High-dimensional Statistical Methods

Skoltech

Q. Paris*

National Research University HSE
Faculty of Computer Science
Moscow, Russia

A1. Sub-gaussian random variables

1 Definitions 2

2 Tail behavior and moments 2

3 Maximal inequalities 3
3.1 Finite maximum 4
3.2 Maximum over a convex polytope 6
3.3 The ε-net argument 7

4 Exercises 10

*email:qparis@hse.ru, teaching material: http://www.qparis-math.com/teaching.

The material gathered below is borrowed from various sources including Rigollet (2015) and van Handel (2016).

1 Definitions

Definition 1.1. *A centered and real-valued random variable ξ is said to be sub-gaussian with variance proxy $\sigma^2 > 0$ if, for all $\lambda \in \mathbf{R}$,*

$$\log \mathbf{E} \exp(\lambda \xi) \leq \frac{\lambda^2 \sigma^2}{2}.$$

In this case, we denote $\xi \in \text{SG}(\sigma^2)$.

Some basic intuitions on this notion and examples of sub-gaussian random variables are provided in Exercise (4.1) below. The previous definition extends to the case of random vectors and matrices as follows.

Definition 1.2. (1) *An $\mathbf{R}^d$-valued random vector $\boldsymbol{\xi}$ is said sub-gaussian with variance proxy $\sigma^2 > 0$ if*

$$u^\top \boldsymbol{\xi} \in \text{SG}(\sigma^2),$$

for all $u \in \mathbf{R}^d$ such that $\|u\|_2 \leq 1$. In this case, we denote $\boldsymbol{\xi} \in \text{SG}_d(\sigma^2)$.

(2) *An $\mathrm{M}_{p,q}(\mathbf{R})$-valued random matrix $\boldsymbol{A}$ is said sub-gaussian with variance proxy $\sigma^2 > 0$ if*

$$u^\top \boldsymbol{A} v \in \text{SG}(\sigma^2),$$

for all $(u, v) \in \mathbf{R}^p \times \mathbf{R}^q$ such that $\|u\|_2 \leq 1$ and $\|v\|_2 \leq 1$. In this case, we denote $\boldsymbol{A} \in \text{SG}_{p,q}(\sigma^2)$.

Simple but useful observations about these definitions are discussed in Exercise (4.2) at the end of the appendix.

2 Tail behavior and moments

Next, we describe some fundamental properties of sub-gaussian variables that will be often used in the sequel. We start by an upper bound on the tails of a sub-gaussian variable.

Theorem 2.1. *Let ξ be a centered and sub-gaussian random variable with variance proxy $\sigma^2 > 0$. Then, for all $t \geq 0$,*

$$\mathbf{P}(\xi \geq t) \vee \mathbf{P}(\xi \leq -t) \leq e^{-\frac{t}{2\sigma^2}} \quad \text{and} \quad \mathbf{P}(|\xi| \geq t) \leq 2e^{-\frac{t}{2\sigma^2}}.$$

The proof of this result is left as an exercise and is based on the important Chernoff bound discussed in Exercise (4.3) below. The next result builds upon the previous to provide upper bounds on the moments of sub-gaussian variables.

Theorem 2.2. *Let ξ be a centered and sub-gaussian random variable with variance proxy $\sigma^2 > 0$. Then, $\mathbf{E}|\xi| \le \sigma\sqrt{2\pi}$ and*

$$k^{-\frac{1}{2}}(\mathbf{E}|\xi|^k)^{\frac{1}{k}} \le e^{\frac{1}{e}}\sigma,$$

for all $k \ge 2$.

