

Concentration inequalities for separately convex functions

ANTOINE MARCHINA

Laboratoire de Mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035 Versailles, France. E-mail: antoine.marchina@uvsq.fr

We provide new comparison inequalities for separately convex functions of independent random variables. Our method is based on the decomposition in Doob martingale. However, we only impose that the martingale increments are stochastically bounded. For this purpose, building on the results of Bentkus (*Lith. Math. J.* **48** (2008) 237–255; *Lith. Math. J.* **48** (2008) 137–157; Bounds for the stop loss premium for unbounded risks under the variance constraints (2010) Preprint), we establish comparison inequalities for random variables stochastically dominated from below and from above. We illustrate our main results by showing how they can be used to derive deviation or moment inequalities for functions which are both separately convex and separately Lipschitz, for weighted empirical distribution functions, for suprema of randomized empirical processes and for chaos of order two.

Keywords: concentration inequalities; deviation inequalities; generalized moments; martingale method; suprema of empirical processes

1. Introduction

Let E be a vector space. A function F from E^n into \mathbb{R} is said to be separately convex if it is convex in each coordinate. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X_1, \dots, X_n be a finite sequence of independent and centered random variables with values in E . Throughout the paper, F is a measurable separately convex function from E^n to \mathbb{R} . In this work, we are concerned with deviation inequalities for the random variable

$$Z := F(X_1, \dots, X_n). \quad (1.1)$$

Before going further, let us introduce some notations which are used in this paper. Set $\mathcal{F}_0 := \{\emptyset, \Omega\}$ and for all $k = 1, \dots, n$, $\mathcal{F}_k := \sigma(X_1, \dots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. Let \mathbb{E}_k (respectively \mathbb{E}_n^k) denote the conditional expectation operator associated to \mathcal{F}_k (resp. \mathcal{F}_n^k). Set also

$$Z^{(k)} := F(X_1, \dots, X_{k-1}, 0, X_{k+1}, \dots, X_n), \quad (1.2)$$

$$Z_k := \mathbb{E}_k[Z - \mathbb{E}[Z]]. \quad (1.3)$$

Our approach to obtain deviation inequalities is based on the martingale method. The idea is to decompose the random variable $Z - \mathbb{E}[Z]$ as a sum of martingale increments. Precisely, the

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

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

The main problem is to control the increments Δ_k . Classical concentration inequalities for martingales assume that their increments are bounded (see, for example, Chapter 3 of Bercu, Delyon and Rio [8]). In this paper, our hypotheses on F and on the random variables X_1, \dots, X_n do not imply a deterministic boundedness condition on the martingale increments, but only a symmetric two-sided stochastic one: $-\xi_k \leq Z_k - Z^{(k)} \leq \xi_k$, for some stochastic order \leq , where ξ_1, \dots, ξ_n are real-valued nonnegative random variables. Δ_k and $Z_k - Z^{(k)}$ are linked by the following observation:

$$\Delta_k = Z_k - \mathbb{E}_k[Z^{(k)}] - \mathbb{E}_{k-1}[Z_k - \mathbb{E}_k[Z^{(k)}]]. \tag{1.4}$$

Note that this observation was already made by Pinelis and Sakhnenko [24] (see their Inequality (9)) when the function F is the norm of the sum. Let us now explain which stochastic order we work with. Let $\alpha > 0$. We define the class \mathcal{H}_+^α of functions φ from \mathbb{R} into \mathbb{R} as follows:

$$\mathcal{H}_+^\alpha := \left\{ \varphi : \varphi(u) = \int_{-\infty}^\infty (u - t)_+^\alpha \mu(dt) \text{ for some Borel measure } \mu \geq 0 \text{ on } \mathbb{R} \right. \\ \left. \text{and all } u \in \mathbb{R} \right\}.$$

Here, as usual, $x_+ := x \vee 0 := \max(0, x)$ and $x_+^\alpha := (x_+)^alpha$ for all real x . Using the family \mathcal{H}_+^α , we define a family of stochastic order by the formula

$$X \underset{\mathcal{H}_+^\alpha}{\preceq} \xi \quad \text{if } \mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\xi)] \text{ for all } \varphi \in \mathcal{H}_+^\alpha, \tag{1.5}$$

where X and ξ are real-valued random variables. We refer the reader to Pinelis [22] for more on this stochastic order. Our main results in this paper will be expressed in terms of comparison inequality with respect to $\underset{\mathcal{H}_+^\alpha}{\preceq}$ between $Z - \mathbb{E}[Z]$ and a function of ξ_1, \dots, ξ_n .

Concerning general functions of independent random variables, Boucheron, Bousquet, Lugosi and Massart [9] provided general moment inequalities, using an extension of the entropy method proposed by Ledoux [17]. They derived moment inequalities for various functions such as homogeneous tetrahedral polynomials in Rademacher variables or unbounded empirical processes. Recently, Adamczak and Wolff [2] (see Theorem 1.4) gave a concentration inequality for polynomials of independent sub-Gaussian random variables.

Moreover, if F is separately Lipschitz (E is then assumed equipped with a norm), $Z_k - Z^{(k)}$ satisfies naturally our stochastic boundedness conditions. When F is only separately Lipschitz, a corollary of a result of Pinelis [18] gives that

$$F(X_1, \dots, X_n) \underset{\mathcal{H}_+^1}{\preceq} \sum_{k=1}^n \varepsilon_k \|X_k\|,$$

where $\varepsilon_1, \dots, \varepsilon_n$ is a sequence of independent Rademacher random variables. Kontorovich [15] gave extensions of McDiarmid’s inequality for metric spaces with unbounded diameter. He required a sub-Gaussian control of the symmetrized of $\|X_k - X'_k\|$ where X'_k is an independent copy of X_k .

A particular case of separately convex functions is suprema of empirical processes: $F(x_1, \dots, x_n) = \sup_{t \in \mathcal{T}} \sum_{i=1}^n x_{i,t}$, where \mathcal{T} is a countable index set. Only few results concern concentration inequalities for suprema of unbounded empirical processes: assuming weak tails with respect to suitable Orlicz norms, Adamczak [1], and van de Geer and Lederer [27] obtained exponential bounds. Later van de Geer and Lederer [16] required only weak moment conditions on an envelope of the class of functions and obtained generalized moment inequalities. In this paper, we will also treat the case of $F(x_1, \dots, x_n) = \sup_{t \in \mathcal{T}} \sum_{1 \leq i < j \leq n} x_{i,t} x_{j,t}$, which is a particular case of supremum of polynomials in independent random variables.

We shall use the following notation throughout the paper. The quantile function of a real-valued random variable X which is the general inverse of the nonincreasing and left continuous tail function of X , $\mathbb{P}(X > t)$, is denoted by Q_X . It is defined by

$$Q_X(u) := \inf\{x \in \mathbb{R} : \mathbb{P}(X > x) \leq u\}.$$

Moreover, for $p \geq 1$, let \mathbb{L}^p be the space of real-valued random variables with a finite absolute moment of order p and we denote by $\|X\|_p$ the \mathbb{L}^p -norm of X . Let $(a_1, \dots, a_n) \in \mathbb{R}^n$. As usual for any $r \geq 1$, we write

$$\|a\|_r = \left(\sum_{k=1}^n |a_k|^r \right)^{1/r}, \quad \text{and} \quad \|a\|_\infty = \max_{1 \leq k \leq n} |a_k|.$$

Finally, for any real function f , we denote by $f(a+)$ (respectively, $f(a-)$) the right (resp. left) limit of f at point a .

The paper is organized as follows. In Section 2, we state the main results of this paper. In Section 3, we explain how we can extract a tail comparison inequality from a comparison inequality with respect to the stochastic order associated with the class \mathcal{H}_+^α . In Section 4, new comparison inequalities for unbounded real-valued random variables are given. The results in this section will allow us to control the increments of the Doob martingale associated to $Z - \mathbb{E}[Z]$. We provide detailed proofs of Sections 2 and 4 in Section 9. We give some applications of the main results in other sections: in Section 5 we examine the special case where F is also separately Lipschitz. Section 6 considers the weighted empirical distribution functions, Section 7 deals with the suprema of randomized empirical processes. Finally, Theorems 2.1 and 2.3 are applied to chaos of order two in Section 8.

2. Main results

Theorem 2.1. *Let Z and $Z^{(k)}$ be defined respectively by (1.1) and (1.2). Assume that for all $k = 1, \dots, n$, there exist nonnegative, square integrable and $\sigma(X_k)$ -measurable random variables T_k and W_k such that,*

$$-T_k \leq Z - Z^{(k)} \leq W_k, \quad \text{almost surely.} \tag{2.1}$$

Let ξ_1, \dots, ξ_n be any finite sequence of nonnegative random variables such that, for any real t ,

$$\max(\mathbb{E}[(T_k - t)_+], \mathbb{E}[(W_k - t)_+]) \leq \mathbb{E}[(\xi_k - t)_+]. \tag{2.2}$$

Then

$$Z - \mathbb{E}[Z] \underset{\mathcal{H}_+^2}{\preceq} \sum_{k=1}^n \varepsilon_k Q_{\xi_k}(U_k/2), \tag{2.3}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent Rademacher random variables, U_1, \dots, U_n are independent random variables distributed uniformly on $[0, 1]$ and these two families are independent.

Remark 2.2. Using new results of Pinelis [23] (see his Corollary 5.8), it is straightforward to extend (2.3) to the larger class of differentiable convex nondecreasing function with a convex derivative.

In the following result, we relax the assumption (2.1) and we instead assume that the bounds have a \mathcal{F}_n^k -measurable component.

