EXPONENTIAL INEQUALITIES FOR SUPREMA OF PROCESSES WITH STOCHASTIC NORMALIZATION

Julien Aubert, Luc Lehéricy

Exponential inequalities for suprema of processes with stochastic normalization

Julien Aubert
Université Côte d'Azur, CNRS, LJAD, France
julien.aubert@univ-cotedazur.fr

Luc Lehéricy
Université Côte d'Azur, CNRS, LJAD, France
luc.lehericy@univ-cotedazur.fr

Abstract

We prove a deviation bound for the supremum of a normalized process derived from a real-valued random process $(Y_t)_{t\in S}$. Assuming the increments $Y_t - Y_s$ satisfy Bernstein deviation bounds with random fluctuation terms makes it possible to use a stochastic normalization term. This improves on classical versions where the control of the fluctuation is deterministic.

Keywords Talagrand's inequality · supremum of a random process · martingales · peeling argument · chaining argument

1 Introduction

Consider the following result, proved by Baraud in [1]. Let $(X_t)_{t\in S}$ be a family of centered real-valued random variables indexed by a countable set S embedded in a vector space E of dimension D that satisfies the following Bernstein bound: there exists two (deterministic) distances d and δ on E and a nonnegative constant c such that for all $s, t \in S$, for all $\lambda \in \big[0, (c\delta(s,t))^{-1}\big)$,

$$\mathbb{E}\left[\exp\left(\lambda(X_t - X_s) - \frac{\lambda^2 d^2(s,t)}{2(1-\lambda c\delta(s,t))}\right)\right] \leqslant 1.$$

If moreover, d and δ are norms bounded on S respectively by $v > 0$ and b/c for some finite v, b, then for any $t_0 \in S$ and $x \geqslant 0$,

$$\mathbb{P}\left(\sup_{t\in S}(X_t - X_{t_0}) \geqslant 18\left(\sqrt{v^2(x+D)} + b(x+D)\right)\right) \leqslant e^{-x}. \tag{1}$$

From the two equations above, we can deduce an upper bound on the renormalized process, of the form: for any $t_0 \in S$ and $\sigma, x > 0$, letting $\Psi(\sigma, x) = 18(\sigma\sqrt{x+D} + b(x+D))$,

$$\mathbb{P}\left(\sup_{t\in S}\frac{X_t - X_{t_0}}{d(t,t_0)^2 + \sigma^2} \geqslant 4\sigma^{-2}\Psi(\sigma, x)\right) \leqslant 2(\log(v/\sigma) \vee 0 + 1)e^{-x}. \tag{2}$$

(2) improves on (1) since σ can be chosen much smaller than the upper bound v on S. In particular, it allows to control the fluctuation of $X_t - X_{t_0}$ jointly for all $t \in S$ by a term comparable to the variance of $X_t - X_{t_0}$, namely $d^2(t, t_0)$.

Inequalities such as (2) are key to obtain oracle inequalities in model selection and empirical risk minimization (see e.g., [1] , [2] and [3]), where d is the error on the parameter and X_t is the difference between the empirical risk and the true risk for the parameter t. In the existing literature, this distance d is deterministic. However, it is sometimes possible to control the fluctuations of $X_t - X_s$ with smaller, albeit random, distances d and δ. Being able to obtain results such as (2) with these random terms could therefore provide significant improvements.

To our knowledge, there exists no similar result when the distances $d(s,t)$ and $\delta(s,t)$ are replaced with random quantities $R_2(s,t)$ and $R_\infty(s,t)$. Working with random fluctuation bounds is a natural idea that arises in many situations (e.g., [4, Theorem 3]). For instance, [5] and [6] study the process $Y_t(\lambda) = \exp(\lambda A_t - \lambda^2 B_t^2/2)$, for random variables $B_t > 0$ and A_t, to develop inequalities for the moments of A_t/B_t or $\sup_{t\geqslant 0} A_t/(B_t(\log\log B_t)^{1/2})$ when

$\mathbb{E}[Y_t(\lambda)] \leqslant 1$ and to obtain deviation inequalities for self-normalized martingales (see [7] for a comprehensive review). [8] improve the constants in (1) for processes of the form $\{s^T X, s \in T\}$ where X is a random vector in $\mathbb{R}^n$ with independent components satisfying a Bernstein type condition.

The goal of this note is to obtain a result similar to (2) but for random quantities instead of the deterministic distances d and δ.

2 Main result

We say that a nonnegative function $R : S \times S \to \mathbb{R}_+$ satisfies the triangle inequality if for all $(s, t, u) \in S^3$, $R(s, u) \leqslant R(s, t) + R(t, u)$. A semi-norm over a vector space E is a function $N : E \to \mathbb{R}_+$ such that $N(x + y) \leqslant N(x) + N(y)$ for all $x, y \in E$ and $N(\lambda x) = |\lambda| N(x)$ for all $x \in E$ and $\lambda \in \mathbb{R}$. For any two real numbers x, y, we write $x \wedge y$ the minimum of x and y, and $x \vee y$ their maximum.

The typical setting for applying our result is when a Bernstein type inequality with control of the fluctuations is available, as in the following proposition.