Proof of Theorem 2.2. First observe that, for all $k \ge 1$,

$$\begin{aligned}
\mathbf{E}|\xi|^k = \int_0^{+\infty} \mathbf{P}(|\xi|^k > t)\mathrm{d}t &= \int_0^{+\infty} \mathbf{P}(|\xi| > t^{1/k})\mathrm{d}t \\
&\le 2\int_0^{+\infty} \exp(-t^{\frac{2}{k}}/2\sigma^2)\mathrm{d}t \\
&= k(2\sigma^2)^{\frac{k}{2}} \int_0^{+\infty} u^{\frac{k}{2}-1}e^{-u}\mathrm{d}t \\
&= (2\sigma^2)^{\frac{k}{2}} k\Gamma(k/2).
\end{aligned}$$

The first statement follows from the fact that $\Gamma(1/2) = \sqrt{\pi}$. Next, for $k \ge 2$, using the fact that $\Gamma(x) \le x^x$ and $x^{1/x} \le e^{1/e}$ for all $x > 0$, we obtain

$$((2\sigma^2)^{\frac{k}{2}} k\Gamma(k/2))^{\frac{1}{k}} = \sigma\sqrt{2}k^{\frac{1}{k}}\Gamma(k/2)^{\frac{1}{k}} \le \sigma\sqrt{2}e^{\frac{1}{e}}\sqrt{k/2} = \sigma e^{\frac{1}{e}}\sqrt{k},$$

which completes the proof. $\square$

3 Maximal inequalities

In this section, we are interested in describing the behavior of quantities of the type

$$\sup_{v \in \mathcal{K}} v^\top \boldsymbol{\xi}, \tag{1}$$

where $\boldsymbol{\xi} \in \mathrm{SG}_d(\sigma^2)$ and $\mathcal{K}$ is a bounded set in $\mathbf{R}^d$. These quantities appear naturally in the analysis of constrained least squares. In practice, the quantity (1) will be of particular importance when $\mathcal{K}$ is a convex body, *i.e.* a convex and compact set with non-empty interior. The first paragraph below studies first the case of a finite set. This simple case will be very useful for moving towards more general situations where the set $\mathcal{K}$, of more complex nature, can be approximated by a finite set, in some sense.

3.1 Finite maximum

We start with a simple but important result.

Theorem 3.1. *Let $\xi_1, \dots, \xi_d$ be centered and sub-gaussian random variables with variance proxy $\sigma^2 > 0$. Then the following statements hold.*

(1) *We have*

$$\mathbf{E} \max_{1 \le i \le d} \xi_i \le \sigma \sqrt{2 \log d}.$$

(2) *For all $\delta \in (0, 1)$,*

$$\max_{1 \le i \le d} \xi_i \le \sigma \sqrt{2 \log(d/\delta)},$$

with probability at least $1 - \delta$.

Proof of Theorem 3.1. (1) For any $\lambda \neq 0$, using the convexity of $t \mapsto \exp(\lambda t)$, we deduce from Jensen's inequality that

$$\exp(\lambda \mathbf{E} \max_i \xi_i) \le \mathbf{E} \exp \lambda \max_i \xi_i.$$

Taking the log on both sides of the previous inequality, and provided $\lambda > 0$, we deduce from the continuity of exp that

$$\begin{aligned} \mathbf{E} \max_i \xi_i &\le \frac{1}{\lambda} \log \mathbf{E} \exp \lambda \max_i \xi_i \\ &= \frac{1}{\lambda} \log \mathbf{E} \max_i \exp \lambda \xi_i \\ &\le \frac{1}{\lambda} \log \sum_i \mathbf{E} \exp \lambda \xi_i. \end{aligned}$$

Using the sub-exponential property on the right hand-side of the last inequality, it follows that

$$\begin{aligned} \mathbf{E} \max_i \xi_i &\le \frac{1}{\lambda} \log \sum_i \exp\left(\frac{\lambda^2 \sigma^2}{2}\right) \\ &= \frac{\log d}{\lambda} + \frac{\lambda^2 \sigma^2}{2}. \end{aligned}$$

Optimizing this upper bound in $\lambda > 0$ leads to the desired result.

