Concentration Inequalities for Functions of Independent Variables

Andreas Maurer
Adalbertstr. 55
D-80799 München
andreasmaurer@compuserve.com

June 23, 2005

Abstract

Following the entropy method this paper presents general concentration inequalities, which can be applied to combinatorial optimization and empirical processes. The inequalities give improved concentration results for optimal travelling salesmen tours, Steiner trees and the eigenvalues of random symmetric matrices.

1 Introduction

Since its appearance in 1995 Talagrand's convex distance inequality [17] has been very successful as a tool to prove concentration results for functions of independent variables in cases which were previously inaccessible, or could be handled only with great difficulties. The now classical applications (see McDiarmid [12] and Steele [15]) include concentration inequalities for configuration functions, such as the length of the longest increasing subsequence in a sample, or for geometrical constructions, such as the length of an optimal travelling salesman tour or an optimal Steiner tree.

Another recently emerged technique to prove concentration results is the *entropy method*. Originating in the work of Leonard Gross on logarithmic Sobolev inequalities for Gaussian measures [6], the method has been developed and refined by Ledoux, Bobkov, Massart, Boucheron, Lugosi, Rio, Bousquet and others (see [8], [10], [11], [2], [3], etc) to become an important tool in the study of empirical processes and learning theory. In [2, Boucheron at al] a general theorem on configuration functions is presented, which improves on the results obtained from the convex distance inequality. In [3, Boucheron et al] more results of this type are given and a weak version of the convex distance inequality itself is derived.

Technically the core of the entropy method is a tensorisation inequality bounding the entropy of an n variable function Z in terms of a sum of entropies with respect to the individual variables. Together with an elementary bound on the logarithm this leads to a differential inequality (Proposition 4 below) for a quantity related to the logarithm of the Laplace transform of Z. Integration of this inequality (the so-called Herbst argument) leads to concentration inequalities. We use this technique in a straightforward manner only slightly different from [3] to obtain stronger results under more restrictive conditions.

The principal contribution of this paper are two theorems giving general concentration inequalities for functions on product spaces. When applied to the geometric problems mentioned above, the upwards deviation bound of Theorem 1 improves on the results obtained from the convex distance inequality. Theorem 2 refers to self-bounding functions and can be applied to configuration functions and empirical processes.

Concentration inequalities for functions on product spaces typically bound the deviation probability for a random variable Z in terms of the deviation t and some other random variable $\Psi(Z)$ which somehow delimits the sensitivity Z has to changes in its individual arguments. In most cases the expectation of $\Psi(Z)$ can be used to bound the variance of Z by way of the Efron-Stein inequality [14]. The sensitivity function $\Psi(Z) = \Delta_{+,Z}$ which we use in this paper is more restrictive than the $\Psi(Z) = V_{+,Z}$ used in [3]. Consequently everything which can be done with our Δ_+ can also be done with V_+, and there can be cases which can be handled with V_+ where our method fails. Where our method does work however, it often gives better bounds. This is the case for several classical applications of the convex distance inequality, as shown in section 3, where Δ_+ provides a simple and natural measure of sensitivity

The next section states our two principal theorems and explains how the sensitivity function Δ_+ enters in the entropy method. Section 3 gives applications and an appendix contains detailed proofs of the main theorems and auxilliary results which would have interrupted the main line of development in the other parts of the paper.

2 Main Results

Throughout this paper we let $\{(\Omega_i, \Sigma_i, \mu_i)\}_{i=1}^n$ be a collection of probability spaces and (Ω, Σ, μ) their product with expectation $E[f] = \int f d\mu$. For measurable $A \subseteq \Omega$ we write $\Pr(A) = \mu(A)$ and for a measurable function f on Ω we use $\|f\|_\infty$ to denote the supremum of $|f|$ (our results can be easily extended to include exception sets of measure zero, which would slightly complicate the presentation). For $\mathbf{x} = (x_1, ..., x_n) \in \Omega$, $k \in \{1, ..., n\}$ and $y \in \Omega_k$ define the modification $\mathbf{x}_{y,k}$ by

$$\mathbf{x}_{y,k} = (x_1, ..., x_{k-1}, y, x_{k+1}, ..., x_n).$$

We use $\mathcal{A}_k$ to denote the algebra of bounded measurable functions independent of x_k, that is

$$\mathcal{A}_k = \left\{ f : \|f\|_\infty < \infty \text{ and } \forall \mathbf{x} \in \Omega, y \in \Omega_k, f(\mathbf{x}) = f(\mathbf{x}_{y,k}) \right\}.$$

In all of the following Z will be some random variable on Ω, which we assume to be bounded for simplicity. For $1 \leq k \leq n$ we define a random variable $\inf_k Z$ on Ω by

$$\left(\inf_k Z \right)(\mathbf{x}) = \inf_{y \in \Omega_k} Z(\mathbf{x}_{y,k}).$$

Then $\inf_k Z \in \mathcal{A}_k$ and $\inf_k Z$ is the largest member f of $\mathcal{A}_k$ satisfying $f \leq Z$ (Lemma 14). We denote

$$\Delta_{+,Z} = \sum_{k=1}^{n} \left(Z - \inf_k Z \right)^2$$

and usually drop the second subscript indicating the dependence on Z, unless there is some possible ambiguity. Using the Efron-Stein inequality [14] we obtain at once $Var(Z) \leq E[\Delta_{+,Z}]$, so that $\Delta_{+,Z}$ can always be used for a crude estimate of the variance of Z. Note that $\Delta_{+,Z}$ is distribution independent.

We give exponential concentration inequalities for Z in terms of the random variable Δ_+.

Theorem 1 *For $t > 0$ we have*

$$\Pr\left\{ Z - E[Z] \geq t \right\} \leq \exp\left(\frac{-t^2}{2\|\Delta_+\|_\infty} \right) \tag{1}$$

and, if $Z - \inf_k Z \leq 1$ for all k, then

$$\Pr\left\{ E[Z] - Z \geq t \right\} \leq \exp\left(\frac{-t^2}{2\left(\|\Delta_+\|_\infty + t/3\right)} \right). \tag{2}$$

The simplifying assumption of boundedness does not restrict the use of this result, since a nontrivial right side forces $\|\Delta_+\|_\infty$ to be finite, which in turn implies $\|Z\|_\infty < \infty$.

