



HAL
open science

Concentration inequalities for suprema of unbounded empirical processes

Antoine Marchina

► **To cite this version:**

Antoine Marchina. Concentration inequalities for suprema of unbounded empirical processes. 2017. hal-01545101v1

HAL Id: hal-01545101

<https://hal.science/hal-01545101v1>

Preprint submitted on 22 Jun 2017 (v1), last revised 9 Dec 2022 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Concentration inequalities for suprema of unbounded empirical processes

Antoine Marchina*

June 22, 2017

Abstract

Using martingale methods, we obtain some Fuk-Nagev type inequalities for suprema of unbounded empirical processes associated with independent and identically distributed random variables. We then derive weak and strong moment inequalities. Next, we apply our results to suprema of empirical processes which satisfy a power-type tail condition.

1 Introduction

Let us consider a sequence X_1, X_2, \dots of independent random variables valued in some measurable space $(\mathcal{X}, \mathcal{F})$. Let P_n denote for every integer n the empirical probability measure $P_n := n^{-1}(\delta_{X_1} + \dots + \delta_{X_n})$. Let \mathcal{F} be a countable class of measurable functions from \mathcal{X} into \mathbb{R} such that $\mathbb{E}[f(X_k)] = 0$ for all f in \mathcal{F} and all $k = 1, \dots, n$. We assume that \mathcal{F} has a square integrable envelope function Φ , that is

$$|f| \leq \Phi \text{ for any } f \in \mathcal{F}, \text{ and } \Phi \in \mathbb{L}^2. \quad (1.1)$$

As in Boucheron, Lugosi and Massart [4], we define the wimpy variance σ^2 and the weak variance Σ^2 by

$$\sigma^2 := \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[f^2(X_k)], \text{ and } \Sigma^2 := \mathbb{E} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^n f^2(X_k) \right]. \quad (1.2)$$

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

Let us also define

$$E_k := \mathbb{E} \sup_{f \in \mathcal{F}} P_k(f) \quad \text{for any } k \in \{1, \dots, n\}. \quad (1.3)$$

The purpose of this paper is to provide concentration inequalities around its mean for the random variable

$$Z := \sup\{nP_n(f) : f \in \mathcal{F}\}, \quad (1.4)$$

involving σ^2 , and under the additional assumption of identically distributed data. Our approach is based on a decomposition of Z into a sum of two martingales. Then, we control each martingale separately by Fuk-Nagaev type inequalities : in a first part, by one found by Courbot [7], which allows us to derive a strong moment inequality (following Petrov [14]), and in a second part, by one found recently by Rio [17], which allows us to derive a weak moment inequality. We stress out that we only require that the envelope function Φ has an ℓ th weak or strong moment, while classical concentration inequalities for suprema of empirical processes assume uniform boundedness of \mathcal{F} . Let us recall a main result in this direction : the following Bennett type inequality obtained by Bousquet [5], which is an improvement of Theorem 1.1 in Rio [16] :

Theorem 1.1 ([4], Theorem 12.5). *Let X_1, \dots, X_n be a sequence of independent random variables with values in \mathcal{X} and distributed according to P . Assume that $P(f) = 0$ and $f \leq 1$ for all $f \in \mathcal{F}$. Let Z be defined by (1.4) and set $v_n := n\sigma^2 + 2\mathbb{E}[Z]$, where σ^2 is defined in (1.2). Let h be the function defined, for any $u \geq -1$, by $h(u) := (1 + u) \log(1 + u) - u$. Then, for all $t \geq 0$,*

$$\begin{aligned} \mathbb{P}(Z - \mathbb{E}[Z] \geq t) &\leq \exp\left(-v_n h\left(\frac{t}{v_n}\right)\right) && \text{(Bousquet's inequality)} \\ &\leq \exp\left(-\frac{t}{2} \log\left(1 + \frac{t}{v_n}\right)\right). && \text{(Rio's inequality)} \end{aligned}$$

We refer the reader to Section 12 of [4] for an overview of the bounded case. Here we are interested in unbounded functions. Few results in the literature concern concentration inequalities for suprema of unbounded empirical processes. Let us first mention the considerable work of Boucheron, Bousquet, Lugosi and Massart [3], concerning moment inequalities for general functions of independent random variables. Their methods are based on an extension of the entropy method proposed by Ledoux [11]. In particular, they establish the following generalized moment inequality for suprema of (possibly unbounded) empirical processes involving σ^2 and Σ^2 :

Theorem 1.2 ([4], Theorems 15.14 and 15.5). *Let X_1, \dots, X_n be a sequence of independent random variables with values in \mathcal{X} . Assume that $\mathbb{E}[f(X_k)] = 0$ for all $f \in \mathcal{F}$ and all $k = 1, \dots, n$. Let Z be defined by*

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

Let $M := \max_{k=1, \dots, n} \Phi(X_k)$, where Φ is defined in (1.1). Then, for all $\ell \geq 2$,

$$\|(Z - \mathbb{E}[Z])_+\|_\ell \leq \sqrt{n\kappa(\ell - 1)(\Sigma + \sigma)} + \kappa(\ell - 1) \left(\|M\|_\ell + \sup_{\substack{f \in \mathcal{F} \\ k=1, \dots, n}} \|f(X_k)\|_2 \right),$$

where σ^2, Σ^2 are defined in (1.2) and $\kappa := \sqrt{e}/(\sqrt{e} - 1)$.

For several reasons (see, for instance, discussion after Theorem 3 in [1]) one would like to express the variance factor in terms of σ^2 rather than Σ^2 . First, observe that Σ^2 is greater than σ^2 . In the bounded case, an application of the contraction principle gives $n\Sigma^2 \leq n\sigma^2 + 16\mathbb{E}[Z]$, when $|f| \leq 1$ for any $f \in \mathcal{F}$ (see Corollary 15 in [13]). However, in the unbounded case, Σ^2 is more difficult to compare to σ^2 . In the setting of Theorem 1.2, one can only prove the much less efficient inequality

$$n\Sigma^2 \leq n\sigma^2 + 32\sqrt{\mathbb{E}[M^2]}\mathbb{E}[Z] + 8\mathbb{E}[M^2],$$

(see Theorem 11.17 and Section 15 in [4]). Similarly to the bounded case, the bounds that we will obtain in this paper will involve σ^2 , and the expectations E_k rather than the weak variance Σ^2 . Furthermore, we shall prove in a particular case that our bounds provide a much more accurate estimate of the variance.

