Concentration inequalities for suprema of unbounded empirical processes

Antoine Marchina

Annales Henri Lebesgue
4 (2021) 831-861

ANNALES
HENRI LEBESGUE

ANTOINE MARCHINA

CONCENTRATION INEQUALITIES FOR SUPREMA OF UNBOUNDED EMPIRICAL PROCESSES

INÉGALITÉS DE CONCENTRATION POUR LES SUPREMA DE PROCESSUS EMPIRIQUES NON BORNÉS

Abstract. — In this paper, we provide new concentration inequalities for suprema of (possibly) non-centered and unbounded empirical processes associated with independent and identically distributed random variables. In particular, we establish Fuk–Nagaev type inequalities with the optimal constant in the moderate deviation bandwidth. The proof builds on martingale methods and comparison inequalities, allowing to bound generalized quantiles as the so-called Conditional Value-at-Risk. Importantly, we also extent the left concentration inequalities of Klein (2002) to classes of unbounded functions.

Résumé. — Dans cet article, nous donnons des inégalités de concentration pour des suprema de processus empiriques non bornés et (éventuellement) non centrés, associés à des variables aléatoires indépendantes et identiquement distribuées. En particulier, nous établissons des inégalités de type Fuk–Nagaev avec constantes optimales dans la bande des moyennes déviations. Notre approche est basée sur des techniques de martingales et des inégalités de comparaison permettant de majorer des quantiles généralisés comme la CVaR. Nous étendons également les inégalités de concentration à gauche de Klein (2002) à des classes de fonctions non bornées.

Keywords: concentration inequalities, empirical processes, martingale method, generalized moments.
2020 *Mathematics Subject Classification:* 60E15.
DOI: `https://doi.org/10.5802/ahl.90`
(*) Many thanks to Emmanuel Rio for enlightening discussions and his careful reading of this work.

1. Introduction

Let $\mathcal{T}$ be a countable index set. For each $k = 1, \ldots, n$, let $X_k := (X_{k,t})_{t\in\mathcal{T}}$ be a collection of centered real-valued random variables such that $X_1, \ldots, X_n$ are independent and identically distributed according to a law P. Define the envelope of the collection of coordinates by

$$M_k := \sup_{t\in\mathcal{T}} |X_{k,t}| \qquad \text{for all } k = 1, \ldots, n\,. \tag{1.1}$$

Throughout the paper, we assume that

$$\mathbb{E}\left[M_1^2\right] < \infty\,. \tag{1.2}$$

Set now for all $t \in \mathcal{T}$, $S_{n,t} := X_{1,t} + \ldots + X_{n,t}$. For a given (deterministic) vector $c := (c_t)_{t\in\mathcal{T}}$, we define

$$Z := \sup\{S_{n,t} + nc_t : t \in \mathcal{T}\}\,. \tag{1.3}$$

The vector c is introduced to consider possibly non-centered empirical processes.

Remark 1.1. — In this paper, as in [BLM13, LvdG14], we express the empirical process in terms of random vector. If the indices $t \in \mathcal{T}$ are associated with measurable functions $f_t : \mathcal{X} \to \mathbb{R}$ defined on some measurable space $(\mathcal{X}, \mathcal{F})$, and if $Y_1, \ldots, Y_n$ is a sequence of centered iid random variables, then defining $X_{k,t} = f_t(Y_k)$ leads to the other classical notation (in the centered case),

$$Z = \sup\left\{\sum_{k=1}^{n} f(Y_k) : f \in \mathscr{F}\right\},$$

where $\mathscr{F} := \{f_t : t \in \mathcal{T}\}$.

The question asked in this paper is: without the classical boundedness condition on the random variables $X_{k,t}$, how to provide an upper bound on the upper tail quantiles of $Z - \mathbb{E}[Z]$ and $\mathbb{E}[Z] - Z$?

As a direct consequence of such bounds, we will then able to derive deviation inequalities and moment inequalities which have their own interest in practical applications such as statistical learning. Let us emphasize that in the literature, there are only few studies in the unbounded setting. First, we recall that the bounded case (that is, for all $t \in \mathcal{T}$, $X_{k,t} \leqslant 1$ a.s. or $|X_{k,t}| \leqslant 1$ a.s) is handled by Talagrand's inequality. Precisely, it is the merit of Talagrand [Tal96] to obtain the first a functional Bennett inequality for suprema of empirical processes, but with unspecified constants. Ledoux [Led97] introduces a new method based on entropic inequalities to recover more directly Talagrand's result. This method is the starting point of a series of paper, mainly to reach optimal constants in Talagrand's inequality: Massart [Mas00], Rio [Rio01, Rio02, Rio12], Bousquet [Bou03], Klein [Kle02] and Klein and Rio [KR05]. For an overview of the results in the bounded case, we refer the reader to Section 12 in the book of Boucheron, Lugosi, and Massart [BLM13].

The different attempts to get tail bounds for suprema of unbounded empirical processes I am aware of, have been investigated by Adamczak [Ada08], Lederer and van de Geer [LvdG14, L13], and Dirksen [Dir15]. Under the moment condition on the envelope $\|M_1\|_{\psi_\alpha} < +\infty$ for some $\alpha \in \,]0\,,1]$, where $\|.\|_{\psi_\alpha}$ is the Orlicz norm associated with $\psi_\alpha(x) = \exp(x^\alpha) - 1$ (the case $\alpha = 1$ corresponds to having sub-exponential tails), Adamczak [Ada08, Theorem 4] proves the following deviation inequality: let $\widetilde{Z} = \sup_{t\in\mathcal{T}} |\sum_{k=1}^n X_{k,t}|$, then for all $0 < \varepsilon < 1$, $\delta > 0$, and $t > 0$,

$$\text{(1.4)} \quad \mathbb{P}\left(\widetilde{Z} \geqslant (1+\varepsilon)\mathbb{E}[\widetilde{Z}] + t\right) \leqslant \exp\left(-\frac{t^2}{2(1+\delta)n\sigma^2}\right) + 3\exp\left(-\left(\frac{t}{C_{\alpha,\varepsilon,\delta}\|\max_k M_k\|_{\psi_\alpha}}\right)^\alpha\right),$$

where $\sigma^2 = \sup_{t\in\mathcal{T}} \operatorname{Var}(X_{1,t})$ is the wimpy variance, and $C_{\alpha,\varepsilon,\delta}$ is an unspecified constant. In [LvdG14, Corollary 3.1], Lederer and van de Geer only require that $\|M_1\|_p < \infty$ for some $p > 2$, and obtain that for all $\varepsilon > 0$ and $t > 0$,

$$\text{(1.5)} \quad \mathbb{P}\left(\widetilde{Z} \geqslant (1+\varepsilon)\mathbb{E}[\widetilde{Z}] + t\right) \leqslant \min_{1\leqslant r\leqslant p} t^{-r}\left(4\sigma\sqrt{rn} + \left(\frac{64}{\varepsilon} + 7 + \varepsilon\right) r^{1-r/p} n^{r/p}\|M_1\|_p\right)^r .$$

These two results are based on truncation arguments and on Talagrand type inequalities for suprema of bounded empirical processes. That explains the additional factor $(1+\varepsilon)$ in front of $\mathbb{E}[Z]$. The same authors in [L13, Theorem 8] provide an exponential tail bound for $\widetilde{Z}$ under a uniform Bernstein condition. To do this, they introduce the Bernstein–Orlicz norm that they combine with a variation of the classical chaining. In [Dir15, Corollary 5.2], assuming the same uniform Bernstein condition as in [L13], Dirksen uses the generic chaining to upper bound all p^{th} moments of $\widetilde{Z}$ and thus he derives an exponential tail bound by a standard optimization argument.

In the present paper, we relax the previous tail assumptions by only assuming the mild condition that $\|M_1\|_p$ is finite for some $p > 2$. We will prove in particular the following deviation inequality around $\mathbb{E}[Z]$, without extra centering term $\varepsilon\mathbb{E}[Z]$ and with explicit constants (see Theorem 2.1 below): for any $u \in]0\,,1[$,

$$\mathbb{P}\left(Z > \mathbb{E}[Z] + \sqrt{2n\log(1/u)}\ (\sigma + \sqrt{v_n}\,) +\ n^{1/p}(p+8)\,\|M_1\|_p u^{-1/p}\right) \leqslant u,$$

where v_n satisfies $v_n = o(1)$ as n tends to infinity. We emphasize that the constant 1 in front of $\sigma\sqrt{2n\log(1/u)}$ is optimal and, in addition, we show that we cannot expect to have a nonasymptotic upper bound with σ instead of $\sigma + v_n$ (see Section 1.3 below). Moreover, our term v_n is of a correct order. The proofs of our results rely on martingales methods, comparison inequalities, and recent concentration inequalities for real-valued martingales proved by Rio [Rio17a]. Another important point is that we do not seek to directly control the upper tail quantile as usual, but instead we control the Conditional Value-At-Risk of $Z - \mathbb{E}[Z]$. Finally, let us mention that we already use martingale methods in a previous work to obtain concentration inequalities for (general) separately convex functions [Mar18]. However when we apply the results to the case of unbounded empirical processes, the upper bound suffer from the fact that they do not involve the wimpy variance σ^2 and then are less interesting for the statistical applications we have in mind.

In order to explain our results, let us give some definitions and notation.

1.1. Upper tail quantile and Conditional Value-At-Risk

Let X be a real-valued random variable. As usual, we denote by F_X its distribution function and by F_X^{-1} the càdlàg inverse of F_X.

Definition 1.2. — *The "upper tail" quantile function of X, which is the càdlàg inverse of the tail function $t \mapsto 1 - F_X(t)$, is denoted by Q_X. It is defined by*

$$Q_X(u) := \inf\{x \in \mathbb{R} : 1 - F_X(x) \leqslant u\}\,.$$

Note that $Q_X(u)$ is the value of the usual quantile function at point $1-u$. The basic property of Q_X is: $x < Q_X(u)$ if and only if $1 - F_X(x) > u$. This ensures that $Q_X(U)$ has the same distribution as X for any random variable U uniformly distributed over $[0\,,1]$.

Definition 1.3. — *Assume that X is integrable. The Conditional Value-at-Risk (CVaR for short) of X is defined by*

$$\widetilde{Q}_X(u) := \frac{1}{u}\int_0^u Q_X(s)ds \quad \text{for any } u \in\,]0\,,1]\,. \tag{1.6}$$

It is worth noticing that, for any $u \in\,]0\,,1[$, if the distribution of X has no atom at $Q_X(u)$, then (see, for instance, [Pin14b, Section 5.9] in Pinelis)

$$\mathbb{E}\left[X \mid X > Q_X(u)\right] = u^{-1} \sup_{\mathbb{P}(A)\leqslant u} \mathbb{E}\left[X \mathbb{1}_A\right] = \widetilde{Q}_X(u)\,.$$

When X has a finite Laplace transform on a right neighborhood of 0, we denote by ℓ_X the log-Laplace transform of X, given by

$$\ell_X(t) := \log \mathbb{E}\left[\exp(tX)\right] \quad \text{for any } t \geqslant 0\,. \tag{1.7}$$

The Legendre transform ℓ_X^* of X is defined by

$$\ell_X^*(\lambda) := \sup\left\{\lambda t - \ell_X(t) : t > 0\right\} \quad \text{for any } \lambda \geqslant 0\,. \tag{1.8}$$

We recall that the inverse function of ℓ_X^* admits the following variational expression (see, for instance, Rio [Rio17b, Lemma A.2]):

$$\ell_X^{*-1}(x) = \inf\left\{t^{-1}(\ell_X(t) + x) : t > 0\right\} \quad \text{for any } x \geqslant 0\,. \tag{1.9}$$

Clearly, one has $Q_X \leqslant \widetilde{Q}_X$. Moreover,

$$\widetilde{Q}_X(u) \leqslant \ell_X^{*-1}(\log(1/u)) \quad \text{for any } u \in\,]0\,,1]\,. \tag{1.10}$$

This result may be found in Pinelis [Pin14b, Theorem 3.4]. In this paper, we will focus on upper bounds on the CVaR of $Z - \mathbb{E}[Z]$ instead of on its upper tail quantile function (also known as the Value-at-Risk). To a certain extent, we can consider that the CVaR is superior than the Value-at-Risk since it has stronger mathematical properties which are very useful in optimization modeling or statistics. We refer the reader to Pinelis [Pin14b] or Rockafellar and Uryasev [RU00] (see also the references therein) for more on the CVaR. One property that will be of interest to us is the subadditivity of the CVaR (see [Pin14b, Theorem 3.4]):

Proposition 1.4. — *Let X and Y be real-valued and integrable random variables (X and Y may not be independent). Then, for any $u \in\,]0\,,1]$,*

$$\widetilde{Q}_{X+Y}(u) \leqslant \widetilde{Q}_X(u) + \widetilde{Q}_Y(u)\,.$$

1.2. Aim and organization of the paper

In a well known paper [CIS76], Cirel'son, Ibragimov and Sudakov proved a comparison inequality between the quantiles of a supremum of a Gaussian process and those of a standard Gaussian random variable (see [CIS76, pages 22-23]). Precisely, let $(G_t)_{t\in\mathcal{T}}$ be a centered Gaussian process indexed by $\mathcal{T}$ such that $Z_G := \sup_{t\in\mathcal{T}} G_t < \infty$ almost surely. Define $\sigma_G^2 := \sup_{t\in\mathcal{T}} \mathbb{E}[G_t^2]$, and let Y be a standard Gaussian random variable. They proved that for any $u, v \in]0\,,1]$,

$$\left|Q_{Z_G-\mathbb{E}[Z_G]}(u) - Q_{Z_G-\mathbb{E}[Z_G]}(v)\right| \leqslant \sigma_G \left|Q_Y(u) - Q_Y(v)\right| . \tag{1.11}$$

Now, combining this inequality with (1.10) yields that for any $u \in]0\,,1]$,

$$\widetilde{Q}_{Z_G-\mathbb{E}[Z_G]}(u) \leqslant \sigma_G\, \widetilde{Q}_Y(u) \leqslant \sigma_G\, \ell_Y^{*-1}(\log(1/u)) = \sigma_G \sqrt{2\log(1/u)}\,. \tag{1.12}$$

For the convenience of the reader, we provide a proof of the first inequality in (1.12) in Appendix A. Note that this inequality could be also deduced from Bobkov [Bob03, Corollary 2].