An inequality related to (1) appears in Corollary 3 of [3]:

$$\Pr\left\{ Z - E[Z] \geq t \right\} \leq \exp\left(\frac{-t^2}{4\|V_+\|_\infty} \right), \tag{3}$$

where

$$V_+(\mathbf{x}) = \sum_k \int_{\Omega_k \cap \{y : Z(\mathbf{x}) > Z(\mathbf{x}_{y,k})\}} \left(Z(\mathbf{x}) - Z(\mathbf{x}_{y,k}) \right)^2 d\mu_k(y). \tag{4}$$

Evidently $V_+ \leq \Delta_+$, so (3) implies (1), but with an exponent worse by a factor of 2. In every application of (3), which relies on uniform bounds of the integrands in (4), it is better to use Theorem 1, which is also applicable, because Δ_+ is the smallest such bound, is simpler than V_+, and gives a better exponent.

In [3] inequality (3) is used to derive a weaker version of Talagrand's convex distance inequality. The inequalities of Theorem 1 can be used to improve on this result, without however fully reproducing Talagrand's result. They can also be applied to sharpen the upwards deviation bounds in a theorem in [12] which can be used to derive concentration inequalities for optimal travelling salesman tours, Steiner trees and the eigenvalues of random symmetric matrices. See section 3 for these applications.

The random variable Δ_+ can also play a role in concentration inequalities for self-bounding functions. We have

Theorem 2 *Suppose that*

$$\Delta_+ \leq Z. \tag{5}$$

Then for $t > 0$

$$\Pr\{Z - E[Z] \geq t\} \leq \exp\left(\frac{-t^2}{2E[Z] + t}\right), \tag{6}$$

and, if $Z - \inf_k Z \leq 1 \, \forall k$*, then*

$$\Pr\{E[Z] - Z \geq t\} \leq \exp\left(\frac{-t^2}{2E[Z]}\right).$$

One candidate for comparison is the result proved in [2], which has almost the same conclusion (slightly better with $(2/3)\,t$ in the denominator in (6) instead of t), but requires the much stronger self-boundedness conditions

$$0 \leq Z - Z_k \leq 1 \tag{7}$$

and

$$\sum_k (Z - Z_k) \leq Z, \tag{8}$$

where $Z_k \in \mathcal{A}_k$. Under these conditions we have

$$\Delta_+ = \sum_{k=1}^{n} \left(Z - \inf_k Z\right)^2 \leq \sum_{k=1}^{n} (Z - Z_k)^2 \leq \sum_k (Z - Z_k) \leq Z,$$

so our result is also applicable.

The other similar result is Theorem 5 in [3] where self-boundedness is required only in the form

$$V_+ \leq Z,$$

and it implies our result, but with an exponent worse by a factor of 2.

The above results are asymmetrical due to the nature of Δ_+. We define a random variable $\Delta_{-,Z}$ by

$$\Delta_{-,Z} = \sum_{k=1}^{n} \left(\sup_k Z - Z \right)^2 ,$$

where $\sup_k Z = \sup_{y \in \Omega_k} Z(\mathbf{x}_{y,k})$. Then clearly $\Delta_{+,-Z} = \Delta_{-,Z}$, so we can use our results to derive bounds for $-Z$ in terms of $\Delta_{+,-Z} = \Delta_{-,Z}$.

In the remainder of this section we show how the sensitivity function $\Delta_{+,Z}$ enters the mechanism of the entropy method.

Definition 3 *For any real β and a r.v. f on Ω define the Gibbs expectation $E_{\beta Z}[f]$ by*

$$E_{\beta Z}[f] = \frac{E\left[f e^{\beta Z}\right]}{E\left[e^{\beta Z}\right]}.$$

We also denote for $\beta \neq 0$

$$H(\beta) = \frac{1}{\beta} \ln E\left[e^{\beta Z}\right].$$

If $-Z$ is viewed as the energy function of a physical system, then $E_{\beta Z}$ is the thermal expectation and $H(\beta)$ is the Helmholtz free energy at inverse temperature β. The formulas

$$\frac{d}{d\beta} \ln E\left[e^{\beta Z}\right] = E_{\beta Z}[Z] \tag{9}$$

$$\frac{d}{d\beta} H(\beta) = \frac{1}{\beta} E_{\beta Z}[Z] - \frac{1}{\beta^2} \ln E\left[e^{\beta Z}\right] \tag{10}$$

$$\lim_{\beta \to 0} H(\beta) = E[Z]$$

follow from straightforward computation and l'Hospital's rule. We also define two real functions ϕ and g by $\phi(t) = e^{-t} + t - 1$ and

$$g(t) = \sum_{k=0}^{\infty} \frac{(-t)^k}{(k+2)!}.$$

Then

$$\phi(t) = t^2 g(t) \text{ and } g(t) = \begin{cases} \phi(t)/t^2 & \text{for } t \neq 0 \\ 1/2 & \text{for } t = 0 \end{cases}. \tag{11}$$

Furthermore the function g is positive, nonincreasing, and for $t \leq 0$ and $a > 0$ we have

$$\frac{a g(t)}{1 - a t g(t)} \leq \frac{\max\{1, a\}}{2}. \tag{12}$$

A principal tool of the entropy method is the following fundamental inequality (see [11] and [2]).

Proposition 4 *Let $Z_k \in \mathcal{A}_k$ for every $k \in \{1, ..., n\}$. Then for all β*

$$\beta E\left[Ze^{\beta Z}\right] - E\left[e^{\beta Z}\right] \ln E\left[e^{\beta Z}\right] \leq E\left[\sum_{k=1}^{n} e^{\beta Z} \phi\left(\beta\left(Z - Z_k\right)\right)\right]. \qquad (13)$$

For $\beta \neq 0$, dividing this inequality by $\beta^2 E\left[e^{\beta Z}\right]$ and using (10) we get

$$H'\left(\beta\right) \leq \frac{1}{\beta^2} E_{\beta Z}\left[\sum_{k=1}^{n} \phi\left(\beta\left(Z - Z_k\right)\right)\right]. \qquad (14)$$

Using the properties of ϕ, g and Δ_+ one obtains the following simple proposition, which is the key to our results:

Proposition 5 *For $\beta > 0$ we have*

$$H'\left(\beta\right) \leq \frac{1}{2} E_{\beta Z}\left[\Delta_+\right]. \qquad (15)$$

For $\beta < 0$, if $Z - \inf_k Z \leq 1$ for all k, then

$$H'\left(\beta\right) \leq g\left(\beta\right) E_{\beta Z}\left[\Delta_+\right]. \qquad (16)$$

