1. Introduction

For a stochastic process $(X_t : t \in T)$, we define

$$\mathbb{E}\sup_{t\in T} X_t := \sup\big(\mathbb{E}\sup_{t\in F} X_t : F \subset T, F \text{ finite}\big)$$

The relevant object to study to bound a Gaussian process $(X_t : t \in T)$ is the metric space (T, d) where

$$d(s,t) := \big(\mathbb{E}(X_s - X_t)^2\big)^{1/2}.$$

A very important fact is that for any $u > 0$,

$$\mathbb{P}\big[|X_s - X_t| \geq u d(s,t)\big] \leq \exp(-u^2/2).$$

Bounding processes defined on abstract spaces T is in most of the case a succession of combination of concentration's inequality with the so called union-bound. What is heard by *union bound* is the simple fact that for any familly of events $(A_i)_i$, the probability measure of the union is smaller than the sum of all the probability:

$$\mathbb{P}\Big[\cup_i A_i\Big] \leq \sum_i \mathbb{P}[A_i].$$

Applying this union bound to familly of events related to sequences of partitions of a metric space is the heart of this section. Undertsanding how to construct convenient sequences of partitions of the metric space (d, T) is the core of the proof of upper and lower bounds on the quantity $\mathbb{E}\max_{t\in T} X_t$.

The heart of this section is to understand the trade-off between the concentration properties of the increments of the process $(X_t : t \in T)$ and the complexity of the size t (measured w.r.t. the canonical distance (or sequence of distance (cf. the sequence of interpolated norms associated with Rademacher processes, i.e. Theorem**??** and Theorem**??**)).

2. ϵ-net argument of Pisier

We will present this argument through the following problem: let $X, X_1, \dots, X_n$ be $n+1$ i.i.d. random vectors of $\mathbb{R}^d$. We assume that

(1) *X is isotrope:* i.e. $\forall u \in \mathbb{R}^d, \mathbb{E}\langle u, X\rangle^2 = \|u\|_2$;
(2) *X is ψ_2 w.r.t. $\|\cdot\|_2$:* i.e. for all $u \in \mathbb{R}^d$ and any $t > 0$,

$$\mathbb{P}[|\langle u, X\rangle| > t\|u\|_2] \leq 2\exp(-ct^2);$$

(3) X is mean zero.

We want to study the *sampling problem* in this setup: given is $0 < \kappa, \eta < 1$, we want to know what is the minimal sample size n needed to have, with probability greater than $1 - \kappa$,

$$\sup_{u\in\mathcal{S}^{d-1}} \Big|\frac{1}{n}\sum_{i=1}^n \langle u, X_i\rangle^2 - 1\Big| \leq \eta,$$

where $\mathcal{S}^{d-1}$ is the unit euclidean ball of $\mathbb{R}^d$.

The solution to this problem is $n \geq C(\kappa, \eta)d$, where $C(\kappa, \eta)$ is a constant depending only on κ and η. This means that the sample size has to be of the order of the dimension d.

Now, we turn to a proof of this fact using the ϵ-net argument.

Let $0 < \epsilon < 1/2$ and N_ϵ be a maximal ϵ-net of $\mathcal{S}^{d-1}$ w.r.t. the euclidean metric. This means that N_ϵ is a subset of $\mathcal{S}^{d-1}$ of maximal cardinality such that every elements $x \neq y$ in N_ϵ are ϵ-far w.r.t. the $\|\cdot\|_2$-norm.

Let $u \in N_\epsilon$. We want a concentration inequality for $|n^{-1}\sum\langle u, X_i\rangle^2 - 1|$. We know that $\langle u, X\rangle$ is a ψ_2 random variable thus, $\langle u, X\rangle^2$ is a ψ_1 random variable such that $\Big\|\langle u, X\rangle^2\Big\|_{\psi_1} =$

$\|\langle u, X\rangle\|_{\psi_2}^2 \leq c\|u\|_2^2$. Now, we use Bernstein's inequality for ψ_1 random variables (cf. Theorem **??**) to get, for every $t > 0$, with probability greater than $1 - 2\exp(-cnt^2 \wedge t)$

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\langle u, X_i\rangle^2 - 1\Big| \leq ct\|u\|_2^2 = ct. \tag{0.1}$$

We use an union bound to obtain the last result uniformly over the finite set N_ϵ: for every $t > 0$, with probability greater than $1 - 2|N_\epsilon|\exp(-cnt^2 \wedge t)$,

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\langle u, X_i\rangle^2 - 1\Big| \leq ct, \forall u \in N_\epsilon. \tag{0.2}$$

Since $|N_\epsilon| \leq (c/\epsilon)^{d-1}$ (cf.[1]), Equation (0.2) holds with probability at least $1 - 2\exp\big((d-1)\log(c/\epsilon) - cnt^2 \wedge t\big)$.

Now, we want to obtain the result of (0.2) uniformly over $\mathcal{S}^{d-1}$. For this task we use the ϵ-*net argument*: let $u \in \mathcal{S}^{d-1}$; we want to write u as

$$u = \sum_{i=0}^{\infty}\delta_i u_i \text{ s.t. } u_i \in N_\epsilon \text{ and } \delta_0 = 1, |\delta_i| \leq \epsilon^i, \forall i \geq 1. \tag{0.3}$$

There exists $u_0 \in N_\epsilon$ such that $\|u - u_0\|_2 \leq \epsilon$. If $u = u_0$ the claim follows otherwise

$$\frac{u - u_0}{\|u - u_0\|_2} \in \mathcal{S}^{d-1}.$$

Thus, there exists $u_1 \in \mathcal{S}^{d-1}$ such that

$$\Big\|\frac{u - u_0}{\|u - u_0\|_2} - u_1\Big\|_2 \leq \epsilon.$$

We continue this argument to obtain (0.3).