Einmahl and Li [8] prove a Fuk-Nagaev type inequality for suprema of empirical processes involving σ^2 and the ℓ th strong moment of the envelope function. They use an improvement of Bousquet's inequality for suprema of bounded empirical processes to nonnecessarily identically distributed random variables obtained by Klein and Rio [9], a truncation argument and the so-called Hoffman-Jørgensen inequality. Using similar techniques, Adamczak [1] provides a concentration inequality for suprema of empirical processes under a semi-exponential tail condition on the envelope function Φ of \mathcal{F} :

Theorem 1.3 ([1], Theorem 4). *Let X_1, \dots, X_n be a sequence of independent random variables with values in \mathcal{X} . Assume that $\mathbb{E}[f(X_k)] = 0$ for all $f \in \mathcal{F}$ and all $k = 1, \dots, n$. For all $\alpha \in]0, 1]$, let ψ_α be the function defined, for any*

$x > 0$, by $\psi_\alpha(x) := \exp(x^\alpha) - 1$ and let $\|\cdot\|_{\psi_\alpha}$ denote the associated Orlicz norm, which is defined by

$$\|X\|_{\psi_\alpha} := \inf\{\lambda > 0 : \mathbb{E}[\psi_\alpha(|X|/\lambda)] \leq 1\}, \text{ for all random variable } X.$$

Assume now that for some $\alpha \in]0, 1]$, $\|\Phi(X_k)\|_{\psi_\alpha} < \infty$ for all $k = 1, \dots, n$, where Φ is defined in (1.1). Let Z be defined by $Z := \sup_{f \in \mathcal{F}} |\sum_{k=1}^n f(X_k)|$. Then, for all $0 < \eta < 1$ and $\delta > 0$, there exists a constant $C = C(\alpha, \eta, \delta)$, such that, for all $t \geq 0$,

$$\begin{aligned} & \mathbb{P}(Z - (1 + \eta)\mathbb{E}[Z] \geq t) \\ & \leq \exp\left(-\frac{t^2}{2(1 + \delta)n\sigma^2}\right) + 3 \exp\left(-\left(\frac{t}{C\|\max_{k=1, \dots, n} \Phi(X_k)\|_{\psi_\alpha}}\right)^\alpha\right). \end{aligned}$$

Let us point out that the upper bound in the inequality above (and also in [8]) do not involve $\mathbb{E}[Z]$ or the entropy of the class \mathcal{F} . The price to be paid is the additional factor $1 + \eta$ in front of $\mathbb{E}[Z]$ and the non explicit constant $C(\alpha, \eta, \delta)$. More recently, van de Geer and Lederer [18] introduce a new Orlicz norm (called Bernstein-Orlicz norm), and under some Bernstein conditions satisfied by the envelope function Φ , they derive exponential inequalities. Their upper bounds involve the constant K in the Bernstein conditions and $\mathbb{E}[Z]$ (which is bounded up in terms of complexity of \mathcal{F} and K). Next, the same authors in [10], require only that the envelope function Φ has an ℓ th strong moment and obtained deviation and moment inequalities involving σ^2 . However, it concerns the deviation of Z around $(1 + \eta)\mathbb{E}[Z]$. Finally, Marchina [12] provides deviation inequalities around $\mathbb{E}[Z]$ for suprema of randomized unbounded empirical processes involving only the envelope function Φ , see for example the following Proposition :

Proposition 1.4 ([12], Proposition 7.14). *Let X_1, \dots, X_n be a sequence of independent random variables with values in \mathcal{X} . Let 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 functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $-G \leq f \leq H$ for all $f \in \mathcal{F}$, where G and H are nonnegative functions. Define $Z = \sup_{f \in \mathcal{F}} \sum_{k=1}^n Y_k f(X_k)$. Let*

$$s_k^2 := \mathbb{E}[Y_k^2] \mathbb{E}[H^2(X_k) + G^2(X_k)] \text{ and } s^2 := \sum_{k=1}^n s_k^2.$$

Then, for any $2 \leq \ell \leq 4$,

$$\|(Z - \mathbb{E}[Z])_+\|_\ell^\ell \leq \frac{1}{2} \sum_{k=1}^n \mathbb{E}\left[|Y_k|^\ell (H^\ell(X_k) + G^\ell(X_k))\right] + \frac{1}{2} s^\ell \|g\|_\ell^\ell,$$

where g is a standard Gaussian random variable.

The results of [12] are based on martingale techniques. The purpose of the present paper is to introduce the wimpy variance in the concentration inequalities derived from the martingale approach. We shall give deviation inequalities around $\mathbb{E}[Z]$, without extra centering term $\eta\mathbb{E}[Z]$ and with explicit constants.

The paper is organized as follows : we first recall some definitions and notations in Section 2. In Section 3, we state Fuk-Nagaev type inequalities for $Z - \mathbb{E}[Z]$ and the resulting corollaries concerning the weak and strong moments of order $\ell > 2$. We shall also apply the Fuk-Nagaev inequalities to bound up the generalized moment $\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+]$. Finally, in Section 4, we apply the main results to the special case $Z = \sup_{g \in \mathcal{G}} \sum_{k=1}^n Y_k g(X_k)$ where Y_k satisfies a power-type tail condition and \mathcal{G} is a class of bounded functions.

2 Definitions and notations

In this section, we give the notations and definitions which we will use all along the paper. Let us start with the classical notations $x_+ := \max(0, x)$ and $x_+^\alpha := (x_+)^alpha$ for all real x and α . Next, we define the tail function, the quantile function and the Conditional Value-at-Risk.