Proof. Since $\inf_k Z \in \mathcal{A}_k$ we have by inequality (14) and (11) for $\beta \neq 0$

$$\begin{aligned} H'\left(\beta\right) &\leq \frac{1}{\beta^2} E_{\beta Z}\left[\sum_{k=1}^{n} \phi\left(\beta\left(Z - \inf_k Z\right)\right)\right] \qquad (17) \\ &= E_{\beta Z}\left[\sum_{k=1}^{n} g\left(\beta\left(Z - \inf_k Z\right)\right)\left(Z - \inf_k Z\right)^2\right]. \end{aligned}$$

Since g is nonincreasing we have for positive β

$$g\left(\beta\left(Z - \inf_k Z\right)\right) \leq g\left(0\right) = \frac{1}{2},$$

so (17) implies (15). For $\beta < 0$, if $Z - \inf_k Z \leq 1$ we have $\beta \leq \beta\left(Z - \inf_k Z\right)$, whence

$$g\left(\beta\left(Z - \inf_k Z\right)\right) \leq g\left(\beta\right),$$

so again (17) gives (16). ■

The proofs of Theorem 1 and Theorem 2 are now straightforward and follow the path taken in [9] and [3]. The next step is to integrate inequalities (15) and (16) to obtain a bound on $H\left(\beta\right) - H\left(0\right)$ in terms of an integral of the thermal expectation of Δ_+. Uniform bounds on the integrand lead to Theorem 1, substitution of the self-bounding condition (5) gives Theorem 2. For the readers convenience detailed proofs are given in the appendix.

3 Applications

We first derive a version of the convex distance inequality from Theorem 1, then we give a 'packaged' result leading to improved bounds for optimal travelling salesmen tours, Steiner trees and the eigenvalues of random symmetric matrices.

3.1 The Convex Distance Inequality

Talagrand's convex distance inequality (see [16], [17], [12], [15] or [3]) asserts that (questions of measurability aside), for $A \subseteq \Omega$ and $t > 0$ we have

$$\Pr(A) \Pr\{d_T(\mathbf{x}, A) \geq t\} \leq \exp\left(-\frac{t^2}{4}\right). \tag{18}$$

The 'convex distance' $d_T(\mathbf{x}, A)$ is defined as follows:

Definition 6 *For* $\mathbf{x}, \mathbf{y} \in \Omega$ *and* $\alpha \in \mathbb{R}_+^n$ *write*

$$d_\alpha(\mathbf{x}, \mathbf{y}) = \sum_{i: x_i \neq y_i} \alpha_i.$$

For $A \subseteq \Omega$ *and* $\mathbf{x}, \mathbf{y} \in \Omega$ *define*

$$d_T(\mathbf{x}, A) = \sup_{\alpha: \|\alpha\|=1, \alpha_i \geq 0} \inf_{\mathbf{y} \in A} d_\alpha(\mathbf{x}, \mathbf{y}).$$

In [3] it is shown that (18) can be partially derived from the entropy method: If $\Pr A \geq 1/2$ and $t > \sqrt{2}$ then

$$\Pr\{d_T(\mathbf{x}, A) \geq t\} \leq 2 \exp\left(-\frac{t^2}{8}\right).$$

Using a similar argument together with Theorem 1 we can somewhat improve this result, still without fully recovering Talagrand's inequality:

If $\Pr A \geq \exp(-3/4)$ (this is about $0.472 < 1/2$), then

$$\Pr(A) \Pr\{d_T(\mathbf{x}, A) \geq t\} \leq \exp\left(-\frac{t^2}{5}\right). \tag{19}$$

The strategy is to first show that for $Z = d_T(., A)$ we have $\|\Delta_+\|_\infty \leq 1$, which is the hard part and implicitely contained in [3], where Sion's minimax theorem is used. A self-contained proof is given in the appendix.

Lemma 7 *If* $Z(\mathbf{x}) = d_T(\mathbf{x}, A)$ *then* $\|\Delta_+\|_\infty \leq 1$.

Given the lemma, for $\eta \geq 0$ define

$$h(\eta) = \frac{2\eta}{9} + \frac{2}{3}\sqrt{2\eta + \frac{1}{9}\eta^2}.$$

By Lemma 15 we have for any real s and $0 \leq \eta \leq 3/4$

$$\left(\frac{3}{2}h(\eta)\right)^2 \leq 3\eta \tag{20}$$

and

$$\frac{s^2}{2} \geq \frac{\left(s + \frac{3}{2}h(\eta)\right)^2}{5} - \eta. \tag{21}$$

Now write $Z = d_T(., A)$, $\rho = E[Z]$ and $\eta = \ln(1/\Pr A) \leq 3/4$. By Lemma 7 and Theorem 1 we have

$$\Pr A = \Pr\{\rho - Z \geq \rho\} \leq \exp\left(\frac{-\rho^2}{2 + 2\rho/3}\right)$$

or $\rho \leq \eta(2 + 2\rho/3)$. Elementary algebra gives $\rho \leq (3/2)h(\eta)$.

Now let $t > 0$. If $t \leq (3/2)h(\eta)$ then

$$\frac{t^2}{5} \leq \frac{1}{5}\left(\frac{3}{2}h(\eta)\right)^2 \leq \frac{3}{5}\eta \leq \eta,$$

so (19) follows from

$$\exp\left(-\frac{t^2}{5}\right) \geq \exp(-\eta) = \Pr A.$$

If on the other hand $t > (3/2)h(\eta)$ then, writing $s = t - (3/2)h(\eta)$ and invoking the upper tail bound in Theorem 1 and (21), we have

$$\begin{aligned}
\Pr\{Z \geq t\} &= \Pr\{Z \geq (3/2)h(\eta) + s\} \\
&\leq \Pr\{Z \geq \rho + s\} \leq e^{-\frac{s^2}{2}} \\
&\leq \exp\left(-\frac{(s + (3/2)h(\eta))^2}{5} + \eta\right) \\
&= \frac{1}{\Pr A} e^{\frac{-t^2}{5}},
\end{aligned}$$

which concludes the proof of (19).

3.2 A "Packaged" Concentration Result

The following result sharpens the upwards deviation bound of a general theorem in [12] (Theorem 4.5), therein derived from the convex distance inequality and used to give concentration inequalities for geometrical constructions such as minimal length travelling salesman tours and Steiner trees. Our result is a weak corollary of Theorem 1 and allows us to make a comparison with consequences of the convex distance inequality.

Recall definition 6 above and observe that for $\mathbf{x},\mathbf{y}\in\Omega$, $\alpha\in\mathbb{R}_+^n$, $k\in\{1,...,n\}$ and $y\in\Omega_k$ we have $d_\alpha\left(\mathbf{x},\mathbf{x}_{y,k}\right)\leq\alpha_k$.