Theorem 2.3. Let $r > 2$. Let Z and $Z^{(k)}$ be defined respectively by (1.1) and (1.2). Assume that for all $k = 1, \dots, n$, there exist nonnegative, \mathbb{L}^r -integrable and $\sigma(X_k)$ -measurable random variables T_k and W_k and nonnegative, \mathbb{L}^r -integrable and \mathcal{F}_n^k -measurable random variable ψ_k such that

$$-T_k \psi_k \leq Z - Z^{(k)} \leq W_k \psi_k, \quad \text{almost surely.} \tag{2.4}$$

Let ξ_1, \dots, ξ_n be any finite sequence of nonnegative random variables such that, for any real t ,

$$\max(\mathbb{E}[(T_k - t)_+], \mathbb{E}[(W_k - t)_+]) \leq \mathbb{E}[(\xi_k - t)_+]. \tag{2.5}$$

Then

$$\|(Z - \mathbb{E}[Z])_+\|_r^2 \leq (p - 1) \sum_{k=1}^n \|\mathbb{E}_{k-1}[\psi_k]\|_r^2 \|Q_{\xi_k}(U_k/2)\|_r^2, \tag{2.6}$$

where U_1, \dots, U_n are independent random variables uniformly distributed on $[0, 1]$.

3. Concentration inequalities from comparison inequalities in \mathcal{H}_+^α

In this section, we repeat the relevant materials from [20] and [21] without proofs, of how one obtains a deviation inequality from a comparison inequality with respect to the stochastic order associated with the class \mathcal{H}_+^α , $\alpha > 0$, such as in Theorem 2.1.

First, let us mention some facts about the class \mathcal{H}_+^α . It is easy to see that $0 \leq \beta < \alpha$ implies $\mathcal{H}_+^\alpha \subset \mathcal{H}_+^\beta$. Moreover, for any real t and any positive λ , the functions $x \mapsto (x - t)_+^\alpha$ and $x \mapsto e^{\lambda(x-t)}$ belong to \mathcal{H}_+^α . Finally, the following assertions are equivalent:

- (i) $X \underset{\mathcal{H}_+^\alpha}{\preceq} \xi$,
- (ii) $\mathbb{E}[(X - t)_+^\alpha] \leq \mathbb{E}[(\xi - t)_+^\alpha]$ for all real t .

The following is a special case of Theorem 4 of Pinelis [21].

Theorem 3.1. *Suppose that $\alpha > 0$, X and ξ are real-valued random variables, and the tail function $x \mapsto \mathbb{P}(\xi \geq x)$ is log-concave on \mathbb{R} . Then the comparison inequality $X \underset{\mathcal{H}_+^\alpha}{\preceq} \xi$ implies that, for all real x ,*

$$\mathbb{P}(X \geq x) \leq P_\alpha(\xi; x) := \inf_{t < x} \frac{\mathbb{E}[(\xi - t)_+^\alpha]}{(x - t)^\alpha} \tag{3.1}$$

$$\leq c_{\alpha,0} \mathbb{P}(\xi \geq x), \tag{3.2}$$

where the constant factor $c_{\alpha,0} := \Gamma(\alpha + 1)(e/\alpha)^\alpha$ is the best possible.

Remark 3.2. A thorough study of $P_\alpha(\xi; x)$ can be found in Pinelis [22]. See also Bentkus, Kalosha and van Zuijlen [7] for a description of the calculation for specific α and specific families of distribution.

Remark 3.3. Since the class \mathcal{H}_+^α contains all increasing exponential functions, $P_\alpha(\xi; x)$ is also majorized by the exponential bound $\inf_{\lambda > 0} e^{-\lambda x} \mathbb{E}[e^{\lambda \xi}]$. For all small enough x , the exponential bound is better than (3.2). However, for large values of x , the latter will be significantly better than the exponential one.

4. New comparison inequalities

The purpose of this section is to obtain extensions of an inequality of Hoeffding to unbounded random variables. In particular, Lemma 4.6 below will be our main tool to control the increments of the Doob martingale associated to $Z - \mathbb{E}[Z]$. First, let us recall the definition of the usual stochastic order. Let X and Y be two real-valued random variables. X is said to be smaller than Y in the usual stochastic order, denoted by $X \leq_{st} Y$, if $\mathbb{P}(X \geq x) \leq \mathbb{P}(Y \geq x)$ for all real x .

Throughout this section, η and ψ are random variables such that

$$\eta \in L^1, \quad \psi \in L^2, \quad \text{and} \quad \eta \leq_{st} \psi. \tag{4.1}$$

We introduce a family of probability distribution related to the distributions of η and ψ . We recall first some classical notations. The distribution function of a real-valued random variable X is denoted by F_X . The generalized inverse of F_X is defined by

$$F_X^{-1}(u) := \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq u\}.$$

Definition 4.1. Let assumption (4.1) hold. For every q in $(0, 1)$, set $a_q := F_\eta^{-1}(1 - q)$, $b_q := F_\psi^{-1}(1 - q)$ and let F_q be the distribution function defined by

$$F_q(x) := \begin{cases} F_\eta(x) & \text{if } x < a_q, \\ 1 - q & \text{if } a_q \leq x < b_q, \\ F_\psi(x) & \text{if } x \geq b_q. \end{cases}$$

We also set $F_0 := F_\eta$ and $F_1 := F_\psi$. In the following, we always denote by ζ_q a random variable having the distribution function F_q .

Remark 4.2. A similar construction can be found in Bentkus [4,5] and [6].

The following bound was obtained by Bentkus [6] (see Theorem 1) with a little stronger assumption on the stochastic boundedness condition. Indeed Bentkus supposed that $\eta \leq_{st} X \leq_{st} \psi$, which implies our hypothesis (4.2).

Lemma 4.3. Let assumption (4.1) hold. Let ζ_q be as in Definition 4.1 and let X be an integrable random variable such that for any real t ,

$$\mathbb{E}[(X - t)_+] \leq \mathbb{E}[(\psi - t)_+], \quad \mathbb{E}[(t - X)_+] \leq \mathbb{E}[(t - \eta)_+]. \tag{4.2}$$

Let q_0 be the highest real in $[0, 1]$ such that

$$\int_{1-q_0}^1 (F_\psi^{-1}(u) - F_\eta^{-1}(u)) du = \mathbb{E}[X] - \mathbb{E}[\eta]. \tag{4.3}$$

Then, X and ζ_{q_0} have the same expectation and for any real t ,

$$\mathbb{E}[(X - t)_+] \leq \mathbb{E}[(\zeta_{q_0} - t)_+]. \tag{4.4}$$

Consequently, for any convex function φ ,

$$\mathbb{E}[\varphi(X - \mathbb{E}[X])] \leq \mathbb{E}[\varphi(\zeta_{q_0} - \mathbb{E}[\zeta_{q_0}])]. \tag{4.5}$$

Remark 4.4. As noticed by Bentkus [4–6], we can see this lemma as an extension of an inequality of Hoeffding. Indeed, if η and ψ are two constants, respectively equal to a and b , it is easy to see that (4.1) and (4.2) imply that $a \leq X \leq b$ a.s. Then we obtain for all convex function φ that $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\theta)]$ where θ is a two-valued random variable taking the values a and b , and such that $\mathbb{E}[X] = \mathbb{E}[\theta]$.

Remark 4.5. The special case $0 \leq X \leq_{st} \psi$ was considered by Bentkus [4,5]. In [6], Bentkus obtained similar results in the situation where $X \leq_{st} \psi$ and the variance of X is known.

The right-hand side of (4.5) still depends on the expectation of X by the term $\mathbb{E}[\zeta_{q_0}]$. The next lemma provides a bound in the symmetric case $\eta = -\psi$, which does not depend on $\mathbb{E}[X]$. The drawback is that we have to pick φ in the smaller class of functions \mathcal{H}_+^2 .

Lemma 4.6. *Let ψ and η be two random variables, respectively nonnegative and nonpositive, satisfying (4.1). Let ζ_q, a_q and b_q be given by Definition 4.1.*

(i) *Let $\tilde{q} := \inf\{q \geq 1/2 : b_q + a_q \leq 2\mathbb{E}[\zeta_q]\}$. Then for all real t ,*

$$q \mapsto \mathbb{E}\left[\left(\zeta_q - \mathbb{E}[\zeta_q] - t\right)_+^2\right]$$

is nonincreasing on $[\tilde{q}, 1]$.

(ii) *Assume that $\eta = -\psi$ and let X be an integrable random variable satisfying (4.2). If $\mathbb{E}[X] \geq 0$, then*

$$X - \mathbb{E}[X] \preceq_{\mathcal{H}_+^2} \zeta_{1/2}. \tag{4.6}$$

Remark 4.7. We have a better understanding of the random variable $\zeta_{1/2}$ if we observe that it has the same distribution than $\varepsilon Q_\psi(U/2)$, where ε is a Rademacher random variable, U is a random variable distributed uniformly on $[0, 1]$ and these random variables are independent.

The following result is a corollary of a result obtained by Pinelis [19]. It will be needed in the proof of Theorem 2.3 and in Section 8.

Proposition 4.8. *Let $r > 2$ and let X and Y be random variables in \mathbb{L}^r such that $\mathbb{E}[Y | X] = 0$ almost surely. Then $\|(X + Y)_+\|_r^2 \leq \|(X)_+\|_r^2 + (r - 1)\|Y\|_r^2$.*

Exactly as in Rio [26] (see Theorem 2.1), we deduce from Proposition 4.8 the following inequality by induction on n .

Corollary 4.9. *Let $r > 2$ and $(M_n)_{n \geq 0}$ be a sequence of random variables in \mathbb{L}^r . Set $\Delta M_k := M_k - M_{k-1}$. Assume that $\mathbb{E}[\Delta M_k | M_{k-1}] = 0$ almost surely for any positive k . Then*

$$\|(M_n)_+\|_r^2 \leq \|(M_0)_+\|_r^2 + (r - 1) \sum_{k=1}^n \|\Delta M_k\|_r^2. \tag{4.7}$$

5. Lipschitz functions of independent random vectors

Throughout this section, we assume that $(E, \|\cdot\|)$ is a separable Banach space. In addition to being separately convex, we suppose that F is separately 1-Lipschitz. Precisely, F satisfies the following Lipschitz type condition:

$$\left| F(x_1, \dots, x_n) - F(y_1, \dots, y_n) \right| \leq \sum_{k=1}^n \|x_k - y_k\|.$$

Now, $Z = F(X_1, \dots, X_n)$ naturally fulfills the hypotheses (2.1)–(2.2) of Theorem 2.1 with $\xi_k = \|X_k\|$.

