About the rate function in concentration inequalities for suprema of bounded empirical processes

Antoine Marchina*

November 18, 2018

Abstract

We provide new deviation inequalities in the large deviations bandwidth for suprema of empirical processes indexed by classes of uniformly bounded functions associated with independent and identically distributed random variables. The improvements we get concern the rate function which is, as expected, the Legendre transform of the suprema of the log-Laplace transform of the pushforward measure by the functions of the considered class (up to an additional corrective term). Our approach is based on a decomposition in martingale together with some comparison inequalities.

Keywords: concentration inequality, empirical process, large deviation bandwidth, comparison inequality, martingale method.

1 Introduction

Let $X_1, \ldots, X_n$ 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 P_n denote the empirical probability measure $P_n := n^{-1}(\delta_{X_1} + \ldots + \delta_{X_n})$. Let $\mathscr{F}$ be a countable class of measurable functions $f : \mathcal{X} \to \mathbb{R}$ such that $P(f) = 0$ and $|f(x)| \leq 1$ for all $x \in \mathcal{X}$ and all $f \in \mathscr{F}$. We are concerned with exponential deviation inequalities with precise rate functions in the large deviations bandwidth for the random variable

$$Z := \sup\{nP_n(f) : f \in \mathscr{F}\}, \tag{1.1}$$

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

around its mean. First, let us briefly recall known results on concentration of Z around its mean for uniform bounded classes $\mathscr{F}$. Talagrand [23] obtains a Bennett-type inequality by means of isoperimetric inequalities for product measures. Ledoux [13] introduces a new method based on entropic inequalities to recover more directly Talagrand's inequalities. This method, which allows to bound above the Laplace transform of Z, is the starting point of a series of papers, mainly to reach optimal constants in Talagrand's inequalities. Let us cite, among others, Massart [15], Rio [18, 19, 20], Bousquet [7], Klein [11], Klein and Rio [12]. In the large deviations bandwidth, as rate function, we expect the Legendre transform of $t \mapsto \sup_{f\in\mathscr{F}} \ell_f(t)$, denoted by $\ell^*_{\mathscr{F}}$, where ℓ_f is the log-Laplace transform $\ell_f(t) := \log P(e^{tf})$ for all $t \geq 0$ and all $f \in \mathscr{F}$. Indeed, one has

$$\frac{1}{n}\log \mathbb{E}[e^{tZ}] \geq \sup_{f\in\mathscr{F}} \ell_f(t) =: \ell_{\mathscr{F}}(t), \tag{1.2}$$

which implies

$$\frac{1}{n}\log \mathbb{E}[e^{t(Z-\mathbb{E}[Z])}] \geq \ell_{\mathscr{F}}(t) - t\,\frac{\mathbb{E}[Z]}{n}. \tag{1.3}$$

Now, if $\mathbb{E}[Z]/n$ tends to 0 (for example, this condition is satisfied when $\mathscr{F}$ is a Glivenko-Cantelli class), then

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{E}[e^{t(Z-\mathbb{E}[Z])}] \geq \ell_{\mathscr{F}}(t). \tag{1.4}$$

This elementary lower bound shows that the large deviations rate function $\ell^*_{\mathscr{F}}$ cannot be improved. To the best of our knowledge, the only result in this direction is obtained in Rio [18] and concerns the particular case of set-indexed empirical processes. Rio gets as rate function, for the right-hand side deviations for sets with large measure under P and for the left-hand side deviations, that of a Bernoulli random variable which actually corresponds to $\ell^*_{\mathscr{F}}$. In this paper, we obtain as rate function for the general case, the function $\ell^*_{\mathscr{F}}$ with an additional corrective term which tends to 0 as n tends to infinity as soon as $\mathscr{F}$ is a weak Glivenko-Cantelli class (see Remark 3.3). Our methods are only based on martingale techniques and comparison inequalities.

The paper is organized as follows. First, in Section 2 we recall some definitions and preliminary results on the Conditional Value-At-Risk and some comparison inequalities. In Section 3 we state the main results of this paper. We study the rate function $\ell^*_{\mathscr{F}}$ in Section 4. Finally, we provide detailed proofs in Section 5.

2 Notation and preliminary results

In this section, we give notation and definitions which we will use all along the paper. Let us start by the definition of the Conditional Value-at-Risk (CVaR for short).

Definition 2.1. *Let X be a real-valued integrable random variable. Let the function Q_X be the càdlàg inverse of $x \mapsto \mathbb{P}(X > x)$. The Conditional Value-at-Risk is defined by*

$$\tilde{Q}_X(u) := u^{-1} \int_0^u Q_X(s) ds \quad \text{for any } u \in \,]0\,,1]. \tag{2.1}$$

Let us now recall the definition of the Legendre transform of a convex function.

Definition 2.2. *Let $\phi : [0\,,\infty[\; \to [0\,,\infty[$ be a convex, nondecreasing and càdlàg function such that $\phi(0) = 0$. The Legendre transform ϕ^* of the function ϕ is defined by*

$$\phi^*(\lambda) := \sup\{\lambda t - \phi(t) : t > 0\} \quad \text{for any } \lambda \geq 0. \tag{2.2}$$

The inverse function of ϕ^* admits the following variational expression (see, for instance, Rio [21, Lemma A.2]).

$$\phi^{*-1}(x) = \inf\{t^{-1}(\phi(t) + x) : t > 0\} \quad \text{for any } x \geq 0. \tag{2.3}$$

A particular function ϕ satisfying conditions in Definition 2.2 is the log-Laplace transform of a random variable:

Notation 2.3. *Let X be a real-valued integrable random variable with a finite Laplace transform on a right neighborhood of 0. The log-Laplace transform of X, denoted by ℓ_X, is defined by*

$$\ell_X(t) := \log \mathbb{E}[\exp(tX)] \quad \text{for any } t \geq 0. \tag{2.4}$$

The function Q_X and the CVaR satisfy the following elementary properties, which are given and proved in Pinelis [16, Theorem 3.4].

Proposition 2.4. *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) $\tilde{Q}_{X+Y}(u) \leq \tilde{Q}_X(u) + \tilde{Q}_Y(u)$.

(iv) Assume that X has a finite Laplace transform on a right neighborhood of 0. Then $\tilde{Q}_X(u) \leq \ell_X^{*-1}(\log(1/u))$.

