1. Introduction

Let $(X_i : i \in \mathbb{N})$ be a sequence of independent and identically distributed (i.i.d.) real-valued random variables. We assume that $\mathbb{E}X_1$ exists. In this section, we will be interested in the behaviour of the empirical mean $\bar{X}_n := n^{-1}\sum_{i=1}^n X_i$ w.r.t. the actual mean $\mathbb{E}X_1$. We will call this behaviour *concentration of $\bar{X}_n$ around its mean.*

The first "level" of concentration is given by the *Law of Large Number* (LLN) saying that, almost surely,

$$\bar{X}_n \longrightarrow \mathbb{E}X_1 \text{ when } n \to \infty. \tag{0.1}$$

Note that this result is very sharp, in the sens that, if $\bar{X}_n$ converges a.s. then $\mathbb{E}X_1$ is finite and $\bar{X}_n$ converges a.s. to $\mathbb{E}X_1$.

For the second "level" of concentration, we need to assume that $\mathbb{E}X_1^2$ exists. Then, the behaviour of $\bar{X}_n - EX_1$ is provided by the Central Limit Theorem (CLT) which says that, after renormalizing, $\sqrt{n}(\bar{X}_n - \mathbb{E}X_1)$ behaves asymptotically like a Gaussian variable:

$$\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \rightsquigarrow \sqrt{\mathbb{V}(X_1)}G \text{ when } n \to \infty, \tag{0.2}$$

where $\mathbb{V}(X_1)$ is the variance of X_1, $G \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable and the symbol $\rightsquigarrow$ stands for the *convergence in probability.*

The two first level of concentration are asymptotic (n has to tend to infinity to make these results relevant). In general, we have at hand only a finite number of random variables. Nevertheless, we still want to know how behaves $\bar{X}_n$ around its mean. We thus need non-asymptotic concentration results (results which hold for any n). The first result going into that direction is the classical Berry-Essen's theorem (cf.[15]), which yields

$$\sup_{t\in\mathbb{R}}\Big|\mathbb{P}\big[\sqrt{n}(\bar{X}_n - \mathbb{E}X_1) \geq t\big] - \mathbb{P}\big[\sqrt{\mathbb{V}(X_1)}G \geq t\big]\Big| \leq \frac{c\mathbb{E}|X_1|^3}{\sqrt{n}}, \forall n \in \mathbb{N}, \tag{0.3}$$

provided that $\mathbb{E}|X_1|^3 < \infty$. Other Berry-Essen type of results can be found in the Section 7.

Berry-Essen theorem provides a non-asymptotic concentration result for the empirical mean under the only assumption that a third moment exists. It leads to the first *concentration's inequality* of the type, for every $t > 0$,

$$\mathbb{P}\big[\bar{X}_n - \mathbb{E}X_1 \geq t\big] \leq \mathbb{P}\big[\sqrt{\mathbb{V}(X_1)}G \geq \sqrt{n}t\big] + \frac{c\mathbb{E}|X_1|^3}{\sqrt{n}}. \tag{0.4}$$

Concentration's inequalities of the form of Equation (0.4) are one of the main tools of empirical processes theory. Nevertheless, the one obtained here, using Berry-Essen theorem is too weak to be interesting. Indeed, the upper bound of Equation (0.4) behaves like the second term which is of the order of $n^{-1/2}$. Berry-Essen's theorem is too general (only a third moment is needed) to be applicable in many other setup where we have more than a finite third moment. In some cases (cf. Exercise 0.4), we can even have an exponential decrease (in function of n and t) of the tail $\mathbb{P}\big[\bar{X}_n - \mathbb{E}X_1 \geq t\big]$, which then, behaves like the term $\mathbb{P}\big[\sqrt{\mathbb{V}(X_1)}G \geq \sqrt{n}t\big]$. In the following section, we will explore some special cases (finite Orlicz's norm, self-bounded, conditions on the moment, assumption on the Legendre's transform,..) which can lead to such exponential decay of the tail behaviour of $\bar{X}_n - \mathbb{E}X_1$.

2. Orlicz spaces

Definition 0.1 (cf. [4, 18]). *A* ***Young-Orlicz modulus*** *is a convex increasing function ψ from $[0,\infty)$ onto $[0,\infty)$ (in particular $\psi(0) = 0$ and $\psi(x) \to \infty$ when $x \to \infty$). Let $(\mathcal{X}, \tau, \mu)$ be a measure space and ψ be a Young-Orlicz modulus. Denote by $\mathcal{L}_\psi(\mathcal{X}, \tau, \mu)$ the space of all real-valued measurable functions f onto $\mathcal{X}$ such that*

$$\|f\|_\psi := \inf\big(c > 0 : \mathbb{E}_\mu\psi\big(|f|/c\big) \leq 1\big) < \infty.$$

Define $L_\psi(\mathcal{X},\tau,\mu)$ *to be the set of all the equivalence classes of functions in* $\mathcal{L}_\psi(\mathcal{X},\tau,\mu)$ *for the almost everywhere equality w.r.t.* μ*.* $L_\psi(\mathcal{X},\tau,\mu)$ *is called an* ***Orlicz space****.*

Classical examples of Young-Orlicz moduli are

$$\phi_p(x) := x^p, p \geq 1 \text{ and } \psi_\alpha(x) := \exp(x^\alpha) - 1, \alpha \geq 1, \tag{0.5}$$

we can also define Young-Orlicz modulus ψ_α for $0 < \alpha < 1$ by $\psi_\alpha(x) = \exp(x^\alpha) - 1$ for $x \geq x_\alpha$ large enough and take ψ_α to be linear on $[0, x_\alpha]$.

Theorem 0.1 ([4])**.** *For any space* $(\mathcal{X},\tau,\mu)$ *and any Young-Orlicz modulus* ψ*, the space* $L_\psi(\mathcal{X},\tau,\mu)$ *is a Banach space.*

Let ψ be a convex function. We define the convex conjugate of ψ by

$$\phi(y) := \sup\big(xy - \psi(x) : x > 0\big), \forall y > 0.$$

Note that the convex conjugate of a Young-Orlicz modulus is also a Young-Orlicz modulus.

Theorem 0.2. *Let* ψ *be a Young-Orlicz modulus and* ϕ *be its convex conjugate. Let* $f \in \mathcal{L}_\psi(\mathcal{X},\tau,\mu)$ *and* $g \in \mathcal{L}_\phi(\mathcal{X},\tau,\mu)$ *then* $fg \in \mathcal{L}_1(\mathcal{X},\tau,\mu)$ *and*

$$\mathbb{E}_\mu|fg| \leq 2\left\|f\right\|_\phi \left\|g\right\|_\psi.$$

Proof. By homogeneity, we can assume $\|f\|_\psi = \|g\|_\phi = 1$. By definition of the convex conjugate, we have, for any $x \in \mathcal{X}$,

$$|f(x)g(x)| \leq \psi(|f(x)|) + \phi(|g(x)|).$$

Taking the expectation and using that $\mathbb{E}\psi(|f(x)|), \mathbb{E}\phi(|g(x)|) \leq 1$ leads to the result. ■

In particular, ϕ_p is an Orlicz modulus with convex conjugate $c_p\phi_q$, where $p^{-1} + q^{-1} = 1$. In this case, Theorem 0.2 is, up to a multiplying constant, Hölder's inequality.

For our concentration purpose of this section, we will restrict ourselves to the study of the $L_\psi(\Omega,\sigma,\mathbb{P})$ Orlicz spaces. We will denote these spaces by L_ψ. These spaces are set of equivalence classes of random variables defined on an abstract probability space $(\Omega,\sigma,\mathbb{P})$. For instance, given is a real-valued random variable X and $\alpha \geq 1$, the ψ_α-norm of X is defined by

$$\|X\|_{\psi_\alpha} := \inf\big(c > 0 : \mathbb{E}\exp\big(|X|^\alpha / c^\alpha\big) \leq 2\big).$$

We first start with a maximal inequality.

Proposition 0.1 ([18])**.** *There exists an absolute constant* c_0 *such that the following holds. Let* ψ *be Young-Orlicz modulus such that there exists an absolute constant* $c > 0$ *such that*

$$\limsup_{x,y\to\infty} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty.$$

Then, for any real-valued random variables $X_1, \ldots, X_n$*,*

$$\left\|\max_{1\leq i\leq n} X_i\right\|_\psi \leq c_0 \psi^{-1}(n) \max_{1\leq i\leq n} \|X_i\|_\psi$$

Orlicz-spaces L_ψ are very useful to characterize the tail behavior of random variables. For instance, we say that X has a *sub-gaussian behavior* when $\|X\|_{\psi_2} < \infty$, we say that X has a *sub-exponential behavior* when $\|X\|_{\psi_1} < \infty$ and in general, we say that X has a ψ_α *behavior* when $\|X\|_{\psi_\alpha} < \infty$. It is easy to get $L_\infty \subset L_{\psi_{\alpha'}} \subset L_{\psi_\alpha} \subset L_2$ when $\alpha' \leq \alpha$.

Proposition 0.2. *Let* X *be a real-valued random variable and* $\alpha \geq 1$*. All the following points are equivalent:*

(1) $\exists K_1 > 0 : \|X\|_{\psi_\alpha} \leq K_1$*;*
(2) $\exists K_2 > 0 : \mathbb{E}\exp\big(|X|^\alpha / K_2^\alpha\big) \leq 2$*;*
(3) $\exists K_3 > 0 : \big[\mathbb{E}|X|^p\big]^{1/p} \leq K_3 p^{1/\alpha}, \forall p \in \mathbb{N}$*;*

(4) $\exists K_4, K_4' > 0 : \mathbb{P}\big[|X| \geq t\big] \leq K_4' \exp\big(-t^\alpha/K_4^\alpha\big), \forall t > 0;$
(5) there exists $c > 0$ *such that* $\exists K_5, K_5' > 0 : \mathbb{P}\big[|X| \geq t\big] \leq K_5' \exp\big(-t^\alpha/K_5^\alpha\big), \forall t \geq c;$
(6) $\exists K_6, K_6' > 0, \forall \lambda > 0, \mathbb{E}\exp\big(\lambda|X|\big) \leq K_6' \exp\big((\lambda K_6)^{\alpha/(\alpha-1)}\big).$

Moreover, all the constants $K_1, \dots, K_6$ *are proportional, up to some multiplying constants which depend only on* α*.*

Proof.(1) $\iff$ (2): Assume (1). By monotone convergence, it is easy to check that $\mathbb{E}\exp\big(|X|^\alpha/K_1^\alpha\big) \leq 2$. Point (2) follows easily. By definition of the Orlicz norm, (2) implies (1).

(4) implies (3): Let $p \in \mathbb{N}$.

$$\begin{aligned}\mathbb{E}|X|^p &= \int_0^\infty pt^{p-1}\mathbb{P}\big[|X| \geq t\big]dt \leq 2p\int_0^\infty t^{p-1}\exp\big(-t^\alpha/K_4^\alpha\big)dt \\ &= \frac{2pK_4^{p-1}}{\alpha}\Gamma(p/\alpha) \leq \frac{2pK_4^{p-1}}{\alpha}(p/\alpha)^{p/\alpha-1},\end{aligned}$$

where $\Gamma(u) = \int_0^\infty t^{u-1}\exp(-t)dt$ satisfies $\Gamma(u+1) = u\Gamma(u)$. Thus, $\big(\mathbb{E}|X|^p\big)^{1/p} \leq (2K_4)\alpha^{-1/\alpha}p^{1/\alpha}$.

