Elect. Comm. in Probab. **13** (2008), 628–640

ELECTRONIC COMMUNICATIONS in PROBABILITY

EXPONENTIAL BOUNDS FOR MULTIVARIATE SELF-NORMALIZED SUMS

PATRICE BERTAIL
Laboratory of Statistics, CREST and MODALX, University Paris X, France
email: bertail@ensae.fr

EMMANUELLE GAUTHERAT
Laboratory of Statistics, CREST and Economic Faculty of Reims, France
email: gauthera@ensae.fr

HUGO HARARI-KERMADEC
Laboratory of Statistics, CREST, Timbre J340, 3 av. P. Larousse, 92241 Malakoff Cedex and Université Paris-Dauphine, France
email: harari@ensae.fr

Submitted April 9, 2008, *accepted in final form* November 25, 2008

AMS 2000 Subject classification: Primary 62G15 ; secondary 62E17, 62H15.
Keywords: Exponential inequalities; Self-normalization; multivariate; Hoeffding inequality.

Abstract

In a non-parametric framework, we establish some non-asymptotic bounds for self-normalized sums and quadratic forms in the multivariate case for symmetric and general random variables. This bounds are entirely explicit and essentially depends in the general case on the kurtosis of the Euclidean norm of the standardized random variables.

1 Introduction

Let $Z, Z_1, ..., Z_n$ be i.i.d. random centered vectors from a probability space $(\Omega, \mathscr{A}, \Pr)$ to $(\mathbb{R}^q, \mathscr{B}, \mathbb{P})$. We denote E the expectation under $\mathbb{P}$. In the following we put $\overline{Z}_n = n^{-1}\sum_{i=1}^n Z_i$. Define S a square root of the matrix $S^2 = E(ZZ')$ and similarly S_n a square root of $S_n^2 = n^{-1}\sum_{i=1}^n Z_i Z_i^{'}$. We assume in the following that S^2 is full rank and therefore S_n^2 is also full rank with probability 1 as soon as $n > p$. For further use, we define $\gamma_r = E(\|S^{-1}Z\|_2^r)$, $r > 0$, where $\|\ \|_2$ is the Euclidean norm on $\mathbb{R}^q$. Now consider the self-normalized sum

$$n^{1/2}S_n^{-1}\overline{Z}_n = \left(\sum_{i=1}^n Z_i Z_i^{'}\right)^{-1/2}\sum_{i=1}^n Z_i. \tag{1}$$

and its Euclidean norm

$$n\overline{Z}_n^{'}S_n^{-2}\overline{Z}_n \tag{2}$$

Self-normalized sums have recently given rise to an important literature : see for instance [13, 6] or [4] for self-normalized processes. It has been proved that non-asymptotic exponential bounds can be obtained for these quantities under very weak conditions on the underlying moments of the variables Z_i. Unfortunately, except in the symmetric case, these bounds established in the real case ($q = 1$) are not universal and depend on the skewness $\gamma_3 = E|S^{-1}Z|^3$ or even an higher moments for instance $\gamma_{10/3} = E|S^{-1}Z|^{10/3}$, see [13]. Actually, uniform bounds in $\mathbb{P}$ are impossible to obtain, otherwise this would contradict Bahadur and Savage's Theorem, see [2, 18]. Recall that the behaviour of self-normalized sums is closely linked to the behaviour of the statistics of Student, which is the basic asymptotic root for constructing confidence intervals (see Remark 2 below). Moreover, available bounds are not explicit and only valid for $n \geq n_0$, n_0 large and unknown. To our knowledge, non-asymptotic exponential bounds with **explicit constants** are only available for symmetric distribution [12, 9, 17], in the unidimensional case ($q = 1$). In this paper, we obtain generalizations of these bounds for (2) in the multivariate case by using a multivariate extension of the symmetrization method developed in [16] as well as arguments taken from the literature on self-normalized process, see [4]. Our bounds are explicit but depend on the kurtosis γ_4 of the Euclidean norm of $S^{-1}Z$ rather than on the skewness. They hold for any value of the parameter size q. One technical difficulty in the multidimensional case is to obtain an explicit exponential bound for the smallest eigenvalue of the empirical variance which allows to control the deviation of S_n^2 from S^2, a result which has its own interest.

2 Exponential bounds for self-normalized sums

Some bounds for self-normalized sums may be quite easily obtained in the symmetric case (that is for random variables having a symmetric distribution) and are well-known in the unidimensional case. In non-symmetric and/or multidimensional case theses bounds are new and not trivial to prove. One of the main tools for obtaining exponential inequalities in various setting is the famous Hoeffding inequality (see [12]) yielding that for independent real random variables (r.v.) Y_i, $i = 1, ..., n$, with finite support say $[0, 1]$, we have

$$\Pr\left(n^{-1}\left(\sum_{i=1}^{n} Y_i\right)^2 \geq t\right) \leq 2\exp\left(-\frac{t}{2}\right).$$

A direct application of this inequality to self-normalized sums (via a randomization step introducing Rademacher r.v.'s) yields (see [9, 8]) that, for n independent random variables Z_i symmetric about 0, and not necessarily bounded (nor identically distributed), we have

$$\Pr\left(\frac{\left(\sum_{i=1}^{n} Z_i\right)^2}{\sum_{i=1}^{n} Z_i^2} \geq t\right) \leq 2\exp\left(-\frac{t}{2}\right). \tag{3}$$