Theorem 8 *Suppose that there is a constant $c>0$ such that for each $\mathbf{x}\in\Omega$ there is vector $\alpha\in\mathbb{R}_+^n$ with $\|\alpha\|\leq 1$ such that*

$$Z\left(\mathbf{x}\right)\leq Z\left(\mathbf{y}\right)+cd_\alpha\left(\mathbf{x},\mathbf{y}\right)\text{ for each }\mathbf{y}\in\Omega. \tag{22}$$

Then

$$\Pr\left\{Z-E\left[Z\right]\geq t\right\}\leq\exp\left(\frac{-t^2}{2c^2}\right) \tag{23}$$

and

$$\Pr\left\{E\left[Z\right]-Z\geq t\right\}\leq\exp\left(\frac{-t^2}{2\left(c^2+ct/3\right)}\right). \tag{24}$$

Proof. Fix any $\mathbf{x}\in\Omega$ and chose α such that (22) holds. Thus for any k and $y\in\Omega_k$ we have

$$Z\left(\mathbf{x}\right)\leq Z\left(\mathbf{x}_{y,k}\right)+cd_\alpha\left(\mathbf{x},\mathbf{x}_{y,k}\right)\leq Z\left(\mathbf{x}_{y,k}\right)+c\alpha_k.$$

Taking the infimum over $y\in\Omega_k$ we see that

$$Z\left(\mathbf{x}\right)-\inf_k Z\left(\mathbf{x}\right)\leq c\alpha_k.$$

Thus, since $\|\alpha\|\leq 1$,

$$\Delta_+\left(\mathbf{x}\right)=\sum_k\left(Z\left(\mathbf{x}\right)-\inf_k Z\left(\mathbf{x}\right)\right)^2\leq c^2\sum_k\alpha_k^2\leq c^2$$

and the upper tail bound follows directly from Theorem 1. To ensure the required bound $Z-\inf_k Z\leq 1$ we apply the lower tail bound of Theorem 1 to Z/c with deviation t/c to arrive at the second conclusion. ■

Under the same conditions the convex distance inequality (as in [12]) gives the conclusions

$$\Pr\left\{\pm\left(Z-M\right)\geq t\right\}\leq 2\exp\left(\frac{-t^2}{4c^2}\right), \tag{25}$$

where M refers to a median of Z.

Our upper tail bound (23) certainly looks better than (25), while the advantage of our lower tail bound (24) is questionable. For a more detailed comparison we set $c = 1$ and use Lemma 4.6 in [12], which asserts that under these conditions $|E[Z] - M| \leq 2\sqrt{\pi}$. We assume this to be our only available estimate on the difference of median and expectation, so we replace t with $\max\{0, t - 2\sqrt{\pi}\}$ in our bounds, if we discuss deviations from a median, and make the same substitution in (25) when we consider deviations from the expectation.

If one is interested in the probability of deviations from the expectation (perhaps the more frequent case) then our upper tail bound is superior for all deviations and decays more rapidly than (25). The lower tail bound in (25) has the value 2 for $t \leq 2\sqrt{\pi}$ and becomes nontrivial for $t > 2\sqrt{\pi} + 2\sqrt{\ln 2}$. For all deviations larger than $t \approx 10.9$ it will be smaller and decay more rapidly than our lower tail bound (24). At deviation $t \approx 10.9$ both results bound the deviation probability with $\approx 2.7 \times 10^{-6}$, for smaller deviations (and larger deviation probabilities) our bound is better.

By the Efron-Stein-inequality [14] and the above proof we have $Var(Z) \leq E[\Delta_+] \leq \|\Delta_+\|_\infty \leq 1$, so the Chebychev inequality gives a deviation bound of $1/t^2$, which improves on our lower tail bound between the deviation probabilities ≈ 0.12 and ≈ 0.52.

If one is interested in the probability of deviations from the median, our lower tail bound (24) is inferior to (25) for all deviations. Our upper tail bound (23) remains trivial for $t \leq 2\sqrt{\pi}$, while (25) already becomes nontrivial at $2\sqrt{\ln 2}$. For all deviations larger than $t \approx 11.8$ our upper tail bound will be smaller and decay more rapidly than (25). At $t \approx 11.8$ both bound the deviation probability with $\approx 1.5 \times 10^{-15}$, for smaller deviations (and larger deviation probabilities) the bound (25) is better.

The above considerations remain valid for general values of c if we replace t by t/c.

3.3 Applications to Geometry

Let Ω_k be the unit square $[0,1]^2$ and for $\mathbf{x} = (x_1, ..., x_n) \in \Omega = \left([0,1]^2\right)^n$ let $tsp(\mathbf{x})$ be the minimum length of a 'travelling salesman tour' passing through all the points $x_i = (x_i^1, x_i^2)$ in a closed sequence. We use $\pi(\mathbf{x})$ to denote an optimal sequence corresponding to the set of positions $\mathbf{x} \in \mathbf{\Omega}$ so that

$$tsp(\mathbf{x}) = \sum \left\| \pi(\mathbf{x})_i - \pi(\mathbf{x})_{i-1} \right\|,$$

where it is understood that the first and last elements in the summation are identified.

Using space-filling curves it can be shown (see [15]), that there is a constant c_2 such that for all n and $\mathbf{x} \in \Omega$ there is a tour $\pi^c(\mathbf{x})$ with

$$\sum \left\| \pi^c(\mathbf{x})_i - \pi^c(\mathbf{x})_{i-1} \right\|^2 \leq c_2. \tag{26}$$

It is shown in [15] and [12], that tsp satisfies the conditions of Theorem 8, with c replaced by $2\sqrt{c_2}$, so that

$$\Pr\{tsp - E[tsp] \geq t\} \leq \exp\left(\frac{-t^2}{8c_2}\right) \tag{27}$$

and

$$\Pr\{E[tsp] - tsp \geq t\} \leq \exp\left(\frac{-t^2}{8c_2 + 4\sqrt{c_2}t/3}\right). \tag{28}$$

If we are interested in the deviation from the expectation then, as explained above, inequality (27) is an improvement over the result obtained from the convex distance inequality (in [12]), and (28) is an improvement for deviation probabilities $\gtrapprox 2.7 \times 10^{-6}$.

