On the boundedness of Bernoulli processes

Rafał Latała (based on the joint work with Witold Bednorz)

University of Warsaw

Berkeley, October 1 2013

Introduction

One of the fundamental issues of probability theory is the study of suprema of stochastic processes. Besides various practical motivations it is closely related to such theoretical problems as

- regularity of sample paths of stochastic processes,
- convergence of othogonal and random series,
- estimates of norms of random vectors and random matrices,
- limit theorems for random vectors and empirical processes,
- combinatorial matching theorems.

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.

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.

To avoid measurability problems one may either assume that T is countable or define

$$\mathbb{E}\sup_{t\in T} X_t := \sup_{F} \mathbb{E}\sup_{t\in F} X_t,$$

Gaussian Processes

Let $(G_t)_{t\in T}$ be a centered Gaussian process and

$$g(T) := \mathbb{E}\sup_{t\in T} G_t.$$

In this case the boundedness of the process (which by the concentration properties of Gaussian processes is equivalent to the condition $g(T) < \infty$) is related to the geometry of the metric space (T, d), where $d(t,s) := (\mathbb{E}(G_t - G_s)^2)^{1/2}$.

R.Dudley'67 derived an upper bound for $g(T)$ in terms of entropy numbers:

$$g(T) \leq L \int_0^\infty \log^{1/2} N(T,d,r)dr,$$

where L is a universal constant and $N(T,d,r)$ is the minimal number of balls with radius r that cover T.
Dudley's bound may be reversed for stationary processes, but not in general.

Majorizing Measure Theorem

Two-sided bound for $g(T)$ was found by X.Fernique'74 (upper bound) and M.Talagrand'87 (lower bound) in terms of *majorizing measures*. Let

$$\gamma_2(T) := \inf \sup_{t \in T} \int_0^\infty \log^{1/2} \left(\frac{1}{\mu(B(t,x))} \right) dx,$$

where the infimum is taken over all probability measures on T and $B(t, \varepsilon)$ denotes the closed ball in (T, d) with radius x, centered at t.

Theorem (Fernique-Talagrand)

There exists a universal constant $L < \infty$ such that for all centered Gaussian processes,

$$\frac{1}{L}\gamma_2(T) \le g(T) \le L\gamma_2(T).$$

In particular $g(T) < \infty$ if and only if $\gamma_2(T) < \infty$.

Majorizing measures without measures

In general finding a majorizing measure in a concrete situation is a highly nontrivial task. Talagrand proposed a more combinatorial approach to this problem.

An increasing sequence $(\mathcal{A}_n)_{n\geq 0}$ of partitions of the set T is called *admissible* if $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \leq N_n := 2^{2^n}$. Talagrand showed that the quantity $\gamma_2(T)$ is equivalent to

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

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.

So to bound a supremum of some Gaussian process one may either construct a measure or build a partition.

Bernoulli Processes

Any separable Gaussian process has a canonical Karhunen-Loève type representation $(\sum_{i\geq 1} t_i g_i)_{t\in T}$, 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.

Let $(\varepsilon_i)_{i\geq 1}$ be a Bernoulli sequence, i.e. a sequence of i.i.d. symmetric r.v.'s taking values ± 1. For any $t \in \ell^2$ the series $\sum_{i\geq 1} t_i\varepsilon_i$ converges a.s. and we may define a *Bernoulli process*

$$\Big(\sum_{i\geq 1} t_i\varepsilon_i\Big)_{t\in T}, \quad T \subset \ell^2.$$

It is natural to ask when the Bernoulli process is a.s. bounded or, equivalently, when

$$b(T) := \mathbb{E}\sup_{t\in T}\sum_{i\geq 1} t_i\varepsilon_i < \infty?$$

Two easy upper bounds

The first bound follows by the uniform bound $|\sum_i t_i\varepsilon_i| \le \sum_i |t_i| = \|t\|_1$, so that

$$b(T) \le \sup_{t\in T} \|t\|_1.$$

Another bound is based on the domination by the canonical Gaussian process $G_t := \sum_i t_i g_i$. Assuming independence of (g_i) and (ε_i), Jensen's inequality implies

$$\begin{aligned} g(T) = \mathbb{E}\sup_{t\in T}\sum_i t_i g_i &= \mathbb{E}\sup_{t\in T}\sum_i t_i\varepsilon_i|g_i| \\ &\ge \mathbb{E}\sup_{t\in T}\sum_i t_i\varepsilon_i\mathbb{E}|g_i| = \sqrt{\frac{2}{\pi}}\,b(T). \end{aligned}$$

