The Annals of Probability
2000, Vol. 28, No. 2, 863–884

ABOUT THE CONSTANTS IN TALAGRAND'S CONCENTRATION INEQUALITIES FOR EMPIRICAL PROCESSES

BY PASCAL MASSART

Université Paris-Sud

We propose some explicit values for the constants involved in the exponential concentration inequalities for empirical processes which are due to Talagrand. It has been shown by Ledoux that deviation inequalities for empirical processes could be obtained by iteration of logarithmic Sobolev type inequalities. Our approach follows closely that of Ledoux. The improvements that we get with respect to Ledoux's work are based on refinements of his entropy inequalities and computations.

1. Introduction. The concentration of measure phenomenon for product measures has been investigated in depth by Talagrand in a most remarkable series of works (see in particular [20] for an overview and [21] for recent advances). One of the first striking results illustrating this phenomenon was obtained in the seventies. It is the concentration of the standard Gaussian measure on $\mathbb{R}^N$. Consider some Lipschitz function ζ on the Euclidean space $\mathbb{R}^N$ with Lipschitz constant L, if P denotes the canonical Gaussian measure on $\mathbb{R}^N$; then, for every $x \geq 0$,

$$P[|\zeta - M| \geq x] \leq 2\exp\left(-\frac{x^2}{2L^2}\right)$$

and

$$P[\zeta - M + x] \leq \exp\left(-\frac{x^2}{2L^2}\right), \tag{1}$$

where M denotes either the mean or the median of ζ with respect of P. These inequalities were independently established in [9] and [6] when M is a median and in [8] when M is the mean (we refer to [15] for various proofs and numerous applications of these statements). Usually the first inequality is called a concentration inequality while the latter is called a deviation inequality. A very interesting feature of these inequalities is that they do not depend on the dimension N which allows using them for controlling suprema of Gaussian processes, for instance (see [15]). Extending such results to more general product measures appears to be a very difficult task. Talagrand's methods rely on isoperimetric ideas in the sense that concentration inequalities for functionals around their median are derived from probability inequalities for enlargements of sets with respect to various distances (or more general measures

Received January 1998; revised October 1999.
AMS 1991 *subject classifications*. Primary 60E15; secondary 60F10, 94A17.
Key words and phrases. Concentration of measure, concentration inequalities, deviation inequalities, empirical processes.

of proximity) which are proved by induction on the number of coordinates. Due to the variety of applications of these concentration inequalities and the high level of technicity of Talagrand's proofs, several recent papers have been devoted to the development of new approaches for deriving such results with a view to simplify the proofs (as in [11] or [16]) or to extend them to Markov chains (as in [18]). Among other fields of applications, Talagrand's approach can be used for controlling empirical processes.

1.1. *Talagrand's concentration inequalities for empirical processes.* In [21] (see Theorem 4.1), Talagrand obtained some striking concentration inequalities for the suprema of empirical processes. His result can be stated as follows.

THEOREM 1 (Talagrand [21]). *Consider n independent and identically distributed random variables $\xi_1, \dots, \xi_n$ with values in some measurable space $(\mathbb{X}, \mathscr{X})$. Let $\mathscr{F}$ be some countable family of real-valued measurable functions on $(\mathbb{X}, \mathscr{X})$, such that $\|f\|_\infty \le b < \infty$ for every $f \in \mathscr{F}$. Let $Z = \sup_{f\in\mathscr{F}} \sum_{i=1}^n f(\xi_i)$ and $v = \mathbb{E}[\sup_{f\in\mathscr{F}} \sum_{i=1}^n f^2(\xi_i)]$. Then for every positive number x,*

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le K \exp\left[-\frac{1}{K'}\frac{x}{b}\log\left(1 + \frac{xb}{v}\right)\right] \tag{2}$$

and

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le K \exp\left[-\frac{x^2}{2(c_1 v + c_2 bx)}\right], \tag{3}$$

where K, K', c_1 and c_2 are universal positive constants. Moreover, the same inequalities hold when replacing Z by $-Z$.

Of course, inequality (3) easily derives from (2) but it has its own interest. Theorem 1 can be viewed as a functional version of Bennett's or Bernstein's inequalities for sums of independent and bounded real-valued random variables. These classical inequalities (see [2]) apply under the assumptions of Theorem 1 when $\mathscr{F}$ is reduced to a single function. Bennett's inequality ensures that

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-\frac{v}{b^2} h\left(\frac{bx}{v}\right)\right], \tag{4}$$

where $h(u) = (1+u)\log(1+u) - u$ for all positive u, which in particular implies, since $2h(u) \ge u\log(1+u)$, the following bound which is directly comparable to (2):

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-\frac{x}{2b}\log\left(1 + \frac{xb}{v}\right)\right]. \tag{5}$$

Bernstein's inequality, which follows from Bennett's inequality by noticing that $2h(u) \ge u^2(1 + u/3)^{-1}$, ensures that

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-\frac{x^2}{2(v + bx/3)}\right]. \tag{6}$$

By analogy to the Gaussian case, it is natural to raise the following question.

(Q) Do Talagrand's inequalities (2) and (3) hold with the same constants as in the one-dimensional case; that is $K = 1$, $K' = 2$, $c_1 = 1$ and $c_2 = 1/3$?

Talagrand's proof of Theorem 1 is rather intricate and does not lead to very attractive values for the constants K, K', c_1 and c_2. It is the merit of Ledoux's work in [16] to provide a much simpler approach leading to deviation inequalities which are close to Theorem 1. There is therefore some hope that the answer to question (Q) could be given or at least that this question could be better understood. To be precise, it should be noticed that Ledoux failed to recover exactly Theorem 1, in the sense that his statement (see Theorem 2.5 in [16]) is analogous to that of Theorem 1 but with v taken as

$$v = \mathbb{E}\left[\sup_{f\in\mathscr{F}} \sum_{i=1}^{n} f^2(\xi_i)\right] + Cb\mathbb{E}[Z], \tag{7}$$

where C is an adequate constant ($C = 4/21$ works). Moreover, he did not provide an analogous inequality for $-Z$, which means that his inequality allows analyzing the concentration of Z around its means only from one side. However, although he did not pretend to present optimized computations, Ledoux could give sensible values for some of the constants involved in his probability bounds. In particular he could show that, taking v as in (7), (3) holds with $K = 2$, $c_1 = 42$ and $c_2 = 8$. Ledoux's approach is based on entropy inequalities for product measures which are obtained by iteration of logarithmic Sobolev type inequalities. We first would like to recall why such an approach leads to the optimal deviation inequality (1) in the Gaussian framework.

1.2. *Logarithmic Sobolev inequalities.* The connection between the concentration of measure phenomenon and logarithmic Sobolev inequalities is known as Herbst's argument was apparently pointed out for the first time by Davies and Simon in [10] (for more details on this topic we refer to [15]). Let us state Gross's logarithmic Sobolev inequality (see [12]) for the standard Gaussian measure on $\mathbb{R}^N$ and then show how it implies (1).

THEOREM 2 (Gross [12]). *Let P be the standard Gaussian measure on the Euclidean space $\mathbb{R}^N$ and u be any continuously differentiable function on $\mathbb{R}^N$. Then*

$$E_P(u^2|\log u^2|) - E_P(u^2)\log E_P(u^2) \le 2E_P(\|\nabla u\|^2). \tag{8}$$

If we now consider some Lipschitz functional ζ on the Euclidean space $\mathbb{R}^N$ with Lipschitz constant L and if we furthermore assume ζ to be continuously differentiable, we have for all x in $\mathbb{R}^N$, $\|\nabla\zeta(x)\| \le L$ and given $\lambda > 0$, we can apply (8) to $u = e^{\lambda\zeta/2}$. Since for all x in $\mathbb{R}^N$ we have

$$\|\nabla u(x)\|^2 = \frac{\lambda^2}{4}\|\nabla\zeta(x)\|^2 e^{\lambda\zeta(x)} \le \frac{\lambda^2 L^2}{4} e^{\lambda\zeta(x)},$$

we derive from (8) that

$$\lambda E_P\big[\zeta e^{\lambda\zeta}\big] - E_P\big[e^{\lambda\zeta}\big]\log E_P\big[e^{\lambda\zeta}\big] \le \frac{\lambda^2 L^2}{2} E_P\big[e^{\lambda\zeta}\big]. \tag{9}$$

This inequality holds for all positive λ and therefore yields the differential inequality

$$\frac{1}{\lambda}\frac{F'(\lambda)}{F(\lambda)} - \frac{1}{\lambda^2}\log F(\lambda) \le \frac{L^2}{2},$$

where $F(\lambda) = E_P\big[\exp(\lambda(\zeta - E_P(\zeta)))\big]$. Setting $H(\lambda) = \lambda^{-1}\log F(\lambda)$, we see that the differential inequality simply becomes $H'(\lambda) \le L^2/2$, which in turn implies since $H(\lambda)$ tends to 0 as λ tends to 0, $H(\lambda) \le \lambda L^2/2$. Hence for any positive λ,

$$E_P\big[\exp(\lambda(\zeta - E_P(\zeta)))\big] \le \exp\left(\frac{\lambda^2 L^2}{2}\right). \tag{10}$$

Using a regularization argument (by convolution), this inequality remains valid when ζ is only assumed to be Lipschitz and (1) follows by Markov's inequality.

