CHAPITRE 3

NORMES ψ_α ET INÉGALITÉS DE TYPE KHINTCHINE

3.1. Normes ψ_α : formulations équivalentes

Dans ce chapître $\alpha \in [1, +\infty[$ est une constante donnée. Les deux cas qui nous intéresserons le plus dans la suite sont

$$\alpha = 1 \quad \text{et} \quad \alpha = 2.$$

On se donne une probabilité μ sur un espace mesurable X.

On introduit

$$L_{\psi_\alpha}(\mu) := \left\{ f : X \to \mathbb{R} \ ; \ \exists c > 0, \ \int e^{c|f|^\alpha} \, d\mu < +\infty \right\}$$

et pour $f : X \to \mathbb{R}$,

$$\|f\|_{\psi_\alpha(\mu)} := \inf \left\{ \lambda > 0 \ ; \ \int e^{\left(\frac{|f|}{\lambda}\right)^\alpha} \, d\mu \leq 2 \right\}$$

avec la convention $\inf \emptyset = +\infty$. Remarquez qu'on a

$$L_{\psi_\alpha}(\mu) = \left\{ f : X \to \mathbb{R} \ ; \ \|f\|_{\psi_\alpha(\mu)} < +\infty \right\}.$$

Plus précisément :

Exercice. — *Soit $f : X \longrightarrow \mathbb{R}$ et $c, C > 0$ deux constantes tel que*

$$\int e^{\left(\frac{|f|}{c}\right)^\alpha} \, d\mu \leq C.$$

Montrer qu'il existe une constante $\tilde{c}$ dépendant seulement de $\{c, C, \alpha\}$ (et donc ni de f ou de μ) tel que

$$\int e^{\left(\frac{|f|}{\tilde{c}}\right)^\alpha} \, d\mu \leq 2.$$

Mieux : montrer que $\|f\|_{\psi_\alpha(\mu)} \leq c \, (\log(C))^{1/\alpha}$.

Remarquez que pour $f \in L_{\psi_\alpha}(\mu)$ non nul (μ-pp), l'infimum dans la définition de la norme devient un un minimum (par le théorème de convergence monotone) :

$$\int e^{\left(\frac{|f|}{\|f\|_{\psi_\alpha(\mu)}}\right)^\alpha} \, d\mu \leq 2$$

Si $f \in L_{\psi_\alpha}(\mu)$ on dit parfois que f est $\psi_\alpha(\mu)$.

Les notations précédentes proviennent de la théorie de espaces de Orlicz. L'espace L_{ψ_α} est en effet l'espace de Orlicz associé à la fonction

$$\psi_\alpha(t) := e^{t^\alpha} - 1,$$

qui vérifie les propriétés suivantes : $\psi_\alpha : [0,+\infty] \to [0,+\infty]$ est continue, strictement croissante et convexe avec $\psi_\alpha(0) = 0$ et $\psi_\alpha(+\infty) = +\infty$. On a alors

$$\|f\|_{\psi_\alpha(\mu)} = \inf \left\{ \lambda > 0 \; ; \; \int \psi_\alpha\left(\frac{|f|}{\lambda}\right) d\mu \leq 1 \right\}.$$

À titre culturel, mentionnons le résultat classique suivant :

Proposition. — *Avec les définitions ci-dessus, $L_{\psi_\alpha}(\mu)$ est un espace vectoriel et $\|\cdot\|_{\psi_\alpha(\mu)}$ est une norme sur cet espace. En fait, $(L_{\psi_\alpha}(\mu), \|\cdot\|_{\psi_\alpha(\mu)})$ est un Banach.*

Introduisons

$$N_{\psi_\alpha(\mu)}(f) := \sup_{p\in\mathbb{N}^*} \frac{\|f\|_{L_p(\mu)}}{p^{1/\alpha}}.$$

Notez que, puisque les normes $\|\cdot\|_{L_p(\mu)}$ sont croissantes (comme fonction de $p \in]0,+\infty[$), on peut aussi bien définir le supremum pour les réels $p \in [1,+\infty[$, quitte à perdre une constante :

$$N_{\psi_\alpha(\mu)}(f) \leq \sup_{p\geq 1} \frac{\|f\|_{L_p(\mu)}}{p^{1/\alpha}} \leq 2^{1/\alpha} N_{\psi_\alpha(\mu)}(f) \leq 2N_{\psi_\alpha(\mu)}(f). \tag{3.1}$$

Il pourra nous arriver dans la suite de confondre les suprema sur $\mathbb{N}^*$ et sur $[1,+\infty[$.

On va montrer que (de manière universelle)

$$\|\cdot\|_{\psi_\alpha(\mu)}(\cdot) \simeq N_{\psi_\alpha(\mu)}(\cdot).$$

Plus précisemment :

Lemme. — *Pour toute probabilité μ et toute fonction $f \in L_{\psi_\alpha}(\mu)$ on a*

$$\frac{1}{2} N_{\psi_\alpha(\mu)}(f) \leq \|f\|_{\psi_\alpha(\mu)} \leq 4e\, N_{\psi_\alpha(\mu)}(f).$$

En particulier, pour toute fonction f on a

$$\forall p \geq 1, \qquad \|f\|_{L^p(\mu)} \leq 2\, p^{1/\alpha}\, \|f\|_{\psi_\alpha(\mu)}.$$

On voit qu'on a les inclusions (continues) suivantes : pour tout $p \geq 1$,

$$L^\infty(\mu) \subset L_{\psi_\alpha}(\mu) \subset L^p(\mu).$$

Démonstration. — Nous allons montrer les inégalités suivantes :

$$\frac{1}{2^{1/\alpha}} N_{\psi_\alpha(\mu)}(f) \leq \|f\|_{\psi_\alpha(\mu)} \leq (4e\alpha)^{1/\alpha} N_{\psi_\alpha(\mu)}(f),$$

ce qui implique le résultat annoncé puisque, α étant plus grand que 1, on a $(\alpha e)^{1/\alpha} \leq e$ et $\frac{1}{2^{1/\alpha}} \geq \frac{1}{2}$. Pour l'inégalité de gauche, fixons $p \geq 1$ quelconque et remarquons qu'en utilisant l'inégalité de Jensen (μ proba) on a :

$$2 \geq \int e^{(f/\|f\|_{\psi_\alpha})^\alpha} d\mu \geq \int \frac{|f|^{\alpha p}}{p!\|f\|_{\psi_\alpha}^{\alpha p}} d\mu \geq \frac{\|f\|_{L_p(\mu)}^{\alpha p}}{p!\|f\|_{\psi_\alpha}^{\alpha p}}$$

et donc, en utilisant que $p^p \geq p!$,

$$2^{1/\alpha} \geq 2^{1/\alpha p} \geq \frac{\|f\|_{L_p(\mu)}}{p^{1/\alpha}\|f\|_{\psi_\alpha}},$$

ce qui donne le résultat annoncé. Pour l'inégalité inverse, fixons $\lambda \in \mathbb{R}$ et écrivons

$$\int e^{(f/\lambda)^\alpha}\, d\mu = \sum_{n=0}^{\infty} \frac{\|f\|_{L_{\alpha n}(\mu)}^{\alpha n}}{\lambda^{\alpha n}\, n!}.$$

On utilise alors l'estimation facile suivante (exercice)

$$\forall n \geq 1, \qquad n! \geq \left(\frac{n}{e}\right)^n.$$

On a alors (en toute rigueur on utilise (3.1) pour passer à un p non entier)

$$\int e^{(f/\lambda)^\alpha}\, d\mu \leq \sum_{n=0}^{\infty} \left(\frac{2\, e\, \alpha\, N_{\psi_\alpha(\mu)}(f)^\alpha}{\lambda^\alpha}\right)^n,$$

et cette série géométrique est de somme inférieure à 2 si

$$\frac{2e\, \alpha\, N_{\psi_\alpha(\mu)}(f)^\alpha}{\lambda^\alpha} \leq \frac{1}{2}$$

ce qui est bien le cas pour

$$\lambda = (4e\alpha)^{1/\alpha} N_{\psi_\alpha(\mu)}(f).$$

□

On reformule et on complète le Lemme précédent comme suit.