Bernoulli Conjecture

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

$$b(T) \le \inf\Big\{ \sup_{t\in T_1} \|t\|_1 + \sqrt{\frac{\pi}{2}} g(T_2) : \ T \subset T_1 + T_2 \Big\}$$
$$\le \inf\Big\{ \sup_{t\in T_1} \|t\|_1 + L\gamma_2(T_2) : \ T \subset T_1 + T_2 \Big\}.$$

It was open for about 25 years (under the name of Bernoulli conjecture) whether the above estimate may be reversed. Next theorem provides a positive solution.

Theorem (Bednorz, L.'13+)

For any set $T \subset \ell^2$ with $b(T) < \infty$ we may find a decomposition $T \subset T_1 + T_2$ with $\sup_{t\in T_1} \sum_{i\ge 1} |t_i| \le Lb(T)$ and $g(T_2) \le Lb(T)$.

Kwapień's Conjecture

Problem

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

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

The positive solution of the Bernoulli Conjecture shows that the answer is positive for $F = \ell^\infty$, in general one 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, 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.

Another bound for Bernoulli processes

For a random variable X and $p > 0$ we set $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$.

Corollary

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

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

$$\leq \sum_{n\geq 1} u^{-\log(n+2)}$$

and integration by parts easily yields for any $t_0 \in T$,

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

Proof of the Gaussian case

Proof of the lower bound for $g(T)$ is based on two facts - Gaussian concentration and Sudakov minoration.

Theorem (Gaussian concentration)

Let $(G_t)_{t\in T}$ be a centered Gaussian process, then for any $u > 0$,

$$\mathbb{P}\Big(\Big|\sup_{t\in T} G_t - g(T)\Big| \geq u\Big) \leq 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big),$$

where $\sigma^2 := \sup_{t\in T}\mathbb{E}|G_t|^2$.

Theorem (Sudakov minoration)

Suppose that $t_1, \ldots, t_m$ are such that $\|G_{t_l} - G_{t_{l'}}\|_2 \geq a$ for all $l \neq l'$. Then

$$\mathbb{E}\sup_{l\leq m} G_{t_l} \geq \frac{1}{L} a\sqrt{\log m}.$$

Concentration and minoration for Bernoulli processes

Theorem (Talagrand'88)

Let $T \subset \ell^2$ be such that $b(T) < \infty$, then for any $u > 0$,

$$\mathbb{P}\Big(\Big|\sup_{t\in T}\sum_{i\geq 1} t_i\varepsilon_i - b(T)\Big| \geq u\Big) \leq L\exp\Big(-\frac{u^2}{L\sigma^2}\Big),$$

where $\sigma^2 := \sup_{t\in T}\|t\|_2^2$.

Theorem (Talagrand'93)

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

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

Then

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

About the proof of BC

It is however not an easy task to combine in a right way concentration and minoration properties of Bernoulli processes. Our proof builds on many ideas developed over the years by Michel Talagrand.

One of difficulties lies in the fact that there is no direct way of producing a decomposition $t = t^1 + t^2$ for $t \in T$ such that $\sup_{t \in T} \|t^1\|_1 \leq Lb(T)$ and $\gamma_2(\{t^2 : \ t \in T\}) \leq Lb(T)$. Following Talagrand we connect decompositions of the set T with suitable sequences of its admissible partitions $(\mathcal{A}_n)_{n \geq 0}$.

To each set $A \in \mathcal{A}_n$ we will assign an integer $j_n(A)$ and a point $\pi_n(A) \in T$. The main novelty in the next statement is the introduction of sets $I_n(A)$.

Partitions and Bernoulli decomposition

Theorem

Suppose that $(\mathcal{A}_n)_{n\geq 0}$ *is an admissible sequence of partitions s.t.*
i) $\|t-s\|_2 \leq \sqrt{M} r^{-j_0(T)}$ *for* $t,s\in T$,
ii) if $n\geq 1$, $\mathcal{A}_n \ni A \subset A' \in \mathcal{A}_{n-1}$ *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*

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

where for any $t\in A$,

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

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

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

Partition construction

To build such partition we use another idea of Talagrand and construct functionals on nonempty subsets of T and related distances that posess a Talagrand-type decomposition property, which roughly says that each index set may be decomposed into a pieces that either have small diameter (with respect to a suitable distance) or a small value of a suitable functional on subsets of small diameter.