In the general non-symmetric case, the master result of [13] for $q = 1$ states that if $\gamma_{10/3} < \infty$, then for some $A \in \mathbb{R}$ and some $a \in]0, 1[$,

$$\Pr\left(\frac{\left(\sum_{i=1}^{n} Z_i\right)^2}{\sum_{i=1}^{n} Z_i^2} \geq t\right) \leq 2\overline{F}_1(t) + A\gamma_{10/3} n^{-1/2} e^{-at/2}, \tag{4}$$

where $\overline{F}_q$ is the survival function of a $\chi^2(q)$ distribution defined by $\overline{F}_q(t) = \int_t^{+\infty} f_q(y)dy$ with $f_q(y) = \dfrac{1}{2^{q/2}\Gamma(q/2)} y^{q/2-1} e^{-\frac{y}{2}}$ and $\Gamma(p) = \int_0^{+\infty} y^{p-1} e^{-y} dy$.
However the constants A and a are not explicit and, despite of its great interest to understand the large deviation behaviour of self-normalized sums, the bound is of no direct practical use. In the non-symmetric case our bounds are worse than (4) as far as the control of the approximation by a $\chi^2(q)$ distribution are concerned, but entirely explicit.

Theorem 1. *Let Z, $(Z_i)_{1\le i\le n}$, be an i.i.d. sample in $\mathbb{R}^q$ with probability $\mathbb{P}$. Suppose that S^2 is of rank q. Then the following inequalities hold,* ***for finite*** *$n > q$ and for $t < nq$,*

a) *if Z has a symmetric distribution, then, without any moment assumption,*

$$\Pr\left(n\overline{Z}_n' S_n^{-2} \overline{Z}_n \ge t\right) \le 2q e^{-\frac{t}{2q}}; \tag{5}$$

b) *for general distribution of Z, with $\gamma_4 < \infty$, for any $a > 1$,*

$$\begin{aligned} \Pr\left(n\overline{Z}_n' S_n^{-2} \overline{Z}_n \ge t\right) &\le 2q e^{1-\frac{t}{2q(1+a)}} + C(q)\, n^{3\tilde{q}} \gamma_4^{-\tilde{q}} e^{-\frac{n}{\gamma_4(q+1)}\left(1-\frac{1}{a}\right)^2} \\ &\le 2q e^{1-\frac{t}{2q(1+a)}} + K(q)\, n^{3\tilde{q}} e^{-\frac{n}{\gamma_4(q+1)}\left(1-\frac{1}{a}\right)^2} \end{aligned} \tag{6}$$

with $\tilde{q} = \frac{q-1}{q+1}$ and

$$C(q) = \frac{(2e\pi)^{2\tilde{q}}(q+1)}{2^{2/(q+1)}(q-1)^{3\tilde{q}}} \quad \textit{and} \quad K(q) = \frac{C(q)}{q^{2\tilde{q}}} \le 8.$$

Moreover for $nq \le t$, we have

$$\Pr\left(n\overline{Z}_n' S_n^{-2} \overline{Z}_n \ge t\right) = 0.$$

The proof is postponed to Appendix (1). Part a) in the symmetric multidimensional case follows by an easy but crude extension of [12] or [9, 8]. It is also given under a different form in [10]. The exponential inequality (5) is classical in the unidimensional case. Other type of inequalities with suboptimal rate in the exponential term have also been obtained by [14].
In the general multidimensional framework, the main difficulty is actually to keep the self-normalized structure when symmetrizing the original sum. We first establish the inequality in the symmetric case by an appropriate diagonalization of the estimated covariance matrix, which reduces the problem to q -unidimensional inequalities. The next step is to use a multidimensional version of Panchenko's symmetrization lemma (see [16]). However this symmetrization lemma destroys partly the self-normalized structure (the normalization is then $S_n^2 + S^2$ instead of the expected S_n^2), which can be retrieved by obtaining a lower tail control of the distance between S_n^2 and S^2. This is done by studying the behavior of the smallest eigenvalue of the normalizing empirical variance. The second term in the right hand side of inequality (6) is essentially due to this control.
However, for $q > 1$, the bound of part a) is clearly not optimal. A better bound, which has not exactly an exponential form, has been obtained by [17] following previous works by [7]. Pinelis' result gives a much more precise evaluation of the tail for moderate q. It essentially says that in the symmetric case the tail of the self-normalized sum can essentially be bounded by the tail of a

$\chi^2(q)$ distribution. Notice that this tail $\overline{F}_q$ satisfies the following approximation (see [1], p. 941, result 26.4.12)

$$\overline{F}_q(t) \underset{t\to\infty}{\sim} \frac{1}{\Gamma(\frac{q}{2})}\left(\frac{t}{2}\right)^{\frac{q}{2}-1}\exp(-\frac{t}{2}).$$

This bounds gives the right behavior of the tail (in q) as t grows, which is not the case for a). However, in the unidimensional case a) still gives a better approximation than [17]. a) can still be used in the multidimensional case to get crude but exponential bounds. We expect however Pinelis' inequality to give much better bounds for moderate q and moderate sample size n in the symmetric case. For these reason, we will extend the results of Theorem 1 by using a $\chi^2(q)$ type of control. This essentially consists in extending Lemma 1 of [16] to non exponential bound.

