality $d+1$ we have $|\{C\cap A\colon\ C\in\mathcal{C}\}|<2^{d+1}$.

Theorem 8.2. *Let $(X_i)_{i\in I}$ be independent random variables in a separable Banach space $(F,\|\ \|)$ such that $|\{i\colon\ X_i\neq 0\}|<\infty$ a.s. and $\mathcal{C}$ be a countable VC-class of subsets of I of order d. Then*

$$\mathbb{P}\Big(\sup_{C\in\mathcal{C}}\Big\|\sum_{i\in C}X_i\Big\|\geq u\Big)\leq K(d)\sup_{C\in\mathcal{C}\cup\{I\}}\mathbb{P}\Big(\Big\|\sum_{i\in C}X_i\Big\|\geq\frac{u}{K(d)}\Big)\quad\text{for } u>0,$$

where $K(d)$ is a constant that depends only on d. Moreover if the variables X_i are symmetric then

$$\mathbb{P}\Big(\sup_{C\in\mathcal{C}}\Big\|\sum_{i\in C}X_i\Big\|\geq u\Big)\leq K(d)\mathbb{P}\Big(\Big\|\sum_{i\in I}X_i\Big\|\geq\frac{u}{K(d)}\Big)\quad\text{for } u>0.$$

Remark. Analysis of the proof shows that $K(d)\leq L\sqrt{d}$.

It is easy to see (taking $F=\mathbb{R}$, $X_i=\varepsilon_i v$ for $i\in I_0$ and $X_i=0$ otherwise, where I_0 is a finite subset of I and v is any nonzero vector in F) that being a VC-class is a necessary assumption even in the scalar case.

Maximal inequalities of this type may be used to derive Itô-Nisio type theorems reducing almost sure statements to statements in probability and as a consequence obtain various limit type theorems for VC-classes. As an example of application we present a uniform Strong Law of Large Numbers.

Corollary 8.3. *Let $(X_i)_{i\geq 1}$ be independent symmetric r.v's with values in a separable Banach space $(F,\|\ \|)$ such that $\frac{1}{a_n}\sum_{i=1}^n X_i\to 0$ a.s.. Then for any VC-class $\mathcal{C}$ of subsets of $\mathbb{N}$ we have*

$$\lim_{n\to\infty}\frac{1}{a_n}\max_{C\in\mathcal{C}}\Big\|\sum_{i\in C\cap\{1,\ldots,n\}}X_i\Big\|=0\ \text{a.s.}.$$

Proof. Let n_0 be a fixed positive integer. Then for any $A\subset\mathbb{N}$

$$\max_{n\geq n_0}\frac{1}{a_n}\Big\|\sum_{i\in A\cap\{1,\ldots,n\}}X_i\Big\|=\Big\|\sum_{i\in A}Y_i\Big\|,$$

where Y_i are random variables in $\ell^\infty(F)$ given by $Y_i(n)=0$ for $n<n_0$ or $i>n$ and $Y_i(n)=\frac{1}{a_n}X_i$ for $i\leq n\geq n_0$. Applying Theorem 8.2 to random variables Y_i we get for any $u>0$,

$$\mathbb{P}\Big(\max_{n\geq n_0}\frac{1}{a_n}\max_{C\in\mathcal{C}}\Big\|\sum_{i\in C\cap\{1,\ldots,n\}}X_i\Big\|\geq u\Big)\leq K\mathbb{P}\Big(\max_{n\geq n_0}\frac{1}{a_n}\Big\|\sum_{i=1}^n X_i\Big\|\geq\frac{u}{K}\Big)$$

where K is a constant that depends only on $\mathcal{C}$ and the assertion easily follows. □

Sketch of the proof of Theorem 8.2. It is rather a standard exercise (c.f. [9]) to reduce to the case when I is finite and $X_i = v_i\varepsilon_i$ for some vectors $v_i \in F$. Using concentration properties of Bernoulli processes it is enough to show that for any bounded symmetric set $T \subset \mathbb{R}^I$ and any VC-class of order d,

$$\mathbb{E}\sup_{C\in\mathcal{C}}\sup_{t\in T}\Big|\sum_{i\in C} t_i\varepsilon_i\Big| \le K(d)\mathbb{E}\sup_{t\in T}\Big|\sum_{i\in I} t_i\varepsilon_i\Big| = K(d)b(T). \qquad (32)$$

