C. R. Acad. Sci. Paris, Ser. I 351 (2013) 131–134

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability Theory

On the suprema of Bernoulli processes ☆

Witold Bednorz, Rafał Latała

Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland

ARTICLE INFO

Article history:
Received 25 February 2013
Accepted 27 February 2013
Available online 21 March 2013

Presented by Michel Talagrand

ABSTRACT

In this note we announce the affirmative solution of the so-called Bernoulli Conjecture concerning the characterization of the sample boundedness of Bernoulli processes. We also present some applications and discuss related open problems.

RÉSUMÉ

Dans cette note, nous annonçons la solution positive de la «conjecture de Bernoulli» concernant la caractérisation des processus de Bernoulli bornés. Nous en présentons des applications et discutons de questions ouvertes qui lui sont liées.

1. Introduction and main result

One of the fundamental issues of the probability theory is the study of suprema of stochastic processes. In particular, in many situations one needs to estimate the quantity $\mathbb{E}\sup_{t\in T} X_t$, where $(X_t)_{t\in T}$ is a stochastic process. (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, present already in the works of Kolmogorov and successfully developed over the last 40 years (see the monographs [13] and [15]).

The most important case of centered Gaussian processes is well understood. In this case, the boundedness of the process is related to the geometry of the metric space (T,d), where $d(t,s) := (\mathbb{E}(X_t - X_s)^2)^{1/2}$. In 1967, R. Dudley [3] obtained an upper bound for $\mathbb{E}\sup_{t\in T} X_t$ in terms of entropy numbers and in 1975 X. Fernique [4] improved Dudley's bound using so-called majorizing measures. In the seminal paper [10], M. Talagrand showed that Fernique's bound may be reversed and that for centered Gaussian processes (X_t),

$$\frac{1}{L}\gamma_2(T,d) \leqslant \mathbb{E}\sup_{t\in T} X_t \leqslant L\gamma_2(T,d),$$

where here and in the sequel L denotes a universal constant (which value may differ at each occurrence). There are many equivalent definitions of the functional γ_2 [12], for example one may set:

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

where the infimum runs over all sequences T_n of subsets of T such that $|T_0| = 1$ and $|T_n| \leqslant 2^{2^n}$.

☆ Research supported by the NCN grant DEC-2012/05/B/ST1/00412.
E-mail addresses: wbednorz@mimuw.edu.pl (W. Bednorz), rlatala@mimuw.edu.pl (R. Latała).

1631-073X/$ – see front matter
http://dx.doi.org/10.1016/j.crma.2013.02.013

Another fundamental class of processes is based on the Bernoulli sequence, i.e. the sequence $(\varepsilon_i)_{i\geqslant 1}$ of i.i.d. symmetric r.v.'s taking values ± 1. For $t \in l_2$, the series $X_t := \sum_{i\geqslant 1} t_i\varepsilon_i$ converges a.s. and for $T \subset l_2$, 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 one is a consequence of the uniform bound $|X_t| \leqslant \|t\|_1 = \sum_{i\geqslant 1}|t_i|$, so that $b(T) \leqslant \sup_{t\in T}\|t\|_1$. Another is based on the domination by the canonical Gaussian process $G_t := \sum_{i\geqslant 1} t_i g_i$, where g_i are i.i.d. $\mathcal{N}(0,1)$ r.v.'s. Indeed, assuming the independence of (g_i) and (ε_i), Jensen's inequality implies:

$$g(T) := \mathbb{E}\sup_{t\in T}\sum_{i\geqslant 1} t_i g_i = \mathbb{E}\sup_{t\in T}\sum_{i\geqslant 1} t_i\varepsilon_i|g_i| \geqslant \mathbb{E}\sup_{t\in T}\sum_{i\geqslant 1} t_i\varepsilon_i\mathbb{E}|g_i| = \sqrt{\frac{2}{\pi}}b(T).$$