Given (26) the inequalities (27) and (28) can be obtained from Theorem 1 directly, which leads to a considerable simplification of the geometrical arguments involved, because to apply Theorem 1 we need to verify the condition (22) only for those $\mathbf{y} \in \Omega$ which differ from $\mathbf{x}$ in *only one coordinate*:

Proposition 9

$$\|\Delta_{+,tsp}\|_\infty \leq 4c_2.$$

Proof. Fix $\mathbf{x} \in \Omega$ and write $\theta_k = \left\|\pi^c(\mathbf{x})_k - \pi^c(\mathbf{x})_{k-1}\right\|$. To bound $\Delta_+(\mathbf{x})$ we consider an arbitrary index k and an arbitrary position $y \in [0,1]^2$ to replace x_k. The conclusion is trivial if $n = 1$ so we assume $n > 1$. Now the tour $\pi(\mathbf{x}_{y,k})$ visits all the points of $\mathbf{x}$ except for x_k, so somewhere along $\pi(\mathbf{x}_{y,k})$ is the position x^*, which on $\pi^c(\mathbf{x})$ precedes x_k. Let's start there, go to x_k (distance θ_k) and back to x^* (again θ_k) and then continue along $\pi(\mathbf{x}_{y,k})$ until we arrive at our starting point x^*. We have then visited all the x_i and the entire trip had length $tsp(\mathbf{x}_{y,k}) + 2\theta_k$. By the triangle inequality $tsp(\mathbf{x}) \leq tsp(\mathbf{x}_{y,k}) + 2\theta_k$ or $tsp(\mathbf{x}) - \inf_y tsp(\mathbf{x}_{y,k}) \leq 2\theta_k$. The result follows from summing the squares and using (26). ■

Obvious modifications of this argument lead to concentration inequalities for optimal Steiner trees.

3.4 Eigenvalues of random symmetric matrices

As an application of Theorem 8 we consider the concentration of eigenvalues of random symmetric matrices as analysed in [1]. Here we take $\Omega_k = [-1,1]$ and $n = \binom{m+1}{2}$. A randomly chosen vector $\mathbf{x} = (x_{ij})_{1 \leq i \leq j \leq m} \in \Omega$ defines a symmetric $m \times m$-matrix $M(\mathbf{x})$ by setting $M(\mathbf{x})_{ij} = x_{ij}$ for $i \leq j$ and $M(\mathbf{x})_{ij} = x_{ji}$ for $j < i$. We order the eigenvalues of $M(\mathbf{x})$ by decreasing values $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$, and let $Z(\mathbf{x})$ be the d-th eigenvalue of $M(\mathbf{x})$ in this order.

Theorem 10 *For $t > 0$*

$$\Pr\{Z - E[Z] \geq t\} \leq \exp\left(\frac{-t^2}{16d^2}\right)$$

and

$$\Pr\{E[Z]-Z\geq t\}\leq \exp\left(\frac{-t^2}{16d^2+2dt}\right).$$

Again the upper tail bound compares favourably with the result in [1] derived from the convex distance inequality

$$\Pr\{Z-M\geq t\}\leq 2\exp\left(\frac{-t^2}{32d^2}\right),$$

where M refers to the median.

To prove our result we plagiarise the argument in [1], which is modified only slightly to make Theorem 8 applicable.

Proof. Fix $\mathbf{x}=(x_{ij})_{1\leq i\leq j\leq m}\in\Omega$ and let $v^1,...,v^d$ be normalized eigenvectors corresponding to the d largest eigenvalues of the matrix $M(\mathbf{x})$. Use V to denote the span of the v^k for $k\in\{1,...,d\}$. We define $\theta\in\mathbb{R}^n$ by

$$\begin{aligned}\theta_{ii} &= \sum_{p=1}^d \left(v_i^p\right)^2 \text{ for } 1\leq i\leq m\\ \theta_{ij} &= 2\sqrt{\sum_{p=1}^d \left(v_i^p\right)^2}\sqrt{\sum_{p=1}^d \left(v_j^p\right)^2} \text{ for } 1\leq i<j\leq m.\end{aligned}$$

Then

$$\begin{aligned}\sum_{1\leq i\leq j\leq m}\theta_{ij}^2 &= \sum_{i=1}^n\left(\sum_{p=1}^d\left(v_i^p\right)^2\right)^2+4\sum_{1\leq i<j\leq m}\left(\sum_{p=1}^d\left(v_i^p\right)^2\right)\left(\sum_{p=1}^d\left(v_j^p\right)^2\right)\\ &\leq 2\left(\sum_{i=1}^n\sum_{p=1}^d\left(v_i^p\right)^2\right)^2=2\left(\sum_{p=1}^d\sum_{i=1}^n\left(v_i^p\right)^2\right)^2=2d^2,\end{aligned}$$

so, if we define $\alpha=\theta/\left(\sqrt{2}d\right)$, we have $\|\alpha\|\leq 1$.

Now take $\mathbf{y}\in\Omega$ and use W to denote the span of the eigenvectors corresponding to the $d-1$ largest eigenvalues of $M(\mathbf{y})$. Since $\dim W^\perp+\dim V=m+1$ we must have $W^\perp\cap V\neq\{0\}$. Let $u=\sum_{p=1}^d c_p v^p$ be a unit vector in $W^\perp\cap V$. Since $u\in W^\perp$ we must have $\langle M(\mathbf{y})u,u\rangle\leq Z(\mathbf{y})$. Also

$$Z(\mathbf{x})=\sum_{p=1}^d c_p^2\lambda_d(\mathbf{x})\leq\sum_{p=1}^d c_p^2\lambda_p(\mathbf{x})=\langle M(\mathbf{x})u,u\rangle.$$

If X denotes the set of pairs $1\leq i,j\leq m$ with $x_{ij}\neq y_{ij}$, then

$$\begin{aligned}Z(\mathbf{x})-Z(\mathbf{y}) &\leq \langle(M(\mathbf{x})-M(\mathbf{y}))u,u\rangle=\sum_{(i,j)\in X}(x_{ij}-y_{ij})\sum_{p=1}^d c_pv_i^p\sum_{p=1}^d c_pv_j^p\\ &\leq 2\sum_{(i,j)\in X}\left|\sum_{p=1}^d c_pv_i^p\right|\left|\sum_{p=1}^d c_pv_j^p\right|,\end{aligned}$$

where we have used the boundedness condition $|x_{ij} - y_{ij}| \leq 2$. The Cauchy-Schwartz inequality then gives