As compared to Talagrand's approach, Ledoux's method which is based on entropy inequalities, naturally produces probability controls for the deviation of a functional from its mean rather than its median as is the case for the isoperimetric approach. Another advantage is that the proofs are much simpler, mainly because the induction argument is contained in a single tensorization inequality for entropy (see Proposition 7 below). As a counterpart, as mentioned by Ledoux himself in [16], his approach does not clearly lead to concentration inequalities but only to a deviation inequality on the right tail. Fortunately, as quoted for the first time by Samson (see [19] and, more precisely, the logarithmic Sobolev inequality for separately concave functionals, inequality (1.20) therein) in his study of the concentration of a separately convex functional around its mean, Ledoux's method also applies to deriving deviation inequalities on the left tail and therefore concentration inequalities. Although it could be possible to use Samson's new logarithmic Sobolev inequality to derive the type of concentration inequalities that we have in view here, we shall avoid it in order to get better constants.

1.3. *Some new results and their motivations.* This paper is largely inspired by Ledoux's work [16]. Our approach will consist of refining his entropy inequalities in order to get sharper probability bounds. The main argument is Lemma 8, which plays the same role in the derivation of our probability bounds as Gross's logarithmic Sobolev inequality in the Gaussian case. In particular, we exactly recover Theorem 1 and show that (3) holds with $K = 1, c_1 = 8$ and $c_2 = 2.5$. Of course we are far from providing a positive answer to question (Q) but it is of some interest to show that the constants K, c_1 and c_2 are not ridiculously large. Indeed, it turns out that precise deviation inequalities from the mean are very useful for some statistical applications and in particular for model selection in density estimation (see [3] and [4] and also [1] for an application of (2) to model selection for regression on a fixed design). In such cases one needs a more tractable formulation of (3), involving $\sup_{f\in\mathscr{F}} \mathbb{E}\left[\sum_{i=1}^n f^2(\xi_i)\right]$ instead of $\mathbb{E}\left[\sup_{f\in\mathscr{F}} \sum_{i=1}^n f^2(\xi_i)\right]$. The following theorem provides such a result.

THEOREM 3. *Consider n independent random variables $\xi_1, \dots, \xi_n$ with values in some measurable space $(\mathbb{X}, \mathscr{X})$. Let $\mathscr{F}$ be some countable family of real-valued measurable functions on $(\mathbb{X}, \mathscr{X})$, such that $\|f\|_\infty \le b < \infty$ for every $f \in \mathscr{F}$. Let Z denote either*

$$\sup_{f\in\mathscr{F}} \left| \sum_{i=1}^n f(\xi_i) \right| \quad or \quad \sup_{f\in\mathscr{F}} \left| \sum_{i=1}^n f(\xi_i) - \mathbf{E}[f(\xi_i)] \right|.$$

Let $\sigma^2 = \sup_{f\in\mathscr{F}} \sum_{i=1}^n \mathrm{Var}(f(\xi_i))$, then, for any positive real numbers ε and x,

$$\mathbb{P}\left[Z \ge (1+\varepsilon)\,\mathbb{E}[Z] + \sigma\sqrt{2\kappa x} + \kappa(\varepsilon)bx\right] \le \exp(-x), \tag{11}$$

where κ and $\kappa(\varepsilon)$ can be taken equal to $\kappa = 4$ and $\kappa(\varepsilon) = 2.5 + 32\varepsilon^{-1}$. Moreover, one also has

$$\mathbb{P}\left[Z \le (1-\varepsilon)\,\mathbb{E}[Z] - \sigma\sqrt{2\kappa' x} - \kappa'(\varepsilon)bx\right] \le \exp(-x) \tag{12}$$

where $\kappa' = 5.4$ and $\kappa'(\varepsilon) = 2.5 + 43.2\varepsilon^{-1}$

It appears that for the statistical application we have in mind (see [3] or [4] for illustrations), the crucial point is to have a minimal value for κ. It would also be of interest, although it is comparatively of minor importance, to get a better value for $\kappa(\varepsilon)$. It seems to us that $\kappa = 1$ is a reasonable conjecture, but unfortunately we were not able to prove it.

It is important to understand that concentration inequalities like those stated in Theorem 1 or Theorem 3 immediately derive from corresponding inequalities for random vector via the following device. Indeed, since one has to deal with a class of functions $\mathscr{F}$ which is at most countable, one can always write $\mathscr{F}$ as $\mathscr{F} = \{f_t, t \in \mathbb{N}^*\}$ and setting $X_{i,t} = f_t(\xi_i)$ for all $i \le n$ and all $t \in \mathbb{N}^*$, one has

$$\sup_{f\in\mathscr{F}} \sum_{i=1}^n f(\xi_i) = \lim_N \uparrow \sup_{1\le t\le N} \sum_{i=1}^n X_{i,t}$$

and

$$\sup_{f\in\mathscr{F}} \left| \sum_{i=1}^n f(\xi_i) \right| = \lim_N \uparrow \sup_{1\le t\le N} \left| \sum_{i=1}^n X_{i,t} \right|.$$

Hence, by monotone convergence it is enough to prove concentration inequalities for the independent $\mathbb{R}^N$-valued random vectors $X_1, \dots, X_n$, provided that these inequalities involve absolute constants (i.e., constants which do not depend on the dimension N). To be more concrete, using this device, one readily sees that Theorem 3 is a consequence of the following result to be proved in Section 4 (and which therefore should be considered as a more general result than Theorem 3).

THEOREM 4. *Let $X_1, \dots, X_n$ be independent random variables with values in $[-b, b]^N$ for some positive number b. Let Z denote either*

$$\sup_{1\le t\le N}\left|\sum_{i=1}^{n} X_{i,t}\right| \quad or \quad \sup_{1\le t\le N}\left|\sum_{i=1}^{n} X_{i,t} - \mathbb{E}\left[X_{i,t}\right]\right|$$

and $\sigma^2 = \sup_{1\le t\le N}\sum_{i=1}^{n}\mathrm{Var}(X_{i,t})$. Then, for any positive real numbers ε and x the following inequality holds:

$$\mathbb{P}\left[Z \ge (1+\varepsilon)\,\mathbb{E}[Z] + \sigma\sqrt{2\kappa x} + \kappa(\varepsilon)bx\right] \le \exp(-x), \tag{13}$$

where κ and $\kappa(\varepsilon)$ can be taken equal to $\kappa = 4$ and $\kappa(\varepsilon) = 2.5+32\varepsilon^{-1}$. Moreover, one also has

$$\mathbb{P}\left[Z \le (1-\varepsilon)\mathbb{E}[Z] - \sigma\sqrt{2\kappa' x} - \kappa'(\varepsilon)bx\right] \le \exp(-x), \tag{14}$$

where $\kappa' = 5.4$ and $\kappa'(\varepsilon) = 2.5 + 43.2\varepsilon^{-1}$.

So from now on, we shall focus on the heart of the problem, that is, the derivation of concentration inequalities for random vectors rather than empirical processes.

The paper is organized as follows. In Section 2 we present the main entropy inequality (Lemma 8) and derive a functional Hoeffding-type deviation inequality as a first illustration. Section 3 is devoted to the proof of Talagrand's concentration inequalities with explicit constants in the probability bounds, while the proof of Theorem 4 is given in Section 4.

2. Entropy and tensorization. The purpose of this section is to establish a simple but powerful entropy inequality. It is based on a tensorization argument which, as quoted in [16], is at the heart of Ledoux's approach and has been already used by Gross to prove his logarithmic Sobolev inequality for the standard Gaussian measure on a Euclidean space. Before introducing this argument, let us recall some well-known facts on entropy.

DEFINITION 5. Let Φ denote the function defined on $\mathbb{R}_+$ by $\Phi(u) = u\log u$. Let $(\Omega, \mathscr{A})$ be some measurable space. For any nonnegative measurable function g on $(\Omega, \mathscr{A})$ and any probability measure P such that g is P-integrable, we define the entropy of g with respect to P by

$$H_P(g) = E_P[\Phi(g)] - \Phi[E_P(g)].$$

Note that since Φ is bounded from below by $-e^{-1}$ one can always give a sense to $E_P[\Phi(g)]$ even if $\Phi(g)$ is not P-integrable and $H_P(g) < \infty$ if and only if $\Phi(g)$ is P-integrable. Moreover, since Φ is a convex function, Jensen's inequality warrants that entropy is a nonnegative quantity. A classical alternative definition of entropy (see [14]) which derives from an elementary computation will be helpful.