Now, we consider the random matrice

$$\Gamma := \frac{1}{\sqrt{n}}\begin{pmatrix} X_1^t \\ \vdots \\ X_n^t \end{pmatrix}.$$

Let $u \in \mathcal{S}^{d-1}$ and consider its decomposition (0.3). We have

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\langle u, X_i\rangle^2\Big)^{1/2} = \|\Gamma u\|_2 \leq \sum_i |\delta_i|\,\|\Gamma u_i\|_2 \leq \frac{1}{1-\epsilon}\max\big(\|\Gamma u\|_2 : u \in N_\epsilon\big) \tag{0.4}$$

and, similarly,

$$\|\Gamma u\|_2 \geq \|\Gamma u_0\|_2 - \sum_{i\geq 1}|\delta_i|\,\|\Gamma u_i\|_2 \geq \min\big(\|\Gamma u\|_2 : u \in N_\epsilon\big) - \frac{\epsilon}{1-\epsilon}\max\big(\|\Gamma u\|_2 : u \in N_\epsilon\big). \tag{0.5}$$

Using Equation (0.2), with probability greater than $1 - 2\exp\big((d-1)\log(c/\epsilon) - cnt^2 \wedge t\big)$,

$$\big(\sqrt{1-ct} - 2\epsilon\sqrt{1+ct}\big)^2 \leq \frac{1}{n}\sum_{i=1}^{n}\langle u, X_i\rangle^2 \leq (1+2\epsilon)^2(1+ct), \forall u \in \mathcal{S}^{d-1}. \tag{0.6}$$

Note that the lower estimate holds only when $ct < 1$ and that the case $ct \geq 1$ is trivial for the lower bound.

Let $0 < \kappa, \eta < 1$. We can choose ϵ and t depending only on κ and η such that Equation (0.6) implies

$$\sup_{u \in \mathcal{S}^{d-1}}\Big|\frac{1}{n}\sum_{i=1}^{n}\langle u, X_i\rangle^2 - 1\Big| \leq \eta.$$

To make the probability $1 - 2\exp\big((d-1)\log(c/\epsilon) - cnt^2 \wedge t\big)$ greater than $1-\eta$ we have to take n at least of the same order of d.

3. Chaining and the Koltchinskii-Dudley entropy's integral

In this Section, we present some upper bounds for the supremum $\sup_{t\in T} X_t$ obtained using the entropy integral. We start with the classical chaining argument under the subgaussian assumption on the increment of the process:

$$\mathbb{P}\Big[|X_s - X_t| > td(s,t)\Big] \le 2\exp\big(-ct^2\big), \forall s,t \in T \tag{0.7}$$

where d is a semi-metric on T.

Then, we follow the line of [3] to explore the case where the process $(X_t : t \in T)$ is such that the increment have a ψ behaviour w.r.t. some distance d:

$$\|X_s - X_t\|_\psi \le cd(s,t) \forall s,t \in T. \tag{0.8}$$

We first start by introducing the metric quantities and complexities of a semi-metric space which are at the heart of this approach.

Definition 0.1. *Let (T,d) be a semi-metric space and $\epsilon > 0$. The **ϵ-covering number** $N(\epsilon,T,d)$ of (T,d) is the minimal number of balls of radius ϵ needed to cover T. The **ϵ-packing number** $D(\epsilon,T,d)$ is the maximal number of ϵ-separated points in T. The **entropy numbers** are the logarithms of the covering and packing numbers respectively.*

Note that

$$N(\epsilon,T,d) \le D(\epsilon,T,d) \le N(\epsilon/2,T,d), \forall \epsilon > 0$$

and by definition a semi-metric space (T,d) is *totally bounded* when covering and packing numbers are finite for every $\epsilon > 0$.

Finally, in all the following, we will need the following continuity assumption on the process $(X_t : t \in T)$ w.r.t. the semi-metric d:

(0.9)
almost surely, for every $t \in T$ for every sequences $(t_n : n \in \mathbb{N})$ of T such that

(0.10)
$d(t_n,t)$ tends to zero when n tends to infinity, the process $(X_{t_n} : n \in \mathbb{N})$ tends to X_t when n tends to i

Theorem 0.1. *There exists some absolute constants c_0, c_1, c_2, c_3 and c_4 such that the following holds. Let (T,d) be a semi-metric space. Let $(X_t : t \in T)$ be a stochastic process satisfying the continuity assumption (0.9) and the subgaussian condition (0.7).*

For every $v \ge c_0$, with probability greater than $1 - c_1\exp(-c_2 v^2)$

$$\sup_{s,t\in T} |X_t - X_s| \le c_3 v \int_0^\infty \log^{1/2} N(\epsilon,T,d)d\epsilon$$

and

$$\mathbb{E}\sup_{s,t\in T} |X_t - X_s| \le c_4 \int_0^\infty \log^{1/2} N(\epsilon,T,d).$$

Proof. We define, for every integer $i \ge 1$,

$$\epsilon_i := \inf\{\epsilon > 0 : N(\epsilon,T,d) \le 2^{2^i}\} \text{ and } \epsilon_0 := \mathrm{diam}(T,d).$$

Consider a sequence $(T_i : i \ge 0)$ of subsets of T such that, for every integer $i \ge 1$, T_i is a set of minimal cardinality satisfying that for every point $t \in T$ there exists $t_i \in T_i$ such that $d(t,t_i) \le \epsilon_i$. By definition, $|T_i| = N(\epsilon_i,T,d) \le 2.2^{2^i}$ and $|T_0| = 1$. For each point $t \in T$ and any level $i \in \mathbb{N}$, we denote by $\pi_i(t) \in T_i$ one of the points of T_i which are ϵ_i-close to t.