Obviously also if $T \subset T_1 + T_2 = \{t^1 + t^2\colon\ t^i \in T_i\}$ then $b(T) \leqslant b(T_1) + b(T_2)$, hence:

$$b(T) \leqslant \inf\left\{\sup_{t\in T_1}\|t\|_1 + \sqrt{\frac{\pi}{2}}g(T_2)\colon\ T \subset T_1 + T_2\right\} \leqslant \inf\left\{\sup_{t\in T_1}\|t\|_1 + L\gamma_2(T_2)\colon\ T \subset T_1 + T_2\right\},$$

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 [9] or Chapter 4 in [13]). The next theorem provides an affirmative answer.

Theorem 1. *For any nonempty set $T \subset l_2$ with $b(T) < \infty$, one may find a decomposition $T \subset T_1 + T_2$ with $\sup_{t\in T_1}\sum_{i\geqslant 1}|t_i| \leqslant Lb(T)$ and $g(T_2) \leqslant Lb(T)$.*

As a corollary we obtain another useful criteria of boundedness for Bernoulli processes. For a random variable X and $p > 0$ we set $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$.

Corollary 2. *Suppose that $(X_t)_{t\in T}$ is a Bernoulli process with $b(T) < \infty$. Then there exist $t^1, t^2, \ldots \in l_2$ such that $T \subset \overline{\mathrm{conv}}\{t^n\colon\ n \geqslant 1\}$ and $\|X_{t^n}\|_{\log(n+2)} \leqslant Lb(T)$ for all $n \geqslant 1$.*

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

$$\mathbb{P}\Big(\sup_{t\in T} X_t \geqslant uM\Big) \leqslant \mathbb{P}\Big(\sup_{n\geqslant 1} X_{t^n} \geqslant uM\Big) \leqslant \sum_{n\geqslant 1}\mathbb{P}\big(X_{t_n} \geqslant u\|X_{t^n}\|_{\log(n+2)}\big) \leqslant \sum_{n\geqslant 1} u^{-\log(n+2)}$$

and integration by parts easily yields $\mathbb{E}\sup_{t\in T} X_t \leqslant LM$.

2. Ideas of the proof

In this section we try to briefly describe a few ideas behind the proof of Theorem 1, presented in details in [2]. The proof is not very long, but is quite technical. It builds on many ideas developed over the years by Michel Talagrand. As in the case of Gaussian processes it uses both concentration and minorization properties.

The main difficulty lies in the fact that there is no direct way of producing decomposition $t = t^1 + t^2$ for $t \in T$ such that $\sup_{t\in T}\|t^1\|_1 \leqslant Lb(T)$ and $\gamma_2(\{t^2\colon\ t \in T\}) \leqslant Lb(T)$. Following Talagrand, we connect decompositions of the set T with suitable sequences of its partitions in the next theorem. We recall that an increasing sequence $(\mathcal{A}_n)_{n\geqslant 0}$ of partitions of T is called *admissible* if $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \leqslant 2^{2^n}$. For such partitions and $t \in T$, we denote by $A_n(t)$ a unique set in $\mathcal{A}_n$ which contains t.

Theorem 3. *Suppose that $M > 0$, $r \geqslant 2$, $(\mathcal{A}_n)_{n\geqslant 0}$ is an admissible sequence of partitions of T, 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) $\|t - s\|_2 \leqslant Mr^{-j_0(T)}$ *for* $t, s \in T$,

ii) *if* $n \geqslant 1$, $\mathcal{A}_n \ni A \subset A' \in \mathcal{A}_{n-1}$ *then either* $j_n(A) = j_{n-1}(A)$ *and* $\pi_n(A) = \pi_{n-1}(A)$ *or* $j_n(A) > j_{n-1}(A')$, $\pi_n(A) \in A'$ *and*

$$\sum_{i\in I_n(A)}\min\left\{\big(t_i - \pi_n(A)_i\big)^2, r^{-2j_n(A)}\right\} \leqslant M2^n r^{-2j_n(A)} \quad \textit{for all } t \in A,$$

where for any $t \in A$,

$$I_n(A) = I_n(t) := \left\{i \geqslant 1\colon\ \left|\pi_{k+1}\big(A_{k+1}(t)\big)_i - \pi_k\big(A_k(t)\big)_i\right| \leqslant r^{-j_k(A_k(t))} \textit{ for } 0 \leqslant k \leqslant n-1\right\}.$$