PROPOSITION 6. *Let $(\Omega, \mathscr{A}, P)$ be some probability space. For any nonnegative measurable function g on $(\Omega, \mathscr{A})$ such that $\Phi(g)$ is P-integrable, the following identity holds*:

$$H_P(g) = \inf_{u>0} E_P[g(\log g - \log u) - (g - u)]. \tag{15}$$

H_P is actually a convex functional. As pointed out by Bobkov in [5], this is a key property for deriving the following tensorization inequality for entropy (see [5] or [16] for a proof).

PROPOSITION 7. *Let $(\Omega_i, \mathscr{A}_i, \mu_i)_{1\le i\le n}$ be probability spaces. We consider the product probability space*

$$(\Omega, \mathscr{A}, P) = \left(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathscr{A}_i, \bigotimes_{i=1}^n \mu_i\right),$$

and some nonnegative measurable function g on $(\Omega, \mathscr{A})$ such that $\Phi(g)$ is integrable with respect to P. Given $x \in \Omega$ and $1 \le i \le n$, we denote by $g_{i,x}$ the function defined on Ω_i by $g_{i,x}(y) = g(x^i)$, where $x^i_j = x_j$ for any $j \ne i$ and $x^i_i = y$. Then

$$H_P(g) \le \sum_{i=1}^n \int_\Omega H_{\mu_i}(g_{i,x})\, dP(x). \tag{16}$$

We are now in position to prove the entropy inequality for functionals on a product probability space which is at the center of the present paper.

2.1. *Functionals of independent random variables.* It is worth mentioning that a similar inequality is already present in [16] although not explicitly stated.

LEMMA 8. *Let $(\Omega_1, \mathscr{A}_1), \dots, (\Omega_n, \mathscr{A}_n)$ be some measurable spaces and $X_1, \dots, X_n$ be independent random variables with values in $\Omega_1, \dots, \Omega_n$, respectively. Let ζ be some real-valued measurable function on $(\Omega, \mathscr{A}) = (\prod_{i=1}^n \times \Omega_i, \otimes_{i=1}^n \mathscr{A}_i)$ and $Z = \zeta(X_1, \dots, X_n)$. Given some independent random variables $X'_1, \dots, X'_n$ with values in $\Omega_1, \dots, \Omega_n$ and independent of $X_1, \dots, X_n$, let for all $1 \le i \le n$, $Z^{\vee i}$ denote the random variable $\zeta(X^i_1, \dots, X^i_n)$ where $X^i_k = X_k$, for $k \ne i$ and $X^i_i = X'_i$. Let for any real number z, $\phi(z) = \exp(z) - z - 1$ and $\psi(z) = z(\exp(z) - 1)$. If the Laplace transform $\lambda \to \mathbb{E}[e^{\lambda Z}]$ is finite on some nonempty open interval I then, for any $\lambda \in I$,*

$$\lambda \mathbb{E}\big[Ze^{\lambda Z}\big] - \mathbb{E}\big[e^{\lambda Z}\big] \log \mathbb{E}\big[e^{\lambda Z}\big] \le \sum_{i=1}^n \mathbb{E}\big[e^{\lambda Z}\phi\big(-\lambda(Z - Z^{\vee i})\big)\big]. \tag{17}$$

If, moreover $(X_1', \dots, X_n')$ *has the same distribution as* $(X_1, \dots, X_n)$, *then one also has for any* $\lambda \in I$,

$$\text{(18)}\quad \lambda \mathbb{E}\big[Ze^{\lambda Z}\big] - \mathbb{E}\big[e^{\lambda Z}\big] \log \mathbb{E}\big[e^{\lambda Z}\big] \le \sum_{i=1}^{n} \mathbb{E}\big[e^{\lambda Z} \psi\big(-\lambda(Z - Z^{\vee i})\big) \mathbb{1}_{Z - Z^{\vee i} \ge 0}\big].$$

PROOF. The main arguments of the proof are borrowed from [16]. The first step consists in using the tensorization inequality (16) that we write in a slightly different and somehow more probabilistic language. Let $X = (X_1, \dots, X_n)$, $X' = (X_1', \dots, X_n')$ and g be some nonnegative function on Ω such that $G = g(X)$ satisfies to $\mathbb{E}[G|\log G|] < \infty$. For any $1 \le i \le n$, let $\mathbb{E}^{\vee i}$ denote the expectation operator conditionally to the σ-field $\mathscr{A}_i$ generated by the variables $\{X_k, 1 \le k \le n\} \backslash \{X_i\}$. Then, recalling that $\Phi(u) = u \log u$,

$$\text{(19)}\qquad \mathbb{E}[\Phi(G)] - \Phi(\mathbb{E}[G]) \le \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{E}^{\vee i}[\Phi(G)] - \Phi\big(\mathbb{E}^{\vee i}[G]\big)\right].$$

The second step consists of using the variational definition of entropy (15). Namely, given $1 \le i \le n$,

$$\mathbb{E}^{\vee i}[\Phi(G)] - \Phi\big(\mathbb{E}^{\vee i}[G]\big) = \inf_u \mathbb{E}^{\vee i}[G(\log G - \log u) - (G - u)],$$

where the infimum in the identity above is extended to all nonnegative measurable functions u of $\{X_k, 1 \le k \le n\} \backslash \{X_i\}$. Therefore, given $\omega \in \Omega$ and choosing $u = g(X_1^i(\omega), \dots, X_n^i(\omega))$ where $X_k^i(\omega) = X_k$, for $k \ne i$ and $X_i^i(\omega) = \omega_i$, we get

$$\mathbb{E}^{\vee i}[\Phi(G)] - \Phi\big(\mathbb{E}^{\vee i}[G]\big) \le \mathbb{E}^{\vee i}[G(\log G - \log u) - (G - u)].$$

Applying the above inequality to the variable $G = e^{\lambda Z}$, with $\lambda \in I$ and integrating with respect to ω according to the distribution of X' we get by Fubini's theorem,

$$\mathbb{E}^{\vee i}[\Phi(G)] - \Phi\big(\mathbb{E}^{\vee i}[G]\big) \le \mathbb{E}^{\vee i}\big[e^{\lambda Z} \phi(-\lambda(Z - Z^{\vee i}))\big].$$

Combining this inequality with (19) leads to

$$\text{(20)}\qquad \lambda \mathbb{E}\big[Ze^{\lambda Z}\big] - \mathbb{E}\big[e^{\lambda Z}\big] \log \mathbb{E}\big[e^{\lambda Z}\big] \le \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{E}^{\vee i}\big[e^{\lambda Z} \phi\big(-\lambda(Z - Z^{\vee i})\big)\big]\right],$$

which of course implies inequality (17). Moreover, for all $i \le n$ we can write

$$\begin{aligned} e^{\lambda Z} \phi\big(-\lambda(Z - Z^{\vee i})\big) &= e^{\lambda Z} \phi\big(\lambda(Z^{\vee i} - Z)\big) \mathbb{1}_{Z^{\vee i} - Z \ge 0} \\ &\quad + e^{\lambda Z} \phi\big(-\lambda(Z - Z^{\vee i})\big) \mathbb{1}_{Z - Z^{\vee i} \ge 0}. \end{aligned}$$

However, if $(X_1', \dots, X_n')$ has the same distribution as $(X_1, \dots, X_n)$, then, conditionally to $\mathscr{A}_i$, Z and $Z^{\vee i}$ are independent and have the same distribution; hence

$$\mathbb{E}^{\vee i}\big[e^{\lambda Z} \phi\big(\lambda(Z^{\vee i} - Z)\big) \mathbb{1}_{Z^{\vee i} - Z \ge 0}\big] = \mathbb{E}^{\vee i}\big[e^{\lambda Z^{\vee i}} \phi\big(\lambda(Z - Z^{\vee i})\big) \mathbb{1}_{Z - Z^{\vee i} \ge 0}\big],$$

that we can also write

$$\begin{aligned}&\mathbb{E}^{\vee i}\big[e^{\lambda Z}\phi\big(\lambda(Z^{\vee i}-Z)\big)\mathbb{1}_{Z^{\vee i}-Z\geq 0}\big]\\&\qquad=\mathbb{E}^{\vee i}\big[e^{\lambda Z}\exp\big(-\lambda(Z-Z^{\vee i})\big)\phi\big(\lambda(Z-Z^{\vee i})\big)\mathbb{1}_{Z-Z^{\vee i}\geq 0}\big]\end{aligned}$$

and therefore, since for any $x\in\mathbb{R}$, $\psi(x)=e^x\phi(-x)+\phi(x)$,

$$\mathbb{E}^{\vee i}\big[e^{\lambda Z}\phi\big(-\lambda(Z-Z^{\vee i})\big)\big]=\mathbb{E}^{\vee i}\big[e^{\lambda Z}\psi\big(-\lambda(Z-Z^{\vee i})\big)\mathbb{1}_{Z-Z^{\vee i}\geq 0}\big].$$

Combining these identities with inequality (20) readily implies (18) and achieves the proof of the Lemma. □

Let us develop a first example illustrating the power of Lemma 8.