Remark 2.5. *Since we use different notation from that in Pinelis [16], let us mention that his notation $Q_0(X;u)$, $Q_1(X;u)$ and $Q_\infty(X;u)$ correspond respectively to $Q_X(u)$, $\tilde{Q}_X(u)$ and $\ell_X^{*-1}(\log(1/u))$.*

We now recall comparison inequalities which will be used in the proof of the main result. Let us first give a notation for a family of distribution probability.

Notation 2.6. *Let α, β be two reals such that $\alpha < \beta$. We say that a random variable θ follows a Bernoulli distribution if it assumes exactly two values and we write $\theta \sim \mathcal{B}_m(\alpha, \beta)$ if*

$$\mathbb{P}(\theta = \beta) = 1 - \mathbb{P}(\theta = \alpha) \in \,]0\,, 1[, \ \text{ and } \mathbb{E}[\theta] = m. \tag{2.5}$$

Notice that

$$\mathrm{Var}(\theta) = (m - \alpha)(\beta - m). \tag{2.6}$$

The following classical convex comparison inequality between a bounded random variable X and a Bernoulli random variable with values of the bounds of X was first proved by Hoeffding (see Inequalities (4.1) and (4.2) in [10]); it straight follows by the property of convexity.

Proposition 2.7. *Let a, b be two positive reals and let X be a bounded random variable such that $-a \leq X \leq b$ and $\mathbb{E}[X] = m$. Let $\theta \sim \mathcal{B}_m(-a, b)$. Then, for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\theta)].$$

In particular, since $\mathbb{E}[\theta] = m$, $\mathrm{Var}(X) \leq \mathrm{Var}(\theta) = (a + m)(b - m)$.

Next, Bentkus (see Lemmas 4.4 and 4.5 in [3]) proved that a martingale with bounded from above increments is more concentrate with respect to a certain class of convex functions than a sum of independent and identically distributed Bernoulli random variables.

Proposition 2.8. *Let $b, s_1^2, \ldots, s_n^2$ be positive reals. Let $M_n := \sum_{k=1}^n X_k$ be a martingale with respect to a nondecreasing filtration $(\mathcal{F}_k)$ such that $M_0 = 0$,*

$$X_k \leq b, \ \text{ and } \mathbb{E}[X_k^2 \mid \mathcal{F}_{k-1}] \leq s_k^2 \quad a.s. \tag{2.7}$$

Let $s^2 := n^{-1}(s_1^2 + \ldots + s_n^2)$ and $S_n := \vartheta_1 + \ldots + \vartheta_n$ be a sum of n independent copies of a random variable ϑ with distribution $\mathcal{B}_0(-s^2/b, b)$. Then, for any convex nondecreasing function $\varphi : \mathbb{R} \to \mathbb{R}$, differentiable and with convex derivative,

$$\mathbb{E}[\varphi(M_n)] \leq \mathbb{E}[\varphi(S_n)].$$

Remark 2.9. *Actually Bentkus obtains the above inequality in a smaller class of functions. This generalization is due to Pinelis (see Corollary 5.8 in [17]).*

3 Main results

Let us first introduce one more notation. We denote for any $k = 1, \ldots, n$ the expectations

$$E_k := \mathbb{E} \sup_{f \in \mathscr{F}} P_k(f). \tag{3.1}$$

The main result of the paper is the following theorem:

Theorem 3.1. *Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1, 1]$ such that $P(f) = 0$ for all $f \in \mathscr{F}$. Let Z be defined by (1.1). For any $f \in \mathscr{F}$, let ℓ_f and $\ell_{\mathscr{F}}$ be the functions defined by*

$$\ell_f(t) := \log P(e^{tf}) \text{ and } \ell_{\mathscr{F}}(t) := \sup_{f \in \mathscr{F}} \ell_f(t) \quad \text{for any } t \geq 0. \tag{3.2}$$

Denote $\bar{E}_n := n^{-1}(E_1 + \ldots + E_n)$ and define

$$v_n := \frac{\bar{E}_n}{2}\Big(1 - \frac{\bar{E}_n}{2}\Big). \tag{3.3}$$

Let $\theta^{(n)}$ be a Bernoulli random variable with distribution $\mathcal{B}_0(-v_n, 1)$. We denote by ℓ_{v_n} the log-Laplace transform of $\theta^{(n)}$. Then, for any $x \geq 0$,

$$n^{-1}\tilde{Q}_{Z-\mathbb{E}[Z]}(e^{-nx}) \leq \ell_{\mathscr{F}}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x). \tag{a}$$

Consequently, for any $x \geq 0$,

$$\mathbb{P}(Z - \mathbb{E}[Z] > n(\ell_{\mathscr{F}}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x))) \leq e^{-nx}. \tag{b}$$

The inverse function of $\ell^*_{v_n}$ cannot be explicitly computed. For this reason we provide below a tractable bound.

Corollary 3.2. *Let ψ be the function defined by $\psi(0) = 0$ and for any positive x by*

$$\psi(x) := \frac{\sqrt{2x} + 4x/3}{\log(1 + x/3 + \sqrt{2x})} - 1. \tag{3.4}$$

Then

$$\ell_{v_n}^{*-1}(x) \leq v_n \psi\Big(\frac{x}{v_n}\Big) \quad \text{for any } x \geq 0. \tag{a}$$

Consequently, for any $x \geq 0$,

$$\mathbb{P}\Big(Z - \mathbb{E}[Z] > n\Big(\ell_{\mathscr{F}}^{*-1}(x) + 2\,v_n\psi\Big(\frac{x}{v_n}\Big)\Big)\Big) \leq e^{-nx}. \tag{b}$$

Remark 3.3. *If the class $\mathscr{F}$ is a weak Glivenko-Cantelli class, that is $\sup_{f\in\mathscr{F}}|P_n(f)|$ converges to 0 in probability, then E_n decreases to 0 (see, for instance, Section 2.4 of van der Vaart and Wellner [24]) and so v_n also decreases to* 0 *(we recall that $\bar{E}_n \leq 1$). Consequently, recalling that $\ell_{v_n}(t) = \log(v_n e^t + e^{-tv_n}) - \log(1+v_n)$, we assert by the variational formula* (2.3) *that*

$$\lim_{n\to\infty} \ell_{v_n}^{*-1}(x) = 0 \quad \textit{for all } x \geq 0. \tag{3.5}$$

