CHAPITRE 3

FONCTION DE CONCENTRATION ET INÉGALITÉS DE DÉVIATION SUR UN ESPACE MÉTRIQUE MESURÉ

3.1. Introduction

Dans ce chapitre on travaillera sur un espace métrique (X, d) qui sera muni de sa tribu borélienne (comme toujours, il sera implicite que les ensembles considérés sont boréliens). Le *diamètre* de cet espace est défini par $\mathrm{Diam}(X, d) := \sup_{x,y\in X} d(x, y) \in [0, +\infty]$.

On suppose également que l'on a une mesure (borélienne) de probabilité μ sur X. On dit que (X, d, μ) est un espace métrique mesuré (et ici par mesure on sous-entend probabilité).

Definition 3.1 **(e.m.p.).** — *Un triplet (X, d, μ) est un* **e.m.p.***(espace métrique de probabilité) si (X, d) est un espace métrique et μ une probabilité borélienne sur X.*

On utilisera parfois un vocabulaire probabiliste : au lieu de travailler avec μ on supposera qu'on a un élément aléatoire Y, défini sur un espace de probabilité $(\Omega, \mathbb{P})$, à valeur dans X et tel que sa loi soit égale à μ. Ainsi, pour $A \subset X$ borélien,

$$\mu(A) = \mathbb{P}(Y \in A)$$

et pour $f : X \longrightarrow \mathbb{R}$ (mesurable)

$$\mathbb{E}f(Y) = \int f\, d\mu.$$

Exemple 3.2 **(Espace Gaussien).** — *$X = \mathbb{R}^n$, muni de la norme (et de la distance associée) euclidienne $|\cdot|$ (aussi notée $\|\cdot\|_2$) et de la mesure gaussienne γ_n.*

Exemple 3.3 **(Cube discret ou « booléen » Ω_n).** — *On considère l'espace à deux points $\{-1, 1\}$ dont on fait le n-produit : $\Omega_n = \{-1, 1\}^n$. On considère sur X la mesure uniforme normalisée σ_n : pour $A \subset \Omega_n$, $\sigma_n(A) = \frac{|A|}{2^n}$. Noter que cette mesure est aussi la mesure de probabilité produit $\sigma_n = \sigma^{\otimes n}$ de la mesure $\sigma(\{1\}) = \sigma(\{-1\}) = \frac{1}{2}$. Ainsi, un élément de Ω_n peut aussi se voir comme le résultat d'un tirage de n variables de Bernoulli ± 1 indépendantes. Sur Ω_n, on s'intéressera à la distance suivante, dite distance de Hamming (non normalisée) : pour $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$ dans $\{-1, 1\}^n$,*

$$d(x, y) := \big|\{i \le n\ ;\ x_i \ne y_i\}\big| \tag{3.1}$$

On remarque que (Ω_n, d) est alors de diamètre n.

Si μ est une probabilité borélienne sur un espace métrique (M,d), la *fonction de concentration* $\alpha_{(X,d,\mu)}$ de (X,d,μ) est définie, pour tout $r>0$ par,

$$\begin{aligned} \alpha_{(X,d,\mu)}(r) &:= \sup\{1-\mu(A_r)\ ;\ A\subset X,\ \mu(A)\geq \frac{1}{2}\} \\ &= \sup_{\mu(A)\geq\frac{1}{2}} \mu(A_r^c) \end{aligned} \tag{3.2}$$

où $A_r := \{x\in X\ ;\ d(x,A)<r\}$ désigne le r-voisinage (ouvert) de A (par rapport à la distance d) et

$$A_r^c := X\setminus A_r = \{x\in X\ ;\ d(x,A)\geq r\}.$$

Ainsi, la fonction de concentration est la meilleure (i.e. la plus petite) fonction $\alpha:\mathbb{R}^+\to\mathbb{R}^+$ telle que

$$\forall A\subset X,\quad \text{avec } \mu(A)\geq\frac{1}{2},\quad \text{on a :}\quad \mu(A_r^c)\leq\alpha(r). \tag{3.3}$$

Notons que $\alpha_{(X,d,\mu)}(r)=0$ pour $r>\mathrm{Diam}(X,d)$. En règle générale, on a $\alpha_{(X,d,\mu)}(r)\to 0$ lorsque $r\to\infty$. En effet, fixons $x_0\in X$ et $\varepsilon\in]0,\frac{1}{2}[$. Soit r_0 telle que le complémentaire de la boule $B(x_0,r_0)$ de centre x_0 et de rayon r_0 soit de mesure inférieure à ε ; alors si A est de mesure $\geq\frac{1}{2}$ il doit intersecter $B(x_0,r_0)$ et donc, pour $r\geq 2r_0$, A_r couvre $B(x_0,r_0)$, d'où $1-\mu(A_r)\leq 1-\mu(B)\leq\varepsilon$.