5.1. Moment inequality

Proposition 5.1. *Let $r \geq 2$ and define the function Q by*

$$Q^2(u) := \sum_{k=1}^n Q_{\|X_k\|}^2(u).$$

Then

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] \leq \|g\|_r^r \int_0^{1/2} Q^r(u) du, \tag{5.1}$$

where g is standard Gaussian random variable.

Example 5.2. Let X be a centered random vector with values in E and a_1, \dots, a_n be deterministic reals. Let X_1, \dots, X_n be n independent copies of X . Define the function F by

$$Z := F(\tilde{X}_1, \dots, \tilde{X}_n) := \left\| \sum_{k=1}^n a_k X_k \right\|,$$

where $\tilde{X}_k := a_k X_k$. Then Proposition 5.1 yields for any $r \geq 2$ that

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] \leq \|g\|_r^r \left(\sum_{k=1}^n a_k^2 \right)^{r/2} \int_0^{1/2} Q_{\|X\|}^r(u) du \tag{5.2}$$

$$\leq \|g\|_r^r \left(\sum_{k=1}^n a_k^2 \right)^{r/2} \|X\|^r. \tag{5.3}$$

We now apply this result to suprema of empirical processes, that is,

$$Z = \sup_{f \in \mathcal{F}} \left| \sum_{k=1}^n f(X_k) \right|,$$

where \mathcal{F} is a countable class of measurable real-valued functions. To do this, we first assume that \mathcal{F} is finite and we then conclude by the monotonous convergence theorem. We suppose that \mathcal{F} has an r -integrable envelop function Φ and we set $M^r := \mathbb{E}[\Phi^r(X)]$. Thus (5.3) yields

$$\|(Z - \mathbb{E}[Z])_+\|_r \leq \pi^{-1/2r} \sqrt{2 \sum_{k=1}^n a_k^2} M \left(\Gamma\left(\frac{r+1}{2}\right) \right)^{1/r}, \tag{5.4}$$

where Γ is the usual Gamma function. This result improves Theorem 4.1 of Lederer and van de Geer [16].

Proof of Proposition 5.1. Theorem 2.1 applied with $\xi_k = \|X_k\|$ and (1.5) specified to $\varphi(x) = x_+^r$ yield that

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] \leq \mathbb{E}\left[\left(\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\right)_+^r\right]. \tag{5.5}$$

Since the random variables $\varepsilon_k Q_{\|X_k\|}(U_k/2)$ are symmetric,

$$\mathbb{E}\left[\left(\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\right)_+^r\right] = \frac{1}{2} \mathbb{E}\left[\left|\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\right|^r\right]. \tag{5.6}$$

Conditioning by \mathcal{F}_n and using the classical Khintchine inequality with the best possible constant founded by Whittle (for $r \geq 3$) and Haagerup (for $r > 0$) (see the Introduction of Figiel et al. [13] and references therein for a statement of these results), one has

$$\frac{1}{2} \mathbb{E}\left[\left|\sum_{k=1}^n \varepsilon_k Q_{\|X_k\|}(U_k/2)\right|^r\right] \leq \frac{1}{2} \|g\|_r^r \mathbb{E}\left[\left(\sum_{k=1}^n Q_{\|X_k\|}^2(U_k/2)\right)^{r/2}\right], \tag{5.7}$$

where g is a standard Gaussian random variable. Next, it is an easy exercise which is left to the reader, to see that Riesz Representation Theorem and Lemma 2.1(a) and (c) of Rio [25] imply that

$$\mathbb{E}\left[\left(\sum_{k=1}^n Q_{\|X_k\|}^2(U_k/2)\right)^{r/2}\right] \leq \mathbb{E}\left[\left(\sum_{k=1}^n Q_{\|X_k\|}^2(U/2)\right)^{r/2}\right], \tag{5.8}$$

where U is a random variable distributed uniformly on $[0, 1]$. Finally, combining (5.5)–(5.8), one has (5.1) which ends the proof. \square

5.2. A deviation inequality for the bounded case

Consider the bounded case $\|X_k\| \leq a_k$ a.s., for some positive reals a_k . Theorem (2.1) implies that

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \sum_{k=1}^n a_k \varepsilon_k, \tag{5.9}$$

where ε_k are i.i.d. Rademacher random variables.

Remark 5.3. Inequality (5.9) can be obtained directly via Lemma 4.4. of Bentkus [3].

Proposition 5.4. Let ℓ be the function defined by $\ell(t) := \log(\cosh(t))$ and let ℓ^* denote the Legendre–Fenchel transform of ℓ , which is defined for any positive x by $\ell^*(x) := \sup_{t>0} \{xt - \ell(t)\}$. Then for any x in $[0, 1]$,

$$\ell^*(x) = \frac{1}{2}((1+x)\log(1+x) + (1-x)\log(1-x)),$$

and

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \|a\|_1 x) \leq \exp\left(-\frac{\|a\|_1^2}{\|a\|_2^2} \ell^*(x)\right). \tag{5.10}$$

Proof. Since the exponential function $x \mapsto e^{tx}$ belongs to \mathcal{H}_+^2 for any positive t , (5.9) ensures that

$$\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \log \mathbb{E}\left[\exp\left(t \sum_{k=1}^n \varepsilon_k a_k\right)\right] = \sum_{k=1}^n \ell(a_k t). \tag{5.11}$$

Note that $\ell'(\cdot) = \tanh(\cdot)$ is a concave function on $[0, \infty)$. Now, from (5.11), proceeding exactly as in Bercu, Delyon and Rio [8] (see Inequality (2.98)) we get

$$\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \frac{\|a\|_1^2}{\|a\|_2^2} \ell\left(\frac{\|a\|_2^2 t}{\|a\|_1}\right). \tag{5.12}$$

Finally, (5.10) follows from Markov's inequality together with (5.12). □

6. Weighted empirical distribution functions

Let U be a random variable uniformly distributed on $[0, 1]$, $\tilde{U}_1, \dots, \tilde{U}_n$ be n independent copies of U , and denote the uniform empirical process by

$$e_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (\mathbb{1}_{\tilde{U}_k \leq t} - t).$$

Let $q : [0, 1] \rightarrow \mathbb{R}$ be a weight function such that

$$q(t) = q(1-t), \quad q(t) > 0 \quad \text{on } (0, 1), \quad \int_0^{1/2} \frac{dt}{q^2(t)} < \infty,$$

$$t \mapsto \frac{q(t)}{t} \quad \text{is nonincreasing,} \quad \text{and} \quad t \mapsto \frac{q(t)}{1-t} \quad \text{is nondecreasing.}$$

Example 6.1. The most common such weight functions q are

$$q(t) = (\sqrt{t(1-t)})^\alpha, \quad \text{for any } \alpha \text{ in } (0, 1),$$

$$q(t) = \max(\sqrt{t(1-t)}, \sqrt{\delta(1-\delta)}), \quad \text{for some } 0 < \delta < 1.$$

In this section, the quantity of interest is

$$Z := \sup_{0 \leq t \leq 1} \frac{e_n(t)}{q(t)}.$$

We refer the reader to Csörgő and Horváth [11] for asymptotic results on this object. Setting now the class of function $\mathcal{F} := \{\frac{\mathbb{1}_{[0,t]}-t}{q(t)} : t \in [0, 1]\}$ and $\tilde{X}_k := (f(\tilde{U}_k))_{f \in \mathcal{F}}$, we can write Z as

$$Z = F(\tilde{X}_1, \dots, \tilde{X}_n) := \frac{1}{\sqrt{n}} \sup_{f \in \mathcal{F}} \sum_{k=1}^n f(\tilde{U}_k).$$

Proposition 6.2. *We have*

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \frac{1}{\sqrt{n}} \sum_{k=1}^n \varepsilon_k \frac{1 - U_k/2}{q(U_k/2)}. \tag{6.1}$$

Remark 6.3. Define $\ell^\infty(\mathcal{F}) := \{x : \mathcal{F} \rightarrow \mathbb{R} : \sup_{f \in \mathcal{F}} |x(f)| < \infty\}$ equipped with the norm $\|x\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |x(f)|$. Then, the summands in the right-hand side of (6.1) are equal to $\|\tilde{X}_k\|_{\mathcal{F}}$, leading to

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \frac{1}{\sqrt{n}} \sum_{k=1}^n \varepsilon_k \|\tilde{X}_k\|_{\mathcal{F}}.$$

Remark 6.4. The uniform case also treats the general one. Precisely, let X_1, \dots, X_n be n independent copies of a real-valued random variable X with a continuous distribution function F_X . Then

$$Z := \frac{1}{\sqrt{n}} \sup_{t \in \mathbb{R}} \frac{\sum_{k=1}^n (\mathbb{1}_{X_k \leq t} - F_X(t))}{q(F_X(t))} = \sup_{t \in \mathbb{R}} \frac{e_n(F_X(t))}{q(F_X(t))}.$$

Proceeding in the same way as in the proof of Proposition 5.1, we obtain the following moment inequality.

Corollary 6.5. *Let $\alpha \in (0, 1)$, $q(t) = (\sqrt{t(1-t)})^\alpha$, and $r \geq 2$ such that $r\alpha < 2$. Then*

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^r] \leq \|g\|_r^r \int_0^{1/2} (1-u)^{(2-\alpha)(r/2)} u^{-\alpha r/2} du, \tag{6.2}$$

where g is a standard Gaussian random variable.