2.2. *A functional Hoeffding type inequality.* The following result generalizes on Ledoux's inequality (1.9) in [16].

THEOREM 9. *Let $X_1,\dots,X_n$ be independent random variables with values in $\mathbb{R}^N$. We assume that for some real numbers $a_{i,t}$ and $b_{i,t}$ such that $a_{i,t}\leq X_{i,t}\leq b_{i,t}$, for all $i\leq n$ and all $t\leq N$. Let*

$$Z=\sup_{1\leq t\leq N}\sum_{i=1}^{n}X_{i,t}$$

and define $L^2=\sup_{1\leq t\leq N}\sum_{i=1}^{n}(b_{i,t}-a_{i,t})^2$. Then, for any positive x,

$$\mathbb{P}[Z\geq\mathbb{E}[Z]+x]\leq\exp\left(-\frac{x^2}{2L^2}\right). \tag{21}$$

PROOF. We apply Lemma 8 with $X'=a$, so that inequality (17) holds. Let τ be defined as

$$\tau=\min\left\{k\leq N\colon\max_{1\leq t\leq N}\sum_{i=1}^{n}X_{i,t}=\sum_{i=1}^{n}X_{i,k}\right\}\quad\text{if } Z=\max_{1\leq t\leq N}\sum_{i=1}^{n}X_{i,t}.$$

Then for any $1\leq i\leq n$, one has

$$0\leq Z-Z^{\vee i}\leq X_{i,\tau}-a_{i,\tau}\leq b_{i,\tau}-a_{i,\tau}.$$

Now the function $u\to\phi(u)/u^2$ is nondecreasing on $\mathbb{R}$, hence $\phi(u)\leq u^2/2$ for $u\leq 0$. Therefore, for every $\lambda>0$, we derive that

$$\sum_{i=1}^{n}\mathbb{E}\big[e^{\lambda Z}\phi\big(-\lambda(Z-Z^{\vee i})\big)\big]\leq\sum_{i=1}^{n}\frac{\lambda^2}{2}\mathbb{E}\big[(b_{i,\tau}-a_{i,\tau})^2e^{\lambda Z}\big],$$

which yields via inequality (17),

$$\lambda\mathbb{E}\big[Ze^{\lambda Z}\big]-\mathbb{E}\big[e^{\lambda Z}\big]\log\mathbb{E}\big[e^{\lambda Z}\big]\leq\frac{\lambda^2L^2}{2}\mathbb{E}\big[e^{\lambda Z}\big]. \tag{22}$$

This differential inequality has the same structure as (9) so that we can derive exactly as in the Gaussian case that, for any $\lambda \geq 0$,

$$\mathbb{E}\big[\exp(\lambda(Z - \mathbb{E}[Z]))\big] \leq \exp\left(\frac{\lambda^2 L^2}{2}\right)$$

and the result follows by Markov's inequality. $\square$

Let us make some comments about Theorem 9.

REMARK 1. (i) This result in particular applies to the situation where the variables ξ_i's are valued in $[-1, 1]$, $\{\alpha_{i,t}, 1 \leq t \leq N,\ 1 \leq i \leq n\}$ is a family of real numbers and $X_{i,t} = \alpha_{i,t}\xi_i$ for all $t \leq N$ and all $i \leq n$. Then the random variable

$$Z = \sup_{1 \leq t \leq N} \sum_{i=1}^{n} \alpha_{i,t}\xi_i$$

obeys the deviation inequality

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + x] \leq \exp\left(-\frac{x^2}{8\sigma^2}\right),$$

for any positive x, where $\sigma^2 = \sup_{1 \leq t \leq N}(\sum_{i=1}^{n} \alpha_{i,t}^2)$. We then recover Ledoux's inequality (1.9) in [16].

(ii) The classical Hoeffding inequality (see [13]) ensures that under the assumptions of Theorem 9, when $N = 1$, then the variable $Z = \sum_{i=1}^{n} X_i$ satisfies

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + x] \leq \exp\left(-\frac{2x^2}{L^2}\right) \tag{23}$$

for every positive x. When we compare it to inequality (23), we see that a factor 4 is lost in the exponent of the upper bound in (21). Whether inequality (23) holds in the general framework covered by Theorem 9 remains an open problem.

We pass now to the main issue of the paper which consists in providing a rather elementary proof of Theorem 1 with explicit values for the constants in the probability bounds.

3. Talagrand's concentration inequalities.

As announced, the method of proof that we present follows closely the approach proposed by Ledoux in [16] but we add some new calculations with respect to Ledoux's work in order to get better constants in the probability bounds. We shall use the fundamental entropy lemma (Lemma 8) twice. The first step consists in applying Lemma 8 to nonnegative variables.

3.1. *Nonnegative variables.* It might seem useless to build deviation inequalities for sums of nonnegative random variables when the target is to deal with centered random variables. However, the behavior of the supremum of sums of centered random variables will turn out to depend heavily on that of the associated sums of the squared variables. This is indeed the main motivation for the following Poissonian bound (see also [7] for extensions of this result to some other nonnegative functionals of independent variables with applications to random combinatorics).

THEOREM 10. *Let $X_1, \dots, X_n$ be independent random variables with values in $[0,1]^N$. We consider $Z = \sup_{1\le t\le N} \sum_{i=1}^n X_{i,t}$ and define the function h as $h(u) = (1+u)\log(1+u) - u$, for $u \ge 0$. Then, for every positive λ, setting $\phi(\lambda) = \exp(\lambda) - \lambda - 1$,*

$$\log \mathbb{E}\big[\exp(\lambda(Z - \mathbb{E}[Z]))\big] \le v\phi(\lambda), \tag{24}$$

which implies that for any positive number x,

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-\mathbb{E}[Z] h\left(\frac{x}{\mathbb{E}[Z]}\right)\right]. \tag{25}$$

PROOF. We apply Lemma 8 with $X_i' = 0$ for all $i \le n$ so that inequality (17) holds for any positive λ. Now, defining

$$\tau = \min\left\{k \le N \max_{1\le t\le N} \sum_{i=1}^n X_{i,t} = \sum_{i=1}^n X_{i,k}\right\},$$

we have for any $1 \le i \le n$,

$$0 \le Z - Z^{\vee i} \le X_{i,\tau} \le 1. \tag{26}$$

Since the function ϕ is convex, for any positive λ and any $u \in [0,1]\phi(-\lambda u) \le u\phi(-\lambda)$, it follows from (17) and (26) that for any positive λ,

$$\begin{aligned}\lambda\mathbb{E}\big[Ze^{\lambda Z}\big] - \mathbb{E}\big[e^{\lambda Z}\big]\log\mathbb{E}\big[e^{\lambda Z}\big] &\le \mathbb{E}\left[\phi(-\lambda)e^{\lambda Z}\sum_{i=1}^n X_{i,\tau}\right]\\ &\le \phi(-\lambda)\mathbb{E}\big[Ze^{\lambda Z}\big].\end{aligned}$$

If we introduce $\widetilde{Z} = Z - \mathbb{E}[Z]$, then $F(\lambda) = \mathbb{E}[e^{\lambda\widetilde{Z}}]$ for any positive λ and setting $v = \mathbb{E}[Z]$, the preceding inequality becomes

$$[\lambda - \phi(-\lambda)]\frac{F'(\lambda)}{F(\lambda)} - \log F(\lambda) \le v\phi(-\lambda), \tag{27}$$

which in turn implies

$$(1 - e^{-\lambda})\Psi'(\lambda) - \Psi(\lambda) \le v\phi(\lambda) \quad \text{with } \Psi(\lambda) = \log F(\lambda).$$

Now observe that the function $\Psi_0 = v\phi$ is a solution of the ordinary differential equation $(1-e^{-\lambda})\Psi'(\lambda) - \Psi(\lambda) = v\phi(-\lambda)$. We want to show that $\Psi \le \Psi_0$. In fact, if $\Psi_1 = \Psi - \Psi_0$, then

$$(1-e^{-\lambda})\Psi_1'(\lambda) - \Psi_1(\lambda) \le 0. \tag{28}$$

Hence, defining $f(\lambda) = \log(e^{\lambda}-1)$ and $g(\lambda) = e^{-f(\lambda)}\Psi_1(\lambda)$, we have

$$(1-e^{-\lambda})[f'(\lambda)g(\lambda) + g'(\lambda)] - g(\lambda) \le 0,$$

which means since $f'(\lambda)(1-e^{-\lambda}) = 1$, that g' is nonpositive and therefore that g is nonincreasing on $(0,\infty)$. Now, since $\widetilde{Z}$ is centered, $\Psi_1'(0) = 0$. Using the fact that $\lambda e^{-f(\lambda)}$ tends to 1 as λ goes to 0, we conclude that $g(\lambda)$ tends to 0 as λ goes to 0. This shows that g is nonpositive; therefore $\Psi \le \Psi_0$ and inequality (24) is proved. This readily implies (25) by Markov's inequality. Indeed,

$$\mathbb{P}[Z - \mathbb{E}[Z] \ge x] \le \exp\Big[-\sup_{\lambda>0}(x\lambda - v\phi(\lambda))\Big]$$

and we use the easy-to-check (and well-known) relation $\sup_{\lambda>0}[x\lambda - v\phi(\lambda)] = vh(x/v)$. $\square$

Let us comment on this result.