(3) implies (1): Let $c > 0$.

$$\mathbb{E}\exp\big(|X|^\alpha/c^\alpha\big) = \sum_{k\geq 0}\frac{1}{k!c^{\alpha k}}\mathbb{E}\big[|X|^{\alpha k}\big] \leq \sum_{k\geq 0}\frac{1}{k!}\Big(\frac{K_3^\alpha}{c^\alpha}\alpha k\Big)^k.$$

Now, using the Stirling's formula it is easy to get $\frac{1}{k!}\Big(\frac{K_3^\alpha}{c^\alpha}\alpha k\Big)^k \sim \frac{1}{\sqrt{2\pi k}}\Big(\frac{\alpha e K_3^\alpha}{c^\alpha}\Big)^k$. Thus, there exists an absolute constant c_0 such that for $c \geq c_0\alpha^{1/\alpha}K_3$, we have $\mathbb{E}\exp\big(|X|^\alpha/c^\alpha\big) \leq 2$. This implies that $\|X\|_{\psi_\alpha} \leq c_0\alpha^{1/\alpha}K_3$.

(1) implies (4): Let $t > 0$. By Markov's inequality:

$$\begin{aligned}\mathbb{P}\big[|X| \geq t\big] &= \mathbb{P}\Big[\exp\big(|X|^\alpha/K_1^\alpha\big) > \exp\big(t^\alpha/K_1^\alpha\big)\Big] \\ &\leq \mathbb{E}\Big[\exp\big(|X|^\alpha/K_1^\alpha\big)\Big]\exp\big(-t^\alpha/K_1^\alpha\big) \leq 2\exp\Big(-t^\alpha/K_1\alpha\Big).\end{aligned}$$

(4) $\iff$ (5): (4) implies (5) is trivial. Assume (5), for $K_4 \geq (c(\log 2)^{1/\alpha}) \wedge K_5$, (4) is satisfied.

(3) implies (6): Let $\lambda > 0$. We have

$$\mathbb{E}\exp\big(\lambda|X|\big) \leq \sum_{k\geq 0}(\lambda K_3)^k\frac{k^{k/\alpha}}{k!}. \tag{0.6}$$

Now, we use the following approximation

$$k! = \sqrt{2\pi k}\big(k/e\big)^k\exp(\lambda_k) \text{ where } \frac{1}{12k+1} < \lambda_k < \frac{1}{12k} \tag{0.7}$$

in Equation (0.6), to get, for $i_k := \Big\lceil\frac{k(\alpha-1)}{\alpha}\Big\rceil$ (the integer part of $k(\alpha-1)/\alpha$),

$$\mathbb{E}\exp\big(\lambda|X|\big) \leq c_\alpha\sum_{k\geq 0}\frac{\big[(\lambda K_3)^{\alpha/\alpha-1}\big]^{i_k}}{i_k!} \leq c_\alpha'\exp\big((\lambda K_3)^{\alpha/(\alpha-1)}\big).$$

(6) implies (4): Let $t > 0$. For every $\lambda > 0$, by Markov's inequality,

$$\begin{aligned}\mathbb{P}\big[|X| > t\big] &= \mathbb{P}\big[\exp(\lambda|X|) > \exp(\lambda t)\big] \leq \exp(-\lambda t)\mathbb{E}\exp\big(\lambda|X|\big) \\ &\leq K_6'\exp\big((\lambda K_6)^{\alpha/(\alpha-1)} - \lambda t\big).\end{aligned}$$

We obtain (4) by optimizing λ in the last inequality. ■

In particular, moments and concentration properties of a random variable are closely related:

Corollary 0.1. *Let $\alpha \geq 1$ and $X \in L_{\psi_\alpha}$, we have*

$$\|X\|_{\psi_\alpha} \sim \sup_{p\geq 1} \frac{\|X\|_p}{p^{1/\alpha}}.$$

Now, we take a look of sums of ψ_α random variables.

Proposition 0.3. *There exists an absolute constant $c > 0$ such that the following holds. Let $X, X_1, \ldots, X_n$ be i.i.d. mean-zero random variables. Then, for every $a = (a_1, \ldots, a_n)^t \in \mathbb{R}^n$,*

$$\left\| \sum_{i=1}^n a_i X_i \right\|_{\psi_2} \leq c \|X\|_{\psi_2} \|a\|_2 .$$

Proof.

Lemma 0.1 (cf.[17]). *Let X be a ψ_2 mean zero random variable. Then,*

$$\mathbb{E}\exp(\lambda|X|) \leq \exp\big(8\lambda^2 \|X\|_{\psi_2}^2\big), \forall \lambda > 0.$$

Proof. For every $t > 0$, $\mathbb{P}\big(|X| \geq t\big) \leq 2\exp\big(-t^2/\|X\|_{\psi_2}^2\big)$. Thus, for any integer $k \geq 2$, $\mathbb{E}|X|^k \leq 2\|X\|_{\psi_2}^k \Gamma\big(k/2+1\big)$, where we set $\Gamma(u) := \int_0^\infty t^{u-1}\exp(-u)du$ for any $u \geq 1$. Using the last inequality, the fact that $\mathbb{E}X = 0$, that Γ is a non-decreasing function and that $\forall k \geq 2, \Gamma(k+1) = k\Gamma(k) = k!$, we obtain, for every $\lambda > 0$,

$$\begin{aligned}
\mathbb{E}\exp(\lambda|X|) &= 1 + \sum_{k\geq 2} \frac{\lambda^k \mathbb{E}|X|^k}{k!} \leq 1 + 2\sum_{k\geq 2} \frac{\Gamma(k/2+1)}{\Gamma(k+1)} \big(\lambda \|X\|_{\psi_2}\big)^k \\
&\leq 1 + 2\sum_{k\geq 2} \frac{\big(\lambda\|X\|_{\psi_2}\big)^k}{\Gamma(k/2+1)} = 1 + 2\sum_{k\geq 1} \frac{\big(\lambda\|X\|_{\psi_2}\big)^{2k}}{\Gamma(k+1)} + \frac{\big(\lambda\|X\|_{\psi_2}\big)^{2k+1}}{\Gamma(k+3/2)} \\
&\leq 1 + 2\sum_{k\geq 1} \frac{\big(\lambda^2\|X\|_{\psi_2}^2\big)^k \big(1 + \big(\lambda^2\|X\|_{\psi_2}^2\big)^{1/2}\big)}{\Gamma(k+1)},
\end{aligned}$$

if $\lambda^2\|X\|_{\psi_2}^2 \geq 1$, the sumand is smaller than $2(2\lambda^2\|X\|_{\psi_2}^2)^{k+1}/[(k+1)!]$ otherwise, it is smaller than $2(\lambda^2\|X\|_{\psi_2}^2)^k/k!$. The claim follows by summing. ■

To prove Proposition 0.3, it suffices to upper bound the Legendre transform of $\big|\sum_i a_i X_i\big|$. Indeed, by independence, for every $\lambda > 0$,

$$\mathbb{E}\exp\big(\lambda\sum_i a_i X_i\big) = \prod_i \mathbb{E}\exp(\lambda a_i X_i) \leq \exp\big(8\lambda^2\|X\|_{\psi_2}^2 \|a\|_2^2\big).$$

For the same reason, $\mathbb{E}\exp\big(-\lambda\sum_i a_i X_i\big) \leq \exp\big(8\lambda^2\|X\|_{\psi_2}^2\|a\|_2^2\big)$. Thus,

$$\mathbb{E}\exp\Big(\lambda\Big|\sum_i a_i X_i\Big|\Big) \leq 2\exp\big(8\lambda^2\|X\|_{\psi_2}^2\|a\|_2^2\big).$$

We conclude with Proposition 0.2. ■

Now, we turn on p-th moment of sum of independent ψ_α r.v.. It appears that the behavior of a general centred ψ_α r.v. can be easily reduced to the symmetric Weibull variable. Meaning that, when considering p-th moments of sum of independent r.v. with a given ψ_α tail decay, the worst case scenario is when these r.v. are the symmetric Weibull variables (the one which are exactly ψ_α (and not $\psi_{\alpha+\epsilon}$ for all $\epsilon > 0$). We first recalll the definition of Weibull and symmetric Weibull variables.

A *Weibull random variable with shape parameter* $\alpha > 0$ *and scale parameter* $\lambda > 0$ is a r.v. with probability density function

$$f_{\alpha,\lambda}(x) := \begin{cases} \frac{\alpha}{\lambda}\Big(\frac{x}{\lambda}\Big)^{\alpha-1}\exp\big(-(x/\lambda)^\alpha\big) & \text{if} \quad x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that $\mathbb{P}[X \geq t] = \exp\big(-(t/\lambda)^\alpha\big)$ for all $t \geq 0$ and $\mathbb{P}[X \geq t] = 1$ for all $t \leq 0$. A *symmetric Weibull r.v. with shape* $\alpha > 0$ *and scale* $\lambda > 0$ X is a variable satisfying $\mathbb{P}[|X| > t] = \exp\big(-(t/\lambda)^\alpha\big)$ for all $t \in \mathbb{R}$.

Theorem 0.3. *There exists an absolute constant* $C > 0$ *such that the following holds. Let* $1 \leq \alpha \leq 2$ *and* $X_1, \ldots, X_n$ *be independent mean zero random variables with* $\|X_i\|_{\psi_\alpha} \leq 1, \forall i$. *Then, for every* $a = (a_1, \ldots, a_n)^t \in \mathbb{R}^n$ *and* $p \geq 1$,

$$\Big\|\sum_{i=1}^n a_i X_i\Big\|_p \leq C\big(\sqrt{p}\,\|a\|_2 + p^{1/\alpha}\,\|a\|_{\alpha^*}\big),$$

where $\alpha^{-1} + (\alpha^*)^{-1} = 1$.

Proof. Let $p \geq 1$. We first start with a **symmetrization argument**. For that we introduce a *ghost sample* $X_1', \ldots, X_n'$ (that is n independent r.v., independent of $X_1, \ldots, X_n$ and such that X_i' has the same distribution as X_i for every $i = 1, \ldots, n$) and $\epsilon_1, \ldots, \epsilon_n$ n i.i.d. Rademacher variables independent of $X_1, \ldots, X_n$ and $X_1', \ldots, X_n'$. The symmetrization argument is based on the fact that $(X_i' - X_i)$ and $\epsilon_i(X_i' - X_i)$ have the same distribution. Indeed, for any measurable function f, we have

$$\mathbb{E}f(\epsilon_i(X_i'-X_i)) = \mathbb{E}\mathbb{E}_\epsilon\big[f(\epsilon_i(X_i-X_i'))|X_i, X_i'\big] = \frac{1}{2}\mathbb{E}\big[f(X_i-X_i')+f(X_i-X_i')\big] = \mathbb{E}f(X_i'-X_i).$$

Using this fact, the fact that $X_1, \ldots, X_n$ have mean zero and Jensen's inequality, we get

$$\begin{aligned}
\Big\|\sum_{i=1}^n a_i X_i\Big\|_p &= \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i X_i\Big|^p\Big)^{1/p} = \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i X_i - \mathbb{E}\sum_{i=1}^n a_i X_i'\Big|^p\Big)^{1/p} \\
&= \Big(\mathbb{E}\Big|\mathbb{E}\Big[\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i X_i'\Big|X_1', \ldots, X_n'\Big]\Big|^p\Big)^{1/p} \\
&\leq \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i X_i'\Big|^p\Big)^{1/p} = \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i(X_i - X_i')\Big|^p\Big)^{1/p} \\
&= \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i\epsilon_i(X_i - X_i')\Big|^p\Big)^{1/p} \\
&\leq \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i\epsilon_i X_i\Big|^p\Big)^{1/p} + \Big(\mathbb{E}\Big|\sum_{i=1}^n a_i\epsilon_i X_i'\Big|^p\Big)^{1/p} = 2\Big(\mathbb{E}\Big|\sum_{i=1}^n a_i\epsilon_i X_i\Big|^p\Big)^{1/p} \qquad (0.8)
\end{aligned}$$

Now, we turn to a **contraction principle** (cf.[12] p.95). We use the first version of this principle saying that if (α_i) and (β_i) are two sequences of real numbers such that $|\alpha_i| \leq |\beta_i|, \forall i$ and (ϵ_i) is a sequence of i.i.d. Rademacher variables then, for every $p \geq 1$,

$$\Big\|\sum_{i=1}^n \epsilon_i\alpha_i\Big\|_p \leq \Big\|\sum_{i=1}^n \epsilon_i\beta_i\Big\|_p. \qquad (0.9)$$

Before using the contraction principle, we have to "pre-conditioned" the variables in such a way that we can compare them. Let $\beta := (\log 2)^{1/\alpha}$ and set $U_i := (|X_i| - \beta)_+, \forall i = 1, \ldots, n$.