Since the empirical process $S_{n,t}$, $t \in \mathcal{T}$, converges fidi (after suitable normalization) towards a Gaussian process, the aim is then to reach an extension of the Gaussian bound (1.12) for $Z - \mathbb{E}[Z]$ in the unbounded case.

Let us mention the considerable work of Boucheron, Bousquet, Lugosi, and Massart [BBLM05] concerning moment inequalities for general functions of independent random variables. Their methods are based on an extension of the entropy method proposed by Ledoux, which allows to derive a moment inequality on $\|(Z-\mathbb{E}[Z])_+\|_p$, $p > 2$, for centered processes (that is $c \equiv 0$ in (1.3)). However, considering the upper bound that their inequality provides on the quantile, we may notice that it is suboptimal in the moderate deviations bandwidth. Precisely, there is a power $u^{-1/p}$ instead of the term $\sqrt{\log(1/u)}$ we would like to have by comparing with the right-hand side in (1.12). Furthermore, it seems beyond the scope of classical functional analysis tools to handle non-centered empirical processes.

Our method is based on a martingale decomposition of $Z - \mathbb{E}[Z]$ into a sum of two martingales. The subadditivity of the CVaR allows us to treat each martingale separately. The difficulties lie in the control of their increments, in particular their conditional variance. To this end, we shall establish comparison inequalities with respect to a class of convex functions, similar to Hoeffding's classical inequalities concerning bounded random variables [Hoe63].

In addition, we want our inequalities on $Z - \mathbb{E}[Z]$ to generalize those known for the sums of independent random variables, corresponding to a class $\mathcal{T}$ reduced to one element. When the random variables are unbounded but with finite variances, their deviations are handled by Fuk–Nagaev inequalities. The best known constants are given recently by Rio [Rio17a] in the context of real-valued martingales. In the setting of real-valued iid centered random variables $X_1, \ldots, X_n$, he proves that, for any $u \in]0\,,1]$,

$$\widetilde{Q}_{S_n}(u) \leqslant \sigma\sqrt{2n\log(1/u)} + n^{1/p} K_p \left\|X_1\right\|_p u^{-1/p}, \tag{1.13}$$

where $p > 2$, $S_n := X_1 + \ldots + X_n$, the constant K_p, depending only on p, is explicit and $\sigma^2 := \mathrm{Var}(X_1)$. The reader may notice that the constant 1 in front of the term $\sigma\sqrt{2n\log(1/u)}$ is optimal.

The organization of the paper is as follows. In Section 2, we present our main results which are Fuk–Nagaev and Rosenthal type inequalities for $Z - \mathbb{E}[Z]$ and $\mathbb{E}[Z] - Z$. A detailed application under the assumption that the $X_{k,t}$'s have power tails is given. In Section 3, we derive upper bounds for $\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+^\alpha]$, $\alpha \geqslant 1$. In Section 4, we study left deviation inequalities which turn out to be easier to obtain than the right ones. All the proofs are deferred to Section 5. Before starting, let us do some comments on the variance factor in our inequalities.

1.3. About the variance factor

Let $(G_t)_{t\in\mathcal{T}}$ be a centered Gaussian process such that $Z_G := \sup_{t\in\mathcal{T}} G_t < \infty$ almost surely. As a consequence of (1.12), one can derive the following upper bound on the variance:

$$\mathrm{Var}(Z_G) \leqslant \sigma_G^2, \tag{1.14}$$

where $\sigma_G^2 := \sup_{t\in\mathcal{T}} \mathbb{E}[G_t^2] = \sup_{t\in\mathcal{T}} \mathrm{Var}(G_t)$. Then, a natural question is whether (1.14) is satisfied by Z with

$$\sigma^2 := \sup\left\{\mathrm{Var}(X_{1,t}) : t \in \mathcal{T}\right\} = \sup\left\{\mathbb{E}\left[X_{1,t}^2\right] : t \in \mathcal{T}\right\}. \tag{1.15}$$

The answer is no, even in the bounded case (see, for instance, [BLM13, Exercise 11.1] for a simple counterexample). Let us give another example in the unbounded case that will be useful in Section 2 to comment on the variance factor appearing in our results.

Example 1.5. — Let $U, U_1, \ldots, U_n$ be a sequence of iid random variables uniformly distributed on $[0\,,1]$. Let $\ell > 2$, $\Delta > 0$ and $p \in \,]0\,,1[$. Let $\mathcal{S}$ be the set of all finite unions of disjoint intervals with rational endpoints, which measure is lower than p, and included in $[0\,,\Delta]$. Define

$$Z := \sup_{S\in\mathcal{S}} \sum_{k=1}^{n} \left(U_k^{-1/\ell}\mathbb{1}_S(U_k) - \int_S u^{-1/\ell}du\right),$$

and $\sigma^2 := \sup_{S\in\mathcal{S}} \mathrm{Var}(U^{-1/\ell}\mathbb{1}_S(U))$. Then, for p and Δ small enough, one can show that there exists $K > 0$ such that

$$\frac{1}{n}\mathrm{Var}(Z) - \sigma^2 \geqslant K\left(\frac{\mathbb{E}[Z]}{n}\right)^{\frac{\ell-2}{\ell-1}}. \tag{1.16}$$

The details of the proof of (1.16) are deferred to Appendix A. Therefrom, since we want to provide nonasymptotic inequalities, we cannot expect, as in (1.13), the quantity $\sigma\sqrt{2n\log(1/u)}$ in the moderate deviation part: a corrective term to σ is required. Let us now describe this quantity which appears in our inequalities.

First, we set for all $k = 1, \ldots, n$,

$$E_k := \mathbb{E}\left[\sup_{t \in \mathcal{T}} \frac{1}{k} \sum_{j=1}^{k} X_{j,t}\right]. \tag{1.17}$$

Next, we define the following class of distribution functions:

Definition 1.6. — *Let $q \in [0, 1]$. Let ψ be a nonnegative random variable and set $b_{\psi,q} := F_{\psi}^{-1}(1-q)$. We denote by $F_{\psi,q}$ the distribution function of $\psi \mathbb{1}_{\psi \geqslant b_{\psi,q}}$, that is*

$$F_{\psi,q}(x) := (1-q)\mathbb{1}_{0 \leqslant x < b_{\psi,q}} + F_{\psi}(x)\mathbb{1}_{x \geqslant b_{\psi,q}} \quad \textit{for all } x \in \mathbb{R}. \tag{1.18}$$

Let ψ and X be two nonnegative random variables such that X is first-order stochastically dominated by ψ, that is $\mathbb{P}(X > x) \leqslant \mathbb{P}(\psi > x)$ for all $x > 0$. Let $\zeta_{\psi,q}$ be a random variable with distribution function $F_{\psi,q}$, where q is such that $\mathbb{E}[X] \leqslant \mathbb{E}[\zeta_{\psi,q}]$. Then [Ben08, Lemma 1] of Bentkus (see also [Mar18, Lemma 4.3] in Marchina) ensures that

$$\mathbb{E}[\varphi(X)] \leqslant \mathbb{E}[\varphi(\zeta_{\psi,q})] \quad \text{for any function } \varphi \in \mathcal{H}_+^1, \tag{1.19}$$

where $\mathcal{H}_+^1 := \{\varphi : \varphi \text{ is convex, differentiable, and } \lim_{x \to -\infty} \varphi(x) = 0\}$.

If $\mathbb{E}[X] = \mathbb{E}[\zeta_{\psi,q}]$, then the above inequality is true for any convex function. In order to have a better understanding of this result, see that the distributions $F_{\psi,q}$, $q \in [0, 1]$, are the extremal ones satisfying $0 \leqslant X \leqslant_{\mathrm{st}} \psi$, where $\leqslant_{\mathrm{st}}$ denote the first-order stochastic dominance. If $\psi = b$ is a constant, the extremal distributions are the two-valued ones $\mu_q := (1-q)\delta_0 + q\delta_b$, $q \in [0, 1]$. And it is well known that for any convex function φ, $\mathbb{E}[\varphi(X)] \leqslant \int \varphi d\mu_{q_0}$, where q_0 is such that $\mathbb{E}[X] = \int x d\mu_{q_0}$ (see Hoeffding [Hoe63]). Then (1.19) is an extension of Hoeffding's comparison inequality to unbounded random variables.

Notation 1.7. — Throughout the rest of the paper, ζ_k denotes a random variable with distribution function F_{2M_1, q_k} where q_k is the greatest real in $[0, 1]$ such that $\mathbb{E}[\zeta_k] = E_k$. Define also

$$V_n := \sum_{k=1}^{n} \mathbb{E}\left[\zeta_k^2\right], \text{ and } v_n := \frac{V_n}{n}. \tag{1.20}$$

The variance factor in our inequalities is $\sqrt{n}(\sigma + \sqrt{v_n})$. We can see that $\sqrt{v_n}$ is indeed a "corrective" term. If the class $\mathcal{T}$ satisfies the uniform law of large numbers, that is $\sup_{t \in \mathcal{T}} |n^{-1} S_{n,t}|$ converges to 0 in probability, then E_n decreases to 0 (see, for instance, [W96, Section 2.4] of van der Vaart and Wellner). Now, from the square integrability of M_1 and the definition of the random variable ζ_n, the convergence of E_n to 0 implies the convergence of $\mathbb{E}[\zeta_n^2]$ to 0, which ensures that $V_n = o(n)$ as n tends to infinity. Thus,

$$\sqrt{n}(\sigma + \sqrt{v_n}) \sim \sigma\sqrt{n} \quad \text{as } n \text{ tends to infinity}. \tag{1.21}$$

2. Fuk–Nagaev and Rosenthal type inequalities

In this section we provide Fuk–Nagaev and Rosenthal type inequalities for $Z-\mathbb{E}[Z]$ and $\mathbb{E}[Z]-Z$. We first introduce some more definitions and notations. For any real-valued integrable random variable X and any $r \geqslant 1$, define

$$\Lambda_r^+(X) := \sup_{t>0} t\,(\mathbb{P}(X>t))^{1/r}\,. \tag{2.1}$$

We say that X has a weak moment of order r if $\Lambda_r^+(|X|)$ is finite. We denote by $\mathbb{L}_r^w$ the space of real-valued random variables with a finite weak moment of order r. Note that from the definition of Q_X, we have (see, for instance, [BS88, Chapter 4] of Bennett and Sharpley)

$$\Lambda_r^+(X) = \sup_{u\in\,]0,\,1]} u^{1/r} Q_X(u)\,. \tag{2.2}$$

2.1. Fuk–Nagaev type inequalities

We will now formulate the main results of this part, starting by right-hand side deviations.

Theorem 2.1. — *Let Z be defined by* (1.3). *Let $p>2$ and assume that $M_1 \in \mathbb{L}_p^w$. Let $\mu_p := 2+\max(4/3,p/3)$. Then for any $u\in\,]0\,,1[$,*

$$\widetilde{Q}_{Z-\mathbb{E}[Z]}(u) \leqslant \sqrt{2n\log(1/u)}\;(\sigma+\sqrt{v_n}) + 3\,n^{1/p}\mu_p\,\Lambda_p^+(M_1)u^{-1/p}. \tag{a}$$

Consequently,

$$\mathbb{P}\left(Z > \mathbb{E}[Z] + \sqrt{2n\log(1/u)}\;(\sigma+\sqrt{v_n}) + 3\,n^{1/p}\mu_p\,\Lambda_p^+(M_1)u^{-1/p}\right) \leqslant u. \tag{b}$$

Remark 2.2. — For any real-valued random variable X, by Markov's inequality, $\Lambda_p^+(X) \leqslant \|X\|_p$. Then, if the envelope M_1 has a finite p^{th} moment, one can replace the weak moment $\Lambda_+^p(M_1)$ by the strong moment $\|M_1\|_p$.

Remark 2.3. — Note that an upper bound on the CVaR of Z immediately gives an upper bound on the upper tail quantile of

$$Z^* := \max_{k\leqslant n}\sup_{t\in\mathcal{T}}\left\{\sum_{j=1}^{k} X_{j,t} + kc_t\right\}. \tag{2.3}$$

Indeed, for any $u\in\,]0\,,1[$,

$$\begin{aligned} Q_{Z^*}(u) &\leqslant \widetilde{Q}_Z(u) \\ &\leqslant \mathbb{E}[Z] + \sqrt{2n\log(1/u)}\;(\sigma+\sqrt{v_n}) + 3\,n^{1/p}\mu_p\,\Lambda_p^+(M_1)u^{-1/p}. \end{aligned} \tag{2.4}$$

The first inequality follows from a byproduct of Doob's maximal inequality which can be found in Gilat and Meilijson [GM88].

