C. R. Acad. Sci. Paris, Ser. I 357 (2019) 537–544

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability theory/Statistics

An exponential inequality for suprema of empirical processes with heavy tails on the left

Une inégalité exponentielle pour les suprema de processus empiriques avec queues lourdes sur la gauche

Antoine Marchina [a,b]

[a] *Université Paris-Est, LAMA (UMR 8050), UPEM, UPEC, CNRS, 77454 Marne-la-Vallée, France*
[b] *Laboratoire de mathématiques de Versailles, UVSQ, CNRS, Université Paris-Saclay, 78035 Versailles, France*

A R T I C L E I N F O

Article history:
Received 11 October 2018
Accepted after revision 17 June 2019
Available online 3 July 2019

Presented by Michel Talagrand

A B S T R A C T

In this Note, we provide exponential inequalities for suprema of empirical processes with heavy tails on the left. Our approach is based on a martingale decomposition, associated with comparison inequalities over a cone of convex functions originally introduced by Pinelis. Furthermore, the constants in our inequalities are explicit.

R É S U M É

Dans cette Note, nous donnons des inégalités exponentielles pour les suprema de processus empiriques avec queues lourdes sur la gauche. Notre approche est basée sur une décomposition en martingale, associée à des inégalités de comparaison sur un cône de fonctions convexes, initialement introduit par Pinelis. Les constantes données sont explicites.

1. Introduction

Let $X, X_1, \dots, X_n$ be a sequence of independent random variables valued in some measurable space $(\mathcal{X}, \mathcal{F})$, and identically distributed according to a law P. Let $\mathcal{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $]-\infty, 1]$ such that $P(f) = 0$ for all $f \in \mathcal{F}$. Let $1 < p < 2$. We suppose that, for all $f \in \mathcal{F}$, $f(X)$ satisfies the following behavior on the left:

$$\mathbb{P}(f(X) \le -t) \le \left(\frac{c}{t}\right)^p \quad \text{for any } t > 0, \tag{1.1}$$

for some $c > 0$. Let Z_n be the real-valued random variable defined by

E-mail address: antoine.marchina@u-pem.fr.

https://doi.org/10.1016/j.crma.2019.06.008
1631-073X/

$$Z_n := \sup\left\{\sum_{k=1}^{n} f(X_k) : f \in \mathcal{F}\right\}. \tag{1.2}$$

The aim of this Note is to give an exponential bound for the probability of deviation of Z_n above its mean. Throughout the paper, we denote by η a Pareto (on the left) random variable with parameters (c, p), that is, its distribution function F_η is defined by

$$F_\eta(-t) = \left(\frac{c}{t}\right)^p \wedge 1 \ \text{and}\ F_\eta(t) = F_\eta(0) = 1,\ \text{for any}\ t > 0. \tag{1.3}$$

The control of the random fluctuations of an empirical process has a central role in mathematical statistics and machine learning. For general presentations of these connections between empirical process theory and statistics, see, for instance, the books of Van der Vaart and Wellner [24], Massart [15], or Koltchinskii [9]. For example, in the context of nonparametric estimation, the calibration of adaptive methods is strongly related to the control of an empirical process: see, for instance, [10] for the Empirical Risk Minimization method, [2] for the Cross-Validation method and [7] for the Goldenshluger–Lepski method.

Since Talagrand's [23] and Ledoux's [12] pioneering work, concentration inequalities for suprema of empirical processes have been the subject of intense research. Mainly, the aim is to reach optimal counterparts for Z_n of classical Hoeffding's, Bernstein's, and Bennett's exponential inequalities for sums of i.i.d. random variables, which correspond to classes $\mathcal{F}$ reduced to one element. The reader is referred to Chapter 12 of the book by Boucheron, Lugosi, and Massart [5] for an overview of this subject. Recently, some efforts have been made to consider heavier tails by only assuming that $\sup_{f\in\mathcal{F}} |f(X)|$ is $\mathbb{L}^r$-integrable for some $r > 2$: see, Boucheron, Bousquet, Lugosi, and Massart [4], Adamczak [1], van de Geer and Lederer [11], and Marchina [13]. However, the common point of all these results is that $\sup_{f\in\mathcal{F}} \mathrm{Var}(f(X))$ is finite, which we do not want to assume in this work. Instead, we assume (1.1) which states the first-order stochastic dominance of $-\eta$ over $-f(X)$, for all $f \in \mathcal{F}$. In other words, F_η is the extremal (in the sense of first-order stochastic dominance) distribution that $f(X)$ could have. The Pareto distribution is the prototype of "power-tailed" distributions, which are important particular cases of heavy-tailed distributions. More precisely, a distribution is said power-tailed if its tail function is of the form $x^{-\alpha}$, $\alpha > 0$, for large x. It has the singular property that every moment greater than the α-th is infinite (see, for instance, the book of Foss et al. [6]). Here, since we assume $1 < p < 2$, the first moment of η is finite, and the second one is infinite. In particular, combined with (1.1), it implies that $\mathrm{Var}(f(X))$ could be infinite for every $f \in \mathcal{F}$. Moreover, we emphasize that heavy-tailed data are commonly encountered, for example, in the areas of computer science, finance, biology, and astronomy, among others. Then the assumption (1.1) is also of interest for applications.