Therefrom $2\,\ell_{v_n}^{*-1}(x)$ is just a correctional term. Moreover, note that $\psi(x)/x$ tends to 0 *as x tends to infinity and thus,*

$$\lim_{n\to\infty} v_n \psi\Big(\frac{x}{v_n}\Big) = 0 \quad \textit{for all } x \geq 0. \tag{3.6}$$

Hence, the bound given in Corollary 3.2 is still relevant in the large deviations bandwidth.
Consider now the classical bound $\ell_{v_n}(t) \leq v_n\, t^2/(2-2t/3)$ for any $t \in [0\,,3[$, which follows from the domination by a centered Poisson distribution. This leads to

$$\ell_{v_n}^{*-1}(x) \leq \sqrt{2v_n x} + \frac{x}{3}. \tag{3.7}$$

Note that the right-hand side does not tend to 0 contrary to the other bounds which makes (3.7) *non efficient in the large deviations bandwidth.*

Remark 3.4. *Let us mention another possibly upper bound for $\ell_{v_n}^{*-1}$ which is more reader friendly than Inequality (a) of Corollary 3.2 and still relevant in the large deviations bandwidth. Lemma 2.26 of Bercu, Delyon and Rio [4] gives that $\ell_{v_n}^{*}(x) \geq x^2/\varphi(v_n)$ where $\varphi(v) = (1-v^2)/|\log(v)|$ for any $v \in\,]0\,,1[$ (recall that $v_n \leq 1/4$). Hence, this yields the upper bound*

$$\ell_{v_n}^{*-1}(x) \leq \sqrt{x\varphi(v_n)} \textit{ for any } x \geq 0. \tag{3.8}$$

Notice that $\sqrt{x\varphi(v_n)} \leq v_n\psi(x/v_n)$ only for large values of x.

Remark 3.5 (On the large deviations on the left)**.** *Assume that $\mathscr{F}$ is a Glivenko-Cantelli class and that the identically zero function belongs to $\mathscr{F}$. Then $\mathbb{E}[Z]$ is small with respect to n and $Z \geq 0$. Thus for any $x > 0$, $\mathbb{P}(Z - \mathbb{E}[Z] \leq -nx) = 0$ for n large enough.*

Remark 3.6 (Explicit bound for v_n)**.** *In view of* (3.3)*, since the function $x \mapsto x(1-x)$ is increasing between 0 and* $1/2$*, in order to provide a more explicit bound for v_n, we only have to provide a bound for $\bar{E}_n$ (which is lower than* 1 *and tends to 0 as n tends to infinity). To this end, we shall use the* *recent results of Baraud [1] who provides (see his Theorems 2.1 and 2.2) upper bounds with explicit constants for the expectations of suprema of empirical processes, under the hypothesis that $\mathscr{F}$ is a weak VC-major class.*

Assume then that $\mathscr{F}$ is a weak VC-major class with dimension d. Let $\sigma^2 := \sup_{f\in\mathscr{F}} P(f^2)$ denote the wimpy variance. Then Inequality (2.8) in [1] implies the following proposition (the proof is postponed to Section 5).

Proposition 3.7. *Assume that $n \geq d$. Then*

$$\bar{E}_n \leq 2\sqrt{2}\,\sigma\log(e/\sigma)n^{-1/2}\Big(\sqrt{C_1(d)}+\sqrt{C_2(n,d)}\,\Big)+8\,n^{-1}\Big(C_1(d)+C_2(n,d)\Big), \tag{a}$$

where

$$C_1(d) := \log(2)\sum_{k=1}^{d}(1+1/k) \quad \textit{and}$$
$$C_2(n,d) := (d/2)\log\left(\frac{n+1/2}{d+1/2}\right)\log\left(4\,\frac{e^2}{d^2}\,(n+1/2)(d+1/2)\right).$$

As n tends to infinity, the right-hand side of (a) admits the following behavior

$$2\sqrt{2}\,\sigma\log(e/\sigma)n^{-1/2}\log(n)+4\,d\,n^{-1}\log^2(n). \tag{b}$$

We end this section by giving a simple example where the function $\ell_{\mathscr{F}}$ is explicit.

Example 3.8. *Let $\mathscr{S}$ be a countable class of sets. Let $\varepsilon_1,\ldots,\varepsilon_n$ be a sequence of independent Rademacher random variables and independent of $X_1,\ldots,X_n$. Define*

$$Z := \sup_{S\in\mathscr{S}}\sum_{k=1}^{n}\varepsilon_k \mathbb{1}_S(X_k). \tag{3.9}$$

For any $S\in\mathscr{S}$ and any $k=1,\ldots,n$, we get by a straightforward calculation

$$\ell_S(t) := \log\mathbb{E}[e^{t\varepsilon_k\mathbb{1}_S(X_k)}] = \log(1+P(S)(\cosh(t)-1)) \quad \textit{for any } t\geq 0. \tag{3.10}$$

Clearly the right-hand side is increasing with respect to $P(S)$. Then

$$\ell_{\mathscr{S}}(t) := \sup_{S\in\mathscr{S}}\ell_S(t) = \log(1+p(\cosh(t)-1)) \quad \textit{for any } t\geq 0, \tag{3.11}$$

where $p := \sup\{P(S) : S\in\mathscr{S}\}$. By (2.3), $\ell_{\mathscr{S}}^{*-1}$ is then given by the variational formula

$$\ell_{\mathscr{S}}^{*-1}(x) = \inf_{t>0}\Big\{t^{-1}\Big(x+\log(1+p(\cosh(t)-1))\Big)\Big\} \quad \textit{for any } x\geq 0. \tag{3.12}$$

We also refer the reader to Bennett [2], p. 532, for an explicit formula of $\ell_{\mathscr{S}}^*$.

4 About the rate function $\ell^*_{\mathscr{F}}$

4.1 Comments on Large Deviation Principle

In this subsection we explain how the rate function $\ell^*_{\mathscr{F}}$ arises in the large deviations theory for suprema of bounded empirical processes.