Definition 2.1. *Let X be a real-valued random variable.*

(i) *The distribution function of X is denoted by F_X and the càglàd inverse of F_X is denoted by F_X^{-1} .*

(ii) *The quantile function of X , which is the càdlàg inverse of the tail function $t \mapsto 1 - F_X(t)$, is denoted by Q_X .*

(iii) *Assume that X is integrable. The integrated quantile function \tilde{Q}_X of X , which is also known as the Conditional Value-at-Risk (CVaR for short), is defined by $\tilde{Q}_X(u) := u^{-1} \int_0^u Q_X(s) ds$.*

We recall the following elementary properties of these quantities, which are given and proved by Pinelis [15].

Proposition 2.2. *Let X and Y be real-valued and integrable random variables. Then, for any $u \in]0, 1]$,*

- (i) $\mathbb{P}(X > Q_X(u)) \leq u$,
- (ii) $Q_X(u) \leq \tilde{Q}_X(u)$,
- (iii) $\tilde{Q}_{X+Y}(u) \leq \tilde{Q}_X(u) + \tilde{Q}_Y(u)$.

Let us now define the following class of distribution functions.

Notation 2.3. Let $q \in [0, 1]$. Let ψ be a nonnegative random variable and set $b_{\psi,q} := F^{-1}(1 - q)$. We denote by $F_{\psi,q}$ the distribution function defined by

$$F_{\psi,q}(x) := (1 - q)\mathbb{1}_{0 \leq x < b_{\psi,q}} + F_{\psi}(x)\mathbb{1}_{x \geq b_{\psi,q}}. \quad (2.1)$$

These distribution functions will be used to bound up the generalized moments of nonnegative random variables which are dominated by ψ . Precisely, let X be a nonnegative random variable stochastically dominated by ψ , that is $\mathbb{P}(X > x) \leq \mathbb{P}(\psi > x)$ for all $x > 0$. Let $\zeta_{\psi,q}$ be a random variable with distribution function $F_{\psi,q}$, where q is such that $\mathbb{E}[X] = \mathbb{E}[\zeta_{\psi,q}]$. Then Lemma 1 of Bentkus [2] (see also Lemma 2.1 of Marchina [12]) ensures that for any function $\varphi \in \mathcal{H}_+^1$, $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\zeta_{\psi,q})]$, where \mathcal{H}_+^1 is the class of numerical functions φ defined by

$$\mathcal{H}_+^1 := \left\{ \varphi : \varphi \text{ is convex, differentiable, and } \lim_{x \rightarrow -\infty} \varphi(x) = 0 \right\}. \quad (2.2)$$

Now, we recall the definitions of strong and weak norms of a real-valued random variable X . For all $r \geq 1$, let \mathbb{L}^r be the space of real-valued random variables with a finite absolute moment of order r and we denote by $\|X\|_r$ the \mathbb{L}^r -norm of X . Let

$$\Lambda_r^+(X) := \sup_{t>0} t (\mathbb{P}(X > t))^{1/r}. \quad (2.3)$$

We say that X have a weak moment of order r if $\Lambda_r^+(|X|)$ is finite. Define also

$$\tilde{\Lambda}_r^+(X) := \sup_{u \in]0,1]} u^{(1/r)-1} \int_0^u Q_X(s) ds. \quad (2.4)$$

From the definition of Q_X , we have

$$\Lambda_r^+(X) = \sup_{u \in]0,1]} u^{1/r} Q_X(u). \quad (2.5)$$

Hence, we get that

$$\Lambda_r^+(X) \leq \tilde{\Lambda}_r^+(X) \leq \left(\frac{r}{r-1}\right) \Lambda_r^+(X). \quad (2.6)$$

Furthermore, from Proposition 2.2 (iii), $\tilde{\Lambda}_r^+(\cdot)$ is sub-additive.

3 Statement of results

Let us first recall the assumptions we work with. Let X_1, \dots, X_n be a sequence of independent and identically random variables distributed according

to P with values in \mathcal{X} . Let \mathcal{F} be a countable class of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $P(f) = 0$ for all $f \in \mathcal{F}$, and we suppose that \mathcal{F} has a square integrable envelope function Φ defined in (1.1). In this situation, the wimpy variance is $\sigma^2 = \sup_{f \in \mathcal{F}} P(f^2)$. We consider the random variable

$$Z = \sup_{f \in \mathcal{F}} \sum_{k=1}^n f(X_k). \quad (3.1)$$

Throughout the rest of the paper, ζ_k denotes a random variable with distribution function $F_{2\Phi(X_1), q_k}$ defined in (2.1) where q_k is the real in $[0, 1]$ such that $\mathbb{E}[\zeta_k] = E_k$ (E_k is defined in (1.3)). We also set

$$V_n := \sum_{k=1}^n \mathbb{E}[\zeta_k^2]. \quad (3.2)$$

Remark 3.1. *If the class \mathcal{F} satisfies the uniform law of large numbers, that is $\sup_{f \in \mathcal{F}} |P_n(f)|$ converge to 0 in probability, then E_n decreases to 0 (see, for instance, Section 2.4 of van der Vaart and Wellner [19]). Now, from the integrability of Φ^2 and (2.1), the convergence of E_n to 0 implies the convergence of $\mathbb{E}[\zeta_n^2]$ to 0, which ensures that V_n/n tends to 0. More precise estimates of V_n will be proved for particular cases in Section 4.*

We first derive a Fuk-Nagaev type inequality for $Z - \mathbb{E}[Z]$ from one obtained by Courbot [7] concerning martingales.

Theorem 3.2. *Let $x > 0$. For any $s > 0$, we have*

$$\begin{aligned} \mathbb{P}((Z - \mathbb{E}[Z])_+ \geq x) &\leq \left(1 + \frac{x^2}{sn\sigma^2}\right)^{-s/2} + \left(1 + \frac{x^2}{sV_n}\right)^{-s/2} \\ &\quad + 2n \mathbb{P}\left(\Phi(X_1) \geq \frac{x}{2s}\right). \end{aligned}$$

Next, under weak moment conditions, we derive from a Fuk-Nagaev type inequality for martingales with efficient constants obtained recently by Rio [17], an other Fuk-Nagaev type inequality for $Z - \mathbb{E}[Z]$.