Example 6.6. With $r = 2$ and $\alpha = 1/2$,

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^2] \leq \frac{1}{2} + \frac{3}{16}\pi \approx 1.089.$$

Proof of Proposition 6.2. For any function f in \mathcal{F} and for all $x \in [0, 1]$,

$$-\frac{x}{q(x)} \leq f(x) \leq \frac{1-x}{q(x)}.$$

Since $q(t) = q(1 - t)$, $W := (1 - U)/q(1 - U)$ and $T := U/q(U)$ have the same distribution. Moreover, $Q_T(U_k/2) = (1 - U_k/2)/q(U_k/2)$. Then Theorem 2.1 implies (6.1), and the proof is completed. \square

7. Suprema of randomized empirical processes

Let X_1, \dots, X_n be a sequence of independent random variables with values in some Polish space \mathcal{X} and Y_1, \dots, Y_n be a sequence of independent real-valued symmetric random variables such that the two sequences are independent. Let \mathcal{F} be a countable class of measurable real-valued functions and define the function F by

$$Z := F(\tilde{X}_1, \dots, \tilde{X}_n) := \sup_{f \in \mathcal{F}} \sum_{k=1}^n Y_k f(X_k), \tag{7.1}$$

where $\tilde{X}_k := (Y_k f(X_k))_{f \in \mathcal{F}}$. Assume that there exist nonnegative functions G and H such that for any function f in \mathcal{F} , $-G \leq f \leq H$. It thus follows that $-T_k \leq Z - Z^k \leq W_k$, where

$$W_k := (Y_k)_+ H(X_k) + (Y_k)_- G(X_k), \quad T_k := (Y_k)_- H(X_k) + (Y_k)_+ G(X_k).$$

Throughout, we assume that $\mathbb{E}[G^2(X_k)] < \infty$ and $\mathbb{E}[H^2(X_k)] < \infty$ for any $k = 1, \dots, n$. Since Y is symmetric, W_k and T_k have the same distribution. Then Theorem 2.1 yields

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \sum_{k=1}^n \varepsilon_k Q_{W_k}(U_k/2). \tag{7.2}$$

Throughout this section, we will use the following notation:

$$s^2 := \sum_{k=1}^n s_k^2 := \sum_{k=1}^n \mathbb{E}[Y_k^2],$$

$$\sigma^2 := \sum_{k=1}^n \sigma_k^2 := \sum_{k=1}^n \mathbb{E}[Q_{W_k}^2(U_k/2)].$$

In the rest of this section, we present how (7.2) may be used to derive concentration inequalities through several examples. However, in some cases, this bound can prove difficult to manipulate. Now, we show that, due to the symmetry of the Y_k , we can derive a more tractable comparison moment inequality, which is, however, less efficient. Precisely, set $\xi_k := (Y_k)_+(G(X_k) + H(X_k))$. For any nonnegative t , the superadditivity of the function $x \mapsto (x - t)_+$ on $[0, \infty)$ yields

$$\mathbb{E}[(W_k - t)_+] \leq \mathbb{E}[(\xi_k - t)_+].$$

This inequality remains trivially true for negative t . Moreover, since Y_k is symmetric and independent of X_k , $\varepsilon_k Q_{\varepsilon_k}(U_k/2)$ and $Y_k(G(X_k) + H(X_k))$ have the same distribution. Consequently,

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \sum_{k=1}^n Y_k(G(X_k) + H(X_k)). \tag{7.3}$$

Example 7.1. Let $F_0 : \mathbb{R} \rightarrow [0, 1]$ be a nondecreasing function. Let X_1, \dots, X_n be independent real-valued random variables and let $q : [0, 1] \rightarrow \mathbb{R}$ be a weight function such that

$$q(t) > 0 \quad \text{on } (0, 1), \quad \int_0^{1/2} \frac{dt}{q^2(t)} < \infty,$$

$$t \mapsto \frac{q(t)}{t} \quad \text{is nonincreasing, and} \quad t \mapsto \frac{q(t)}{1-t} \quad \text{is nondecreasing.}$$

Define now

$$Z := \sup_{t \in \mathbb{R}} \sum_{k=1}^n Y_k \frac{\mathbb{1}_{X_k \leq t} - F_0(t)}{q(F_0(t))}.$$

In this case

$$H(x) = \frac{1 - F_0(x)}{q(F_0(x))}, \quad \text{and} \quad G(x) = \frac{F_0(x-)}{q(F_0(x-))},$$

whence

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \sum_{k=1}^n Y_k \left(\frac{1 - F_0(X_k)}{q(F_0(X_k))} + \frac{F_0(X_k-)}{q(F_0(X_k-))} \right)$$

$$\preceq_{\mathcal{H}_+^2} \sum_{k=1}^n \frac{Y_k}{q(F_0(X_k))}. \tag{7.4}$$

Let us now give a relevant example. We assume that X_1, \dots, X_n are n independent copies of a random variable U distributed uniformly on $[0, 1]$. Let $F_0 = F_U$ be the distribution function of U and let $q(t) = \sqrt{\max(t, \delta)}$ for some $0 < \delta < 1$. Then (7.4) gives

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^2] \leq \frac{1}{2} \sum_{k=1}^n \mathbb{E} \left[\frac{Y_k^2}{\max(X_k, \delta)} \right] = \frac{1}{2} s^2 \log \left(\frac{e}{\delta} \right).$$

7.1. Case $G = 0$

Here $W_k = (Y_k)_+ H(X_k)$. Since Y_k is symmetric and independent of X_k , $\varepsilon_k Q_{\varepsilon_k}(U_k/2)$ and $Y_k H(X_k)$ have the same distribution. Then

$$Z - \mathbb{E}[Z] \preceq_{\mathcal{H}_+^2} \sum_{k=1}^n Y_k H(X_k), \quad \text{and} \quad \sigma^2 = \sum_{k=1}^n \mathbb{E}[Y_k^2] \mathbb{E}[H^2(X_k)]. \tag{7.5}$$

7.1.1. Chebyshev type inequality

Proposition 7.2. For any positive x ,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \sigma x) \leq \min\left(\frac{1}{1+x^2}, \frac{1}{2x^2}\right). \tag{7.6}$$

Proof. From (7.5) and (3.1), we derive

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq P_2(\xi; x) = \inf_{t < x} \frac{\mathbb{E}[(\xi - t)_+^2]}{(x - t)^2}, \tag{7.7}$$

where $\xi := \sum_{k=1}^n Y_k H(X_k)$. Since $z_+^2 \leq z^2$ for all real z , we obtain the Cantelli inequality $P_2(\xi; x) \leq \sigma^2/(\sigma^2 + x^2)$. Moreover, taking $t = 0$ in (7.7) gives the other bound $P_2(\xi; x) \leq \sigma^2/(2x^2)$. \square

7.1.2. $H = 1$ and Gaussian case

Proceeding as in the proof of Proposition 7.2 and adding the bound provided by (3.2), we obtain the following inequality.

Proposition 7.3. Let g be a standard Gaussian random variable and suppose that Y_1, \dots, Y_n is a sequence of independent centered Gaussian random variables. Let σ denote the standard deviation of $\sum_{k=1}^n Y_k$. Then for any positive x ,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq \sigma x) \leq \min\left\{\frac{1}{1+x^2}, \frac{1}{2x^2}, \frac{e^2}{2}\mathbb{P}(g \geq x)\right\} := h(x). \tag{7.8}$$

Remark 7.4. Note that $h(x) = 1/(1+x^2)$ for any $0 < x \leq 1$, $h(x) = 1/2x^2$ for any $1 < x \leq x_0$ and $h(x) = (e^2/2)\mathbb{P}(g \geq x)$ for any $x > x_0$, where x_0 is the unique root of the equation $(e^2/2)\mathbb{P}(g \geq x) = 1/2x^2$. A numerical calculation gives $x_0 \approx 1.6443$. Furthermore the function h is always better than the usual exponential bound (i.e., $h(x) \leq \exp(-x^2/2)$).

7.1.3. $0 \leq H \leq 1$ and Gaussian case

Here we suppose that X_1, \dots, X_n are identically distributed according to some distribution P . Let g_1, \dots, g_n be an independent sequence of standard Gaussian random variables and $\sigma_1, \dots, \sigma_n$ be a sequence of positive deterministic reals and set $Y_k = \sigma_k g_k$.

Proposition 7.5. Let $v := \mathbb{E}[H^2(X_1)]$ and let γ be the function defined on $(0, \infty)$ by

$$\gamma(x) := x\sqrt{2}/\sqrt{\log(1 + v^{-1}(e^x - 1))}.$$

Then for any positive x ,

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > \frac{\|\sigma\|_2^2}{\|\sigma\|_\infty} \gamma\left(\frac{\|\sigma\|_\infty^2 x}{\|\sigma\|_2^2}\right)\right) \leq \exp(-x). \tag{7.9}$$

Remark 7.6. As x goes to zero, the function γ has the asymptotic expansion

$$\gamma(x) = \sqrt{2vx}(1 + O(x)),$$

and as x goes to infinity, $\gamma(x) \sim \sqrt{2x}$.

Proof of Proposition 7.5. Starting as in the proof of Proposition 5.4 and conditioning by X_k , one has

$$\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \sum_{k=1}^n \log \mathbb{E} \left[\exp \left(\frac{\sigma_k^2 t^2}{2} H^2(X_k) \right) \right].$$

Define next the function ℓ_v by

$$\ell_v(t) := \log(1 + v(e^{t^2/2} - 1)). \tag{7.10}$$

Now by the convexity of the function $\lambda \mapsto e^{\alpha\lambda}$,

$$\log \left(\mathbb{E} \left[\exp \left(\frac{\sigma_k^2 t^2}{2} H^2(X_k) \right) \right] \right) \leq \ell_v(\sigma_k t). \tag{7.11}$$

In order to bound up the right-hand side term, we will use the property below concerning ℓ_v .