Let $Y_1, \dots, Y_n$ be n i.i.d. Weibull r.v. with shape α and scale 1. Let $i = 1, \dots, n$. For every $t > 0$,

$$\mathbb{P}[U_i \geq t] = \mathbb{P}[|Z_i| \geq t + \beta] \leq 2\exp(-(t+\beta)^\alpha) \leq 2\exp\big(-t^\alpha - \beta^\alpha\big) = \mathbb{P}[Y_i \geq t].$$

Using the last inequality, we can construct $\tilde{U}_1, \dots, \tilde{U}_n$ independent r.v. and $\tilde{Y}_1, \dots, \tilde{Y}_n$ independent random variables such that for every $i = 1, \dots, n$, $U_i \sim \tilde{U}_i$, $Y_i \sim \tilde{Y}_i$ and $\tilde{U}_i \leq \tilde{Y}_i$. Indeed, for any random variable X we define the cumulative distribution function and its generalized inverse by

$$F_X(t) := \mathbb{P}[X \leq t], \forall t \in \mathbb{R} \text{ and } F_X^{-1}(y) := \inf\big(t \in \mathbb{R} : F_X(t) \geq y\big), \forall y \in [0,1].$$

Now, take $\mathcal{U}_1, \dots, \mathcal{U}_n$ n independent random variable uniformly distributed on $[0,1]$. Then, set for every $i = 1, \dots, n$, $\tilde{U}_i := F_{U_i}^{-1}(\mathcal{U}_i)$ and $\tilde{Y}_i := F_{Y_i}^{-1}(\mathcal{U}_i)$.

We are now in position to apply the contraction principle (conditionally to $\tilde{U}_1, \dots, \tilde{U}_n$ and $\tilde{Y}_1, \dots, \tilde{Y}_n$):

$$\Big(\mathbb{E}_\epsilon \Big|\sum_{i=1}^n a_i \epsilon_i \tilde{U}_i\Big|^p\Big)^{1/p} \leq \Big(\mathbb{E}_\epsilon \Big|\sum_{i=1}^n a_i \epsilon_i \tilde{Y}_i\Big|^p\Big)^{1/p}. \tag{0.10}$$

Taking the last inequality to the power p, then the expectation, and using the fact that $\tilde{U}_i \sim U_i$ and $\tilde{Y}_i \sim Y_i$ for every $i = 1, \dots, n$, we get

$$\Big\|\sum_{i=1}^n a_i \epsilon_i U_i\Big\|_p \leq \Big\|\sum_{i=1}^n a_i \epsilon_i Y_i\Big\|_p . \tag{0.11}$$

Now, we combine the symmetrization argument of (0.8), the contraction principle of (0.11) and the fact that $\epsilon_i X_i$ and $\epsilon_i |X_i|$ have the same probability distribution for every $i = 1, \dots, n$: for every $p \geq 1$,

$$\begin{aligned}
\Big\|\sum_{i=1}^n a_i X_i\Big\|_p &\leq 2\Big\|\sum_{i=1}^n a_i \epsilon_i X_i\Big\|_p \quad \text{(symmetrization)}\\
&= 2\Big\|\sum_{i=1}^n a_i \epsilon_i |X_i|\Big\|_p \quad (\forall i, \epsilon_i X_i \sim \epsilon_i |X_i|)\\
&\leq 2\Big\|\sum_{i=1}^n a_i \epsilon_i (\beta + U_i)\Big\|_p \quad \text{(the contraction principle)}\\
&\leq 2\Big\|\sum_{i=1}^n a_i \epsilon_i \beta\Big\|_p + 2\Big\|\sum_{i=1}^n a_i \epsilon_i U_i\Big\|_p\\
&\leq C\sqrt{p}\,\|a\|_2 + 2\Big\|\sum_{i=1}^n a_i \epsilon_i Y_i\Big\|_p \quad \text{(the contraction principle)},
\end{aligned}$$

where in the last inequality, we used Khintchine's inequality.

To finish the proof, we use the fact that $\epsilon_i Y_i$ is distributed like a symmetric Weibull r.v. with shape α and scaling parameter $\lambda = 1$. Thus,

$$\Big\|\sum_{i=1}^n a_i \epsilon_i Y_i\Big\|_p = \Big\|\sum_{i=1}^n a_i Z_i\Big\|_p$$

and to compute the p-th moment of $\sum a_i Z_i$ where $Z_1, \dots, Z_n$ are n i.i.d. symmetric Weibull variables we refer to Corollaries 2.9 and 2.10 of [16]. ∎

Thanks to the upper estimate on the p-th moments of the sum $\sum a_i X_i$ obtained in Theorem 0.3 for every $p \geq 1$, we can obtain a deviation result for $\sum a_i X_i$.

Theorem 0.4 ([16]). *There exists an absolute constant $c > 0$ such that the following holds. Let $1 \leq \alpha \leq 2$ and $X_1, \dots, X_n$ be independent random variables such that $\|X_i\|_{\psi_\alpha} \leq A, \forall i$. Then, for every $a = (a_1, \dots, a_n)^t \in \mathbb{R}^n$ and every $t \geq 0$,*

$$\mathbb{P}\Big[\Big|\sum_{i=1}^n a_i X_i\Big| \geq tA\Big] \leq 2\exp\Big(-c\min\Big(\frac{t^2}{\|a\|_2^2}, \frac{t^\alpha}{\|a\|_{\alpha^*}^\alpha}\Big)\Big),$$

where $\alpha^{-1} + (\alpha^*)^{-1} = 1$.

Proof. Without loss of generality, we can assume that $\|X_i\|_{\psi_\alpha} = 1$ for all $i = 1, \dots, n$. Take $a = (a_1, \dots, a_n)^t \in \mathbb{R}^n$ and set $Z := \big|\sum_{i=1}^n a_i X_i\big|$. Using Chebitchev's inequality, we obtain for every $p \geq 1$,

$$\mathbb{P}[Z \geq \|Z\|_p e] \leq \frac{\mathbb{E}Z^p}{\|Z\|_p^p e^p} = \exp(-p). \tag{0.12}$$

Take $1 \leq p \leq \big(\|a\|_2 / \|a\|_{\alpha^*}\big)^{2\alpha/(2-\alpha)}$. Then, using Theorem 0.3 and (0.12), we obtain

$$\mathbb{P}[Z \geq cp^{1/\alpha}\|a\|_{\alpha^*}] \leq \exp(-p)$$

and thus, for $t = p^{1/\alpha}\|a\|_{\alpha^*} \in \big[\|a\|_{\alpha^*}, \big(\|a\|_2^2 / \|a\|_{\alpha^*}^\alpha\big)^{1/(2-\alpha)}\big]$,

$$\mathbb{P}[Z \geq ct] \leq \exp\Big(-\frac{t^\alpha}{\|a\|_{\alpha^*}^\alpha}\Big). \tag{0.13}$$

For every $p \geq \big(\|a\|_2 / \|a\|_{\alpha^*}\big)^{2\alpha/(2-\alpha)}$, Theorem 0.3 and (0.12) yield $\mathbb{P}[Z \geq c\sqrt{p}\|a\|_2] \leq \exp(-p)$ and thus, for $t = \sqrt{p}\|a\|_2 \geq \big(\|a\|_2^2 / \|a\|_{\alpha^*}^\alpha\big)^{1/(2-\alpha)}$,

$$\mathbb{P}[Z \geq ct] \leq \exp\Big(-\frac{t^2}{\|a\|_2^2}\Big). \tag{0.14}$$

The claim follows by combining the results of (0.13) and (0.14) for the case $t \geq \|a\|_{\alpha^*}$. The case $0 \leq t \leq \|a\|_{\alpha^*}$ follows by chosing the absolute constant c in a convenient way. ■

Theorem 0.5 (cf.[14]). *Let X be a nonnegative random variable satisfying*

$$\mathbb{E}\exp\big(X^\alpha\big) < \infty, \text{ for some } \alpha \in (0,1).$$

Let $X_1, \dots, X_n$ be n i.i.d. copies of X. There exists $t_0 > 0$ depending only on α and $\mathbb{E}X$ and $c > 0$ depending only on α such that, for all $t > t_0$,

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}X > t\mathbb{E}X\Big] \leq \exp\big(-c(nt\mathbb{E}X)^\alpha\big).$$

Now, we take a particular look to Rademacher series:

Proposition 0.4 ([12]). *Let $a_1, \dots, a_n$ be n real numbers, $\epsilon_1, \dots, \epsilon_n$ be n independent Rademacher variable and $0 < p < \infty$. There exists some constants $A_{p,j} < B_{p,j} : j = 1, 2, 3$ such that the following holds:*

$$A_{p,1}\|a\|_2 \leq \Big\|\sum_{i=1}^n a_i\epsilon_i\Big\|_p \leq B_{p,1}\|a\|_2 \text{ and } \Big\|\sum_{i=1}^n a_i\epsilon_i\Big\|_\infty = \|a\|_1.$$

If $0 < p \leq 2$ then,

$$A_{p,2}\|a\|_2 \leq \Big\|\sum_{i=1}^n a_i\epsilon_i\Big\|_{\psi_p} \leq B_{p,2}\|a\|_2,$$

and for $p > 2$,

$$A_{p,3} \|a\|_{q,\infty} \le \left\| \sum_{i=1}^n a_i \epsilon_i \right\|_{\psi_p} \le B_{p,3} \|a\|_{q,\infty},$$

where $\|a\|_{q,\infty} := \sup_{t>0} t |\{i : |a_i| > t\}|^{1/q}$ *and* $p^{-1} + q^{-1} = 1$.

The first inequality is the called *Khintchin's inequality.* The best possible constants $A_{p,1}$ and $B_{p,1}$ are known (cf.[7]). We retain that $B_{p,1} \le K\sqrt{p}$ ($p \ge 1$) for some numerical constant K.