Theorem 3.3. *Let $\ell > 2$. Assume that $\Phi(X_1)$ have a weak moment of order ℓ . Then for any $u \in]0, 1[$,*

$$Q_{Z - \mathbb{E}[Z]}(u) \leq \tilde{Q}_{Z - \mathbb{E}[Z]}(u) \quad (a)$$

$$\leq \sqrt{2 \log(1/u)} (\sigma\sqrt{n} + \sqrt{V_n}) + 3n^{1/\ell} \mu_\ell \Lambda_\ell^+(\Phi(X_1)) u^{-1/\ell}, \quad (b)$$

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

$$\mathbb{P}\left(Z - \mathbb{E}[Z] > \sqrt{2 \log(1/u)} (\sigma\sqrt{n} + \sqrt{V_n}) + 3n^{1/\ell} \mu_\ell \Lambda_\ell^+(\Phi(X_1)) u^{-1/\ell}\right) \leq u. \quad (c)$$

In the two following results, we derive from Theorems 3.2 and 3.3, strong and weak moment inequalities for $Z - \mathbb{E}[Z]$.

Corollary 3.4. *Let $\ell > 2$. Assume that $\Phi(X_1)$ is \mathbb{L}^ℓ -integrable. Then*

$$\|(Z - \mathbb{E}[Z])_+\|_\ell \leq \ell^{1/\ell} \sqrt{\ell + 1} (\sigma\sqrt{n} + \sqrt{V_n}) + 2^{1+1/\ell} n^{1/\ell} (\ell + 1) \|\Phi(X_1)\|_\ell.$$

Remark 3.5. *Note that $\ell^{1/\ell} \leq e^{1/e} \simeq 1.4447$.*

Remark 3.6. *By analyzing the proofs of Theorem 3.2 and Corollary 3.4, we can slightly improve the constant $2^{1+1/\ell}$ to $(1 + 2^\ell)^{1/\ell}$.*

Corollary 3.7. *Let $\ell > 2$. Assume that $\Phi(X_1)$ have a weak moment of order ℓ . Then*

$$\Lambda_\ell^+(Z - \mathbb{E}[Z]) \leq \tilde{\Lambda}_\ell^+(Z - \mathbb{E}[Z]) \tag{a}$$

$$\leq \sqrt{(\ell/e)} (\sigma\sqrt{n} + \sqrt{V}) + 3n^{1/\ell} \mu_\ell \Lambda_\ell^+(\Phi(X_1)), \tag{b}$$

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

3.1 Bound of generalized moment of $Z - \mathbb{E}[Z]$

In this section, we apply Theorem 3.3 to bound up $\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+]$ for every $t > 0$. We emphasize that it is of interest to obtain such bounds in various situations coming from statistical applications, such the study of rates of convergence for estimators (see, for instance, Comte and Lacour [6]).

Proposition 3.8. *Let Z , σ , V_n be defined as in Section 3. Let $\ell > 2$ and $\mu_\ell = 2 + \max(4/3, \ell/3)$. Set also*

$$s_n := \sigma\sqrt{n} + \sqrt{V_n}, \text{ and } b_{n,\ell} := 3n^{1/\ell} \mu_\ell \Lambda_\ell^+(\Phi(X_1)).$$

Then, for any $t > 0$,

$$\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+] \leq s_n \frac{e^{-\frac{1}{2}(1+t^2/s_n^2)}}{\sqrt{1 + t^2/s_n^2}} + b_{n,\ell}.$$

Proof. Let us start by recalling the variational expression of $\mathbb{E}[(X - t)_+]$ involving \tilde{Q}_X . Since $x < Q_X(u)$ if and only if $1 - F_X(x) > u$, we get for any real t ,

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

Now, Inequality (3.3) and Theorem 3.3 (b) imply

$$\begin{aligned}\mathbb{E}[(Z - \mathbb{E}[Z] - t)_+] &\leq \sup_{u \in]0,1]} u \left(s_n \sqrt{2 \log(1/u)} + b_{n,\ell} u^{-1/\ell} - t \right) \\ &\leq \sup_{u \in]0,1]} u \left(s_n \sqrt{2 \log(1/u)} - t \right) + b_{n,\ell},\end{aligned}\quad (3.4)$$

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

$$y_0 := \frac{t}{2 s_n} + \sqrt{1 + \frac{t^2}{4 s_n^2}}.\quad (3.5)$$

Then, the supremum in (3.4) is equal to $s_n e^{-y_0^2/2}/y_0$. Observing now that $y_0 \geq \sqrt{1 + t^2/s_n^2}$, we finally get the desired inequality which concludes the proof. \square

Remark 3.9. *As starting point of the proof, in place of (3.3), we can use the equality $Z - \mathbb{E}[Z] \stackrel{D}{=} Q_{Z - \mathbb{E}[Z]}(U)$, where U is a random variable distributed uniformly on $[0, 1]$. However, contrary to the above proof, we then need to integrate Inequality (b) of Theorem 3.3, which shows the interest of the CVaR.*

3.2 Proofs of the main results

Our method is based on a martingale decomposition of Z which we now recall. We suppose that \mathcal{F} is a finite class of functions, that is $\mathcal{F} = \{f_i : i \in \{1, \dots, m\}\}$. The results in the countable case are derived from the finite case using the monotone convergence theorem. 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) denotes the conditional expectation operator associated with \mathcal{F}_k (resp. \mathcal{F}_n^k). Set also

$$Z^{(k)} := \sup\{nP_n(f) - f(X_k) : f \in \mathcal{F}\},\quad (3.6)$$

$$Z_k := \mathbb{E}_k[Z - \mathbb{E}[Z]].\quad (3.7)$$

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

$$Z - \mathbb{E}[Z] = \sum_{k=1}^n \Delta_k, \quad \text{where } \Delta_k := Z_k - Z_{k-1}.\quad (3.8)$$