To the best of our knowledge, the only result without the assumption of square integrability of $f(X)$, $f \in \mathcal{F}$, is provided in Rio [20, Theorem 2]. The author gives an upper bound of the log-Laplace transform of $Z_n - \mathbb{E}[Z_n]$ involving squares of positive parts and truncated negative parts of $f(X)$ for all $f \in \mathcal{F}$. His proof relies on a martingale decomposition of $Z_n - \mathbb{E}[Z_n]$ associated with an exponential inequality for positive self-bounding functions (based on Ledoux's entropy method) proved in Rio [21]. Our approach uses the same martingale decomposition used by Rio. However, the difference lies in the control of the martingale increments. To give an upper bound on their log-Laplace transform, we resort to convex comparison inequalities, similar to those of Hoeffding [8] for bounded random variables. To our knowledge, this result has no counterpart in the existing literature. Then it is not that easy to study the optimality of the constants appearing in our inequalities. Nevertheless, we think that the method used may be encouraging for further works. Actually, it seems beyond the scope of traditional functional analysis tools to handle the case of nonfiniteness of $\sup_{f\in\mathcal{F}} \mathrm{Var}(f(X))$. Furthermore, we have recently shown that martingale methods can be used to relax classical hypotheses (as the uniform boundedness conditions) in concentration inequalities for separately convex functions of independent random variables, especially for suprema of empirical processes (see [14,13]).

2. Result

Let us first give some notation. For any real-valued random variable X, F_X and F_X^{-1} denote respectively the distribution function of X and the càdlàg inverse of F_X. For all reals x and α, $x_+ := \max(0, x)$ and $x_+^\alpha := (x_+)^\alpha$. We denote by $\Gamma(.)$ the usual gamma function.

For the sake of clarity, let us recall the setting we work with. Let $X, X_1, \ldots, X_n$ be a sequence of i.i.d. random variables valued in $(\mathcal{X}, \mathcal{F})$ with common distribution P. Let $1 < p < 2$. Let η be a random variable with distribution function F_η defined by (1.3). We recall that F_η depends on $c > 0$ and p. Let $\mathcal{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $\mathbb{R}$. We make the following assumptions.

Assumption 2.1. For all $f \in \mathcal{F}$ and all $x \in \mathcal{X}$,

$$f(x) \le 1 \quad \text{and} \quad P(f) := \mathbb{E}[f(X)] = 0. \tag{2.1}$$

Assumption 2.2. The parameter c is such that

$$c \geq \left(\frac{p-1}{2p-1}\right)^{1/p}. \tag{2.2}$$

This condition on c is purely technical. Note that the function $h : p \mapsto \left(\frac{p-1}{2p-1}\right)^{1/p}$ is increasing on $[1\,,2]$. Since $h(2) = \sqrt{1/3}$, if $c \geq \sqrt{1/3}$, then the condition (2.2) is satisfied.

Assumption 2.3. For any $t > 0$,

$$\mathbb{E}[(-t - f(X))_+] \leq \mathbb{E}[(-t-\eta)_+]. \tag{2.3}$$

This assumption corresponds to the second-order stochastic dominance of $-\eta$ over $-f(X)$, for all $f \in \mathcal{F}$. It is weaker than the condition (1.1), which corresponds to the first-order stochastic dominance, and is sufficient for our result.

The following result then holds.