Theorem 2. *The following inequalities hold, **for finite** $n > q$ and for $t < nq$:*

- ***a)*** *(Pinelis 1994) if Z has a symmetric distribution, without any moment assumption, then we have*

$$\Pr\left(n\overline{Z}_n' S_n^{-2}\overline{Z}_n \geq t\right) \leq \frac{2e^3}{9}\overline{F}_q(t), \tag{7}$$

- ***b)*** *for general distribution of Z with kurtosis $\gamma_4 < \infty$, for any $a > 1$ and for $t \geq 2q(1+a)$ and $\tilde{q} = \frac{q-1}{q+1}$ we have*

$$\begin{aligned}
&\Pr\left(n\overline{Z}_n' S_n^{-2}\overline{Z}_n \geq t\right) \\
&\leq \frac{2e^3}{9\Gamma(\frac{q}{2}+1)}\left(\frac{t-q(1+a)}{2(1+a)}\right)^{\frac{q}{2}} e^{-\frac{t-q(1+a)}{2(1+a)}} + C(q)\left(\frac{n^3}{\gamma_4}\right)^{\tilde{q}} e^{-\frac{n\left(1-\frac{1}{a}\right)^2}{\gamma_4(q+1)}} \\
&\leq \frac{2e^3}{9\Gamma(\frac{q}{2}+1)}\left(\frac{t-q(1+a)}{2(1+a)}\right)^{\frac{q}{2}} e^{-\frac{t-q(1+a)}{2(1+a)}} + K(q)\,n^{3\tilde{q}} e^{-\frac{n\left(1-\frac{1}{a}\right)^2}{\gamma_4(q+1)}}
\end{aligned} \tag{8}$$

For $t \geq nq$, we have $\Pr\left(n\overline{Z}_n' S_n^{-2}\overline{Z}_n \geq t\right) = 0.$

Remark 1. *Notice that the constant $K(q)$ does not increase with large q as it can be seen on Figure 1. A close examination of the bounds shows that essentially $\gamma_4(q+1)$ has to be small compared to n*

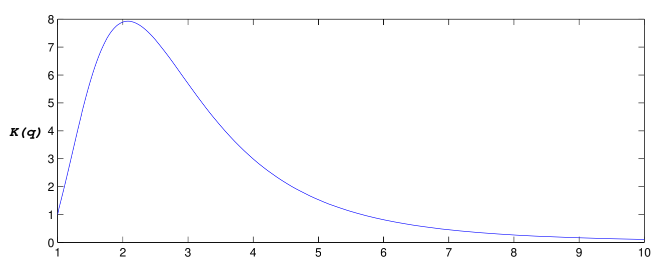

Figure 1: Value of $K(q)$ as a function of q

for practical use of these bounds. Of course practically γ_4 is not known, however one may use an estimator or an upper bound for this quantity to get some insight on a given estimation problem.

Remark 2. *It can be tempting to compare our bounds with some more classical results in statistics. We recall that, in an unidimensional framework, the studentized ratio is given by* $\widetilde{T}_n = \sqrt{n}\widetilde{S}_n^{-1}\bar{Z}_n$ *where* $\widetilde{S}_n$ *is the unbiased estimator of the variance* $\widetilde{S}_n = (\frac{1}{n-1}\sum_{i=1}^n (Z_i - \bar{Z}_n)^2)^{-1/2}$*. In a Gaussian framework,* $\widetilde{T}_n$ *has a Student distribution with* $(n-1)$ *degrees of freedom. In opposition, our self-normalized sum is defined by* $T_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Z_i^2\right)^{-1/2}\bar{Z}_n$*. It is related to* $\widetilde{T}_n$ *by the relation* $T_n = f_n(\widetilde{T}_n)$ *with* $f_n(x) = \sqrt{\frac{n}{n-1}}\left(1 + \frac{x^2}{n-1}\right)^{-1/2} x$*. As a consequence, one gets in an unidimensional symmetric case, for* $t > 0$*,*

$$\Pr(\widetilde{T}_n \geq t) \leq \exp\left\{ -\frac{1}{2}\frac{n}{n-1}\frac{t^2}{1 + \frac{t^2}{n-1}} \right\}.$$

For large n we recover an sub-gaussian type of inequality. At fixed n, , this inequality is noninformative for $t \to \infty$ *since the right-hand side tends to 1. Recall that, in a Gaussian framework, the tail* $Pr(\widetilde{T}_n > t)$ *is of order* $O(\frac{1}{t^{n-1}})$ *as* $t \to \infty$*.*

Remark 3. *In the best case, past studies give some bounds for n sufficiently large, without an exact value for "sufficiently large". Here, the bounds are valid and explicit for any* $n > q$*.*

These bounds are motivated by some statistical applications to the construction of non-asymptotic confidence intervals with conservative coverage probability in a semi-parametric setting. Self-normalized sums appear naturally in the context of empirical likelihood and its generalization to Cressie-Read divergences, see [11, 15]. In particular, [5] shows how the bounds obtained here may be used to construct explicit non asymptotic confidence regions, even when q depends on n.

A Proofs of the main results

A.1 Some lemmas

The first lemma is a direct extension of Panchenko, 2003, Corollary 1 to the multidimensional case, which will be used both in theorem 1 and 2.