Throughout this section, we assume that for all $f \in \mathscr{F}$, $0 \leq f \leq 1$. We denote by $l_\infty(\mathscr{F})$ the space of all bounded real functions on $\mathscr{F}$ equipped with the norm $\|F\|_{\mathscr{F}} := \sup_{f\in\mathscr{F}} |F(f)|$, making $(l_\infty(\mathscr{F}), \|.\|_\infty)$ a Banach space. For each finite measure ν on $(\mathcal{X}, \mathcal{F})$ corresponds to an element $\nu^{\mathscr{F}} \in l_\infty(\mathscr{F})$ defined by $\nu^{\mathscr{F}}(f) := \nu(f) = \int f d\nu$ for any $f \in \mathscr{F}$. With a slight abuse of notation, we will keep the notation ν instead of $\nu^{\mathscr{F}}$. Wu [25] gives necessary and sufficient conditions with respect to $\mathscr{F}$ which ensure that P_n satisfies the Large Deviation Principle (LDP for short) in $l_\infty(\mathscr{F})$. We refer the reader to the paper of Wu for these conditions (for example, if $\mathscr{F}$ is a Donsker class then the required conditions are satisfied). The (good) rate function is given by

$$h_{\mathscr{F}}(F) := \inf\{H(\nu \mid P) : \nu \text{ is a probability and } \nu = F \text{ on } \mathscr{F}\}, \tag{4.1}$$

where $H(\nu \mid P)$ is the relative entropy of ν with respect to P given, as soon as ν is absolutely continuous with respect to P, by

$$H(\nu, P) := \int \frac{d\nu}{dP} \log\left(\frac{d\nu}{dP}\right) dP. \tag{4.2}$$

Therefrom, a direct application of the contraction principle (see, for instance, Theorem 4.2.1 in Dembo and Zeitouni [9]) ensures that $\|P_n\|_{\mathscr{F}}$ satisfies the LDP with rate function given by

$$J(y) := \inf\{H(\nu \mid P) : \nu \text{ is a probability and } \|\nu\|_{\mathscr{F}} = y\} \text{ for any } y \in [0\,,1[. \tag{4.3}$$

We prove the following lemma which gives a better understanding of the rate function J:

Lemma 4.1. *$J(y) = \inf_{f\in\mathscr{F}} \ell^*_f(y)$, where $\ell_f(t) := \log P(e^{tf})$ for any $t \geq 0$.*

The important remark is that if we can invert the infimum and the supremum in $\inf_{f\in\mathscr{F}} \sup_{t>0}\{ty - \ell_f(y)\}$, we get that $\inf_{f\in\mathscr{F}} \ell^*_f(y) = \ell^*_{\mathscr{F}}(y)$. It seems not possible to invert the infimum and the supremum in general. However, note that we always have the inequality $\inf_{f\in\mathscr{F}} \ell^*_f(y) \geq \ell^*_{\mathscr{F}}(y)$. In the following proposition, we describe a particular case in which the inversion is valid, which then simplifies the calculation of $\ell^*_{\mathscr{F}}$. Since it directly follows from a minimax theorem (see, for instance, Corollary 3.3 in Sion [22]), we omit the proof.

Proposition 4.2. *Let X be a random variable valued in $(\mathcal{X}, \mathcal{F})$ with distribution P. Let $\mathscr{F}$ be a countable class of measurable functions from $\mathcal{X}$ into $[-1\,,1]$ such that $P(f) = 0$ for all $f \in \mathscr{F}$. Let Θ be a convex compact subset of a vector space. Let $\{\mu_\theta : \theta \in \Theta\}$ be a family of probability distribution on $[-1\,,1]$ such that, for any $t \geq 0$, $\theta \mapsto \ell_{\mu_\theta}(t) := \log \int e^{tz}\mu_\theta(dz)$ is concave and upper semi-continuous. We assume that for all $f \in \mathscr{F}$, there exists $\theta \in \Theta$ such that $f(X)$ has the distribution μ_θ. Then,*

$$\ell^*_{\mathscr{F}}(x) \geq \inf_{\theta\in\Theta} \ell^*_{\mu_\theta}(x) \quad \text{for any } x \geq 0.$$

Example 4.3 (Set-indexed empirical processes)**.** *Let X be a random variable valued in $(\mathcal{X}, \mathcal{F})$ with distribution P and let $\mathscr{S}$ be a countable class of measurable sets of $\mathcal{X}$. We consider the class of functions $\mathscr{F} := \{\mathbb{1}_S - P(S) : S \in \mathscr{S}\}$. Define $p := \sup\{P(S) : S \in \mathscr{S}\}$ and assume that $p < 1/2$. For any $\theta \in [0\,,p]$, let us define the function $\ell_\theta(t) := \log(1 + \theta(e^t - 1)) - \theta t$ for any $t \geq 0$. Then Proposition 4.2 yields*

$$\ell^*_{\mathscr{F}}(x) \geq \inf_{\theta\in[0,p]} \ell^*_\theta(x) \quad \text{for any } x \geq 0. \tag{4.4}$$

The computation of the right-hand side of (4.4) is performed by Rio (see p. 175 in [18]): for any $x \leq 1 - 2p$,

$$\inf_{\theta\in[0,p]} \ell^*_\theta(x) = \ell^*_p(x) = (p+x)\log(1+x/p) + (1-p-x)\log(1-x/(1-p)). \tag{4.5}$$

Furthermore, for any $x \geq 1 - 2p$,

$$\begin{aligned}\inf_{\theta\in[0,p]} \ell^*_\theta(x) &\geq \ell^*_p(1-2p) + \int_{1-2p}^{x} (\ell^*_{(1-y)/2})'(y)dy \\ &= 2(1+x)\log(1+x) + 2(1-x)\log(1-x) \\ &\qquad - (1+2p)\log(2p) - (3-2p)\log(2-2p).\end{aligned} \tag{4.6}$$

Remark 4.4. *Proceeding as in the proof of Theorem 4.2 in Rio [18], one can derive from Theorem 6.3 in Bousquet [7] that, if $p_0 := p + E_n$ satisfies $p_0 < 1/2$, then for any $t > 0$ such that $p_0 < (te^t - e^t + 1)(e^t - 1)^{-2}$,*

$$n^{-1}\log \mathbb{E}[\exp(t(Z - \mathbb{E}[Z]))] \leq \log(1 + p_0(e^t - 1)) - tp_0. \tag{4.7}$$