Proposition. — *Étant donné une constante $c > 0$ et une fonction $f : X \to \mathbb{R}$, considérons les deux propriétés suivantes :*

$$\begin{aligned}
P(c) \quad &: \quad \int e^{(|f|/c)^\alpha}\, d\mu \leq 2 \quad (\textit{c'est-à-dire } \|f\|_{\psi_\alpha(\mu)} \leq c) \\
Q(c) \quad &: \quad \forall p \geq 1, \quad \|f\|_{L_p(\mu)} \leq c\, p^{1/\alpha} \\
R(\lambda_0, C, c) \quad &: \quad \forall r \geq \lambda_0, \quad \mu\left(\{|f| \geq r\}\right) \leq C\, e^{-r^\alpha/c^\alpha}
\end{aligned}$$

Alors ces propriétés sont équivalentes au sens suivant :

$$P(c) \Longrightarrow Q(2c) \quad \textit{et} \quad Q(c) \Longrightarrow P(4e\, c),$$

ainsi que

$$P(c) \Longrightarrow R(0, 2, c) \quad \textit{et} \quad R(\lambda_0, C, c) \Longrightarrow P(K) \quad \textit{où } K \textit{ ne dépend que de } \{\lambda_0, C, c\}.$$

Démonstration. — Seule les dernières implications reste à démontrer. L'implication $P(c) \Rightarrow R(0, 2, c)$ est immédiate. pour la dernière on rappelle le fait élémentaire suivant.

Fait. — *Soit (X, μ) un espace mesuré.*

1. *Si $g : X \to \mathbb{R}^+$ est une fonction mesurable positive et $q > 0$, alors on a*

$$\int g^q\, d\mu = q \int_0^{+\infty} \mu\left(\{g \geq r\}\right) r^{q-1}\, dr.$$

2. *Soit $\phi : [0,+\infty[\to \mathbb{R}$ une fonction continue sur $[0,+\infty[$ et C^1 sur $]0,+\infty[$. Si $g : X \longrightarrow \mathbb{R}^+$ est une fonction positive (avec $\phi(g)-\phi(0)$ μ-intégrable), alors on a*
$$\int [\phi(g)-\phi(0)]\,d\mu = \int_0^{+\infty} \mu(\{g \ge r\})\,\phi'(r)\,dr.$$

Soit $\kappa > c$ que l'on fixera plus tard. On a donc
$$\begin{aligned}
\int e^{\left(\frac{|f|}{\kappa}\right)^\alpha} d\mu &= 1 + \kappa^{-\alpha}\alpha \int_0^{+\infty} \mu(\{|f| \ge r\})\, r^{\alpha-1} e^{\kappa^{-\alpha} r^\alpha}\, dr \\
&\le e^{\kappa^{-\alpha}\lambda_0} + \kappa^{-\alpha}\alpha \int_{\lambda_0}^{+\infty} \mu(\{|f| \ge r\})\, r^{\alpha-1} e^{\kappa^{-\alpha} r^\alpha}\, dr \\
&\le e^{c^{-\alpha}\lambda_0} + \kappa^{-\alpha}\alpha C \int_{\lambda_0}^{+\infty} r^{\alpha-1} e^{-(c^{-\alpha}-\kappa^{-\alpha}) r^\alpha}\, dr. \\
&= e^{\kappa^{-\alpha}\lambda_0} + \frac{\kappa^{-\alpha}\alpha C e^{-(c^{-\alpha}-\kappa^{-\alpha})\lambda_0^\alpha}}{c^{-\alpha}-\kappa^{-\alpha}},
\end{aligned}$$
et cette dernière quantité est ≤ 2 à condition de prendre κ suffisamment grand (plus grand qu'une borne – que l'on pourrait calculer explicitement, mais cela n'aurait guère d'intérêt – ne dépendant que de $C, \tilde{C}, \lambda_0$). On pourrait aussi invoquer directement le résultat de l'exercice 3.1. □

Enfin, il y a une quatrième formulation équivalente du caractère ψ_α, qui repose sur l'estimation de la transformée de Laplace. La méthode de Laplace est justement une des méthodes principales pour obtenir des estimées ψ_α. Nous la verrons en action un peu plus loin.

Il faut être particulièrement vigilant au fait que ces équivalences sont données sous forme « non normalisées ». En pratique, on cherche des inégalités avec des constantes universelles pour une certaine classe de fonctions. Or typiquement, dans la Proposition, la constante c dépendra de f. Cela n'est pas très grave car les dépendances sont ici linéaires (on passe de c à $2c$, par exemple). Cependant, la bonne manière de faire est d'appliquer la proposition à une forme normalisée de f (pour laquelle la constante c sera universelle). Par exemple, en appliquant la Proposition à
$$\frac{f}{\|f\|_{L_1(\mu)}}.$$
on obtient :

Proposition. — *Soit $c > 0$ une constante donnée. On se donne une fonction f intégrable telle que*
$$\int \exp\left[\left(\frac{|f|}{c\|f\|_{L_1(\mu)}}\right)^\alpha\right] d\mu \le 2, \tag{3.2}$$
c'est-à-dire $\|f\|_{\psi_\alpha(\mu)} \le c\|f\|_{L_1(\mu)}$. Alors on a :
$$\forall p \ge 1, \qquad \|f\|_{L_p(\mu)} \le 2c\,p^{1/\alpha}\,\|f\|_{L_1(\mu)} \tag{3.3}$$
La réciproque est également vraie, modulo un ajustement de la constante (comme dans le Lemme précédent).

Rappelons qu'on a toujours, par Jensen ou Hölder, puisque μ est une probabilité,
$$\forall p \ge 1, \qquad \|f\|_{L_1(\mu)} \le \|f\|_{L_p(\mu)},$$

Ainsi, l'inégalité (3.3) s'apparente à une forme inverse de l'inégalité de Hölder ; on dit que c'est une inégalité de « type Khintchine ». On cherche donc en général le plus grand $\alpha > 0$ possible pour lequel ces inégalités sont vérifiées.