REMARK 2. (i) Theorem 10 provides a more explicit version of one of Ledoux's inequalities (see precisely Theorem 2.4 in [16]), which ensures that

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-\eta x \log\left(1 + \frac{x}{\mathbb{E}[Z]}\right)\right],$$

where η is some numerical constant. As a matter of fact, since $h(u) \ge (u/2)\times \log(1+u)$ for any positive u, inequality (25) ensures that η can be taken as 1/2 in Ledoux's inequality.

(ii) It is important to notice that inequality (25) is in some sense unimprovable. Indeed, let us consider the situation where $N = 1$ and $X_1, \dots, X_n$ are independent Bernoulli trials with probability of success $p = 1 - q$. Then the classical Bennett inequality [see inequality (4)] ensures that, setting

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-npqh\left(\frac{x}{npq}\right)\right]$$

and since $npq \le \mathbb{E}[Z]$, one gets

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + x] \le \exp\left[-\mathbb{E}[Z]h\left(\frac{x}{\mathbb{E}[Z]}\right)\right],$$

which is exactly (25). Given $\theta > 0$, taking $p = \theta/n$ and setting $x = \theta\varepsilon$, this inequality can be written as

$$\mathbb{P}[Z \ge \theta + \theta\varepsilon] \le \exp[-\theta h(\varepsilon)] \quad \text{for any positive } \varepsilon. \tag{29}$$

But Z follows the binomial distribution $\mathscr{B}(n, \theta/n)$ and therefore follows asymptotically the Poisson distribution with parameter θ as n goes to infinity. Moreover the right-hand side of (29) is known to be the Cramér–Chernoff deviation upper bound for a Poisson random variable with parameter θ. This implies that the exponent in this upper bound cannot be improved since Cramér's large deviation asymptotic ensures that for any positive ε,

$$\liminf_{\theta\to\infty} \lim_{n\to\infty} \frac{1}{\theta} \log \mathbb{P}[Z \geq \theta + \theta\varepsilon] \geq -h(\varepsilon).$$

Applying Lemma 8 again together with Theorem 10 it is possible to prove concentration inequalities for sums of bounded (and possibly centered) random vectors.

3.2. *Explicit constants in Talagrand's inequalities.* We shall need the following technical result which might be of independent interest.

LEMMA 11. *Let V and Y be some random variables, some probability space $(\Omega, \mathscr{A}, \mathscr{P})$ and $\lambda > 0$ such that $e^{\lambda V}$ and $e^{\lambda Y}$ are P-integrable. Then, one has*

$$\frac{\lambda \mathbb{E}[Ve^{\lambda Y}]}{\mathbb{E}[e^{\lambda Y}]} \leq \frac{\lambda \mathbb{E}[Ye^{\lambda Y}]}{\mathbb{E}[e^{\lambda Y}]} + \log \mathbb{E}[e^{\lambda V}] - \log \mathbb{E}[e^{\lambda Y}]. \tag{30}$$

PROOF. Let Q be the probability distribution defined by

$$dQ = \frac{e^{\lambda Y}}{\mathbb{E}[e^{\lambda Y}]} dP,$$

then by Jensen's inequality,

$$\lambda \mathbb{E}_Q[V - Y] \leq \log \mathbb{E}_Q\bigl[e^{\lambda(V-Y)}\bigr],$$

which is exactly equivalent to (30). □

We begin by proving functional Bernstein-type inequalities.

THEOREM 12. *Let $X_1, \dots, X_n$ be independent random variables with values in $[-b, b]^N$ for some positive number b. Let Z denote either*

$$\sup_{1\leq t\leq N} \sum_{i=1}^n X_{i,t} \quad or \quad \sup_{1\leq t\leq N} \left| \sum_{i=1}^n X_{i,t} \right|.$$

Let $(X'_1, \dots, X'_n)$ be independent from $(X_1, \dots, X_n)$ and have the same distribution as $(X_1, \dots, X_n)$. Setting

$$v = \mathbb{E}\left[\sum_{1\leq t\leq N} \sum_{i=1}^n (X_{i,t} - X'_{i,t})^2 \right],$$

we get for any positive number x,

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + 2\sqrt{vx} + 2cbx] \leq \exp(-x) \tag{31}$$

and

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + x] \leq \exp\left(- \frac{x^2}{4(v + cbx)}\right) \tag{32}$$

where the constant c can be taken equal to 5/4. *Moreover, setting* $c' = 7/4$ *and* $\gamma = \sqrt{c'(\exp(1/c') - 1)}$, *one also has*

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - 2\gamma\sqrt{vx} - 2c'bx] \leq \exp(-x) \tag{33}$$

and

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - x] \leq \exp\left(- \frac{x^2}{4(v\gamma^2 + c'bx)}\right). \tag{34}$$

(*Note that* $\gamma^2 < 1.35$).

PROOF. By homogeneity we can take $b = 1/2$. Let τ be defined either as

$$\tau = \min\left\{k \leq N \colon \max_{1\leq t\leq N} \sum_{i=1}^{n} X_{i,t} = \sum_{i=1}^{n} X_{i,k}\right\} \quad \text{if } Z = \max_{1\leq t\leq N} \sum_{i=1}^{n} X_{i,t}$$

or as

$$\tau = \min\left\{k \leq N \colon \max_{1\leq t\leq N} \left|\sum_{i=1}^{n} X_{i,t}\right| = \left|\sum_{i=1}^{n} X_{i,k}\right|\right\} \quad \text{if } Z = \max_{1\leq t\leq N} \left|\sum_{i=1}^{n} X_{i,t}\right|.$$

We apply Lemma 8. Then inequality (18) holds for any $\lambda \in \mathbb{R}$ and we therefore have to control

$$\sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda Z}\psi\left(- \lambda(Z - Z^{\vee i})\right) \mathbb{1}_{Z - Z^{\vee i} \geq 0}\right]$$

for any $\lambda \in \mathbb{R}$, where we recall that $\psi(x) = x(e^x - 1)$. To do so, it is useful to note the following crucial, though elementary, property of the function ψ:

$$\text{the function } x \to x^{-2}\psi(x) \text{ is increasing on } \mathbb{R} \tag{35}$$

(taking the value of this function at point 0 as 1).

Let us now first focus on the proof of the deviation inequalities (31) and (32) and assume that $\lambda > 0$. We observe that property (35) implies that $\psi(-u) \leq u^2$ for any $u \geq 0$ and since for any $1 \leq i \leq n$, $Z - Z^{\vee i} \leq |X_{i,\tau} - X'_{i,\tau}|$, we derive that

$$\psi(-\lambda(Z - Z^{\vee i}))\mathbb{1}_{Z - Z^{\vee i} \geq 0} \leq \lambda^2\left(X_{i,\tau} - X'_{i,\tau}\right)^2,$$

which leads via (18) to

$$\lambda\mathbb{E}[Ze^{\lambda Z}] - \mathbb{E}[e^{\lambda Z}]\log\mathbb{E}[e^{\lambda Z}] \leq \lambda^2 \sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda Z}\left(X_{i,\tau} - X'_{i,\tau}\right)^2\right].$$

This implies that, for any $\lambda > 0$,

$$\lambda\mathbb{E}[Ze^{\lambda Z}] - \mathbb{E}[e^{\lambda Z}]\log\mathbb{E}[e^{\lambda Z}] \leq \lambda^2\mathbb{E}[Ve^{\lambda Z}], \tag{36}$$

with $V = \sup_{1\le t\le N} \sum_{i=1}^{n} \left(X_{i,t} - X'_{i,t}\right)^2$. Setting $\widetilde{Z} = Z - \mathbb{E}[Z]$, inequality (36) becomes

$$\frac{\lambda \mathbb{E}\big[\widetilde{Z} e^{\lambda \widetilde{Z}}\big]}{\mathbb{E}\big[e^{\lambda \widetilde{Z}}\big]} - \log \mathbb{E}\big[e^{\lambda \widetilde{Z}}\big] \le \lambda^2 \frac{\mathbb{E}\big[V e^{\lambda \widetilde{Z}}\big]}{\mathbb{E}\big[e^{\lambda \widetilde{Z}}\big]}.$$