Now (4.7) and the usual Cramér-Chernoff calculation, imply that $\mathbb{P}(Z - \mathbb{E}[Z] \geq nx) \leq \exp(-n\ell^*_{p_0}(x))$, for any $x > 0$ such that

$$x \leq (x + p_0)(1 - x - p_0)\log\Big(\frac{(1-p_0)}{p_0}\frac{(t+p_0)}{(1-t-p_0)}\Big). \tag{4.8}$$

Bousquet [7] tells without proof that (4.8) *holds for any* $x \le (3/4)(1-2p_0)$. *If* $x = x_0 := 1-2p_0$, (4.8) *is equivalent to*

$$p_0(1-p_0) \ge (1-2p_0)/2\log(1/p_0 - 1),$$

which is wrong (see Hoeffding [10] p. 19). Recall now that Bousquet's results are derived from the entropy method introduced by Ledoux [13] on the context of concentration inequalities. It appears here that this method does not provide the exact rate function for large values of x, *including* $x = 1-2p_0$.

4.2 The case of nondecreasing 1-Lipschitz functions

Here we study the special case of $\mathscr{F}$ included in the set of nondecreasing 1-Lipschitz functions. We can then bound above $\ell_{\mathscr{F}}^{*-1}$ by a more tractable quantity.

Corollary 4.5. *Let* X *be a random variable valued in* $(\mathcal{X}, \mathcal{F})$ *with distribution* P *and* $X_1, \ldots, X_n$ *be* n *independent copies of* X. *Let* $\mathscr{F}$ *be a countable class of measurable functions from* $\mathcal{X}$ *into* $[-1\,,1]$, *nondecreasing, 1-Lipschitz and such that* $P(f) = 0$ *for all* $f \in \mathscr{F}$. *Let* Z *be defined by* (1.1). *Moreover, we assume that the distribution* P *satisfies that for any* $t \in \mathbb{R}$, $\int e^{tx} P(dx) < \infty$. *Then*

$$\ell_{\mathscr{F}}^{*-1}(x) \le \ell_{X-\mathbb{E}[X]}^{*-1}(x) \quad \textit{for any } x \ge 0. \tag{a}$$

Consequently, for any $x \ge 0$,

$$\mathbb{P}\Big(Z - \mathbb{E}[Z] > n\big(\ell_{X-\mathbb{E}[X]}^{*-1}(x) + 2\,\ell_{v_n}^{*-1}(x)\big)\Big) \le e^{-nx}. \tag{b}$$

Example 4.6. *Let* P *be the uniform distribution on* $[-1\,,1]$. *Then, by* (2.3), ℓ_X^{*-1} *is given by the following variational formula whose values in every point is computable:*

$$\ell_X^{*-1}(x) = \inf_{t>0}\left\{\frac{1}{t}\Big(x + \log\Big(\frac{\sinh(t)}{t}\Big)\Big)\right\} \quad \textit{for any } x \ge 0. \tag{4.9}$$

Let us also provide a bound of ℓ_X^{*-1} *which is relevant for large values of* x. *Since* $\sinh(t) \le e^t/2$ *for any* $t > 0$, *one has*

$$\ell_X^{*-1}(x) \le 1 + \inf_{t>0}\left\{\frac{1}{t}(x - \log(2t))\right\} \quad \textit{for any } x \ge 0. \tag{4.10}$$

Then, for each $x \ge 0$, *the infimum in* (4.10) *is reached at* $t_x := e^{x+1}/2$, *which leads to*

$$\ell_X^{*-1}(x) \le 1 - \frac{2}{e}\,e^{-x} \quad \textit{for any } x \ge 0. \tag{4.11}$$

Furthermore, one can prove that $\ell_X^{*-1}(x)$ *is equivalent to* $1 - \frac{2}{e}\,e^{-x}$ *as* x *tends to infinity.*

5 Proofs

5.1 Proofs of Section 3

Proof of Theorem 3.1. First, notice that (b) follows immediately from (a) by Proposition 2.4 (i) and (ii). Let us now prove (a). Our method is based on a martingale decomposition of Z which we now recall. We suppose that $\mathscr{F}$ is a finite class of functions, that is $\mathscr{F} = \{f_i : i \in \{1, \ldots, m\}\}$. The results in the countable case are derived from the finite case using the monotone convergence theorem. Set $\mathcal{F}_0 := \{\varnothing, \Omega\}$ and for all $k = 1, \ldots, n$, $\mathcal{F}_k := \sigma(X_1, \ldots, X_k)$ and $\mathcal{F}_n^k := \sigma(X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n)$. Let $\mathbb{E}_k$ (respectively $\mathbb{E}_n^k$) denote the conditional expectation operator associated with $\mathcal{F}_k$ (resp. $\mathcal{F}_n^k$). Set also

$$Z_k := \mathbb{E}_k[Z], \quad Z^{(k)} := \sup\{nP_n(f) - f(X_k) : f \in \mathscr{F}\}. \tag{5.1}$$

The sequence (Z_k) is an $(\mathcal{F}_k)$-adapted martingale and

$$Z - \mathbb{E}[Z] = \sum_{k=1}^{n} \Delta_k, \quad \text{where } \Delta_k := Z_k - Z_{k-1}. \tag{5.2}$$

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

$$\tau := \inf\{i \in \{1, \ldots, m\} : nP_n(f_i) = Z\}, \tag{5.3}$$

$$\tau_k := \inf\{i \in \{1, \ldots, m\} : nP_n(f_i) - f_i(X_k) = Z^{(k)}\}. \tag{5.4}$$

Notice first that

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

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

$$\mathbb{E}_k[f_{\tau_k}(X_k)] \le Z_k - \mathbb{E}_k[Z^{(k)}] \le \mathbb{E}_k[f_\tau(X_k)]. \tag{5.5}$$

Set now $\xi_k := \mathbb{E}_k[f_{\tau_k}(X_k)]$ and let $\varepsilon_k \ge r_k \ge 0$ 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 (5.5) becomes

$$\xi_k \le \xi_k + r_k \le \xi_k + \varepsilon_k. \tag{5.6}$$

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

$$\mathbb{E}_n^k[f_{\tau_k}(X_k)] = P(f_{\tau_k}) = 0 \quad \text{a.s.,} \tag{5.7}$$

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 (5.2) yield the decomposition of $Z - \mathbb{E}[Z]$ in a sum of two martingales:

$$Z - \mathbb{E}[Z] = \Xi_n + R_n, \tag{5.8}$$

where

$$\Xi_n := \sum_{k=1}^{n} \xi_k \text{ and } R_n := \sum_{k=1}^{n} (r_k - \mathbb{E}_{k-1}[r_k]). \tag{5.9}$$

Now, we bound above separately the log-Laplace transforms of Ξ_n and R_n.