One drawback of the Orlicz norms for the characterization of concentration's properties of a random variable is that it only allows one type of "concentration behavior". In the sens that, saying that $Z \in L_{\psi_\alpha}$ tells that the tail decay of Z behaves like $\exp(-ct^\alpha), \forall t > 0$. Whereas some variables have more subtle behavior depending on the level of deviation t. For instance, a Rademacher series $Z = \sum_i a_i \epsilon_i$ satisfies $\|Z\|_{\psi_2} \sim \|a\|_2$. Thus, we can think that Rademacher series and Gaussian series $Y = \sum_i a_i g_i$ have the same tail behavior (since $\|Y\|_{\psi_2} \sim \|a\|_2$). Nevertheless, there is a gap between Rademacher series and Gaussian series and the tail behavior of Rademacher series is 'level dependent', the $\|a\|_2$-behavior being the worse gaussian case. To describe the exact tail behavior of Rademacher series we introduce the following interpolated norm

$$K_{1,2}(x,t) := \inf\big(\|x'\|_1 + t \|x''\|_2 : x' \in \ell_1^n, x'' \in \ell_2^n, x = x' + x'', x = x' + x'' \big).$$

Theorem 0.6 ([13])**.** *There exists an absolute constant $c > 0$ such that the following holds. For every vector $a = (a_1, \ldots, a_n)^t \in \mathbb{R}^n$ and every $t > 0$,*

$$\mathbb{P}\Big[\sum_{i=1}^n a_i \epsilon_i > K_{1,2}(a,t) \Big] \le \exp\big(-t^2/2 \big)$$

and

$$\mathbb{P}\Big[\sum_{i=1}^n a_i \epsilon_i > c^{-1} K_{1,2}(a,t) \Big] \ge c^{-1} \exp\big(-t^2 c \big).$$

Note that this result has been extended to Banach-valued Rademacher chaos in [3]. We can also obtain precise estimates of the L_p norms of Rademacher sums.

Theorem 0.7 (cf.[11, 8])**.** *Let $p \ge 1$, $a_1, \ldots, a_n$ be real numbers and $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher random variables. We have*

$$\left\| \sum_{i=1}^n a_i \epsilon_i \right\|_p \sim \sum_{i \le p} a_i^* + \sqrt{p} \Big(\sum_{i > p} a_i^{*2} \Big)^{1/2} \sim K_{1,2}(a, \sqrt{p}),$$

where $(a_i^* : i = 1, \ldots, n)$ is the non-increasing rearrangement of the absolute values of $(a_1, \ldots, a_n)$.

Note that Theorem 0.7 is stronger than the result on the L_p moments of Rademacher sums of Theorem 0.5 because the equivalence in Theorem 0.7 is up to an absolute constant whereas, in Theorem 0.6, the multiplying constant depend on p. Moreover, it is obvious that $c_p \|a\|_2 \le K_{1,2}(a, \sqrt{p}) \le \sqrt{p} \|a\|_2, \forall a \in \mathbb{R}^n$.

Proof.

■

3. Bernstein's inequalities

We first start with bounded random variables.

Theorem 0.8 (Prokhorov-Bennett's inequality)**.** *Let $X_1, \dots, X_n$ be n independent mean-zero random variables such that $|X_i| \le c$ a.s.. Set $\sigma^2 := n^{-1}\sum_i \mathbb{V}(X_i)$. For any $t > 0$, we have*

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n X_i \ge t\Big] \le \exp\big(-\frac{n\sigma^2}{c^2}h\Big(\frac{ct}{\sigma^2}\Big)\big),$$

where $h(u) := (1+u)\log(1+u) - u, \forall u > 0$.

Proof. Let $t > 0$. For every $\lambda > 0$,

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n X_i \ge t\Big] \le \exp(-\lambda t)\mathbb{E}\exp\Big(\frac{\lambda}{n}\sum_{i=1}^n X_i\Big) = \exp(-\lambda t)\prod_{i=1}^n\Big[1 + \sum_{k\ge 2}\frac{\lambda^k \mathbb{E}X_i^k}{n^k k!}\Big]$$

$$\le \exp\big(-\lambda t\big)\exp\Big(n\sigma^2\sum_{k\ge 2}\frac{\lambda^k c^{k-2}}{n^k k!}\Big) \le \exp\Big(\frac{n\sigma^2}{c^2}\Big(\exp\Big(\frac{\lambda c}{n}\Big) - 1 - \frac{c\lambda}{n}\Big) - \lambda t\Big).$$

The claim follows by optimizing in λ in the last inequality. ■

The most famous corollary of Prokhorov-Bennett's inequality is Bernstein's inequality for bounded variables which follows since $h(u) \ge u^2/(2 + 2u/3), \forall u > 0$.

Theorem 0.9 (Bernstein's inequality)**.** *There exists an absolute constant $c > 0$ for which the following holds. Let $X_1, \dots, X_n$ be n independent mean-zero random variables such that $X_i \in L_\infty, \forall i$. Then, for every $t > 0$,*

$$\mathbb{P}\Big[\Big|\frac{1}{n}\sum_{i=1}^n X_i\Big| \ge t\Big] \le 2\exp\Big(-cn\min\Big(\frac{t^2}{\sigma^2}, \frac{t}{M}\Big)\Big),$$

where $M := \max_i \|X_i\|_\infty$ and $\sigma^2 := n^{-1}\sum_i \mathbb{V}(X_i)$.

From Bernstein's inequality, we can deduce that the concentration behavior of a mean of bounded variable has two regimes. On one hand, there is a subexponential regime w.r.t. M for large values of t ($t \ge \sigma^2/M$). On the other hand, the mean has a subgaussian concentration behaviour w.r.t. σ^2. Bernstein's inequality can be compared with Hoeffding's inequality (cf. Exercise 0.18) which provides a subgaussian concentration with respect to M.

Now, we turn on to the "ψ_1" version of Bernstein's inequality.

Theorem 0.10 (Bernstein's inequality for sub-exponential r.v.)**.** *There exists an absolute constant $c > 0$ for which the following holds. Let $X_1, \dots, X_n$ be n independent mean-zero and ψ_1 random variables. Then, for every $t > 0$,*

$$\mathbb{P}\Big[\Big|\frac{1}{n}\sum_{i=1}^n X_i\Big| \ge t\Big] \le 2\exp\Big(-cn\min\Big(\frac{t^2}{\bar{v}}, \frac{t}{M}\Big)\Big),$$

where $M := \max_i \|X_i\|_{\psi_1}$ and $\bar{v} := n^{-1}\sum_i \|X_i\|_{\psi_1}^2$.

Proof. Let $X \in L_{\psi_1}$. Using that for every integer $k \ge 2$, the following decomposition $\mathbb{E}|X|^k \le Kk!e^k \|X\|_{\psi_1}^2 \|X\|_{\psi_1}^{k-2}$ (cf. Equation (0.7) combined with Proposition 0.2), it is easy to get, for every $\lambda \ge 0$ such that $e\lambda \|X\|_{\psi_1} < 1$,

$$\mathbb{E}(\exp(\lambda X) - 1 - \lambda X) \le \frac{K(e\lambda)^2 \|X\|_{\psi_1}^2}{1 - e\lambda \|X\|_{\psi_1}}. \tag{0.15}$$

We denote by $\bar{X}_n$ the empirical mean $n^{-1}\sum_i X_i$. The Legendre's transform of $\bar{X}_n$ is defined by $\psi_{\bar{X}_n}(\lambda) := \log\mathbb{E}\exp(\lambda\bar{X}_n)$ for every $\lambda \ge 0$. Using Equation (0.15) and independence, we obtain, for every $\lambda \geq 0$ such that $e\lambda \|X_i\|_{\psi_1} < 1, \forall i = 1, \dots, n$,

$$\psi_{\bar{X}_n}(\lambda) \leq K\lambda^2 \bar{v}\Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\|X_i\|_{\psi_1}^2}{\bar{v}} \phi\big(\|X_i\|_{\psi_1} \big)\Big], \tag{0.16}$$

where $\phi(x) := (1 - e\lambda x)^{-1}, \forall x \in [0, (e\lambda)^{-1})$. Since ϕ is convex, we get, for every $0 \leq \lambda < (eM)^{-1}$ where $M := \max_i \|X_i\|_{\psi_1}$,

$$\psi_{\bar{X}_n}(\lambda) \leq \frac{K(e\lambda)^2 \bar{v}^2}{\bar{v} - e\lambda\bar{\mu}} \leq \frac{K(e\lambda)^2 \bar{v}}{1 - e\lambda M}, \tag{0.17}$$

where $\bar{\mu} := n^{-1} \sum_i \|X_i\|_{\psi_1}^3 \leq M\bar{v}$. We denote by $\psi_{\bar{X}_n}^*$ the convex conjugate of $\psi_{\bar{X}_n}$ defined by

$$\psi_{\bar{X}_n}^*(t) := \sup\big(t\lambda - \psi_{\bar{X}_n}(\lambda) : \lambda \in [0, \lambda^{-1})\big), \forall t \geq 0.$$

Using Markov's inequality and Equation (0.17), it is easy to get

$$\mathbb{P}\Big[\bar{X}_n \geq t\Big] \leq \exp(-\psi_{\bar{X}_n}(t)) \leq \exp\Big(\frac{-cnt^2}{\bar{v} + tM}\Big).$$

We obtain the same result for $-\bar{X}_n$ and then, the claim follows. ∎

4. Self-bounded variables

Now, we come to another type of concentration's inequalities obtained using logarithmic Sobolev inequalities. We will restrict ourselves to recall the results from [1].

We begin by introducing some notation that is used throughout this section. We assume that $X_1, \dots, X_n$ are independent r.v. taking values in a measurable space $\mathcal{X}$. Denote by X_1^n the vector of these n random variables. Let $f : \mathcal{X}^n \longmapsto \mathbb{R}$ be some measurable function. We are concerned with the concentration of

$$Z := f(X_1, \dots, X_n).$$

Throughout, $X_1', \dots, X_n'$ denote independent copies of $X_1, \dots, X_n$, and we write

$$Z^{(i)} := f(X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n).$$

Define the random variables V_+ and V_- by

$$V_+ := \mathbb{E}\Big[\sum_i (Z - Z^{(i)})^2 \mathbb{1}_{Z > Z^{(i)}} | X_1^n\Big] \text{ and } V_- := \mathbb{E}\Big[\sum_i (Z - Z^{(i)})^2 \mathbb{1}_{Z < Z^{(i)}} | X_1^n\Big].$$

Theorem 0.11. *For all $\theta > 0$ and $\lambda \in (0, 1/\theta)$,*

$$\log \mathbb{E} \exp\big(\lambda(Z - \mathbb{E}Z)\big) \leq \frac{\lambda\theta}{1 - \lambda\theta} \log \mathbb{E}\Big[\exp\Big(\frac{\lambda V_+}{\theta}\Big)\Big]$$

and

$$\log \mathbb{E} \exp\big(-\lambda(Z - \mathbb{E}Z)\big) \leq \frac{\lambda\theta}{1 - \lambda\theta} \log \mathbb{E}\Big[\exp\Big(\frac{\lambda V_-}{\theta}\Big)\Big].$$

Corollary 0.2. *Assume that there exists a positive constant c such that $V_+ \leq c$ a.s.. Then, for all $t > 0$,*

$$\mathbb{P}[Z > \mathbb{E}Z + t] \leq \exp(-t^2/(4c)).$$

Moreover, if $V_- \leq c$ a.s. then, for all $t > 0$,

$$\mathbb{P}[Z < \mathbb{E}Z - t] \leq \exp(-t^2/(4c)).$$

Corollary 0.3. *Assume the random variable V_+ to be such that there exists a positive constant c such that for all $\lambda \in (0, c^{-1})$,*

$$\log \mathbb{E}\big[\exp\big(\lambda(V_+ - \mathbb{E}V_+)\big)\big] \le \frac{\lambda^2 c \mathbb{E}V_+}{1 - c\lambda}.$$

Then,

$$\mathbb{P}[Z > \mathbb{E}Z + t] \le \exp\Big(\frac{-t^2}{4\mathbb{E}V_+ + 2(c+1)t/3}\Big).$$

The same result holds for the lower tail when the Legendre transform of V_- satisfies the same type of inequality.