Lemma 1. *Let* $\mathcal{S}_q$ *be the unit sphere of* $\mathbb{R}^q$*,* $\mathcal{S}_q = \{\lambda \in \mathbb{R}^q,\ \|\lambda\|_2 = 1\}$*. Let* $Z^{(n)} = (Z_i)_{1\leq i\leq n}$ *and* $Y^{(n)} = (Y_i)_{1\leq i\leq n}$ *be i.i.d. centered random vectors in* $\mathbb{R}^q$ *with* $Z^{(n)}$ *independent of* $Y^{(n)}$*. We denote, for any random vector* $W = (W_i)_{1\leq i\leq n}$ *with coordinates in* $\mathbb{R}^q$*,* $S^2_{n,W} = \frac{1}{n}\sum_i^n W_i W_i'$*.*
If there exists $D > 0$ *and* $d > 0$ *such that, for all* $t \geq 0$*,*

$$\Pr\left(\sup_{\lambda\in\mathcal{S}_q} \left(\frac{\sqrt{n}\lambda'(\overline{Z}_n - \overline{Y}_n)}{\sqrt{\lambda' S^2_{n,(Z^{(n)}-Y^{(n)})}\lambda}} \right) \geq \sqrt{t} \right) \leq De^{-dt}, \tag{9}$$

then, for all $t \geq 0$*,*

$$\Pr\left(\sup_{\lambda\in\mathcal{S}_q} \frac{\sqrt{n}\lambda'\overline{Z}_n}{\sqrt{\lambda' S_n^2\lambda + \lambda' S^2\lambda}} \geq \sqrt{t} \right) \leq De^{1-dt}. \tag{10}$$

Proof. This proof follows Lemma 1 of [16] with some adaptations to the multidimensional case. Denote

$$A_n(Z^{(n)}) = \sup_{\lambda \in \mathscr{I}_q} \sup_{b>0} \left\{ E\left[4b(\lambda'(\overline{Z}_n - \overline{Y}_n) - b\lambda' S^2_{n,Z^{(n)}-Y^{(n)}}\lambda) | Z^{(n)} \right] \right\}$$

$$C_n(Z^{(n)}, Y^{(n)}) = \sup_{\lambda \in \mathscr{I}_q} \sup_{b>0} \left\{ 4b(\lambda'(\overline{Z}_n - \overline{Y}_n) - b\lambda' S^2_{n,Z^{(n)}-Y^{(n)}}\lambda) \right\}.$$

By Jensen inequality, we have Pr-almost surely

$$A_n(Z^{(n)}) \leq E[C_n(Z^{(n)}, Y^{(n)})|Z^{(n)}]$$

and, for any convex function Φ, by Jensen inequality, we also get

$$\Phi(A_n(Z^{(n)})) \leq E[\Phi(C_n(Z^{(n)}, Y^{(n)}))|Z^{(n)}].$$

We obtain

$$E(\Phi(A_n(Z^{(n)}))) \leq E(\Phi(C_n(Z^{(n)}, Y^{(n)}))). \tag{11}$$

Now remark that

$$\begin{aligned} A_n(Z^{(n)}) &= \sup_{\lambda \in \mathscr{I}_q} \sup_{b>0} \left\{ 4b\left(\lambda'\overline{Z}_n - b\lambda' S_n^2 \lambda - b\lambda' S^2 \lambda \right) \right\} \\ &= \sup_{\lambda \in \mathscr{I}_q} \left(\frac{\lambda'\overline{Z}_n}{\sqrt{\lambda' S_n^2 \lambda + \lambda' S^2 \lambda}} \right)^2 \end{aligned}$$

and

$$C_n(Z^{(n)}, Y^{(n)}) = \sup_{\lambda \in \mathscr{I}_q} \left(\frac{\lambda'(\overline{Z}_n - \overline{Y}_n)}{\sqrt{\lambda' S^2_{n,Z^{(n)}-Y^{(n)}}\lambda}} \right)^2.$$

Now, notice that $\sup_{\lambda \in \mathscr{I}_q} \frac{\lambda'\overline{Z}_n}{\sqrt{\lambda' S_n^2 \lambda}} > 0$ and apply the arguments of the proof of [16]'s Corollary 1 applied to inequality (11) to obtain the result. □

The next lemma allows to establish an non exponential version of the preceding lemmas. Indeed a consequence of this lemma is that, if the tail of the symmetrized version in inequality (9) is controlled by a chi-square tail, then the non symmetrized version may be controlled by an exponential multiplied by a polynomial. The rate in the exponential is asymptotically correct.

Lemma 2. *For any $t > q$, let $\Phi_t(x) = \max(x - t + q; 0)$. Let ν and ξ be two r.v.'s, such that for any $t > q$, $E\left[\Phi_t(\xi)\right] \leq E\left[\Phi_t(\nu)\right]$. Suppose that, there exists a constant $C > 0$ such that, for $t > 0$,*

$$\Pr(\nu > t) \leq C\overline{F}_q(t).$$

Then, for $t \geq 2q$, we have

$$\Pr(\xi > t) \leq C \left(\frac{(t-q)}{2} \right)^{\frac{q}{2}} \frac{e^{-\frac{(t-q)}{2}}}{\Gamma(q/2+1)}.$$

and for $t > q$, we have

$$\Pr(\xi > t) \leq C\overline{F}_{q+2}(t-q).$$

Proof. We follow the lines of the proof of Panchenko's lemma, with function Φ_t. Remark that $\Phi_t(0) = 0$ and $\Phi_t(t) = q$, then we have

$$\begin{aligned}\Pr(\xi \geq t) &\leq \frac{1}{\Phi_t(t)}\left(\Phi_t(0) + \int_0^{+\infty} \Phi_t'(x)\Pr(\nu \geq x)dx\right) \\ &\leq \frac{C}{q}\int_{t-q}^{+\infty} \overline{F}_q(x)dx.\end{aligned}$$

By integration by parts, we have

$$\int_{t-q}^{+\infty} \overline{F}_q(x)dx = \int_{t-q}^{+\infty} x f_q(x)dx - (t-q)\int_{t-q}^{+\infty} f_q(x)dx.$$