Define now the stopping times τ and τ_k , respectively \mathcal{F}_n -measurable and \mathcal{F}_n^k -measurable, by

$$\tau := \inf\{i \in \{1, \dots, m\} : nP_n(f_i) = Z\}, \quad (3.9)$$

$$\tau_k := \inf\{i \in \{1, \dots, m\} : nP_n(f_i) - f_i(X_k) = Z^{(k)}\}. \quad (3.10)$$

Notice first that

$$Z^{(k)} + f_{\tau_k}(X_k) \leq Z \leq Z^{(k)} + f_{\tau}(X_k).$$

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

$$\mathbb{E}_k[f_{\tau_k}(X_k)] \leq Z_k - \mathbb{E}_k[Z^{(k)}] \leq \mathbb{E}_k[f_{\tau}(X_k)]. \quad (3.11)$$

Set now $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and let $\varepsilon_k \geq r_k \geq 0$ be random variables such that

$$\xi_k + r_k = Z_k - \mathbb{E}_k[Z^{(k)}] \quad \text{and} \quad \xi_k + \varepsilon_k = \mathbb{E}_k[f_{\tau}(X_k)].$$

Thus (3.11) becomes

$$\xi_k \leq \xi_k + r_k \leq \xi_k + \varepsilon_k. \quad (3.12)$$

Since the stopping time τ_k is \mathcal{F}_n^k -measurable, we have by the centering assumption on the elements of \mathcal{F} ,

$$\mathbb{E}_n^k[f_{\tau_k}(X_k)] = P(f_{\tau_k}) = 0, \quad (3.13)$$

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

$$\Delta_k = Z_k - \mathbb{E}_k[Z_k] - \mathbb{E}_{k-1}[Z_k - \mathbb{E}_k[Z_k]] = \xi_k + r_k - \mathbb{E}_{k-1}[r_k],$$

which, combined with (3.8), yields the decomposition of $Z - \mathbb{E}[Z]$ in a sum of two martingales :

$$Z - \mathbb{E}[Z] = \left(\sum_{k=1}^n \xi_k \right) + \left(\sum_{k=1}^n (r_k - \mathbb{E}_{k-1}[r_k]) \right). \quad (3.14)$$

Before proving the results, we provide bounds for their quadratic variations which will be needed in the proofs.

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

Notice that the same argument as (3.13) yields $\mathbb{E}_n^k[f_{\tau_k}^2(X_k)] = P(f_{\tau_k}^2)$. It follows from the conditional Jensen inequality that $\sum_{k=1}^n \mathbb{E}_{k-1}[\xi_k^2] \leq n\sigma^2$.

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

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

Lemma 3.10. *For any integer $j \geq k$, $\mathbb{E}_{k-1}[f_\tau(X_k)] = \mathbb{E}_{k-1}[f_\tau(X_j)]$.*

Proof. By the definition of the stopping time τ , for every permutation on n elements σ , $\tau(X_1, \dots, X_n) = \tau \circ \sigma(X_1, \dots, X_n)$ almost surely. Applying now this fact to $\sigma = (k j)$ (the transposition which exchanges k and j), it suffices to use Fubini's theorem (recalling that $j \geq k$) to complete the proof. \square

Hence,

$$\begin{aligned} \mathbb{E}_{k-1}[\varepsilon_k] &= \mathbb{E}_{k-1}[f_\tau(X_k)] \\ &= \mathbb{E}_{k-1}[f_\tau(X_k) + \dots + f_\tau(X_n)] / (n - k + 1) \\ &\leq \mathbb{E}_{k-1} \sup_{f \in \mathcal{F}} \{f(X_k) + \dots + f(X_n)\} / (n - k + 1) = E_{n-k+1}. \end{aligned} \quad (3.15)$$

Since $0 \leq r_k \leq \varepsilon_k$, we thus get that $0 \leq \mathbb{E}_{k-1}[r_k] \leq E_{n-k+1}$.

Moreover, (3.12) implies that $0 \leq r_k \leq 2\Phi(X_k)$. Then Lemma 1 of Bentkus [2] ensures that for any function $\varphi \in \mathcal{H}_+^1$, $\mathbb{E}[\varphi(r_k)] \leq \mathbb{E}[\varphi(\zeta_k)]$, where \mathcal{H}_+^1 is defined in (2.2). Notice that $x \mapsto x_+^2$ belongs to \mathcal{H}_+^1 and $r_{k+} = r_k$, whence,

$$\sum_{k=1}^n \mathbb{E}_{k-1}[(r_k - \mathbb{E}_{k-1}[r_k])^2] \leq \sum_{k=1}^n \mathbb{E}_{k-1}[r_{k+}^2] \leq \sum_{k=1}^n \mathbb{E}[\zeta_k^2].$$

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

Proof of Theorem 3.2. The key result is the following Fuk-Nagaev inequality for martingales obtained by Courbot :

Theorem 3.11 ([7], Theorem 1). *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 $\|\mathbb{E}[X_k^2 | \mathcal{F}_{k-1}]\|_\infty < \infty$. Define*

$$\langle M \rangle_n := \sum_{k=1}^n \mathbb{E}[X_k^2 | \mathcal{F}_{k-1}].$$

Then, for any $x, s, v > 0$,

$$\mathbb{P}(M_{n+} \geq x) \leq \sum_{k=1}^n \mathbb{P}(sX_{k+} > x) + \mathbb{P}(\langle M \rangle_n > v) + \exp\left(-\frac{s^2 v}{x^2} h\left(\frac{x^2}{sv}\right)\right),$$

where $h(x) = (1+x)\log(1+x) - x$.