Remarquez que pour toute fonction f,

$$(3.2) \implies \forall p \in [1,2], \quad \|f\|_{L_1(\mu)} \leq \|f\|_{L_p(\mu)} \leq 4c\, \|f\|_{L_1(\mu)} \tag{3.4}$$

ce qui donne en particulier l'équivalence entre les normes L_1 et L_2.

Parfois, on préfère normaliser par la norme L_2 car le calcul de $\|f\|_{L_2(\mu)}$ peut s'avérer plus aisé. Supposons que l'on ait

$$\int \exp\left[\left(\frac{|f|}{c\|f\|_{L_2(\mu)}}\right)^\alpha\right] d\mu \leq 2,$$

ce qui est une condition plus faible que (3.2). La Proposition (3.1) dit qu'on a alors

$$\forall p \geq 1, \qquad \|f\|_{L_p(\mu)} \leq 2c\, p^{1/\alpha}\, \|f\|_{L_2(\mu)}.$$

On peut remarquer que pour $p \in [1,2]$, on a toujours l'inégalité triviale $\|f\|_{L_p(\mu)} \leq \|f\|_{L_2(\mu)}$. En ce sens, l'inégalité (3.3) est plus précise puisqu'elle donne aussi une inégalité inverse pour $p \in [1,2]$. Cependant, on peut récupérer une inégalité inverse pour $p \in [1,2]$ en interpolant entre les valeurs $p, 2, 4$ comme suit. Pour $p \in [1,2]$, soit $\theta \in [0,1]$ tel que

$$\frac{1}{2} = \frac{1-\theta}{p} + \frac{\theta}{4}$$

(On a $\theta = 2(2-p)/(4-p)$). Alors par l'inégalité de Hölder et l'inégalité précédente pour $p = 4$, on a

$$\|f\|_{L_2(\mu)} \leq \|f\|_{L_p(\mu)}^{1-\theta} \|f\|_{L_4(\mu)}^{\theta} \leq (2c\, 4^{1/\alpha})^\theta \|f\|_{L_p(\mu)}^{1-\theta} \|f\|_{L_2(\mu)}^{\theta},$$

soit encore

$$\|f\|_{L_p(\mu)} \geq \frac{1}{(2c\, 4^{1/\alpha})^{\theta/(1-\theta)}} \|f\|_{L_2(\mu)} \geq \frac{1}{(8c)^{\frac{4}{p}-2}} \|f\|_{L_2(\mu)} \geq \frac{1}{(8c)^2} \|f\|_{L_2(\mu)}.$$

Notez qu'on a encore l'équivalence des normes L_1 et L_2. Enonçons la version L_2 sous forme de Proposition :

Proposition. — *Soit $c > 0$ une constante donnée. On se donne une fonction $f \in L_2(\mu)$ telle que*

$$\int \exp\left[\left(\frac{|f|}{c\|f\|_{L_2(\mu)}}\right)^\alpha\right] d\mu \leq 2, \tag{3.5}$$

c'est-à-dire $\|f\|_{\psi_\alpha(\mu)} \leq c\|f\|_{L_2(\mu)}$. Alors on a :

$$\begin{cases} \forall p \geq 2, & \|f\|_{L_p(\mu)} \leq 2c\, p^{1/\alpha}\, \|f\|_{L_2(\mu)} \\ \forall p \in [1,2], & \frac{1}{(8c)^2} \|f\|_{L_2(\mu)} \leq \|f\|_{L_p(\mu)} \end{cases} \tag{3.6}$$

Puisque (3.6) entraîne (3.3) (avec des constantes légèrement différentes), on voit que la condition (3.5), qui est plus faible que la condition (3.2), lui est en fait équivalente. En d'autres termes

$$\|f\|_{\psi_2(\mu)} \leq c\|f\|_{L_2(\mu)} \Longrightarrow \|f\|_{\psi_2(\mu)} \leq C(c)\|f\|_{L_1(\mu)},$$

mais je ne vois pas de manière directe de le voir, sans passer par le comportement des puissances p.

Notez que l'argument précédent indique que la deuxième inégalité de (3.6) se déduit de la première (appliqué avec $p = 4$) et de Hölder. En fait, cet argument montre que la deuxième inégalité est valable pour $p \in]0, 2]$, mais la constante $\frac{1}{(8c)^{\frac{4}{p}-2}}$ explose lorsque $p \to 0$ (on ne peut trouver par cette méthode une constante numérique valable pour tous les $p \in [0, 1]$).

Exemple. — *Le calcul de la norme $\psi_2(\gamma_n)$ donne, en utilisant que $e^t \geq 1 + t/2$ pour $t \in [-1.5, 0]$*

$$\|f\|_{\psi_2(\gamma_n)} = \sqrt{2}\big[1 - e^{-2\log(2)/n}\big]^{-1/2} \leq \sqrt{\frac{2n}{\log(2)}} \leq \sqrt{2}\sqrt{n} = \sqrt{2}\|f\|_{L_2(\gamma_n)}.$$

On a donc en particulier, si X est un vecteur gaussien standard dans $\mathbb{R}^n$,

$$\mathbb{E}|X| \simeq \sqrt{n}.$$

Les propositions précédentes donnent la meilleure dépendance (asymptotique) possible en p dans (3.3) sous l'hypothèse (3.2) ou (3.5), comme le montre l'exercice suivant :

Exercice. — *On travaille avec la mesure gaussienne γ_n sur l'espace euclidien $\mathbb{R}^n$. On étudiera successivement le cas des fonctions*

$$f(x) = x \cdot \theta \ (|\theta| = 1 \text{ fixé}) \qquad \text{et} \qquad f(x) = |x|.$$

1. *Quel est le plus grand α pour lequel $f \in L_{\psi_\alpha}(\gamma_n)$.*
2. *Montrer que pour $p \to \infty$:*

$$\log\left(\frac{\|f\|_{L_{2p}(\gamma_n)}}{\|f\|_{L_2(\gamma_n)}}\right) \sim \log(\sqrt{2p})$$

Nous insistons encore une fois que, dans les équivalences entre différentes caractérisations du caractère ψ_α, il convient de se ramener au cas où les constantes $C, \tilde{C}, \lambda_0$ sont « universelles » (en général ce sont de braves constantes numériques). Voici par exemple une reformutation avec une normalisation L_2.