It follows by straightforward calculations that, for $t > q$,

$$\Pr(\xi \geq t) \leq \frac{C}{q}\int_{t-q}^{+\infty} \overline{F}_q(x)dx \leq C\left(\overline{F}_{q+2}(t-q) - \frac{t-q}{q}\overline{F}_q(t-q)\right).$$

For $t \geq 2q$, and using the recurrence relation 26.4.8 of [1], page 941.

$$\Pr(\xi \geq t) \leq C\left(\overline{F}_{q+2}(t-q) - \overline{F}_q(t-q)\right) = \left(\frac{(t-q)}{2}\right)^{q/2}\frac{Ce^{-\frac{(t-q)}{2}}}{\Gamma(\frac{q}{2}+1)}.$$

Moreover, for $t > q$ we have $\Pr(\xi \geq t) \leq C\overline{F}_{q+2}(t-q)$. $\square$

We now extend a result of [3], which controls the behavior of the smallest eigenvalue of the empirical variance. In the following, for a given symmetric matrix A, we denote $\mu_1(A)$ its smallest eigenvalue.

Lemma 3. *Let $(Z_i)_{1\leq i\leq n}$ be i.i.d. random vectors in $\mathbb{R}^q$ with common mean 0. Assume $1 \leq \widetilde{\gamma}_4 = E(\|Z_1\|_2^4) < +\infty$. Then, for any $n > q$ and $0 < t \leq \mu_1(S^2)$,*

$$\Pr\left(\mu_1(S_n^2) \leq t\right) \leq C(q)\,\frac{n^{3\tilde{q}}\mu_1(S^2)^{2\tilde{q}}}{\widetilde{\gamma}_4^{\tilde{q}}}\exp\left(-\frac{n(\mu_1(S^2)-t)^2}{\widetilde{\gamma}_4(q+1)}\right) \wedge 1,$$

with $\widetilde{q} = \frac{q-1}{q+1}$ and

$$\begin{aligned}C(q) &= \pi^{2\tilde{q}}(q+1)e^{2\tilde{q}}(q-1)^{-3\tilde{q}}2^{2\tilde{q}-\frac{2}{q+1}} &(12)\\ &\leq 4\pi^2(q+1)e^2(q-1)^{-3\tilde{q}}. &(13)\end{aligned}$$

Proof. The proof of this result is adapted from [3] and makes use of some idea of [4] .
We first have by a truncation argument and applying Markov's inequality on the last term in the inequality (see the proof of [3], Lemma 4), for every $M > 0$,

$$\Pr\left(\mu_1\left(\sum_{i=1}^n Z_i Z_i'\right) \leq nt\right) \leq \Pr\left(\inf_{v\in\mathscr{S}_q}\sum_{i=1}^n (v'Z_i)^2 \leq nt,\ \sup_{i=1,\ldots,n}\|Z_i\|_2 \leq M\right) + n\frac{\widetilde{\gamma}_4}{M^4} \qquad (14)$$

We call I the first term on the right hand side of this inequality.

Notice that by symmetry of the sphere, we can always work with the northern hemisphere of the sphere rather than the sphere. In the following, we denote by $\mathcal{N}_q$ the northern hemisphere of the sphere. Notice that, if the supremum of the $||Z_i||_2$ is smaller than M, then for u, v in $\mathcal{N}_q$, we have

$$\left| \sum_{i=1}^{n} (v'Z_i)^2 - \sum_{i=1}^{n} (u'Z_i)^2 \right| \leq 2n||u-v||_2 M^2.$$

Thus if u and v are apart by at most $t\eta/(2M^2)$ then $|\sum_{i=1}^{n}(v'Z_i)^2 - \sum_{i=1}^{n}(u'Z_i)^2| \leq \eta n t$. Now let $N(\mathcal{N}_q, \varepsilon)$ be the smallest number of caps of radius ε centered at some points on $\mathcal{N}_q$ (for the $||.||_2$ norm) needed to cover $\mathcal{N}_q$. Now we follow the same arguments as [3] to control I: I is bounded by the sum of the probabilities that the infimum of $\sum_{i=1}^{n}(v'Z_i)^2$ over each cap is smaller thant nt and that $\sup_{i=1,\ldots,n} ||Z_i||_2 \leq M$. We bound this sum by the number of caps times the larger probability: for any $\eta > 0$,

$$I \leq N\left(\mathcal{N}_q, \frac{t\eta}{2M^2}\right) \max_{u \in \mathcal{N}_q} \Pr\left(\sum_{i=1}^{n} (u'Z_i)^2 \leq (1+\eta)nt\right).$$

The proof is now divided in three steps, i) control of $N(\mathcal{N}_q, \frac{t\eta}{2M^2})$, ii) control of the maximum over $\mathcal{N}_q$ of the last expression in I, iii) optimization over all the free parameters.

i) On the one hand, we have, for some constant $b(q) > 0$,

$$N(\mathcal{N}_q, \varepsilon) \leq b(q)\varepsilon^{-(q-1)} \vee 1. \tag{15}$$

For instance, we may choose $b(q) = \pi^{q-1}$. Indeed, following [3], the northern hemisphere can be parameterized in polar coordinates, realizing a diffeomorphism with $\mathcal{J}_{q-1} \times [0, \pi]$. Now proceed by induction, notice that for $q = 2$, $\mathcal{N}_q$, the half circle can be covered by $[\pi/2\varepsilon] \vee 1 + 1 \leq 2([\pi/2\varepsilon] \vee 1) \leq \pi/\varepsilon \vee 1$ caps of diameter 2ε, that is, we can choose the caps with their center on a $\varepsilon-$grid on the circle. Now, by induction we can cover the cylinder $\mathcal{J}_{q-1} \times [0, \pi]$ with $[\pi/2\varepsilon\ (\pi)^{q-2}/\varepsilon^{q-2}] \vee 1 + 1 \leq \pi^{q-1}/\varepsilon^{q-1}$ intersecting cylinders which in turn can be mapped to region belonging to caps of radius ε, covering the whole sphere (this is still a covering because the mapping from the cylinder to the sphere is contractive).