Theorem 2.1. *Let Z_n be defined by (1.2). Let q_0 be the real in $]0\,,1[$ such that*

$$q_0 - c\,\frac{p}{p-1}(1-q_0)^{1-1/p} = 0. \tag{2.4}$$

Under Assumptions 2.1, 2.2 and 2.3, for any $x \leq q_0\,\Gamma(2-p)$,

$$\mathbb{P}(Z_n - \mathbb{E}[Z_n] \geq nx) \leq \exp\left(-n\gamma_p\, x^{p/(p-1)}(1-\varepsilon_p(x))\right), \tag{a}$$

where

$$\gamma_p = (p\alpha_p)^{-1/(p-1)}\,\frac{p-1}{p}, \qquad \alpha_p = \frac{c^p}{p-1}\Gamma(2-p), \tag{2.5}$$

$$\text{and } \varepsilon_p(x) = \frac{p}{p-1}\sum_{k=2}^{\infty}\frac{q_0}{k!}x^{(k-p)/(p-1)}(p\alpha_p)^{-(k-1)/(p-1)}. \tag{2.6}$$

Moreover, for any $x > q_0\,\Gamma(2-p)$,

$$\mathbb{P}(Z_n - \mathbb{E}[Z_n] \geq nx) \leq \exp\big(-n(x(1-q_0)^{1/p}c^{-1} - \beta_p)\big), \tag{b}$$

where

$$\beta_p = \frac{-\Gamma(2-p)}{p-1} - q_0\left(-\frac{\Gamma(2-p)}{p-1} - 1 + \exp((1-q_0)^{1/p}c^{-1}) - (1-q_0)^{1/p}c^{-1}\right). \tag{2.7}$$

Roughly speaking, Inequality (a) states that, for small enough $x > 0$,

$$\mathbb{P}(Z_n - \mathbb{E}[Z_n] \geq n^{1/p}x) \leq \exp\left(-K_p\, x^q\big(1 + O\big(x^{(2-p)/(p-1)}n^{-2/p}\big)\big)\right), \tag{2.8}$$

where $q = p/(p-1)$ is the Hölder exponent conjugate of p and K_p is a constant depending only on p. Under Assumption 2.1 and $\sup_{f\in\mathcal{F}} \mathrm{Var}(f(X)) < \infty$, it is known that Z_n satisfies a Bennett-type inequality (see, for instance, Rio [22, Theorem!1.1]). It leads to the following inequality for small enough $x > 0$:

$$\mathbb{P}(Z_n - \mathbb{E}[Z_n] \geq \sqrt{n}x) \leq \exp\left(-\frac{x^2}{2v}\big(1 + O\big(xn^{-1/2}\big)\big)\right), \tag{2.9}$$

where $v := \sigma^2 + 2n^{-1}\mathbb{E}[Z_n]$ and $\sigma^2 := \sup_{f\in\mathcal{F}} P(f^2)$. Thus, Inequality ($a$) may be regarded as an extension of (2.9).

Remark 2.2 *(Explanation of (2.4)).* For any bounded random variable $a \leq Y \leq b$, $a, b \in \mathbb{R}$, Hoeffding [8] shows that Y is more concentrate for the convex functions than the two-valued random variable θ taking the values a and b and such that $\mathbb{E}[\theta] = \mathbb{E}[Y]$ (see his inequalities (4.1) and (4.2)). It means that $\mathbb{E}[\varphi(Y)] \leq \mathbb{E}[\varphi(\theta)]$ for all convex functions φ. This result has been extended to unbounded random variables: the reals a, b are replaced by some random variables α, β and $a \leq Y \leq b$ is replaced by $\alpha \preccurlyeq Y \preccurlyeq \beta$ for some stochastic order $\preccurlyeq$ (see Bentkus [3] and Marchina [14]). Here, one has $\eta \preccurlyeq_2 f(X) \leq 1$, where $\preccurlyeq_2$ denotes the usual second-order stochastic dominance. Then, with respect to a class of convex functions, the distribution of $f(X)$ is more concentrate than the distribution $\mu^{[q_0]}$ defined by

$$\mu^{[q_0]}(A) = \mu(A \cap\,]F_\eta^{-1}(1-q_0)\,,+\infty[\,) + q_0\,\delta_1(A), \tag{2.10}$$