By the continuity assumption on the process $(X_t : t \in T)$, we have, almost surely, for every $t \in T$,

$$X_t - X_{\pi_0(t)} = \sum_{i=0}^{\infty} X_{\pi_{i+1}(t)} - X_{\pi_i(t)}. \tag{0.11}$$

Let be given a level $i \in \mathbb{N}$. Let $t \in T$. By using the subgaussian assumption (0.7), for every $u > 0$, with probability greater than $1 - 2\exp(-cu^2)$

$$|X_{\pi_{i+1}(t)} - X_{\pi_i(t)}| \leq u d(\pi_{i+1}(t), \pi_i(t)) \leq u(\epsilon_i + \epsilon_{i+1}) \leq 2u\epsilon_i \tag{0.12}$$

To get this result uniformly over all links $\{(\pi_{i+1}(t), \pi_i(t)), \forall t \in T\}$, we use an union bound: with probability greater than $1 - 2|T_{i+1}||T_i| \exp(-cu^2) \geq 1 - 2\exp\big(3.2^i \log 2 - cu^2\big)\big)$,

$$|X_{\pi_{i+1}(t)} - X_{\pi_i(t)}| \leq 2u\epsilon_i, \forall t \in T.$$

To make this result interesting the term "cu^2" has to defy the term "$3.2^i \log 2$" in the probability estimate. Thus, we apply the last result to $u := v2^{i/2}$ where v has to be larger than $\sqrt{6 \log 2/c}$. Finally, for the level i, we obtain with probability greater than $1 - 2\exp\big(-(c/2)v^2 2^i\big)$,

$$|X_{\pi_{i+1}(t)} - X_{\pi_i(t)}| \leq 2v2^{i/2}\epsilon_i, \forall t \in T,$$

for every v larger than an absolute constant.

We apply Equation (3) combined with an union bound on all the level $i \in \mathbb{N}$, to get, with probability greater than $1 - 2\sum_{i=0}^{\infty} \exp\big(-(c/2)v^2 2^i\big)$,

$$|X_t - X_{\pi_0(t)}| \leq 2v \sum_{i=0}^{\infty} 2^{i/2}\epsilon_i, \forall t \in T. \tag{0.13}$$

The sum in the probability estimate is geometric, so it is comparable to its first term. Thus, Equation0.13 holds with probability greater than $1 - c_0 \exp(-c_1 v^2)$.

The right hand term in Equation0.13 can be written as an integral: the Koltchinskii-Dudley entropy integral, in the following way. Let $i \in \mathbb{N}$, if $\epsilon < \epsilon_i$ then $N(\epsilon, T, d) > 2^{2^i}$ and so $N(\epsilon, T, d) \geq 2^{2^i} + 1$. So we have

$$\sqrt{\log(1 + 2^{2^i})}(\epsilon_i - \epsilon_{i+1}) \leq \int_{\epsilon_{i+1}}^{\epsilon_i} \sqrt{\log N(\epsilon, T, d)} d\epsilon.$$

Since $\log(1 + 2^{2^i}) \geq 2^i \log 2$, summing over all $i \geq 0$ yields

$$\sqrt{\log 2} \sum_{i=0}^{\infty} 2^{i/2}(\epsilon_i - \epsilon_{i+1}) \leq \int_0^{\epsilon_0} \sqrt{\log N(\epsilon, T, d)} d\epsilon$$

and

$$\sum_{i=0}^{\infty} 2^{i/2}(\epsilon_i - \epsilon_{i+1}) = \sum_{i=0}^{\infty} 2^{i/2}\epsilon_i - \sum_{i=1}^{\infty} 2^{(i-1)/2}\epsilon_i \geq \Big(1 - \frac{1}{\sqrt{2}}\Big) \sum_{i=0}^{\infty} 2^{i/2}\epsilon_i.$$

Finally, we obtain, for every v larger than an absolute constant, with probability greater than $1 - c_0 \exp(-c_1 v^2)$,

$$\sup_{t \in T} |X_t - X_{\pi_0(t)}| \leq c_2 v \int_0^{\infty} \log^{1/2} \big(N(\epsilon, T, d)\big) d\epsilon.$$

By a classical integration argument, we obtain

$$\mathbb{E} \sup_{t \in T} |X_t - X_{\pi_0(t)}| = \int_0^{\infty} \mathbb{P}\big[\sup_{t \in T} |X_t - X_{\pi_0(t)}| > u\big] du \leq c_4 \int_0^{\infty} \log^{1/2} \big(N(\epsilon, T, d)\big) d\epsilon.$$

To conclude, we use the fact that $|T_0| = 1$, thus, for every $t, s \in T$,

$$|X_t - X_s| \leq |X_t - X_{\pi_0(t)}| + |X_s - X_{\pi_0(s)}|.$$

■

The subgaussian assumption (0.7) can be written as $\|X_s - X_t\|_{\psi_2} \leq cd(s,t), \forall s,t \in T$. This assumption has been generalized to the one of Equation (0.8). The result is given in the following theorem.