Proposition 1. *Let A be an event, Z be a random variable and R_2, R_∞ be nonnegative random variables such that*

$$\forall \lambda \geqslant 0 \qquad \mathbb{E}\left[\exp\left(\lambda Z - \frac{(\lambda R_2)^2}{2} \sum_{k \geqslant 0} (\lambda R_\infty)^k\right) \mathbf{1}_A\right] \leqslant 1. \tag{3}$$

Then, for all $\sigma, \sigma', x \geqslant 0$,

$$\mathbb{P}\left(\left\{Z \geqslant \sigma\sqrt{2x} + \sigma' x,\ R_2 \leqslant \sigma \text{ and } R_\infty \leqslant \sigma'\right\} \cap A\right) \leqslant e^{-x}. \tag{4}$$

Proof. By (3), for all $\sigma, \sigma', \lambda \geqslant 0$,

$$\mathbb{E}\left[\exp\left(\lambda Z - \frac{(\lambda R_2)^2}{2} \sum_{k \geqslant 0} (\lambda R_\infty)^k\right) \mathbf{1}_{R_2 \leqslant \sigma} \mathbf{1}_{R_\infty \leqslant \sigma'} \mathbf{1}_A\right] \leqslant 1,$$

and therefore for all $\lambda \in [0, (\sigma')^{-1})$,

$$\mathbb{E}[\exp(\lambda Z) \mathbf{1}_{R_2 \leqslant \sigma} \mathbf{1}_{R_\infty \leqslant \sigma'} \mathbf{1}_A] \leqslant \exp\left(\frac{\lambda^2 \sigma^2}{2(1 - \lambda\sigma')}\right),$$

after which the usual proof of Bernstein's inequality using the Chernoff bound produces the desired result. □

For martingales, (3) can be checked as follows, as a direct consequence of [9, Lemma 3.3], in which case R_2^2 is an upper bound of their quadratic variation.

Proposition 2. *Let $(M_n)_{n \geqslant 0}$ be a $(\mathcal{F}_n)_{n \geqslant 0}$-martingale with $M_0 = 0$. Let $n \geqslant 1$ and assume that there exists nonnegative random variables R_2 and R_∞ such that for all $\ell \geqslant 2$,*

$$\sum_{i=1}^{n} \mathbb{E}\left[(M_i - M_{i-1})^\ell \,\middle|\, \mathcal{F}_{i-1}\right] \leqslant \frac{\ell!}{2} R_2^2 R_\infty^{\ell-2}.$$

Then (3) *holds for $Z = M_n$ and any event A.*

A more classical way of expressing the latter for martingales (see [7], [9]) is the following corollary.

Corollary 3. *Let $n \geqslant 1$. Let $(M_i)_{0 \leqslant i \leqslant n}$ be a $(\mathcal{F}_i)_{0 \leqslant i \leqslant n}$-martingale with $M_0 = 0$, and let $(\langle M \rangle_i)_{0 \leqslant i \leqslant n}$ be its predictable quadratic variation. For all $i \in \{0, \dots, n-1\}$, let B_i be a $\mathcal{F}_i$-measurable random variable, and let $A = \{\forall i \in \{0, \dots, n-1\}, |M_{i+1} - M_i| \leqslant B_i\}$.*

Then (3) *holds for $Z = M_n$, $R_2 = \langle M \rangle_n$, $R_\infty = \max_{0 \leqslant i \leqslant n-1} B_i$, and this event A.*

Proof of Proposition 2. Let $C_0^\ell = 0$ and $C_i^\ell = \sum_{i=1}^n \mathbb{E}[(M_i - M_{i-1})^\ell \,|\, \mathcal{F}_{i-1}]$ for all $n \geqslant i \geqslant 1$ and $\ell \geqslant 2$. Lemma 3.3 of [9] gives that for all $\lambda > 0$, the sequence $(\mathcal{E}_i)_{i \geqslant 0}$ defined by

$$\mathcal{E}_i = \exp\left(\lambda M_i - \sum_{\ell \geqslant 2} \frac{\lambda^\ell}{\ell!} C_i^\ell\right)$$

is a supermartingale. In particular, $\mathbb{E}(\mathcal{E}_n) \leqslant \mathbb{E}(\mathcal{E}_0) = 1$. The result follows immediately. □

Example 4. Let $n \geqslant 1$. Let $X_1, \dots, X_n$ be centered independent real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that there exists $m > 0$ such that for all $j \in \{1, \dots, n\}$, for all $\ell \geqslant 2$, $\mathbb{E}[X_j^\ell] \leqslant \ell! m^\ell$. For instance, this is satisfied as soon as $|X_j|$ is stochastically dominated by an exponential random variable with parameter $1/m$.

Let $h_1 = 1$, for all $i \in \{2, \dots, n\}$, let $h_i : x \in \mathbb{R}^{i-1} \mapsto h_i(x) \in \mathbb{R}$ be a Borel measurable function, and let $h = (h_i)_{1 \leqslant i \leqslant n}$. Consider the process $(M_i(h))_{0 \leqslant i \leqslant n}$ defined by $M_0(h) = 0$ and for all $i \in \{1, \dots, n\}$,

$$M_i(h) = \sum_{j=1}^{i} h_j(X_1, \dots, X_{j-1}) X_j. \tag{5}$$

For any $i \in \{1, \dots, n\}$, let $\mathcal{F}_i$ be the sigma algebra generated by $\{X_1, \dots, X_i\}$, and let $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then $(M_i(h))_{0 \leqslant i \leqslant n}$ is an $(\mathcal{F}_i)_{0 \leqslant i \leqslant n}$-martingale. Moreover, Proposition 2 holds for $M_n(h)$ with

$$R_\infty = m \sup_{1 \leqslant i \leqslant n} |h_i(X_1, \dots, X_{i-1})| \quad \text{and} \quad R_2 = m \left(2 \sum_{i=1}^{n} h_i(X_1, \dots, X_{i-1})^2 \right)^{1/2}.$$

Our main result is a uniform control over the deviation of a renormalization of any process whose increments satisfy a Bernstein concentration inequality with random fluctuation bounds, as long as these bounds are upper bounded by deterministic seminorms. It can in particular be applied to any kind of supremum of martingales. The supremum does not have to be taken in time as is usually the case for martingales, but can possibly be taken in other parameters when considering families of martingales as shown in Example 7.