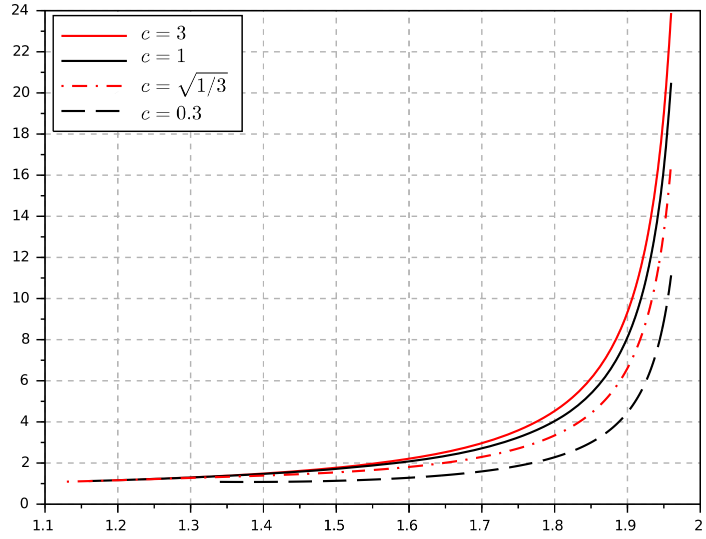

Fig. 1. Impact of the value of c on the function $p \mapsto q_0\Gamma(2-p)$.

where μ denotes the distribution of η, $A \subset \mathbb{R}$ (measurable), δ_1 stands for the Dirac measure centered on 1 and $q_0 \in [0, 1]$ is such that $\int t\mu^{[q_0]}(dt) = \mathbb{E}[f(X)] = 0$. In fact, this last equation is equivalent to (2.4).

By (2.4), q_0 depends on c. The impact of c on the window $[0, q_0\,\Gamma(2-p)]$, in which (a) is valid, is shown in Fig. 1. The function $p \mapsto q_0\,\Gamma(2-p)$ for the values $c=3$, $c=1$, $c=\sqrt{1/3}$ and $c=0.3$ is represented.

3. Proof

We start in the same way as in the proof of the main results in [13], that is, by a martingale decomposition of $Z_n - \mathbb{E}[Z_n]$. Let us recall the main points. The reader is referred to [13] for more details. First, by virtue of the monotone convergence theorem, we can suppose that $\mathcal{F}$ is a finite class of functions. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \ldots, n$, $\mathcal{F}_k := \sigma(X_1, \ldots, X_k)$, and $\mathcal{F}_n^k := \sigma(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also $Z_n^{(k)} := \sup_{f\in\mathcal{F}} \sum_{j\neq k} f(X_j)$ and $Z_k := \mathbb{E}_k[Z_n]$. Let us number the functions of the class $\mathcal{F}$ and consider the random indices

$$\tau := \inf\left\{ i > 0 : \sum_{k=1}^{n} f_i(X_k) = Z_n \right\}$$

$$\tau_k := \inf\left\{ i > 0 : \sum_{j=1}^{n} f_i(X_j) - f_i(X_k) = Z_n^{(k)} \right\}.$$

Define $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and $r_k := (Z_k - \mathbb{E}_k[Z_n^{(k)}]) - \xi_k \geq 0$. Note that we have

$$\xi_k \leq \xi_k + r_k \leq \mathbb{E}_k[f_\tau(X_k)]. \tag{3.1}$$

By the centering assumption on the elements of $\mathcal{F}$, $\mathbb{E}_{k-1}[\xi_k] = 0$, leading to

$$Z_n - \mathbb{E}[Z_n] = \sum_{k=1}^{n} \Delta_k \quad \text{where } \Delta_k := \xi_k + r_k - \mathbb{E}_{k-1}[r_k]. \tag{3.2}$$

The proof is made in three steps:

1. Using the results of Section 3 in Marchina [14], we compare generalized (conditional) moments of Δ_k with those of a random variable ζ_{q_0} with distribution $\mu^{[q_0]}$ given in (2.10). In particular, the class of generalized moments on which we obtain a comparison inequality contains increasing exponential functions $x \mapsto e^{tx}$ for every $t \geq 0$.
2. We give an upper bound on the exponential moments of ζ_{q_0}.
3. We conclude the proof by the usual Cramér–Chernoff calculation.