Theorem 0.2. *Let ψ be a Young-Orlicz modulus such that there exists an absolute constant c satisfying*

$$\text{limsup}_{x,y\longmapsto\infty} \frac{\psi(x)\psi(y)}{\psi(xy)} \leq c.$$

Let $(X_t : t \in T)$ be a separable stochastic process with

$$\|X_s - X_t\|_\psi \leq Cd(s,t), \forall s,t \in T$$

for some semi-metric d on T and a constant C. Then, for any $\delta, \eta > 0$,

$$\Big\| \sup_{d(s,t)\leq\delta} |X_s - X_t| \Big\|_\psi \leq K\Big[\int_0^\eta \psi^{-1}\big(D(\epsilon,T,d)\big)d\epsilon + \delta\psi^{-1}\big(D^2(\eta,T,d)\big)\Big],$$

for a constant K depending only on ψ and C.

Proof. Without loss of generality we can assume that $D(\eta,T,d)$ and $\int_0^\eta \psi^{-1}\big(D(\epsilon,T,d)\big)d\epsilon$ are finite.

We construct a sequence $(T_i : i \geq 0)$ of nested sets $T_0 \subset T_1 \subset T_2 \subset \cdots \subset T$ such that T_j is a maximal set of $\eta 2^{-j}$-separated points in T. By definition, we have

$$|T_j| \leq D(\eta 2^{-j}, T, d). \tag{0.14}$$

We construct "links" between the elements of the sequence $(T_i : i \geq 0)$: for every point $t_{j+1} \in T_{j+1}$ we define a unique point $t_j \in T_j$ such that $d(t_{j+1}, t_j) \leq \eta 2^{-j}$. So that every point t_{j+1} is associated with a sequence, called a chain: $t_{j+1}, t_j, t_{j-1}, \ldots, t_0$.

Given a level $k \in \mathbb{N}$, we can control uniformly all the increments of the process at this level: let $s_{k+1}, t_{k+1} \in T_{k+1}$ and $s_0 \in T_0$ (respectively $t_0 \in T_0$) the corresponding beginning of the chain associated with s_{k+1} (respectively t_{k+1}). We have

$$\big|(X_{s_{k+1}} - X_{s_0}) - (X_{t_{k+1}} - X_{t_0})\big| \leq \Big|\sum_{j=0}^k (X_{s_{j+1}} - X_{s_j}) - \sum_{j=0}^k (X_{t_{j+1}} - X_{t_j})\Big|$$

$$\leq 2\sum_{j=0}^k \max\big(|X_u - X_v| : (u,v) \in T_{j+1} \times T_j, d(u,v) \leq \eta 2^{-j} \text{ and } (u,v) \text{ is a link}\big).$$

Now, we apply the maximal inequality of Proposition**??** to every level j to get

$$\begin{aligned}
&\big\|\max\big(|X_u - X_v| : (u,v) \in T_{j+1} \times T_j, d(u,v) \leq \eta 2^{-j} \text{ and } (u,v) \text{ is a link}\big)\big\|_\psi \\
&\leq K_0 \psi^{-1}(|T_{j+1}|) \max\big(\|X_u - X_v\|_\psi : (u,v) \in T_{j+1} \times T_j, d(u,v) \leq \eta 2^{-j}\big) \\
&\leq K_1 \psi^{-1}(D(\eta 2^{-j-1}, T, d))\eta 2^{-j}
\end{aligned}$$

and so

$$\big\|\max\big(\big|(X_s - X_{s_0}) - (X_t - X_{t_0})\big| : s,t \in T_{k+1}\big)\big\|_\psi \leq K_2 \int_0^\eta \psi^{-1}\big(D(\epsilon,T,d)\big)d\epsilon,$$

where in this bound, s_0 and t_0 are the endpoints of the chains starting at s and t respectively.

Using the triangle inequality, it remains to upper bound the increments $|X_{s_0} - X_{t_0}|$. For every pair of endpoints s_0, t_0 of chains starting at two points of T_{k+1} within distance δ of each other, choose exactly one pair s_{k+1}, t_{k+1} in T_{k+1} with $d(s_{k+1}, t_{k+1}) < \delta$, whose chains end at s_0, t_0. By definition of T_0, this gives at most $D^2(\eta, T, d)$ such pairs. By the triangle inequality,

$$|X_{s_0} - X_{t_0}| \leq |(X_{s_0} - X_{s_{k+1}}) - (X_{t_0} - X_{t_{k+1}})| + |X_{s_{k+1}} - X_{t_{k+1}}|.$$

where

$$\big\|(X_{s_0} - X_{s_{k+1}}) - (X_{t_0} - X_{t_{k+1}})\big\|_\psi \leq K_2 \int_0^\eta \psi^{-1}\big(D(\epsilon, T, d)\big) d\epsilon$$

and, applying Proposition**??**,

$$\big\|X_{s_{k+1}} - X_{t_{k+1}}\big\|_\psi \leq K\psi^{-1}(D^2(\eta, T, d))\delta.$$