Let $T \subset T_1 + T_2$ be a decomposition given by Theorem 1.1. We may also assume that T_1 and T_2 are symmetric. Obviously $|\sum_{i\in C} t_i^1\varepsilon_i\Big| \le \sum_{i\in C}|t_i^1| \le \|t^1\|_1$, hence

$$\mathbb{E}\sup_{C\in\mathcal{C}}\sup_{t^1\in T_1}\Big|\sum_{i\in C} t_i^1\varepsilon_i\Big| \le \sup_{t^1\in T_1}\|t^1\|_1 \le Lb(T). \qquad (33)$$

Inequality (2) gives

$$\mathbb{E}\sup_{C\in\mathcal{C}}\sup_{t^2\in T_2}\Big|\sum_{i\in C} t_i^2\varepsilon_i\Big| \le \sqrt{\frac{\pi}{2}}\mathbb{E}\sup_{C\in\mathcal{C}}\sup_{t^2\in T_2}\Big|\sum_{i\in C} t_i^2 g_i\Big|. \qquad (34)$$

The result of Krawczyk [7] and the choice of T_2 yields

$$\mathbb{E}\sup_{C\in\mathcal{C}}\sup_{t^2\in T_2}\Big|\sum_{i\in C} t_i^2 g_i\Big| \le K(d)g(T_2) \le K(d)b(T). \qquad (35)$$

Estimates (33)-(35) imply (32). □

Remark. Alternatively one may prove (32) using Corollary 1.2 and the fact that maximal inequalities hold for $F = \mathbb{R}$.

9 Further Questions

It is natural to ask for bounds on suprema for another classes of stochastic processes. The majorizing measure upper bound works in quite general situations, cf. [1]. Two-sided estimates are known however only in very few cases. For "canonical processes" of the form $X_t = \sum_{i\ge1} t_i X_i$, where X_i are independent centered r.v's results in the spirit of Corollary 1.2 were obtained for certain symmetric variables with log-concave tails [18, 8].

A basic important class of canonical processes worth investigation is a class of "selector processes" of the form

$$X_t = \sum_{i\ge1} t_i(\delta_i - \delta), \quad t \in \ell^2,$$

where $(\delta_i)_{i\ge1}$ are independent random variables such that $\mathbb{P}(\delta_i = 1) = \delta = 1 - \mathbb{P}(\delta_i = 0)$. We may bound the quantity

$$\delta(T) := \mathbb{E}\sup_{t\in T}\Big|\sum_{i\ge1} t_i(\delta_i - \delta)\Big|, \quad T \subset \ell^2$$

in two ways.

First bound for $\delta(T)$ follows by a pointwise estimate. Namely let $(\delta_i')_{i\geq 1}$ be an independent copy of $(\delta_i)_{i\geq 1}$, then by Jensen's inequality,

$$\delta(T) \leq \mathbb{E}\sup_{t\in T}\Big|\sum_{i\geq 1} t_i(\delta_i - \delta_i')\Big| \leq 2\mathbb{E}\sup_{t\in T}\Big|\sum_{i\geq 1} t_i\delta_i\Big| \leq 2\mathbb{E}\sup_{t\in T}\sum_{i\geq 1}|t_i|\delta_i.$$