Theorem 5. *Let $(Y_t)_{t \in S}$ be a family of real-valued random variables, indexed by a countable subset S of a linear space E of finite dimension D. Let R_2 and R_∞ be two nonnegative random functions on $S \times S$ that satisfy the triangle inequality. Assume that there exists an event A such that for all $s, u \in S$, the increments $Y_s - Y_u$ satisfy* (4). *Assume that there exists two deterministic seminorms N_2 and N_∞ on E and nonnegative constants v, w, such that on the event A, for all $s, u \in S$,*

$$\begin{cases} R_2(s,u) \leqslant N_2(s-u) \leqslant v, \\ R_\infty(s,u) \leqslant N_\infty(s-u) \leqslant w. \end{cases} \tag{6}$$

Let $c \geqslant 0$. Define for all $x, \sigma \geqslant 0$,

$$\Psi(\sigma, x) = 20 \left[(\sigma \wedge v) \sqrt{x + D \log\left(\frac{v \vee (cw)}{\sigma} \vee e \right)} + (\frac{\sigma}{c} \wedge w) \left(x + D \log\left(\frac{v \vee (cw)}{\sigma} \vee e \right) \right) \right],$$

with the convention $\frac{\sigma}{0} = +\infty$. Finally, define for all $t, t_0 \in S$, $\Delta_c(t, t_0) = R_2(t, t_0) \vee (c R_\infty(t, t_0))$. Then, for any $t_0 \in S$, $x \geqslant 0$ and $\sigma > 0$,

$$\mathbb{P}\left(\left\{ \sup_{t \in S} \frac{Y_t - Y_{t_0}}{\Delta_c(t,t_0)^2 + \sigma^2} \geqslant 4\sigma^{-2} \Psi(\sigma, x) \right\} \cap A \right) \leqslant 2 \left(\log\left(\frac{v \vee (cw)}{\sigma} \right) \vee 0 + 1 \right) e^{-x}. \tag{7}$$

A typical application of this result is the following corollary:

Corollary 6. *Let α be a positive real number. Under the same assumptions and notations as in Theorem 5, for all $x > 0$, with probability at least $1 - 2(\log((v \vee (cw))/\sqrt{\alpha D}) \vee 0 + 1) e^{-x}$,*

$$\forall t \in S, \qquad Y_t - Y_{t_0} \leqslant 80(\sqrt{\alpha} + w) \left[\alpha^{-1} \Delta_c(t, t_0)^2 + x + D \log\left(\frac{v \vee (cw)}{\sqrt{\alpha D}} \vee e \right) \right].$$

Proof. The result is obtained by taking $\sigma^2 = \alpha \left(x + D \log\left(\frac{v \vee (cw)}{\sigma} \vee e \right) \right)$ in (7). □

Example 7. Take the assumptions and notations of Example 4. Let J be a subset of a linear space E of finite dimension D. Consider a class of vectors of measurable functions $\mathcal{H} = \{h_u : u \in J\}$ such that there exists $u_0 \in J$ satisfying $h_{u_0} = 0 \in \mathcal{H}$ and for each $u \in J$, $h_u = (h_{u,1}, \dots, h_{u,n})$ with $h_{u,i} : \mathbb{R}^{i-1} \to \mathbb{R}$. Assume that there exists $L > 0$ and $M > 0$ for some norm $\|\cdot\|$ on E such that for all $u, v \in J$, $\sup_{1 \leqslant i \leqslant n} \|h_{u,i} - h_{v,i}\|_\infty \leqslant L\|u - v\| \leqslant LM$. Replace h in (5) by the differences $h_u - h_v$, and take

$$\begin{cases} R_\infty(u,v) = m \sup_{1 \leqslant i \leqslant n} |h_{u,i}(X_1, \dots, X_{i-1}) - h_{v,i}(X_1, \dots, X_{i-1})| \leqslant m \sup_{1 \leqslant i \leqslant n} \|h_{u,i} - h_{v,i}\|_\infty \\ \frac{R_2(u,v)}{\sqrt{2}} = m \left(\sum_{i=1}^{n} (h_{u,i}(X_1, \dots, X_{i-1}) - h_{v,i}(X_1, \dots, X_{i-1}))^2 \right)^{1/2} \leqslant \sqrt{n} m \sup_{1 \leqslant i \leqslant n} \|h_{u,i} - h_{v,i}\|_\infty. \end{cases}$$

Since,

$$\begin{cases} R_\infty(u,v) \leqslant mL\|u-v\|, \\ \frac{R_2(u,v)}{\sqrt{2}} \leqslant \sqrt{n}mL\|u-v\|, \end{cases}$$

take,

$$\begin{cases} N_\infty(u-v) = mL\|u-v\| \leqslant mLM, \\ \frac{N_2(u-v)}{\sqrt{2}} = \sqrt{n}mL\|u-v\| \leqslant \sqrt{n}mLM, \end{cases}$$

and Theorem 5 with $c=0$, $\sigma = m\sqrt{2n}LM$, $t_0 = u_0$ leads to, for all $x>0$, with probability at least $1-2e^{-x}$,

$$\sup_{u\in J} \frac{M_n(h_u)}{R_2(u,u_0)^2 + 2(mLM)^2 n} \leqslant 57(mLM)^{-1}\left[\sqrt{\frac{x+D}{n}} + \frac{x+D}{n}\right].$$

3 Proof

The first step is to control the deviation of the supremum of the process over the (random) balls of R_2 and R_∞ using a similar technique to [1, Theorems 2.1 and 5.1], before using a peeling argument similar to [2, Lemma 4.23] in a second phase.

In all that follows, we omit the intersection with the event A in order to lighten the notations. Every event considered should be understood as its intersection with A.