Lemma 7.7. *Let h_v be the function defined by $h_v(t) := \ell'_v(t)/t$ for any positive t . Then h_v is nondecreasing.*

Proof of Lemma 7.7. A straightforward calculation leads to

$$\ell''_v(t) = \ell'_v(t) \left(\frac{1}{t} + \frac{t(1-v)}{1 + v(e^{t^2/2} - 1)} \right).$$

Then

$$h'_v(t) = \frac{t\ell''_v(t) - \ell'_v(t)}{t^2} = \ell'_v(t) \left(\frac{1-v}{1 + v(e^{t^2/2} - 1)} \right).$$

Since $v \leq 1$, we get $h'_v(t) \geq 0$ and the lemma follows. □

Next, from (7.11) and Lemma 7.7, proceeding exactly as in Bercu, Delyon and Rio [8] (see Inequality (2.97)), one has

$$\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \frac{\|\sigma\|_2^2}{\|\sigma\|_\infty^2} \ell_v(\|\sigma\|_\infty t).$$

From the inversion formula for ℓ_v^* given in [8] (see Exercise 1, page 57)

$$\ell_v^{*-1}(x) = \inf\{t^{-1}(\ell_v(t) + x) : t > 0\}, \tag{7.12}$$

it is easy to see that for any positive x ,

$$\inf_{t>0} \frac{1}{t} \left(\frac{\|\sigma\|_2^2}{\|\sigma\|_\infty^2} \ell_v(\|\sigma\|_\infty t) + x \right) = \frac{\|\sigma\|_2^2}{\|\sigma\|_\infty} \ell_v^{*-1} \left(\frac{\|\sigma\|_\infty^2 x}{\|\sigma\|_2^2} \right). \tag{7.13}$$

Then (see Lemma 2.7 of [8]),

$$\mathbb{P} \left(Z - \mathbb{E}[Z] > \frac{\|\sigma\|_2^2}{\|\sigma\|_\infty} \ell_v^{*-1} \left(\frac{\|\sigma\|_\infty^2 x}{\|\sigma\|_2^2} \right) \right) \leq \exp(-x). \tag{7.14}$$

However, it seems difficult to calculate the inverse function of ℓ_v^* . Then to obtain a “ready-to-use” inequality, we will bound up $\ell_v^{*-1}(x)$.

Let $t_x := \sqrt{2 \log(1 + v^{-1}(e^x - 1))}$. Hence, putting t_x in (7.12), we get $\ell_v^{*-1}(x) \leq \gamma(x)$ and the proposition follows. □

7.1.4. Unbounded function H

In the following result, we suppose that Y_k and $H(X_k)$ are \mathbb{L}^r -integrable random variables with $2 < r \leq 4$.

Proposition 7.8. *Let $2 < r \leq 4$. Then*

$$\|(Z - \mathbb{E}[Z])_+\|_r^r \leq \frac{1}{2} \sum_{k=1}^n \|Y_k\|_r^r \|H(X_k)\|_r^r + \frac{1}{2} \sigma^r \|g\|_r^r,$$

where g is a standard Gaussian random variable.

Proof. We already noticed in the proof of Proposition 5.1 that the symmetry of $Y_k H(X_k)$ allows us to write

$$\mathbb{E} \left[\left(\sum_{k=1}^n Y_k H(X_k) \right)_+^r \right] = \frac{1}{2} \sum_{k=1}^n \mathbb{E}[|Y_k H(X_k)|^r].$$

Then the result follows directly from Corollary 6.2 of Figiel et al. [13] (see also Theorem 6.1 and Theorem 7.1). □

7.2. Case $G \neq 0$

First, we present a duality formula for the r th moments of $\varepsilon_k Q_{W_k}(U_k/2)$ for $r \geq 2$. It will allow us to derive a simpler bound of these moments which we will use thereafter to obtain concentration inequalities.

7.2.1. Duality formula

Lemma 7.9. Let $r \geq 2$. One has

$$\begin{aligned} & \mathbb{E}[|\varepsilon_k Q_{W_k}(U_k/2)|^r] \\ &= \sup\{\mathbb{E}[|Y_k|^r (H^r(X_k)\mathbb{1}_A + G^r(X_k)\mathbb{1}_B)] : A, B \in \mathcal{F}, \mathbb{P}(A) + \mathbb{P}(B) = 1\}. \end{aligned} \tag{7.15}$$

Remark 7.10. The duality formula gives us directly a more tractable bound

$$\mathbb{E}[|\varepsilon_k Q_{W_k}(U_k/2)|^r] \leq \mathbb{E}[|Y_k|^r] \mathbb{E}[H^r(X_k) + G^r(X_k)]. \tag{7.16}$$

Proof of Lemma 7.9. Let us recall the general following fact. Let $\alpha \in (0, 1)$ and let θ_α be a Bernoulli random variable with parameter α . Let X be an integrable random variable. Then

$$\int_0^\alpha Q_X^r(u) du = \int_0^1 Q_X^r(u) Q_{\theta_\alpha}(u) du = \sup_\theta \mathbb{E}[X\theta],$$

where the supremum is taken over the set of all Bernoulli random variable with parameter α . Consequently,

$$\begin{aligned} \mathbb{E}[|\varepsilon_k Q_{W_k}(U_k/2)|^r] &= 2 \sup_C \{\mathbb{E}[W_k^r \mathbb{1}_C]\} \\ &= 2 \sup_C \{\mathbb{E}[|Y_k|^r (H^r(X_k)\mathbb{1}_{C \cap \{Y>0\}} + G^r(X_k)\mathbb{1}_{C \cap \{Y<0\}})]\}, \end{aligned}$$

where the suprema are taken over the set of all measurable set C such that $\mathbb{P}(C) = 1/2$, and (7.15) follows. □

7.2.2. Chebyshev type inequality

Proposition 7.11. Define

$$V = \sum_{k=1}^n \mathbb{E}[Y_k^2 (G \vee H)^2(X_k)], \quad \text{and} \quad V_1 = \sum_{k=1}^n \mathbb{E}[Y_k^2 (G^2(X_k) + H^2(X_k))].$$

Then for any positive x ,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq x) \leq \min\left(\frac{V}{V + x^2}, \frac{\sigma^2}{2x^2}\right) \tag{7.17}$$

$$\leq \min\left(\frac{V}{V + x^2}, \frac{V_1}{2x^2}\right). \tag{7.18}$$

Remark 7.12. Note that $V_1/2x^2 \leq V/(V + x^2)$ for all x such that

$$x^2 \geq \sum_{k=1}^n \mathbb{E}[Y_k^2 (G \vee H)(X_k)] \frac{\sum_{k=1}^n \mathbb{E}[Y_k^2 (G^2(X_k) + H^2(X_k))]}{\sum_{k=1}^n \mathbb{E}[Y_k^2 |G^2(X_k) - H^2(X_k)|]}.$$

Example 7.13. Let \mathcal{S} be a countable class of sets such that for any $S \in \mathcal{S}$, $\mathbb{P}(S) \leq p$. We consider the class of function $\mathcal{F} = \{1_S - \mathbb{P}(S) : S \in \mathcal{S}\}$. Here $H = 1$ and $G = p$. Proposition 7.11 yields for any positive x ,

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq sx) \leq \begin{cases} 1 & \text{if } x < x_0, \\ \frac{1+x^2}{1+p^2} & \text{if } x \geq x_0, \\ \frac{1}{2x^2} & \end{cases}$$

where $x_0^2 = (1 + p^2)/(1 - p^2)$.

Proof of Proposition 7.11. Since $\sigma^2 \leq V_1$ by Remark 7.12, (7.18) follows from (7.17). Let us prove now (7.17). Theorem 4 in Pinelis and Sakhnenko [24] (or Theorem 11.1 of Boucheron, Lugosi and Massart [10]) implies that $\text{Var}(Z) \leq V$ and the bound $V/(V + x^2)$ follows then from the Cantelli inequality. Proceeding as in the proof of Proposition 7.2, we get the bound $\sigma^2/2x^2$ which ends the proof. \square

7.2.3. *Moment inequality*

The following moment inequality is similar to Proposition 7.8. We assume that the random variables Y_k , $G(X_k)$ and $H(X_k)$ are \mathbb{L}^r -integrable with $2 < r \leq 4$.

Proposition 7.14. *Let $2 < r \leq 4$. Then*

$$\|(Z - \mathbb{E}[Z])_+\|_r^r \leq \frac{1}{2} \sum_{k=1}^n \mathbb{E}[|\varepsilon_k Q_{W_k}(U_k/2)|^r] + \frac{1}{2} \sigma^r \|g\|_r^r, \tag{7.19}$$

where g is a standard Gaussian random variable. Consequently, using Remark 7.10, we get

$$\|(Z - \mathbb{E}[Z])_+\|_r^r \leq \frac{1}{2} \sum_{k=1}^n \mathbb{E}[|Y_k|^r (H^r(X_k) + G^r(X_k))] + \frac{1}{2} \sigma^r \|g\|_r^r. \tag{7.20}$$

Example 7.13 (continued). *We derive immediately by Markov's inequality, for any $2 < r \leq 4$,*

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq sx) \leq \frac{1}{2} \left(\|g\|_r^r \left(\frac{\sqrt{1+p^2}}{x} \right)^r + \frac{1+p^r}{s^r x^r} \sum_{k=1}^n \mathbb{E}[Y_k^r] \right).$$

7.2.4. *Exponential inequality*

Proposition 7.15. *For any positive t ,*

$$\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \sum_{k=1}^n \log(\mathbb{E}[\cosh(tY_k H(X_k)) + \cosh(tY_k G(X_k)) - 1]).$$

Example 7.13 (continued). Here, we make the additional assumptions that the Y_k are standard Gaussian random variables. In this situation, the inequality above implies that

$$L(t) \leq n \log(e^{t^2/2} + e^{p^2 t^2/2} - 1) \leq n \ell_{1+p^2}(t),$$

where $\ell_v(t) := \log(1 + v(e^{t^2/2} - 1))$. Hence, for any positive x ,

$$\mathbb{P}\left(Z - \mathbb{E}[Z] \geq \frac{n \ell_{1+p^2}(t) + x}{t}\right) \leq \exp(-x). \tag{7.21}$$

Let $t_x := \sqrt{2 \log(1 + (1 + p^2)^{-1}(e^{x/n} - 1))}$. Then putting t_x in (7.21), we obtain for any positive x ,

$$\mathbb{P}\left(Z - \mathbb{E}[Z] \geq \frac{x \sqrt{2}}{\sqrt{\log(1 + (1 + p^2)^{-1}(e^{x/n} - 1))}}\right) \leq \exp(-x). \tag{7.22}$$