ii) On the other hand, for all $t > 0$, we have by exponentiation and Markov's inequality, and independence of $(Z_i)_{1 \leq i \leq n}$, for any $\lambda > 0$

$$\max_{u \in \mathcal{N}_q} \Pr\left(\sum_{i=1}^{n} u'Z_i Z_i' u \leq nt\right) \leq e^{n\lambda t} \max_{u \in \mathcal{N}_q} \left(E\left[e^{-\lambda u' Z_1 Z_1' u}\right]\right)^n.$$

Now, using the classical inequalities, $\log(x) \leq x-1$ and $e^{-x}-1 \leq -x+x^2/2$, both valid for $x>0$, we have

$$\begin{aligned}
\max_{u\in\mathcal{N}_q}\left(E\left[e^{-\lambda u' Z_1 Z_1' u}\right]\right)^n &= \max_{u\in\mathcal{N}_q}\exp\left\{n\log\left(E\left[e^{-\lambda u' Z_1 Z_1' u}\right]\right)\right\} \\
&\leq \max_{u\in\mathcal{N}_q}\exp\left\{nE\left[e^{-\lambda u' Z_1 Z_1' u}-1\right]\right\} \qquad (16)\\
&\leq \max_{u\in\mathcal{N}_q}\exp\left\{n\left(-\lambda u' S^2 u+\frac{\lambda^2}{2}\widetilde{\gamma}_4\right)\right\} \\
&= \exp\left(\frac{\lambda^2}{2}n\widetilde{\gamma}_4-\lambda n\mu_1(S^2)\right). \qquad (17)
\end{aligned}$$

iii) From (17) and (15), we deduce that, for any $t>0, \lambda>0,\ \eta>0$,

$$I \leq b(q)\left(\frac{2M^2}{t\eta}\right)^{q-1}\exp\left(\lambda(1+\eta)nt+\frac{\lambda^2}{2}n\widetilde{\gamma}_4-\lambda n\mu_1(S^2)\right).$$

Optimizing the expression $\exp(-(q-1)\log(\eta)+\lambda\eta nt)$ in $\eta>0$, yields immediately, for any $t>0$, any $M>0$, any $\lambda>0$

$$I \leq b(q)\left(\frac{2enM^2\lambda}{q-1}\right)^{q-1}\exp\left(\lambda n\left(t-\mu_1(S^2)\right)+\frac{\lambda^2}{2}n\widetilde{\gamma}_4\right).$$

The infimum in λ in the exponential term is attained at $\lambda=\frac{\mu_1(S^2)-t}{\widetilde{\gamma}_4}$, provided that $0<t<\mu_1(S^2)$. Therefore, for such t and all $M>0$, we get that $\Pr(\mu_1(\sum_{i=1}^n Z_i Z_i')\leq nt)$ is less than

$$b(q)\left(\frac{2enM^2\mu_1(S^2)}{\widetilde{\gamma}_4(q-1)}\right)^{q-1}\exp\left(-\frac{n}{2\widetilde{\gamma}_4}\left(\mu_1(S^2)-t\right)^2\right)+n\frac{\widetilde{\gamma}_4}{M^4}.$$

We now optimize in $M^2>0$ and the optimum is attained at

$$M_*^2=\left(\frac{2n\widetilde{\gamma}_4}{(q-1)b(q)}\right)^{\frac{1}{q+1}}\left(\frac{2en}{q-1}\frac{\mu_1(S^2)}{\widetilde{\gamma}_4}\right)^{-\frac{(q-1)}{q+1}}\exp\left(\frac{n(\mu_1(S^2)-t)^2}{2\widetilde{\gamma}_4(q+1)}\right),$$

yielding the bound

$$\Pr\left(\mu_1\left(n^{-1}\sum_{i=1}^n Z_i Z_i'\right)\leq t\right)\leq \tilde{C}(q)\,n^{3\frac{q-1}{q+1}}\mu_1(S^2)^{\frac{2(q-1)}{q+1}}\,\widetilde{\gamma}_4^{-\frac{q-1}{q+1}}\exp\left(-\frac{n\left(\mu_1(S^2)-t\right)^2}{\widetilde{\gamma}_4(q+1)}\right),$$

with

$$\tilde{C}(q)=b(q)^{\frac{2}{q+1}}(q+1)e^{\frac{2(q-1)}{q+1}}(q-1)^{-3\frac{q-1}{q+1}}2^{\frac{2q-4}{q+1}}.$$

Using the constant $b(q)=\pi^{q-1}$ we get the expression of $C(q)$, which is bounded by the simpler bound (for large q this bound will be sufficient) $4\pi^2(q+1)e^2(q-1)^{-3\frac{q-1}{q+1}}$, using the fact that $\widetilde{\gamma}_4\geq 1$. The result of the Lemma follows by using this inequality combined with inequality 14. □