Lorsque le contexte ne prête pas à confusion, on utilisera des notations allégées du type

$$\alpha_\mu = \alpha_{(X,\mu)} = \alpha_{(X,d,\mu)}$$

en gardant en tête que cette quantité dépend intimement de la métrique d choisie.

Le fait d'introduire la fonction de concentration $\alpha_{(X,d,\mu)}$ est utile d'un point de vue théorique. En pratique, elle intervient pour majorer $\mu(A_r^c)$ comme dans (3.3) et donc on cherche surtout une bonne majoration de $\alpha_{(X,d,\mu)}$. **On ne cherche pas à calculer de manière exacte $\alpha_{(X,d,\mu)}$ mais à en obtenir une bonne majoration**. Plus précisément, c'est la manière dont $\alpha_{(X,d,\mu)}(r)$ tend vers 0 pour r « relativement grand »qui nous intérèsse.

On dira que (X,d,μ) a une concentration gaussienne (ou normale) si pour certaines constantes $c,C>0$ on a

$$\alpha_{(X,d,\mu)}(r)\leq Ce^{-cr^2}.$$

On dira que (X,d,μ) a une concentration exponentielle si pour certaines constantes $c,C>0$ on a

$$\alpha_{(X,d,\mu)}(r)\leq Ce^{-cr}.$$

Exemple 3.4. — *Pour l'espace gaussien $(\mathbb{R}^n,\gamma_n)$ on a vu que*

$$\forall A\subset\mathbb{R}^n,\qquad \gamma_n(A_r^c)\leq 2e^{-r^2/4}.$$

Ainsi l'espace gaussien a une concentration gaussienne (!)

On a vu aussi (exercice) plus généralement que si μ est une mesure de proba ayant une densité e^{-V} avec $\mathrm{Hess}\,V\geq\lambda\mathrm{Id}$ $(\lambda>0)$, alors

$$\forall A\subset\mathbb{R}^n,\qquad \mu(A_r^c)\leq 2e^{-\lambda r^2/4},$$

ce qui exprime encore une concentration gaussienne : $\alpha_{(\mathbb{R}^n,|\cdot|,\mu)}(r)\leq 2e^{-\lambda r^2/4}$.

3.2. Inégalités de déviation pour les fonctions lipschitziennes

On se fixe donc comme précédemment un espace métrique mesuré (X, d, μ). Dans ce cadre une fonction lipschitzienne sur (X, d) est une fonction $f : X \longrightarrow \mathbb{R}$ pour laquelle il existe $C > 0$ tel que

$$\forall x, y \in X, \qquad |f(y) - f(x)| \leq C\, d(x, y).$$

La meilleure constante C ci-dessus s'appelle la constante de lipschitz de f et sera notée $\|f\|_{\text{lip}}$,

$$\|f\|_{\text{lip}} = \sup_{x \neq y} \frac{|f(y) - f(x)|}{d(x, y)}.$$

On dit que f est 1-lipschitzienne si $\|f\|_{\text{lip}} \leq 1$.

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

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

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

Proposition 3.5 **(Concentration sur les ensembles $\Rightarrow$ concentration pour les fonctions lipschitziennes autour d'une moyenne)**

Soit (X, d, μ) un **e.m.p.** *avec fonction de concentration notée α_μ. Alors pour toute fonction $f : X \longrightarrow \mathbb{R}$ lipschitzienne de médiane m_f pour μ on a, pour tout $r \geq 0$:*

$$\mu\left(\{f \leq m_f - r\}\right) \leq \alpha_\mu\left(r/\|f\|_{\text{lip}}\right) \quad \textit{et} \quad \mu\left(\{f \geq m_f + r\}\right) \leq \alpha_\mu\left(r/\|f\|_{\text{lip}}\right)$$

ainsi que

$$\mu\left(\{|f - m_f| \geq r\}\right) \leq 2\alpha_\mu\left(r/\|f\|_{\text{lip}}\right).$$

Démonstration. — On peut supposer que f est 1-lipschitzienne (notez que m médiane de f entraîne que $k\,m$ est médiane de $k\,f$). Introduisons $A = \{f \leq m_f\}$ qui est de mesure au moins $1/2$. Alors, pour $r > 0$,

$$A_r = \{x \in X;\ d(x, A) < r\} \subset \{\, x \in X;\ f(x) < m_f + r\}$$

et donc $\mu\left(\{f \geq m_F + r\}\right) \leq \mu(A_r^c) \leq \alpha_\mu(r)$.