Step 1: Comparison inequality

Let us denote by ζ_q, for any $q \in [0, 1]$, a random variable with distribution function given by

$$F_q(x) := \begin{cases} F_\eta(x) & \text{if } x < a_q, \\ 1-q & \text{if } a_q \le x < 1, \\ 1 & \text{if } x \ge 1, \end{cases} \tag{3.3}$$

where $a_q := F_\eta^{-1}(1-q)$. Notice that

$$\mathbb{E}[\zeta_q] = q - c\,\frac{p}{p-1}\,(1-q)^{1-1/p}, \tag{3.4}$$

which ensures that (2.4) is equivalent to $\mathbb{E}[\zeta_{q_0}] = 0$. Let us also define

$$\mathcal{H}_+^2 := \left\{\varphi \in \mathcal{C}^1(\mathbb{R}) : \varphi' \text{ is convex, and } \lim_{x\to-\infty} \varphi(x) = \lim_{x\to-\infty} \varphi'(x) = 0\right\}$$

This part is devoted to prove the following lemma.

Lemma 3.1. For any $\varphi \in \mathcal{H}_+^2$ and any $k = 1, \ldots, n$,

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \le \mathbb{E}[\varphi(\zeta_{q_0})]. \tag{a}$$

Consequently, for any $t \ge 0$,

$$\log \mathbb{E}[\exp(t(Z_n - \mathbb{E}[Z_n]))] \le n \log \mathbb{E}[\exp(t\,\zeta_{q_0})]. \tag{b}$$

Remark 3.2. Comparison inequalities with respect to the class of functions $\mathcal{H}_+^2$ (or more generally $\mathcal{H}_+^\alpha$, $\alpha > 0$) have been widely studied by Pinelis (these include, among others, [16–19], and we refer the reader to these papers for more details). We only recall that we have the following equivalence:

(i) $\mathbb{E}[\varphi(X)] \le \mathbb{E}[\varphi(Y)]$ for any $\varphi \in \mathcal{H}_+^2$
(ii) $\mathbb{E}[(X-t)_+^2] \le \mathbb{E}[(Y-t)_+^2]$ for any $t \in \mathbb{R}$.

The proof of Lemma 3.1 lies on the following results, which were established in Marchina [14].

Lemma 3.3 *(Lemmas 4.3 and 4.6 (i) in [14]).* Let η be defined by (1.3).

(i) Let X be an integrable random variable such that $X \le 1$ and for any real t, $\mathbb{E}[(t-X)_+] \le \mathbb{E}[(t-\eta)_+]$. Then for any convex function φ,

$$\mathbb{E}[\varphi(X - \mathbb{E}[X])] \le \mathbb{E}[\varphi(\zeta_q - \mathbb{E}[\zeta_q])],$$

where $q \in [0, 1]$ is such that $\mathbb{E}[\zeta_q] = \mathbb{E}[X]$.

(ii) Let $\tilde{q} := \inf\{q \ge 1/2 : 1 + F_\eta^{-1}(1-q) \le 2\,\mathbb{E}[\zeta_q]\}$. For all $t \in \mathbb{R}$, the function

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

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

Proof of Lemma 3.1. Since $\mathcal{H}_+^2$ contains all increasing exponential functions, taking $\varphi(x) = \mathrm{e}^{tx}$ with $t \ge 0$ in (a) leads to (b) by an induction on n. Moreover, in view of Remark 3.2, we only have to prove (a) for the functions $\varphi(x) = (x-t)_+^2$, with $t \in \mathbb{R}$. Let $t > 0$. Since $r_k \ge 0$, Jensen's inequality implies that

$$\begin{aligned} \mathbb{E}_{k-1}[(-t - (\xi_k + r_k))_+] &\le \mathbb{E}_{k-1}[(-t - \xi_k)_+] \\ &\le \mathbb{E}_{k-1}[(-t - f_{\tau_k}(X_k))_+] \\ &\le \mathbb{E}[(-t-\eta)_+], \end{aligned} \tag{3.5}$$

where the last inequality follows from Assumption 2.3. Furthermore, we can directly verify that (3.5) holds for $t \le 0$. Since $f \le 1$ for any $f \in \mathcal{F}$, (3.1) implies that $\xi_k + r_k \le 1$. Hence, applying Lemma 3.3 (i) conditionally to $\mathcal{F}_{k-1}$, with $X = \xi_k + r_k$, yields that

$$\mathbb{E}_{k-1}[\varphi(\Delta_k)] \le \mathbb{E}_{k-1}[\varphi(\zeta_{\hat{q}} - \mathbb{E}[\zeta_{\hat{q}}])], \tag{3.6}$$

for any convex function φ, where $\hat{q} \in [0,1]$ is such that $\mathbb{E}_{k-1}[\xi_k + r_k] = \mathbb{E}_{k-1}[r_k] = \mathbb{E}_{k-1}[\zeta_{\hat{q}}]$.