3.2. Lemme de Borell et estimées ψ_1 pour les mesures log-concaves

***Proposition* (Lemme de Borell).** — *Soit μ une probabilité log-concave sur $\mathbb{R}^n$. Alors pour tout ensemble convexe symétrique $A \subset \mathbb{R}^n$ avec $\mu(A) =: a > 0$ on a,*

$$\forall r \geq 1, \qquad \mu\big(\mathbb{R}^n \setminus rA\big) \leq a\left(\frac{1-a}{a}\right)^{\frac{r+1}{2}}.$$

En particulier, si A est un ensemble convexe symétrique tel que $\mu(A) \geq \frac{3}{4}$, on a, pour tout $r \geq 1$,

$$\mu\big(\mathbb{R}^n \setminus rA\big) \leq e^{-r/2} \tag{3.7}$$

Démonstration. — Pour $B \subset \mathbb{R}^n$, notons $B^c := \mathbb{R}^n \setminus B$ le complémentaire. Soit donc A un ensemble convexe symétrique (d'intérieur non-vide). Sans perte de généralité, on peut supposer que A est de plus fermé. On a, pour tout $r > 1$,

$$A^c \supset \frac{2}{r+1}(rA)^c + \frac{r-1}{r+1}A.$$

Pour le voir, on peut considérer la jauge p_A de A, qui est une semi-norme paire. On a que $A = \{p_A \leq 1\}$ et plus généralement que $rA = \{p_A \leq r\}$. Pour tout $x \in (rA)^c$ et $y \in A$ on a

$$p_A\left(\frac{2}{r+1}x + \frac{r-1}{r+1}y\right) \geq p_A\left(\frac{2}{r+1}x\right) - p_A\left(-\frac{r-1}{r+1}y\right) > \frac{2}{r+1} \times r + \frac{r-1}{r+1} \geq 1.$$

L'inégalité de Brunn-Minkowski pour μ nous donne

$$\mu(A^c) \geq \big(\mu(rA)\big)^{2/(r+1)}\big(\mu(A)\big)^{(r-1)/(r+1)}$$

ce qui revient au résultat annoncé. Lorsque $a \in [\frac{3}{4}, 1]$ on a

$$a\left(\frac{1-a}{a}\right)^{\frac{r+1}{2}} \leq \left(\frac{1}{3}\right)^{\frac{r+1}{2}} \leq e^{-\frac{r+1}{2}\log(3)} \leq e^{-r/2}.$$

□

On va en déduire que les semi-normes sont "universellement" ψ_1 pour les mesures log-concaves.

Théorème. — *Soit μ une probabilité log-concave sur $\mathbb{R}^n$. Alors, pour toute semi-norme paire $F\mathbb{R}^n \to \mathbb{R}^+$ on a, pour tout*

$$\forall q \geq p \geq 1, \qquad \left(\int F^q \, d\mu\right)^{1/q} \leq c\,\frac{q}{p}\left(\int F^p \, d\mu\right)^{1/p}$$

où $c > 0$ est une constante numérique. En particulier, on a

$$\|F\|_{\psi_1(\mu)} \leq c\|F\|_{L^1(\mu)},$$

et si on note $m_F = \int F \, d\mu$,

$$\forall r \geq 1, \qquad \mu\left(\{F \geq C\, m_F\, r\}\right) \leq e^{-r/2}.$$

avec $C = 2ce$.

Ce qui est remarquable, c'est que l'estimée ne dépende ni de n, μ ou F. En fait, on va établir l'inégalité de déviation correspondante :

Lemme. — *Soit μ une probabilité log-concave sur $\mathbb{R}^n$. Si $F : \mathbb{R}^n \to \mathbb{R}^+$ est une semi-norme paire. Alors, pour tout $p \geq 1$, si on note pour $m_p := \big(\int F^p \, d\mu\big)^{1/p}$ on a*

$$\forall r \geq 1, \qquad \mu\left(\{F \geq 4m_p\, r\}\right) \leq e^{-pr/2}.$$

Cette inégalité, appliquée avec $p = 1$, donne directement le caractère ψ_1 du théorème ci-dessus. La même méthode permet d'avoir l'estimation plus précise pour $q \geq p \geq 1$ (Exercice).

Preuve du Lemme. — Introduisons l'ensemble $A := \{F < \frac{4m}{p}\}$. Alors A est convexe symétrique. De $m^p \geq \displaystyle\int_{A^c} F^p \, d\mu \geq \frac{4^p p}{m}\,\mu(A^c)$ on tire que

$$a := \mu(A) \geq 1 - \frac{1}{4^p} \geq \frac{3}{4}$$

et donc que

$$\frac{1-a}{a} \leq \frac{4^{-p}}{1-4^{-p}} \leq 3^{-p}.$$

Par ailleurs, pour $r \geq 1$ on a, par homogénéité de F,

$$\{F < 4mr\} = \{x \in \mathbb{R}^n \ ; \ F(\frac{x}{r}) < 4m\} = rA$$

et donc d'après le Lemme de Borell, on a

$$\mu\left(\{F \geq 4m\, r\}\right) = \mu((rA)^c) \leq a\left(\frac{1-a}{a}\right)^{\frac{r+1}{2}} \leq 1 \times \left(3^{-p}\right)^{\frac{r+1}{2}} = e^{-p\log(3)(r+1)/2} \leq e^{-pr/2}.$$

□

Le cas typique d'application est lorsqu'on prend pour μ la mesure de Lebesque restreinte à un corps convexe et pour F une des fonction suivantes :

$$F(x) = |x| \quad \text{ou} \quad F(x) = x \cdot \theta \ (\theta \text{ fixé}).$$

Par exemple, si K est un corps convexe de volume 1 ($K| = 1$) on a : pour toute direction $\theta \in S^{n-1}$

$$\left(\int_K |x\cdot\theta|^p\, dx\right)^{1/p} \leq cp\int_K |x\cdot\theta| dx \leq cp\left(\int_K |x\cdot\theta|^2\, dx\right)^{1/2}. \tag{3.8}$$

et

$$\left(\int_K |x|^p\, dx\right)^{1/p} \leq cp\int_K |x|\, dx \leq cp\left(\int_K |x|^2\, dx\right)^{1/2}.$$

En plaçant K dans une « bonne » position, on peut remplacer cette dernière estimée ψ_1 par une estimée ψ_2.

3.3. Inégalités gaussiennes et sphériques

Tout d'abord, on peut utiliser que γ_n, la mesure gaussienne standard sur $\mathbb{R}^n$ est une mesure log-concave sur $\mathbb{R}^n$. Ainsi, on sait que pour toute semi-norme paire on a, pour $q \geq p \geq 1$,

$$\|f\|_{L^p(\gamma_n)} \leq \|f\|_{L^q(\gamma_n)} \leq c\frac{q}{p}\|f\|_{L^p(\gamma_n)}.$$

Voici une application intéressante de ce résultat.

Proposition. — *Soit F une semi-norme paire sur $\mathbb{R}^n$. Alors pour $q \geq p \geq 1$ on a*

$$\left(\int_{S^{n-1}} F(u)^q\, d\sigma(u)\right)^{1/q} \leq c\,\frac{q}{p}\sqrt{\frac{n+p}{n+q}}\left(\int_{S^{n-1}} F(u)^p\, d\sigma(u)\right)^{1/p}$$

et en particulier pour tout $p \geq 2$

$$\|F\|_{L^p(\sigma)} \leq C\min\{p, \sqrt{n}\sqrt{p}\}\|F\|_{L^2(\sigma)}.$$

La borne pour $\|F\|_{L^p(\sigma)}$ est d'ordre p pour $p \leq n$ (comportement ψ_1) et d'ordre $\sqrt{p}$ pour $p \geq n$ (comportement ψ_2) mais avec une borne qui dépend de n. Tout comme dans le Lemme de Borell, il est remarquable que l'estimation ne dépende pas de F. Prenons par exemple $F(u) = |Au|$ pour une matrice symétrique positive A. On peut donc estimer $\|F\|_{L^p(\sigma)}$ en fonction de la norme de Hilbert-Schmidt,

$$\|F\|_2 = \left(\int |Au|^2\, d\sigma(u)\right)^{1/2} = \frac{1}{n}\|A\|_{HS}.$$

Les inégalités de concentration donneront une dépendance en fonction de la constante de Lipschitz, qui est la norme d'opérateur $\|A\|_{op}$, en général plus grande que $\frac{1}{n}\|Au\|_{HS}$, mais ce comportement sera toujours ψ_2. Donc il faut regarder en détail suivant les valeurs de p...