De même, si on prend $B = \{f \geq m_f\}$ on a, pour $r > 0$,

$$B_r = \{x \in X;\ d(x, B) < r\} \subset \{\, x \in X;\ f(x) > m_f - r\}$$

et donc $\mu\left(\{f \leq m_F - r\}\right) \leq \mu(B_r^c) \leq \alpha_\mu(r)$. On aurait aussi pu remplacer f par $-f$ dans l'estimée précédente (puisque $-f$ est aussi 1-lipschitzienne et admet $-m_f$ comme médiane).

La dernière inégalité s'obtient en sommant les deux inégalités précédentes.

$\square$

Proposition 3.6 **(Réciproque)**. — *Soit (X, d, μ) un* **e.m.p.** *et $\alpha : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ telle que, pour toute fonction lipschitzienne $f : X \longrightarrow \mathbb{R}$ de médiane m_f pour μ on ait, pour tout $r \geq 0$:*

$$\mu\left(\{f \geq m_f + r\}\right) \leq \alpha\left(r/\|f\|_{\text{lip}}\right).$$

Alors, pour tout ensemble $A \subset X$ avec $\mu(A) \geq \frac{1}{2}$ on a

$$1 - \mu(A_r) \leq \alpha(r),$$

et par conséquent, $\alpha_{(X,d,\mu)} \leq \alpha$.

Démonstration. — Cela repose sur le fait bien connu que pour un ensemble $A \subset X$, la fonction distance

$$f(x) := d(x, A)$$

est 1-lipschitzienne sur X. De plus, pour $r > 0$ on a par définition,

$$A_r = \{f < r\}.$$

Pour un ensemble A avec $\mu(A) \geq \frac{1}{2}$ on a

$$\mu\left(\{f = 0\}\right) \geq \mu(A) \geq \frac{1}{2}$$

et donc 0 est une médiane de f. On en déduit que

$$\mu(A_r^c) = \mu\left(\{f \geq r\}\right) \leq \alpha(r).$$

□

Il arrive souvent qu'on préfère travailler avec la « μ-moyenne » de f,

$$\int f \, d\mu$$

plutôt qu'avec une médiane. On a encore que la concentration des fonctions lipschitziennes autour de leur moyennes est équivalente à la concentration autour de leurs médianes, qui est équivalente à la concentration ensembliste $\alpha_{(X,d,\mu)} \leq \ldots$ Commençons par voir l'analogue de la « réciproque » précédente pour la moyenne.

***Proposition 3.7* (Concentration autour de la moyenne implique concentration ensembliste)**

Soit (X, d, μ) un **e.m.p.** *et $\alpha : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ telle que, pour toute fonction lipschitzienne f sur (X, d) on ait*

$$\forall r \geq 0, \qquad \mu\left(\left\{f \geq \int f \, d\mu + r\right\}\right) \leq \alpha\big(r/\|f\|_{\mathrm{lip}}\big).$$

Alors, pour tout $A \subset X$ avec $\mu(A) > 0$, on a, pour tout $r \geq 0$,

$$\mu(A_r^c) \leq \alpha\big(\mu(A)r\big),$$

et donc, si α est décroissante,

$$\alpha_{(X,d,\mu)}(r) \leq \alpha\Big(\frac{r}{2}\Big).$$

Démonstration. — Soit donc $A \subset X$ avec $\mu(A) > 0$ et $r > 0$. On considère la fonction définie sur X par $F(x) = \min(d(x, A), r)$. Alors F est encore 1-lipschitzienne et on a

$$\int F \, d\mu = \int_{A^c} F \, d\mu \leq \mu(A^c))r.$$

Par construction on a $\mu(A_r^c) = \mu\left(\{F \geq r\}\right)$ et donc

$$\mu(A_r^c) \leq \mu\left(\left\{F \geq \int F\, d\mu + \mu(A)r\right\}\right) \leq \alpha(\mu(A)r).$$