Let k tends to infinity to conclude the proof. ■

Corollary 0.1. *There exists an absolute constant $c > 0$ such that the following holds.*

$$\Big\|\max_{s,t\in T} |X_s - X_t|\Big\|_\psi \leq c \int_0^{\mathrm{Diam}(T,d)} \psi^{-1}\big(D(\epsilon, T, d)\big) d\epsilon.$$

Proof. Apply Theorem 0.2 to $\eta = \delta = \mathrm{Diam}(T, d)$. ■

4. Generic chaining and the γ-functional of M.Talagrand

In this section, we present an improvement upon the chaining argument. This argument is called the *generic chaining* (cf. [2]). The Koltchinski-Dudley entropy integral is the natural metric complexity coming out of the chaining approach. For the generic chaining argument, the natural metric complexity measure is given by the γ-functional of M.Talagrand that we introduce now:

Definition 0.2. *Let (T, d) be a semi-metric space. We say that a sequence $(T_n : n \geq 0)$ of subsets of T is* ***admissible*** *when $|T_0| \leq 1$ and $|T_n| \leq 2^{2^n} \forall n \geq 1$.*

Let $\alpha > 0$. We define the ***γ-functional of Talagrand*** *by*

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

where the infimum is taken over all admissible sequences $(T_n)_n$ and $d(t, T_n) := \inf_{s\in T_n} d(t, s)$.

The functions γ_α are purely metric and are upper bounded by the Koltchinski-Dudley entropy integral:

$$\gamma_\alpha(T, d) \leq c \int_0^\infty \log^{1/\alpha}(\epsilon, T, d) d\epsilon.$$

Indeed, for every $n \in \mathbb{N}$, take T_n to be a minimal ϵ_n-net of T, w.r.t. d, where ϵ_n is defined by $N(\epsilon_n, T, d) \leq 2^{2^n}$. By minimality $(T_n : n \in \mathbb{N})$ is an admissible sequence and for every $n \in \mathbb{N}$, if $\epsilon < \epsilon_n$ then $N(\epsilon, T, d) > 2^{2^n}$ and so $N(\epsilon, T, d) \geq 2^{2^n} + 1$. So we have

$$\log^{1/\alpha}(1 + 2^{2^n})(\epsilon_n - \epsilon_{n+1}) \leq \int_{\epsilon_{n+1}}^{\epsilon_n} \log^{1/\alpha} N(\epsilon, T, d) d\epsilon.$$

Since $\log^{1/\alpha}(1 + 2^{2^n}) \geq 2^{n/\alpha} \log^{1/\alpha}(2)$, summing over all $n \geq 0$ yields

$$\log^{1/\alpha}(2) \sum_{n=0}^\infty 2^{n/\alpha}(\epsilon_n - \epsilon_{n+1}) \leq \int_0^{\epsilon_0} \log^{1/\alpha} N(\epsilon, T, d) d\epsilon$$

and

$$\sum_{n=0}^\infty 2^{n/\alpha}(\epsilon_n - \epsilon_{n+1}) = \sum_{n=0}^\infty 2^{n/\alpha}\epsilon_n - \sum_{n=0}^\infty 2^{(n-1)/\alpha}\epsilon_n \geq \Big(1 - \frac{1}{\sqrt{2}}\Big) \sum_{n=0}^\infty 2^{n/\alpha}\epsilon_n.$$

We conclude by using

$$\sup_{t\in T}\sum_{n\geq 0} 2^{n/\alpha} d(t,T_n) \leq \sum_{n\geq 0} 2^{n/\alpha}\sup_{t\in T} d(t,T_n) \leq \sum_{n\geq 0} 2^{n/\alpha}\epsilon_n. \tag{0.15}$$

Note that, in the first inequality of (0.15), the gap between the right and left sides can be very large.

Now, we turn to the upper bound of the supremum of empirical processes under the ψ_2 assumption of (0.7).

Theorem 0.3. *There exists some absolute constants c_0, c_1, c_2, c_3 and c_4 such that the following holds. Let (T,d) be a semi-metric space. Let $(X_t : t\in T)$ be a stochastic process satisfying the continuity assumption (0.9) and the subgaussian condition (0.7).*

For every $v\geq c_0$, with probability greater than $1-c_1\exp(-c_2 v^2)$

$$\sup_{s,t\in T}|X_t - X_s| \leq c_3 v\gamma_2(T,d)$$

and

$$\mathbb{E}\sup_{s,t\in T}|X_t - X_s| \leq c_4\gamma_2(T,d).$$

Proof. The proof follows the same lines as the proof of Theorem 0.3. We sketch here the proof. Let $(T_n : n\in\mathbb{N})$ be an admissible sequence. For every $t\in T$ and $n\in\mathbb{N}$ denote by $\pi_n(t)$ one of the closest element of T_n to t. The union bound and the subgaussian assumption yield the following probability bound: for every v greater than an absolute constant, with probability greater than $1-2\sum_{i=0}^\infty \exp(-c_1 v^2 2^i)$,

$$|X_t - X_{\pi_0(t)}| \leq c_3\sum_{n=0}^\infty 2^{n/2} d(\pi_n(t),\pi_{n+1}(t)), \forall t\in T.$$

The claim follows easily. ∎

Note that one can replace the subgaussian assumption (0.7) by a ψ_α assumption ($\alpha\geq 1$) on the increments of the process:

$$\mathbb{P}\Big[|X_s - X_t| > t d(s,t)\Big] \leq 2\exp\big(-ct^\alpha\big), \forall s,t\in T. \tag{0.16}$$

In this case, Theorem 0.3 is still true when replacing the complexity measure $\gamma_2(T,d)$ by the quantity $\gamma_\alpha(T,d)$.