Then there exist sets $T_1, T_2 \subset l_2$ *such that* $T \subset T_1 + T_2$ *and*

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

The main novelty here is the introduction of sets $I_n(A)$. To build such a partition, we use another idea of Talagrand and construct functionals on nonempty subsets of T that satisfy certain growth conditions. Our functionals depend not only on two integer-valued parameters, but also on the subset. Their construction combines Talagrand's "chopping maps" and ideas from [7]. The key ingredient to show the growth condition is the following modification of Proposition 1 in [7], which is based on concentration and minorization properties of Bernoulli processes.

Proposition 4. *Let $J\subset\mathbb{N}^*$, an integer $m\geqslant 2$, $\sigma>0$ and $T\subset l_2$ be such that*:

$$\Big(\sum_{i\in J}(t_i-s_i)^2\Big)^{1/2}\leqslant\frac{1}{L}\sigma \quad \text{and} \quad \|t-s\|_\infty\leqslant\frac{\sigma}{\sqrt{\log m}} \quad \text{for all } t,s\in T.$$

Then there exist $t^1,\dots,t^m\in T$ such that either $T\subset\bigcup_{l\leqslant m}B(t^l,\sigma)$ or the set $S:=T\setminus\bigcup_{l\leqslant m}B(t^l,\sigma)$ satisfies:

$$b_J(S):=\mathbb{E}\sup_{t\in S}\sum_{i\in J}t_i\varepsilon_i\leqslant b(T)-\frac{1}{L}\sigma\sqrt{\log m}.$$

The crucial point here is that we make no assumption about the diameter of the set T with respect to the l_2 distance.

3. Selected applications

In this section, we present two consequences of Theorem 1. The first shows that if a Bernoulli vector Y weakly dominates random vector X then Y strongly dominates X (cf. [8]).

Corollary 5. *Let X,Y be random vectors in a separable Banach space $(F,\|\cdot\|)$ such that $Y=\sum_{i\geqslant 1}u_i\varepsilon_i$ for some vectors $u_i\in F$ and*

$$\mathbb{P}\big(|\varphi(X)|\geqslant u\big)\leqslant\mathbb{P}\big(|\varphi(Y)|\geqslant u\big) \quad \text{for all } \varphi\in F^*,\ u>0.$$

Then there exists universal constant L such that:

$$\mathbb{P}\big(\|X\|\geqslant u\big)\leqslant L\mathbb{P}\big(\|Y\|\geqslant u/L\big) \quad \text{for all } u>0.$$

Another result is a Levy–Ottaviani type maximal inequality for VC-classes (see [6] 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 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 6. *Let $(X_i)_{i\in I}$ be independent random variables in a separable Banach space $(F,\|\cdot\|)$ 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*

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

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

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

4. Further questions

The following generalization of the Bernoulli Conjecture was formulated by S. Kwapień (private communication).

Problem 7. Let $(F,\|\cdot\|)$ be a normed space and (u_i) be a sequence of vectors in F such that the series $\sum_{i\geqslant 1}u_i\varepsilon_i$ converges a.s. Does there exist a universal constant L and a decomposition $u_i=v_i+w_i$ such that

$$\mathbb{E}\Big\|\sum_{i\geqslant 1}v_ig_i\Big\|\leqslant L\mathbb{E}\Big\|\sum_{i\geqslant 1}u_i\varepsilon_i\Big\| \quad \text{and} \quad \sup_{\eta_i=\pm 1}\Big\|\sum_{i\geqslant 1}w_i\eta_i\Big\|\leqslant L\mathbb{E}\Big\|\sum_{i\geqslant 1}u_i\varepsilon_i\Big\|?$$

Theorem 1 states that the answer is positive for $F = l_\infty$.