(2) For the second statement, observe that, for all $t > 0$ and all $\lambda > 0$,

Markov's inequality easily implies that

$$\begin{aligned}
\mathbf{P}(\max_i \xi_i > t) &= \mathbf{P}(\exp(\lambda \max_i \xi_i) > \exp(\lambda t)) \\
&\le \exp(-\lambda t)\mathbf{E}\exp(\lambda \max_i \xi_i) \\
&= \exp(-\lambda t)\mathbf{E}\max_i \exp(\lambda \xi_i) \\
&\le \exp(-\lambda t)\sum_i \mathbf{E}\exp(\lambda \xi_i).
\end{aligned}$$

Hence, using the sub-gaussian assumption, it follows that

$$\mathbf{P}(\max_i \xi_i > t) \le d\exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2}\right).$$

As in the first part of the proof, optimizing in $\lambda > 0$ yields

$$\mathbf{P}(\max_i \xi_i > t) \le d\exp\left(-\frac{t^2}{2\sigma^2}\right),$$

which, after solving $\delta = d\exp(-t^2/2\sigma^2)$, leads to the desired result. □

Slightly reformulated, the above result provides upper bounds in expectation and in probability for the quantity (1) when $\mathcal{K}$ is any finite set of non zero vectors in $\mathbf{R}^d$ and $\boldsymbol{\xi} \in \mathrm{SG}_d(\sigma^2)$.

Corollary 3.1. *Let $\boldsymbol{\xi} \in \mathrm{SG}_d(\sigma^2)$ and $\mathcal{K}$ be any finite set in $\mathbf{R}^d$. Then the following statements hold.*

(1) *We have*

$$\mathbf{E}\max_{u\in\mathcal{K}}(u^\top \boldsymbol{\xi}) \le \sigma \max_{u\in\mathcal{K}} \|u\|_2 \sqrt{2\log|\mathcal{K}|},$$

where $|\mathcal{K}|$ stands for the cardinality of $\mathcal{K}$.

(2) *For all $\delta \in (0,1)$,*

$$\max_{u\in\mathcal{K}}(u^\top \boldsymbol{\xi}) \le \sigma \max_{u\in\mathcal{K}} \|u\|_2 \sqrt{2\log(|\mathcal{K}|/\delta)},$$

with probability at least $1-\delta$.

Proof of Theorem 3.1. By definition of a sub-gaussian vector, and for each $u \in \mathcal{K}$, variable $u^\top \boldsymbol{\xi}/\|u\|_2$ is sub-gaussian with variance proxy σ^2. The results then follow immediately from Theorem 3.1. □

3.2 Maximum over a convex polytope

Next, we consider a more general case where $\mathcal{K}$ is a convex polytope. Recall that for any subset $A \subset \mathbf{R}^d$, its convex hull $\operatorname{conv}(A)$ is defined by

$$\operatorname{conv}(A) = \bigcup_{q=1}^{+\infty} \operatorname{conv}_q(A),$$

where

$$\operatorname{conv}_q(A) = \left\{ \sum_{j=1}^{q} \lambda_j a_j : a_j \in A, \lambda_j \geq 0, \sum_{j=1}^{q} \lambda_j = 1 \right\}.$$

Note in particular that $\operatorname{conv}(A) = \operatorname{conv}_q(A)$ if A has at most q elements.

Definition 3.1. *A convex polytope $\mathcal{K} \subset \mathbf{R}^d$, of order $q \geq 2$, and with vertices $\mathcal{V}(\mathcal{K}) = \{v_1, \ldots, v_q\}$ is a set of the form*

$$\mathcal{K} = \operatorname{conv}_q(v_1, \ldots, v_q),$$

such that, for all $s < q$,

$$\mathcal{K} \neq \operatorname{conv}_s(v_1, \ldots, v_s).$$

From brevity, we have denoted $\operatorname{conv}_q(v_1, \ldots, v_q) = \operatorname{conv}_q(\{v_1, \ldots, v_q\})$. As demonstrated in the next result, convex functions behave nicely on a convex polytope.