Démonstration. — En intégrant en coordonnées polaires, on remarque (Exercice) que pour une fonction F 1-homogène on a

$$\left(\int_{R^n} |F(x)|^q \, d\gamma_n(x)\right)^{1/q} \simeq \sqrt{n+q} \left(\int_{S^{n-1}} |F(u)|^q \, d\sigma(u)\right)^{1/q},$$

où par $\simeq$ on entend un encadrement par des constantes unverselles. □

Le cas gaussien n'est pas toujours intéressant car les quantités se calculent explicitement, comme vu à l'exercice 3.1. Nous le réécrivons rapidement simplement car cela permet de voir le type d'inégalité qui nous intéresse et le raisonnement pour y aboutir.

Soit $u \in \mathbb{R}^n$ un vecteur non-nul fixé. Dans la suite, $g_1, \ldots, g_n$ désigneront des variables aléatoires gaussiennes standard indépendantes, de sorte $X = (g_1, \ldots, g_n)$ est un vecteur gaussien standard de $\mathbb{R}^n$. La fonction

$$f_u(x) = x \cdot u$$

appartient à $L_{\psi_2}(\gamma_n)$ et on a

$$\int e^{(x\cdot u/(2|u|))^2} \, d\gamma_n(x) = \int e^{(x\cdot \frac{u}{|u|})^2/4} \, d\gamma_n(x) = \int_{\mathbb{R}} e^{x_1^2/4} \gamma(x_1) = \sqrt{2} \le 2,$$

ce qui donne que

$$\|f_u\|_{\psi_2(\gamma_n)} \le 2|u| = 2\|f_u\|_{L_2(\gamma_n)}.$$

En fait, le calcul donne exactement

$$\|f_u\|_{\psi_2(\gamma_n)} = \frac{2}{\sqrt{3}}|u| = \frac{2}{\sqrt{3}}\|f_u\|_{L_2(\gamma_n)}.$$

On peut écrire

$$\|f_u\|_{L_p(\gamma_n)} = (\mathbb{E}\,|X\cdot u|^p)^{1/p} = \left(\mathbb{E}\left|\sum_{i=1^n} u_i \, g_i\right|^p\right)^{1/p}$$

Ainsi, par la Proposition 3.1 on a :

Proposition. — *Soit $g_1, \ldots, g_n$ des variables aléatoires gaussiennes standard indépendantes. Alors pour tout $a_1, \ldots, a_n \in \mathbb{R}$ on a :*

$$\forall p \ge 2, \quad \left(\mathbb{E}\left|\sum_{i=1}^n a_i \, g_i\right|^p\right)^{1/p} \le 4\sqrt{p} \left(\mathbb{E}\left|\sum_{i=1}^n a_i \, g_i\right|^2\right)^{1/2}.$$

Évidemment, ce que nous venons de faire est iun peu ridicule, puisque ce résultat s'obtient par un calcul direct ; on est même ramené à un calcul en dimension 1 (ce qui limite en fait l'intérêt de cette observation) puisque

$$\|f_u\|_{L_p(\gamma_n)} = \gamma(p)|u| \quad \text{avec} \quad \gamma(p) = \|f_{e_1}\|_{L_p(\gamma_n)} = \left(\int |t|^p \, d\gamma_1(t)\right)^{1/p}.$$

Exercice. — *Soit* $\{g_1, \dots, g_n\}$ *des variables aléatoires gaussiennes standards (pas nécessairement indépendantes). En utilisant le caractère* ψ_2 *de* g_i *(c'est-à-dire que de* $t \to t$ *est dans* $L_{\psi_2}(\gamma_1)$ *avec une norme* ψ_2 *contrôlée par la norme* L_2*) et le fait que* ℓ_∞^n *est universellement équivalent à une certaine norme* ℓ_p^n*, retrouver que*

$$\mathbb{E} \sup_{1 \le i \le n} |g_i| \le C \sqrt{\log n}$$

pour une certaine constante numérique $C > 0$.

On peut suivre la même chaîne de déduction pour la fonction

$$f(x) = |x|.$$

Proposition. — *Soit* $g_1, \dots, g_n$ *des variables aléatoires gaussiennes standard indépendantes. Alors si* $e_1, \dots, e_n$ *est une base orthonormée de* $\mathbb{R}^n$ *on a*

$$\forall p \ge 2, \quad \left(\mathbb{E} \left| \sum_{i=1}^n g_i\, e_i \right|^p \right)^{1/p} \le 4\sqrt{p} \left(\mathbb{E} \left| \sum_{i=1}^n g_i\, e_i \right|^2 \right)^{1/2} = 4\sqrt{p}\,\sqrt{n}.$$

Ici encore, tout se calcule explicitement, mais cette fois-ci les calculs sont n-dimensionnels. On voit ici un trait important des inégalités de type Khintchine : l'inégalité est indépendante de n (c'est-à-dire du nombre de termes dans la sommes, ce qui indique qu'on peut passer à des séries aléatoires). Par contre, dans le cas ci-dessous, on peut en fait montrer des inégalités bien meilleures pour $p \le \sqrt{n}$ (car le caractère ψ_2, que l'on avait déjà obtenu dans le premier exercice, est plus faible que ce que donne la concentration, par exemple).

Remarquons que le résultat l'inégalité ci-dessus reste vrai si on remplace les e_i par des vecteurs v_i quelconque de $\mathbb{R}^n$ ou plus généralement d'un espace de Hilbert $(H, \|\cdot\|, \langle\cdot,\cdot\rangle)$. En effet, on est ramené à étudier le comportement ψ_2 de la fonction

$$f(x) = \|\sum_{i=1}^n x_i v_i\| = \sqrt{\sum_{i,j=1}^n x_i x_j \langle v_i, v_j\rangle} = \sqrt{\sum_{i=1}^n \lambda_i^2 y_i^2}$$

où les ${\lambda_i}^2$ sont les valeurs propres de la matrice symétrique positive $(\langle v_i, v_j\rangle)_{i,j\le n}$ et $y = P^{-1}x$ avec P la matrice orthogonale de passage qui la diagonalise. On a donc $f(x) = |\Lambda P^{-1}x|$ où Λ est la matrice diagonale donnée par les λ_i. Par invariance par rotation, elle a la même norme $\psi_2(\gamma_n)$ que la fonction $g(y) = |\Lambda y|^2$. Le calcul est alors facile et donne une norme ψ_2 controlée par $\sqrt{\sum {\lambda_i}^2}$, qui est aussi la norme L_2.