Theorem 0.12. *Assume that there exist positive constants a and b such that*

$$V_+ \le aZ + b.$$

Then, for all $t > 0$,

$$\mathbb{P}[Z > \mathbb{E}Z + t] \le \exp\Big(\frac{-t^2}{4a\mathbb{E}Z + 4b + 2at}\Big).$$

Theorem 0.13. *Assume that for some non-decreasing function g,*

$$V_- \le g(Z).$$

Then, for all $t > 0$,

$$\mathbb{P}[Z < \mathbb{E}Z - t] \le \exp\Big(\frac{-t^2}{4(e-1)\mathbb{E}g(Z)}\Big).$$

Theorem 0.14. *Assume that for some non-decreasing function g, $V_- \le g(Z)$ and for any value of X_1^n and X_i', $|Z - Z^{(i)}| \le 1$ for all $i = 1, \dots, n$. Then, for all $t > 0$, with $t \le (e-1)\mathbb{E}g(Z)$ we have*

$$\mathbb{P}[Z < \mathbb{E}Z - t] \le \exp\Big(\frac{-t^2}{4\mathbb{E}g(Z)}\Big).$$

Theorem 0.15. *Assume that f is nonnegative. Assume that there exists a random variable W, such that*

$$V_+ \le WZ.$$

Then, for all $\theta > 0$ and $\lambda \in (0, 1/\theta)$,

$$\log \mathbb{E}\exp\big(\lambda(\sqrt{Z} - \mathbb{E}\sqrt{Z})\big) \le \frac{\lambda\theta}{1 - \lambda\theta} \log \mathbb{E}\exp\Big(\frac{\lambda W}{\theta}\Big).$$

Bounds for the upper tail probability of Z may be derived using Theorem 0.15, since for any $\lambda > 0$,

$$\mathbb{P}[Z > \mathbb{E}Z + t] \le \mathbb{E}\exp\big(\lambda(\sqrt{Z} - \mathbb{E}\sqrt{Z})\big)\exp(-\lambda x),$$

for $x := \sqrt{\mathbb{E}Z + t} - \sqrt{\mathbb{E}Z}$.

An immediate application of Corollary 0.2 is the well known McDiarmind inequality.

Theorem 0.16 (Bounded difference inequality)**.** *Assume that for all $i = 1, \dots, n$, there exists a constant c_i such that*

$$\sup_{x_1,\dots,x_n,x_i'} |f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n)| \le c_i.$$

Then, for all $t > 0$,

$$\mathbb{P}[Z > \mathbb{E}Z + t], \mathbb{P}[Z < \mathbb{E}Z - t] \le \exp\Big(\frac{-2t^2}{\sum_i c_i^2}\Big)$$

5. Martingale inequalities

In this section, we recall three useful results of [12]. Recall that $L_1 = L_1(\Omega, \mathcal{A}, \mathbb{P})$ is the space of all measurable functions f on Ω such that $\mathbb{E}|f| < \infty$. Assume that we are given a filtration

$$\{\emptyset, \Omega\} := \mathcal{A}_0 \subset \mathcal{A}_1 \subset \cdots \subset \mathcal{A}_n = \mathcal{A}$$

of sub-σ-algebras of $\mathcal{A}$. The symbol $\mathbb{E}^{\mathcal{A}_i}$ denotes the conditional operator w.r.t. $\mathcal{A}_i$. Given $f \in L_1$, we set, for each $i = 1, \dots, n$,

$$d_i := \mathbb{E}^{\mathcal{A}_i} f - \mathbb{E}^{\mathcal{A}_{i-1}} f$$

so that $f - \mathbb{E}f = \sum_i d_i$. $(d_i : i \leq n)$ defines a so-called *martingale difference sequence* characterized by the property $\mathbb{E}^{\mathcal{A}_i} d_i = 0, i \leq n$.

One of the typical martingale difference that we have in mind is a sequence $(X_i : i \leq n)$ of independent mean zero random variables for the filtration $(\mathcal{A}_i : i \leq n)$ defined by $\mathcal{A}_i := \sigma(X_1, \dots, X_i)$. Hence, all the results present for $f - \mathbb{E}f = \sum d_i$ can be applied to the sum $\sum X_i$.

Theorem 0.17. *Assume that $\|d_i\|_\infty < \infty$ and let $a := \big(\sum_{i=1}^n \|d_i\|_\infty^2 \big)^{1/2}$. Then, for every $t > 0$,*

$$\mathbb{P}[f - \mathbb{E}f > t] \leq \exp\big(- t^2/(2a^2)\big).$$

Theorem 0.18. *Set $a := \max_i \|d_i\|_\infty$ and $b \geq \big(\sum \Big\| \mathbb{E}^{\mathcal{A}_{i-1} d_i^2} \Big\| \big)^{1/2}$. Then, for every $t > 0$,*

$$\mathbb{P}[f - \mathbb{E}f > t] \leq \exp \left[\frac{-t^2}{2b^2} \Big(2 - \exp(at/b^2) \Big) \right]$$

Theorem 0.19. *Let $1 < p < 2$ and denote by q its conjugate ($p^{-1} + q^{-1} = 1$). There exists a constant c_q depending only on q such that for $a := \max_i i^{1/p} \|d_i\|_\infty$ and any $t > 0$*

$$\mathbb{P}[f - \mathbb{E}f > t] \leq \exp\big(- c_q t^q / a^q \big).$$

6. Talagrand's concentration inequalities

In this section, we recall the very useful results on the concentration of supremum of sum of independent random variables due to M.Talagrand.

Theorem 0.20. *There exists an absolute constant $K > 0$ such that the following holds. Let $(\mathcal{X}, \sigma)$ be a measurable space, $\mathcal{F}$ be a set of real-valued functions defined on $\mathcal{X}$ and $X, X_1, \dots, X_n$ be $n+1$ i.i.d. $\mathcal{X}$-valued random variables. Assume that $\mathbb{E}f(X) = 0, \forall f \in \mathcal{F}$, that $\|\mathcal{F}\|_\infty := \sup\big(\|f(X)\|_\infty : f \in \mathcal{F} \big)$ and $\sigma^2(\mathcal{F}) := \sup\big(\mathbb{V}(f(X)) : f \in \mathcal{F} \big)$ are finite. Denote*

$$Z := \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n f(X_i) \text{ and } \bar{Z} := \sup_{f \in \mathcal{F}} \Big| \frac{1}{n} \sum_{i=1}^n f(X_i) \Big|.$$

For every $x > 0$ and $\alpha > 0$, with probability greater than $1 - \exp(-x)$,

$$Z \leq (1 + \alpha)\mathbb{E}Z + K\sigma(\mathcal{F}) \sqrt{\frac{x}{n}} + K\big(1 + \alpha^{-1}\big) \frac{\|\mathcal{F}\|_\infty x}{n}.$$

For every x, with probability greater than $1 - \exp(-x)$,

$$Z \geq (1 - \alpha)\mathbb{E}Z - K\sigma(\mathcal{F}) \sqrt{\frac{x}{n}} - K\big(1 + \alpha^{-1}\big) \frac{\|\mathcal{F}\|_\infty x}{n}.$$

The same inequalities hold for $\bar{Z}$.

7. Exercises and related results

Exercise 0.1 (Stein's approximation method for Berry-Essen type theorem)**.** *Let* $(X_i : i \geq 1)$ *be a sequence of i.i.d. mean zero random variables with variance* 1*. Let* d_L *denote the Lipschitz distance for probability distribution defined by* $d_L(P,Q) := \sup\big(|Ph - Qh| : h \in \mathrm{Lip}\big)$ *for any probability distribution* P *and* Q*, where* Lip *is the set of all Lipschitz functions with Lipschitz constant* 1*. We have*

$$d_L\Big(\mathcal{L}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big), \mathcal{N}(0,1)\Big) \leq \frac{5\mathbb{E}|X_1|^3}{\sqrt{n}}, \forall n \in \mathbb{N}.$$

Proof. The proof is based on Stein's method which is at the heart of the approximation theory in probability. We present here the general scheme of this method together with the proof of the exercise (cf. Wikipedia article).

Stein's method is a way to bound the distance of two probability distributions in a specific probability metric. To be tractable with the method, the metric must be given in the form

$$d(P,Q) := \sup_{h\in\mathcal{H}} \big|Ph - Qh\big| = \sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Y)|, \tag{0.18}$$

where P and Q are two probability distributions on the measurable space $(\mathcal{X}, \sigma)$, W and Y are random variables with probability distribution P and Q respectively and $\mathcal{H}$ is a set of functions from $\mathcal{X}$ to $\mathbb{R}$. This set has to be large enough, so that the above definition indeed yields a metric. Important examples are the **total variation metric**, where $\mathcal{H}$ consists of all the indicator functions of measurable sets, the **Kolmogorov (uniform) metric**, where $\mathcal{H}$ is the set of all the half-line indicator functions, and the **Lipschitz (first order Wasserstein; Kantorovich) metric**, where $\mathcal{H}$ is itself a metric space made of all Lipschitz-continuous functions with Lipschitz-constant 1. However, note that not every metric can be represented in the form (0.18).

In what follows we think of P as a complicated distribution (e.g. a sum of dependent random variables), which we want to approximate by a much simpler and tractable distribution Q (e.g. the standard normal distribution to obtain a central limit theorem).

The Stein operator: We assume now that the distribution Q is a fixed distribution; in what follows we shall in particular consider the case where Q is the standard normal distribution, which serves as a classical example of the application of Stein's method.

First of all, we need an operator $\mathcal{A}$ which acts on a class $\mathcal{C}$ of functions f from $\mathcal{X}$ to $\mathbb{R}$, and which 'characterizes' the distribution Q in the sense that the following equivalence holds:

$$\mathbb{E}(\mathcal{A}f)(Y) = 0, \forall f \in \mathcal{C} \Longleftrightarrow Y \text{ has probability distribution } Q. \tag{0.19}$$

Such an operator is called **the Stein operator**. For the standard normal distribution $(\mathcal{N}(0,1))$, **Stein's lemma** exactly yields such an operator:

$$\mathbb{E}(Yf(Y) - f'(Y)) = 0, \forall f \in \mathcal{C}_b^1 \Longleftrightarrow Y \sim \mathcal{N}(0,1) \tag{0.20}$$

where $\mathcal{C}_b^1$ is the set of all continuously differentiable functions with bounded derivative. Thus, for the standard Gaussian probability distribution we can take

$$\mathcal{A}f(x) = xf(x) - f'(x), \forall f \in \mathcal{C}_b^1, \forall x \in \mathbb{R} \tag{0.21}$$

as a Stein Operator.

There are different ways to find Stein operators. But by far the most important one is via generators (cf. works of Barbour and Götze). Assume that $Z := (Z_t)_{t\geq 0}$ is a (homogeneous) continuous time Markov process taking values in $\mathcal{X}$. If Z has the stationary distribution Q it is easy to see that, if $\mathcal{A}$ is the generator of Z, we have $\mathbb{E}(\mathcal{A}f)(Y) = 0$ for a large set of functions f. Thus, generators are natural candidates for Stein operators and this approach will also help us for later computations.

Setting up the Stein equation: Observe now that saying that P is close to Q with respect to d is equivalent to saying that the difference of expectations in (0.18) is close to 0, and indeed if $P = Q$ it is equal to 0. We hope now that the operator exhibits the same behavior: clearly if $P = Q$ we have $\mathbb{E}(\mathcal{A}f)(W) = 0$ and hopefully if $P \approx Q$ we have $\mathbb{E}(\mathcal{A}f)(W) \approx 0$.