A.2 Proof of Theorem 1

Proof. Notice that we have always $\bar{Z}_n' S_n^{-2} \bar{Z}_n \leq q$. Indeed, there exists an orthogonal transformation O_n and a diagonal matrix $\Lambda_n^2 := \text{diag}[\hat{\mu}_j]_{1\leq j\leq q}$ with $\hat{\mu}_j > 0$ being the eigenvalues of S_n^2, such that $S_n^2 = O_n^{'} \Lambda_n^2 O_n$. Now put $Y_i := [Y_{i,j}]_{1\leq j\leq q} = O_n Z_i$. It is easy to see that by construction the empirical variance of the Y_i is

$$\frac{1}{n}\sum_{i=1}^{n} Y_i Y_i' = \frac{1}{n}\sum_{i=1}^{n} O_n Z_i Z_i' O_n' = O_n S_n^2 O_n^{'} = \Lambda_n^2.$$

It also follows from this equality that, for all $j = 1, \cdots, q$, $\frac{1}{n}\sum_{i=1}^{n} Y_{i,j}^2 = \hat{\mu}_j$, and

$$\bar{Z}_n' S_n^{-2} \bar{Z}_n = \bar{Y}_n' \Lambda_n^{-2} \bar{Y}_n = \sum_{j=1}^{q} \left(\frac{1}{n} \sum_{i=1}^{n} Y_{i,j} \right)^2 / \hat{\mu}_j.$$

This quantity is lower than q by Cauchy-Schwartz inequality. So, it follows that, for all $t > qn$

$$\Pr\left(n\bar{Z}_n' S_n^{-2} \bar{Z}_n \geq t \right) = 0.$$

a) In the symmetric and unidimensional framework ($q = 1$), this bound follows from Hoeffding inequality (see [9]). Consider now the symmetric multidimensional framework ($q > 1$). Let $\sigma_i, 1 \leq i \leq n$ be Rademacher random variables, independent from $(Z_i)_{1\leq i\leq n}$, $\mathbb{P}(\sigma_i = -1) = \mathbb{P}(\sigma_i = 1) = 1/2$. We denote $\sigma_n(Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \sigma_i Z_i$ and remark that $S_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i Z_i Z_i' \sigma_i$. Since the Z_i's have a symmetric distribution, meaning that $-Z_i$ has the same distribution as Z_i, we make use of a first symmetrization step:

$$\Pr\left(n\overline{Z}_n^{'} S_n^{-2} \overline{Z}_n \geq t \right) = \Pr(\sigma_n(Z)^{'} S_n^{-2} \sigma_n(Z) \geq t).$$

Now, we have

$$\sigma_n(Z)^{'} S_n^{-2} \sigma_n(Z) = \sigma_n(Y)^{'} \Lambda_n^{-2} \sigma_n(Y) = \sum_{j=1}^{q} \left(\sum_{i=1}^{n} \sigma_i Y_{i,j} \right)^2 / \sum_{i=1}^{n} Y_{i,j}^2.$$

It follows that, for $t > 0$,

$$\begin{aligned} \Pr(\sigma_n(Z)^{'} S_n^{-2} \sigma_n(Z) \geq t) &\leq \sum_{j=1}^{q} \Pr\left(\frac{|\sum_{i=1}^{n} \sigma_i Y_{i,j}|}{\sqrt{\sum_{i=1}^{n} Y_{i,j}^2}} \geq \sqrt{t/q} \right) \\ &\leq 2\sum_{j=1}^{q} E \Pr\left(\frac{\sum_{i=1}^{n} \sigma_i Y_{i,j}}{\sqrt{\sum_{i=1}^{n} Y_{i,j}^2}} \geq \sqrt{t/q} \,\middle|\, (Z_i)_{1\leq i\leq n} \right). \end{aligned}$$

Apply now Hoeffding inequality to each unidimensional self-normalized term in this sum to conclude.

b) The Z_i's are not anymore symmetric. Define

$$B_n = \sup_{\lambda \in \mathscr{S}_q} \left\{ \frac{\lambda' \overline{Z}_n}{\sqrt{\lambda' S_n^2 \lambda}} \right\} \text{ and } D_n = \sup_{\lambda \in \mathscr{S}_q} \left\{ \sqrt{1 + \frac{\lambda' S^2 \lambda}{\lambda' S_n^2 \lambda}} \right\}.$$

First of all, remark that the following events are equivalent

$$\left\{ n \overline{Z}_n' S_n^{-2} \overline{Z}_n \geq t \right\} = \left\{ B_n \geq \sqrt{\frac{t}{n}} \right\}. \tag{18}$$

Indeed, the supremum in the definition of B_n is reached at $\lambda = S_n^{-2} \overline{Z}_n'$ and then $B_n = \sqrt{\overline{Z}_n' S_n^{-2} \overline{Z}_n}$. Notice that

$$\Pr\left(B_n \geq \sqrt{\frac{t}{n}} \right) \leq \inf_{a > -1} \left\{ \Pr\left(B_n D_n^{-1} \geq \sqrt{\frac{t}{n(1+a)}} \right) + \Pr(D_n \geq \sqrt{1+a}) \right\}.$$

The control of the first term on the right hand side is obtained in two steps. First apply part a) of Theorem 1 to $n^{1/2} \sup_{\lambda \in \mathscr{S}_q} \frac{\lambda'\left(\overline{Z}_n - \overline{Y}_n\right)}{\sqrt{\lambda' S_{n,Z-Y}^2 \tilde{\lambda}}}$. Then, by application of Lemma 1 and (18), we get

$$\sqrt{n} B_n D_n^{-1} \leq n^{1/2} \sup_{\lambda \in \mathscr{S}_q} \frac{\lambda' \overline{Z}_n}{\sqrt{\lambda' S_n^2 \tilde{\lambda} + \lambda' S^2 \tilde{\lambda}}},$$

and then we have for all $t > 0$,

$$\Pr\left(B_n D_n^{-1} \geq \sqrt{\frac{t}{n(1+a)}} \right) \leq 2q e^{1 - \frac{t}{2q(1+a)}}.$$