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 [1]; however, two-sided estimates are only known in very few cases. For "canonical processes" of the form $X_t = \sum_{i\geqslant 1} t_i X_i$, where X_i are independent, centered r.v.'s results in the spirit of Corollary 2 were obtained for certain symmetric variables with log-concave tails [11,5]. A basic important class of canonical processes worth investigation is based on two-point-valued r.v.'s. The following conjecture was stated by M. Talagrand ($d_p(t,s) = \|t-s\|_p$ denotes the l_p-distance).

Conjecture 8. *Let $0 < \delta \leqslant 1/2$, δ_i be independent random variables such that $\mathbb{P}(\delta_i = 1) = \delta = 1 - \mathbb{P}(\delta_i = 0)$ and $\delta(T) := \mathbb{E}\sup_{t\in T} |\sum_{i\geqslant 1} t_i(\delta_i - \delta)|$ for $T \subset l_2$. Then, for any set T with $\delta(T) < \infty$, one may find a decomposition $T \subset T_1 + T_2$ such that:*

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

It may be shown that for $\delta = 1/2$ the above conjecture follows from Theorem 1. Moreover, Bernstein's inequality and a generic chaining argument imply that $\delta(T) \leqslant L(\sqrt{\delta}\gamma_2(T, d_2) + \gamma_1(T, d_\infty))$ (see [14]), so Conjecture 8 would result in a two-sided bound on $\delta(T)$.

References

[1] W. Bednorz, A theorem on majorizing measures, Ann. Probab. 34 (2006) 1771–1781.
[2] W. Bednorz, R. Latała, On the boundedness of Bernoulli processes, preprint.
[3] R.M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, J. Funct. Anal. 1 (1967) 290–330.
[4] X. Fernique, Regularité des trajectoires des fonctions aléatoires gaussiennes, in: École d'été de probabilités de Saint-Flour, IV-1974, in: Lectures Notes in Math., vol. 480, Springer, Berlin, 1975, pp. 1–96.
[5] R. Latała, Sudakov minoration principle and supremum of some processes, Geom. Funct. Anal. 7 (1997) 936–953.
[6] R. Latała, A note on the maximal inequalities for VC classes, in: Advances in Stochastic Inequalities, Atlanta, GA, 1997, in: Contemp. Math., vol. 234, Amer. Math. Soc., Providence, RI, 1999, pp. 125–134.
[7] R. Latała, On the boundedness of Bernoulli processes over thin sets, Electron. Commun. Probab. 13 (2008) 175–186.
[8] R. Latała, On weak tail domination of random vectors, Bull. Pol. Acad. Sci. Math. 57 (2009) 75–80.
[9] M. Ledoux, M. Talagrand, Probability in Banach Spaces. Isoperimetry and Processes, Springer-Verlag, Berlin, 1991.
[10] M. Talagrand, Regularity of Gaussian processes, Acta Math. 159 (1987) 99–149.
[11] M. Talagrand, The supremum of some canonical processes, Amer. J. Math. 116 (1994) 284–325.
[12] M. Talagrand, Majorizing measures without measures, Ann. Probab. 29 (2001) 411–417.
[13] M. Talagrand, The Generic Chaining. Upper and Lower Bounds of Stochastic Processes, Springer-Verlag, Berlin, 2005.
[14] M. Talagrand, Chaining and the geometry of stochastic processes, in: Proceedings of the 6th European Congress of Mathematics, in press.
[15] M. Talagrand, Upper and Lower Bounds for Stochastic Processes, Modern Methods and Classical Problems, Ergebnisse der Mathematik, Springer-Verlag, in press.