Now, we show that an upper bound of the right-hand side of (3.6) can be obtained by the use of Lemma 3.3 (ii). Let us recall the notation $\tilde{q} := \inf\{q \geq 1/2 : 1 + F_\eta^{-1}(1-q) \leq 2\,\mathbb{E}[\zeta_q]\}$. We shall prove the following lemma.

Lemma 3.4. We have

(i) $1 + F_\eta^{-1}(1-q_0) \leq 0 = 2\,\mathbb{E}[\zeta_{q_0}]$. Consequently, $\tilde{q} \leq q_0$.
(ii) $q_0 \leq \hat{q}$.

Remark 3.5. Note that $q \mapsto 1 + F_\eta^{-1}(1-q)$ is nonincreasing and tends to $-\infty$ as q tends to 1 and $q \mapsto \mathbb{E}[\zeta_q]$ is nondecreasing and tends to 1 as q tends to 1. Moreover, a calculation shows that $1 + F_\eta^{-1}(1/2) > 2\,\mathbb{E}[\zeta_{1/2}]$. Thus $\tilde{q}$ exists and if q is such that $1 + F_\eta^{-1}(1-q) \leq 2\,\mathbb{E}[\zeta_q]$, then $\tilde{q} \leq q$.

Proof of Lemma 3.4. Let us first prove (i). Define $q_1 := 1 - \min(c^p, 1)$. Starting from (3.4), one can verify that condition (2.2) on c implies $\mathbb{E}[\zeta_{q_1}] \leq 0$. Then, since $\mathbb{E}[\zeta_{q_0}] = 0$, one has $q_1 \leq q_0$ by the monotonicity of $q \mapsto \mathbb{E}[\zeta_q]$. Next, since $F_\eta(-1) = 1 - q_1$, it implies $F_\eta^{-1}(1-q_1) \leq -1$. Therefrom, since F_η^{-1} is nondecreasing,

$$1 + F_\eta^{-1}(1-q_0) \leq 1 + F_\eta^{-1}(1-q_1) \leq 0.$$

Thus, since $2\,\mathbb{E}[\zeta_{q_0}] = \mathbb{E}[\zeta_{q_0}] = 0$, one has $\tilde{q} \leq q_0$, which ends the proof of (i). The point (ii) follows directly from the monotonicity of $q \mapsto \mathbb{E}[\zeta_q]$ since

$$\mathbb{E}_{k-1}[\zeta_{\hat{q}}] = \mathbb{E}_{k-1}[r_k] \geq 0 = \mathbb{E}[\zeta_{q_0}].$$

This concludes the proof of Lemma 3.4. $\square$

Now, Lemma 3.3 (ii), associated with Lemmas 3.4 and (3.6), applied to $\varphi(x) = (x-t)_+^2$, yields that, for any $t \in \mathbb{R}$,

$$\mathbb{E}_{k-1}[(\Delta_k - t)_+^2] \leq \mathbb{E}_{k-1}[(\zeta_{\hat{q}} - \mathbb{E}[\zeta_{\hat{q}}] - t)_+^2] \leq \mathbb{E}[(\zeta_{q_0} - \mathbb{E}[\zeta_{q_0}] - t)_+^2],$$

which then implies Inequality (a) of Lemma 3.1 and finishes the proof. $\square$

Step 2: Upper bound of exponential moments of ζ_{q_0}

In this part, we give an upper bound of $\mathbb{E}[\exp(t\zeta_{q_0})]$ for any $t \geq 0$. First, recall the notation $a_q = F_\eta^{-1}(1-q) \leq 0$. The aim of this part is to prove the following lemma:

Lemma 3.6. Let $t \geq 0$ such that $-ta_{q_0} \leq 1$. Then

$$\log \mathbb{E}[\mathrm{e}^{t\zeta_{q_0}}] \leq q_0(\mathrm{e}^t - t - 1) + \alpha_p t^p.$$

Proof of Lemma 3.6. Let $t > 0$. Starting from the definition of the random variable ζ_q, one has