4.1. Generic Chaining for processes with non-homogenous tail behaviour. It is usual to meet process having two different concentration behaviours. For instance, in Theorems **??**, **??** and **??**, the empirical mean $\bar{X}_n$ has a subgaussian behaviour for small concentration level (values of t in $(0,c]$, for some c depending only on the tail behaviour of X) and, in general, a ψ_α behaviour for large deviation (values of t larger than c). The subgaussian behaviour comes from the asymptotic behaviour of $\bar{X}_n$ given by the CLT (cf. Berry-Esseen theorem for a lower bound for small values of t). This subgaussian non-asymptotic behaviour of the mean is the "beginning" of the asymptotic normality of the mean. On the opposite, the ψ_α behaviour of $\bar{X}_n$ is related to the behaviour of a generic element X. Indeed, for a realisation $\omega\in\Omega$, most of the elements $X_i(\omega)$ are around the mean $\mathbb{E}X$ (providing the subgaussian concentration) whereas only few of them are far from the mean (providing the same behaviour of a single realisation $X_i(\omega)$ to $\bar{X}_n(\omega)$).

In this subsection, we study the maximum of processes with increments having the following concentration behaviour for some $\alpha>0$:

$$\mathbb{P}\big[|X_s - X_t|\geq u\big] \leq 2\exp\Big(-\Big(\frac{u^2}{d_2^2(s,t)}\Big)\wedge\Big(\frac{u^\alpha}{d_\alpha^\alpha(s,t)}\Big)\Big), \forall u>0, s,t\in T. \tag{0.17}$$

Theorem 0.4. *There exists absolute constants c_0 and c_1 such that the following holds. Let $\alpha > 0$ and T be a set endowed with two semi-metrics d_α and d_2. Consider a process $(X_t : t \in T)$ having the continuity property (0.9) w.r.t. d_2 and d_α and satisfying the concentration condition (0.17). Then, for every $u \geq c_0$, with probability greater than $1 - c_1 \exp(-c_2 u^2 \wedge u^\alpha)$,*

$$\sup_{t,s\in T} |X_t - X_s| \leq c_3 u\big(\gamma_2(T, d_2) + \gamma_\alpha(T, d_\alpha)\big)$$

and

$$\mathbb{E} \sup_{t,s\in T} |X_t - X_s| \leq c_4\big(\gamma_2(T, d_2) + \gamma_\alpha(T, d_\alpha)\big).$$

Proof. Take $(A_n : n \in \mathbb{N})$ and $(B_n : n \in \mathbb{N})$ be two admissible sequences of T satisfying

$$\sup_{t\in T} \sum_{n\geq 0} 2^{n/2} d_2(t, A_n) \leq 2\gamma_2(T, d_2) \text{ and } \sup_{t\in T} \sum_{n\geq 0} 2^{n/\alpha} d_\alpha(t, B_n) \leq 2\gamma_\alpha(T, d_\alpha).$$

We construct the admissible sequence $(T_n : n \in \mathbb{N})$ by setting

$$T_0 := \{t_0\} \text{ and } T_n := A_{n-1} \cup B_{n-1}, \forall n \geq 1,$$

where t_0 is one element of T. We also define $\pi_n(t)$ to be the closest point to $t \in T$ in T_n for each $n \in \mathbb{N}$ and $t \in T$.

Let $t \in T$ and $n \in \mathbb{N}$. Using the estimate on the concentration behaviour of the increments (cf. Equation (0.16)), we have for all $u > 0$, with probability greater than $1 - 2\exp(-2^n[u^2 \wedge u^\alpha])$,

$$|X_{\pi_{n+1}(t)} - X_{\pi_n(t)}| \leq u2^{n/\alpha} d_\alpha(\pi_{n+1}(t), \pi_n(t)) + u2^{n/2} d_2(\pi_{n+1}(t), \pi_n(t)). \tag{0.18}$$

Using an union bound, we extend the last inequality to all links $\{(\pi_{n+1}, \pi_n(t)) : t \in T\}$ and then to all level $n \in \mathbb{N}$. We have, for every $u \geq c_0$, with probability greater than $1 - c_1 \exp(-c_2 u^2 \wedge u^\alpha)$, for every $s, t \in T$,

$$\begin{aligned} |X_t - X_s| &\leq |X_t - X_{\pi_0(t)}| + |X_s - X_{\pi_0(s)}| \\ &\leq \sum_{n=0}^{\infty} |X_{\pi_{n+1}(t)} - X_{\pi_n(t)}| + \sum_{n=0}^{\infty} |X_{\pi_{n+1}(s)} - X_{\pi_n(s)}| \\ &\leq u \sup_{t\in T} \sum_{n\geq 0} \big(2^{n/\alpha} d_\alpha(\pi_{n+1}(t), \pi_n(t)) + 2^{n/2} d_2(\pi_{n+1}(t), \pi_n(t))\big). \end{aligned}$$

By definition of $(A_n : n \in \mathbb{N})$, we have, for each $n \geq 1$, $T_n \subset A_{n-1}$, so

$$d_2(\pi_n(t), \pi_{n+1}(t)) \leq d_2(\pi_n(t), t) + d_2(t, \pi_{n+1}(t)) \leq d_2(t, A_{n-1}) + d_2(t, A_n).$$