Lemma 5.1. *We have*

$$\log \mathbb{E}\Big[\exp\Big(t\,\Xi_n\Big)\Big] \leq n\,\ell_{\mathscr{F}}(t) \quad \textit{for any } t \geq 0.$$

Proof of Lemma 5.1. The $\mathscr{F}_n^k$-measurability of τ_k gives

$$\mathbb{E}_n^k[\exp(tf_{\tau_k}(X_k))] = P(e^{tf_{\tau_k}}). \tag{5.10}$$

This ensures, with an application of the conditional Jensen inequality, that

$$\mathbb{E}_{k-1}[e^{t\xi_k}] = \mathbb{E}_{k-1}\mathbb{E}_n^k[e^{t\xi_k}] \leq \mathbb{E}_{k-1}[P(e^{tf_{\tau_k}})] \leq \sup_{f\in\mathscr{F}} P(e^{tf}), \tag{5.11}$$

almost surely. Then Lemma 5.1 follows by an immediate induction on n. $\square$

Lemma 5.2. *We have*

$$\log \mathbb{E}\Big[\exp\Big(t\,R_n\Big)\Big] \leq n\,\ell_{v_n}(2\,t) \quad \textit{for any } t \geq 0.$$

Proof of Lemma 5.2. Actually, the inequality follows by taking $\varphi(x) = e^{tx}$ with $t \geq 0$ in the more general comparison inequality below:

Lemma 5.3. *Let $\theta_1^{(n)}, \ldots, \theta_n^{(n)}$ be a sequence of n independent copies of $\theta^{(n)}$ with $\mathcal{B}_0(-v_n, 1)$ distribution. Then, for any convex nondecreasing function φ from $\mathbb{R}$ into $\mathbb{R}$, differentiable and with convex derivative,*

$$\mathbb{E}[\varphi(R_n)] \leq \mathbb{E}\Big[\varphi\Big(2\sum_{k=1}^{n} \theta_k^{(n)}\Big)\Big].$$

Proof of Lemma 5.3. We start the proof by showing that

$$r_k - \mathbb{E}_{k-1}[r_k] \leq 2, \text{ and } \operatorname{Var}(r_k \mid \mathcal{F}_{k-1}) \leq E_{n-k+1}(2 - E_{n-k+1}) \quad \text{a.s.} \tag{5.12}$$

The first inequality above is straightforward by (5.6) and the uniform boundedness condition on $\mathscr{F}$. Let us prove now the second inequality. We start by bounding above $\operatorname{Var}(r_k \mid \mathcal{F}_{k-1})$ in terms of $\mathbb{E}_{k-1}[r_k]$. Since $0 \leq r_k \leq 2$, Proposition 2.7, applied conditionally to $\mathcal{F}_{k-1}$, yields

$$\operatorname{Var}(r_k \mid \mathcal{F}_{k-1}) \leq \mathbb{E}_{k-1}[r_k](2 - \mathbb{E}_{k-1}[r_k]) \quad \text{a.s.} \tag{5.13}$$

Next, we prove that $E_{k-1}[r_k]$ is bounded above by a deterministic constant.

Lemma 5.4. *We have* $0 \leq \mathbb{E}_{k-1}[r_k] \leq E_{n-k+1} \leq 1$ *a.s.*

Proof of Lemma 5.4. The proof is based on the following result on exchangeability of variables, proved in Marchina [14]. Since it is the fundamental tool of the paper, we give again the proof for sake of completeness.

Lemma 5.5. *For any integer* $j \geq k$*,* $\mathbb{E}_{k-1}[f_\tau(X_k)] = \mathbb{E}_{k-1}[f_\tau(X_j)]$ *a.s.*

Proof of Lemma 5.5. By the definition of the random index τ, for every permutation on n elements σ, $\tau(X_1, \ldots, X_n) = \tau \circ \sigma(X_1, \ldots, X_n)$ 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. □

Then,

$$\begin{aligned}
\mathbb{E}_{k-1}[\varepsilon_k] &= \mathbb{E}_{k-1}[f_\tau(X_k)] \\
&= \mathbb{E}_{k-1}[f_\tau(X_k) + \ldots + f_\tau(X_n)]/(n-k+1) \\
&\leq \mathbb{E}_{k-1} \sup_{f \in \mathscr{F}} \{f(X_k) + \ldots + f(X_n)\}/(n-k+1) = E_{n-k+1}.
\end{aligned} \tag{5.14}$$

Recalling that $0 \leq r_k \leq \varepsilon_k$, we get $0 \leq \mathbb{E}_{k-1}[r_k] \leq E_{n-k+1}$. The bound $E_{n-k+1} \leq 1$ is straightforward by the uniform boundedness condition on the elements of $\mathscr{F}$, which ends the proof of Lemma 5.4. □

Finally, (5.13) together with Lemma 5.4 and the fact that $x \mapsto x(2-x)$ is increasing between 0 and 1, imply

$$\operatorname{Var}(r_k \mid \mathcal{F}_{k-1}) \leq E_{n-k+1}(2 - E_{n-k+1}) \quad \text{a.s.,} \tag{5.15}$$