We now have to control the quantity $\mathbb{E}[Ve^{\lambda \widetilde{Z}}]$. To do so we use Lemma 11 and get

$$\frac{\lambda \mathbb{E}\big[V e^{\lambda \widetilde{Z}}\big]}{\mathbb{E}\big[e^{\lambda \widetilde{Z}}\big]} \le \frac{\lambda \mathbb{E}\big[\widetilde{Z} e^{\lambda \widetilde{Z}}\big]}{\mathbb{E}\big[e^{\lambda \widetilde{Z}}\big]} + \log \mathbb{E}\big[e^{\lambda V}\big] - \log \mathbb{E}\big[e^{\lambda \widetilde{Z}}\big].$$

Hence

$$\lambda(1-\lambda)\frac{\mathbb{E}\big[\widetilde{Z} e^{\lambda \widetilde{Z}}\big]}{\mathbb{E}\big[e^{\lambda \widetilde{Z}}\big]} - (1-\lambda)\log \mathbb{E}\big[e^{\lambda \widetilde{Z}}\big] \le \lambda \log \mathbb{E}\big[e^{\lambda V}\big]$$

and setting, for any positive λ, $F(\lambda) = \mathbb{E}[e^{\lambda \widetilde{Z}}]$, this inequality means that for any $0 < \lambda < 1$,

$$\lambda^{-1}\frac{F'(\lambda)}{F(\lambda)} - \lambda^{-2}\log F(\lambda) \le \frac{\log \mathbb{E}\big[e^{\lambda V}\big]}{\lambda(1-\lambda)}.$$

Integrating this inequality (taking into account that $\lambda^{-1}\log F(\lambda)$ tends to 0 as λ goes to 0) yields to

$$\lambda^{-1}\log F(\lambda) \le \int_0^\lambda \frac{\log \mathbb{E}\big[e^{uV}\big]}{u(1-u)}\,du \le (1-\lambda)^{-1}\int_0^\lambda \frac{\log \mathbb{E}\big[e^{uV}\big]}{u}\,du. \tag{37}$$

Now recalling that $b = 1/2$, the results of Theorem 10 apply to V. In particular we derive from (24) that, setting $v = \mathbb{E}[V]$, one has $\log \mathbb{E}[e^{uV}] \le vu + v\phi(u)$ which implies via (37) that

$$(1-\lambda)\log F(\lambda) \le \lambda^2 v\left[1 + \frac{1}{\lambda}\int_0^\lambda \frac{\phi(u)}{u}\,du\right]. \tag{38}$$

Let us now check that

$$B(\lambda) = \left[1 + \frac{1}{\lambda}\int_0^\lambda \frac{\phi(u)}{u}\,du\right] \le (1-\lambda)(1-c\lambda)^{-1} = C(\lambda)$$

for any $0 < \lambda \le c^{-1}$. Expanding ϕ in power series easily yields the following expansion for B: $B(\lambda) = \sum_{k=0}^{\infty} b_k \lambda^k$, with $b_0 = 1$ and

$$b_k = \frac{1}{(k+1)(k+1)!} \quad \text{for } k \ge 1.$$

Moreover, the power series expansion of C can be written as $C(\lambda) = \sum_{k=0}^{\infty} c_k \lambda^k$, with $c_0 = 1$ and for $k \ge 1$, $c_k = c^k(1 - c^{-1})$. We note that $b_0 = c_0$ and that c can be chosen in a way that $b_1 = c_1$, that is $c = 5/4$. This is in fact enough to imply that $b_k \le c_k$ for any $k \ge 1$ because $(b_k)_{k\ge 1}$ is a nonincreasing sequence while $(c_k)_{k\geq 1}$ is nondecreasing. Therefore $B(\lambda) \leq C(\lambda)$ for any $0 < \lambda \leq c^{-1}$ which implies by (38) that

$$\log F(\lambda) \leq \frac{2\lambda^2 v}{2(1 - c\lambda)}.$$

Such a control for the Laplace transform of $\widetilde{Z} = Z - \mathbb{E}[Z]$ yields via Chernoff's inequality,

$$\mathbb{P}[Z - \mathbb{E}[Z] \geq z] \leq \exp\left[-\sup_{0<\lambda<c^{-1}}\left(z\lambda - \frac{2\lambda^2 v}{2(1-c\lambda)}\right)\right],$$

for any positive z, with

$$\sup_{0<\lambda<c^{-1}}\left(z\lambda - \frac{2\lambda^2 v}{2(1-c\lambda)}\right) = \frac{2v}{c^2} h_1\left(\frac{cz}{2v}\right),$$

where $h_1(u) = 1 + u - \sqrt{1+2u}$ for $u > 0$. So we get

$$\mathbb{P}[Z - \mathbb{E}[Z] \geq z] \leq \exp\left[-\frac{2v}{c^2} h_1\left(\frac{cz}{2v}\right)\right]. \tag{39}$$

Now it is easy to check that on the one hand the inverse function of h_1 is equal to $h_1^{-1}(u) = u + \sqrt{2u}$ for $u > 0$ and on the other hand $2h_1(u) \geq u^2/(1+u)$ for $u > 0$. This immediately leads to (31) and (32) via (39). As for the deviation inequality on the left tail, the preceding proof for the right tail needs to be slightly modified. We use inequality (18) again (changing λ into $-\lambda$) and get for any $\lambda \geq 0$,

$$\lambda\mathbb{E}[-Ze^{-\lambda Z}] - \mathbb{E}[e^{-\lambda Z}]\log\mathbb{E}[e^{-\lambda Z}] \leq \mathbb{E}\left[e^{-\lambda Z}\sum_{i=1}^{n}\psi\big(\lambda(Z - Z^{\vee i})\big)\mathbb{1}_{Z - Z^{\vee i}\geq 0}\right].$$

Recalling again that $b = 1/2$, we observe that $Z - Z^{\vee i} \leq |X_{i,\tau} - X'_{i,\tau}| \leq 1$ and using the monotonicity property (35), we get for all $0 \leq \lambda < c'^{-1}$,

$$\psi\big(\lambda(Z - Z^{\vee i}\big)\mathbb{1}_{Z - Z^{\vee i}\geq 0} \leq c'^2\psi(c'^{-1})\lambda^2\big(X_{i,\tau} - X'_{i,\tau}\big)^2.$$

Hence, noting that $\gamma = c'\sqrt{\psi(c'^{-1})}$,

$$\lambda\mathbb{E}[-Ze^{-\lambda Z}] - \mathbb{E}[e^{-\lambda Z}]\log\mathbb{E}[e^{-\lambda Z}] \leq \gamma^2\lambda^2\mathbb{E}\left[e^{-\lambda Z}\sum_{i=1}^{n}\big(X_{i,\tau} - X'_{i,\tau}\big)^2\right]$$

and therefore

$$\lambda\mathbb{E}[-Ze^{-\lambda Z}] - \mathbb{E}[e^{-\lambda Z}]\log\mathbb{E}[e^{-\lambda Z}] \leq \gamma^2\lambda^2\mathbb{E}[e^{-\lambda Z}V] \quad \text{for } 0 \leq \lambda < c'^{-1}.$$

This inequality is the analogue to (36) apart from the extra factor γ^2. We can therefore use the previous computations to derive the following control on the Laplace transform F of $\widetilde{Z} = -Z + \mathbb{E}[Z]$ [which is the analogue of (37)]:

$$(1 - \gamma^2\lambda)\log F(\lambda) \leq \gamma^2\lambda^2 v\left[1 + \frac{1}{\lambda}\int_0^{\lambda}\frac{\phi(u)}{u}\,du\right]. \tag{40}$$

Expanding in power series as before, we derive from (40) that

$$\log F(\lambda) \leq \frac{2\lambda^2 v\gamma^2}{2(1-c'\lambda)} \quad \text{for any } 0 \leq \lambda < c'^{-1}$$

provided that c' satisfies to the condition $c' - \gamma^2 = c' - c'^2\psi(c'^{-1}) \geq 1/4$, which actually holds true whenever $c' = 7/4$. So we get, using the same arguments as in the proof of the deviation inequality for the right tail,

$$\mathbb{P}[-Z + \mathbb{E}[Z] \geq z] \leq \exp\left[-\frac{2v\gamma^2}{c'^2} h_1\left(\frac{c'z}{2v\gamma^2}\right)\right],$$

which finishes the proof since the inverse function of h_1 is equal to $h_1^{-1}(u) = u + \sqrt{2u}$ for $u > 0$ and $2h_1(u) \geq u^2/(1+u)$ for $u > 0$.

It is quite easy to derive Talagrand's deviation inequalities (3) and (2) from Theorem 10 and Theorem 12.