For all $a > 0$ and all $t > 0$, we have

$$\begin{aligned} \left\{ D_n \geq \sqrt{1+a} \right\} &= \left\{ \sup_{\lambda \in \mathscr{S}_q} \left(1 + \frac{\lambda' S^2 \lambda}{\lambda' S_n^2 \lambda} \right) \geq 1 + a \right\} \\ &= \left\{ \inf_{\lambda \in \mathscr{S}_q} \left(\lambda' S^{-1} S_n^2 S^{-1} \lambda \right) \leq \frac{1}{a} \right\} \subset \left\{ \mu_1 (S^{-1} S_n^2 S^{-1}) \leq \frac{1}{a} \right\}. \end{aligned}$$

We now use Lemma 3 applied to the r.v.'s $(S^{-1} Z_i)_{1 \leq i \leq n}$ with covariance matrix equal to Id_q. It is easy to check that $\gamma_4 = \widetilde{\gamma}_4$. For all $1 < a$, we have,

$$\Pr(D_n > \sqrt{1+a}) \leq C(q) \left(\frac{n^3}{\gamma_4} \right)^{\tilde{q}} e^{-\frac{n}{(q+1)\gamma_4}(1 - \frac{1}{a})^2}.$$

Since $\inf_{a > -1} \leq \inf_{a > 1}$, we conclude that, for any $t > 0$,

$$\Pr\left(B_n > \sqrt{\frac{t}{n}} \right) \leq \inf_{a > 1} \left\{ 2q e \, e^{-\frac{t}{2q(1+a)}} + C(q) \left(\frac{n^3}{\gamma_4} \right)^{\tilde{q}} e^{-\frac{n}{(q+1)\gamma_4}(1 - \frac{1}{a})^2} \right\}.$$

We achieve the proof by noticing that $\gamma_4 \geq q^2$ from Jensen's inequality and $E(\|S^{-1} Z\|_2^2) = q$. $\square$

A.3 Proof of Theorem 2.

Part a) is proved in [17]. Now, the proof of part b) follows the same lines as the Theorem 1 combining Lemmas 1, 2 and 3.

References

[1] M. Abramovitch and L. A. Stegun. *Handbook of Mathematical Tables*. National Bureau of Standards, Washington, DC, 1970.

[2] R. R. Bahadur and L. J. Savage. The nonexistence of certain statistical procedures in non-parametric problems. *Annals of Mathematical Statistics*, 27:1115–1122, 1956. MR0084241

[3] P. Barbe and P. Bertail. Testing the global stability of a linear model. Working Paper nř46, CREST, 2004.

[4] B. Bercu, E. Gassiat, and E. Rio. Concentration inequalities, large and moderate deviations for self-normalized empirical processes. *Annals of Probability*, 30(4):1576–1604, 2002. MR1944001

[5] P. Bertail, E. Gauthérat, and H. Harari-Kermadec. Exponential bounds for quasi-empirical likelihood. Working Paper nř34, CREST, 2005.

[6] G. P. Chistyakov and F. Götze. Moderate deviations for Student's statistic. *Theory of Probability & Its Applications*, 47(3):415–428, 2003.

[7] M. L. Eaton. A probability inequality for linear combinations of bounded random variables. *Annals of Statistics*, 2:609–614, 1974.

[8] M. L. Eaton and B. Efron. Hotelling's t^2 test under symmetry conditions. *Journal of american statistical society*, 65:702–711, 1970. MR0269021

[9] B. Efron. Student's t-test under symmetry conditions. *Journal of american statistical society*, 64:1278–1302, 1969. MR0251826

[10] E. Giné and F. Götze. On standard normal convergence of the multivariate Student t-statistic for symmetric random vectors. *Electron. Comm. Probab.*, 9:162–171 (electronic), 2004. ISSN 1083-589X. MR2108862

[11] H. Harari-Kermadec. *Vraisemblance empirique généralisée et estimation semi-paramétrique*. PhD thesis, Université Paris X, 2006.

[12] W. Hoeffding. Probability inequalities for sums of bounded variables. *Journal of the American Statistical Association*, 58:13–30, 1963. MR0144363

[13] B.-Y. Jing and Q. Wang. An exponential nonuniform Berry-Esseen bound for self-normalized sums. *Annals of Probability*, 27(4):2068–2088, 1999. MR1742902

[14] P. Major. A multivariate generalization of Hoeffding's inequality. Arxiv preprint math. PR/0411288, 2004.

[15] A. B. Owen. *Empirical Likelihood*. Chapman and Hall/CRC, Boca Raton, 2001.

[16] D. Panchenko. Symmetrization approach to concentration inequalities for empirical processes. *Annals of Probability*, 31(4):2068–2081, 2003. MR2016612

[17] I. Pinelis. Probabilistic problems and Hotelling's t^2 test under a symmetry condition. *Annals of Statistics*, 22(1):357–368, 1994. MR1272088

[18] J. P. Romano and M. Wolf. Finite sample nonparametric inference and large sample efficiency. *Annals of Statistics*, 28(3):756–778, 2000. MR1792786