$\square$

Pour boucler la boucle, montrons maintenant qu'une borne sur la concentration ensembliste implique une inégalité de concentration pour les fonctions lipschitziennes autour de leur moyennes. Le cas général suppose de faire quelques hypothèses techniques. Nous allons nous contenter du cas où on a une inégalité de concentration du type $\alpha_\mu(r) \leq Ce^{-cr^p}$, qui est le seul cas qui nous intéresse. On va en fait montrer un énoncé assez général (valable pour chaque fonction prise individuellement). Rappelons auparavant le fait classique suivant qui découle du théorème de Fubini (ou, dans les cas usuels, de la définition de l'intégrale de Lebesgue) :

Fait 3.8. — *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*
$$\int \left[\phi(g) - \phi(0)\right] d\mu = \int_0^{+\infty} \mu\left(\{g \geq r\}\right) \phi'(r)\, dr.$$

Proposition 3.9. — *Soit f une fonction (mesurable) sur un espace de probabilité (X, μ). On suppose qu'il existe $p \geq 1$, des constantes $C, c > 0$, et $a_f \in \mathbb{R}$ tel que*

$$\forall r > 0, \qquad \mu\left(\{|f - a_f| \geq r\}\right) \leq Ce^{-c\, r^p}.$$

Alors

$$\forall r > 0, \qquad \mu\left(\left\{\left|f - \int f\, d\mu\right| \geq r\right\}\right) \leq \tilde{C}\, e^{-c\, k_p\, r^p},$$

où $\tilde{C}$ dépend seulement de $\{c, C, p\}$ et k_p seulement de p.

Ainsi, si on a une concentration autour de la médiane (et donc si on a une concentration pour les ensembles), alors on aussi une concentration identique pour les fonctions lipschitziennes autour de leur moyennes (avec des constantes dépendant seulement des anciennes constantes, et pas de la fonction). Cela dit, en pratique, on obtient souvent directement une inégalité de concentration autour la moyenne.

Démonstration. — On a

$$\left|\int f\, d\mu - a_f\right| \leq \int |f - a_f|\, d\mu = \int_0^{+\infty} \mu\left(\{|f - a_f| \geq r\}\right)\, dr \leq C \int_0^{+\infty} e^{-cr^p}\, dr \;=: \kappa,$$

avec κ dépendant seulement de $\{c, C, p\}$. On en déduit que,

$$\forall r > 0, \qquad \mu\left(\left\{\left|f - \int f\, d\mu\right| \geq r + \kappa\right\}\right) \leq Ce^{-c\, r^p}.$$

Posons $C_1 := e^{c\,\kappa^p}$, qui ne dépend que de $\{c, C, p\}$. On a alors

$$\forall r \in]0, \kappa], \qquad \mu\left(\left\{\left|f - \int f\,d\mu\right| \geq r\right\}\right) \leq 1 = C_1 e^{-c\,\kappa^p} \leq C_1\, e^{-c\,r^p}.$$

D'autre part,

$$\forall r > \kappa, \qquad \mu\left(\left\{\left|f - \int f\,d\mu\right| \geq r\right\}\right) \leq Ce^{-c\,(r-\kappa)^p} \leq C_2\, e^{-ck_p\,r^p}.$$

avec des constantes bien choisies, C_2 dépendant seulement de $\{c, C, p\}$ et k_p seulement de p (on peut utiliser par exemple que $(r-\kappa)^p + \kappa^p \geq 2^{1-p} r^p$). Il suffit alors de mettre les deux bornes ensembles (en prenant des max/min des constantes obtenues). □

Le résultat suivant montre qu'une inégalité de concentration implique de l'intégrablité.

Proposition 3.10. — *Soit f une fonction (mesurable) sur un espace de probabilité (X, μ). On suppose qu'il existe $p \geq 1$, des constantes $C, c > 0$, et $a_f \in \mathbb{R}$ tel que*

$$\forall r > 0, \qquad \mu\left(\{|f - a_f| \geq r\}\right) \leq Ce^{-c\,r^p}.$$

Alors, pour tout $\rho \leq c$ ona $\displaystyle\int e^{\rho f^p} < +\infty$.

Avec un vocabulaire qu'on verra plus loin, ce résultat dit qu'une concentration en e^{-cr^p} pour f implique que f est $\psi_p(\mu)$.