The term v_n depends only on E_k, $k=1,\ldots,n$, and on the tail distribution of M_1 (see (1.20)). We provide below a useful and explicit upper bound on v_n:

Lemma 2.4. — *One has*

$$v_n \leqslant \left(2\Lambda_p^+(M_1)\right)^{\frac{p}{p-1}} \frac{p}{p-2}\left(1-\frac{1}{p}\right)^{\frac{p-2}{p-1}} \frac{1}{n}\sum_{k=1}^{n} E_k^{\frac{p-2}{p-1}}.$$

If a Donsker theorem holds, one can consider that $E_n \asymp n^{-1/2}$ (see, for instance, van der Vaart and Wellner [W96]). Then Lemma 2.4 yields that

$$v_n \leqslant K_p n^{-\frac{1}{2}\frac{p-2}{p-1}}, \tag{2.5}$$

where K_p is a constant depending only on p. This upper bound has to be linked with inequality (1.16). Indeed, we have shown in Example 1.5, which is a particular case of power-type tail, that we have to add to σ^2 in the variance factor, a term which is at least of order $n^{-\frac{1}{2}\frac{p-2}{p-1}}$ when $E_n \asymp n^{-1/2}$. Thus, it supports the idea that our additional term v_n is of a correct order.

For left-hand side deviations, the concentration bounds are similar but the proofs are simpler than for the right-hand side. In fact, it has already been noted by Samson [Sam07] in the context of transport methods, that martingale like techniques allow to obtain left deviations more easily as regards to the entropy method introduced by Ledoux.

Theorem 2.5. — *Let Z be defined by* (1.3). *Let $p > 2$. Assume that $M_1 \in \mathbb{L}_p^w$ and that for some $r \geqslant p$,*

$$m_r^r := \sup_{t\in\mathcal{T}} \mathbb{E}\left[(-X_{1,t})_+^r\right] < \infty\,.$$

Let $\mu_p := 2 + \max(4/3, p/3)$. Then for any $u \in]0\,,1[$,

$$\text{(a)} \qquad \widetilde{Q}_{\mathbb{E}[Z]-Z}(u) \leqslant \sqrt{2n\log(1/u)}\;(\sigma + \sqrt{v_n}\,) + n^{1/r}\mu_\ell\, m_r u^{-1/r}.$$

Consequently,

$$\text{(b)} \qquad \mathbb{P}\left(Z < \mathbb{E}[Z] - \sqrt{2n\log(1/u)}\;(\sigma + \sqrt{v_n}\,) - n^{1/r}\mu_r\, m_r u^{-1/r}\right) \leqslant u.$$

2.2. Weak and strong Rosenthal type inequalities

We start with weak Rosenthal inequalities derived from the inequalities of the previous section. We first introduce some more notation. Define

$$\widetilde{\Lambda}_r^+(Y) := \sup_{u\,\in\,]0,\,1]} u^{(1/r)-1}\int_0^u Q_Y(s)ds = \sup_{u\,\in\,]0,\,1]} u^{1/r}\widetilde{Q}_Y(u)\,. \tag{2.6}$$

Hence, we get that

$$\Lambda_r^+(Y) \leqslant \widetilde{\Lambda}_r^+(Y) \leqslant \left(\frac{r}{r-1}\right)\Lambda_r^+(Y)\,. \tag{2.7}$$

Furthermore, from the subadditivity of the CVaR (Proposition 1.4), $\widetilde{\Lambda}_r^+(.)$ is subadditive, which implies that $\widetilde{\Lambda}_r^+(.)$ is a norm on the space $\mathbb{L}_r^w$.

Now, when the envelope M_1 has a weak moment of order $p > 2$, proceeding as in Rio [Rio17a], we derive from Theorems 2.1 and 2.5 the following inequalities:

Corollary 2.6. — *Let Z be defined by* (1.3). *Let $p > 2$ and assume that M_1 have a weak moment of order p. Let $\mu_p := 2 + \max(4/3, p/3)$. Then*

$$\Lambda_p^+(Z - \mathbb{E}[Z]) \leqslant \widetilde{\Lambda}_p^+(Z - \mathbb{E}[Z]) \tag{a}$$

$$\leqslant \sqrt{(p/e)}\sqrt{n}\ (\sigma + \sqrt{v_n}\,) + 3\, n^{1/p}\mu_p\, \Lambda_p^+(M_1)\,. \tag{b}$$

Moreover,

$$\Lambda_p^+(\mathbb{E}[Z] - Z) \leqslant \widetilde{\Lambda}_p^+(\mathbb{E}[Z] - Z) \tag{c}$$

$$\leqslant \sqrt{(p/e)}\sqrt{n}\ (\sigma + \sqrt{v_n}\,) + n^{1/p}\mu_p\, \Lambda_p^+(M_1)\,. \tag{d}$$

The variational formula (2.6) ensures that these inequalities above are the optimal ones that can be achieved from the Fuk–Nagaev type inequalities in Theorems 2.1-2.5. Unfortunately, such a formula linking strong moments of order $p > 2$ of a real-valued random variable and its CVaR does not exist. For this reason, to obtain an upper bound on $\|Z - \mathbb{E}[Z]\|_p$ for $p > 2$, we directly reinvest the martingale decomposition of $Z - \mathbb{E}[Z]$ used in the proofs of previous theorems, that we associate with Rosenthal inequalities for real-valued martingales. Those with best known constants are given by Pinelis [Pin15, Corollary 1]. His result holds for any $p > 2$, but for $p \in \,]2\,,4]$, the constants are close to optimality (see the discussion [Pin15, pages 701-702]) and are easy to express. We obtain the following inequality:

Theorem 2.7. — *Let Z be defined by* (1.3). *Let $p \in \,]2\,,4]$ and assume that $\|M_1\|_p < \infty$. Then*

$$\|Z - \mathbb{E}[Z]\|_p \leqslant (p-1)^{1/p}\sqrt{n}(\sigma + \sqrt{v_n}\,) + (2^{1/p} + 1)n^{1/p}\|M_1\|_p\,.$$

Let us now recall the moment inequality for suprema of empirical processes obtained by Boucheron, Bousquet, Lugosi and Massart [BBLM05] (see [BLM13, Theorems 15.14 and 15.5]): let $p > 2$ and define $\widetilde{Z} := \sup_{t\in\mathcal{T}} |\sum_{k=1}^n X_{k,t}|$. Then

$$\left\|(\widetilde{Z} - \mathbb{E}[\widetilde{Z}])_+\right\|_p \leqslant \sqrt{\kappa(p-1)}\sqrt{n}\,(\sigma + \Sigma) + \kappa(p-1)\left(\left\|\max_{k=1,\,\dots,\,n} M_k\right\|_p + \sup_{t\in\mathcal{T}}\|X_{1,t}\|_2\right), \tag{2.8}$$

where $\kappa := \sqrt{e}/(\sqrt{e} - 1)$ and $\Sigma^2 := \mathbb{E}[\sup_{t\in\mathcal{T}} n^{-1}\sum_{k=1}^n X_{k,t}^2]$. Let us now comment the differences with our result Theorem 2.7:

- In fact, inequality (2.8) does not require the identical distribution of the sequence $X_1, \dots, X_n$. However, contrary to Theorem 2.7, inequality (2.8) concerns centered empirical processes, that is $c \equiv 0$ in (1.3).
- The moment inequality (2.8) is only given for the positive part of $\widetilde{Z} - \mathbb{E}[\widetilde{Z}]$.
- To compare the constant in front of the variance factor, let us consider the upper bound that the authors provided on Σ^2 (see [BLM13, Theorem 11.17]):

$$n\Sigma^2 \leqslant n\sigma^2 + 32\left\|\max_{k=1,\,\dots,\,n} M_k\right\|_2 \mathbb{E}[\widetilde{Z}] + 8\left\|\max_{k=1,\,\dots,\,n} M_k\right\|_2^2. \tag{2.9}$$

Combining (2.9) with (2.8) leads to the following behavior:

$$\left\|(\widetilde{Z}-\mathbb{E}[\widetilde{Z}])_+\right\|_p \leqslant 2\kappa\sqrt{p-1}\,\sigma\sqrt{n}+o\left(\sqrt{n}\right). \tag{2.10}$$

Note that $B_p := 2\kappa\sqrt{p-1}$ is increasing in p, $B_2 \approx 3.1184$ and $B_4 \approx 5.5225$. By comparison, our constant $C_p := (p-1)^{1/p}$ is increasing on $[2\,,4]$, $C_2 = 1$ and $C_4 \approx 1.3161$.

- Let us now compare the variance factor by comparing v_n to $\Sigma^2 - \sigma^2$. To this end, we will use a version of Pisier lemma which can be found in Rio [Rio17b, Appendix D]. It states that for any $0 < r < p$,

$$\left\|\max_{k=1,\,\dots,\,n} M_k\right\|_r \leqslant \left(n\int_0^{1/n} Q_{M_1}^r(u)du\right)^{1/r} \leqslant \Lambda_p^+(M_1)\left(\frac{p}{p-r}\right)^{1/r} n^{1/p}. \tag{2.11}$$

Putting this upper bound in (2.9) yields that

$$\Sigma^2 - \sigma^2 \leqslant K_p\, n^{-\frac{1}{2}\left(1-\frac{2}{p}\right)} + o\left(n^{-\frac{1}{2}\left(1-\frac{2}{p}\right)}\right).$$

Now, see that

$$n^{-\frac{1}{2}\frac{p-2}{p-1}} = o\left(n^{-\frac{1}{2}\left(1-\frac{2}{p}\right)}\right).$$

Thus, recalling (2.5), one can say that $\Sigma^2 - \sigma^2$ is of a larger order than v_n.

2.3. Application to power-type tail

Let $Y_1, \dots, Y_n$ be a finite sequence of nonnegative iid random variables and let $X_1, \dots, X_n$ be a finite sequence of iid random variables with values in some measurable space $(\mathcal{X}, \mathcal{F})$ such that the two sequences are independent. Let P denote the common distribution of the X_k's. Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1\,,1]$ such that for all $f \in \mathscr{F}$,

$$P(f) = 0 \quad \text{and} \quad P(f^2) < \delta^2 \ \text{ for some } \delta \in\,]0\,,1[\,. \tag{2.12}$$

Let F be a measurable envelope function of $\mathscr{F}$, that is

$$|f| \leqslant F \text{ for any } f \in \mathscr{F}, \text{ and } F(x) \leqslant 1 \text{ for all } x \in \mathcal{X}\,. \tag{2.13}$$

We suppose furthermore that for some constant $\ell > 2$,

$$\mathbb{P}(Y_1 > t) \leqslant t^{-\ell} \quad \text{ for any } t > 0\,. \tag{2.14}$$

Define now

$$Z := \sup_{f\in\mathscr{F}} \sum_{k=1}^n Y_k\, f(X_k)\,. \tag{2.15}$$

We associate to each $f \in \mathscr{F}$ an unique index t, and we define $\mathcal{T}$ the set of all these indices. Set for all $k = 1, \dots, n$, $\widetilde{X}_{k,t} := Y_k f(X_k)$. Then, we have $Z = \sup_{t\in\mathcal{T}} \sum_{k=1}^n X_{k,t}$. This allows us to apply results of the previous section. The envelope of the collection of coordinates M_k is defined by $M_k := Y_k F(X_k)$.

Recalling Lemma 2.4, the remaining task to provide a useful bound on v_n is to upper bound the quantities E_k, $k = 1, \dots, n$. This will be done by using local maximal inequalities for empirical processes due to van der Vaart and Wellner [W11, Theorem 2.1]. The upper bound is expressed in terms of uniform entropy integral. Let us first recall some classical definitions.

Definition 2.8 (Covering number and uniform entropy integral). — *The covering number $N(\epsilon, \mathscr{G}, \nu)$ is the minimal number of balls of radius ϵ in $\mathbb{L}^2(\nu)$ needed to cover the set $\mathscr{G}$. The uniform entropy integral is defined by*

$$J(\delta, \mathscr{G}) := \int_0^\delta \sup_\nu \sqrt{1 + \log N\left(\epsilon \|F\|_{\nu,2}, \mathscr{F}, \nu\right)}\, d\epsilon\,.$$

Here, the supremum is taken over all finitely discrete probability distributions ν on $(\mathcal{X}, \mathcal{F})$ and $\|f\|_{\nu,2}$ denotes the norm of a function f in $\mathbb{L}^2(\nu)$.

Lemma 2.9. — *There exists a universal constant K such that for any integer $k \geqslant 1$,*

$$\frac{1}{k}\,\mathbb{E} \sup_{g\in\mathscr{G}} \left|\sum_{j=1}^k Y_j g(X_j)\right| \leqslant K\,\frac{\ell}{\ell-2}\left(k^{-1/2} J(\delta,\mathscr{G}) + \ell\, k^{-(1-1/\ell)}\left(\frac{J^2(\delta,\mathscr{G})}{\delta^2}\right)^{1-1/\ell}\right).$$

Remark 2.10. — For a numerical value of the (universal) constant K, we refer the reader to [GN16, Section 3.5.1] in the book of Giné and Nickl. Moreover, the terms $\ell/(\ell-2)$ and ℓ are certainly not optimal and could be improved. They are given for the sake of completeness.