3.4. Variables ψ_2, somme de bernouillis, méthode de Laplace et inégalité de Hoeffding

On commence par un peu de vocabulaire. Étant donné une variable aléatoire X (sur un espace de probabilité $(\Omega, \mathbb{P})$), on définit

$$\|X\|_{\psi_\alpha} := \|t \to t\|_{L_{\psi_\alpha}(\mu_X)}$$

où μ_X est la loi sur $\mathbb{R}$ de X. En d'autres termes

$$\|X\|_{\psi_\alpha} = \inf\left\{\lambda > 0\,;\ \mathbb{E}\, e^{\left(\frac{|X|}{\lambda}\right)^\alpha} \le 2\right\}.$$

On dit que X est ψ_α lorsque $\|X\|_{\psi_\alpha} < +\infty$ (pour $\alpha = 2$, on dit aussi que X est sous-gaussienne). On a bien une norme sur l'espace des variables ψ_2. En particulier $\|aX\|_{\psi_\alpha} = |a|\|X\|_{\psi_\alpha}$ pour $a \in \mathbb{R}$. On retiendra que la définition de la norme ψ_α entraine que, pour une fonction constante,

$$\|1\|_{\psi_\alpha} = \log(2)^{-1/\alpha} \in [1,2].$$

On peut alors réécrire toutes les équivalences vue ci-dessus, avec la fonction $f(x) = x$. On retiendra que cela équivaut à un contrôle des queues du type $\mathbb{P}(\{|X| \geq t\|X\|_{\psi_\alpha}) \leq Ce^{-ct^\alpha}$ ou une croissance des moments controlées $\|X\|_p \leq cp^{1/\alpha}\|X\|_{\psi_\alpha}$. Dans la suite, on va utiliser une quatrième forme équivalente : le comportement de la transformée de Laplace.

On dit que la variable aléatoire ε est une variable de Bernouilli (sous entendu centrée de variance 1) si

$$\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$$

Étant bornée, cette variable est-sous gaussienne, et un calcul donne $\|\varepsilon\|_{\psi_2} = \log(2)^{-1/2}$ (en fait $\varepsilon^2 = 1$), mais cette valeur n'a pas d'importance, on utilisera seulement

$$\|a\varepsilon\|_{\psi_2} = c\|a\varepsilon\|_2$$

pour une constante numerique $c > 0$.

Proposition. — *Soit $\varepsilon_1, \ldots, \varepsilon_N$ des variables de Bernouilli indépendantes, et $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$. Alors, pour tout $t > 0$,*

$$\mathbb{P}\big(\big\{\sum_{i=1}^N a_i\varepsilon_i \geq t\big\}\big) \leq e^{-t^2/(2\|a\|_2^2)}$$

et

$$\mathbb{P}\big(\big\{\Big|\sum_{i=1}^N a_i\varepsilon_i\Big| \geq t\big\}\big) \leq 2e^{-t^2/(2\|a\|_2^2)}.$$

Remarquons que cela implique $\sum_{i=1}^N a_i\varepsilon_i$ est ψ_2 avec

$$\Big\|\sum_{i=1}^N a_i\varepsilon_i\Big\|_{\psi_2} \leq c\|a\|_2 = c\sqrt{a_1^2 + \ldots + a_N^2}.$$

Par ailleurs, comme

$$\Big\|\sum_{i=1}^N a_i\varepsilon_i\Big\|_2 = \|a\|_2$$

(en fait la famille $(\varepsilon_1, \ldots, \varepsilon_N)$ est orthonormée dans $L^2(\mathbb{P})$), on peut aussi écrire pour la variable $Y := \sum_{i=1}^N a_i\varepsilon_i$ que

$$\|Y\|_{\psi_2} \leq c\|Y\|_2.$$

Démonstration. — **(Méthode de Laplace)** Pour $\lambda > 0$ on a (truc de Markov)

$$\mathbb{P}\big(\big\{\sum_{i=1}^N a_i\varepsilon_i \geq t\big\}\big) = \mathbb{P}\big(\exp\big[\lambda\sum_{i=1}^N a_i\varepsilon_i\big] \geq \exp[\lambda t]\big) \leq e^{-\lambda t}\,\mathbb{E}\exp\big[\lambda\sum_{i=1}^N a_i\varepsilon_i\big].$$

L'indépendance donne

$$\mathbb{E}\exp\big[\lambda\sum_{i=1}^N a_i\varepsilon_i\big] = \prod_{i=1}^N \mathbb{E}e^{\lambda a_i\varepsilon}.$$

Or on a

$$\mathbb{E}e^{\lambda a_i\varepsilon} = \frac{1}{2}e^{-\lambda a_i} + \frac{1}{2}e^{\lambda a_i} = \cosh(\lambda a_i) \le e^{\lambda^2 a_i^2/2},$$

et donc

$$\mathbb{P}\big(\big\{\sum_{i=1}^N a_i\varepsilon_i \ge t\big\}\big) \le e^{-\lambda t}\, e^{\lambda^2\|a\|_2^2/2}.$$

On conclut en optimisant en $\lambda > 0$ (c'est à dire en prenant $\lambda = t\|a\|_2^2$). □

On a donc obtenu :

***Théorème* (Inégalités de Khintchine).** — *Il existe une constante numérique $C > 0$ tel que, pour tout $n \ge 1$, si $\varepsilon_1, \dots, \varepsilon_n$ sont n variables de Bernoulli ± 1 indépendantes et $a_1, \dots a_n \in \mathbb{R}$, alors*

$$\forall p \ge 2, \qquad \left(\mathbb{E}\left|\sum_{i=1}^n a_i\,\varepsilon_i\right|^p\right)^{1/p} \le C\sqrt{p}\left(\mathbb{E}\left|\sum_{i=1}^n a_i\,\varepsilon_i\right|^2\right)^{1/2} = C\sqrt{p}\,\|a\|_2,$$

$$\forall p \in [1,2], \qquad \frac{1}{C}\left(\mathbb{E}\left|\sum_{i=1}^n a_i\,\varepsilon_i\right|^2\right)^{1/2} \le \left(\mathbb{E}\left|\sum_{i=1}^n a_i\,\varepsilon_i\right|^p\right)^{1/p}.$$

Ce qu'il y a de particulièrement intéressant dans ce résultat, c'est que l'équivalence des normes est indépendante de n. Soit $(\omega \to \varepsilon_i(\omega))_{i\in\mathbb{N}}$ une suite de variables de Bernoulli indépendantes, définies sur un espace de probabilité $(\Omega, \mathbb{P})$. Notons que les $(\varepsilon_i)_{i\in\mathbb{N}}$ forment une famille orthonormée de $L_2(\Omega, \mathbb{P})$. On peut considérer l'espace vectoriel des variables aléatoires engendrées par ces variables :

$$\mathrm{vect}(\{\varepsilon_i\}_{i\in\mathbb{N}}) = \bigcup_{n\ge 0}\{\sum_{i=1}^n a_i\varepsilon_i\ ;\ a_1, \dots, a_n \in \mathbb{R}\} \subset L_\infty(\Omega, \mathbb{P}).$$