Theorem 3.2. *Let $\mathcal{K}$ be a convex polytope in $\mathbf{R}^d$ and $f : \mathcal{K} \to \mathbf{R}$ be a convex function. Then,*

$$\max_{x \in \mathcal{K}} f(x) = \max_{x \in \mathcal{V}(\mathcal{K})} f(x).$$

Proof of Theorem 3.2. First, since $\mathcal{V}(\mathcal{K}) \subset \mathcal{K}$, we obviously have

$$\max_{x \in \mathcal{V}(\mathcal{K})} f(x) \leq \max_{x \in \mathcal{K}} f(x).$$

Conversely, denote $\{v_1, \ldots, v_q\} = \mathcal{V}(\mathcal{K})$ the vertices of $\mathcal{K}$. By definition, any $x \in \mathcal{K}$ can be written as a convex combination $x = \sum_{i=1}^{q} \lambda_i v_i$. Therefore,

$$f(x) \leq \sum_{i=1}^{q} \lambda_i f(v_i) \leq \max_{1 \leq j \leq q} f(v_j) \sum_{i=1}^{q} \lambda_i = \max_{v \in \mathcal{V}(\mathcal{K})} f(v),$$

where the first inequality follows by convexity of f. The upper bound holds for all $x \in \mathcal{K}$ and therefore implies the opposite inequality. $\square$

The combination of Theorems 3.1 and 3.2 implies the following result. The proof is left as an exercise.

Corollary 3.2. *Let $\mathcal{K} = \mathrm{conv}_q(v_1, \dots, v_q)$ be a convex polytope of order $q \geq 2$ in $\mathbf{R}^d$. Then, for any $\boldsymbol{\xi} \in \mathrm{SG}_d(\sigma^2)$, the following statements hold.*

(1) *We have*

$$\mathbf{E} \max_{u \in \mathcal{K}} (u^\top \boldsymbol{\xi}) \leq \sigma \max_{1 \leq j \leq q} \|v_j\|_2 \sqrt{2 \log q}.$$

(2) *For all $\delta \in (0, 1)$,*

$$\max_{u \in \mathcal{K}} (u^\top \boldsymbol{\xi}) \leq \sigma \max_{1 \leq j \leq q} \|v_j\|_2 \sqrt{2 \log(q/\delta)},$$

with probability at least $1 - \delta$.

We attract the attention of the reader on the fact that, apriori, the previous upper bounds do not depend on the dimension d of the sub-gaussian vector $\boldsymbol{\xi}$ but only on the order q of the polytope. However, the reader should notice that in $\mathbf{R}^d$, a convex polytope with at most $d-1$ vertices is obviously included in the linear span of the vertices, which is a strict linear subspace, and therefore has Lebesgue measure 0. In particular, if the distribution of $\boldsymbol{\xi}$ has a density with respect to the Lebesgue measure in $\mathbf{R}^d$, then

$$\max_{u \in \mathcal{K}} (u^\top \boldsymbol{\xi}) = 0, \quad \text{a.s.}$$

In the context of this course, convex polytopes of central importance are balls for the $\|.\|_1$-norm, *i.e.*

$$\mathcal{K} = \{x \in \mathbf{R}^d : \|x\|_1 \leq \lambda\},$$

for $\lambda > 0$. These sets are, in $\mathbf{R}^d$, convex polytopes of order $2d$. The previous Corollary therefore indicates that, for example,

$$\mathbf{E} \max_{\|u\|_1 \leq \lambda} (u^\top \boldsymbol{\xi}) \leq \sigma \lambda \sqrt{2 \log(2d)},$$

provided $\boldsymbol{\xi} \in \mathrm{SG}_d(\sigma^2)$.

3.3 The ε-net argument

This paragraph is devoted to the case where $\mathcal{K}$ is a Euclidean ball. Here, the convexity of $\mathcal{K}$ will play no role in the mathematical analysis. In the context of convex polytopes, convexity was a means to reduce the maximum over a complex set to a maximum over a finite number of points (the vertices of the polytope). Here, the approach relies on the fact that, as any bounded set in finite dimensions, Euclidean balls can be approximated by a finite number of points (in a different sense than polytopes).