Proof of Proposition 7.15. Let L denote the logarithm of the Laplace transform of $Z - \mathbb{E}[Z]$. Next, applying (7.2) with $\varphi(x) = e^{tx}$, $t > 0$, we get

$$L(t) \leq \sum_{k=1}^n \log(\mathbb{E}[\exp(t \varepsilon_k Q_{W_k}(U_k/2))]). \tag{7.23}$$

Now, using Remark 7.10,

$$\begin{aligned} \mathbb{E}[\exp(t \varepsilon_k Q_{W_k}(U_k/2))] &= \sum_{j=0}^{\infty} \frac{t^{2j}}{(2j)!} \mathbb{E}[(Q_{W_k}(U_k/2))^{2j}] \\ &\leq \mathbb{E}[\cosh(t Y_k H(X_k)) + \cosh(t Y_k G(X_k)) - 1]. \end{aligned}$$

Putting then this inequality in (7.23) ends the proof. □

8. Chaos of order two

Let \mathcal{X} be a Polish space and \mathcal{F} be a countable class of measurable functions from \mathcal{X} into \mathbb{R} and let Γ be a subset of $\mathcal{F} \times \mathcal{F}$. Let $A = (a_{i,j})_{1 \leq i,j \leq n}$ be a symmetric real matrix with zero diagonal entries (i.e., $a_{i,i} = 0$ for all i) and let $\|\cdot\|$ denote the Hilbert–Schmidt norm which is $\|A\|_{\text{HS}} = \sqrt{\text{Tr}(A^T A)}$. Let X be a random variable with values in \mathcal{X} such that for any function f of \mathcal{F} , $f(X)$ is a centered random variable. Let X_1, \dots, X_n be n independent copies of X . Define now the function F by

$$Z := F(\tilde{X}_1, \dots, \tilde{X}_n) := \sup_{(f,g) \in \Gamma} \left\{ \sum_{1 \leq i < j \leq n} a_{ij} f(X_i) g(X_j) \right\}, \tag{8.1}$$

where $\tilde{X}_k := (f(X_k))_{f \in \mathcal{F}}$. We say that \mathcal{F} is a Vapnik–Červonenkis (VC for short) subgraph class if the collection of all subgraph of the functions in \mathcal{F} (i.e., the collection of sets $\{(x, s) \in \mathcal{X} \times \mathbb{R} : s < f(x)\}$ for $f \in \mathcal{F}$) forms a VC-class of sets in $\mathcal{X} \times \mathbb{R}$ (see, for instance, van der Vaart and Wellner [28]).

Proposition 8.1. *Let Z be defined by (8.1) and $p > 2$. Assume that \mathcal{F} is a VC-subgraph class of functions with square integrable envelope function Φ . Then there exists a constant $K(\mathcal{F})$ depending only on \mathcal{F} such that*

$$\begin{aligned} & \| (Z - \mathbb{E}[Z])_+ \|_p \\ & \leq (p - 1) \frac{\|A\|_{\text{HS}}}{\sqrt{2}} \|Q_{\Phi(X)}(U/2)\|_p^2 \left(\sqrt{1 + \frac{K(\mathcal{F})}{(p - 1)} \frac{\|\Phi(X)\|_2^2}{\|Q_{\Phi(X)}(U/2)\|_p^2}} \right), \end{aligned} \tag{8.2}$$

where U is a random variable distributed uniformly on $[0, 1]$.

Remark 8.2. See that

$$\|\Phi(X)\|_p \leq \|Q_{\Phi(X)}(U/2)\|_p \leq 2^{1/p} \|\Phi(X)\|_p. \tag{8.3}$$

Suppose now that $\Phi(X)$ is in \mathbb{L}^p for all $p > 2$ and $\|\Phi(X)\|_p$ tends to infinity as p tends to infinity. Then as p tends to infinity, we obtain the following behavior of the right-hand side of (8.2)

$$\begin{aligned} & (p - 1) \frac{\|A\|_{\text{HS}}}{\sqrt{2}} \|Q_{\Phi(X)}(U/2)\|_p^2 \left(\sqrt{1 + \frac{K(\mathcal{F})}{(p - 1)} \frac{\|\Phi(X)\|_2^2}{\|Q_{\Phi(X)}(U/2)\|_p^2}} \right) \\ & = (p - 1) \frac{\|A\|_{\text{HS}}}{\sqrt{2}} \|\Phi(X)\|_p^2 \left(1 + O\left(\frac{1}{p}\right) \right). \end{aligned} \tag{8.4}$$

Example 8.3. Here we assume that $\Phi \leq 1$ and we show how (8.2) can be used to obtain an exponential bound for the tail probability.

Let $p > 2$, $x \geq 0$, and define the function f_x on $(2, \infty)$ by

$$f_x(q) := \left((q - 1) C_x \|Q_{\Phi(X)}(U/2)\|_q^2 \left(\sqrt{1 + \frac{K(\mathcal{F})}{(q - 1)} \frac{\|\Phi(X)\|_2^2}{\|Q_{\Phi(X)}(U/2)\|_q^2}} \right) \right)^q,$$

where $C_x := \|A\|_{\text{HS}}/x\sqrt{2}$. Using (8.3) and $\Phi \leq 1$, one has

$$\begin{aligned} f_x(q) & \leq \left((q - 1) C_x 2^{2/q} \|\Phi\|_q^2 \left(\sqrt{1 + \frac{K(\mathcal{F})}{(q - 1)} \frac{\|\Phi\|_2^2}{\|\Phi\|_q^2}} \right) \right)^q \\ & \leq 4(C_x q \sqrt{1 + K(\mathcal{F})})^q := 4 \exp(h_x(q)). \end{aligned}$$

Clearly, $\inf_{q>0} h_x(q) = h_x(1/(eC_x\sqrt{1+K(\mathcal{F})}))$. Now, by Markov's inequality, for any $x \geq 0$ such that $2eC_x\sqrt{1+K(\mathcal{F})} \leq 1$,

$$\begin{aligned} \mathbb{P}(Z - \mathbb{E}[Z] \geq x) &\leq \inf_{p>2} \frac{\mathbb{E}[(Z - \mathbb{E}[Z])_+^p]}{x^p} \\ &\leq 4 \exp\left(-\frac{1}{eC_x\sqrt{1+K(\mathcal{F})}}\right). \end{aligned}$$

Proof of Proposition 8.1. For any m and l belonging to $\{1, \dots, n\}$, we set $S_k(l, m) := \sup_{f \in \mathcal{F}} |\sum_{i=l}^m a_{ik} f(X_i)|$. Noting that

$$|Z - Z^{(k)}| \leq \Phi(X_k)(S_k(1, k-1) + S_k(k+1, n)),$$

it follows from Theorem 2.3 that

$$\|(Z - \mathbb{E}[Z])_+\|_p^2 \leq (p-1) \|Q_{\Phi(X)}(U/2)\|_p^2 \sum_{k=1}^n (\|S_k(1, k-1) + \mathbb{E}[S_k(k+1, n)]\|_p^2). \tag{8.5}$$

Define the function \tilde{F} such that $\tilde{Z} := \tilde{F}(\tilde{X}_1, \dots, \tilde{X}_{k-1}) := S(1, k-1)$. Define also for each $l \in \{1, \dots, k-1\}$,

$$\tilde{Z}^{(l)} := \tilde{F}(\tilde{X}_1, \dots, \tilde{X}_{l-1}, 0, \tilde{X}_{l+1}, \dots, \tilde{X}_{k-1}).$$

Hence, it follows that

$$|\tilde{Z} - \tilde{Z}^{(l)}| \leq |a_{lk}| \Phi(X_l).$$

Since $S_k(1, k-1) + \mathbb{E}[S_k(k+1, n)]$ is a nonnegative random variable, we can replace its p -norm in (8.5) by $\|(S_k(1, k-1) + \mathbb{E}[S_k(k+1, n)])_+\|_p$. Then an application of Theorem 2.1 leads to

$$\begin{aligned} &\|(S_k(1, k-1) + \mathbb{E}[S_k(k+1, n)])_+\|_p^2 \\ &\leq \left\| \left(\sum_{i=1}^{k-1} a_{ik} \varepsilon_i Q_{\Phi(X)}(U/2) + \mathbb{E}[S_k(1, k-1) + S_k(k+1, n)] \right) \right\|_p^2. \end{aligned} \tag{8.6}$$

Now by Corollary 4.9, Inequality (8.6) becomes

$$\begin{aligned} &\|S_k(1, k-1) + \mathbb{E}[S_k(k+1, n)]\|_p^2 \\ &\leq (p-1) \left(\sum_{i=1}^{k-1} a_{ik}^2 \right) \|Q_{\Phi(X)}(U/2)\|_p^2 + (\mathbb{E}[S_k(1, k-1) + S_k(k+1, n)])^2. \end{aligned} \tag{8.7}$$

Combining (8.5) and (8.7), we get

$$\begin{aligned} & \| (Z - \mathbb{E}[Z])_+ \|_p^2 \\ & \leq (p - 1)^2 \left(\sum_{1 \leq i < k \leq n} a_{ik}^2 \right) \| Q_{\Phi(X)}(U/2) \|_p^4 \\ & \quad + (p - 1) \| Q_{\Phi(X)}(U/2) \|_p^2 \sum_{k=1}^n (\mathbb{E}[S_k(1, k - 1) + S_k(k + 1, n)])^2. \end{aligned} \tag{8.8}$$