Le théorème précédent entraine que toutes les normes L_p, $p \in [1, +\infty[$, sont équivalentes sur cet espace. Plus précisément, considérons l'application (linéaire)

$$\begin{aligned} \ell_2(\mathbb{R}) &\longrightarrow L_2(\Omega, \mathbb{P}) \\ (a_n) &\longrightarrow \phi\big((a_n)\big) := \sum_{n=0}^\infty a_n\varepsilon_n. \end{aligned}$$

Cette application est bien définie. La convergence n'est pas normale (car $\|a_n\varepsilon_n\|_{L_2(\mathbb{P})} = |a_n|$), mais le critère de Cauchy est immédiat, puisque

$$\left\|\sum_{n=p}^q a_n\varepsilon_n\right\|_{L_2(\mathbb{P})}^2 = \sum_{n=p}^q a_n^2.$$

On voit que ϕ est une <u>isométrie</u> sur son image :

$$\left\|\sum_{n=0}^{\infty} a_n\varepsilon_n\right\|^2_{L_2(\mathbb{P})} = \sum_{n=0}^{\infty} a_n^2.$$

Posons

$$\mathcal{R} := \phi\big(\ell_2(\mathbb{R})\big).$$

Alors, $\mathcal{R}$ est un sous-espace fermé de $L_2(\mathbb{P})$ isométrique à $\ell_2(\mathbb{R})$, sur lequel les normes $\|\cdot\|_{L_p(\mathbb{P})}$ avec $p \geq 1$ sont toutes équivalentes (en particulier $\mathcal{R}$ est fermé dans tous les $L_p(\mathbb{P})$). De manière équivalente, on peut dire que, pour tout $p \geq 1$, les espaces $(\mathcal{R}, \|\cdot\|_{L_p(\mathbb{P})})$ et $(\mathcal{R}, \|\cdot\|_{L_2(\mathbb{P})})$ sont isomorphes. Le cas $p = \infty$ est évidemment (?) à part :

Exercice. — *Montrer que* $\sum_{n=0}^{\infty} a_n\varepsilon_n \in L_\infty(\mathbb{P})$ *si et seulement si* $(a_n) \in \ell_1(\mathbb{R})$.

Exercice. — *Soit* $g_1, \dots, g_n, \varepsilon_1, \dots, \varepsilon_n$ *des variables indépendantes (définies sur un certain espace de probabilité* $(\Omega, \mathbb{P})$*) avec pour* $i \leq n$, g_i *gaussienne standard et* ε_i *Bernoulli* ± 1.

1. *Montrer que* $\varepsilon_1|g_1|$ *est encore une variable gaussienne standard.*
2. *Montrer que pour* $a_1, \dots, a_n \in \mathbb{R}$ *on a, pour tout* $p \geq 1$

$$\|\sum_{i=1}^{n} g_i a_i\|_{L_p(\Omega,\mathbb{P})} \geq \sqrt{\frac{2}{\pi}}\,\|\sum_{i=1}^{n} \varepsilon_i a_i\|_{L_p(\Omega,\mathbb{P})},$$

et expliquer comment l'inégalité de Khintchine pour les Bernoulli peut se déduire de celle pour les variables gaussiennes.

Les inégalités de Khintchine peuvent aussi s'interpréter comme des plongements isomorphes, par exemple de ℓ_2 dans ℓ_1. Supposons que $n = 2^k$ $(k = \log_2(n))$ et considérons l'application

$$T : \mathbb{R}^k \longrightarrow \mathbb{R}^n = \mathbb{R}^{\{-1,1\}^k}$$

définie, pour $a = (a_1, \dots, a_k) \in \mathbb{R}^k$, par $T(a) \in \mathbb{R}^n = \mathbb{R}^{\{-1,1\}^k}$ donné par

$$T(a)(\varepsilon) := \sum_{i=1}^{k} a_i\varepsilon_i. \tag{3.9}$$

où l'on voit chaque $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k) \in \{-1, 1\}^k$ comme un indice de coordonnée. On a alors

$$\|T(a_1, \dots, a_k)\|_{\ell_p^n} = n^{1/p}\|T(a_1, \dots, a_k)\|_{L_p(\sigma_n)}.$$

L'inégalité de Khitchine pour $p = 1$ nous dit qu'il existe une constante $C > 0$ telle que pour tout k :

$$\forall a \in \mathbb{R}^k, \quad \frac{\sqrt{n}}{C}\|Ta\|_{\ell_2^n} \leq \|Ta\|_{\ell_1^n} \leq \sqrt{n}\|Ta\|_{\ell_2^n}$$

Remarquons que pour un n général on peut considérer $k = \lfloor \log_2(n) \rfloor$ puis, pour $n' = 2^k \leq n$, plonger trivialement $\mathbb{R}^{n'}$ dans $\mathbb{R}^n$. On a donc :

Proposition. — *Il existe une constante numérique $C > 0$ telle que : pour tout $n \geq 1$, l'application $T : \mathbb{R}^k \longrightarrow \mathbb{R}^{2^k} \hookrightarrow \mathbb{R}^n$ définie pour $k = \lfloor \log_2(n) \rfloor$ par (3.9) est telle que sur l'espace k-dimensionnel $E = T(\mathbb{R}^k) \subset \mathbb{R}^n$ on a*

$$\forall x \in E, \qquad \frac{\sqrt{n}}{C} \|x\|_{\ell_2^n} \leq \|x\|_{\ell_1^n} \leq \sqrt{n} \|x\|_{\ell_2^n}$$

Comme, pour $a \in \mathbb{R}^k$, $\|T(a_1, \ldots, a_k)\|_{\ell_2^n} = \sqrt{n}\|a\|_{\ell_2^k}$, on peut aussi dire que T est un C-plongement de ℓ_2^k dans ℓ_1^n :

$$\forall a \in \mathbb{R}^k, \quad \frac{n}{C} \|a\|_{\ell_2^k} \leq \|Ta\|_{\ell_1^n} \leq n \|a\|_{\ell_2^k}.$$

Théoriquement, ce résultat est moins fort que le théorème de Dvoretzky pour ℓ_1^n que l'on verra plus loin, puisque d'une part le résultat n'est pas presque isométrique (on ne peut obtenir une distortion de type $1 + \varepsilon$ mais seulement une distorsion bornée, c'est-à-dire tout de même indépendante de n) et surtout parce que la dimension est beaucoup plus petite : $\log(n)$ contre λn. Mais ce qu'on a gagné c'est que la construction est explicite. Dans le théorème de Dvoretzky, l'application T est construite par un procédé aléatoire, alors qu'ici T est déterministe et donnée explicitement par (3.9). Très récemment, d'autres constructions explicites ont été obtenues permettant d'arriver presque jusqu'à la dimension optimale λn.