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{t\zeta_{q_0}}] &= q_0\,\mathrm{e}^t + p(tc)^p \int_{-ta_{q_0}}^{\infty} \mathrm{e}^{-u} u^{-(p+1)}\,\mathrm{d}u \\ &= q_0\,\mathrm{e}^t + p(tc)^p \Big(\int_0^{\infty} u^{-(p+1)}(\mathrm{e}^{-u} - 1 + u)\,\mathrm{d}u \\ &\quad + \int_{-ta_{q_0}}^{\infty} u^{-(p+1)}(1-u)\,\mathrm{d}u - \int_0^{-ta_{q_0}} u^{-(p+1)}(\mathrm{e}^{-u} - 1 + u)\,\mathrm{d}u \Big).\end{aligned} \tag{3.7}$$

Now, for any $1 < p < 2$,

$$\int_0^{\infty} u^{-(p+1)}(\mathrm{e}^{-u} - 1 + u)\,\mathrm{d}u = \Gamma(-p), \tag{3.8}$$

where $\Gamma(-p) = \frac{1}{p(p-1)}\Gamma(2-p)$. Moreover, the expansion $\mathrm{e}^{-u} = \sum_{k=0}^{\infty}(-u)^k/k!$ yields that

$$\int_{-ta_{q_0}}^{\infty} u^{-(p+1)}(1-u)\,\mathrm{d}u - \int_{0}^{-ta_{q_0}} u^{-(p+1)}(\mathrm{e}^{-u}-1+u)\,\mathrm{d}u = -\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\frac{(-ta_{q_0})^{k-p}}{k-p}. \tag{3.9}$$

Thus, since $q_0 = 1-(c/-a_{q_0})^p$, (3.7) becomes

$$\mathbb{E}[\mathrm{e}^{t\zeta_{q_0}}] = q_0\,\mathrm{e}^t + p(tc)^p\Gamma(-p) + (1-q_0)\Big(1 - ta_{q_0}\frac{p}{1-p} + \sum_{k=2}^{\infty}(-1)^{k+1}\frac{(-ta_{q_0})^k}{k!}\frac{p}{k-p}\Big). \tag{3.10}$$

Next, we observe that under the assumption $-ta_{q_0} \le 1$, the sum in (3.10) is an alternating series whose absolute value of the general term decreases to 0. Thus, the sum is of the sign of the term corresponding to $k=2$, which is negative. Hence,

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{t\zeta_{q_0}}] &\le q_0\,\mathrm{e}^t + p(tc)^p\Gamma(-p) + (1-q_0)\Big(1 - ta_{q_0}\frac{p}{1-p}\Big)\\ &\le 1 + q_0(\mathrm{e}^t - t - 1) + p(tc)^p\Gamma(-p) + t\Big(q_0 - (1-q_0)a_{q_0}\frac{p}{1-p}\Big).\end{aligned} \tag{3.11}$$

Observe now that the last term in the right-hand side is equal to zero. Indeed, since $a_{q_0} = -c\,(1-q_0)^{-1/p}$,

$$q_0 - (1-q_0)a_{q_0}\frac{p}{1-p} = q_0 - c\frac{p}{p-1}(1-q_0)^{1-1/p} = \mathbb{E}[\zeta_{q_0}] = 0.$$

Hence, taking the logarithm and using the inequality $\log(1+x) \le x$ for any $x > 0$, conclude the proof of Lemma 3.6. □

Step 3: Conclusion by the Cramér–Chernoff calculation

We now complete the proof of Theorem 2.1. From Lemma 3.1 (b) and Lemma 3.6, by the usual Cramér–Chernoff calculation, we get

$$\mathbb{P}(Z_n - \mathbb{E}[Z_n] \ge nx) \le \exp\left(-n\,\phi^*_{\zeta_{q_0}}(x)\right), \tag{3.12}$$

where

$$\phi^*_{\zeta_{q_0}}(x) = \sup_{t\in]0,-1/a_{q_0}]}\left\{tx - \alpha_p t^p - q_0(\mathrm{e}^t - t - 1)\right\}. \tag{3.13}$$

In order to prove (a), we give a lower bound of $\phi^*_{\zeta_{q_0}}(x)$ by taking the real $t_x \in]0, -1/a_{q_0}]$ that maximizes $t \mapsto tx - \alpha_p t^p$. A straightforward calculation yields

$$t_x = x^{1/(p-1)}(p\alpha_p)^{-1/(p-1)}, \tag{3.14}$$

and $-t_x a_{q_0} \le 1$ is equivalent to $x \le q_0\,\Gamma(2-p)$. We then have

$$t_x x - \alpha_p t_x^p = \gamma_p\, x^{p/(p-1)} \text{ and } q_0(\mathrm{e}^{t_x} - t_x - 1) = \gamma_p\,\varepsilon_p(x) x^{p/(p-1)}, \tag{3.15}$$

which concludes the proof of (a). Inequality (b) follows directly by putting $t = -a_{q_0}^{-1}$ in the right-hand side of (3.13). This concludes the proof of Theorem 2.1.