We apply the above result to the two martingales in (3.14). The lower bound $h(x) \geq x \log(1+x)/2$ gives us the two first terms of the desired inequality. It then remains us to bound up the tail functions $\mathbb{P}((\xi_k)_+ > x)$ and $\mathbb{P}((r_k - \mathbb{E}_{k-1}[r_k])_+ > x)$, which is simply done using the fact that $\xi_k \leq \Phi(X_k)$ and $0 \leq r_k \leq 2\Phi(X_k)$. Moreover, observe that

$$\mathbb{P}(\Phi(X_1) > x/s) + \mathbb{P}(2\Phi(X_1) > x/s) \leq 2\mathbb{P}(\Phi(X_1) > x/2s),$$

which completes the proof. \square

Proof of Theorem 3.3. First observe that (a) is the property (ii) of Proposition 2.2 and that (c) follows immediately from (b) by the point (i) of the same Proposition 2.2. Let us now prove (b). Recalling the decomposition (3.14), the property (iii) of Proposition 2.2 imply

$$\tilde{Q}_{Z-\mathbb{E}[Z]}(u) \leq \tilde{Q}_{\sum_{k=1}^n \xi_k}(u) + \tilde{Q}_{\sum_{k=1}^n (r_k - \mathbb{E}_{k-1}[r_k])}(u). \quad (3.16)$$

Next, to control the terms in the right-hand side, the key result is the following new Fuk-Nagaev inequality obtained by Rio :

Theorem 3.12 ([17], Theorem 4.1). *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 for some constant $r > 2$,*

$$\|\mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}]\|_\infty < \infty \quad \text{and} \quad \left\| \sup_{t>0} \left(t^r \mathbb{P}(X_{k+} > t \mid \mathcal{F}_{k-1}) \right) \right\|_\infty < \infty.$$

Define

$$\sigma = \left\| \sum_{k=1}^n \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \right\|_\infty^{1/2} \quad \text{and} \quad C_r^w(M) = \left\| \sup_{t>0} \left(t^r \sum_{k=1}^n \mathbb{P}(X_{k+} > t \mid \mathcal{F}_{k-1}) \right) \right\|_\infty^{1/r}.$$

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

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

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

As in the proof of Theorem 3.2, we bound up ξ_k and $r_k - \mathbb{E}_{k-1}[r_k]$ respectively by $\Phi(X_k)$ and $2\Phi(X_k)$ to get

$$C_\ell^w \left(\sum_{k=1}^n \xi_k \right) + C_\ell^w \left(\sum_{k=1}^n (r_k - \mathbb{E}_{k-1}[r_k]) \right) \leq 3 n^{1/\ell} \Lambda_\ell^+(\Phi(X_1)). \quad (3.17)$$

Recalling the bounds of the quadratic variations of the two martingales that we found previously, we conclude then the proof by combining (3.16), Theorem 3.12 and (3.17). \square

Proof of Corollary 3.4. First, we have (see Petrov [14], p.61–62 and Exercice 2.26) that for any $\ell \geq 1$,

$$\mathbb{E}[(Z - \mathbb{E}[Z])_+^\ell] = \ell \int_0^\infty \mathbb{P}((Z - \mathbb{E}[Z])_+ \geq x) x^{\ell-1} dx. \quad (3.18)$$

Hence, using Theorem 3.2, we get

$$\begin{aligned} \mathbb{E}[(Z - \mathbb{E}[Z])_+^\ell] &\leq \frac{\ell}{2} s^{\ell/2} B\left(\frac{s-\ell}{2}, \frac{\ell}{2}\right) ((n\sigma^2)^{\ell/2} + (V_n)^{\ell/2}) \\ &\quad + 2n\ell \int_0^\infty x^{\ell-1} \mathbb{P}(\Phi(X_1) > x/2s) dx, \end{aligned} \quad (3.19)$$

where $B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$ is the usual Beta function. See now that for $\ell \geq 2$ and $s := \ell + 1$, we have

$$\frac{1}{2} B\left(\frac{s-\ell}{2}, \frac{\ell}{2}\right) = \frac{\sqrt{\pi}}{2} \frac{\Gamma(\ell/2)}{\Gamma((\ell+1)/2)} \leq 1.$$

Finally, we conclude the proof by the change of variables $x/2s = y$ in the integral term in (3.19) and the subadditivity of the function $x \mapsto x^{1/\ell}$. \square

Proof of Corollary 3.7. First, observe that (a) follows directly from (2.6). Let us now prove (b). We proceed exactly as in Rio [17, Theorem 5.1]. Both (2.4) and Theorem 3.3 (b) imply

$$\tilde{\Lambda}_\ell^+(Z - \mathbb{E}[Z]) \leq (\sigma\sqrt{n} + \sqrt{V_n}) \sup_{u \in]0,1[} \left(u^{1/\ell} \sqrt{2 \log(1/u)} \right) + 3n^{1/\ell} \mu_\ell \Lambda_\ell^+(\Phi(X_1)). \quad (3.20)$$

Next, observe that $u^{1/\ell} \sqrt{2 \log(1/u)} \leq \sqrt{(\ell/e)}$, which concludes the proof. \square

4 Application to power-type tail

Let Y_1, \dots, Y_n be a finite sequence of nonnegative, independent and identically distributed random variables and X_1, \dots, X_n a finite sequence of independent and identically distributed 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 . Let \mathcal{G} be a countable class of measurable functions from \mathcal{X} into $[-1, 1]$ such that for all $g \in \mathcal{G}$,

$$P(g) = 0 \quad \text{and} \quad P(g^2) < \delta^2 \quad \text{for some } \delta \in]0, 1[. \quad (4.1)$$

Let G be a measurable envelope function of \mathcal{G} that is

$$|g| \leq G \quad \text{for any } g \in \mathcal{G}, \quad \text{and } G(x) \leq 1 \quad \text{for all } x \in \mathbb{R}. \quad (4.2)$$

We suppose furthermore that for any positive t ,

$$\mathbb{P}(Y_1 > t) \leq t^{-p} \quad \text{for some } p > 2. \quad (4.3)$$

Define now

$$Z := \sup_{g \in \mathcal{G}} \sum_{k=1}^n Y_k g(X_k). \quad (4.4)$$

Setting $\tilde{X}_k := (X_k, Y_k)$ and \mathcal{F} the class of functions from $\mathcal{X} \times \mathbb{R}_+$ into \mathbb{R} which verified that for any $f \in \mathcal{F}$ there exists a unique $g \in \mathcal{G}$ such that $f(x, y) = yg(x)$, we then have $Z = \sup_{f \in \mathcal{F}} \sum_{k=1}^n f(\tilde{X}_k)$. Hence, this allows us to apply results of the previous section. The envelope function of \mathcal{F} is defined by $F(x, y) := yG(x)$. Moreover, we can obtain a more precise bound for V_n . Indeed, we will use an upper bound for the mean of the suprema of empirical processes, expressed in terms of the uniform entropy integral, proved by van der Vaart and Wellner [20]. Let us first recall some classical definitions.