To make this statement rigorous we could find a function f, such that, for a given function h,

$$\mathbb{E}(\mathcal{A}f)(W) = \mathbb{E}h(W) - \mathbb{E}h(Y) \tag{0.22}$$

so that the behavior of the right hand side is reproduced by the operator $\mathcal{A}$ and f. However, this equation is too general. We solve instead the more specific equation

$$(\mathcal{A}f)(x) = h(x) - \mathbb{E}h(Y), \forall x$$

which is called the **Stein equation**. Replacing x by W and taking expectation with respect to W, we are back to (0.22), which is what we effectively want. Now all the effort is worth only if the left hand side of (0.22) is easier to bound than the right hand side. This is, surprisingly, often the case.

If Q is the standard normal distribution and we use (0.21), the corresponding Stein equation is

$$xf(x) - f'(x) = h(x) - \mathbb{E}h(Y), \forall x \in \mathbb{R},$$

which is just an ordinary differential equation having for solution

$$f(x) = \exp(x^2/2) \int_{-\infty}^{x} [h(s) - \mathbb{E}h(Y)] \exp(-s^2/2) ds. \tag{0.23}$$

In particular, this solution satisfies

$$\|f\|_\infty, \|f'\|_\infty, \|f''\|_\infty \leq c \max\{\|h\|_\infty, \|h'\|_\infty\}, \tag{0.24}$$

where this bound is of course only applicable if h is differentiable (or at least Lipschitz-continuous, which, for example, is not the case if we regard the total variation metric of the Kolmogorov metric). As the standard normal distribution has no extra parameters, in this specific case, the constants are free of additional parameters.

Note that, up to this point, we did not make use of the random variable W. So, the steps up to here in general have to be calculated only once for a specific probability distribution Q, metric d and Stein operator $\mathcal{A}$.

Now, we turn to the proof of the Berry-Essen type of result for the Lipschitz distance. We set $W := \sum_{i=1}^{n} X_i$. By a Taylor expansion, we get

$$\left|\mathbb{E}(f'(W) - Wf(W))\right| \leq n \|f''\| \left[(1/2)\mathbb{E}|X_i|^3 + \mathbb{E}|X_i|^2 \mathbb{E}|X_i|\right].$$

We conclude with Equation (0.24). ■

Exercise 0.2 (cf. [4])**.** *Let ψ be a Young-Orlicz modulus. By convexity, ψ has a right derivative in every point $x \geq 0$, we set $\Psi(x) := \psi'(x^+)$. Assume that Ψ is unbounded and $\Psi(x)$ tends to zero when x tends to zero. Denote $\Phi(y) := \inf\big(x \geq 0 : \Psi(x) \geq y\big)$ and $\phi(y) := \int_0^y \Phi(t)dt$.*

(1) ϕ is a Young-Orlicz modulus;
(2) $\forall x, y \geq 0, xy \leq \psi(x) + \phi(y)$.

In particular, for all $x, y \geq 0$,

$$xy \leq \exp\left(x^2/\beta^2\right) - 1 + \beta y \log^{1/2}\left(1 + y\right).$$

Exercise 0.3. *Let $X \in L_{\psi_2}$. We have $\left\|X^2\right\|_{\psi_1} = \|X\|_{\psi_2}$ and $\|X\|_{\psi_2} \le \mathbb{E}\exp(X^2)$. If $X \sim \mathcal{N}(0,1)$ then $\|X\|_{\psi_2} \le \sqrt{8/3}$.*

Exercise 0.4 (Bernoulli estimate and Chernoff's bound)**.** *Let $\delta_1, \dots, \delta_n$ be n i.i.d. Bernoulli random variables with mean δ (also called* ***selectors****).* ***Chernoff's bound*** *says that for every $t > 0$,*

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n \delta_i - \delta \ge t\Big], \mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n \delta_i - \delta \le -t\Big] \le \exp(-nh_\delta(t)),$$

where $h_\delta(t) := (1-\delta-t)\log\Big(\frac{1-\delta-t}{1-\delta}\Big) + (\delta+t)\log\Big(\frac{\delta+t}{\delta}\Big)$.

Let $(A_i : i \ge 1)$ be a sequence of independent sets such that $a := \sum_i \mathbb{P}(A_i) < \infty$. For every n, the ***Bernoulli estimate*** *says that*

$$\mathbb{P}\Big[\sum_i \mathbb{I}_{A_i} \ge n\Big] \le \frac{a^n}{n!} \le \Big(\frac{ea}{n}\Big)^n.$$

Proof. We use the classical Cramer-Chernoff method to obtain, for all $t > 0$,

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n \delta_i - \delta \ge t\Big] \le \exp\big(-\psi^*_{\bar{\delta}_n}(t)\big),$$

where $\psi^*_{\bar{\delta}_n}$ is the convex conjugate of the Legendre transform of $\bar{\delta}_n := n^{-1}\sum_{i=1}^n \delta_i - \delta$ defined by $\psi_{\bar{\delta}_n}(\lambda) := \log \mathbb{E}\exp(\lambda\bar{\delta}_n)$. Using independence, we have $\psi_{\bar{\delta}_n}(\lambda) = n\psi_{\delta_1-\delta}(\lambda/n)$ and so $\psi^*_{\bar{\delta}_n} = n\psi^*_{\delta_1-\delta}(\lambda)$, for all $\lambda > 0$. It is easy to get, for every $\lambda \ge 0$,

$$\psi_{\delta_1-\delta}(\lambda) = \log\big(\delta\exp(\lambda) + 1 - \delta\big) - \lambda\delta$$

and, for every $0 \le t < 1-\delta$,

$$\psi^*_{\delta_1-\delta}(t) = (1-\delta-t)\log\Big(\frac{1-\delta-t}{1-\delta}\Big) + (\delta+t)\log\Big(\frac{\delta+t}{\delta}\Big).$$

We also have, for every $\delta < t \le 1$,

$$\psi^*_{\delta_1}(t) = (1-t)\log\Big(\frac{1-t}{1-\delta}\Big) + t\log\Big(\frac{t}{\delta}\Big).$$

Now, we turn to the proof of the Bernoulli estimate. Using the union-bound and the independence, we get

$$\begin{aligned}\mathbb{P}\big[\sum_i \mathbb{I}_{A_i} \ge n\big] &= \mathbb{P}\Big[\cup_{i_1<\ldots<i_n} A_{i_1}\cap\ldots\cap A_{i_n}\Big] \le \sum_{i_1<\ldots<i_n}\prod_{j=1}^n \mathbb{P}(A_j)\\ &= \frac{1}{n!}\sum_{i_1\neq\cdots\neq i_n}\prod_j \mathbb{P}(A_{i_j}) \le \frac{1}{n!}\sum_{(i_1,\ldots,i_n)}\prod_j \mathbb{P}(A_{i_j}) \le \frac{a^n}{n!}.\end{aligned}$$

We conclude with Stirling's formula. ∎

Exercise 0.5 (Borell's inequality)**.** *Let $(X_t : t \in T)$ be a centered Gaussian process then, for every $x > 0$,*

$$\mathbb{P}\Big[\sup_{t\in T} X_t \ge \mathrm{Med}\Big(\sup_{t\in T} X_t\Big) + x\Big] \le \mathbb{P}[G > x],$$

where G is a standard Gaussian variable and Med *denotes the median.*

Exercise 0.6. *Let $a_1, \dots, a_n$ be n real numbers and $\epsilon_1, \dots, \epsilon_n$ be n independent Rademacher. Then,*

$$\Big\|\sum_{i=1}^n a_i\epsilon_i\Big\|_{\psi_2} \le \sqrt{6}\,\|a\|_2$$

Proof. First, note that for every $x \in \mathbb{R}$, $\exp(x) + \exp(-x) \le 2\exp(x^2/2)$. Now, using the symmetry of $\sum_i a_i\epsilon_i$, we have, for every $\lambda > 0$,

$$\mathbb{E}\exp\big(\lambda|\sum_i a_i\epsilon_i|\big) \le 2\mathbb{E}\exp\big(\lambda\sum_i a_i\epsilon_i\big) \le 2\prod_i \frac{\exp(\lambda a_i) + \exp(-\lambda a_i)}{2}$$
$$\le 2\exp\big(\lambda^2 \|a\|_2^2\big).$$

We finish the proof with Proposition 0.2. ■

Exercise 0.7. *Let Z be a nonnegative random variable such that there exists $c > 0$ satisfying*

$$\mathbb{P}\big[Z - \mathbb{E}Z \ge m\mathbb{E}Z\big] \le c^{-1}\exp(-cm).$$

Then, for every $1 < p < \infty$, there exists c_p such that

$$c_p\big(\mathbb{E}Z^p\big)^{1/p} \le \mathbb{E}Z \le \big(\mathbb{E}Z^p\big)^{1/p}.$$

Proof. Let $1 < p < \infty$. Jensen inequality yields $\mathbb{E}Z \le \big(\mathbb{E}Z^p\big)^{1/p}$. Set $a := \mathbb{E}Z$. We have

$$\begin{aligned}\mathbb{E}Z^p &= \mathbb{E}Z^p\mathbb{1}_{Z\le a} + \sum_{k=1}^{\infty}\mathbb{E}Z^p\mathbb{1}_{ka<Z\le(k+1)a}\\ &\le a^p + \sum_{k\ge1}(k+1)^p a^p\mathbb{P}\big[Z > ka\big]\\ &\le a^p + c^{-1}a^p\sum_{k\ge1}(k+1)^p\exp(-c(k-1)) \le c_p' a^p.\end{aligned}$$

■

Exercise 0.8 (Rosenthal's inequality, cf. [9])**.** *Let $p \ge 1$ and $X_1, \dots, X_n$ be independent random variables with finite p-th moment. There exists constants A_p and B_p that grow like $p/\log p$ as $p \to \infty$ such that:*

if $X_1, \dots, X_n$ are symmetric then, for $p > 2$,

$$\max\left(\Big\|\sum_i X_i\Big\|_2, \big(\sum_i \|X_i\|_p^p\big)^{1/p}\right) \le \Big\|\sum_{i=1}^n X_i\Big\|_p \le A_p\max\left(\Big\|\sum_i X_i\Big\|_2, \big(\sum_i \|X_i\|_p^p\big)^{1/p}\right);$$

if the $X_1, \dots, X_n$ are nonnegative then,

$$\max\left(\Big\|\sum_i X_i\Big\|_1, \big(\sum_i \|X_i\|_p^p\big)^{1/p}\right) \le \Big\|\sum_{i=1}^n X_i\Big\|_p \le B_p\max\left(\Big\|\sum_i X_i\Big\|_1, \big(\sum_i \|X_i\|_p^p\big)^{1/p}\right).$$

Let $p \ge 1$ and $Y_1, \dots, Y_n$ be random variables such that $\mathbb{E}Y_i^p = 1, \forall i$. There exists constants C_p and D_p that grow like $p/\log p$ as $p \to \infty$ such that:

if $Y_1, \dots, Y_n$ are symmetrically exchangeable (i.e. the vectors $(Y_1, \dots, Y_n)$ and $(\epsilon_1 Y_{\pi(1)}, \dots, \epsilon_n Y_{\pi(n)})$ have the same probability distribution for every permutation π) then, for $p \ge 2$ and every vector $a = (a_1, \dots, a_n)^t \in \mathbb{R}^n$, then,

$$\max\big(Cn^{-1/2}\|a\|_2, \|a\|_p^p\big) \le \Big\|\sum_{i=1}^n a_iY_i\Big\|_p \le C_p\max\big(Cn^{-1/2}\|a\|_2, \|a\|_p^p\big),$$

where $C := \Big\|\big(\sum_i |Y_i|^2\big)^2\Big\|_p$.

if the $X_1, \dots, X_n$ are nonnegative exchangeable then, for every vector of nonnegative scalars $a = (a_1, \dots, a_n) \in \mathbb{R}^n_+$,

$$\max\left(Cn^{-1/2}\|a\|_1, \|a\|_p^p\right) \le \left\|\sum_{i=1}^n a_i Y_i\right\|_p \le C_p \max\left(Cn^{-1/2}\|a\|_1, \|a\|_p^p\right),$$

where $C := \|\sum_i Y_i\|_p$.