Let us now bound up $\mathbb{E}[S_k(1, k - 1)]$. Define the probability measure

$$\mathbb{P}_{k-1} = \sum_{i=1}^{k-1} a_{i,k}^2 \delta_{X_i}.$$

Exactly as in the proof of Theorem 2.5.2 in van der Vaart and Wellner [28], it can be shown that for some universal constant K ,

$$\mathbb{E}[S_k(1, k - 1)] \leq K \mathbb{E} \left[\int_0^1 \sqrt{\log N(\eta(\mathbb{P}_{k-1} \Phi^2)^{\frac{1}{2}}, \mathcal{F}, L_2(\mathbb{P}_{k-1}))} d\eta \times (\mathbb{P}_{k-1} \Phi^2)^{\frac{1}{2}} \right], \tag{8.9}$$

where for any semimetric space (T, d) , the covering number $N(\eta, T, d)$ is the minimal number of balls of radius η needed to cover T . Then, recalling that a VC-subgraph class satisfies the uniform entropy condition (see for instance [28], Theorem 2.6.7), there exists a constant $C(\mathcal{F})$ which depends only on \mathcal{F} such that

$$\mathbb{E}[S_k(1, k - 1)] \leq C(\mathcal{F}) \mathbb{E} \left[\left(\sum_{i=1}^{k-1} a_{i,k}^2 \Phi^2(X_i) \right)^{\frac{1}{2}} \right]. \tag{8.10}$$

Proceeding in the same way for $\mathbb{E}[S(k + 1, n)]$, we finally obtain that

$$\begin{aligned} & (\mathbb{E}[S_k(1, k - 1) + S_k(k + 1, n)])^2 \\ & \leq 4C^2(\mathcal{F}) \left(\mathbb{E} \left[\left(\sum_{i=1}^{k-1} a_{i,k}^2 \Phi^2(X_i) \right)^{\frac{1}{2}} + \left(\sum_{i=k+1}^n a_{i,k}^2 \Phi^2(X_i) \right)^{\frac{1}{2}} \right] \right)^2. \end{aligned} \tag{8.11}$$

Since $(\sqrt{x} + \sqrt{y})^2 \leq 2(x + y)$ for any nonnegative x and y , and $\sum_{k=1}^n \sum_{i \neq k} a_{ik} = 2 \times \sum_{1 \leq i < k \leq n} a_{ik}$, we then get by Jensen's inequality

$$(\mathbb{E}[S_k(1, k - 1) + S_k(k + 1, n)])^2 \leq 8C^2(\mathcal{F}) \left(\sum_{\substack{i=1 \\ i \neq k}}^n a_{ik}^2 \right) \mathbb{E}[\Phi^2(X)]. \tag{8.12}$$

Combining this inequality with (8.8), one has (8.2) which ends the proof. □

Remark 8.4. If we are concerned with

$$Z := \sup_{f \in \mathcal{F}} \left\{ \sum_{1 \leq i < j \leq n} a_{ij} f(X_i) f(X_j) \right\},$$

the same proof applies and we obtain exactly the same inequality (8.2).

9. Proofs of the results of Sections 2 and 4

9.1. Proofs of Section 4

Proof of Lemma 4.3. The case $q_0 \in \{0, 1\}$ is straightforward. We now turn to the case $q_0 \in (0, 1)$. By the definition of q_0 , it is clear that $\mathbb{E}[X] = \mathbb{E}[\zeta_{q_0}]$. We set in the following $a := a_{q_0}$ and $b := b_{q_0}$. To prove (4.4), we consider the following cases separately:

$$(i) \ t \leq a, \quad (ii) \ a < t < b, \quad (iii) \ t \geq b.$$

Case (i). Let $t \leq a$. Noting that $(y - t)_+ = (y - t) + (t - y)_+$, one has

$$\mathbb{E}[(\zeta_{q_0} - t)_+] = \mathbb{E}[X - t] + \mathbb{E}[(t - \eta)_+].$$

Hence, using the second inequality of (4.2), $\mathbb{E}[(X - t)_+] \leq \mathbb{E}[(\zeta_{q_0} - t)_+]$.

Case (ii). Let $a < t < b$. A direct calculation leads to

$$\mathbb{E}[(\zeta_{q_0} - t)_+] = q(b - t) + \mathbb{E}[(\psi - b)_+].$$

Let $c = (b - t)/(b - a)$, f and g be the functions defined by

$$\begin{aligned} f(x) &:= \max\{c(x - a)_+, (x - t)_+\}, \\ g(x) &:= ((x - t) - c(x - a))\mathbb{1}_{x \geq b} = (1 - c)(x - b)_+, \end{aligned}$$

for all $x \in \mathbb{R}$. Clearly $f(x) = c(x - a)_+ + g(x)$, whence

$$\mathbb{E}[(X - t)_+] \leq c\mathbb{E}[(X - a)_+] + \mathbb{E}[g(X)]. \tag{9.1}$$

Now by the Case (i),

$$\mathbb{E}[(X - a)_+] \leq q(b - a) + \mathbb{E}[(\psi - b)_+], \tag{9.2}$$

and by the first inequality of (4.2),

$$\mathbb{E}[g(X)] \leq \mathbb{E}[g(\psi)] = (1 - c)\mathbb{E}[(\psi - b)_+]. \tag{9.3}$$

Case (ii) follows then from combining (9.1)–(9.3).

Case (iii). Let $t \geq b$. Clearly, $\mathbb{E}[(\zeta_{q_0} - t)_+] = \mathbb{E}[(\psi - t)_+]$ and the first inequality of (4.2) gives the desired inequality.

The proof of (4.4) is completed. The extension (4.5) to convex functions is classical (see, for example, Proposition 3 in Bentkus [4] or the proof of Theorem 3.3 in Klein, Ma and Privault [14]). \square

Proof of Lemma 4.6. For every real t , define the function g_t on $(0, 1)$ by $g_t(q) := \mathbb{E}[(\zeta_q - \mathbb{E}[\zeta_q] - t)_+^2]$. Let (C_q) denote the condition

$$(C_q): \quad 2\mathbb{E}[\zeta_q] - a_q - b_q \geq 0.$$

Remark that the left-hand side of (C_q) is nondecreasing in q and tends to a positive value as q tends to 1. Hence $\tilde{q} := \inf\{q \geq 1/2 : b_q + a_q \leq 2\mathbb{E}[\zeta_q]\}$ exists, $(C_{\tilde{q}})$ is true and for any $q \geq \tilde{q}$, (C_q) is also verified.

In the following, we link the sign of $g'_t(q)$ with the verification of the condition (C_q) . Now,

$$\begin{aligned} g'_t(q) = & - (a_q - \mathbb{E}[\zeta_q] - t)_+^2 - 2(b_q - a_q) \int_0^{1-q} (F_\eta^{-1}(u) - \mathbb{E}[\zeta_q] - t)_+ du \\ & + (b_q - \mathbb{E}[\zeta_q] - t)_+^2 - 2(b_q - a_q) \int_{1-q}^1 (F_\psi^{-1}(u) - \mathbb{E}[\zeta_q] - t)_+ du. \end{aligned} \tag{9.4}$$

We consider the following cases separately:

- (i) $t + \mathbb{E}[\zeta_q] \geq F_\psi^{-1}(u)$ for all $u \in (0, 1)$,
- (ii) $t + \mathbb{E}[\zeta_q] \leq F_\eta^{-1}(u)$ for all $u \in (0, 1)$,
- (iii) $a_q \leq t + \mathbb{E}[\zeta_q] \leq b_q$,
- (iv) $b_q < t + \mathbb{E}[\zeta_q] < F_\psi^{-1}(1-)$,
- (v) $F_\eta^{-1}(0+) < t + \mathbb{E}[\zeta_q] < a_q$.

Case (i). All the terms in the right-hand side of (9.4) are equal to zero.

Case (ii). In this case, ζ_q has a finite second moment and $g'_t(q) = d/dq \text{Var}(\zeta_q)$. Then, it is elementary to see that $g'_t(q) \leq 0$ if and only if (C_q) is true.

Case (iii). One has

$$g'_t(q) \leq (b_q - \mathbb{E}[\zeta_q] - t)(b_q - \mathbb{E}[\zeta_q] - t - 2q(b_q - a_q)). \tag{9.5}$$

The first factor of the right-hand side of (9.5) is nonnegative. Hence the right-hand side of (9.5) is negative if and only if, for all t in $[a_q - \mathbb{E}[\zeta_q], b_q - \mathbb{E}[\zeta_q]]$,

$$-2q(b_q - a_q) \leq -b_q + \mathbb{E}[\zeta_q] + t. \tag{9.6}$$

See now that the right-hand side of (9.6) is nondecreasing in t . It thus follows that $-2q(b_q - a_q) \leq -b_q + \mathbb{E}[\zeta_q] + t$, or equivalently $q \geq 1/2$, implies that $g'_t(q) \leq 0$.

Case (iv). One has directly in this case

$$g'_t(q) = -2(b_q - a_q) \int_{1-q}^1 (F_\psi^{-1}(u) - \mathbb{E}[\zeta_q] - t)_+ du \leq 0.$$

Case (v). Define

$$\varepsilon_t := \sup\{\theta \in (0, 1) : F_\eta^{-1}(\theta) \leq t + \mathbb{E}[\zeta_q]\}.$$

Then

$$g'_t(q) = (b_q - a_q)A_{q,t},$$

where

$$\begin{aligned} A_{q,t} &= b_q + a_q - 2\mathbb{E}[\zeta_q] - 2t \\ &\quad - 2 \int_{1-q}^1 (F_\psi^{-1}(u) - \mathbb{E}[\zeta_q] - t) du - 2 \int_{\varepsilon_t}^{1-q} (F_\eta^{-1}(u) - \mathbb{E}[\zeta_q] - t) du \\ &= b_q + a_q - 2\mathbb{E}[\zeta_q] + 2 \int_0^{\varepsilon_t} (F_\eta^{-1}(u) - \mathbb{E}[\zeta_q] - t) du \\ &\leq b_q + a_q - 2\mathbb{E}[\zeta_q]. \end{aligned}$$

We note that if (C_q) is true, then $A_{q,t} \leq 0$, whence $g'_t(q) \leq 0$.