$$\begin{aligned}
&Z(\mathbf{x}) - Z(\mathbf{y}) \\
&\leq 2 \sum_{(i,j)\in X} \left(\sum_{p=1}^{d} c_p^2 \right)^{1/2} \left(\sum_{p=1}^{d} (v_i^p)^2 \right)^{1/2} \left(\sum_{p=1}^{d} c_p^2 \right)^{1/2} \left(\sum_{p=1}^{d} (v_j^p)^2 \right)^{1/2} \\
&= 2 \sum_{(i,j)\in X} \left(\sum_{p=1}^{d} (v_i^p)^2 \right)^{1/2} \left(\sum_{p=1}^{d} (v_j^p)^2 \right)^{1/2} \\
&= 2 \sum_{\substack{1\leq i\leq j\leq n \\ x_{ij}\neq y_{ij}}} \theta_{ij} = 2^{3/2} d \sum_{\substack{1\leq i\leq j\leq n \\ x_{ij}\neq y_{ij}}} \alpha_{ij},
\end{aligned}$$

and Theorem 8 gives the desired conclusions. ■

Appendix

We give proofs for Theorem 1 and Theorem 2 and other results used in the paper.

Lemma 11 *For $\beta > 0$*

$$\ln E\left[e^{\beta(Z-E[Z])}\right] \leq \frac{\beta}{2} \int_0^\beta E_{\gamma Z}[\Delta_+]\, d\gamma. \tag{29}$$

For $\beta < 0$, if $Z - \inf_k Z \leq 1$ for all k, then

$$\ln E\left[e^{\beta(Z-E[Z])}\right] \leq -\beta g(\beta) \int_\beta^0 E_{\gamma Z}[\Delta_+]\, d\gamma. \tag{30}$$

Proof. We use proposition 5.

Consider first $\beta > 0$. Integrating (15) from 0 to β and using $\lim_{\beta\to 0} H(\beta) = E[Z]$ gives

$$\frac{1}{\beta} \ln E\left[e^{\beta Z}\right] - E[Z] \leq \frac{1}{2} \int_0^\beta E_{\gamma Z}[\Delta_+]\, d\gamma,$$

and multiplication with β leads to (29).

For the case $\beta < 0$, if $Z - \inf_k Z \leq 1$ we integrate (16) from β to 0 to obtain

$$E[Z] - \frac{1}{\beta} \ln E\left[e^{\beta Z}\right] \leq \int_\beta^0 g(\gamma) E_{\gamma Z}[\Delta_+]\, d\gamma \leq g(\beta) \int_\beta^0 E_{\gamma Z}[\Delta_+]\, d\gamma,$$

where we again used the fact that g is nonincreasing. We then multiply the last inequality with the positive number $-\beta$ to arrive at (30). ■

Proof of Theorem 1. For $\beta > 0$ we use (29) to arrive at

$$\ln E\left[e^{\beta(Z-E[Z])}\right] \leq \frac{\beta}{2}\int_0^\beta E_{\gamma Z}\left[\Delta_+\right] d\gamma \leq \frac{\beta^2}{2}\left\|\Delta_+\right\|_\infty .$$

With $\beta > 0$ we get

$$\Pr\left\{Z - E\left[Z\right] \geq t\right\} \leq E\left[e^{\beta(Z-E[Z]-t)}\right] \leq \exp\left(-t\beta + \frac{\beta^2}{2}\left\|\Delta_+\right\|_\infty\right).$$

Substitution of the minimizing value $\beta = t/\left\|\Delta_+\right\|_\infty$ gives (1).

For the other half, by (30), we have

$$\ln E\left[e^{\beta(Z-E[Z])}\right] \leq \beta^2 g\left(\beta\right)\left\|\Delta_+\right\|_\infty = \phi\left(\beta\right)\left\|\Delta_+\right\|_\infty .$$

Setting $\beta = -\ln\left(1 + t/\left\|\Delta_+\right\|_\infty\right)$ we get, for $\beta < 0$,

$$\begin{aligned}
&\Pr\left\{E\left[Z\right] - Z \geq t\right\} \\
&\leq E\left[e^{\beta(Z-E[Z]+t)}\right] \leq \exp\left(\beta t + \phi\left(\beta\right)\left\|\Delta_+\right\|_\infty\right) \\
&= \exp\left(-\left\|\Delta_+\right\|_\infty\left(\left(1+\frac{t}{\left\|\Delta_+\right\|_\infty}\right)\ln\left(1+\frac{t}{\left\|\Delta_+\right\|_\infty}\right) - \frac{t}{\left\|\Delta_+\right\|_\infty}\right)\right) \\
&\leq \exp\left(\frac{-t^2}{2\left(\left\|\Delta_+\right\|_\infty + t/3\right)}\right),
\end{aligned}$$

where we have used Lemma 2.4. in [12] in the last inequality. ■

Lemma 12 *Let C and b denote two positive real numbers, $t > 0$. Then*

$$\inf_{\beta\in[0,1/b)}\left(-\beta t + \frac{C\beta^2}{1-b\beta}\right) \leq \frac{-t^2}{2\left(2C + bt\right)}. \tag{31}$$

Proof. Let $h\left(t\right) = 1 + t - \sqrt{1+2t}$. Then use

$$\begin{aligned}
2h\left(t\right)\left(1+t\right) &= 2\left(1+t\right)^2 - 2\left(1+t\right)\sqrt{1+2t} \\
&= \left(1+t\right)^2 - 2\left(1+t\right)\sqrt{1+2t} + \left(1+2t\right) + t^2 \\
&= \left(1+t-\sqrt{1+2t}\right)^2 + t^2 \\
&\geq t^2,
\end{aligned}$$

so that

$$h\left(t\right) \geq \frac{t^2}{2\left(1+t\right)}. \tag{32}$$

Substituting

$$\beta = \frac{1}{b}\left(1 - \left(1+\frac{bt}{C}\right)^{-1/2}\right)$$

in the left side of (31) we obtain

$$\inf_{\beta\in[0,1/b)}\left(-\beta t+\frac{C\beta^2}{1-b\beta}\right)\leq -\frac{2C}{b^2}h\left(\frac{bt}{2C}\right)\leq\frac{-t^2}{2\left(2C+bt\right)},$$

where we have used (32). ■

We give the self-bounding version in a form slightly more general than Theorem 2:

Theorem 13 *Suppose that $a>0$ and that*

$$\Delta_+\leq aZ. \tag{33}$$

Then for $t>0$

$$\Pr\left\{Z-E\left[Z\right]\geq t\right\}\leq\exp\left(\frac{-t^2}{2aE\left[Z\right]+at}\right),$$

and, if $Z-\inf_k Z\leq 1\,\forall k$, then

$$\Pr\left\{E\left[Z\right]-Z\geq t\right\}\leq\exp\left(\frac{-t^2}{2\max\left\{1,a\right\}E\left[Z\right]}\right). \tag{34}$$