Exercise 0.9 (Paley-Zygmund, cf.[2])**.** *Let X be a nonnegative random variable such that $0 < \mathbb{E}X^q < \infty$ for some $q > 0$. Then, for any $0 \le \lambda \le 1$ and $0 < p \le q < \infty$, we have*

$$\mathbb{P}\Big[X \ge \lambda \|X\|_p\Big] \ge \Big((1-\lambda^p)\Big(\frac{\|X\|_p}{\|X\|_q}\Big)^{p/q}\Big)^{\frac{q}{q-p}}.$$

Proof. Let $0 < \lambda < 1$. By Hölder inequality, we have

$$\mathbb{E}Z^p = \mathbb{E}Z^p \mathbb{1}_{Z \le \lambda \|Z\|_p} + \mathbb{E}Z^p \mathbb{1}_{Z > \lambda \|Z\|_p} \le \left(\lambda \|Z\|_p\right)^p + (\mathbb{E}Z^q)^{p/q} \mathbb{P}[Z > \lambda \|Z\|_p]^{\frac{q-p}{p}}.$$

The claim follows easily. ■

Exercise 0.10 (Hoffmann-Jϕrgensen, cf.[12])**.** *Let $(B, \|\cdot\|)$ be separable Banach space, $0 < p < \infty$ and let $X_1, \dots, X_n$ be independent random variables in $L_p(B)$. Set $S_k = \sum_{i=1}^k X_i, k \le n$. Then, for*

$$t_0 := \inf\{t > 0 : \mathbb{P}[\max_{k \le n} \|S_k\| > t] \le 2^{-2p-1}\}$$

we have

$$\mathbb{E} \max_{k \le n} \|S_k\|^p \le 2^{2p+1} \mathbb{E} \max_{i \le n} \|X_i\|^p + 2(4t_0)^p.$$

If, moreover, the X_i's are symmetric and $t_0 := \inf\{t > 0 : \mathbb{P}[\|S_n\| > t] \le (8 \cdot 3^p)^{-1}\}$, then

$$\mathbb{E}\|S_n\|^p \le 2 \cdot 3^p \mathbb{E} \max_{i \le n} \|X_i\|^p + 2(3t_0)^p.$$

Exercise 0.11 (Khintchine's inequality by the hyper-contractivity argument, cf. [2])**.** *Let $x_1, \dots, x_n$ be n real numbers and $\epsilon_1, \dots, \epsilon_n$ be independent Rademacher variables. Then,*

$$\left\|\sum_{i=1}^n \epsilon_i x_i\right\|_2 \le \sqrt{3} \left\|\sum_{i=1}^n \epsilon_i x_i\right\|_1.$$

Proof. Set $Z = \sum \epsilon_i x_i$. We have, for every $\alpha = 2/3$,

$$\begin{aligned}\mathbb{E}Z^2 &= \mathbb{E}|Z|^\alpha |Z|^{2-\alpha} \le \||Z|^\alpha\|_{1/\alpha} \left\||Z|^{2-\alpha}\right\|_{1/(1-\alpha)} \\ &= [\mathbb{E}|Z|]^\alpha \left[\mathbb{E}|Z|^{\frac{2-\alpha}{1-\alpha}}\right]^{1-\alpha} = [\mathbb{E}|Z|]^{2/3}\left[\mathbb{E}|Z|^4\right]^{1/3}. \qquad (0.25)\end{aligned}$$

A simple computation yields $\mathbb{E}Z^4 \le 3(\mathbb{E}Z^2)^2$. The claim follows by plugging the last inequality in Equation (0.25). ■

Exercise 0.12 (Khintchine's inequality, cf. [2])**.** *Let F be vector space endowed with a norm. Let $\{x_i\}$ be a countable set of elements in F and $\{\epsilon_i : i \in \mathbb{N}\}$ be a set of independent Rademacher variables. Then, for any $1 < p < q < \infty$,*

$$\left\|\sum \epsilon_i x_i\right\|_p \le \left\|\sum \epsilon_i x_i\right\|_q \le \Big(\frac{q-1}{p-1}\Big)^{1/2} \left\|\sum \epsilon_i x_i\right\|_p$$

where $\|X\|_p := (\mathbb{E}\|X\|^p)^{1/P}$ for any F-value random variable X.

Exercise 0.13 (Latala's results on the L_p moments of sum of independent r.v., cf. [11])**.** *Let $X_1, \dots, X_n$ be a sequence of independent real random variables and let $S := \sum X_i$. We want precise bound on $\|S\|_p := \big(\mathbb{E}|\sum_i X_i|^p\big)^{1/p}$ for every $p > 0$. We introduce the following Orlicz-norm*

$$| \|(X_i)\| |_p := \inf \Big\{ t > 0 : \sum_{i=1}^{n} \log \Big(\phi_p\Big(\frac{X_i}{t}\Big)\Big) \leq p \Big\},$$

where $\phi_p(x) := \mathbb{E}\varphi_p(X)$ and $\varphi_p(x) := |1 + x|^p$.

If the X_i's are nonnegative then,

$$\frac{e-1}{2e^2} | \|(X_i)\| |_p \leq \|S\|_p \leq e | \|(X_i)\| |_p, \ \text{for } p \geq 1$$

and

$$\frac{(e-1)^{1/p}}{2e^2} | \|(X_i)\| |_p \leq \|S\|_p \leq e | \|(X_i)\| |_p, \ \text{for } p \leq 1.$$

If the X_i's are symmetric then, for $p \geq 2$,

$$\frac{e-1}{2e^2} | \|(X_i)\| |_p \leq \|S\|_p \leq e | \|(X_i)\| |_p, \ \text{for } p \leq 1. \tag{0.26}$$

If $X, X_1, \dots, X_n$ are i.i.d. and nonnegative then, for $p \geq 1$,

$$\|S\|_p \sim \sup \Big\{ \frac{p}{s} \Big(\frac{n}{p}\Big)^{1/s} \|X\|_s : \max\big(1, p/n\big) \leq s \leq p \Big\}.$$

If $X, X_1, \dots, X_n$ are i.i.d. and symmetric then, for $p \geq 2$,

$$\|S\|_p \sim \sup \Big\{ \frac{p}{s} \Big(\frac{n}{p}\Big)^{1/s} \|X\|_s : \max\big(1, p/n\big) \leq s \leq p \Big\}.$$

Note that for $X_1, \dots, X_n$ independent, mean zero random variables and $\epsilon_1, \dots, \epsilon_n$ independent Rademacher random variables independent of the X_i's, we have

$$\frac{1}{2} \Big\| \sum X_i \Big\|_p \leq \Big\| \sum \epsilon_i X_i \Big\|_p \leq 2 \Big\| \sum X_i \Big\|_p .$$

Hence, we obtain Result (0.26) for mean zero r.v. by setting $\phi_p(X_i) = \phi_p(\epsilon_i X_i) = \mathbb{E}\tilde{\varphi}_p(X_i)$ where $\tilde{\varphi}_p(x) := (1/2)(\varphi_p(x) + \varphi_p(-x))$.

Finally, Result (0.26) can be obtained for $p < 2$ thanks to Khintchine inequality saying that, for symmetric r.v. $X_1, \dots, X_n$,

$$c_p \Big\| \Big(\sum X_i^2 \Big)^{1/2} \Big\|_p \leq \Big\| \sum X_i \Big\|_p \leq \Big\| \Big(\sum X_i^2 \Big)^{1/2} \Big\|_p .$$

Exercise 0.14 (norms in $\mathbb{R}^n$)**.** *interpolated norms, ψ_α^n-norm,*

Exercise 0.15 ([10])**.** *Let $\alpha \geq 1$ and $p \geq \alpha$. There exists a constant $c_{\alpha,p} > 0$, depending only on α and p, such that the following holds. Let $X_1, \dots, X_n$ be ψ_α random variables and let $X_1^*, \dots, X_n^*$ be the non-increasing rearrangement of the absolute values of the X_i's. Then,*

$$\mathbb{E}\Big(\frac{1}{k} \sum_{i=1}^{k} X_i^{*p} \Big)^{1/p} \leq \log^{1/\alpha} \Big(\frac{c_{\alpha,p} n}{k} \Big).$$

Proof. Since the X_i's are ψ_α r.v. then,

$$\mathbb{E}\Big(\frac{1}{k} \sum_{i=1}^{k} \exp(X_i^{*\alpha}) \Big) \leq \mathbb{E}\Big(\frac{1}{k} \sum_{i=1}^{n} \exp(|X_i|^\alpha) \Big) \leq \frac{cn}{k}.$$

For every $a_1, \ldots, a_n$ real numbers such that $a_i \geq \big(p/\alpha - 1\big)^{p/\alpha} := c_0, \forall i$. Since $x \longmapsto \exp(x^{\alpha/p})$ is convex on $[c_0, \infty)$, we have

$$\exp\Big[\Big(\frac{1}{k}\sum_{i=1}^k a_i^p\Big)^{\alpha/p}\Big] \leq \frac{1}{k}\sum_{i=1}^k \exp(a_i^\alpha).$$

Repacing X_i^* by $\max(X_i^*, c_0)$, we get

$$\mathbb{E}\exp\Big[\Big(\frac{1}{k}\sum_{i=1}^k X_i^{*p}\Big)^{\alpha/p}\Big] \leq \mathbb{E}\Big(\frac{1}{k}\sum_{i=1}^k \exp(X_i^{*\alpha} + c_0^\alpha)\Big) \leq \frac{c_{\alpha,p} n}{k}.$$

The claim follows by using the concavity of the logarithm, the convexity of $x \longmapsto |x|^\alpha$ and Jensen's inequality. ∎

Exercise 0.16 (Gaussian variables)**.** *Let $g_1, \ldots, g_n$ be standard Gaussian variables $\mathcal{N}(0,1)$ and let $g_1^*, \ldots, g_n^*$ be the non-increasing rearrangement of the absolute values of the g_i's. Then,*

$\mathbb{E}\max_{1 \leq i \leq n} \frac{g_i}{\sqrt{i+1}} \leq c_0$, $\mathbb{E} g_i^*$, $\|g_i^*\|_{\psi_2}$, $\|g_i^*\|_{L_p}$

Let $X \sim \mathcal{N}_n(0, I_n)$ be a standard Gaussian variable with values in $\mathbb{R}^n$. $\big(\mathbb{E}|X|^p\big)^{1/p} \sim \big(\mathbb{E}|X|^2\big)^{1/2}$ if $p \leq n$ and $\big(\mathbb{E}|X|^p\big)^{1/p} \sim p^{1/2}$ if $p \geq n$.

Counter-example of a ψ_1 vector which is not a ....? Take $g \sim \mathcal{N}(0,1)$ and $X \sim \mathcal{N}_n(0,1)$ and set $Y := gX$. Then $\big(\mathbb{E}|Y|^p\big)^{1/p}/\big(\mathbb{E}|Y|^2\big)^{1/2}$ tends to infinity.

Moreover, if the g_i's are independent then,

$$\mathbb{E} g_i^* \geq c_0\sqrt{\log\big((2k)/i\big)}, \forall i \leq n/2 \text{ and } \mathbb{E} g_i^* \geq 1 - \frac{i}{k+1}, \forall n/2 \leq i \leq n$$

this result can be found in [6].