Second estimate is based on chaining. To introduce it we define for $\alpha > 0$ and a metric space (T,d),

$$\gamma_\alpha(T,d) := \inf\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/\alpha}\Delta(A_n(t)),$$

where as in the definition of γ_2 the infimum runs over all admissible sequences of partitions $(\mathcal{A}_n)_{n\geq 0}$ of the set T. Bernstein's inequality implies that for $X_t = \sum_{i\geq 1} t_i(\delta_i - \delta)$ and $\delta \in (0,1/2]$ we have

$$\mathbb{P}(|X_t - X_s| \geq u) \leq 2\exp\Big(-\min\Big\{\frac{u^2}{L\delta d_2(s,t)^2}, \frac{u}{Ld_\infty(s,t)}\Big\}\Big) \quad \text{for } s,t\in\ell^2,$$

where $d_p(t,s) := \|t-s\|_p$ denotes the ℓ^p-distance. This together with a chaining argument [22, Theorem 1.2.7] yields

$$\delta(T) \leq L(\sqrt{\delta}\gamma_2(T,d_2) + \gamma_1(T,d_\infty)).$$

The next conjecture, formulated by M. Talagrand [24], states that there are no other ways to bound $\delta(T)$ as the combination of the above two estimates and the fact that $\delta(T_1+T_2) \leq \delta(T_1) + \delta(T_2)$.

Conjecture 9.1. *Let $0 < \delta \leq 1/2$, δ_i be independent random variables such that $\mathbb{P}(\delta_i = 1) = \delta = 1 - \mathbb{P}(\delta_i = 0)$ and $\delta(T) := \mathbb{E}\sup_{t\in T}|\sum_{i\geq 1} t_i(\delta_i - \delta)|$ for $T \subset \ell^2$. Then for any set T with $\delta(T) < \infty$ one may find a decomposition $T \subset T_1 + T_2$ such that*

$$\mathbb{E}\sup_{t\in T_1}\sum_{i\geq 1}|t_i|\delta_i \leq L\delta(T), \quad \sqrt{\delta}\gamma_2(T_2,d_2) \leq L\delta(T) \quad \text{and} \quad \gamma_1(T_2,d_\infty) \leq L\delta(T).$$

It may be showed that for $\delta = 1/2$ the above conjecture follows from Theorem 1.1.

Since any mean zero random variable is a mixture of mean zero two-points random variables selector processes are strictly related to empirical processes

$$Z_f := \frac{1}{\sqrt{N}}\sum_{i\leq N}(f(X_i) - \mathbb{E}f(X_i)), \quad f\in\mathcal{F},$$

where $(X_i)_{i\leq N}$ are i.i.d. random variables and $\mathcal{F}$ is a class of measurable functions. Let

$$S_N(\mathcal{F}) := \mathbb{E}\sup_{f\in\mathcal{F}}|Z_f| = \frac{1}{\sqrt{N}}\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\leq N}(f(X_i) - \mathbb{E}f(X_i))\Big|.$$

As for selector processes there are two distinct ways to bound $S_N(\mathcal{F})$. The first one is to use the trivial pointwise bound $|\sum_{i\le N} f(X_i)| \le \sum_{i\le N} |f(X_i)|$. The second is based on chaining and Bernstein's inequality

$$\mathbb{P}\Big(\Big|\sum_{i\le N}(f(X_i)-\mathbb{E}f(X_i))\Big| \ge t\Big) \le 2\exp\Big(-\min\Big\{\frac{t^2}{4N\|f\|_2^2}, \frac{t}{4\|f\|_\infty}\Big\}\Big), \quad (36)$$

where $\|f\|_p$ denotes the L_p-norm of $f(X_i)$. Similar chaining arguments as in the case of selector processes give

$$S_N(\mathcal{F}) \le L\Big(\gamma_2(\mathcal{F}_2, d_2) + \frac{1}{\sqrt{N}}\gamma_1(\mathcal{F}_2, d_\infty)\Big),$$

where $d_p(f,g) := \|f-g\|_p$.

The following conjecture asserts that there are no other ways to bound suprema of empirical processes.

Conjecture 9.2. *Suppose that $\mathcal{F}$ is a countable class of measurable functions. Then one can find a decomposition $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ such that*

$$\mathbb{E}\sup_{f_1\in\mathcal{F}_1}\sum_{i\le N}|f_1(X_i)| \le \sqrt{N} S_N(\mathcal{F}),$$

$$\gamma_2(\mathcal{F}_2, d_2) \le L S_N(\mathcal{F}) \quad \textit{and} \quad \gamma_1(\mathcal{F}_2, d_\infty) \le L\sqrt{N} S_N(\mathcal{F}).$$

Related conjectures with a much more detailed discussion may be found in [23] and [22, Chapter 12].