Proof. Using (29) and (33) we get, for $\beta>0$,

$$\begin{aligned}\ln E\left[e^{\beta(Z-E[Z])}\right] &\leq \frac{a\beta}{2}\int_0^\beta E_{\gamma Z}\left[Z\right]d\gamma\\ &= \frac{a\beta}{2}\int_0^\beta\left(\frac{d}{d\gamma}\ln E\left[e^{\gamma Z}\right]\right)d\gamma=\frac{a\beta}{2}\ln E\left[e^{\beta Z}\right]\\ &= \frac{a\beta}{2}\ln E\left[e^{\beta(Z-E[Z])}\right]+\frac{a\beta^2}{2}E\left[Z\right].\end{aligned}$$

Rearranging this inequality we get for $\beta\in(0,2/a)$

$$\ln E\left[e^{\beta(Z-E[Z])}\right]\leq\frac{\beta^2}{1-\frac{1}{2}a\beta}\left(\frac{a}{2}E\left[Z\right]\right).$$

Using Lemma 12 with $b=a/2$ and $C=(a/2)\,E\left[Z\right]$ we have

$$\begin{aligned}\Pr\left\{Z-E\left[Z\right]\geq t\right\} &\leq \inf_{\beta\in(0,2)}\ln E\left[e^{\beta(Z-E[Z]-t)}\right]\\ &\leq \inf_{\beta\in(0,2)}\exp\left(-\beta t+\frac{\beta^2}{1-\frac{1}{2}a\beta}\left(\frac{a}{2}E\left[Z\right]\right)\right)\\ &\leq \exp\left(\frac{-t^2}{2aE\left[Z\right]+at}\right).\end{aligned}$$

For the other tail-bound we let $\beta < 0$, use (30) and (33) to obtain

$$\begin{aligned}\ln E\left[e^{\beta(Z-E[Z])}\right] &\leq -a\beta g(\beta)\int_\beta^0 E_{\gamma Z}[Z]\,d\gamma \\ &= -a\beta g(\beta)\int_\beta^0 \left(\frac{d}{d\gamma}\ln E\left[e^{\gamma Z}\right]\right)d\gamma = a\beta g(\beta)\ln E\left[e^{\beta Z}\right] \\ &= a\beta g(\beta)\ln E\left[e^{\beta(Z-E[Z])}\right] + a\beta^2 g(\beta)\,E[Z]\,.\end{aligned}$$

Thus

$$(1-a\beta g(\beta))\ln E\left[e^{\beta(Z-E[Z])}\right] \leq a\beta^2 g(\beta)\,E[Z]\,,$$

and, since $1-\beta g(\beta) > 1$ and using (12)

$$\ln E\left[e^{\beta(Z-E[Z])}\right] \leq \beta^2\frac{ag(\beta)}{1-a\beta g(\beta)}E[Z] \leq \frac{\beta^2}{2}\max\{1,a\}\,E[Z]\,.$$

We therefore have

$$\Pr\{E[Z]-Z\geq t\} \leq \ln E\left[e^{\beta(Z-E[Z]+t)}\right] \leq \exp\left(\beta t + \frac{\beta^2}{2}\max\{1,a\}\,E[Z]\right).$$

The result follows from substitution of $\beta = -t/(\max\{1,a\}\,E[Z])$. ■

Lemma 14 *Write $f \leq g$ iff $f(x) \leq g(x)$ for all $x \in \Omega$ and set*

$$\Lambda_k = \{g \in \mathcal{A}_k : g \leq Z\}\,.$$

Then $\inf_k Z \in \Lambda_k$ and $g \leq \inf_k Z$ for every $g \in \Lambda_k$.

Proof. The first assertion is obvious. Also let $g \in \Lambda_k$ and pick $y \in \Omega_k$. Then $g(x) = g(x_{y,k}) \leq Z(x_{y,k})$. Taking the infimum over y gives the result. ■

Proof of Lemma 7. For any $\mathbf{x} \in \Omega$ there is a map $h_{\mathbf{x}} : \Omega \to \{0,1\}^n \subset \mathbb{R}_+^n$ given by $h_{\mathbf{x}}(\mathbf{z})_i = 1$ if $x_i \neq z_i$ and $h_{\mathbf{x}}(\mathbf{z})_i = 0$ if $x_i = z_i$. Also define the sets $U(\mathbf{x}) = h_{\mathbf{x}}(A)$, let $V(\mathbf{x})$ be the convex hull of the finite set $U(\mathbf{x})$. Note that $U(\mathbf{x}) \subset \mathbb{R}_+^n$ so that $V(\mathbf{x}) \subset \mathbb{R}_+^n$.

Now we fix $\mathbf{x} \in \Omega$ and go about to show that $\Delta_+(\mathbf{x}) \leq 1$. This is trivial if $d_T(\mathbf{x}, A) = 0$. We can therefore assume that $d_T(\mathbf{x}, A) > 0$. Let ξ_0 be the vector of minimal norm in $V(\mathbf{x})$. Since $d_T(\mathbf{x}, A) > 0$ we have $\mathbf{x} \notin A$ so $0 \notin U(\mathbf{x})$ so $0 \notin V(\mathbf{x})$ (since 0 is an extremepoint of $\mathbb{R}_+^n$) so $\|\xi_0\| > 0$. Let ω be the associated unit vector $\omega = \xi_0/\|\xi_0\|$. Use $\langle .,. \rangle$ for the inner product in $\mathbb{R}^n$. Since $V(\mathbf{x})$ is convex and $\|\xi_0\|$ is minimal, we have for any $\xi \in V(\mathbf{x})$

$$0 \leq \left(\frac{d}{d\lambda}\right)_{\lambda=0}\|(1-\lambda)\xi_0 + \lambda\xi\|^2 = 2\langle \xi_0, \xi - \xi_0\rangle$$

so that $\|\xi_0\|^2 \leq \langle \xi, \xi_0 \rangle$. In particular we have, since $h_{\mathbf{x}}(A) \subset V(\mathbf{x})$,

$$\forall \mathbf{z} \in A, \|\xi_0\| \leq \langle h_{\mathbf{x}}(\mathbf{z}), \omega \rangle . \tag{35}$$