Moreover, it is easy to see that, for every semi-metric space (T', d') and $\eta > 0$

$$\gamma_\eta(T', d') \geq \inf_{t_0\in T'} \sup_{t\in T'} d(t, t_0) \geq (1/2)\mathrm{diam}(T', d'),$$

thus $d_2(\pi_1(t), t_0) \leq \mathrm{diam}(T, d) \leq 2\gamma_2(T, d_2)$. Then, proceeding similarly for d_α, we get, for every $u \geq c_0$, with probability greater than $1 - c_1 \exp(-c_2 u^2 \wedge u^\alpha)$,

$$\sup_{s,t\in T} |X_s - X_t| \leq 4u\big(\gamma_\alpha(T, d_\alpha) + \gamma_2(T, d_2)\big).$$

The upper bound on the expectation follows by a classical integration argument. ∎

4.2. Sum of square of ψ_2 functions. In this section, we give a particular look to upper bound the supremum

$$\sup_{f\in F}\left|\frac{1}{n}\sum_{i=1}^n f^2(X_i)-\mathbb{E}f^2(X)\right|, \tag{0.19}$$

where $X_1,\dots,X_n$ are n i.i.d. random variables with values in a measurable space $\mathcal{X}$ and F is a class of real-valued functions defined on $\mathcal{X}$. We assume that

$$\mathrm{diam}(F,\|\cdot\|_{\psi_2(\mu)}):=\alpha<\infty, \tag{0.20}$$

where μ is the probability distribution of $X\sim X_1$. In terms of random variables, Assumption (0.20) means that for all $f\in F$, $f(X)$ has a ψ_2 behaviour and its ψ_2 norm is uniformly bounded over F by α.

Theorem 0.5. *There exists absolute constants c_0, c_1 such that the following holds. Let $F\subset L_2(\mu)$ be star-shaped*

We introduce the following notation. For every function $f\in L_2(\mu)$, we set

$$Z(f):=\frac{1}{n}\sum_{i=1}^n f^2(X_i)-\mathbb{E}f^2(X) \text{ and } W(f):=\Big(\frac{1}{n}\sum_{i=1}^n f^2(X_i)\Big)^{1/2}. \tag{0.21}$$

Thanks to the star-shaped assumption, we can work as if all the elements $f\in F$ are such that $\mathbb{E}f^2(X)=1$. The general case can then be handled thanks to a *localisation argument.*

The first thing that one has to obtain when studying upper bounds for supremum of processes as in (0.19) is the concentration behaviour of increments of the process. Namely, we need concentration result for $Z(f)-Z(g)$ for $f,g\in F$. Since we will treat the end of the chain by using a trick, the deviation behaviour of the increments $W(f-g)$ will be of importance as well.

Lemma 0.1. *There exists an absolute constant c_1 such that the following holds. Let $F\subset\mathcal{S}(L_2(\mu))$ (the unit ball of $L_2(\mu)$). Denote $\alpha:=\mathrm{diam}(F,\psi_2)$. For every $f,g\in F$ we have:*

(1) for every $u\geq 1$,

$$\mathbb{P}\Big[W(f-g)\geq u\,\|f-g\|_{\psi_2}\Big]\leq 2\exp\big(-c_1nu^2\big);$$

(2) for every $u>0$,

$$\mathbb{P}\Big[|Z(f)-Z(g)|\geq u\alpha\,\|f-g\|_{\psi_2}\Big]\leq 2\exp\big(-c_1n(u\wedge u^2)\big);$$

and for every $u>0$,

$$\mathbb{P}\Big[|Z(f)|\geq u\alpha^2\Big]\leq 2\exp\big(-c_1n(u\wedge u^2)\big).$$

Proof. Let $f,g\in F$. Since $f,g\in L_{\psi_2}$, we have $\big\|(f-g)^2\big\|_{\psi_1}=\|f-g\|_{\psi_2}^2$. Then, we apply Bernstein's inequality for ψ_1 random variables (cf. Theorem**??**) to get, for every $t>0$, with probability greater than $1-2\exp(-c_1n(t\wedge t^2))$

$$\frac{1}{n}\sum_{i=1}^n (f-g)^2(X_i)-\mathbb{E}(f-g)^2\leq t\,\|f-g\|_{\psi_2}^2. \tag{0.22}$$

Using $\|f-g\|_{\psi_2}\geq\|f-g\|_2$ with Equation (0.22), it is easy to get for every $u\geq 2$,

$$\mathbb{P}\Big[W(f-g)\geq u\,\|f-g\|_{\psi_2}\Big]\leq\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n (f-g)^2(X_i)-\mathbb{E}(f-g)^2\geq(u^2-1)\,\|f-g\|_{\psi_2}^2\Big]$$

$$\leq 2\exp\big(-c_1 n u^2\big).$$

To prove the end of the claim, we use that $\mathbb{E}f^2 = \mathbb{E}g^2$, $|f^2 - g^2| \leq 4(f-g)^2$ so $\left\|f^2 - g^2\right\|_{\psi_1} \leq 4\left\|f-g\right\|_{\psi_2}^2$ ■

After dealing with the concentration properties of the increments of the process, we want to obtain a uniform upper bound. For that we are going to use the generic chaining argument. But, since we work in a very special framework (sum of square of ψ_1 r.v.), we will perform a particular chaining argument which will allow us to avoid the $\gamma_1(F,\psi_2)$ in the upper bound. Indeed, according to Theorem 0.4 and the deviation inequality on the increments of $(Z(f): f \in F)$ of Lemma 0.1, we can obtain an upper bound for the process in (0.19) proportional to $\gamma_2(F,\psi_2) + \gamma_1(F,\psi_2)$.