Definition 3.2. *Let $\mathcal{K} \subset \mathbf{R}^d$ be a bounded set and let $\varepsilon > 0$. Let $\|.\|$ be an arbitrary norm in $\mathbf{R}^d$. An ε-net of $\mathcal{K}$, in $(\mathbf{R}^d, \|.\|)$, is a finite set of points $x_1, \ldots, x_N \in \mathcal{K}$ such that*

$$\mathcal{K} \subset \bigcup_{j=1}^{N} B(x_j, \varepsilon),$$

where $B(x, \varepsilon) = \{u \in \mathbf{R}^d : \|x - u\| < \varepsilon\}$. The minimal cardinality of an ε-net of $\mathcal{K}$ is called the ε-covering number of $\mathcal{K}$ and denoted $\mathcal{N}(\mathcal{K}, \varepsilon)$.

Even though norms are equivalent in $\mathbf{R}^d$, the definition of $\mathcal{N}(\mathcal{K}, \varepsilon)$ depends on the choice of the norm $\|.\|$. In the context of the norm $\|.\|_q$, for $1 \leq q \leq +\infty$, we denote

$$\mathcal{N}(\mathcal{K}, \varepsilon) = \mathcal{N}_q(\mathcal{K}, \varepsilon).$$

For convenience, the previous definition requires the points in the net to be elements of $\mathcal{K}$. In this case, a net is called a *proper* net in contrast to an *improper* net, the points of which are not required to be in $\mathcal{K}$. The notions of covering numbers, based on proper or improper nets, differ very mildly and their differences will be investigated in the seminar sessions. Next, we give upper and lower bounds for the ε-covering number of Euclidean balls in $\mathbf{R}^d$.

Theorem 3.3. *Let $\bar{B}_2(x, r)$ be a closed Euclidean ball of center $x \in \mathbf{R}^d$ and radius $r > 0$, i.e.*

$$\bar{B}_2(x, r) = \{u \in \mathbf{R}^d : \|u - x\|_2 \leq r\}.$$

Then, for any $0 < \varepsilon \leq r$, we have

$$\left(\frac{r}{\varepsilon}\right)^d \leq \mathcal{N}_2(\bar{B}_2(x, r), \varepsilon) \leq \left(\frac{3r}{\varepsilon}\right)^d.$$

Proof of Theorem (3.3). For brevity, we denote below $\bar{B}(x, r) = \bar{B}_2(x, r)$ and $B(x, r) = B_2(x, r)$. The proof is based on simple volume arguments. Recall in particular that the volume (*i.e.* the Lebesgue measure) of a (closed or open) Euclidean ball of radius $r > 0$ in $\mathbf{R}^d$ is $c_d r^d$ where

$$c_d = \frac{\pi^{d/2}}{\Gamma(1 + \frac{d}{2})}. \tag{2}$$

We first address the lower bound. Consider an ε-net $x_1, \ldots, x_N$ of $\bar{B}(x, r)$. Since the open balls with radius $\varepsilon > 0$ and centered in the x_i's cover $\bar{B}(x, r)$, we easily deduce that

$$c_d r^d = \mathrm{vol}(\bar{B}(x, r)) \leq \sum_{j=1}^{N} \mathrm{vol}(B(x_j, \varepsilon)) = c_d N \varepsilon^d, \tag{3}$$

from which it follows that

$$\left(\frac{r}{\varepsilon}\right)^d \le N.$$

Since this lower bound holds for any ε-net $x_1, \dots, x_N$ of $\bar{B}(x, r)$, it holds in particular for a minimal ε-net which completes the proof of the lower bound. Let's now address the upper bound. Introduce a maximal set $x_1, \dots, x_M$ of points in $\bar{B}(x, r)$ such that,