Definition 4.1 (Covering number and uniform entropy integral). *The covering number $N(\epsilon, \mathcal{G})$ is the minimal number of balls of radius ϵ in $\mathbb{L}^2(Q)$ needed to cover the set \mathcal{G} . The uniform entropy integral is defined by*

$$J(\delta, \mathcal{G}) := \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|G\|_{Q,2}, \mathcal{G})} d\epsilon.$$

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

Throughout this section, K denotes an universal constant which may change from line to line.

Theorem 4.2. *Let Z be defined by (4.4). Under conditions (4.1) – (4.3), the following results hold :*

(i) *If Y_1 is \mathbb{L}^p -integrable, then*

$$\begin{aligned} & \|(Z - \mathbb{E}[Z])_+\|_p \\ & \leq p^{1/p} \sqrt{p+1} \left(\sigma \sqrt{n} + K \sqrt{\frac{p}{p-2}} \left(n^{q/4} \sqrt{J(\delta, \mathcal{G})} + \sqrt{p} n^{1/p} \left(\frac{J(\delta, \mathcal{G})}{\delta} \right)^{1/q} \right) \right) \\ & \quad + 2^{1+1/p} n^{1/p} (p+1) \|Y_1\|_p \|G(X_1)\|_p. \quad (a) \end{aligned}$$

(ii) *Moreover,*

$$\begin{aligned} & \tilde{\Lambda}_p^+(Z - \mathbb{E}[Z]) \\ & \leq \sqrt{(p/e)} \left(\sigma \sqrt{n} + K \sqrt{\frac{p}{p-2}} \left(n^{q/4} \sqrt{J(\delta, \mathcal{G})} + \sqrt{p} n^{1/p} \left(\frac{J(\delta, \mathcal{G})}{\delta} \right)^{1/q} \right) \right) \\ & \quad + 3 n^{1/p} \mu_p, \quad (b) \end{aligned}$$

where $q = p/(p-1)$ and $\mu_p = 2 + \max(4/3, p/3)$.

We now compare Inequality (a) above with results in the literature. Set $C_p := p^{1/p} \sqrt{p+1}$. Consider first the bounded case : $Y_k \leq 1$. Integrating the Rio inequality recalled in Theorem 1.1 and bounding up $\mathbb{E}[Z]$ by Proposition 4.5 on next page, one obtains

$$\|(Z - \mathbb{E}[Z])_+\|_p \leq C_p \left(\sigma \sqrt{n} + K \sqrt{\frac{p}{p-2}} \left(n^{1/4} \sqrt{J(\delta, \mathcal{G})} + \sqrt{p} \frac{J(\delta, \mathcal{G})}{\delta} \right) \right). \quad (4.5)$$

Remark 4.3. Note that when p tends to infinity, q tends to 1. This allows us to see (a) as an extension of (4.5) to the unbounded case.

Remark 4.4. In the unbounded case, Theorem 1.2 (of Boucheron & al. [3]) gives that

$$\|(Z - \mathbb{E}[Z])_+\|_p \leq B_p \sigma \sqrt{n} + o(\sqrt{n}), \quad (4.6)$$

where $B_p := 2(1 - e^{-1/2})^{-1/2} \sqrt{p-1}$. Remark that the constant C_p is always better than the constant B_p . For instance, for $p = 4$, $B_4 \simeq 5.5225$ and $C_4 \simeq 3.1623$. Furthermore, when p tends to infinity, B_p is equivalent to $3.1884 \sqrt{p}$ while C_p is equivalent to \sqrt{p} .

Proof of Theorem 4.2. First, we bound up the term $V_n = \sum_{k=1}^n \mathbb{E}[\zeta_k^2]$. We recall that ζ_k is a random variable with distribution function $F_{2Y_1G(X_1), q_k}$ (defined in (2.1)) and q_k is such that $\mathbb{E}[\zeta_k] = E_k$. Let ψ be a random variable with tail function defined by $\mathbb{P}(\psi > t) = t^{-p}$ for all $t \geq 1$ and let $\tilde{\zeta}_k$ be a random variable with distribution function F_{ψ, \tilde{q}_k} where \tilde{q}_k is the real in $[0, 1]$ such that $\mathbb{E}[\tilde{\zeta}_k] = E_k$. Clearly,

$$F_{2Y_1G(X_1), q_k}(x) \geq F_{\psi, \tilde{q}_k}(x) \text{ for any } x \in \mathbb{R}.$$

Then Lemma 1 of Bentkus [2] ensures that for any $\varphi \in \mathcal{H}_+^1$, $\mathbb{E}[\varphi(\zeta_k)] \leq \mathbb{E}[\varphi(\tilde{\zeta}_k)]$. In particular, this implies $\mathbb{E}[\zeta_k^2] \leq \mathbb{E}[\tilde{\zeta}_k^2]$. Therefore, an elementary calculation yields

$$V_n \leq 2^{-(1-1/p)} \frac{p}{p-2} \left(\frac{p-1}{p} \right)^{\frac{p-2}{p-1}} \sum_{k=1}^n E_k^{\frac{p-2}{p-1}}. \quad (4.7)$$

Next we show how we can obtain a bound for E_k in terms of uniform entropy integral.