Proposition 8. *Let $(Y_t)_{t\in S}$ be a family of real-valued random variables, indexed by a countable subset S of a linear space E of finite dimension D. Let R_2 and R_∞ be two nonnegative random functions on $S\times S$ that almost surely satisfy the triangle inequality. Assume that, for all $s,u\in S$, the increments $Y_s - Y_u$ satisfy* (4) *with respect to R_2 and R_∞. Assume that there exists two deterministic seminorms N_2 and N_∞ on E and nonnegative constants v,w satisfying* (6)*. Fix $t_0\in S$. For $\sigma,\sigma'\geqslant 0$, let*

$$\mathcal{B}(\sigma,\sigma') = \{s\in S \;:\; R_2(s,t_0)\leqslant \sigma \text{ and } R_\infty(s,t_0)\leqslant\sigma'\}\,.$$

Then, there exists a numerical constant $\kappa>0$ (for instance $\kappa=20$) such that for all $x\geqslant 0$ and $\sigma,\sigma'>0$,

$$\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0}) \geqslant \kappa\left[(\sigma\wedge v)\sqrt{x + D\log\left(\frac{v}{\sigma}\vee\frac{w}{\sigma'}\vee e\right)} + (\sigma'\wedge w)\left(x + D\log\left(\frac{v}{\sigma}\vee\frac{w}{\sigma'}\vee e\right)\right)\right]\right) \leqslant e^{-x}.$$

Comments.

- Proposition 8 itself is not enough to conclude. Taking $\sigma\geqslant 0$ and $\sigma' = \sigma/c$ with $c>0$ will result in a bound on the supremum of $(Y_t - Y_{t_0})_{t\in\mathcal{B}(\sigma,\sigma')}$, but not on the renormalized process on all the set S.
- Taking $\sigma = v$ and $\sigma' = w$ recovers the original result of [1, Theoreom 2.1], but it is again insufficient to link it to $\Delta_c(t,t_0)$.
- To conclude, a peeling argument is necessary, hence, this is why Proposition (8) is proved for any ball $\mathcal{B}(\sigma,\sigma')$. By controlling locally the variations of the supremum, Lemma 11 allows to derive global maximal inequalities for the renormalized process of interest.

Proof of Proposition 8. The proof follows exactly the steps of the proof of [1] for Theorem 2.1 and Theorem 5.1, only differing in the family of partitions we consider. First, note that it is sufficient to show the result for a finite set S. Indeed, for any sequence $(S_n)_{n\in\mathbb{N}}$ of finite sets such that $S=\cup_n S_n$ and such that for all $n\in\mathbb{N}$, $S_n\subset S_{n+1}$, one has for all $u\geqslant 0$,

$$\left\{\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0})\geqslant u\right\} = \bigcup_{t\in\mathcal{B}(\sigma,\sigma')\cap S}\{Y_t - Y_{t_0}\geqslant u\} = \bigcup_{n\in\mathbb{N}}\bigcup_{t\in\mathcal{B}(\sigma,\sigma')\cap S_n}\{Y_t - Y_{t_0}\geqslant u\}\,.$$

Since the sequence of sets $\left(\bigcup_{t\in\mathcal{B}(\sigma,\sigma')\cap S_n}\{Y_t - Y_{t_0}\geqslant u\}\right)_{n\in\mathbb{N}}$ is non-decreasing, by upward monotone convergence, for all $u\geqslant 0$,

$$\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0})\geqslant u\right) = \lim_{n\to+\infty}\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')\cap S_n}(Y_t - Y_{t_0})\geqslant u\right).$$

Thus, if the result hold for any finite subset of S, it holds also for S. From now on, let us assume that S is finite.

Note that $\mathcal{B}(\sigma,\sigma') = \mathcal{B}(\sigma\wedge v,\sigma'\wedge w)$ by (6), so, in what follows, we assume $\sigma\leqslant v$ and $\sigma'\leqslant w$.

Lemma 9. *There exists a sequence of finite partitions $(\mathcal{A}_k)_{k\in\mathbb{N}}$ of S satisfying $\mathcal{A}_0 = \{S\}$ and*

$$\begin{cases} \forall k \in \mathbb{N},\ \mathcal{A}_{k+1} \subset \mathcal{A}_k \textit{ in the sense that: } \forall B \in \mathcal{A}_{k+1}, \exists C \in \mathcal{A}_k \textit{ s.t. } B \subset C, \\ \forall k \geqslant 1,\ \forall B \in \mathcal{A}_k,\ \forall s,u \in B,\ N_2(s-u) \leqslant 2^{-k}\sigma \quad \textit{and} \quad N_\infty(s-u) \leqslant 2^{-k}\sigma', \\ \forall k \geqslant 1,\ |\mathcal{A}_k| \leqslant \left(1+8\frac{v}{\sigma}\right)^D \left(1+8\frac{w}{\sigma'}\right)^D \cdot 5^{2(k-1)D}. \end{cases}$$

Proof of Lemma 9. Let us recall a result from [10, Lemma 4.5] also used by [1]. This result is originally formulated for norms, but extends naturally to seminorms by applying it to the quotient of E by the kernel of the seminorm.

Lemma 10. *Let N be an arbitrary seminorm on S and $\mathcal{B}_N(0,1)$ its corresponding unit ball. For all $\delta \in (0,1]$, the minimum number of balls of radius δ which is necessary to cover $\mathcal{B}_N(0,1)$ is at most $(1+2\delta^{-1})^D$.*

In the following proof, we build separately for each seminorm N_j a sequence of partitions $(\mathcal{A}_{j,k})_{k\in\mathcal{N}}$ with $j \in \{2,+\infty\}$. The sequence of partitions of Lemma 9 is then obtained by choosing, for $k \geqslant 0$, the partition $\mathcal{A}_k$ defined by

$$\mathcal{A}_k = \{A_2 \cap A_\infty, A_2 \in \mathcal{A}_{2,k}, A_\infty \in \mathcal{A}_{\infty,k}\}.$$

For the seminorm N_2: let $\mathcal{A}_{2,0} = S$. By (6), $S \subset \mathcal{B}_{N_2}(0,v)$. Applying Lemma 10 to the norm $v^{-1}N_2$ and $\delta = 4^{-1}v^{-1}\sigma$ means that the minimum number of balls of radius $4^{-1}\sigma$ which are necessary to cover $\mathcal{B}_{N_2}(0,v)$ is upper bounded by $\left(1+8v\sigma^{-1}\right)^D$. Let $(B_1,\ldots,B_p)$ be such a minimal covering. Let $C_1 = B_1$ and for $j \in \{2,\ldots,p\}$, define the set C_j as

$$C_j = B_j \backslash \bigcup_{i<j} B_i.$$

The sequence $\mathcal{A}_{2,1} = (C_j)_{j\in\{1,\ldots,p\}}$ forms a partition of S, each set of which has a diameter at most $2^{-1}\sigma$.