$$\forall i \neq j, \quad \|x_i - x_j\|_2 > \varepsilon. \tag{4}$$

Here, the maximality is understood in the sense that, for any other point $y \in \bar{B}(x, r)$, the set $\{y, x_1, \dots, x_M\}$ would not anymore satisfy property (4). Observe that, necessarily, the points $x_1, \dots, x_M$ form an ε-net of $\bar{B}(x, r)$. Indeed, if there were a point, say $y \in \bar{B}(x, r)$, not covered by the reunion of the open balls $B(x_j, \varepsilon)$, $j \in \{1, \dots, M\}$, then this would contradict the maximality of $\{x_1, \dots, x_M\}$. In particular, we deduce that $\mathcal{N}_2(\bar{B}(x, r), \varepsilon) \le M$. Next, observe that the open balls $B(x_j, \varepsilon/2)$, $j \in \{1, \dots, M\}$, are pairwise disjoint by construction and that,

$$\bigcup_{j=1}^{M} B\left(x_j, \frac{\varepsilon}{2}\right) \subset \bar{B}(x, r) + B\left(0, \frac{\varepsilon}{2}\right) = B\left(x, r + \frac{\varepsilon}{2}\right).$$

Hence, it follows that

$$c_d M \left(\frac{\varepsilon}{2}\right)^d \le c_d \left(r + \frac{\varepsilon}{2}\right)^d,$$

from which we deduce that

$$\mathcal{N}_2(\bar{B}(x, r), \varepsilon) \le M \le \left(\frac{2r + \varepsilon}{\varepsilon}\right)^d \le \left(\frac{3r}{\varepsilon}\right)^d.$$

The proof is complete. □

Theorem 3.4. *Let $\bar{B}_2 = \bar{B}_2(0, 1)$ and let $\boldsymbol{\xi} \in \mathrm{SG}_d(\sigma^2)$. Then, the following statements hold.*

(1) *We have*

$$\mathbf{E} \max_{u \in \bar{B}_2} (u^\top \boldsymbol{\xi}) \le 4\sigma\sqrt{d}.$$

(2) *For all $\delta \in (0, 1)$,*

$$\max_{u \in \bar{B}_2} (u^\top \boldsymbol{\xi}) \le 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)},$$

with probability at least $1 - \delta$.

Proof of Theorem 3.4. Let $\mathcal{N} = \{x_1, \ldots, x_N\}$ be a minimal 1/2-net of $\bar{B}_2$. Observe that, for any $x \in \bar{B}_2$, there exists a point x_i in the net such that $\|x - x_i\| \leq 1/2$. In particular, we deduce that

$$\begin{aligned} x^\top \boldsymbol{\xi} &= (x - x_i)^\top \boldsymbol{\xi} + x_i^\top \boldsymbol{\xi} \\ &\leq \sup_{u \in \bar{B}_2(0,1/2)} u^\top \boldsymbol{\xi} + \sup_{u \in \mathcal{N}} u^\top \boldsymbol{\xi} \\ &= \frac{1}{2} \sup_{u \in \bar{B}_2} u^\top \boldsymbol{\xi} + \sup_{u \in \mathcal{N}} u^\top \boldsymbol{\xi}. \end{aligned}$$

Taking the supremum over $x \in \bar{B}_2$ on the left hand-side, we therefore deduce that

$$\sup_{x \in \bar{B}_2} x^\top \boldsymbol{\xi} \leq \frac{1}{2} \sup_{u \in \bar{B}_2} u^\top \boldsymbol{\xi} + \sup_{u \in \mathcal{N}} u^\top \boldsymbol{\xi},$$

from which it follows that

$$\sup_{x \in \bar{B}_2} x^\top \boldsymbol{\xi} \leq 2 \sup_{x \in \mathcal{N}} x^\top \boldsymbol{\xi}.$$

As a result, we deduce from Corollary 3.1 and Theorem 3.3 that

$$\mathbf{E} \sup_{x \in \bar{B}_2} (x^\top \boldsymbol{\xi}) \leq 2\sigma \sqrt{2d \log 6},$$

and the first result follows from the observation that $2\sqrt{2 \log 6} \leq 4$. The second point of the theorem is in the same spirit. $\square$

4 Exercises

Exercise 4.1. Examples of sub-gaussian random variables.