The result of van der Vaart and Weller deals with bounded empirical processes. The proof of Lemma 2.9 is based on a truncation argument associated with their result. Chernozhukov, Chetverikov and Kato [CCK14, Theorem 5.2] have extended the result of van der Vaart and Wellner to unbounded empirical processes. In our setting, it provides the following upper bound:

$$\begin{aligned}(2.16)\quad \frac{1}{k}\,\mathbb{E} \sup_{g\in\mathscr{G}} &\left|\sum_{j=1}^k Y_j g(X_j)\right| \\ &\leqslant K\left(k^{-1/2}\sqrt{\frac{\ell}{\ell-2}} J(\delta,\mathcal{G}) + k^{-1}\frac{J^2(\delta,\mathcal{G})}{\delta^2}\left\|\max_{j=1,\dots,k} M_j\right\|_2\right),\end{aligned}$$

where K is a universal constant. Using now Pisier Lemma (2.11) to bound above $\|\max_{j=1,\dots,k} M_j\|_2$, (2.16) becomes

$$(2.17)\qquad \frac{1}{k}\,\mathbb{E} \sup_{g\in\mathscr{G}} \left|\sum_{j=1}^k Y_j g(X_j)\right| \leqslant K\sqrt{\frac{\ell}{\ell-2}}\left(k^{-1/2} J(\delta,\mathcal{G}) + k^{-(1-1/\ell)}\frac{J^2(\delta,\mathcal{G})}{\delta^2}\right).$$

Compared to Lemma 2.9, the constants in ℓ are better. Namely, in place of the terms $\ell/(\ell-2)$ and ℓ in Lemma 2.9, the right-hand side of (2.17) contains the smaller terms $\sqrt{\ell/(\ell-2)}$ and 1 respectively. However, the exponent of the term $J^2(\delta,\mathcal{G})/\delta^2$, which is $(1-1/\ell)$ in Lemma 2.9, leads to a better estimate as δ is small compared to the exponent 1 in (2.17). Indeed, if $\log N(\epsilon\|F\|_{\nu,2}, \mathscr{F}, \nu)$ is not larger than $H(1/\varepsilon)$ for some nondecreasing function H, independent of ν and satisfying some minor conditions, then $J^2(\delta,\mathcal{G})/\delta^2$ is of an order of $H(1/\delta)$ which then increases as δ tends to 0 (see, for instance, [GN16, Theorem 3.5.6]). We stress out that such an hypothesis on the covering number applies in many situations including VC-classes of functions. Nevertheless, the generality of Chernozhukov, Chetverikov and Kato's result makes it possible to upper bound the E_k's in other cases than the one considered here.

We now apply Theorem 2.7 combined with Lemmas 2.4 and 2.9.

Theorem 2.11. — *Let Z be defined by* (2.15).

(i) *Assume that $\ell \in\,]2\,,4]$ and $\|M_1\|_\ell \leqslant 1$. Then*

$$\text{(a)}\quad \|Z-\mathbb{E}[Z]\|_\ell \leqslant (\ell-1)^{1/\ell}\sqrt{n}\left(\sigma + K_\ell\left(n^{-\frac{1}{4}\frac{\ell-2}{\ell-1}}(J(\delta,\mathscr{G}))^{\frac{1}{2}\frac{\ell-2}{\ell-1}} + n^{-\frac{1}{2}(1-2/\ell)}\left(\frac{J(\delta,\mathscr{G})}{\delta}\right)^{1-2/\ell}\right)\right) + \left(2^{1/\ell}+1\right)n^{1/\ell}.$$

(ii) *Moreover, for any $\ell > 2$,*

$$\text{(b)}\quad \widetilde{\Lambda}^+_\ell(Z-\mathbb{E}[Z]) \leqslant \sqrt{(p/e)}\sqrt{n}\left(\sigma + K_\ell\left(n^{-\frac{1}{4}\frac{\ell-2}{\ell-1}}(J(\delta,\mathscr{G}))^{\frac{1}{2}\frac{\ell-2}{\ell-1}} + n^{-\frac{1}{2}(1-2/\ell)}\left(\frac{J(\delta,\mathscr{G})}{\delta}\right)^{1-2/\ell}\right)\right) + 3\,n^{1/\ell}\mu_\ell,$$

and

$$\text{(c)}\quad \widetilde{\Lambda}^+_p(\mathbb{E}[Z]-Z) \leqslant \sqrt{(p/e)}\sqrt{n}\left(\sigma + K_\ell\left(n^{-\frac{1}{4}\frac{\ell-2}{\ell-1}}(J(\delta,\mathscr{G}))^{\frac{1}{2}\frac{\ell-2}{\ell-1}} + n^{-\frac{1}{2}(1-2/\ell)}\left(\frac{J(\delta,\mathscr{G})}{\delta}\right)^{1-2/\ell}\right)\right) + n^{1/\ell}\mu_\ell,$$

where $\mu_\ell = 2+\max(4/3,\ell/3)$.

Let us compare with the bounded case: $Y_k \leqslant 1$ for all $k = 1, \ldots, n$. Integrating Rio's inequality for suprema of bounded empirical processes [Rio02, Theorem 1] and using Lemma 2.9, one obtains for any $\ell > 2$,

$$\|(Z-\mathbb{E}[Z])_+\|_\ell \leqslant C_\ell\sqrt{n}\sqrt{\sigma^2+2\,\mathbb{E}[Z]/n} \tag{2.18}$$

$$\leqslant C_\ell\sqrt{n}\left(\sigma + K_\ell\left(n^{-1/4}\sqrt{J(\delta,\mathscr{G})} + n^{-1/2}\,\frac{J(\delta,\mathscr{G})}{\delta}\right)\right). \tag{2.19}$$

Notice that the term

$$n^{-\frac{1}{4}\frac{\ell-2}{\ell-1}}(J(\delta,\mathscr{G}))^{\frac{1}{2}\frac{\ell-2}{\ell-1}} + n^{-\frac{1}{2}(1-2/\ell)}\left(\frac{J(\delta,\mathscr{G})}{\delta}\right)^{1-2/\ell}$$

in Theorem 2.11 tends to $n^{-1/4}\sqrt{J(\delta,\mathscr{G})} + n^{-1/2}\,J(\delta,\mathscr{G})/\delta$ as ℓ tends to ∞. Thus, we can see our results as extensions of (2.18) to the unbounded case.

3. Upper bounds on $\mathbb{E}[(Z-\mathbb{E}[Z]-t)_+^\alpha]$ and $Q_2(\mathbb{E}[Z]-Z;u)$

In this section, we provide a general upper bound on $\mathbb{E}[(Z-\mathbb{E}[Z]-t)_+^\alpha]$, where $t\in\mathbb{R}$, $\alpha\geqslant 1$, in terms of generalized quantiles. In order to explain the result, we start by considering the case $\alpha=1$ which is based on Theorem 2.1. We emphasize that it is of interest to obtain such bounds in various situations coming from statistical applications, such as study of rates of convergence for estimators (see, for instance, Comte and Lacour [CL13]).

3.1. Case $\alpha=1$

Let X be an integrable real-valued random variable and let $t\in\mathbb{R}$. Since $Q_X(U)$ and X have the same distribution for any random variable U uniformly distributed on $[0\,,1]$,

$$\mathbb{E}\left[(X-t)_+\right]=\int_0^1 (Q_X(s)-t)_+ds\,. \tag{3.1}$$

Now, recalling that $x<Q_X(u)$ if and only if $\mathbb{P}(X>x)>u$, we get

$$\int_0^1 (Q_X(s)-t)_+ds=\int_0^{\mathbb{P}(X>t)}(Q_X(s)-t)ds=\sup_{u\,\in\,]0,\,1]}\int_0^u (Q_X(s)-t)ds\,. \tag{3.2}$$

Note that the right-hand side of (3.2) is equal to $\sup_{u\,\in\,]0,1]} u(\widetilde{Q}_X(u)-t)$. Combining this fact with (3.1) leads to the following variational formula

$$\mathbb{E}[(X-t)_+]=\sup_{u\,\in\,]0,\,1]} u(\widetilde{Q}_X(u)-t)\,. \tag{3.3}$$

Thus, from the upper bound on $\widetilde{Q}_{Z-\mathbb{E}[Z]}$ given in Theorem 2.1 we derive the following:

Proposition 3.1. — *Let Z be defined by* (1.3)*. Let $p>2$ and $\mu_\ell=2+\max(4/3,p/3)$. Set also*

$$s_n:=\sqrt{n}\,(\sigma+\sqrt{v_n})\,,\ \text{and } b_{n,p}:=3\,n^{1/p}\mu_p\,\Lambda_p^+(M_1)\,.$$

Then, for any $t>0$,

$$\mathbb{E}\left[(Z-\mathbb{E}[Z]-t)_+\right]\leqslant s_n\,\frac{e^{-\frac12\left(1+t^2/s_n^2\right)}}{\sqrt{1+t^2/s_n^2}}+b_{n,p}\,.$$

Remark 3.2. — A similar upper bound on $\mathbb{E}[(\mathbb{E}[Z]-Z-t)_+]$ can be obtain by using Theorem 2.5 instead of Theorem 2.1.

The variational formula (3.3) gives a direct advantage of having an upper bound on the CVaR instead of on the log-Laplace transform of $Z-\mathbb{E}[Z]$, since the calculation we have to do is more straightforward. Indeed, to obtain an upper bound on $\mathbb{E}[(Z-\mathbb{E}[Z]-t)_+]$ from an upper bound on the log-Laplace transform, we first have to derive a deviation inequality by Markov's inequality and then to integrate it from t to $+\infty$.

3.2. Case $\alpha \geqslant 1$

In this section, we want to generalize the variational formula (3.3) to $\mathbb{E}[(X-t)_+^\alpha]$, $\alpha \geqslant 1$. To this end, we consider the generalized quantiles $Q_\alpha(X;u)$, $\alpha \geqslant 1$, introduced by Pinelis [Pin14b, Section 3]. In particular, note that $Q_1(X;u) = \widetilde{Q}_X(u)$. The general definition is somehow complicated, so we do not recall it. However, Pinelis proves the following formula (see [Pin14b, Theorem 3.3]):

$$Q_\alpha(X;u) = \inf_{s \in \mathbb{R}} \left\{ s + u^{-1/\alpha} \|(X-s)_+\|_\alpha \right\}. \tag{3.4}$$

For properties on these quantiles, the reader is referred to [Pin14b, Theorem 3.4]. We only recall that $Q_\alpha(X;u)$ is nondecreasing in α, subadditive in X, and $Q_\alpha(X;u) \leqslant \ell_X^{*-1}(\log(1/u))$ for all $\alpha \geqslant 1$. From Sion's minimax theorem, we derive the following variational formula which generalizes (3.3):

Lemma 3.3. — *Let $\alpha \geqslant 1$. Let X be an integrable real-valued random variable. Then*

$$\|(X-t)_+\|_\alpha = \sup_{u \in]0,1]} u^{1/\alpha} \left(Q_\alpha(X;u) - t \right).$$

The interest of these generalized quantiles Q_α in our setting is similar to that of the CVaR $\widetilde{Q}$. Indeed, our method lies in a decomposition of $Z - \mathbb{E}[Z]$ into a sum of two martingales. The subadditivity property allows to analyze each martingale separately. Moreover, some comparison inequalities on real-valued martingales directly give upper bounds on their quantiles Q_α: see, for instance, Bentkus [Ben04] for cases with $\alpha \in \{1,2,3\}$ or Pinelis [Pin06, Pin14a] for cases with $\alpha \in \{3,5\}$. In the next statement, we give an example of such bounds in the case $\alpha = 2$ for $\mathbb{E}[Z] - Z$ and under the additional assumption that the random variables $X_{1,t}$, $t \in \mathcal{T}$, are bounded from below.

Proposition 3.4. — *Assume that for all $t \in \mathcal{T}$, $X_{1,t} \geqslant -1$. Let Z be defined by (1.3). Let $\theta_1, \ldots, \theta_n$ be a sequence of iid two-valued centered random variables taking the values 1 and $-\sigma^2$. Set $B_n := \sum_{k=1}^n \theta_k$. Then for any $u \in]0,1]$,*

$$Q_2\left(\mathbb{E}[Z] - Z; u\right) \leqslant Q_2(B_n;u) + \sqrt{n}\sqrt{2v_n \log(1/u)}. \tag{a}$$

Consequently,

$$\|(\mathbb{E}[Z] - Z - t)_+\|_2 \leqslant \|(B_n - t)_+\|_2 + \sqrt{nv_n}\sqrt{(2/e)}. \tag{b}$$

Remark 3.5. — If $\sigma^2 \geqslant 1$, then (see, for instance, [BDR15, Lemma 2.36] in Bercu, Delyon and Rio)

$$Q_2(B_n;u) \leqslant \ell_{B_n}^{*-1}(\log(1/u)) \leqslant \sigma\sqrt{2n\log(1/u)}. \tag{3.5}$$

Thus, inequality (a) of Proposition 3.4 provides a sub-Gaussian bound. If $\sigma^2 < 1$, the inequality above does not hold. However, one has

$$Q_2(B_n;u) \leqslant \ell_{B_n}^{*-1}(\log(1/u)) \leqslant \sigma\sqrt{2n\log(1/u)} + \frac{1-\sigma^2}{3}\log(1/u). \tag{3.6}$$

We refer the reader to [BDR15, Theorem 2.28 and Exercise 6 in Section 2.9] for a proof.

4. Left deviation inequalities

As mentioned in Section 2, the left tails are easier to handle than the right tails. Notably, the left-hand side deviations of Z heavily depend on the behavior of the left tails of $X_{1,t}$, $t \in \mathcal{T}$. For example, in Theorem 2.1, the term $\Lambda_p^+(M_1)$ involves the behavior of right and left tails, while the term m_r in Theorem 2.5 only involves the behavior of the left tails. In this section, we give left deviation inequalities under the additional assumption on the $X_{k,t}$'s, of sub-Gamma tails on the left, and of sub-Gaussian tails on the left (our terminology follows [BLM13]).