Consider an almost admissible sequence $(F_n : n \in \mathbb{N})$ of F. That is an admissible sequence such that

$$\gamma_2(F,\psi_2) \leq 2 \sup_{f\in F} \sum_{n=0}^{\infty} 2^{n/2} d_{\psi_2}(f, F_n).$$

If $\gamma_2(F,\psi_2) = \infty$ then the upper bound of Theorem 0.5 is trivial, otherwise for every $f \in F$ the sum $\sum_{n=0}^{\infty} 2^{n/2} d_{\psi_2}(f, F_n)$ converges. In particular, $d_{\psi_2}(f, \pi_n(f))$ tends to zero when n tends to infinity. In what follows, we will assume the non trivial case that $\gamma_2(F,\psi_2)$ is finite. In particular $\mathcal{D} := \cup_{n\in\mathbb{N}} F_n$ is a countable dense (for the ψ_2 norm) subset of F. Take $f \in \mathcal{D}$, there exists $\Omega_f \subset \Omega$ a measurable set of measure 1 such that $\forall \omega \in \Omega_f, \forall i = 1, \ldots, n$, $\pi_s(f(X_i(\omega)))$ tends to $f(X_i(\omega))$ when s tends to infinity. Thus, by continuity of the euclidean norm in $\mathbb{R}^n$, $W(\pi_s(f))$ tends to $W(f)$ on Ω_f. Since $\cap_{f\in F}\Omega_f$ is a set of probability measure 1, almost surely $\forall f/inF$, $W(\pi_s(f) : s \in \mathbb{N})$ converges to $W(f)$. By separability this result holds uniformly over F. The same claim follows for Z.

The idea of the proof is, for a given $f \in F$, to treat the links of the chain $(\pi_n(f) : n \in \mathbb{N})$ in three different region depending on the concentration property that we expect:

(1) $f - \pi_{s_0}(f)$: where we work with the process $W(f - \pi_{s_0}(f))$ which is subgaussian (thanks to this trick we can avoid the ψ_1 behaviour of the process $Z(f)$ and thus the term $\gamma_1(F,\psi_1)$);
(2) $\pi_{s_0-1}(f) - \pi_{s_1}(f)$: where we work with process $Z(\pi_{s_0-1}(f)) - Z(\pi_{s_1}(f))$ which is subgaussian in this range;
(3) $\pi_{s_1-1}(f) - \pi_0(f)$: where the complexity is so small that an upper bound is trivial.

Proposition 0.1 (End of the chain)**.** *There exists an absolute constant for which the following holds. Let $F \subset \mathcal{S}(L_2(\mu))$ and $\alpha := \mathrm{diam}(F,\psi_2)$. With probability greater than $1-\exp(-n)$,*

$$\sup_{f\in F} W(f - \pi_{s_0}(f)) \leq \frac{c\gamma_2(F,\psi_2)}{\sqrt{n}},$$

where s_0 is such that $2^{s_0} \sim n$.

Proof. Let f be in F. Since $(\pi_s(f) : s \in \mathbb{N})$ tends to f in $L_{\psi_2}(\mu)$, we have in $L_{\psi_2}(\mu)$,

$$f - \pi_{s_0}(f) = \sum_{s=s_0}^{\infty} \pi_{s+1}(f) - \pi_s(f).$$

On the other hand, W is sub-linear, thus, by the using the uniform continuity of W over F almost surely,

$$W(f - \pi_{s_0}(f)) \leq \sum_{s\geq s_0} W(\pi_{s+1}(f) - \pi_s(f)).$$

Now, fix a level $s \geq s_0$. Using an union bound on the set of links $\{(\pi_{s+1}(f), \pi_s(f)) : f \in F\}$ and the sub-gaussian property of W (i.e. Lemma 0.1), we get, for every $u \geq 1$, with probability greater than $1 - 2\exp(-cnu^2)$,

$$W(\pi_{s+1}(f) - \pi_s(f)) \leq u \left\| \pi_{s+1}(f) - \pi_s(f) \right\|_{\psi_2}$$

■

4.3. Truncation argument.

5. Exercises

Exercise 0.1 (largest singular value of RM with ψ_2, isotrope and independent rows)**.** *Let K be a symmetric convex body of $\mathbb{R}^n$ in an isotropic position. Let $X, X_1, \ldots, X_N$ be independent and uniformly distributed in K random variables. We assume that K is such that X is subgaussian (i.e. $\exists C_0 : \forall t \in \mathbb{R}^n, \left\| \langle X, t \rangle \right\|_{\psi_2} \leq C_0$). Then, the largest singular value of the operator*

$$T := \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} : \mathbb{R}^n \to \mathbb{R}^N$$

denoted by $\|T\|$ satisfies, for every $x > 0$, with probability greater than $1 - 2\exp(-x)$

$$\Big| 1 - \|T\| \Big| \leq \sqrt{\Big(1 + \frac{x}{cn}\Big)\frac{n}{N}}.$$

Bibliography

[1] Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[2] Michel Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005. Upper and lower bounds of stochastic processes.

[3] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.