For $k \geqslant 1$, proceed by induction using Lemma 10. Assume that there exists a partition $\mathcal{A}_{2,k}$ such that $|\mathcal{A}_{2,k}| \leqslant (1+8v\sigma^{-1})^D \cdot 5^{(k-1)D}$ and such that each element of $\mathcal{A}_{2,k}$ is a subset of a ball of radius $2^{-(k+1)}\sigma$ for the norm N_2. By applying Lemma 10 to $2^{k+1}\sigma^{-1}N_2$ and $\delta = 2^{-1}$, each element of $\mathcal{A}_{2,k}$ can be covered by at most 5^D balls of radius $2^{-(k+2)}\sigma$, and therefore be partitioned into at most 5^D sets of diameter $2^{-(k+1)}\sigma$, each contained in a ball of radius $2^{-(k+2)}\sigma$. $\mathcal{A}_{2,k+1}$ is therefore a partition containing at most $(1+8v\sigma^{-1})^D \cdot 5^{kD}$ elements.

For the seminorm N_∞: the reasoning is the same and produces a partition $\mathcal{A}_{\infty,k+1}$ containing at most $(1+8w(\sigma')^{-1})^D \cdot 5^{kD}$ element for all $k \geqslant 0$.

Final partition: By construction, for all $k \in \mathbb{N}$, $\mathcal{A}_{k+1} \subset \mathcal{A}_k$. Moreover, for all $k \geqslant 1$, $A \in \mathcal{A}_k$ and $s,u \in A$, $N_2(s-u) \leqslant 2^{-k}\sigma$ and $N_\infty(s-u) \leqslant 2^{-k}\sigma'$, and finally,

$$|\mathcal{A}_k| \leqslant |\mathcal{A}_{2,k}| \cdot |\mathcal{A}_{\infty,k}| \leqslant (1+8v\sigma^{-1})^D(1+8w(\sigma')^{-1})^D \cdot 5^{2(k-1)D}.$$

□

Let $(\mathcal{A}_k)_{k\geqslant 0}$ be a sequence of partitions as in Lemma 9. For all $k \in \mathbb{N}^*$ and all $A \in \mathcal{A}_k$, pick a (deterministic) element $t_k(A) \in A$. For any $t \in S$ and $k \geqslant 1$, there exists a unique $A \in \mathcal{A}_k$ such that $t \in A$. Let $\pi_k(t) = t_k(A)$. Let also $\pi_0(t) = t_0$. Since S is finite, the following decomposition holds and contains a finite number of non-zero terms:

$$Y_t - Y_{t_0} = \sum_{k\geqslant 0}\left(Y_{\pi_{k+1}(t)} - Y_{\pi_k(t)}\right).$$

For $k \geqslant 0$, let $E_k = \{(\pi_k(t), \pi_{k+1}(t)), t \in S\}$ and for $k \geqslant 1$, let

$$\begin{cases} z_0 = \dfrac{3}{2}\sigma\sqrt{2(x+\log(2|E_0|))} + \dfrac{3}{2}\sigma'(x+\log(2|E_0|)), \\ z_k = 2^{-k}\left(\sigma\sqrt{2(x+\log(2^{k+1}|E_k|))} + \sigma'(x+\log(2^{k+1}|E_k|))\right). \end{cases}$$

Let

$$\begin{aligned} H &= \frac{3}{2}\sigma\sqrt{2\log(2|E_0|)} + \frac{3}{2}\sigma'\log(2|E_0|) + \sum_{k\geqslant 1} 2^{-k}\left(\sigma\sqrt{2\log(2^{k+1}|E_k|)} + \sigma'\log(2^{k+1}|E_k|)\right) \\ &= \frac{1}{2}\sigma\sqrt{2\log(2|E_0|)} + \frac{1}{2}\sigma'\log(2|E_0|) + \sum_{k\geqslant 0} 2^{-k}\left(\sigma\sqrt{2\log(2^{k+1}|E_k|)} + \sigma'\log(2^{k+1}|E_k|)\right). \end{aligned}$$

Finally, let

$$z = H + \frac{5}{2}\sigma\sqrt{2x} + \frac{5}{2}\sigma' x, \tag{8}$$

so that $z \geqslant \sum_{k\geqslant 0} z_k$. By definition,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0}) \geqslant z\right) &\leqslant \mathbb{P}\left(\exists t \in \mathcal{B}(\sigma,\sigma'), \exists k \geqslant 0, Y_{\pi_{k+1}(t)} - Y_{\pi_k(t)} \geqslant z_k\right)\\ &\leqslant \mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}\left(Y_{\pi_1(t)} - Y_{t_0}\right) \geqslant z_0\right) + \sum_{k\geqslant 1}\mathbb{P}\left(\sup_{t\in S}\left(Y_{\pi_{k+1}(t)} - Y_{\pi_k(t)}\right) \geqslant z_k\right).\end{aligned}$$