COROLLARY 13. *Let $X_1, \dots, X_n$ be independent random variables with values in $[-b, b]^N$, for some positive number b. Let Z denote either*

$$\sup_{1\leq t\leq N} \sum_{i=1}^n X_{i,t} \quad \text{or} \quad \sup_{1\leq t\leq N} \left|\sum_{i=1}^n X_{i,t}\right|.$$

Setting $v = \mathbb{E}\big[\sup_{1\leq t\leq N}\sum_{i=1}^n X_{i,t}^2\big]$, one has

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + x] \leq \exp\left[-\frac{x^2}{4(4v + cbx)}\right] \quad \text{for } x > 0 \tag{41}$$

and

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + x] \leq 2\exp\left[-\eta\frac{x}{b}\log\left(1 + \eta\frac{bx}{v}\right)\right] \quad \text{for } x > 0, \tag{42}$$

where c and η are numerical constants. In fact, one can take $c = 5/4$ and $\eta = [4(2+\sqrt{c})]^{-1} > 2/25$. Moreover, setting $c' = 7/4$, $\gamma^2 = c'(\exp(1/c') - 1) < 1.35$ and $\eta' = [4(2+\sqrt{c'})]^{-1} > 1/14$, one also has

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - x] \leq \exp\left[-\frac{x^2}{4(4\gamma^2 v + c'bx)}\right] \quad \text{for } x > 0 \tag{43}$$

and

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - x] \leq 2\exp\left[-\eta'\frac{x}{b}\log\left(1 + \eta'\frac{bx}{\gamma^2 v}\right)\right] \quad \text{for } x > 0. \tag{44}$$

PROOF. The proof of inequalities (41) or (43) is immediate. We simply start from Theorem 12, write

$$\mathbb{E}\left[\sup_{1\leq t\leq N}\sum_{i=1}^n \left(X_{i,t} - X'_{i,t}\right)^2\right] \leq 4\mathbb{E}\left[\sup_{1\leq t\leq N}\sum_{i=1}^n X_{i,t}^2\right]$$

and get the result. The proof of inequality (42) is more involved. By homogeneity we can take $b = 1$. We use the truncation argument introduced in [21] in the same way as in [16]. Let ρ, z and z' be positive numbers to be chosen later. We define Z_ρ as

$$\sup_{1\le t\le N}\sum_{i=1}^{n} X_{i,t}\,\mathbb{1}_{|X_{i,t}|\le\rho} \quad \text{or} \quad \sup_{1\le t\le N}\left|\sum_{i=1}^{n} X_{i,t}\,\mathbb{1}_{|X_{i,t}|\le\rho}\right|$$

and

$$W=\sup_{t\in T}\sum_{i=1}^{n}|X_{i,t}|\,\mathbb{1}_{|X_{i,t}|>\rho}.$$

Then by definition $|Z-Z_\rho|\le W$, hence $Z\le Z_\rho+W$ and $\mathbb{E}[Z]\ge\mathbb{E}[Z_\rho]-\mathbb{E}(W)$. Therefore,

$$\mathbb{P}[Z-\mathbb{E}[Z]\ge z+z'+2\mathbb{E}[W]]\le\mathbb{P}[Z_\rho-\mathbb{E}[Z_\rho]\ge z]+\mathbb{P}[W-\mathbb{E}[W]\ge z'] \tag{45}$$

and applying Theorem 12 to Z_ρ and Theorem 10 to W, we get

$$\mathbb{P}[Z_\rho-\mathbb{E}[Z_\rho]\ge z]\le\exp\left[-\eta x\log\left(1+\eta\frac{x}{v}\right)\right] \tag{46}$$

by setting

$$z=4\sqrt{v\eta x\log\left(1+\eta\frac{x}{v}\right)}+2c\rho\eta x\log\left(1+\eta\frac{x}{v}\right)$$

and

$$\mathbb{P}[W-\mathbb{E}[W]\ge z']\le\exp\left[-\mathbb{E}[W]h\left(\frac{z'}{\mathbb{E}[W]}\right)\right].$$

Recalling that $h(u)\ge(u/2)\log(1+u)$ and noting that $\mathbb{E}[W]\le v/\rho$, we derive that

$$\mathbb{P}[W-\mathbb{E}[W]\ge z']\le\exp\left[-\frac{z'}{2}\log\left(1+\frac{z'\rho}{v}\right)\right].$$

Since $\log(1+u)\ge(1/2)\log(1+u^2)$, it becomes

$$\mathbb{P}[W-\mathbb{E}[W]\ge z']\le\exp\left[-\frac{z'}{4}\log\left(1+\left(\frac{z'\rho}{v}\right)^2\right)\right]$$

and

$$\mathbb{P}[W-\mathbb{E}[W]\ge z']\le\exp\left[-\frac{z'}{4}\log\left(1+\frac{z'}{4v}\right)\right]\quad\text{if } \rho\ge\sqrt{v/4z'}.$$

Let us choose $z'=4\eta x$ and $\rho=[\sqrt{c}\,\log(1+\eta x/v)]^{-1}$ then, since $\log(1+u)\le\sqrt{u}$ and $\sqrt{c}<4$ one has $\rho\ge\sqrt{v/4z'}$ which yields

$$\mathbb{P}[W-\mathbb{E}[W]\ge z']\le\exp\left[-\eta x\log\left(1+\eta\frac{x}{v}\right)\right]. \tag{47}$$

Collecting inequalities (45), (46) and (47) we get

$$\mathbb{P}[Z-\mathbb{E}[Z]\ge z+z'+2\mathbb{E}[W]]\le 2\,\exp\left[-\eta x\log\left(1+\eta\frac{x}{v}\right)\right],$$

so that the proof will be completed if we can show that $z + z' + 2\mathbb{E}[W] \le x$. Recalling that $\mathbb{E}[W] \le v/\rho$ and replacing z, z' and ρ by their values we get

$$z + z' + 2\mathbb{E}[W] \le 4\sqrt{v\eta x \log\Bigl(1 + \eta\frac{x}{v}\Bigr)} + 2\sqrt{c}\eta x + 4\eta x \\ + 2v\sqrt{c}\log\Bigl(1 + \eta\frac{x}{v}\Bigr),$$

which yields $z + z' + 2\mathbb{E}[W] \le 4(2 + \sqrt{c})\eta x$, since $\log(1 + u) \le u$. This allows the conclusion since our choice of η implies that $4(2+\sqrt{c})\eta x = x$. The proof of inequality (44) can be performed exactly in the same way, using (33) instead of (31) to bound Z_ρ. $\square$

We can now compare Corollary 13 with Talagrand's deviation inequalities as stated in Theorem 1.

REMARK 3. (i) The interesting feature of (41) as compared to (3) is that it proposes sensible numerical values for the unknown constants K, c_1 and c_2, namely, $K = 1$, $c_1 = 8$ and $c_2 = 2.5$. Moreover, at the price of enlarging c_1 to $8\gamma^2 < 11$ and c_2 to 3.5, we know from (43) that (3) also holds when replacing Z by $-Z$.

(ii) We do not pretend anything about the optimality of these values which just come out from our calculations, and the question (Q) in Section 1 remains open.

(iii) Although we did not try to actually optimize the constants η and η' involved in (42) we have done our best to preserve legibility while trying to convince the reader that one can hope that Talagrand's deviation inequality (2) holds with reasonable constants.

4. A ready to be used inequality. In order to make the use of Theorem 12 or Corollary 13 more convenient, it would be desirable to replace

$$\mathbb{E}\Biggl[\sup_{1\le t\le N}\sum_{i=1}^{n}(X_{i,t} - X'_{i,t})^2\Biggr] \quad \text{or} \quad \mathbb{E}\Biggl[\sup_{1\le t\le N}\sum_{i=1}^{n} X_{i,t}^2\Biggr]$$

by $\sup_{1\le t\le N}\mathbb{E}[\sum_{i=1}^{n} X_{i,t}^2]$. This can be done at the price of additional technicalities related to classical symmetrization and contraction inequalities that we recall below. For a proof of these inequalities we refer to the book by Ledoux and Talagrand (see [17], Lemma 6.3 and Theorem 4.12 therein). We first recall that a map $\theta\colon \mathbb{R} \to \mathbb{R}$ is called a contraction if

$$|\theta(u) - \theta(v)| \le |u - v| \quad \text{for all } u, v \in \mathbb{R}.$$

LEMMA 14. *Let $F\colon \mathbb{R}_+ \to \mathbb{R}_+$ be convex and increasing. Let $Y_1, \dots, Y_n$ be independent random variables with values in $\mathbb{R}^N$ and let $\varepsilon_1, \dots, \varepsilon_n$ be independent Rademacher variables such that $(\varepsilon_i)_{1\le i\le n}$ is independent from $(Y_i)_{1\le i\le n}$.*

Assume that, for all $i \le n$, Y_i *is almost surely bounded. If*

$$Z = \sup_{1\le t\le N}\left|\sum_{i=1}^{n} Y_{i,t} - \mathbb{E}[Y_{i,t}]\right| \quad \textit{and} \quad \widetilde{Z} = \sup_{1\le t\le N}\left|\sum_{i=1}^{n} \varepsilon_i Y_{i,t}\right|,$$

then

$$\mathbb{E}\big[F\big(\tfrac{1}{2}\widetilde{Z}\big)\big] \le \mathbb{E}[F(Z)] \tag{48}$$

and

$$\mathbb{E}[F(Z)] \le \mathbb{E}\big[F\big(2\widetilde{Z}\big)\big]. \tag{49}$$

If, moreover, θ *is a contraction such that* $\theta(0) = 0$, *then*

$$\mathbb{E}\left[F\left(\tfrac{1}{2}\sup_{1\le t\le N}\left|\textstyle\sum_{i=1}^{n} \varepsilon_i \theta(Y_{i,t})\right|\right)\right] \le \mathbb{E}\big[F\big(\widetilde{Z}\big)\big]. \tag{50}$$

Combining these inequalities we get the corollary.