4.1. Under sub-Gamma tails on the left assumption

Here we assume that $X_{1,t}$ admits a sub-Gamma tail on the left for all $t \in \mathcal{T}$. To the best of our knowledge, there exists only one exponential inequality for the left-hand side deviations of suprema of empirical processes in the unbounded case (from one side) which is due to Klein [Kle02]. More precisely, Klein provides an upper bound on the Laplace transform of $\mathbb{E}[Z] - Z$ which implies the following deviation inequality:

Theorem 4.1 ([Kle02, Theorem 1.1(3)]). — *Assume that $X_{1,t} \leqslant 1$ for all $t \in \mathcal{T}$, and for all integer $p \geqslant 2$, $|\mathbb{E}[X_{1,t}^p]| \leqslant \sigma^2 p!/2$. Then*

$$\mathbb{P}\left(Z < \mathbb{E}[Z] - \sqrt{2x\widetilde{V}_n} - x - \frac{x\sqrt{2\widetilde{V}_n}}{\sqrt{x} + \sqrt{\widetilde{V}_n/2}}\right) \leqslant \exp(-x),$$

where $\widetilde{V}_n = n\sigma^2 + 2\,\mathbb{E}[Z]$.

The objective here, is to relax the boundedness assumption on the right. Our main result is the following:

Theorem 4.2. — *Assume that there exists a positive constant c such that for any $t \in \mathcal{T}$ and any $\lambda \in]0, c[$,*

$$\log \mathbb{E}[\exp(-\lambda X_{1,t})] \leqslant \frac{\sigma^2\lambda^2}{2(1-c\lambda)}\,. \tag{4.1}$$

Let Z be defined by (1.3). Then for any $u \in]0,1[$,

$$\widetilde{Q}_{\mathbb{E}[Z]-Z}(u) \leqslant \sqrt{2n\log(1/u)}\,(\sigma + \sqrt{v_n}\,) + c\log(1/u)\,. \tag{a}$$

Consequently, for any $x > 0$,

$$\mathbb{P}\left(Z < \mathbb{E}[Z] - \sqrt{2nx}\,(\sigma + \sqrt{v_n}\,) - cx\right) \leqslant \exp(-x)\,. \tag{b}$$

Remark 4.3. — If the random variables $X_{1,t}$, $t \in \mathcal{T}$, satisfy Bernstein's moment conditions on the left, then the hypothesis (4.1) above is satisfied. Precisely, assume that there exists a positive constant c such that, for any integer $p \geqslant 3$, and all $t \in \mathcal{T}$,

$$\mathbb{E}\left[(-X_{1,t})_+^p\right] \leqslant \frac{p!c^{p-2}}{2}\sigma^2. \tag{4.2}$$

Then, since $\mathbb{E}[X_{1,t}] = 0$ for all $t \in \mathcal{T}$, $\log \mathbb{E}[\exp(-\lambda X_{1,t})] \leqslant \frac{\sigma^2\lambda^2}{2(1-c\lambda)}$. We refer the reader to Bercu, Delyon and Rio (see [BDR15, the proof of their Theorem 2.1]) for a proof.

4.2. Under sub-Gaussian tails on the left assumption

Here we assume that $X_{1,t}$ admits a sub-Gaussian tail on the left for all $t \in \mathcal{T}$. We show in the following theorem that $Z - \mathbb{E}[Z]$ is sub-Gaussian on the left.

Theorem 4.4. — *Assume that*

$$C(\mathcal{T}) := \sup_{\lambda > 0} \sup_{t \in \mathcal{T}} \frac{2}{\lambda^2} \log \mathbb{E}\left[\exp(-\lambda X_{1,t})\right] < \infty. \tag{4.3}$$

Let Z be defined by (1.3). *Then for any $u \in]0\,,1[$,*

$$\widetilde{Q}_{\mathbb{E}[Z]-Z}(u) \leqslant \sqrt{2n\log(1/u)}\left(\sqrt{C(\mathcal{T})} + \sqrt{v_n}\right). \tag{a}$$

Consequently, for any $x > 0$,

$$\mathbb{P}(Z < \mathbb{E}[Z] - x) \leqslant \exp\left(-\frac{x^2}{2n\left(\sqrt{C(\mathcal{T})} + \sqrt{v_n}\right)^2}\right). \tag{c}$$

If for all $t \in \mathcal{T}$, $X_{1,t}$ is a centered Gaussian random variable with variance equals to σ_t^2, then we can apply the above result with $C(\mathcal{T}) = \sup_{t \in \mathcal{T}} \sigma_t^2 =: \sigma^2$. However we can prove the more precise following result:

Proposition 4.5. — *Assume that for any $t \in \mathcal{T}$, $X_{1,t}$ is a centered Gaussian random variable. Let then $\sigma^2 := \sup_{t \in \mathcal{T}} \mathrm{Var}(X_{1,t})$. Let Y be a standard Gaussian random variable. Then for any $u \in]0\,,1[$,*

$$\widetilde{Q}_{\mathbb{E}[Z]-Z}(u) \leqslant \sigma\sqrt{n}\,\widetilde{Q}_Y(u) + \sqrt{2n\log(1/u)}\,\sqrt{v_n}$$

Note that, even if $X_1, \ldots, X_n$ are iid Gaussian vectors, this result is beyond the scope of Cirel'son, Ibragimov, and Sudakov's paper [CIS76] since $S_{n,t}$, $t \in \mathcal{T}$, is not a Gaussian process in general.

Example 4.6. — Let $Y_1, \ldots, Y_n$ be a finite sequence of nonnegative iid random variables and let $X_1, \ldots, X_n$ be a finite sequence of iid random variables with values in some measurable space $(\mathcal{X}, \mathcal{F})$ such that the two sequences are independent. Let $\mathcal{T}$ be a countable set, $(A_t)_{t \in \mathcal{T}}$ a family of Borel sets in $\mathcal{X}$ and $(\sigma_t)_{t \in \mathcal{T}}$ a family of positive reals such that $\sigma := \sup_{t \in \mathcal{T}} \sigma_t < \infty$. Then Proposition 4.5 applies to

$$Z := \sup_{t \in \mathcal{T}} \sum_{k=1}^{n} Y_k\left(2\,\mathbb{1}_{X_k \in A_t} - 1\right).$$

5. Proofs

5.1. Preliminaries

The starting point of the proofs is based on a martingale decomposition of Z which we now recall. We suppose that $\mathcal{T}$ is a finite class of functions, that is $\mathcal{T} = \{t_i : i \in \{1, \ldots, m\}\}$. The results in the countable case are derived from the finite case using the monotone convergence theorem. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all

$k = 1, \ldots, n$, $\mathcal{F}_k := \sigma(X_1, \ldots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also

$$Z_k := \mathbb{E}_k[Z], \tag{5.1}$$

$$Z^{(k)} := \sup \{S_{n,t} - X_{k,t} + nc_t : t \in \mathcal{T}\}. \tag{5.2}$$

The sequence (Z_k) is an $(\mathcal{F}_k)$-adapted martingale (the Doob martingale associated with $Z - \mathbb{E}[Z]$) and

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k, \quad \text{where } \Delta_k := Z_k - Z_{k-1}. \tag{5.3}$$

Define now the random indices τ and τ_k, respectively $\mathcal{F}_n$-measurable and $\mathcal{F}_n^k$-measurable, by

$$\tau := \inf \{i \in \{1, \ldots, m\} : S_{n,t_i} + nc_{t_i} = Z\}, \tag{5.4}$$

$$\tau_k := \inf \left\{i \in \{1, \ldots, m\} : S_{n,t_i} - X_{k,t_i} + nc_{t_i} = Z^{(k)}\right\}. \tag{5.5}$$

Notice first that

$$Z^{(k)} + X_{k,t_{\tau_k}} \leqslant Z \leqslant Z^{(k)} + X_{k,t_\tau}.$$

From this, conditioning by $\mathcal{F}_k$ gives

$$\mathbb{E}_k\left[X_{k,t_{\tau_k}}\right] \leqslant Z_k - \mathbb{E}_k\left[Z^{(k)}\right] \leqslant \mathbb{E}_k\left[X_{k,t_\tau}\right]. \tag{5.6}$$

Set now $\xi_k := \mathbb{E}_k[X_{k,t_{\tau_k}}]$ and let $\varepsilon_k \geqslant r_k \geqslant 0$ be random variables such that

$$\xi_k + r_k = Z_k - \mathbb{E}_k\left[Z^{(k)}\right] \quad \text{and} \quad \xi_k + \varepsilon_k = \mathbb{E}_k\left[X_{k,t_\tau}\right].$$

Thus (5.6) becomes

$$\xi_k \leqslant \xi_k + r_k \leqslant \xi_k + \varepsilon_k. \tag{5.7}$$

Since the random index τ_k is $\mathcal{F}_n^k$-measurable, we have by the centering assumption on the random variables $X_{k,t}$, $t \in \mathcal{T}$,

$$\mathbb{E}_n^k\left[X_{k,t_{\tau_k}}\right] = 0, \tag{5.8}$$

which ensures that $\mathbb{E}_{k-1}[\xi_k] = 0$. Moreover, $\mathbb{E}_k[Z^{(k)}]$ is $\mathcal{F}_{k-1}$-measurable. Hence we get

$$\Delta_k = Z_k - \mathbb{E}_k\left[Z^{(k)}\right] - \mathbb{E}_{k-1}\left[Z_k - \mathbb{E}_k\left[Z^{(k)}\right]\right] = \xi_k + r_k - \mathbb{E}_{k-1}[r_k],$$

which, combined with (5.3), yields the decomposition of $Z - \mathbb{E}[Z]$ into a sum of two martingales:

$$Z - \mathbb{E}[Z] = \Xi_n + R_n, \tag{5.9}$$

where

$$\Xi_n := \sum_{k=1}^{n} \xi_k \ \text{ and } \ R_n := \sum_{k=1}^{n} \left(r_k - \mathbb{E}_{k-1}[r_k]\right). \tag{5.10}$$

The strategy of the proofs is to treat the two martingales separately. And the main difficulty lies in the control of their quadratic variations, especially that of R_n.

(i) Upper bound on $\sum_{k=1}^{n} \mathbb{E}_{k-1}[\xi_k^2]$.

Lemma 5.1. — *One has*

$$\langle \Xi \rangle_n := \sum_{k=1}^{n} \mathbb{E}_{k-1}\left[\xi_k^2\right] \leqslant n\sigma^2.$$

Proof. — First, notice that the same argument as in (5.8) yields

$$\mathbb{E}_n^k[X^2_{k,\,t_{\tau_k}}] \leqslant \sigma^2.$$

Thus, Lemma 5.1 follows from the conditional Jensen inequality. □

(ii) Upper bound on $\sum_{k=1}^n \mathbb{E}_{k-1}[(r_k - \mathbb{E}_{k-1}[r_k])^2]$.

We recall that ζ_k denotes a random variable with distribution function $F_{2\,M_1,\,q_k}$ (defined in Definition 1.6) where q_k is such that $\mathbb{E}[\zeta_k] = E_k$.

Lemma 5.2. — *One has*

$$\langle R \rangle_n := \sum_{k=1}^{n} \mathbb{E}_{k-1}\left[(r_k - \mathbb{E}_{k-1}[r_k])^2\right] \leqslant \sum_{k=1}^{n} \mathbb{E}\left[\zeta_k^2\right].$$

Proof. — First, we observe that $\mathbb{E}_{k-1}[r_k]$ is bounded by a deterministic constant. This will be a consequence of the following lemma of exchangeability of variables.

Lemma 5.3. — *For any integer $j \geqslant k$,*

$$\mathbb{E}_{k-1}\left[X_{k,\,\tau}\right] = \mathbb{E}_{k-1}\left[X_{j,\,\tau}\right].$$

Proof of Lemma 5.3. — Pointing out that τ is a function of $X_1, \ldots, X_n$, one has for every permutation π on n elements,

$$\tau(X_1, \ldots, X_n) = \tau \circ \pi(X_1, \ldots, X_n) \text{ almost surely,}$$

leading to

$$\mathbb{E}_n^k[X_{k,\,\tau}] = \mathbb{E}_n^k[X_{k,\,\tau\circ\pi}].$$

This implies

$$\mathbb{E}_{k-1}[X_{k,\,\tau}] = \mathbb{E}_{k-1}[X_{k,\,\tau\circ\pi}] \quad \text{for any } k = 1, \ldots, n.$$

Taking now $j \geqslant k$ and applying the previous equality to the transposition $\pi := (k\,j)$ which exchanges k and j, it yields by Fubini's theorem that

$$\mathbb{E}_{k-1}\left[X_{k,\,\tau}\right] = \mathbb{E}_{k-1}\left[X_{j,\,\tau}\right],$$

which concludes the proof. □

Therefrom,

$$\begin{aligned} \mathbb{E}_{k-1}[\varepsilon_k] &= \mathbb{E}_{k-1}[X_{k,\,\tau}] \\ &= \frac{1}{n-k+1}\mathbb{E}_{k-1}\left[X_{k,\tau} + \ldots + X_{n,\,\tau}\right] \\ &\leqslant \frac{1}{n-k+1}\,\mathbb{E}_{k-1}\sup_{t\in\mathcal{T}}\{X_{k,\,t} + \ldots + X_{n,\,t}\} = E_{n-k+1}. \end{aligned} \tag{5.11}$$

Since $0 \leqslant r_k \leqslant \varepsilon_k$, we thus get

$$0 \leqslant \mathbb{E}_{k-1}[r_k] \leqslant E_{n-k+1}.$$

Next, (5.7) implies that $0 \leqslant r_k \leqslant 2M_k$. Therefore, the comparison inequality (1.19) ensures that for any function $\varphi \in \mathcal{H}_+^1$, $\mathbb{E}_{k-1}[\varphi(r_k)] \leqslant \mathbb{E}[\varphi(\zeta_{n-k+1})]$. Since $r_k \geqslant 0$ and $x \mapsto x_+^2$ belongs to $\mathcal{H}_+^1$, we get

$$\mathbb{E}_{k-1}\left[(r_k - \mathbb{E}_{k-1}[r_k])^2\right] \leqslant \mathbb{E}_{k-1}\left[r_k^2\right] = \mathbb{E}_{k-1}\left[r_{k+}^2\right] \leqslant \mathbb{E}\left[\zeta_{n-k+1}^2\right].$$

Finally, summing for $k = 1, \ldots, n$ leads to

$$\sum_{k=1}^{n} \mathbb{E}_{k-1}\left[(r_k - \mathbb{E}_{k-1}[r_k])^2\right] \leqslant \sum_{k=1}^{n} \mathbb{E}\left[\zeta_k^2\right] = V_n, \tag{5.12}$$

which ends the proof of Lemma 5.2. □

Before proving the main results, let us explain why left deviation inequalities are easier to handle than the right ones.