Acknowledgements

This work has been done while I was preparing my Ph.D. thesis, and I would like to thank my adviser Emmanuel Rio for many fruitful discussions. I am grateful to the reviewer for his/her valuable comments, which helped to improve the presentation and the clarity of the paper.

References

[1] R. Adamczak, A tail inequality for suprema of unbounded empirical processes with applications to Markov chains, Electron. J. Probab. 13 (34) (2008) 1000–1034.
[2] S. Arlot, M. Lerasle, Choice of V for V-fold cross-validation in least-squares density estimation, J. Mach. Learn. Res. 17 (208) (2016) 1–50.
[3] V. Bentkus, Bounds for the stop loss premium for unbounded risks under the variance constraints, Preprint on https://www.math.uni-bielefeld.de/sfb701/preprints/view/423, 2010.
[4] S. Boucheron, O. Bousquet, G. Lugosi, P. Massart, Moment inequalities for functions of independent random variables, Ann. Probab. 33 (2) (2005) 514–560, p. 03.
[5] S. Boucheron, G. Lugosi, P. Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence, Oxford University Press, Oxford, UK, 2013.
[6] S. Foss, D. Korshunov, S. Zachary, An Introduction to Heavy-Tailed and Subexponential Distributions, second edition, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2013.

[7] A. Goldenshluger, O. Lepski, Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality, Ann. Stat. 39 (3) (2011) 1608–1632.
[8] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Stat. Assoc. 58 (1963) 13–30.
[9] V. Koltchinskii, Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems, École d'été de probabilités de Saint-Flour XXXVIII-2008, vol. 2033, Springer Science & Business Media, 2011.
[10] C. Lacour, P. Massart, V. Rivoirard, Estimator selection: a new method with applications to kernel density estimation, Sankhya, Ser. A 79 (2) (2017) 298–335.
[11] J. Lederer, S. van de Geer, New concentration inequalities for suprema of empirical processes, Bernoulli 20 (4) (2014) 2020–2038.
[12] M. Ledoux, On Talagrand's deviation inequalities for product measures, ESAIM Probab. Stat. 1 (1995–1997) 63–87.
[13] A. Marchina, Concentration inequalities for suprema of unbounded empirical processes, Preprint on https://hal.archives-ouvertes.fr/hal-01545101, 2017.
[14] A. Marchina, Concentration inequalities for separately convex functions, Bernoulli 24 (2018) 2906–2933.
[15] P. Massart, Concentration inequalities and model selection, in: Lectures from the 33rd Summer School on Probability Theory Held in Saint-Flour, France, 6–23 July 2007, Lecture Notes in Mathematics, vol. 1896, Springer, Berlin, 2003, with a foreword by Jean Picard.
[16] I. Pinelis, Optimal tail comparison based on comparison of moments, in: High Dimensional Probability, Springer, 1998, pp. 297–314.
[17] I. Pinelis, Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities, Adv. Stoch. Inequal. 234 (1999) 149–168.
[18] I. Pinelis, On normal domination of (super)martingales, Electron. J. Probab. 11 (39) (2006) 1049–1070.
[19] I. Pinelis, On the Bennett-Hoeffding inequality, Ann. Inst. Henri Poincaré Probab. Stat. 50 (1) (2014) 15–27.
[20] E. Rio, Inégalités exponentielles pour les processus empiriques, C. R. Acad. Sci. Paris, Ser. I 330 (2000) 597–600.
[21] E. Rio, Inégalités de concentration pour les processus empiriques de classes de parties, Probab. Theory Relat. Fields 119 (2) (2001) 163–175.
[22] E. Rio, Une inégalité de Bennett pour les maxima de processus empiriques. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov, Ann. Inst. Henri Poincaré Probab. Stat. 38 (6) (2002) 1053–1057.
[23] M. Talagrand, New concentration inequalities in product spaces, Invent. Math. 126 (3) (1996) 505–563.
[24] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes: With Applications to Statistics, Springer Series in Statistics, Springer, New York, 1996.