Proposition 4.5. *There exists a universal constant K such that for any integer $k \geq 1$,*

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

Proof of Proposition 4.5. Let U_1, \dots, U_k be k independent copies of a random variable U distributed uniformly on $[0, 1]$ and let Q be the quantile function of ψ . Therefore $Q(u) = u^{-1/p}$ for any $u \in]0, 1[$. Note that (4.3) implies $Q_Y \leq Q$. Let also $\kappa \in \mathbb{R}$ such that

$$2^\kappa = \frac{k \delta^2}{J^2(\delta, \mathcal{G})}. \quad (4.8)$$

Let us now define for every $j = 1, \dots, \lceil \kappa \rceil$,

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

Here, $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote the classical floor and ceiling functions. We recall the basic property of the quantile function Q_X of a random variable X : $Q_X(U)$ has the same distribution as X for any random variable U with the uniform distribution over $[0, 1]$. Then,

$$\mathbb{E} \sup_{g \in \mathcal{G}} \left| \sum_{j=1}^k Y_j g(X_j) \right| \leq \mathbf{E}_1 + \mathbf{E}_2, \quad (4.9)$$

where

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

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

$$\mathbf{E}_2 \leq k \int_0^{2^{-\lceil \kappa \rceil}} Q(u) du \leq k \frac{p}{p-1} 2^{-\kappa(1-1/p)}. \quad (4.10)$$

To bound up \mathbf{E}_1 , we first notice that, since Q is decreasing, for any $m \in I_j$, $|Y_m g(X_m)| \leq Q(2^{-j})$. We can then apply Theorem 2.1 of Van der Vaart and Wellner [20] which leads to

$$\mathbf{E}_1 \leq K \left(J(\delta, \mathcal{G}) \sum_{j=1}^{\lceil \kappa \rceil} \mathbb{E} \left[|I_j|^{\frac{1}{2}} \right] Q(2^{-j}) + \frac{J^2(\delta, \mathcal{G})}{\delta^2} \sum_{j=1}^{\lceil \kappa \rceil} Q(2^{-j}) \right). \quad (4.11)$$

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 $\mathbb{E}[|I_j|^{\frac{1}{2}}] \leq \sqrt{k 2^{-j}}$. Now, recalling that $Q(u) = u^{-1/p}$,

$$\sum_{j=1}^{\lceil \kappa \rceil} 2^{-j/2} Q(2^{-j}) \leq \frac{2^{1/p-1/2}}{1 - 2^{1/p-1/2}} \leq \frac{2}{\log(2)} \frac{p}{p-2}. \quad (4.12)$$

Likewise,

$$\begin{aligned} \sum_{j=1}^{\lceil \kappa \rceil} Q(2^{-j}) &= 2^{\lceil \kappa \rceil/p} \left(2^{1/p} + \sum_{j=0}^{\lceil \kappa \rceil-1} 2^{-j/p} \right) \\ &\leq 2^{\lceil \kappa \rceil/p} \left(2^{1/p} + \frac{1}{1 - 2^{-1/p}} \right) \leq \frac{2^{\kappa/p}}{\log(2)} \frac{p^2}{p-2}. \end{aligned} \quad (4.13)$$

Hence, we derive from (4.11) – (4.13),

$$\mathbf{E}_1 \leq K \frac{p}{p-2} \left(\sqrt{k} J(\delta, \mathcal{G}) + p \frac{J^2(\delta, \mathcal{G})}{\delta^2} 2^{\kappa/p} \right). \quad (4.14)$$

Finally, both (4.9), (4.14), (4.10) and the definition of κ imply Proposition 4.5. \square

Let us continue the proof of Theorem 4.2. Using the subadditivity of the functions $x \mapsto x^a$ for $0 < a < 1$, from (4.7) and Proposition 4.5 we obtain that

$$\sqrt{V_n} \leq K \sqrt{\frac{p}{p-2}} \left(n^{q/4} \sqrt{J(\delta, \mathcal{G})} + \sqrt{p} n^{1/p} \left(\frac{J(\delta, \mathcal{G})}{\delta} \right)^{1/q} \right). \quad (4.15)$$

Injecting this bound into Corollary 3.4 gives (a). Similarly, injecting this bound in Inequality (b) of Corollary 3.7, we conclude the proof of (b) since $\Lambda_p^+(Y_1 G(X_1)) \leq \Lambda_p^+(\psi) = 1$. This ends the proof of Theorem 4.2. \square

Acknowledgments. The author would like to thank Emmanuel Rio for many helpful discussions and his careful reading of this work.

References

- [1] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.
- [2] V. Bentkus. An extension of the Hoeffding inequality to unbounded random variables. *Lithuanian Mathematical Journal*, 48(2):137–157, 2008.

- [3] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 03 2005.
- [4] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.
- [5] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6):495–500, 2002.
- [6] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2):569–609, 2013.
- [7] B. Courbot. Rates of convergence in the functional CLT for martingales. *Académie des Sciences Paris Comptes Rendus Serie Sciences Mathématiques*, 328:509–513, March 1999.
- [8] U. Einmahl and D. Li. Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.*, 360(12):6677–6693, 2008.
- [9] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 05 2005.
- [10] J. Lederer and S. van de Geer. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038, 11 2014.
- [11] M. Ledoux. On Talagrand’s deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1997.
- [12] A. Marchina. Concentration inequalities for separately convex functions. *To appear in Bernoulli Journal*, 2017+.
- [13] P. Massart. About the constants in Talagrand’s concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.
- [14] V.V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford science publications. Clarendon Press, 1995.
- [15] I. Pinelis. 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(3):349–392, 2014.

- [16] E. Rio. Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):1053–1057, 2002. En l’honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.
- [17] E. Rio. About the constants in the Fuk-Nagaev inequalities. *Electron. Commun. Probab.*, 22:12 pp., 2017.
- [18] S. van de Geer and J. Lederer. The Bernstein-Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2):225–250, 2013.
- [19] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer New York, 1996.
- [20] A.W. van der Vaart and J.A. Wellner. A local maximal inequality under uniform entropy. *Electron. J. Statist.*, 5:192–203, 2011.