5.2. Upper bound on $\widetilde{Q}_{-R_n}$

To study the left deviations, we write

$$\mathbb{E}[Z] - Z = \Xi_n^o + R_n^o, \tag{5.13}$$

where $\Xi_n^o = -\Xi_n$ and $R_n^o = -R_n$. We will control the martingales Ξ_n and Ξ_n^o in the same manner. For R_n^o, the above analysis implies that R_n^o is a martingale with bounded from above increments and then it is a sub-Gaussian martingale as shown by the lemma below.

Lemma 5.4. — *For any $t \geqslant 0$, we have*

$$\ell_{R_n^o}(t) := \log \mathbb{E}\left[\exp\left(tR_n^o\right)\right] \leqslant \frac{t^2 V_n}{2}. \tag{a}$$

Consequently, for any $u \in]0\,,1[$,

$$\widetilde{Q}_{R_n^o}(u) \leqslant \ell_{R_n^o}^{*-1}\left(\log(1/u)\right) \leqslant \sqrt{2n\, v_n \log(1/u)}, \tag{b}$$

where $v_n = V_n/n$.

Proof. — We recall that

$$\mathbb{E}_{k-1}[r_k] - r_k \leqslant E_{n-k+1} \text{ and } \mathbb{E}_{k-1}\left[(\mathbb{E}_{k-1}[r_k] - r_k)^2\right] \leqslant \mathbb{E}\left[\zeta_{n-k+1}^2\right]. \tag{5.14}$$

Thus, $\max(\mathbb{E}[\zeta_k^2], E_k^2) = \mathbb{E}[\zeta_k^2]$. Applying now the inequality for real-valued martingales with differences bounded from above proved by Bentkus [Ben03] (see his inequality (2.16)), we get for any $t \geqslant 0$,

$$\mathbb{E}\left[\exp\left(tR_n^o\right)\right] \leqslant \exp\left(t^2 V_n/2\right), \tag{5.15}$$

which then gives (a). Now, it is a classical calculation that

$$\inf_{t>0}\left\{\frac{1}{t}\left(\frac{t^2 V_n}{2} + x\right)\right\} = \sqrt{2\, V_n\, x} = \sqrt{2n\, v_n\, x}, \tag{5.16}$$

where the infimum is given by the optimal value $t_x = \sqrt{2\,x/V_n}$. Finally, (1.10) ends the proof of (b). □

5.3. Proofs of Section 2

First, we recall some results concerning real-valued martingales. Let $M_n := \sum_{k=1}^n X_k$ be a martingale in $\mathbb{L}^2$ with respect to a nondecreasing filtration $(\mathcal{F}_k)$, such that $M_0 = 0$ and $\mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \leqslant b_k^2$ almost surely for all $k = 1, \dots, n$, where b_k's are some positive reals. Define

$$B_n = \sqrt{\sum_{k=1}^n b_k^2}\,.$$

We start by a Fuk–Nagaev type inequality obtained by Rio [Rio17a]:

Theorem 5.5 ([Rio17a, Theorem 4.1]). — *Let $r > 2$ such that*

$$\left\| \sup_{t>0} \left(t^r\, \mathbb{P}(X_{k+} > t | \mathcal{F}_{k-1}) \right) \right\|_\infty < \infty.$$

Define

$$C_r^w(M) = \left\| \sup_{t>0} \left(t^r \sum_{k=1}^n \mathbb{P}\left(X_{k+} > t | \mathcal{F}_{k-1}\right) \right) \right\|_\infty^{1/r}.$$

Then for any $u \in\,]0\,,1[$,

$$\widetilde{Q}_{M_n}(u) \leqslant \sigma\sqrt{2\log(1/u)} + C_r^w(M)\mu_r u^{-1/r},$$

where $\mu_r := 2 + \max(4/3, r/3)$.

Next, we recall the following Rosenthal-type inequality obtain by Pinelis [Pin15]:

Theorem 5.6 ([Pin15, Theorem 1 and inequality (10)]). — *Let $r \in\,]2\,,4]$. Then*

$$\left\|M_n\right\|_r \leqslant (r-1)^{1/r} B_n + \left(\sum_{k=1}^n \mathbb{E}\left[|X_k|^r\right] \right)^{1/r}.$$

We are now in a position to prove the main results.

Proof of Theorem 2.1. — First, observe that (b) follows immediately from (a) since for any real-valued random variable X and any $u \in\,]0\,,1[$, one has $Q_X \leqslant \widetilde{Q}_X$ and $\mathbb{P}(X > Q_X(u)) \leqslant u$. Let us now prove (a). Recalling the decomposition (5.9), Proposition 1.4 implies

$$\widetilde{Q}_{Z-\mathbb{E}[Z]}(u) \leqslant \widetilde{Q}_{\Xi_n}(u) + \widetilde{Q}_{R_n}(u). \tag{5.17}$$

With the notation of Rio's Theorem, since $\xi_k \leqslant M_k$ and $r_k - \mathbb{E}_{k-1}[r_k] \leqslant 2M_k$, one has

$$C_\ell^w(\Xi_n) \leqslant n^{1/\ell}\Lambda_\ell^+(M_1), \tag{5.18}$$

$$C_\ell^w(R_n) \leqslant 2\, n^{1/\ell}\Lambda_\ell^+(M_1). \tag{5.19}$$

Recalling the bounds on the quadratic variations of Ξ_n and R_n given by Lemmas 5.1-5.2, we then conclude the proof of (a) by combining (5.17)–(5.19) and Theorem 5.5. □

Proof of Lemma 2.4. — By homogeneity, we may assume that $\Lambda_p^+(M_1)=1$. Let ψ be a random variable with tail function defined by $\mathbb{P}(\psi>t)=t^{-\ell}$ for all $t\geqslant 1$ and let $\widetilde{\zeta}_k$ be a random variable with distribution function $F_{2\psi,\tilde{q}_k}$, where $\tilde{q}_k$ is the real in $[0,1]$ such that $\mathbb{E}[\widetilde{\zeta}_k]=E_k$. Clearly,

$$F_{2M_1,\,q_k}(x)\geqslant F_{2\psi,\,\tilde{q}_k}(x) \text{ for any } x\in\mathbb{R}.$$

In other words, $\widetilde{\zeta}_k$ dominates ζ_k for the first-order stochastic dominance. Then [Ben08, Lemma 1] of Bentkus ensures that for any convex function φ,

$$\mathbb{E}[\varphi(\zeta_k)]\leqslant\mathbb{E}[\varphi(\widetilde{\zeta}_k)]\,. \tag{5.20}$$

Thus, applying this inequality to $\varphi(x)=x^2$,

$$\mathbb{E}\left[\zeta_k^2\right]\leqslant\mathbb{E}\left[\widetilde{\zeta}_k^2\right]. \tag{5.21}$$

Moreover, a straightforward calculation yields that

$$\mathbb{E}\left[\widetilde{\zeta}_k^2\right]\leqslant 2^{\frac{p}{p-1}}\,\frac{p}{p-2}\left(1-\frac{1}{p}\right)^{(p-2)/(p-1)}E_k^{(p-2)/(p-1)}. \tag{5.22}$$

The claim follows by summing for $k=1,\,\dots,\,n$. □

Proof of Theorem 2.5. — As for the right deviations, we only have to prove (a). We recall that $\mathbb{E}[Z]-Z=\Xi_n^o+R_n^o$. We shall use Rio's inequality again to control Ξ_n^o and the Gaussian upper bound Lemma 5.4 to control R_n^o. Contrary to the previous proof, since we do not use Rio's inequality on R_n^o, it is interesting to provide a better upper bound on $C_r^w(\Xi_n^o)$. First, by Markov's inequality,

$$C_r^w(\Xi_n^o)\leqslant\left\|\sum_{k=1}^n\mathbb{E}_{k-1}\left[(-\xi_k)_+^r\right]\right\|_\infty^{1/r}. \tag{5.23}$$

Furthermore, using the same trick as in (5.8), we obtain

$$\mathbb{E}_{k-1}\left[(-\xi_k)_+^r\right]\leqslant m_r^r\,. \tag{5.24}$$

Combining now (5.23)-(5.24), Theorem 5.5 and Lemma 5.4 ends the proof. □

Proof of Corollary 2.6. — First, observe that (a) and (c) follow directly from (2.7). Let us now prove (b). We proceed exactly as in Rio [Rio17a, Theorem 5.1]. Both (2.6) and Theorem 2.1 (a) imply

$$\widetilde{\Lambda}_\ell^+(Z-\mathbb{E}[Z])\leqslant\left(\sigma\sqrt{n}+\sqrt{V_n}\right)\sup_{u\,\in\,]0,\,1]}\left(u^{1/\ell}\sqrt{2\log(1/u)}\right)+3\,n^{1/\ell}\mu_\ell\,\Lambda_\ell^+\left(\Phi(X_1)\right). \tag{5.25}$$

Next, observe that $\sup_{u\,\in\,]0,\,1]}u^{1/\ell}\sqrt{2\log(1/u)}=\sqrt{(\ell/e)}$, which concludes the proof of (b). The same is done for the proof of (d) by using Theorem 2.5 (a) instead of Theorem 2.1 (a). □

Proof of Theorem 2.7. — First, we apply Theorem 5.6 to the martingale Ξ_n. Then, it follows from Lemma 5.1 and the definition of M_k that

$$\|\Xi_n\|_p\leqslant(p-1)^{1/p}\sigma\sqrt{n}+n^{1/p}\|M_1\|_p\,. \tag{5.26}$$

Next, we do the same for the martingale R_n, using Lemma 5.2 instead of Lemma 5.1. We only need to handle $\mathbb{E}[|r_k - \mathbb{E}_{k-1}[r_k]|^p]$. It is done by the following:

Lemma 5.7. — $\mathbb{E}[|r_k - \mathbb{E}_{k-1}[r_k]|^p] \leqslant \mathbb{E}[|r_k|^p]$.

This result only uses the fact that r_k is a nonnegative random variable. It is a probably known fact, but for sake of completeness, we give a proof in Appendix A. Now, since $0 \leqslant r_k \leqslant 2\, M_k$, we deduce that

$$\|R_n\|_p \leqslant (p-1)^{1/p}\sqrt{nv_n} + 2^{1/p} n^{1/p} \|M_1\|_p \,. \tag{5.27}$$

The claim follows by combining (5.26) and (5.27). □

Proof of Lemma 2.9. — As in the proof of Lemma 2.4, ψ denotes a random variable such that $\mathbb{P}(X > t) = t^{-\ell}$ for all $t \geqslant 1$. Notice that $Q_\psi(u) = u^{-1/\ell}$ for any $u \in \,]0\,,1[$ and that (2.13) implies $Q_Y \leqslant Q_\psi$. Let also $\kappa \in \mathbb{R}$ be such that

$$2^\kappa = \frac{k\,\delta^2}{J^2(\delta, \mathscr{G})} \,. \tag{5.28}$$

Let $U_1, \ldots, U_k$ be k independent copies of a random variable U distributed uniformly on $[0\,,1]$. Let us now define for every $j = 1, \ldots, \lceil \kappa \rceil$,

$$\begin{aligned} I_j &:= \Big\{ m \in \{1, \ldots, k\} : U_m \in \,]2^{-j}\,, 2^{1-j}] \Big\}\,, \\ J_\kappa &:= \Big\{ m \in \{1, \ldots, k\} : U_m \leqslant 2^{-\lceil \kappa \rceil} \Big\}\,. \end{aligned}$$

Here, $\lfloor . \rfloor$ and $\lceil . \rceil$ denote the classical floor and ceiling functions. Recall that $Q_X(U)$ and X have the same distribution. Then,

$$\mathbb{E} \sup_{g \in \mathscr{G}} \left| \sum_{j=1}^{k} Y_j g(X_j) \right| \leqslant \mathbf{E}_1 + \mathbf{E}_2, \tag{5.29}$$

where

$$\mathbf{E}_1 := \sum_{j=1}^{\lceil \kappa \rceil} \mathbb{E} \sup_{g \in \mathscr{G}} \left| \sum_{i \in I_j} Q_\psi(U_i) g(X_i) \right| \quad \text{and} \quad \mathbf{E}_2 := \mathbb{E} \sup_{g \in \mathscr{G}} \left| \sum_{j \in J_\kappa} Q_\psi(U_j) g(X_j) \right| .$$

Let us bound above $\mathbf{E}_2$. Since $G \leqslant 1$, a straightforward calculation gives

$$\mathbf{E}_2 \leqslant k \int_0^{2^{-\lceil \kappa \rceil}} Q_\psi(u)\, du \leqslant k\, \frac{\ell}{\ell - 1}\, 2^{-\kappa(1-1/\ell)}. \tag{5.30}$$