Now, Proposition 2.8 yields that for any convex, nondecreasing function φ differentiable with convex derivative,

$$\mathbb{E}[\varphi(R_n)] \leq \mathbb{E}\Big[\varphi\Big(\sum_{k=1}^{n} \vartheta_k^{(n)}\Big)\Big], \tag{5.16}$$

where $\vartheta_1^{(n)}, \ldots, \vartheta_n^{(n)}$ is a sequence of i.i.d. random variables such that $\vartheta_k^{(n)}$ has the distribution $\mathcal{B}_0(-\tilde{v}_n, 2)$ with $\tilde{v}_n := \sum_{k=1}^n E_k(2-E_k)$. Moreover, since $x \mapsto x(2-x)$ is concave, $\tilde{v}_n \le \bar{E}_n(2-\bar{E}_n)$. Finally Hoeffding's convex comparison inequality (Proposition 2.7) yields that for any convex, nondecreasing function φ differentiable with convex derivative,

$$\mathbb{E}\Big[\varphi\Big(\sum_{k=1}^n \vartheta_k^{(n)}\Big)\Big] \le \mathbb{E}\Big[\varphi\Big(2\sum_{k=1}^n \theta_k^{(n)}\Big)\Big], \tag{5.17}$$

This inequality associated with (5.16) conclude the proof of Lemma 5.3. □

As mentioned at the beginning of the proof, this also concludes the proof of Lemma 5.2 by taking $\varphi(x) = e^{tx}$ with $t \ge 0$. □

Let us now complete the proof of Theorem 3.1. From (2.3) and Lemmas 5.1–5.2 we derive for any $x \ge 0$,

$$\ell_{\Xi_n}^{*-1}(nx) \le n\,\ell_{\mathscr{F}}^{*-1}(x) \text{ and } \ell_{R_n}^{*-1}(nx) \le 2n\,\ell_{v_n}^{*-1}(x). \tag{5.18}$$

Furthermore, from Proposition 2.4 (iii), (iv) and (5.8)

$$\begin{aligned}\tilde{Q}_{Z-\mathbb{E}[Z]}(e^{-nx}) &\le \tilde{Q}_{\Xi_n}(e^{-nx}) + \tilde{Q}_{R_n}(e^{-nx}) \\ &\le \ell_{\Xi_n}^{*-1}(nx) + \ell_{R_n}^{*-1}(nx).\end{aligned} \tag{5.19}$$

Finally, both (5.19) and (5.18) conclude the proof of Theorem 3.1 (a). □

Proof of Corollary 3.2. Let Π_n be a random variable with Poisson distribution with parameter v_n and let $\tilde{\Pi}_n := \Pi_n - v_n$. A classical result gives $\ell_{v_n}(t) \le \ell_{\tilde{\Pi}_n}(t)$ for any $t \ge 0$ (see, for instance, Theorem 2.9 of [6]). Therefore, for any $x \ge 0$,

$$\ell_{v_n}^{*-1}(x) \le \ell_{\tilde{\Pi}_n}^{*-1}(x) = v_n\, h^{-1}\Big(\frac{x}{v_n}\Big), \tag{5.20}$$

where $h(u) := (1+u)\log(1+u) - u$ for any $u \ge 0$. Next, a Newton algorithm performed in Del Moral and Rio (see Appendix A.6 in [8]) allows to derive the bound $h^{-1}(x) \le \psi(x)$, which concludes the proof of Corollary 3.2. □

Proof of Proposition 3.7. Inequality (2.8) in Baraud [1] implies that for any $k = 1, \ldots, n$,

$$E_k \le 2\sqrt{2}\,\sigma \log(e/\sigma)\sqrt{\tilde{\Gamma}_k(d)} + 8\,\tilde{\Gamma}_k(d), \tag{5.21}$$

where

$$\tilde{\Gamma}_k(d) := k^{-1}\log\left(2\sum_{j=0}^{d\wedge k}\binom{k}{j}\right). \tag{5.22}$$

We recall (see p. 1714 in [1]) that, for any $d \geq k$, $\tilde{\Gamma}_k(d) = \log(2)(k+1)/k$ and for any $d < k$,

$$\tilde{\Gamma}_k(d) \leq \frac{d}{k} \log\Big(\frac{2e}{d} k\Big). \tag{5.23}$$

Moreover, observe that by the concavity of $x \mapsto \sqrt{x}$, one has

$$\frac{1}{n} \sum_{k=1}^{n} \sqrt{\tilde{\Gamma}_k(d)} \leq \sqrt{\frac{1}{n} \sum_{k=1}^{n} \tilde{\Gamma}_k(d)}. \tag{5.24}$$

Then, since $\bar{E}_n = n^{-1}(E_1 + \ldots + E_n)$, the previous facts together with the sub-additivity of $x \mapsto \sqrt{x}$ yield

$$\begin{aligned} \bar{E}_n \leq 2\sqrt{2}\,\sigma \log(e/\sigma) n^{-1/2} \Big(\sqrt{C_1(d)} + \Big(\sum_{k=d+1}^{n} \frac{d}{k} \log\Big(\frac{2e}{d} k\Big)\Big)^{1/2} \Big) \\ + 8\, n^{-1} \Big(C_1(d) + \sum_{k=d+1}^{n} \frac{d}{k} \log\Big(\frac{2e}{d} k\Big)\Big). \end{aligned} \tag{5.25}$$

Observe now that the function h defined by $h(x) := x^{-1} \log((2e/d)x)$ is convex (at least) on $[d, +\infty[$. Thus for any integer $k > d$, $h(k) \leq \int_{k-1/2}^{k+1/2} h(x)dx$. Summing then this inequality from $d+1$ to n gives

$$\sum_{k=d+1}^{n} h(k) \leq \frac{1}{2} \log\Big(\frac{n+1/2}{d+1/2}\Big) \log\Big(\frac{4e^2}{d^2}(n+1/2)(d+1/2)\Big). \tag{5.26}$$

Finally injecting (5.26) in (5.25) concludes the proof. □

5.2 Proofs of Section 4

Proof of Lemma 4.1. Throughout the proof, we use the notation $I(y) = \inf_{f \in \mathscr{F}} \ell_f^*(y)$.

(**i**) Proof of $J(y) \leq I(y)$.