Our functionals and distances depend not only on two integer-valued parameters, but also on the subset. Their construction combines Talagrand's "chopping maps" that "add new Bernoulli r.v's to the process" and other types of maps that "remove Bernoulli r.v's from the process".

Key proposition

The key ingredient to show the decomposition property is the following proposition, which is based on concentration and minorization properties of Bernoulli processes.

Proposition

Let $J \subset \mathbb{N}^*$, an integer $m \geq 2$, $\sigma > 0$ and $T \subset \ell^2$ be such that

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

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

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

The crucial point here is that we make no assumption about the diameter of the set T with respect to the ℓ^2 distance.

Fernique's question

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 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.

Fernique's question

An affirmative answer to Fernique's question follows from the Bernoulli Conjecture.

Theorem

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 $\chi_i(e) = 1$ we have

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

Sketch of the proof

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

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

Let $M := \mathbb{E} \sup_{t \in T} \left| \sum_{i=1}^n t_i \varepsilon_i \right|$. BC 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

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

We may also assume that $T_2 \subset T - T_1$.
Obviously

$$\mathbb{E} \sup_{h \in G, t \in T} \Big| \sum_{i=1}^n t_i \varepsilon_i \chi_i(h) \Big|$$
$$\leq \mathbb{E} \sup_{h \in G, t^1 \in T_1} \Big| \sum_{i=1}^n t_i^1 \varepsilon_i \chi_i(h) \Big| + \mathbb{E} \sup_{h \in G, t^2 \in T_2} \Big| \sum_{i=1}^n t_i^2 \varepsilon_i \chi_i(h) \Big|. \tag{3}$$

Skecth of the proof ctd

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

$$\mathbb{E} \sup_{h\in G, t^1\in T_1} \Big|\sum_{i=1}^n t_i^1 \varepsilon_i \chi_i(h)\Big| \le \sup_{t\in T^1} \|t^1\|_1 \le LM. \tag{4}$$

Jensen' inequality and Fernique's theorem imply

$$\mathbb{E} \sup_{h\in G, t^2\in T_2} \Big|\sum_{i=1}^n t_i^2 \varepsilon_i \chi_i(h)\Big| \le \sqrt{\frac{\pi}{2}}\mathbb{E} \sup_{h\in G, t^2\in T_2} \Big|\sum_{i=1}^n t_i^2 g_i \chi_i(h)\Big|$$
$$\le L\Big(\mathbb{E} \sup_{t^2\in T_2} \Big|\sum_{i=1}^n t_i^2 g_i\Big| + \sup_{t^2\in T_2} \mathbb{E} \sup_{h\in G} \Big|\sum_{i=1}^n t_i^2 g_i \chi_i(h)\Big|\Big). \tag{5}$$

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

$$\mathbb{E} \sup_{h\in G} \Big|\sum_{i=1}^n t_i^2 g_i \chi_i(h)\Big| \le L\mathbb{E} \sup_{h\in G} \Big|\sum_{i=1}^n t_i^2 \varepsilon_i \chi_i(h)\Big|. \tag{6}$$

It is not hard to deduce estimate (1) from (2)-(6).

Lévy-Ottaviani maximal inequalities

The Lévy inequality states that for any a.s. convergent series $\sum_{i=1}^{\infty} X_i$ of independent symmetric r.v's with values in some Banach space and $u > 0$ we have

$$\mathbb{P}\Big(\max_n \Big\|\sum_{i=1}^{n} X_i\Big\| \geq u\Big) \leq 2\mathbb{P}\Big(\Big\|\sum_{i=1}^{\infty} X_i\Big\| \geq u\Big).$$

The generalization of the Lévy inequality to a nonsymmetric case is called the Lévy-Ottaviani inequality. It states that for any a.s. convergent series $\sum_{i=1}^{\infty} X_i$ of independent Banach-space valued r.v's and $u > 0$,

$$\mathbb{P}\Big(\max_n \Big\|\sum_{i=1}^{n} X_i\Big\| \geq u\Big) \leq 3 \max_n \mathbb{P}\Big(\Big\|\sum_{i=1}^{n} X_i\Big\| \geq u/3\Big).$$

Both Lévy and Lévy-Ottaviani inequalities have numerous applications. Roughly speaking they enable to reduce an almost sure statement to a statement in probability (like for example in the Itô-Nisio theorem).