Exercise 0.17. *For all integers D and N with $1 \leq D \leq N$, the following inequality holds*

$$\sum_{i=1}^D \binom{N}{i} \leq \Big(\frac{eN}{D}\Big)^D. \tag{0.27}$$

Proof. The right-hand side being increasing with respect to D, it is larger than $(\sqrt{2e})^n > 2^n$ when $D \geq n/2$. Thus Equation (0.27) is trivial whenever $D \geq n/2$. We assume now that $D < n/2$. Take S to be a r.v. with Binomial distribution $\mathrm{Bin}(N, 1/2)$. Chernoff's inequality (cf. Exercise 0.4) implies that

$$\begin{aligned}\sum_{i=1}^D \binom{N}{i} &= 2^N \mathbb{P}[S \leq D] \leq \exp\big(N\log 2 - N h_{1/2}(D/N - \delta)\big)\\ &\leq \exp\big(N(D/N - (D/N)\log(D/N))\big).\end{aligned}$$

where $h_{1/2}$ is defined in Exercise 0.4 and satisfies $h_{1/2}(1-x) \geq \log 2 + x - x\log x, \forall x < 1$. This concludes the proof. ∎

Exercise 0.18 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i, \forall i$. Then, for every $t > 0$,*

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}X_i \geq t\Big], \mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i - X_i \geq t\Big] \leq \exp\Big(\frac{-2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\Big).$$

Proof. Let Y be a mean zero random variable such that $a \le Y \le b$ a.s. and $\lambda \ge 0$. By convexity, for any $x \in [a,b], \exp(\lambda x) \le \frac{x-a}{b-a}\exp(\lambda b) + \frac{b-x}{b-a}\exp(\lambda a)$. Thus,

$$\mathbb{E}\exp(\lambda Y) \le \frac{b}{b-a}\exp(\lambda a) - \frac{a}{b-a}\exp(\lambda b) \le \exp(\phi(\lambda(b-a))),$$

where $\phi(u) := -pu + \log(1-p+p\exp(u))$ for $p = -a/(b-a)$. It is easy to see that $\phi'(0) = 0$ and $\phi''(u) \le 1/4$. By a Taylor expansion, we get, for some $\theta \in [0, \lambda(b-a)]$, $\phi(\lambda(b-a)) = \phi(0) + \lambda(b-a)\phi'(0) + (\lambda^2(b-a)^2/2)\phi''(\theta) \le \lambda^2(b-a)^2/8$. We conclude with the classical Cramer-Chernoff method. ■

Exercise 0.19 (Hoeffding-Azuma's inequality)**.** *Let $(X_n)_{n\ge1}$ be a martingale with bounded differences ($|X_k - X_{k-1}| \le c_k, \forall k$). Then, for any integer n and any real number $t > 0$,*

$$\mathbb{P}\Big[X_n - X_0 > t\Big] \le \exp\Big(\frac{-t^2}{2\sum_{i=1}^n c_i}\Big).$$

Exercise 0.20 (Einmahl and Masson, cf.[5])**.** *Let $X_1, \dots, X_n$ be independent nonnegative random variables such that $\mathbb{E}X_i^2 \le \sigma^2$ then, for any $t > 0$,*

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n \big(\mathbb{E}X_i - X_i\big) \ge t\Big] \le \exp\Big(\frac{-t^2}{2\sigma^2}\Big).$$

Proof. For any $x \ge 0, \exp(-x) \le 1 - x + x^2/2$. Thus, for any $i = 1, \dots, n$ and $t > 0$,

$$\mathbb{E}\exp(-tX_i) \le \mathbb{E}\big(1 - tX_i + t^2X_i^2/2\big) \le \exp\big(-t\mathbb{E}X_i + t^2\mathbb{E}X_i^2/2\big).$$

Thus, for every $\lambda > 0$,

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^n \big(\mathbb{E}X_i - X_i\big) \ge t\Big] \le \exp(-\lambda t)\prod_{i=1}^n \mathbb{E}\exp\big((\lambda/n)(\mathbb{E}X_i - X_i)\big) \le \exp\big(-\lambda t + (\lambda^2/2n)\sigma^2\big).$$

The claim follows by optimizing the last equation in λ. ■

Exercise 0.21 (Maximal inequality for ψ_2 r.v.)**.** *There exists an absolute constant $c > 0$ such that the following holds. Let $X_1, \dots, X_n$ be ψ_2 random variables such that $\max_i \|X_i\|_{\psi_2} \le \sigma$. Then,*

$$\mathbb{E}\max_{i=1,\dots,n}|X_i| \le c\sigma\sqrt{\log(n)}.$$

Proof. By Jensen, for any $\lambda > 0$,

$$\exp\big(\lambda\mathbb{E}\max_i|X_i|\big) \le \mathbb{E}\exp\big(\lambda\max_i|X_i|\big) \le \mathbb{E}\sum_i \exp(\lambda|X_i|) \le c_1 n\exp\Big(c_0\lambda^2\sigma^2\Big).$$

Thus, $\mathbb{E}\max_i|X_i| \le \frac{\log(c_1 n)}{\lambda} + c_0\lambda\sigma^2$. The claim follows by optimizing in λ. ■

Exercise 0.22. *Let $X \in L_{\psi_\alpha}$ and $X_1, \dots, X_n$ be n i.i.d. copies of X. Then, w.h.p.*

$$\|(X_i)_1^n\|_{\psi_\alpha^n} \le c_0\|X\|_{\psi_\alpha}.$$

Exercise 0.23 (cf. [1])**.** *Let $(B, \|\cdot\|)$ be a separable Banach space and let $X_1, \dots, X_n$ be i.i.d. B-valued random variables such that $\|X_1\| \le 1$ almost surely. Let $\epsilon_1, \dots, \epsilon_n$ be independent Rademacher variables. Set*

$$Z := \mathbb{E}\left[\left\|\sum_{i=1}^n \epsilon_i X_1\right\| \Big| X_1^n\right].$$

For any $t > 0$,

$$\mathbb{P}[Z \ge \mathbb{E}Z + t] \le \exp\Big(\frac{-t^2}{4\mathbb{E}Z + 2t}\Big) \text{ and } \mathbb{P}[Z \le \mathbb{E}Z - t] \le \exp\Big(\frac{-t^2}{4(e-1)\mathbb{E}Z}\Big).$$

Proof. We consider the function $f : B^n \longmapsto \mathbb{R}$ defined by $f(x_1^n) := \mathbb{E}\,\|\sum_i \epsilon_i x_i\|$. We have $Z = f(X_1^n)$. Let $X'_1, \dots, X'_n$ be n independent copies of $X_1, \dots, X_n$ and set $Z^{(i)} := f(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)$for every $i = 1, \dots, n$.

Lemma 0.2. *Let $(x_n)_{n\geq 1}$ be a sequence in B and $(\epsilon_n)_{n\geq 1}$ be a sequence of independent Rademacher random variables. Then, the sequence $\left(\mathbb{E}\,\|\sum_{i=1}^n \epsilon_i x_i\| \,:\, n \geq 1\right)$ is non-decreasing.*

Proof. Let n be an integer. Let $(\epsilon_1, \dots, \epsilon_n) \in \{-1, 1\}^n$. By convexity of the norm, we have

$$\left\|\sum_{i=1}^n \epsilon_i x_i\right\| \leq \frac{1}{2}\left\|\sum_{i=1}^n \epsilon_i x_i + x_{n+1}\right\| + \frac{1}{2}\left\|\sum_{i=1}^n \epsilon_i x_i - x_{n+1}\right\| = \mathbb{E}_{\epsilon_{n+1}}\left\|\sum_{i=1}^n \epsilon_i x_i + \epsilon_{n+1} x_{n+1}\right\|.$$

The claim follows by taking the expectation in the last equation w.r.t. $(\epsilon_1, \dots, \epsilon_n)$. ■

Using Lemma 0.2, we have

$$Z^{(i)} \geq \mathbb{E}\left[\left\|\sum_{j=1, j\neq i}^n \epsilon_j X_j\right\| \Big| X_1^n\right]$$

and, since the X_i's are bounded, $Z - Z^{(i)} \leq 1$.

Moreover, let D denote a dense countable set in the unit ball of the dual B^* of B. For every choice of $\epsilon = (\epsilon_1, \dots, \epsilon_n) \in \{-1, 1\}^n$, the Hahn-Banach theorem implies that there exists some element $v_\epsilon \in D$ such that

$$\left\|\sum_{j=1}^n \epsilon_j X_j\right\| = \left\langle v_\epsilon, \sum_{j=1}^n \epsilon_j X_j\right\rangle$$

and for the same realization of the Rademacher variables,

$$\left\|\sum_{j=1, j\neq i}^n \epsilon_j X_j\right\| \geq \left\langle v_\epsilon, \sum_{j=1, j\neq i}^n \epsilon_j X_j\right\rangle$$

Hence, conditionally to X_1^n, we have

$$\begin{aligned}\sum_{i=1}^n (Z - Z^{(i)})^2 \mathbb{1}_{Z > Z^{(i)}} &\leq \sum_{i=1}^n \mathbb{E}_\epsilon \left\|\sum_{j=1}^n \epsilon_j X_j\right\| - \mathbb{E}_\epsilon \left\|\sum_{j=1, j\neq i}^n \epsilon_j X_j\right\| \\ &\leq \mathbb{E}_\epsilon \sum_{i=1}^n \langle v_\epsilon, \epsilon_i X_i\rangle = \mathbb{E}_\epsilon \left\|\sum_{i=1}^n \epsilon_i X_i\right\|.\end{aligned}$$

The first part of the claim follows by applying Theorem 0.12 with $a = 1$ and $b = 0$. The proof of the second part of the claim follows the same line as the first part together with Theorem 0.13. ■

Bibliography

[1] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.

[2] Víctor H. de la Peña and Evarist Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

[3] S. J. Dilworth and S. J. Montgomery-Smith. The distribution of vector-valued Rademacher series. *Ann. Probab.*, 21(4):2046–2052, 1993.

[4] R. M. Dudley. *Uniform central limit theorems*, volume 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999.

[5] Uwe Einmahl and David M. Mason. Some universal results on the behavior of increments of partial sums. *Ann. Probab.*, 24(3):1388–1407, 1996.

[6] Yehoram Gordon, Alexander Litvak, Carsten Schütt, and Elisabeth Werner. Orlicz norms of sequences of random variables. *Ann. Probab.*, 30(4):1833–1853, 2002.

[7] Marjorie G. Hahn. Conditions for sample-continuity and the central limit theorem. *Ann. Probability*, 5(3):351–360, 1977.

[8] Paweł Hitczenko. Domination inequality for martingale transforms of a Rademacher sequence. *Israel J. Math.*, 84(1-2):161–178, 1993.

[9] W. B. Johnson, G. Schechtman, and J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.*, 13(1):234–253, 1985.

[10] B. Klartag. $5n$ Minkowski symmetrizations suffice to arrive at an approximate Euclidean ball. *Ann. of Math. (2)*, 156(3):947–960, 2002.

[11] Rafał Latała. Estimation of moments of sums of independent real random variables. *Ann. Probab.*, 25(3):1502–1513, 1997.

[12] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[13] S. J. Montgomery-Smith. The distribution of Rademacher sums. *Proc. Amer. Math. Soc.*, 109(2):517–522, 1990.

[14] M. Schmuckenschlaeger. Bernstein inequalities for a class of random variables. *Proc. Amer. Math. Soc.*, 117(4):1159–1163, 1993.

[15] Daniel W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993.

[16] Michel Talagrand. The supremum of some canonical processes. *Amer. J. Math.*, 116(2):283–325, 1994.

[17] Sara A. van de Geer. *Applications of empirical process theory*, volume 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2000.

[18] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.