References

[1] W. Bednorz, *A theorem on majorizing measures*, Ann. Probab. **34** (2006), 1771–1781.

[2] W. Bednorz and R. Latała, *On the suprema of Bernoulli Processes*, C. R. Math. Acad. Sci. Paris **351** (2013), 131–134.

[3] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Functional Analysis **1** (1967), 290–330.

[4] R. M. Dudley, *Uniform central limit theorems*, Cambridge University Press, Cambridge, 1999.

[5] X. Fernique, *Regularité des trajectoires des fonctions aléatoires gaussiennes*, École d'Été de Probabilités de Saint-Flour, IV-1974, Lecture Notes in Mathematics 480, 1–96, Springer, Berlin, 1975.

[6] X. Fernique, *Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens*, Université de Montréal, Centre de Recherches Mathématiques, Montreal, QC, 1997.

[7] L. Krawczyk, *Maximal inequality for Gaussian vectors*, Bull. Polish Acad. Sci. Math. **44** (1996), 157–160.

[8] R. Latała, *Sudakov minoration principle and supremum of some processes*, Geom. Funct. Anal. **7** (1997), 936–953.

[9] R. Latała, *A note on the maximal inequalities for VC classes*, Advances in stochastic inequalities (Atlanta, GA, 1997), 125–134, Contemp. Math. **234**, Amer. Math. Soc., Providence, RI, 1999.

[10] R. Latała, *On the boundedness of Bernoulli processes over thin sets*, Electron. Commun. Probab. **13** (2008), 175–186.

[11] M. Ledoux, *The concentration of measure phenomenom*, AMS, Providence, RI, 2001.

[12] M. Ledoux and M. Talagrand, *Probability in Banach spaces. Isoperimetry and processes*, Springer-Verlag, Berlin, 1991.

[13] M. B. Marcus and G. Pisier, *Random Fourier series with applications to harmonic analysis*, Annals of Mathematics Studies **101**, Princeton University Press, Princeton, N.J., University of Tokyo Press, Tokyo, 1981.

[14] M. Talagrand, *Regularity of Gaussian processes*, Acta Math. **159** (1987), 99–149.

[15] M. Talagrand, *An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities*, Proc. Amer. Math. Soc. **104** (1988), 905–909.

[16] M. Talagrand, *A simple proof of the majorizing measure theorem*, Geom. Funct. Anal. **2** (1992), 118–125.

[17] M. Talagrand, *Regularity of infinitely divisible processes*, Ann. Probab. **21** (1993), 362–432.

[18] M. Talagrand, *The supremum of some canonical processes*, Amer. J. Math. **116** (1994), 284–325.

[19] M. Talagrand, *Constructions of majorizing measures, Bernoulli processes and cotype*, Geom. Funct. Anal. **4** (1994), 660–717.

[20] M. Talagrand *Majorizing measures: the generic chaining*, Ann. Probab. **24** (1996), 1049–1103.

[21] M. Talagrand, *Majorizing measures without measures*, Ann. Probab. **29** (2001), 411–417.

[22] M. Talagrand, *The generic chaining. Upper and lower bounds of stochastic processes*, Springer-Verlag, Berlin, 2005.

[23] M. Talagrand, *Are many small sets explicitly small?*, STOC'10–Proceedings of the 2010 ACM International Symposium on Theory of Computing, 13–35, ACM, New York, 2010

[24] M. Talagrand, *Chaining and the geometry of stochastic processes*, in: European Congress of Mathematics, Kraków 2–7 July 2012, 159–178, Eur. Math. Soc., Zürich 2013.

[25] M. Talagrand, *Upper and lower bounds for stochastic processes. Modern methods and classical problems*, Springer-Verlag, to appear.

[26] W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York, 1996.

Institute of Mathematics
University of Warsaw
Banacha 2
02-097 Warszawa
Poland
`wbednorz@mimuw.edu.pl, rlatala@mimuw.edu.pl`