3.5. Compléments

La transformée de Laplace permet également de caractériser le comportement ψ_α. Introduisons la transformée de Legendre de t^α/α (on divise par α pour avoir des formules plus jolies, mais ça n'a aucune importance) : pour $s > 0$,

$$\Lambda_\alpha(s) = \sup_{t>0} \big\{ st - \frac{1}{\alpha} t^\alpha \big\}.$$

On a

$$\Lambda_\alpha(s) = \begin{cases} \frac{1}{\beta} s^\beta & \text{si } \alpha > 1 \text{ avec } \ \frac{1}{\alpha} + \frac{1}{\beta} = 1 \\ 1^\infty_{[0,1]}(s) & \text{si } \alpha = 1. \end{cases}$$

et $\sup_{s>0} \big\{ st - \Lambda_\alpha(s) \big\} = \frac{1}{\alpha} t^\alpha$. En écrivant

$$\begin{aligned} \int e^{\lambda|f|}\, d\mu = \int_0^\infty e^{\lambda t} \mu(\{|f| \geq t\})\, dt &\leq 2 \int_0^\infty e^{\lambda t} e^{-t^\alpha/\|f\|^\alpha_{\psi_\alpha(\mu)}}\, dt \\ &= 2(2/\alpha)^{1/\alpha} \|f\|_{\psi_\alpha(\mu)} \int_0^\infty e^{(2/\alpha)^{1/\alpha}\|f\|_{\psi_\alpha(\mu)} \lambda r} e^{-2r^\alpha/\alpha}\, dr \\ &\leq \Big(2(2/\alpha)^{1/\alpha} \|f\|_{\psi_\alpha(\mu)} \int_0^\infty e^{-r^\alpha/\alpha}\, dr \Big)\, e^{\Lambda_\alpha((2/\alpha)^{1/\alpha}\|f\|_{\psi_\alpha(\mu)}\lambda)} \end{aligned}$$

soit encore

$$\forall \lambda > 0, \qquad \int e^{\lambda|f|}\, d\mu \leq C(\alpha, \|f\|_{\psi_\alpha(\mu)})\, e^{\Lambda_\alpha(c_\alpha \|f\|_{\psi_\alpha(\mu)}\lambda)}.$$

Le fait que la constante devant dépende de la norme ψ_α n'est pas important : seul le comportement dans l'exponentielle compte. En effet, on vérifie réciproquement (Markov) que si

$$\forall \lambda > 0, \qquad \int e^{\lambda|f|}\, d\mu \leq C e^{\Lambda_\alpha(c_0\, \lambda)}$$

alors

$$\|f\|_{\psi_\alpha(\mu)} \le K(C)\, c_0.$$

Noter que le cas $\alpha = 1$ est un peu à part : on est simplement en train de dire qu'il existe $\lambda_0 > 0$ pour lequel l'intégrale de $e^{\lambda_0|f|}$ est finie.

On peut s'inspirer de la preuve faite pour les Bernouilli pour établir, par exemple, la forme plus générale de l'inégalité de Hoeffding suivante :

***Théorème* (Somme de variables ψ_2).** — *Soit $X_1, \dots, X_N$ des variables aléatoires indépendantes sous-gaussiennes. Alors leur somme est aussi sous-gaussienne avec*

$$\big\|X_1 + \ldots + X_N\big\|_{\psi_2}^2 \le C\big(\|X_1\|_{\psi_2}^2 + \ldots + \|X_N\|_{\psi_2}^2\big)$$

pour une constante numérique $C > 0$.

Enfin, plus loin, nous allons établir des déviations non pas pour $|f|$ mais pour f autour de sa moyenne ou de sa médiane. Il sera utile de passer de la moyenne à la médiane et vice-versa.

Pour une fonction f sur Ω μ-intégrable, dit que $m_f \in \mathbb{R}$ est une *médiane* de f pour μ si

$$\mu(\{f \le m_f\}) \ge \frac{1}{2} \quad \text{et} \quad \mu(\{f \ge m_f\}) \ge \frac{1}{2}.$$

Il n'y pas unicité de « la » médiane, mais cela ne pose pas de problème : on travaille avec *une* médiane. (Remarquons que si Ω est connexe et si la mesure de tout ouvert non-vide de Ω est non-nulle, alors il y unicité de la médiane pour une fonction continue sur Ω.)

Pour une variable aléatoire X, "sa" médiane est "la" médiane de la loi de X, c'est donc un nombre M qui vérifie

$$\mathbb{P}(X \ge M) \ge \frac{1}{2} \quad \text{et} \quad \mathbb{P}(X \le M) \ge \frac{1}{2}.$$

Théorème. — *Soit X une variable aléatoire ψ_α de médiane M et d'espérance E. Alors*

$$c_\alpha\|X - E\|_{\psi_\alpha} \le \|X - M\|_{\psi_\alpha} \le C_\alpha\|X - E\|_{\psi_\alpha}.$$

Démonstration. — Par l'inégalité triangulaire on a

$$\|X - E\|_{\psi_\alpha} \le \|X - M\|_{\psi_\alpha} + \|M - E\|_{\psi_\alpha} = \|X - M\|_{\psi_\alpha} + c_\alpha|M - E|,$$

et

$$|M - E| \le \mathbb{E}|X - M| \le \tilde{c}_\alpha\|X - M\|_{\psi_\alpha},$$

ce qui montrer la première inégalité, pour laquelle on a eu besoin de presque rien. Pour l'autre sens, on commence pareil :

$$\|X - M\|_{\psi_\alpha} \le \|X - E\|_{\psi_\alpha} + \|M - E\|_{\psi_\alpha} = \|X - E\|_{\psi_\alpha} + c_\alpha|M - E|.$$

Il s'agit maintenant de contrôle $|M - E|$ en fonction de $\|X - E\|_{\psi_\alpha} =: K$. Quite à changer X en $-X$, on supposera que $M \ge E$. On a, par définition de la médiane et par la déviation ψ_α,

$$\frac{1}{2} \le \mu(\{X \ge M\}) = \mu(\{X - E \ge M - E\}) \le \mu(\{|X - E| \ge M - E\}) \le 2e^{-((M-E)/K)^\alpha},$$

et donc

$$M - E \le \log(4)^{1/\alpha}K.$$

□

On remarque enfin que le centrage ne fait qu'améliorer le comportement ψ_2 :

***Proposition* (Centrage).** — *Soit X une variable aléatoire ψ_α. Alors*

$$\|X - \mathbb{E}X\|_{\psi_\alpha} \leq 5\|X\|_{\psi_\alpha}$$

Démonstration. — Par l'inégalité triangulaire et la borne $\|X\|_1 \leq 2\|X\|_{\psi_\alpha}$, on a

$$\|X - \mathbb{E}X\|_{\psi_\alpha} \leq \|X\|_{\psi_\alpha} + 2|\mathbb{E}X| \leq \|X\|_{\psi_\alpha} + 2\mathbb{E}|X| \leq \|X\|_{\psi_\alpha} + 4\|X\|_{\psi_\alpha}$$

□

Cependant, il est possible que $X - \mathbb{E}X$ ait de bien meilleures propriétés de déviations que X. Par exemple, prenons

$$X = |G|$$

la norme euclidienne d'un vecteur gaussien standard de $\mathbb{R}^n$. On a vu que $\|X\|_{\psi_2} \simeq \sqrt{n}$, mais on peut voir directement (ou comme conséquence de la concentration pour les fonctions lipschtiziennes, qu'on verra plus loin) que $\|X - \mathbb{E}X\|_{\psi_2} \simeq 1$.