To bound above $\mathbf{E}_1$, we first notice that, since Q_ψ is decreasing, for any $m \in I_j$, $|Y_m g(X_m)| \leqslant Q_\psi(2^{-j})$. We can then apply [W11, Theorem 2.1] of Van der Vaart and Wellner which leads to

$$\mathbf{E}_1 \leqslant K \left(J(\delta, \mathscr{G}) \sum_{j=1}^{\lceil \kappa \rceil} \mathbb{E}\left[|I_j|^{\frac{1}{2}} \right] Q_\psi(2^{-j}) + \frac{J^2(\delta, \mathscr{G})}{\delta^2} \sum_{j=1}^{\lceil \kappa \rceil} Q_\psi(2^{-j}) \right) . \tag{5.31}$$

By the definition of I_j, it is easy to see that

$$\mathbb{E}\left[|I_j|\right] = \sum_{i=1}^{k} i \binom{k}{i} (2^{-j})^i (1 - 2^{-j})^{k-i} = k\, 2^{-j}.$$

Then, Jensen's inequality yields $\mathbb{E}[|I_j|^{\frac{1}{2}}] \leqslant \sqrt{k\,2^{-j}}$. Now, since $Q_\psi(u) = u^{-1/\ell}$,

$$\sum_{j=1}^{\lceil\kappa\rceil} 2^{-j/2} Q_\psi(2^{-j}) \leqslant \frac{2^{1/\ell-1/2}}{1-2^{1/\ell-1/2}} \leqslant \frac{2}{\log(2)}\,\frac{\ell}{\ell-2}\,. \tag{5.32}$$

Likewise,

$$\begin{aligned}\sum_{j=1}^{\lceil\kappa\rceil} Q_\psi(2^{-j}) &= 2^{\lfloor\kappa\rfloor/\ell}\left(2^{1/\ell} + \sum_{j=0}^{\lfloor\kappa\rfloor-1} 2^{-j/\ell}\right)\\ &\leqslant 2^{\lfloor\kappa\rfloor/\ell}\left(2^{1/\ell} + \frac{1}{1-2^{-1/\ell}}\right) \leqslant \frac{2^{\kappa/\ell}}{\log(2)}\,\frac{\ell^2}{\ell-2}\,.\end{aligned} \tag{5.33}$$

Hence, we derive from (5.31)–(5.33),

$$\mathbf{E}_1 \leqslant K\,\frac{\ell}{\ell-2}\left(\sqrt{k}\,J(\delta,\mathscr{G}) + \ell\,\frac{J^2(\delta,\mathscr{G})}{\delta^2}\,2^{\kappa/\ell}\right). \tag{5.34}$$

Finally, (5.29), (5.34), (5.30) and the definition of κ imply Lemma 2.9. □

Proof of Theorem 2.11. — Inequality (a) follows from Theorem 2.7, Lemmas 2.4-2.9 and the subadditivity of the functions $x \mapsto x^a$, for $0 < a < 1$. Similarly, (b) and (c) follow by using Corollary 2.6 instead of Theorem 2.7 and the fact that $\Lambda_\ell^+(Y_1 G(X_1)) \leqslant \Lambda_\ell^+(\psi) = 1$. □

5.4. Proof of Section 3

Proof of Proposition 3.1. — Inequality (3.3) and Theorem 2.1 (a) imply

$$\begin{aligned}\mathbb{E}[(Z-\mathbb{E}[Z]-t)_+] &\leqslant \sup_{u\in]0,1]} u\left(s_n\sqrt{2\log(1/u)} + b_{n,p}\,u^{-1/p} - t\right)\\ &\leqslant \sup_{u\in]0,1]} u\left(s_n\sqrt{2\log(1/u)} - t\right) + b_{n,p},\end{aligned} \tag{5.35}$$

since $u^{1-1/p} \leqslant 1$. With the change of variables $y = \sqrt{2\log(1/u)} \in [0\,,+\infty[$, the supremum is achieved at

$$y_0 := \frac{t}{2\,s_n} + \sqrt{1+\frac{t^2}{4\,s_n^2}}\,. \tag{5.36}$$

Then, the supremum in (5.35) is equal to $s_n\,e^{-y_0^2/2}/y_0$. Observing now that $y_0 \geqslant \sqrt{1+t^2/s_n^2}$, we finally get the desired inequality which concludes the proof. □

Proof of Lemma 3.3. — Define $f : [0\,,1]\times\mathbb{R} \to \mathbb{R}$ by

$$f(u,s) := u^{1/\alpha}(s-t) + \|(X-s)_+\|_\alpha\,.$$

Note that, since $\alpha \geqslant 1$, for any $s\in\mathbb{R}$, $u \mapsto f(u,s)$ is concave and for any $u\in[0\,,1]$, $s \mapsto f(u,s)$ is convex. Thus, we can apply Sion's minimax theorem [Sio58] which implies that

$$\sup_{u\in]0,1]}\inf_{s\in\mathbb{R}} f(u,s) = \inf_{s\in\mathbb{R}}\sup_{u\in]0,1]}\,. \tag{5.37}$$

For any $s < t$,

$$\|(X-t)_+\|_\alpha \leqslant \|(X-s)_+\|_\alpha = \sup_{u\in]0,1]} f(u,s). \tag{5.38}$$

Moreover, for any $s > t$,

$$\begin{aligned}\|(X-t)_+\|_\alpha &= \|((X-s)+(s-t))_+\|_\alpha \\ &\leqslant \|(X-s)_+\|_\alpha + (s-t) \\ &= \sup_{u\in]0,1]} f(u,s).\end{aligned} \tag{5.39}$$

Thus, (5.38) and (5.39) give

$$\inf_{s\in\mathbb{R}} \sup_{u\in]0,1]} f(u,s) = \|(X-t)_+\|_\alpha\,. \tag{5.40}$$

Finally, we derive from (3.4) that

$$\sup_{u\in]0,1]} (Q_\alpha(X;u)-t) = \sup_{u\in]0,1]} \inf_{s\in\mathbb{R}} f(u,s) = \sup_{u\in]0,1]} \inf_{s\in\mathbb{R}} f(u,s). \tag{5.41}$$

The claim follows by combining (5.41), (5.37) and (5.40) □

Proof of Proposition 3.4. — Inequality (b) directly follows from Lemma 3.3 and (a). Let us prove (a). Recalling the decomposition (5.13), the subadditivity property of the quantile Q_2 yields that

$$Q_2\left(\mathbb{E}[Z]-Z;u\right) \leqslant Q_2\left(\Xi_n^o;u\right) + Q_2\left(R_n^o;u\right). \tag{5.42}$$

Moreover, from the properties of Q_2 and Lemma 5.4,

$$Q_2\left(R_n^o;u\right) \leqslant \ell_{R_n^o}^{*-1}(\log(1/u)) \leqslant \sqrt{n}\sqrt{2v_n\log(1/u)}. \tag{5.43}$$

Note that $-\xi_k = -\mathbb{E}_k[X_{k,t_{\tau_k}}] \leqslant 1$ since we assume that $X_{1,t} \geqslant -1$ for all $t\in\mathcal{T}$. Thus, Ξ_n^o is a martingale with increments $(-\xi_k)$ satisfying

$$-\xi_k \leqslant 1 \quad \text{and} \quad \mathrm{Var}_{k-1}(-\xi_k) \leqslant \sigma^2. \tag{5.44}$$

Then, one has the following comparison inequality (see, for instance, Bentkus[Ben04, Lemma 4.4]),

$$\mathbb{E}\left[(\Xi_n^o-t)_+^2\right] \leqslant \mathbb{E}\left[(B_n-t)_+^2\right] \quad \text{for any } t\in\mathbb{R}. \tag{5.45}$$

Combined with (3.4), it yields for any $u\in]0,1]$,

$$Q_2(\Xi_n^o;u) \leqslant Q_2(B_n;u). \tag{5.46}$$

Finally, inequality (a) follows from (5.42), (5.43) and (5.46). □

5.5. Proofs of Section 4

Proof of Theorem 4.2. — As previously, we only have to prove (a). Recall that

$$\langle\Xi^o\rangle_n := \sum_{k=1}^n \mathbb{E}_{k-1}\left[(-\xi_k)^2\right] \leqslant n\sigma^2. \tag{5.47}$$

Furthermore, as in (5.8), since τ_k is $\mathcal{F}_n^k$-measurable, the conditional Jensen inequality implies that for any $\lambda \geqslant 0$,

$$\mathbb{E}_{k-1}\left[\exp(\lambda\xi_k^o)\right] \leqslant \mathbb{E}_{k-1}\mathbb{E}_n^k\left[\exp\left(-\lambda X_{k,\,t_{\tau_k}}\right)\right] \leqslant \frac{\sigma^2\lambda^2}{2(1-c\lambda)}, \tag{5.48}$$

where we use the assumption (4.1) in the last inequality. An immediate induction on n gives that

$$\log \mathbb{E}\left[\exp(\lambda\,\Xi_n^o)\right] \leqslant n\,\frac{\sigma^2\lambda^2}{2(1-c\lambda)}\,. \tag{5.49}$$

Now, it is a classical calculation that

$$\inf_{\lambda\,\in\,]0,\,1/c[}\left(\frac{\sigma^2 t}{2(1-\lambda c)}+\frac{x}{\lambda}\right) = cx+\sqrt{2x\sigma^2}, \tag{5.50}$$

where the infimum is given by the optimal value $\lambda_x = \sqrt{2x}/(\sqrt{\sigma^2}+c\sqrt{2x})$. Recalling (1.10), one concludes that for any $u \in\,]0\,,1[$,

$$\widetilde{Q}_{\Xi_n}(u) \leqslant c\log(1/u)+\sigma\sqrt{2n\log(1/u)}. \tag{5.51}$$

Finally, combining Lemma 5.4, (5.51) and the subadditivity property of $\widetilde{Q}$ implies inequality (a) of the theorem and completes the proof of Theorem 4.2. ☐

Proof of Theorem 4.4. — As previously mentioned, we only have to prove (a). By reasoning in the same way as (5.48), the assumption on the $X_{k,\,t}$'s allows us to derive that

$$\log \mathbb{E}[\exp(\lambda\,\Xi_n^o)] \leqslant n\frac{\lambda^2}{2}C(\mathcal{T}), \tag{5.52}$$

for any $\lambda \geqslant 0$. Therefore, the same conclusion as in the proof of Theorem 4.2 yields that

$$\widetilde{Q}_{\Xi_n^o}(u) \leqslant \sqrt{2nC(\mathcal{T})\log(1/u)}, \tag{5.53}$$

for any $u \in\,]0\,,1[$. The claim follows by associating this fact with Lemma 5.4. ☐

Proof of Proposition 4.5. — Let φ be a convex function. Then by Jensen's inequality,

$$\mathbb{E}_{k-1}\left[\varphi(-\zeta_k)\right] \leqslant \mathbb{E}_{k-1}\left[\varphi\left(-X_{k,\,t_{\tau_k}}\right)\right].$$

Now, since τ_k is $\mathcal{F}_n^k$-measurable, conditionally to $\mathcal{F}_n^k$, $-X_{k,\,t_{\tau_k}}$ is a centered Gaussian random variable with variance equals to $\sigma^2_{t_{\tau_k}}$. Thus, since $\sigma^2_{t_{\tau_k}} \leqslant \sigma^2$,

$$\mathbb{E}_{k-1}\mathbb{E}_n^k\left[\varphi\left(-X_{k,\,t_{\tau_k}}\right)\right] \leqslant \mathbb{E}[\varphi(\sigma Y)], \tag{5.54}$$

where Y is standard Gaussian random variable. The proof of (5.54) is deferred to Appendix A. Now, by an induction on n, we derive that for any convex function φ,

$$\mathbb{E}\left[\varphi\left(\Xi_n^o\right)\right] \leqslant \mathbb{E}\left[\varphi\left(\sigma\sum_{k=1}^n Y_k\right)\right], \tag{5.55}$$

where $Y_1, \ldots, Y_n$ is a sequence of iid standard Gaussian random variables, independent of the other random variables. Then, by the variational formula (3.4) for $\alpha = 1$, we have for any $u \in]0\,,1]$,

$$\widetilde{Q}_{\Xi_n^o}(u) \leqslant \widetilde{Q}_{\sigma \sum_{k=1}^{n} Y_k}(u). \tag{5.56}$$

Now, since $\widetilde{Q}_X$ depends only on the distribution of X, the right-hand side in (5.56) is equal to $\sigma\sqrt{n}\,\widetilde{Q}_Y(u)$, where Y is a standard Gaussian random variable. The claim follows by combining this fact with Lemma 5.4. □

Appendix A. Additional proofs

Proof of Inequality (1.12). — Let $X = Z_G - \mathbb{E}[Z_G]$. Recall that $Q_X(U)$ has the same distribution as X for any random variable U with the uniform distribution over $[0\,,1]$. It implies that $\widetilde{Q}_X(1) = \mathbb{E}[X]$. Now, since $\mathbb{E}[X] = 0$, one has

$$\int_0^u Q_X(s)ds = -\int_u^1 Q_X(s)ds.$$

Moreover,

$$\int_u^1 Q_X(y)dy = \frac{1-u}{u}\int_0^u Q_X\left(u + \frac{1-u}{u}s\right)ds.$$