The first term must be handled carefully, since even if $t \in \mathcal{B}(\sigma,\sigma')$, there is no guarantee that $\pi_1(t) \in \mathcal{B}(\sigma,\sigma')$. However, if t is in $\mathcal{B}(\sigma,\sigma')$, since $\pi_1(t)$ and t are in the same element of $\mathcal{A}_1$, by the triangle inequality,

$$R_2(\pi_1(t), t_0) \leqslant R_2(\pi_1(t), t) + R_2(t, t_0) \leqslant \frac{3}{2}\sigma,$$

and likewise $R_\infty(\pi_1(t), t_0) \leqslant \frac{3}{2}\sigma'$. Therefore,

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0}) \geqslant z\right) &\leqslant \mathbb{P}\left(\sup_{u \text{ s.t. } (t_0,u)\in E_0 \text{ and } u\in\mathcal{B}(3\sigma/2,3\sigma'/2)}(Y_u - Y_{t_0}) \geqslant z_0\right) + \sum_{k\geqslant 1}\mathbb{P}\left(\sup_{(s,u)\in E_k}(Y_u - Y_s) \geqslant z_k\right)\\ &\leqslant \sum_{u \text{ s.t. } (t_0,u)\in E_0}\mathbb{P}\left((Y_u - Y_{t_0}) \geqslant z_0 \text{ and } u \in \mathcal{B}(3\sigma/2, 3\sigma'/2)\right) + \sum_{k\geqslant 1}\mathbb{P}\left(\sup_{(s,u)\in E_k}(Y_u - Y_s) \geqslant z_k\right)\\ &\leqslant \sum_{u \text{ s.t. } (t_0,u)\in E_0}\mathbb{P}\left((Y_u - Y_{t_0}) \geqslant z_0 \text{ and } u \in \mathcal{B}(3\sigma/2, 3\sigma'/2)\right) + \sum_{k\geqslant 1}\sum_{(s,u)\in E_k}\mathbb{P}\left(Y_u - Y_s \geqslant z_k\right).\end{aligned}$$

For $k = 0$ Using (4) and the definition of z_0,

$$\begin{aligned}\mathbb{P}\left(Y_u - Y_{t_0} \geqslant z_0 \text{ and } u \in \mathcal{B}(3\sigma/2, 3\sigma'/2)\right) = \mathbb{P}\Bigg(&Y_u - Y_{t_0} \geqslant \frac{3}{2}\sigma\sqrt{2(x + \log(2|E_0|))} + \frac{3}{2}\sigma'(x + \log(2|E_0|))\\ &\text{and } R_2(u, t_0) \leqslant \frac{3}{2}\sigma,\ R_\infty(u, t_0) \leqslant \frac{3}{2}\sigma'\Bigg) \leqslant \frac{1}{2|E_0|}e^{-x}.\end{aligned}$$

For $k \geqslant 1$ Since $\mathcal{A}_{k+1} \subset \mathcal{A}_k$, $\pi_k(t)$ and $\pi_{k+1}(t)$ belong to the same set in $\mathcal{A}_k$. Therefore, for all $(s,u) \in E_k$, $N_2(s-u) \leqslant 2^{-k}\sigma$ and $N_\infty(s-u) \leqslant 2^{-k}\sigma'$. By assumption, $R_2(s,u) \leqslant N_2(s-u)$ and $R_\infty(s,u) \leqslant N_\infty(s-u)$. Thus, for $(s,u) \in E_k$, $R_2(s,u) \leqslant 2^{-k}\sigma$ and $R_\infty(s,u) \leqslant 2^{-k}\sigma'$ almost surely. By definition of z_k and (4), for all $(s,u) \in E_k$,

$$\mathbb{P}(Y_u - Y_s \geqslant z_k) \leqslant 2^{-(k+1)}|E_k|^{-1}e^{-x}.$$

Summing on all $(s,u) \in E_k$ and all $k \geqslant 0$ leads to

$$\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0}) \geqslant z\right) \leqslant e^{-x}.$$

It remains to compute H in (8). By construction of $\mathcal{A}_k$, the choice of $\pi_{k+1}(t)$ entirely determines the choice of $\pi_k(t)$. Therefore, $|E_k| \leqslant |\mathcal{A}_{k+1}|$, that is

$$2^{k+1}|E_k| \leqslant 2^{k+1}\left(1 + 8\frac{v}{\sigma}\right)^D\left(1 + 8\frac{w}{\sigma'}\right)^D \cdot 5^{2kD} \leqslant 2^{k+1}9^{2D} \cdot 5^{2kD} \cdot \left(\frac{v}{\sigma}\right)^D\left(\frac{w}{\sigma'}\right)^D \leqslant \left(162 \cdot 50^k\frac{vw}{\sigma\sigma'}\right)^D,$$

and thus

$$H \leqslant \underbrace{\frac{1}{2}\sigma\sqrt{2D\log(162\frac{vw}{\sigma\sigma'})} + \frac{1}{2}\sigma' D\log(162\frac{vw}{\sigma\sigma'})}_{E} + \underbrace{\sum_{k\geqslant 0}2^{-k}\sigma\sqrt{2D\log\left(162 \cdot 50^k\frac{vw}{\sigma\sigma'}\right)}}_{F} + \underbrace{\sum_{k\geqslant 0}2^{-k}\sigma' D\log\left(162 \cdot 50^k\frac{vw}{\sigma\sigma'}\right)}_{G}.$$

Let us calculate the three terms separately. Firstly,

$$E \leqslant \frac{\sigma}{2}\sqrt{2D\left(6+\log(\frac{vw}{\sigma\sigma'})\right)} + \frac{\sigma'}{2}D\left(6+\log(\frac{vw}{\sigma\sigma'})\right).$$

Secondly,

$$\begin{aligned} G &= \sigma' D \sum_{k\geqslant 0} 2^{-k}\left(\log(162)+k\log(50)+\log(\frac{vw}{\sigma\sigma'})\right) \\ &= 2\sigma' D\left(\log(162)+\log(50)+\log(\frac{vw}{\sigma\sigma'})\right) \leqslant 2\sigma' D\left(9+\log(\frac{vw}{\sigma\sigma'})\right). \end{aligned}$$

Thirdly, by concavity of $x \mapsto \sqrt{x}$ and Jensen's inequality,

$$F \leqslant 2\sigma\sqrt{2\sum_{k\geqslant 0} 2^{-(k+1)} D \log\left(162\cdot 50^k \frac{vw}{\sigma\sigma'}\right)} \leqslant 2\sigma\sqrt{2D\left(9+\log(\frac{vw}{\sigma\sigma'})\right)}.$$