COROLLARY 15. *Let* $Y_1, \dots, Y_n$ *be independent random variables with values in* $[-1, 1]^N$ *such that* $\mathbb{E}[Y_{i,t}] = 0$ *for all* $t \le N$ *and* $i \le n$. *Then*

$$\mathbb{E}\left[\sup_{1\le t\le N}\sum_{i=1}^{n} Y_{i,t}^2\right] \le \sup_{1\le t\le N}\mathbb{E}\left[\sum_{i=1}^{n} Y_{i,t}^2\right] + 16\mathbb{E}\left[\sup_{1\le t\le N}\left|\sum_{i=1}^{n} Y_{i,t}\right|\right]. \tag{51}$$

If, moreover, the distribution of $Y_{i,t}$ *is symmetric around* 0 *for all* $i \le n$ *and all* $t \le N$, *then this inequality can be improved and one has*

$$\mathbb{E}\left[\sup_{1\le t\le N}\sum_{i=1}^{n} Y_{i,t}^2\right] \le \sup_{1\le t\le N}\mathbb{E}\left[\sum_{i=1}^{n} Y_{i,t}^2\right] + 8\mathbb{E}\left[\sup_{1\le t\le N}\left|\sum_{i=1}^{n} Y_{i,t}\right|\right]. \tag{52}$$

PROOF. By the symmetrization inequality (49) we can write

$$\mathbb{E}\left[\sup_{1\le t\le N}\left|\sum_{i=1}^{n} Y_{i,t}^2 - \mathbb{E}[Y_{i,t}^2]\right|\right] \le 2\mathbb{E}\left[\sup_{1\le t\le N}\left|\sum_{i=1}^{n} \varepsilon_i Y_{i,t}^2\right|\right].$$

Now the function $\theta\colon \mathbb{R} \to \mathbb{R}$ defined by $\theta(u) = (u^2 \wedge 1)/2$ is a contraction and we get from (50),

$$\mathbb{E}\left[\sup_{1\le t\le N}\left|\sum_{i=1}^{n} \varepsilon_i Y_{i,t}^2\right|\right] \le 4\mathbb{E}\left[\sup_{1\le t\le N}\left|\sum_{i=1}^{n} \varepsilon_i Y_{i,t}^2\right|\right],$$

which leads to the result either by symmetry or via the symmetrization inequality (48). $\square$

Inequality (51) in Corollary 15 can obviously be used to bound v in inequality (42) while combining inequality (52) with Theorem 12 leads to Theorem 4 which is the easy-to-use inequality announced in Section 1.

4.1. *Proof of Theorem* 4. By homogeneity, we can assume that $b = 1/2$. Applying inequality (31) of Theorem 12 and Corollary 15 if we set for all $i \leq n$ and $t \leq N$,

$$Y_{i,t} = X_{i,t} - X'_{i,t},$$

we get by (52),

$$\mathbb{E}\left[\sup_{1\leq t\leq N} \sum_{i=1}^{n} Y_{i,t}^2\right] \leq 2\sigma^2 + 16\mathbb{E}[Z]$$

and therefore

$$\mathbb{P}\Big[Z \geq \mathbb{E}[Z] + 2\sqrt{x(2\sigma^2 + 16\mathbb{E}[Z])} + cx\Big] \leq \exp(-x).$$

Now we note that

$$\sqrt{x(2\sigma^2 + 16\mathbb{E}[Z])} \leq \sigma\sqrt{2x} + 4\sqrt{x\mathbb{E}[Z]} \leq \sigma\sqrt{2x} + \frac{\varepsilon}{2}\mathbb{E}[Z] + \frac{8}{\varepsilon}x$$

and (13) follows. The proof of (14) can be performed exactly in the same way, starting from (33) instead of (31) which leads to

$$\mathbb{P}\Big[Z \leq \mathbb{E}[Z] - 2\gamma\sqrt{x(2\sigma^2 + 16\mathbb{E}[Z])} - c'x\Big] \leq \exp(-x),$$

which, combined with

$$\sqrt{x(2\sigma^2 + 16\mathbb{E}[Z])} \leq \sigma\sqrt{2x} + 4\sqrt{x\mathbb{E}[Z]} \leq \sigma\sqrt{2x} + \frac{\varepsilon}{2\gamma}\mathbb{E}[Z] + \frac{8\gamma}{\varepsilon}x,$$

yields (14) since $\gamma^2 < 1.35$. $\square$

As mentioned in the Introduction, it would be desirable to know whether (13) holds with a better value for the constant κ than $\kappa = 4$. Unfortunately, we could not answer this question. In particular, we do not know whether the natural candidate $\kappa = 1$ works or not.

Acknowledgments. I gratefully thank Lucien Birgé and Michel Ledoux for encouraging me to write this paper and Gwenaelle Castellan for pointing out a mistake in a preliminary version of this work. The deviation inequalities on the left tail presented in this paper in Section 1.2 were absent from the initial version submitted to the *Annals of Probability*. I am indebted to Paul-Marie Samson for the stimulating conversations which led to the new inequalities included in the revised version. I am grateful to an anonymous referee whose suggestions also helped me to improve on the initial version of this paper.

REFERENCES

[1] BARAUD, Y. (1997). Model selection for regression on a fixed design. *Probab. Theory Related Fields*. To appear.

[2] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[3] BIRGÉ, L. and MASSART, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien LeCam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds.) 55–87. Springer, New York.

[4] BIRGÉ, L. and MASSART, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4** 329–375.

[5] BOBKOV, S. (1995). On Gross' and Talagrand's inequalities on the discrete cube. *Vestnik Syktyvkar Univ. Ser.* **1 1** 12–19. (in Russian.)

[6] BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216.

[7] BOUCHERON, S., LUGOSI, G. and MASSART, P. (1999). A sharp concentration inequality with applications. *Random Structures Algorithms*. To appear.

[8] CIREL'SON, B. S., IBRAGIMOV, I. A. and SUDAKOV, V. N. (1976). Norm of Gaussian sample function. *Proceedings of the Third Japan–U.S.S.R. Symposium on Probability Theory. Lecture Notes in Math.* **550** 20–41. Springer, Berlin.

[9] CIREL'SON, B. S. and SUDAKOV, V. N. (1978). Extremal properties of half spaces for spherically invariant measures. *J. Soviet. Math.* **9** 9–18. [Translated from *Zap. Nauch. Sem. L.O.M.I.* **41** 14–24 (1974).]

[10] DAVIES, E. B. and SIMON, B. (1984). Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.* **59** 335–395.

[11] DEMBO, A. (1997). Information inequalities and concentration of measure. *Ann. Probab.* **25** 927–939.

[12] GROSS, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083.

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

[14] HOLLEY, R. and STROOCK, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J. Statist. Phys.* **46** 1159–1194.

[15] LEDOUX, M. (1996). Isoperimetry and Gaussian analysis. *Lectures on Probability Theory and Statistics. Ecole d'Eté de Probabilitiés de Saint Flour XXIV* 165–294. Springer, Berlin.

[16] LEDOUX, M. (1996). Talagrand deviation inequalities for product measures. *ESAIM: Probab. Statist.* **1** 63–87. Available at http://www.emath.fr/ps/.

[17] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces (Isoperimetry and Processes).* Springer, Berlin.

[18] MARTON, K. (1996). Bounding $\overline{d}$-distance by information divergence: a method to prove measure concentration. *Ann. Probab.* **24** 927–939.

[19] SAMSON, P. M. (1997). Inégalités de concentration de la mesure pour des chaines de Markov et des processus Φ-mélangeants. Thése de l'Université Paul Sabatier de Toulouse.

[20] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.* **81** 73–205.

[21] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563.

CNRS URA D 0743
UNIVERSITÉ PARIS-SUD
MODÉLISATION STOCHASTIQUE ET STATISTIQUE
BÂT. 425
91405 ORSAY CEDEX
FRANCE