Hence, combining this two facts and using the comparison inequality (1.11),

$$\begin{aligned}
\widetilde{Q}_X(u) &= \frac{1-u}{u}\int_0^u Q_X(s)ds + \int_0^u Q_X(s)ds \\
&= \frac{1-u}{u}\int_0^u \left(Q_X(s) - Q_X\left(u + \frac{1-u}{u}s\right)\right)ds \\
&\leqslant \sigma_G\,\frac{1-u}{u}\int_0^u \left(Q_Y(s) - Q_Y\left(u + \frac{1-u}{u}s\right)\right)ds \\
&= \sigma_G\,\widetilde{Q}_Y(u).
\end{aligned}$$

This concludes the proof of (1.12). □

Proof of Inequality (1.16). — First, note that

$$Z = \sup_{S\in\mathcal{S}}\sum_{k=1}^{n}\left(U_k^{-1/\ell}\mathbb{1}_S(U_k) - \int_S u^{-1/\ell}du\right) = \sum_{k=1}^{n}\left(U_k^{-1/\ell}\mathbb{1}_{U_k\leqslant\Delta} - \int_S u^{-1/\ell}du\right).$$

Next,

$$\frac{1}{n}\operatorname{Var}(Z) = \operatorname{Var}\left(U^{-1/\ell}\mathbb{1}_{U\leqslant\Delta}\right) = \frac{\ell}{\ell-2}\,\Delta^{(\ell-2)/\ell} - \frac{\ell}{\ell-1}\,\Delta^{2(\ell-1)/\ell}. \tag{A.1}$$

Moreover, for all $S \in \mathcal{S}$,

$$\sigma_S^2 := \operatorname{Var}\left(U^{-1/\ell}\mathbb{1}_S(U)\right) \leqslant \int_0^p u^{2/\ell}du = \frac{\ell}{\ell-2}\,p^{\ell-2/\ell}. \tag{A.2}$$

Combining (A.1) and (A.2), and choosing p and Δ small enough, we derive

$$\frac{1}{n}\operatorname{Var}(Z) - \sigma^2 \geqslant K\,\frac{\ell}{\ell-2}\Delta^{(\ell-1)/\ell}, \tag{A.3}$$

for a constant $K > 0$. See now that

$$\frac{1}{n}\mathbb{E}[Z] = \mathbb{E}\left[U^{-1/\ell}\mathbb{1}_{U\leqslant\Delta}\right] = \left(1+\frac{1}{\ell-1}\right)\Delta^{(\ell-2)/(\ell-1)}, \tag{A.4}$$

which implies

$$\Delta = \left(\frac{\ell-1}{\ell}\frac{\mathbb{E}[Z]}{n}\right)^{\frac{\ell}{\ell-1}}. \tag{A.5}$$

Putting (A.5) into (A.3) leads to

$$\frac{1}{n}\operatorname{Var}(Z) - \sigma^2 \geqslant K_1\left(\frac{\mathbb{E}[Z]}{n}\right)^{\frac{\ell}{\ell-1}}, \tag{A.6}$$

where $K_1 > 0$, which ends the proof. □

Proof of Lemma 5.7. — The lemma follows from the following general result:

Lemma A.1. — *Let $p > 2$. Let X be a nonnegative, $\mathbb{L}^p$-integrable, random variable. Then*

$$\mathbb{E}[|X-\mathbb{E}[X]|^p] - \mathbb{E}[|X|^p] \leqslant -2(\mathbb{E}[X])^{p-1}\mathbb{E}[X\mathbb{1}_{X\leqslant\mathbb{E}[X]/2}].$$

Assume that $\mathbb{E}[X] = 1$. The general case follows by considering $X/\mathbb{E}[X]$. Let $M = \mathbb{E}[X\mathbb{1}_{X\leqslant 1/2}]$, and define for any $x \geqslant 0$,

$$f_p(x) = |x-1|^p - |x|^p.$$

Let $g_p := f_p\mathbb{1}_{[0,1/2]}$ and $h_p := f_p\mathbb{1}_{[1/2,+\infty[}$. Since g_p is a decreasing and convex function,

$$\begin{aligned}\mathbb{E}[g_p(X)] &\leqslant \mathbb{E}\left[g_p\left(X\mathbb{1}_{X\leqslant 1/2}\right)\right]\\ &\leqslant (1-2M)g_p(0) + 2Mg_p(1) = 1-2M\,.\end{aligned} \tag{A.7}$$

d Next, since h_p is concave,

$$\mathbb{E}\left[h_p(X)\right] \leqslant h_p(1) = -1. \tag{A.8}$$

Now, since $f_p = g_p + h_p$, the claim follows by combining (A.7) and (A.8). □

Proof of Inequality (5.54). — The claim is an application of the following general comparison result:

Lemma A.2. — *Let X be a centered random variable. Let $0 \leqslant a \leqslant b$. Then, for any convex function φ*

$$\mathbb{E}\left[\varphi(aX)\right] \leqslant \mathbb{E}\left[\varphi(bX)\right].$$

Let φ be a $\mathcal{C}^2$ convex function. Using the following version of Taylor's formula

$$\varphi(x) = \varphi(0) + x\varphi'(0) + |x|^2\int_0^1(1-s)\varphi''(sx)ds, \quad x\in\mathbb{R},$$

we get

$$\begin{aligned}\mathbb{E}\left[\varphi(aX)\right] &= \varphi(0) + a^2\,\mathbb{E}\left[|X|^2\int_0^1\varphi''(saX)ds\right]\\ &= \varphi(0) + ab\,\mathbb{E}\left[|X|^2\int_0^{a/b}\left(1-\frac{b}{a}t\right)\varphi''(tbX)dt\right].\end{aligned}$$

Now, since $\varphi'' \geqslant 0$ and $0 \leqslant a \leqslant b$,

$$ab\,\mathbb{E}\left[|X|^2 \int_0^{a/b} \left(1 - \frac{b}{a}t\right) \varphi''(tbX)dt\right] \leqslant b^2\,\mathbb{E}\left[|X|^2 \int_0^1 (1-t)\,\varphi''(tbX)dt\right].$$

Therefrom,

$$\mathbb{E}[\varphi(aX)] \leqslant \mathbb{E}[\varphi(bX)].$$

The general case follows from the monotone convergence theorem since any convex function can be approximated by an increasing sequence of $\mathcal{C}^2$ convex Lipschitz functions. □

BIBLIOGRAPHY

[Ada08] Radoslaw Adamczak, *A tail inequality for suprema of unbounded empirical processes with applications to Markov chains*, Electron. J. Probab. **13** (2008), 1000–1034. ↑832, 833

[BBLM05] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart, *Moment inequalities for functions of independent random variables*, Ann. Probab. **33** (2005), no. 2, 514–560. ↑835, 840

[BDR15] Bernard Bercu, Bernard Delyon, and Emmanuel Rio, *Concentration Inequalities for Sums and Martingales*, SpringerBriefs in Mathematics, Springer, 2015. ↑845, 846

[Ben03] Vidmants-Kastytis Bentkus, *An Inequality for Tail Probabilities of Martingales with Differences Bounded from One Side*, J. Theor. Probab. **16** (2003), no. 1, 161–173. ↑850

[Ben04] ———, *On Hoeffding's inequalities*, Ann. Probab. **32** (2004), no. 2, 1650–1673. ↑845, 855

[Ben08] ———, *An extension of the Hoeffding inequality to unbounded random variables*, Lith. Math. J. **48** (2008), no. 2, 137–157. ↑837, 852

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013. ↑832, 836, 840, 846

[Bob03] Sergey G. Bobkov, *Localization Proof of the Bakry–Ledoux Isoperimetric Inequality and Some Applications*, Theory Probab. Appl. **47** (2003), no. 2, 308–314. ↑835

[Bou03] Olivier Bousquet, *Concentration inequalities for sub-additive functions using the entropy method*, Stochastic inequalities and applications. Selected papers presented at the Euroconference on "Stochastic inequalities and their applications", Barcelona, June 18–22, 2002, Progress in Probability, vol. 56, Birkhäuser, 2003, pp. 213–247. ↑832

[BS88] Colin Bennett and Robert Sharpley, *Interpolation of operators*, Pure and Applied Mathematics, vol. 129, Academic Press Inc., 1988. ↑838

[CCK14] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato, *Gaussian approximation of suprema of empirical processes*, Ann. Stat. **42** (2014), no. 4, 1564–1597. ↑842

[CIS76] Boris S. Cirel'son, Il'dar A. Ibragimov, and Vladimir N. Sudakov, *Norms of Gaussian sample functions*, Proceedings of the Third Japan. USSR Symposium on Probability Theory (Gisiro Maruyama and Jurii V. Prokhorov, eds.), Lecture Notes in Mathematics, Springer, 1976, pp. 20–41. ↑835, 847

[CL13] Fabienne Comte and Claire Lacour, *Anisotropic adaptive kernel deconvolution*, Ann. Inst. Henri Poincaré, Probab. Stat. **49** (2013), no. 2, 569–609. ↑844

[Dir15] Sjoerd Dirksen, *Tail bounds via generic chaining*, Electron. J. Probab. **20** (2015), article no. 53. ↑832, 833

[L13] Sara van de Geer and Johannes Lederer, *The Bernstein–Orlicz norm and deviation inequalities*, Probab. Theory Relat. Fields **157** (2013), no. 1-2, 225–250. ↑832, 833

[GM88] David Gilat and Isaac Meilijson, *A simple proof of a theorem of Blackwell and Dubins on the maximum of a uniformly integrable martingale*, Séminaire de probabilités XXII, Strasbourg/France, Lecture Notes in Mathematics, vol. 1321, Springer, 1988, pp. 214–216. ↑838

[GN16] Evarist Giné and Richard Nickl, *Mathematical foundations of infinite-dimensional statistical models*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 40, Cambridge University Press, 2016. ↑842, 843

[Hoe63] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Am. Stat. Assoc. **58** (1963), 13–30. ↑835, 837

[Kle02] Thierry Klein, *Une inégalité de concentration à gauche pour les processus empiriques*, C. R. Math. Acad. Sci. Paris **334** (2002), no. 6, 501–504. ↑832, 846

[KR05] Thierry Klein and Emmanuel Rio, *Concentration around the mean for maxima of empirical processes*, Ann. Probab. **33** (2005), no. 3, 1060–1077. ↑832

[Led97] Michel Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM, Probab. Stat. **1** (1997), 63–87. ↑832

[LvdG14] Johannes Lederer and Sara van de Geer, *New concentration inequalities for suprema of empirical processes*, Bernoulli **20** (2014), no. 4, 2020–2038. ↑832, 833

[Mar18] Antoine Marchina, *Concentration inequalities for separately convex functions*, Bernoulli **24** (2018), no. 4A, 2906–2933. ↑833, 837

[Mas00] Pascal Massart, *About the constants in Talagrand's concentration inequalities for empirical processes*, Ann. Probab. **28** (2000), no. 2, 863–884. ↑832

[Pin06] Iosif Pinelis, *On normal domination of (super)martingales*, Electron. J. Probab. **11** (2006), 1049–1070. ↑845

[Pin14a] ———, *On the Bennett-–Hoeffding inequality*, Ann. Inst. Henri Poincaré, Probab. Stat. **50** (2014), no. 1, 15–27. ↑845

[Pin14b] ———, *An optimal three-way stable and monotonic spectrum of bounds on quantiles: a spectrum of coherent measures of financial risk and economic inequality*, Risks **2** (2014), no. 3, 349–392. ↑834, 845

[Pin15] ———, *Rosenthal-type inequalities for martingales in 2-smooth Banach spaces*, Theory Probab. Appl. **59** (2015), no. 4, 699–706. ↑840, 851

[Rio01] Emmanuel Rio, *Inégalités de concentration pour les processus empiriques de classes de parties*, Probab. Theory Relat. Fields **119** (2001), no. 2, 163–175. ↑832

[Rio02] ———, *Une inégalité de Bennett pour les maxima de processus empiriques*, Ann. Inst. Henri Poincaré, Probab. Stat. **38** (2002), no. 6, 1053–1057, En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov. ↑832, 843

[Rio12] ———, *Sur la fonction de taux dans les inégalités de Talagrand pour les processus empiriques*, C. R. Math. Acad. Sci. Paris **350** (2012), no. 5-6, 303–305. ↑832

[Rio17a] ———, *About the constants in the Fuk–Nagaev inequalities*, Electron. Commun. Probab. **22** (2017), article no. 28. ↑833, 835, 839, 851, 852

[Rio17b] ———, *Asymptotic theory of weakly dependent random processes*, Probability Theory and Stochastic Modelling, vol. 80, Springer, 2017, Translated from the 2000 French edition [MR2117923]. ↑834, 841

[RU00] Ralph T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, Risks **2** (2000), 21–42. ↑834

[Sam07] Paul-Marie Samson, *Infimum-convolution description of concentration properties of product probability measures, with applications*, Ann. Inst. Henri Poincaré, Probab. Stat. **43** (2007), no. 3, 321–338. ↑839

[Sio58] Maurice Sion, *On general minimax theorems*, Pac. J. Math. **8** (1958), 171–176. ↑854

[Tal96] Michel Talagrand, *New concentration inequalities in product spaces*, Invent. Math. **126** (1996), no. 3, 505–563. ↑832

[W96] Adrianus W. van der Vaart and Jon A. Wellner, *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, Springer, 1996. ↑837, 839

[W11] ———, *A local maximal inequality under uniform entropy*, Electron. J. Stat. **5** (2011), 192–203. ↑842, 853

Manuscript received on 10th October 2019,
accepted on 27th July 2020.

Recommended by Editor A. Philippe.

This journal is a member of Centre Mersenne.

Antoine MARCHINA
Université de Paris,
CNRS, MAP5 UMR 8145,
F-75006 Paris, (France)
antoine.marchina@parisdescartes.fr