Let $y \in [0, 1]$ and let $\varepsilon > 0$. There exists a function $f \in \mathscr{F}$ such that $\ell_f^*(y) \leq I(y) + \varepsilon$. Now, Cramér's Theorem ensures that

$$\lim_{n \to \infty} n^{-1} \log \mathbb{P}(P_n(f) \geq y) = -\ell_f^*(y). \tag{5.27}$$

Since $\|P_n\|_{\mathscr{F}}$ satisfies the LDP with rate function J and since $P_n(f) \leq \|P_n\|_{\mathscr{F}}$ for all $f \in \mathscr{F}$, we get

$$\begin{aligned} -J(y) &\geq \limsup_{n \to \infty} n^{-1} \log \mathbb{P}(\|P_n\|_{\mathscr{F}} \geq y) \\ &\geq \lim_{n \to \infty} n^{-1} \log \mathbb{P}(P_n(f) \geq y) = -\ell_f^*(y). \end{aligned} \tag{5.28}$$

Therefrom $J(y) \leq I(y) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we conclude the proof by letting ε tends to 0.

(**ii**) Proof of $J(y) \geq I(y)$.
Since the infima may be written as the limit of a sequence of infima taken over finite subsets, it is enough to prove the inequality for a finite class of functions $\mathscr{F}$. Let $y \in [0\,,1]$ and $t > 0$. Let ν be a probability measure absolutely continuous with respect to P such that $\|\nu\|_{\mathscr{F}} = y$. Let $d := (d\nu/dP)$ be the Radon-Nikodym derivative of ν with respect to P and set $g_f := tf - \log P(e^{tf})$ for any $f \in \mathscr{F}$. Young's inequality (see, for instance, Equation (A.2) in Rio [21]) implies that

$$t\nu(f) - \log P(e^{tf}) = \int dg_f \, dP \leq \int e^{g_f} dP + \int (d \log d - d)\, dP. \tag{5.29}$$

Since $\int e^{g_f} dP = 1$, (5.29) leads to

$$t\nu(f) - \log P(e^{tf}) \leq H(\nu \mid P). \tag{5.30}$$

In particular (5.30) is valid for the function $\tilde{f} \in \mathscr{F}$ which satisfies $y = \nu(\tilde{f})$ (recall that $\mathscr{F}$ is finite) and for any $t > 0$. Then we have

$$\ell^*_{\tilde{f}}(y) \leq H(\nu \mid P), \tag{5.31}$$

which implies $I(y) \leq J(y)$ and ends the proof. □

Proof of Corollary 4.5. Let X be a random variable with distribution P. Recalling that $P(f) = 0$ for any $f \in \mathscr{F}$, Lemma 2 of Bobkov [5] states that for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and for any $f \in \mathscr{F}$,

$$\mathbb{E}[\varphi(f(X))] \leq \mathbb{E}[\varphi(X - \mathbb{E}[X])].$$

In particular with $\varphi(x) = e^{tx}$, $t \geq 0$,

$$\ell_{\mathscr{F}}(t) = \sup_{f \in \mathscr{F}} \ell_f(t) \leq \log \mathbb{E}[e^{t(X - \mathbb{E}[X])}] = \ell_{X - \mathbb{E}[X]}(t). \tag{5.32}$$

Thus the variational formula (2.3) implies $\ell^{*-1}_{\mathscr{F}}(x) \leq \ell^{*-1}_{X-\mathbb{E}[X]}(x)$ for all $x \geq 0$. An application of Theorem 3.1 completes the proof. □

Acknowledgments. I would like to thank Emmanuel Rio for many insightful discussions and suggestions.

References

[1] Y. Baraud. (2016). Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.* **10**, no. 2, 1709–1728.

[2] G. Bennett. (1963). On the probability of large deviations from the expectation for sums of bounded, independent random variables. *Biometrika* **50**, 528–535.

[3] V. Bentkus. (2004). On Hoeffding's inequalities, *Ann. Probab.* **32**, no. 2, 1650–1673.

[4] B. Bercu, B. Delyon, and E. Rio. (2015). *Concentration inequalities for sums and martingales. SpringerBriefs in Mathematics.* New-York: Springer.

[5] S.G. Bobkov. (2003). Localization proof of the Bakry–Ledoux isoperimetric inequality and some applications. *Theory of Probability & Its Applications*, **47**, no. 2, 308–314.

[6] S. Boucheron, G. Lugosi, and P. Massart. (2013). *Concentration inequalities: A nonasymptotic theory of independence.* Oxford: Oxford Univ. Press.

[7] O. Bousquet. (2003). Concentration inequalities for sub-additive functions using the entropy method. *Stochastic inequalities and applications*, 213–247, Progr. Probab., 56, *Birkhäuser, Basel.*

[8] P. Del Moral and E. Rio. (2011). Concentration inequalities for mean field particle models. *Ann. Appl. Probab.* **21**, no. 3, 1017–1052.

[9] A. Dembo and O. Zeitouni. (2010). Large deviations techniques and applications. Corrected reprint of the second (1998) edition. Stochastic Modelling and Applied Probability, 38. *Springer-Verlag, Berlin.*

[10] W. Hoeffding. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.

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

[12] T. Klein and E. Rio. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, no. 3, 1060–1077.

[13] M. Ledoux. (1995/97). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.* **1**, 63–87.

[14] A. Marchina. (2017). Concentration inequalities for suprema of unbounded empirical processes. Preprint on <hal-01545101>.

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

[16] I. Pinelis. (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**, no. 3, 349–392.

[17] I. Pinelis. (2016). Convex cones of generalized multiply monotone functions and the dual cones. *Banach J. Math. Anal.* **10**, no. 4, 864–897.

[18] E. Rio. (2001). Inégalités de concentration pour les processus empiriques de classes de parties. *Probab. Theory Related Fields* **119**, no. 2; 163–175.

[19] E. Rio. (2002). Une inégalité de Bennett pour les maxima de processus empiriques. *Ann. Inst. H. Poincaré Probab. Statist.* **38**, no. 6, 1053–1057, En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

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

[21] E. Rio. (2017). Asymptotic theory of weakly dependent random processes. Translated from the 2000 French edition [MR2117923]. Probability Theory and Stochastic Modelling, 80. *Springer, Berlin.*

[22] M. Sion. (1958). On general minimax theorems. *Pacific J. Math.* **8**, 171–176.

[23] M. Talagrand. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, no. 3, 505–563.

[24] A.W. van der Vaart and J.A. Wellner. (1996). Weak convergence and empirical processes. With applications to Statistics. Springer Series in Statistics. *Springer-Verlag, New-York.*

[25] L. Wu. (1994). Large deviations, moderate deviations and LIL for empirical processes. *Ann. Probab.* **22**, no. 1, 17–27.