Finally, if $q \geq 1/2$ and (C_q) is verified, then $g'_t(q) \leq 0$ and the proof of (i) is completed. Let us prove now (ii) starting with Lemma 4.3, we get

$$\mathbb{E}[(X - \mathbb{E}[X] - t)_+^2] \leq \mathbb{E}[(\zeta_{q_0} - \mathbb{E}[\zeta_{q_0}] - t)_+^2],$$

where q_0 is given by (4.3). In particular, $\mathbb{E}[\zeta_{q_0}] = \mathbb{E}[X] \geq 0$. Moreover, since $\eta = -\psi$, $\mathbb{E}[\zeta_{1/2}] = 0$. Recalling that $\mathbb{E}[\zeta_q]$ is nondecreasing with respect to q , it implies that $q_0 \geq 1/2$. Now, see that $a_{1/2} + b_{1/2} \leq 0 = \mathbb{E}[\zeta_{1/2}]$. Indeed, for any $u \in (0, 1)$,

$$\begin{aligned} F_\psi^{-1}(u) &\leq -\sup\{t \in \mathbb{R} : F_{-\psi}(-t) \leq 1 - u\} \\ &= -F_{-\psi}^{-1}((1 - u)_+) \\ &\leq -F_{-\psi}^{-1}(1 - u). \end{aligned}$$

Thus, $a_{1/2} + b_{1/2} = F_{-\psi}^{-1}(\frac{1}{2}) + F_\psi^{-1}(\frac{1}{2}) \leq 0$. By the point (i), we obtain (4.6), which concludes the proof. □

Proof of Proposition 4.8. Since $X \leq X_+$,

$$\|(X + Y)_+\|_p^2 \leq \|(X_+ + Y)_+\|_p^2 \leq \|X_+ + Y\|_p^2. \tag{9.7}$$

Moreover, under the same hypotheses of Proposition 4.8, one has the inequality $\|X + Y\|_p^2 \leq \|X\|_p^2 + (p - 1)\|Y\|_p^2$ as a corollary of Proposition 2.1 of Pinelis [19] (see also Lemma 2.4 of [12] and Proposition 2.1 of [26]). Combining this with (9.7) completes the proof. □

9.2. Proofs of Section 2

Proof of Theorem 2.1. Starting from (2.1) and projecting on \mathcal{F}_k , we obtain $-T_k \leq Z_k - \mathbb{E}_k[Z^{(k)}] \leq W_k$ almost surely. Moreover, recalling that the random variables X_k are centered and since F is separately convex, an application of Jensen's inequality ensures that $\mathbb{E}_{k-1}[Z_k - Z^{(k)}] \geq 0$. Thus, conditionally to \mathcal{F}_{k-1} , we can apply the second part of Lemma 4.6 with $X = Z_k - \mathbb{E}_k[Z^{(k)}]$ and $\psi = \xi_k$. Recalling (1.4) and Remark 4.7, it yields that for any function φ in \mathcal{H}_+^2 ,

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \leq \mathbb{E}[\varphi(\varepsilon_k Q_{\xi_k}(U_k/2))]. \tag{9.8}$$

We now prove (2.1) by induction on n . The case $n = 1$ is given by (9.8) with $k = 1$. Let $n > 1$ and assume that (2.3) holds for $n - 1$. We then have

$$\begin{aligned} \mathbb{E}[\varphi(Z - \mathbb{E}[Z])] &= \mathbb{E}[\mathbb{E}_{n-1}[\varphi(Z_{n-1} + \Delta_n)]] \\ &\leq \mathbb{E}[\varphi(Z_{n-1} + \varepsilon_n Q_{\xi_n}(U_n/2))] \\ &\leq \mathbb{E}\left[\varphi\left(\sum_{k=1}^n \varepsilon_k Q_{\xi_k}(U_k/2)\right)\right], \end{aligned}$$

where we use (9.8) in the first inequality and the induction assumption in the second inequality. \square

Proof of Theorem 2.3. As in the proof of Theorem 2.1, we obtain for any function φ in \mathcal{H}_+^2 ,

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \leq \mathbb{E}[\varphi(\varepsilon_k \mathbb{E}_{k-1}[\psi_k] Q_{\xi_k}(U_k/2))]. \tag{9.9}$$

We now prove (2.6) by induction on n . For $n = 1$, it follows from (9.9) for $k = 1$ and (4.8). Let $n > 1$ and assume that (2.6) holds for $n - 1$. Then

$$\begin{aligned} \|(Z - \mathbb{E}[Z])_+\|_p^2 &\leq \|(Z_{n-1} + \varepsilon_n \mathbb{E}_{n-1}[\psi_n] Q_{\xi_n}(U_n/2))_+\|_p^2 \\ &\leq \|(Z_{n-1})_+\|_p^2 + (p-1) \|\mathbb{E}_{n-1}[\psi_n] Q_{\xi_n}(U_n/2)\|_p^2 \\ &\leq (p-1) \sum_{k=1}^n \|\mathbb{E}_{k-1}[\psi_k]\|_p^2 \|Q_{\xi_k}(U_k/2)\|_p^2, \end{aligned}$$

where we use (9.9) in the first inequality, Proposition 4.8 in the second inequality and the induction assumption in the third inequality. \square

Remark 9.1. See that, contrary to (9.8), there is a \mathcal{F}_{k-1} -measurable term in the expectation in the right-hand side of (9.9), which prevents us to proceed as in the proof of Theorem 2.1.

Acknowledgement

I deeply thank my advisor Emmanuel Rio for many fruitful discussions and ideas, and also for his constant support.

References

- [1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. [MR2424985](#)
- [2] Adamczak, R. and Wolff, P. (2015). Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields* **162** 531–586. [MR3383337](#)
- [3] Bentkus, V. (2004). On Hoeffding’s inequalities. *Ann. Probab.* **32** 1650–1673. [MR2060313](#)
- [4] Bentkus, V. (2008). Addendum to: “An extension of an inequality of Hoeffding to unbounded random variables” [Lith. Math. J. **48** (2008), no. 2, 137–157; [MR2425108](#)]: The non-i.i.d. case. *Lith. Math. J.* **48** 237–255. [MR2448328](#)
- [5] Bentkus, V. (2008). An extension of the Hoeffding inequality to unbounded random variables. *Lith. Math. J.* **48** 137–157. [MR2425108](#)
- [6] Bentkus, V. (2010). Bounds for the stop loss premium for unbounded risks under the variance constraints. Preprint. Available at <https://www.math.uni-bielefeld.de/sfb701/preprints/view/423>.
- [7] Bentkus, V., Kalosha, N. and van Zuijlen, M. (2006). On domination of tail probabilities of (super)martingales: Explicit bounds. *Lith. Math. J.* **46** 1–43. [MR2251440](#)
- [8] Bercu, B., Delyon, B. and Rio, E. (2015). *Concentration Inequalities for Sums and Martingales. SpringerBriefs in Mathematics*. Cham: Springer. [MR3363542](#)
- [9] Boucheron, S., Bousquet, O., Lugosi, G. and Massart, P. (2005). Moment inequalities for functions of independent random variables. *Ann. Probab.* **33** 514–560. [MR2123200](#)
- [10] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. [MR3185193](#)
- [11] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics. Wiley Series in Probability and Mathematical Statistics*. Chichester: Wiley. [MR1215046](#)
- [12] Dümbgen, L., van de Geer, S.A., Veraar, M.C. and Wellner, J.A. (2010). Nemirovski’s inequalities revisited. *Amer. Math. Monthly* **117** 138–160. [MR2590193](#)
- [13] Figiel, T., Hitczenko, P., Johnson, W.B., Schechtman, G. and Zinn, J. (1997). Extremal properties of Rademacher functions with applications to the Khintchine and Rosenthal inequalities. *Trans. Amer. Math. Soc.* **349** 997–1027. [MR1390980](#)
- [14] Klein, T., Ma, Y. and Privault, N. (2006). Convex concentration inequalities and forward-backward stochastic calculus. *Electron. J. Probab.* **11** 486–512. [MR2242653](#)
- [15] Kontorovich, A. (2014). Concentration in unbounded metric spaces and algorithmic stability. In *Proceeding of ICML* 28–36.
- [16] Lederer, J. and van de Geer, S. (2014). New concentration inequalities for suprema of empirical processes. *Bernoulli* **20** 2020–2038. [MR3263097](#)
- [17] Ledoux, M. (1995/1997). On Talagrand’s deviation inequalities for product measures. *ESAIM Probab. Stat.* **1** 63–87. [MR1399224](#)
- [18] Pinelis, I. (1994). On a majorization inequality for sums of independent random vectors. *Statist. Probab. Lett.* **19** 97–99. [MR1256696](#)
- [19] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. [MR1331198](#)

- [20] Pinelis, I. (1998). Optimal tail comparison based on comparison of moments. In *High Dimensional Probability (Oberwolfach, 1996)*. *Progress in Probability* **43** 297–314. Basel: Birkhäuser. [MR1652335](#)
- [21] Pinelis, I. (1999). Fractional sums and integrals of r -concave tails and applications to comparison probability inequalities. In *Advances in Stochastic Inequalities (Atlanta, GA, 1997)*. *Contemp. Math.* **234** 149–168. Providence, RI: Amer. Math. Soc. [MR1694770](#)
- [22] Pinelis, I. (2014). 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** 349–392.
- [23] Pinelis, I. (2016). Convex cones of generalized multiply monotone functions and the dual cones. *Banach J. Math. Anal.* **10** 864–897. [MR3555754](#)
- [24] Pinelis, I.F. and Sakhnenko, A.I. (1985). Remarks on inequalities for probabilities of large deviations. *Theory Probab. Appl.* **30** 143–148. [MR0779438](#)
- [25] Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. *Mathématiques & Applications (Berlin) [Mathematics & Applications]* **31**. Berlin: Springer. [MR2117923](#)
- [26] Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163. [MR2472010](#)
- [27] van de Geer, S. and Lederer, J. (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probab. Theory Related Fields* **157** 225–250. [MR3101846](#)
- [28] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. [MR1385671](#)

Received September 2016 and revised April 2017