Thus,

$$H \leqslant \frac{5}{2}\sigma\sqrt{2D\left(9+\log(\frac{vw}{\sigma\sigma'})\right)} + \frac{5}{2}\sigma' D\left(9+\log(\frac{vw}{\sigma\sigma'})\right).$$

Finally, using the concavity of $x \mapsto \sqrt{x}$ again,

$$\begin{aligned} z = H + \frac{5}{2}\sigma\sqrt{2x} + \frac{5}{2}\sigma' x &\leqslant \frac{5}{2}\sigma\sqrt{2D\left(9+\log(\frac{vw}{\sigma\sigma'})\right)} + \frac{5}{2}\sigma' D\left(9+\log(\frac{vw}{\sigma\sigma'})\right) + \frac{5}{2}\sigma\sqrt{2x} + \frac{5}{2}\sigma' x \\ &\leqslant 5\sigma\sqrt{x + \left(9+\log(\frac{vw}{\sigma\sigma'})\right)D} + 5\sigma'\left(x + \left(9+\log(\frac{vw}{\sigma\sigma'})\right)D\right). \end{aligned}$$

This concludes the proof of Proposition 3. □

Proof of Theorem 5. By Proposition 8, for all $\sigma, \sigma' > 0$ and $x \geqslant 0$,

$$\mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma,\sigma')}(Y_t - Y_{t_0}) \geqslant 20\left[(\sigma\wedge v)\sqrt{x + D\log\left(\frac{v}{\sigma}\vee\frac{w}{\sigma'}\vee e\right)} + (\sigma'\wedge w)\left(x + D\log\left(\frac{v}{\sigma}\vee\frac{w}{\sigma'}\vee e\right)\right)\right]\right) \leqslant e^{-x}. \tag{9}$$

To obtain a concentration inequality on the renormalized process, we use an extension of [2, Lemma 4.23] to allow for a random function a.

Lemma 11. *Let S be a countable set, $u \in S$, and $a : S \to \mathbb{R}_+$ be a continuous, possibly random, function such that $\|a\|_\infty \leqslant M$ with $M > 0$, and such that $a(u) = \inf_{t\in S} a(t)$. Let Z be some process indexed by S. Let*

$$\mathcal{B}(\sigma) = \{t \in S, a(t) \leqslant \sigma\}.$$

Let $\sigma_* \geqslant 0$, and $\Psi : (x,y) \in [\sigma_*, +\infty) \times \mathbb{R}_+ \mapsto \Psi(x,y) \in \mathbb{R}_+$ a deterministic function such that for all $y \geqslant 0$, $x \mapsto \Psi(x,y)/x$ is non-increasing on $[\sigma_*, +\infty)$, and assume that for all $y \geqslant 0$,

$$\forall \sigma \geqslant \sigma_* \qquad \mathbb{P}\left(\sup_{t\in\mathcal{B}(\sigma)}[Z(t) - Z(u)] \geqslant \Psi(\sigma, y)\right) \leqslant e^{-y}. \tag{10}$$

Then, for any $\alpha > 0$, $y \geqslant 0$ and $x \geqslant \sigma_*$,

$$\mathbb{P}\left(\sup_{t\in S}\frac{Z(t)-Z(u)}{a^2(t)+x^2} \geqslant x^{-2}\Psi(x,y)\left(1+\frac{(1+\alpha)(2+\alpha)}{2\alpha}\right)\right) \leqslant \left(\frac{\log\left(\frac{M}{x}\right)\vee 0}{\log(1+\alpha)} + 2\right)e^{-y}.$$

In particular, picking $\alpha = \sqrt{2}$ leads to, for any $y \geqslant 0$ and $x \geqslant \sigma_*$,

$$\mathbb{P}\left(\sup_{t\in S}\frac{Z(t)-Z(u)}{a^2(t)+x^2} \geqslant 4x^{-2}\Psi(x,y)\right) \leqslant \left(\frac{\log\left(\frac{M}{x}\right)\vee 0}{\log(1+\sqrt{2})} + 2\right)e^{-y}.$$

Let us conclude the proof of Theorem 5 before proving this lemma. Let $c \geqslant 0$, and take $\sigma' = \sigma/c$ in (9) (or $\sigma' = +\infty$ if $c = 0$). To apply Lemma 11, let $a(t) = (R_2 \vee (cR_\infty))(t, t_0)$, $u = t_0$ and $M = v \vee (cw)$. By definition, $a(t_0) = \inf_{t\in S} a(t) = 0$, and by (6),

$$\|a\|_\infty \leqslant \sup_{t\in S}(N_2 \vee (cN_\infty))(t - t_0) \leqslant v \vee (cw).$$

For $x, \sigma \geqslant 0$, let

$$\Psi(\sigma, x) = 20\left[(\sigma \wedge v)\sqrt{x + D\log\left(\frac{v \vee (cw)}{\sigma} \vee e\right)} + (\frac{\sigma}{c} \wedge w)\left(x + D\log\left(\frac{v \vee (cw)}{\sigma} \vee e\right)\right)\right].$$

By (9), Equation (10) is satisfied with $\sigma_* = 0$. For all $x \geqslant 0$, $\sigma \mapsto \Psi(\sigma, x)/\sigma$ is non-increasing on $\mathbb{R}_+$. Therefore, for all $x \geqslant 0$ and $\sigma > 0$,

$$\mathbb{P}\left(\sup_{t\in S}\frac{Y_t - Y_{t_0}}{a^2(t) + \sigma^2} \geqslant 4\sigma^{-2}\Psi(\sigma, x)\right) \leqslant \left(\frac{\log\left(\frac{v\vee(cw)}{\sigma}\right) \vee 0}{\log(1+\sqrt{2})} + 2\right)e^{-x}.$$