1. *Show that if $\xi \sim \mathcal{N}(0, \sigma^2)$ then $\xi \in \mathrm{SG}(\sigma^2)$.*

2. *Let ξ and ζ be two real-valued and centered random variables. Suppose that their distributions have densities, denoted respectively f and g. Suppose finally that $\zeta \in \mathrm{SG}(\sigma^2)$ and that for some $C > 0$, $f(x) \leq g(x)$ for $|x| \geq C$. Show that ξ is also sub-gaussian and upper bound its variance proxy.*

3. (**Hoeffding's lemma**). *Suppose that, for some $a < b$, the random variable ξ is centered and takes its values in the interval $[a, b]$ with probability one. Then, show that*

$$\log \mathbf{E} \exp(\lambda \xi) \leq \frac{\lambda^2 (b-a)^2}{8},$$

i.e. that $\xi \in \mathrm{SG}((b-a)^2/4)$. Compare (here) the variance proxy to the true variance of ξ.

Exercise 4.2. Sub-gaussian vectors and matrices.

1. *Let* $\boldsymbol{\xi} \in \mathrm{SG}_n(\sigma^2)$ *be an* n*-dimensional sub-gaussian vector with variance proxy* $\sigma^2 > 0$*. Observe that each coordinate* ξ_i *of* $\boldsymbol{\xi}$ *is in* $\mathrm{SG}(\sigma^2)$ *and, more generally, that for any* $u \in \mathbf{R}^n$, $u^\top \boldsymbol{\xi} \in \mathrm{SG}(\sigma^2 \|u\|_2^2)$.

2. *Let* $\xi_1, \ldots, \xi_n$ *be independent sub-gaussian variables with variance proxy* $\sigma^2 > 0$*. Show that* $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^\top \in \mathrm{SG}_n(\sigma^2)$*. Is it true if the variables* $\xi_1, \ldots, \xi_n$ *are not independent ?*

3. *Let* $\boldsymbol{A} \in \mathrm{SG}_{p,q}(\sigma^2)$ *be a* $p \times q$ *sub-gaussian matrix with variance proxy* $\sigma^2 > 0$*. Show that, for any* $(u, v) \in \mathbf{R}^p \times \mathbf{R}^q$, $u^\top \boldsymbol{A} v \in \mathrm{SG}(\sigma^2 \|u\|_2^2 \|v\|_2^2)$.

4. *Let* $\{a_{i,j} : 1 \le i \le p, 1 \le j \le q\}$ *be independent sub-gaussian variables with variance proxy* $\sigma^2 > 0$*. Show that the matrix* $\boldsymbol{A} = (a_{i,j}) \in \mathrm{SG}_{p,q}(\sigma^2)$*. Is it true if the variables* $a_{i,j}$ *are not independent ?*

Exercise 4.3. Chernoff's bound.

1. *Let* ξ *be a centered and real-valued random variable. For all* $\lambda \in \mathbf{R}$*, define its log-Laplace transform* $\psi : \mathbf{R} \to \mathbf{R}$ *by* $\psi(\lambda) = \log \mathbf{E} \exp(\lambda \xi)$ *and its dual* $\psi^\star : \mathbf{R} \to \mathbf{R}$ *by*

$$\psi^\star(t) = \sup_{\lambda \ge 0} \{\lambda t - \psi(\lambda)\} .$$

Prove that, for all $t \ge 0$,

$$\mathbf{P}(\xi \ge t) \le e^{-\psi^\star(t)}.$$

2. *Suppose that* $\xi \in \mathrm{SG}(\sigma^2)$*. Deduce from the previous question that, for all* $t \ge 0$,

$$\mathbf{P}(\xi \ge t) \le e^{-\frac{t}{2\sigma^2}} \quad \text{and} \quad \mathbf{P}(\xi \le -t) \le e^{-\frac{t}{2\sigma^2}}.$$

3. *Conclude in particular that*

$$\mathbf{P}(|\xi| \ge t) \le 2e^{-\frac{t}{2\sigma^2}}.$$

References

P. Rigollet. High-dimensional statistics. MIT lecture notes, 2015.

R. van Handel. Probability in high dimension. Princeton lecture notes, 2016.