General maximal inequalities

However sometimes one has to consider more complicated sets of indices and ways of converging of sums of random variables. This motivates the following definition.

Definition

Let $\mathcal{C}$ be a class of subsets of I and F be a separable Banach space. We say that $\mathcal{C}$ satisfies the maximal inequality in F *if there exist a constant $K < \infty$ such that for any sequence (X_i) of independent r.v.'s with values in F satisfying $\#\{i : X_i \neq 0\} < \infty$ a.s. we have*

$$\forall_{u>0}\ \mathbb{P}\Big(\max_{C\in\mathcal{C}}\Big\|\sum_{i\in C}X_i\Big\| \geq u) \leq K \max_{C\in\mathcal{C}\cup\{I\}}\mathbb{P}\Big(\Big\|\sum_{i\in C}X_i\Big\| \geq u/K\Big).$$

If this is true for any separable Banach space F we say that $\mathcal{C}$ satisfies the maximal inequality*.*

Question. Which classes $\mathcal{C}$ satisfy the maximal inequality?

Maximal inequalities and VC classes

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 $|\{C \cap A\colon\ C \in \mathcal{C}\}| < 2^{d+1}$.

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 class $\mathcal{C}$ that satisfies maximal inequality (even in the scalar case) needs to be a VC-class.

The converse statement follows from the Bernoulli Conjecture.

Maximal inequalities for VC classes

Theorem

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 for $u > 0$,

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

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

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

Empirical processes

Solving the Bernoulli Conjecture is just the first step towards a much more ambitious program of finding two-sided bounds for the suprema of empirical processes.

Let $(X_i)_{i\leq N}$ 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 to relate the quantity

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

with the geometry of the class $\mathcal{F}$.

Let (X_i') be an independent copy of (X_i) then Jensen's inequality and a pointwise estimate imply

$$S_N(\mathcal{F}) \leq \mathbb{E}\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{N}}\Big|\sum_{i\leq N}(f(X_i)-f(X_i'))\Big| \leq 2\mathbb{E}\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{N}}\Big|\sum_{i\leq N}f(X_i)\Big|$$

$$\leq 2\mathbb{E}\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{N}}\sum_{i\leq N}|f(X_i)|$$

Chaining bound

The second bound for $S_N(\mathcal{F})$ is based on chaining and Bernstein's inequality

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

where $\|f\|_p$ denotes the L_p norm of $f(X_i)$. Chaining argument shows that

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

where $d_p(f,g):=\|f-g\|_p$. Here 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 $(\mathcal{A}_n)_{n\ge 0}$ of the set T.

Conjecture for suprema of empirical processes

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

Conjecture (Talagrand)

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

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

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

In other words "chaining explains the part of boundedness due to cancellation".

Selector processes

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

$$X_t = \sum_{i \geq 1} t_i(\delta_i - \delta), \quad t \in \ell^2,$$

where $(\delta_i)_{i \geq 1}$ are independent random variables such that $\mathbb{P}(\delta_i = 1) = \delta = 1 - \mathbb{P}(\delta_i = 0)$.
As for empirical processes we may bound the quantity

$$\delta(T) := \mathbb{E} \sup_{t \in T} \Big| \sum_{i \geq 1} t_i(\delta_i - \delta) \Big|, \quad T \subset \ell^2$$

in two ways.
First bound for $\delta(T)$ follows by a pointwise estimate ($(\delta_i')_i$ denotes an independent copy of $(\delta_i)_i$)

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

Selector processes - chaining bound

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 for $s, t \in \ell^2$,

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

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

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

Conjecture for suprema of selector processes

The next conjecture 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 (Talagrand)

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*

$$\mathbb{E}\sup_{t\in T_1}\sum_{i\geq 1}|t_i|\delta_i \leq L\delta(T),$$

$$\sqrt{\delta}\gamma_2(T_2, d_2) \leq L\delta(T) \quad \textit{and} \quad \gamma_1(T_2, d_\infty) \leq L\delta(T).$$

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

Thank you for your attention!