Since $0 \notin V(\mathbf{x})$ and $U(\mathbf{x})$ is finite any linear functional attains its minimum over $V(\mathbf{x})$ at a point in $U(\mathbf{x})$. Thus for $\alpha \in \mathbb{R}_+^n$ with $\|\alpha\| = 1$, by Cauchy-Schwartz,

$$\begin{aligned} d_\alpha(\mathbf{x}, A) &= \inf_{\mathbf{z}\in A} \sum_{i: x_i \neq y_i} \alpha_i = \inf_{\mathbf{z}\in A} \langle h_{\mathbf{x}}(\mathbf{z}), \alpha \rangle = \inf_{\xi \in U(\mathbf{x})} \langle \xi, \alpha \rangle = \inf_{\xi \in V(\mathbf{x})} \langle \xi, \alpha \rangle \\ &\leq \inf_{\xi \in V(\mathbf{x})} \|\xi\| = \|\xi_0\| , \end{aligned}$$

so that also $d_T(\mathbf{x}, A) \leq \|\xi_0\|$. Using (35) we get

$$\forall \mathbf{z} \in A, d_T(\mathbf{x}, A) \leq \langle h_{\mathbf{x}}(\mathbf{z}), \omega \rangle . \tag{36}$$

Now pick any k and any $y \in \Omega_k$. Then, because minima are achieved on finite sets,

$$d_T(\mathbf{x}_{y,k}, A) = \sup_\alpha \inf_{\xi \in U(\mathbf{x}_{y,k})} \langle \xi, \alpha \rangle \geq \inf_{\xi \in U(\mathbf{x}_{y,k})} \langle \xi, \omega \rangle = \left\langle \xi', \omega \right\rangle$$

for some $\xi' \in U(\mathbf{x}_{y,k})$, say $\xi' = h_{\mathbf{x}_{y,k}}(\mathbf{z}_0)$ for some $\mathbf{z}_0 \in A$. Thus using (36) with $\mathbf{z} = \mathbf{z}_0$

$$d_T(\mathbf{x}, A) - d_T(\mathbf{x}_{y,k}, A) \leq \left\langle h_{\mathbf{x}}(\mathbf{z}_0) - h_{\mathbf{x}_{y,k}}(\mathbf{z}_0), \omega \right\rangle \leq \omega_k,$$

leading to

$$d_T(\mathbf{x}, A) - \inf_{y\in\Omega_k} d_T(\mathbf{x}_{y,k}, A) = \sup_{y\in\Omega} \left(d_T(\mathbf{x}, A) - d_T(\mathbf{x}_{y,k}, A) \right) \leq \omega_k.$$

Squaring this inequality and summing the result over k gives $\Delta_+(\mathbf{x}) \leq \sum_k \omega_k^2$. Observing that ω is a unit vector completes the proof. ■

Lemma 15 *For $\eta \geq 0$ define*

$$h(\eta) = \frac{2\eta}{9} + \frac{2}{3}\sqrt{2\eta + \frac{1}{9}\eta^2}.$$

Then for any real s and $0 \leq \eta \leq 3/4$ we have

$$\left(\frac{3}{2}h(\eta)\right)^2 \leq 3\eta \text{ and } \frac{s^2}{2} \geq \frac{\left(s + \frac{3}{2}h(\eta)\right)^2}{5} - \eta. \tag{37}$$

Proof. h is clearly increasing and $h(3/4) = 1$. Thus, for $\eta \leq 3/4$, we have $h(\eta) \leq 1$ and

$$\left(\frac{3}{2}h(\eta)\right)^2 = \left(\frac{\eta}{3} + \sqrt{2\eta + \frac{1}{9}\eta^2}\right)^2 = \eta h(\eta) + 2\eta \leq 3\eta,$$

which is the first conclusion and also gives $(3/4)\,h(\eta)^2 \leq \eta$. For real s define

$$\psi_\eta(s) = \frac{s^2}{2} - \frac{1}{5}\left(s + \frac{3}{2}h(\eta)\right)^2 + \eta.$$

Then ψ_η has a minimum at $s^* = h(\eta)$. Resubstitution of s^* therefore gives

$$\psi_\eta(s) \geq \psi_\eta(s^*) = \eta - \frac{3}{4}h(\eta)^2 \geq 0,$$

which gives the second conclusion. ■

Acknowledgements

The author is grateful to the reviewers and to Erhard Seiler for comments that improved the paper.

References

[1] A. Alon, M. Krivelevich, V. H. Vu, *On the concentration of eigenvalues of random symmetric matrices*, Israel J. Math, (2002), 131: 259-267.

[2] S. Boucheron, G. Lugosi, P. Massart, *A sharp concentration inequality with applications in random combinatorics and learning*, Random Structures and Algorithms, (2000) 16:277-292.

[3] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities using the entropy method*, Annals of Probability (2003) 31:1583-1614.

[4] O. Bousquet, *A Bennet concentration inequality and its application to suprema of empirical processes,* C. R. Acad. Sci. Paris, (2002) 334: 495-500.

[5] O. Bousquet, *Concentration inequalities for subadditive functions using the entropy method*, Stochastic Inequalities and Applications, (2003) 56: 213-247.

[6] L.Gross, *Logarithmic Sobolev Inequalities,* American Journal of Mathematics, Vol 97:4,(1975) 1061-1083

[7] C.Houdré, *A note on jackknife based estimates of sampling distributions*, available online.

[8] M.Ledoux, *On Talagrand's deviation inequalities for product measures.* ESAIM: Probability and Statistics, (1996) 1:63-87

[9] M.Ledoux, *The Concentration of Measure Phenomenon,* AMS Surveys and Monographs 89 (2001)

[10] P.Massart, *Optimal constants for Hoeffding type inequalities,* Technical report, Mathematiques, Université de Paris-Sud, Report 98.86, 1998

[11] P.Massart, *Some applications of concentration inequalities to statistics*, Annales de la Faculté des Sciences de Toulouse (2000) IX: 245-303

[12] C.McDiarmid, *Concentration*, in *Probabilistic Methods of Algorithmic Discrete Mathematics*, (1998) 195-248. Springer, Berlin

[13] E.Rio, *Inégalités de concentration pour les processus empiriques de classes de parties,* Probability Theory and Related Fields (2000) 119: 163-175

[14] J.M.Steele, An Efron-Stein inequality for nonsymmetric statistics, Annals of Statistics (1986) 14:753-758

[15] J.M.Steele, *Probability Theory and Combinatorial Optimization*, CBMS-NSF, SIAM Philadelphia, 1996.

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

[17] M.Talagrand, *A new look at independence,* Annals of Probability (1996) 24:1-34