□

Proof of Lemma 11. The proof follows the steps of the proof of [2, Lemma 4.23]. We use the notations of Lemma 11. Let $x > 0$, and let $D(x)$ be the first integer $j \geqslant 0$ such that $(1+\alpha)^{j+1}x > M$, so that $D(x) \leqslant \frac{\log(M/x)}{\log(1+\alpha)} \vee 0$. For all $j \in \{0, \dots, D(x) - 1\}$, let

$$\mathcal{C}_j = \left\{t \in S, (1+\alpha)^j x \leqslant a(t) < (1+\alpha)^{j+1}x\right\}.$$

and let $C_{D(x)} = \{t \in S, (1+\alpha)^{D(x)}x \leqslant a(t) \leqslant M\}$. Then, $\left\{\mathcal{B}(x), \{C_j\}_{0\leqslant j\leqslant D(x)}\right\}$ is a partition of S and therefore,

$$\sup_{t\in S}\left[\frac{Z(t) - Z(u)}{a^2(t) + x^2}\right] \leqslant \sup_{t\in\mathcal{B}(x)}\left[\frac{(Z(t) - Z(u))_+}{a^2(t) + x^2}\right] + \sum_{j=0}^{D(x)}\sup_{t\in\mathcal{C}_j}\left[\frac{(Z(t) - Z(u))_+}{a^2(t) + x^2}\right]. \tag{11}$$

For $t \in \mathcal{C}_j$, since $a^2(t) + x^2 \geqslant (1+\alpha)^{2j}x^2 + x^2$,

$$x^2\sup_{t\in S}\left[\frac{Z(t) - Z(u)}{a^2(t) + x^2}\right] \leqslant \sup_{t\in\mathcal{B}(x)}(Z(t) - Z(u))_+ + \sum_{j=0}^{D(x)}\left(1 + (1+\alpha)^{2j}\right)^{-1}\sup_{t\in\mathcal{B}((1+\alpha)^{j+1}x)}(Z(t) - Z(u))_+.$$

Since $a(u) = \inf_{t\in S} a(t)$, $u \in \mathcal{B}((1+\alpha)^k x)$ for every integer k for which $\mathcal{B}((1+\alpha)^k x)$ is non empty and therefore,

$$\sup_{t\in\mathcal{B}((1+\alpha)^k x)}(Z(t) - Z(u))_+ = \sup_{t\in\mathcal{B}((1+\alpha)^k x)}(Z(t) - Z(u)).$$

Let $y \geqslant 0$. By assumption,

$$\begin{cases}\mathbb{P}\left(\sup_{t\in\mathcal{B}(x)}(Z(t) - Z(u)) \geqslant \Psi(x, y)\right) \leqslant e^{-y} \\ \mathbb{P}\left(\sup_{t\in\mathcal{B}((1+\alpha)^{j+1}x)}(Z(t) - Z(u)) \geqslant \Psi((1+\alpha)^{j+1}x, y)\right) \leqslant e^{-y} \quad \text{for all } j \in \{0, \dots, D(x)\}.\end{cases}$$

Recall that $z \mapsto \Psi(z, y)/z$ is non-increasing, so that $(1+\alpha)^{j+1}\Psi(x, y) \geqslant \Psi((1+\alpha)^{j+1}x, y)$, which leads to

$$\mathbb{P}\left(\sup_{t\in\mathcal{B}((1+\alpha)^{j+1}x)}(Z(t) - Z(u)) \geqslant (1+\alpha)^{j+1}\Psi(x, y)\right) \leqslant e^{-y} \quad \text{for all } j \in \{0, \dots, D(x)\}.$$

Taking the union bound leads to

$$\mathbb{P}\left(x^2\sup_{t\in S}\left[\frac{Z(t) - Z(u)}{a^2(t) + x^2}\right] \geqslant \Psi(x, y) + \sum_{j=0}^{D(x)}\frac{(1+\alpha)^{j+1}}{1 + (1+\alpha)^{2j}}\Psi(x, y)\right) \leqslant (D(x) + 2)e^{-y}.$$

To conclude, note that

$$\sum_{j=0}^{D(x)}\frac{(1+\alpha)^{j+1}}{1 + (1+\alpha)^{2j}} \leqslant (1+\alpha)\left(\frac{1}{2} + \sum_{j\geqslant 1}(1+\alpha)^{-j}\right) \leqslant (1+\alpha)\left(\frac{1}{2} + \frac{1}{\alpha}\right) = \frac{(1+\alpha)(2+\alpha)}{2\alpha}.$$

□

References

[1] Y. Baraud, A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression, Bernoulli 16 (4) (2010) 1064 – 1085.

[2] P. Massart, Concentration inequalities and model selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003, Springer, 2007.

[3] V. Spokoiny, T. Dickhaus, et al., Basics of modern mathematical statistics, Springer, 2015.

[4] N. R. Hansen, P. Reynaud-Bouret, V. Rivoirard, Lasso and probabilistic inequalities for multivariate point processes, Bernoulli 21 (1) (2015) 83 – 143.

[5] V. H. de la Peña, M. J. Klass, T. Leung Lai, Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws, The Annals of Probability 32 (3) (Jul. 2004).

[6] B. Bercu, T. Touati, New insights on concentration inequalities for self-normalized martingales, Electronic Communications in Probability 24 (none) (2019) 1 – 12.

[7] B. Bercu, B. Delyon, E. Rio, et al., Concentration inequalities for sums and martingales, Springer, 2015.

[8] V. Spokoiny, M. Zhilova, Sharp deviation bounds for quadratic forms, Mathematical Methods of Statistics 22 (2013) 100–113.

[9] C. Houdré, P. Reynaud-Bouret, Exponential inequalities for u-statistics of order two with constants, Stochastic Inequalities and Applications. Progress in Probability 56 (01 2002).

[10] L. Birgé, Approximation dans les espaces métriques et théorie de l'estimation, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 65 (1983) 181–237.