Démonstration. — On a,

$$\forall r > |a_f|, \quad \mu\left(\{|f| \geq r\}\right) \leq \mu\left(\{|f - a_f| \geq r - |a_f|\}\right) \leq Ce^{-c(r-|a_f|)^p}.$$

On a alors, en utilisant le Fait 3.8, pour $\rho < c$,

$$\begin{aligned}
\int e^{\rho F^2}\, d\mu &= 1 + \rho p \int_0^{+\infty} \mu\left(\{|f| \geq r\}\right) r^{p-1} e^{\rho r^p}\, dr \\
&\leq e^{\rho|a_f|^p} + \rho p \int_{|a_f|}^{+\infty} \mu\left(\{|f| \geq r\}\right) r^{p-1} e^{\rho r^p}\, dr \\
&\leq e^{\rho|a_f|^p} + \rho p C \int_{|a_f|}^{+\infty} e^{-c(r-|a_f|)^p} r^{p-1} e^{\rho r^p}\, dr < +\infty.
\end{aligned}$$

□

Dans le même ordre d'idée, nous pouvons obtenir d'autres caractérisations de la concentration pour les fonctions lipschitziennes en terme de croissance des moments (centrés).

***Exercice 3.1* (Croissance des normes $L_q(\mu)$).** — *Soit $p > 1$ une constante fixée (c'est le cas $p = 2$ qui nous intéresse principalement) et (X, μ) un espace de probabilité.*

1. *Soit $g : X \to \mathbb{R}$ une fonction sur X et $c, C > 0$ tels que*

$$\forall r \geq 0, \mu\left(\left\{|g| \geq r\right\}\right) \leq Ce^{-c\,r^p}.$$

(a) *Montrer que* $\displaystyle\|g\|_{L^q(\mu)}^q \leq qC \int_0^{+\infty} r^{q-1} e^{-cr^p}\, dr$.

(b) *Montrer que pour $q \to +\infty$,*

$$\left(\int_0^{+\infty} r^{q-1} e^{-cr^p}\, dr\right)^{1/q} \sim \left(\frac{q}{cpe}\right)^{1/p}.$$

(c) *En déduire qu'il existe une constante K ne dépendant que des constantes $\{c, C, p\}$ tel que*

$$\forall q \geq 1, \qquad \|g\|_{L^q(\mu)} \leq K\, q^{1/p} \tag{3.4}$$

2. *Montrer que si (X, d, μ) un* **e.m.p.** *ayant une concentration du type*

$$\forall r \geq 0, \qquad \alpha_{(X,d,\mu)}(r) \leq Ce^{-c\, r^p},$$

alors il existe une constante K ne dépendant que de $\{c, C, p\}$ tel que, pour toute fonction lipschitzienne $f : X \longrightarrow \mathbb{R}$ on ait

$$\forall q \geq 1, \qquad \left\| f - \int f\, d\mu \right\|_{L^q(\mu)} \leq K\, q^{1/p}\, \|f\|_{\mathrm{lip}}$$

3. *Réciproquement, soit $g : X \to \mathbb{R}$ une fonction sur X et $K > 0$ une constante tel qu'on ait* (3.4).

(a) *Montrer que, pour tout $r \geq 0$, $\mu(\{|g| \geq r\}) \leq r^{-q} K^q\, q^{q/p}$ pour tout $q \geq 1$.*

(b) *En déduire qu'il existe des constantes $\tilde{C}, \tilde{c}$ dépendant seulement de $\{K, p\}$ tel que*

$$\forall r \geq 0, \qquad \mu\left(\left\{|g| \geq r\right\}\right) \leq \tilde{C} e^{-\tilde{c}\, r^p}.$$

4. *Enoncer une réciproque du résultat obtenu en 2.*

Remarque 3.11 **(Attention « normalisation »).** — *Dans ces derniers résultats (et exercice), nous avons en fait établit des implications au niveau d'une fonction f donnée, en partant d'une inégalité du type*

$$\forall r > 0, \qquad \mu\left(\left\{|f - a_f| \geq r\right\}\right) \leq Ce^{-c\, r^p}, \tag{3.5}$$

et en suivant ce que deviennent les nouvelles constantes en fonction de C et c (qui pourraient tres bien dépendre de f fixé). Cependant, la philosophie des inégalités de concentration est de dire qu'on a une inégalité (3.5) *vérifiée pour toute fonction* 1*-lipschitzienne avec les mêmes constantes c et C pour toute les fonctions (constantes universelles). On a vu que c'est équivalent à $\alpha_\mu(r) \leq \tilde{C}e^{-\tilde{c}\, r^p}$ avec $\tilde{c}, \tilde{C}$ dépendant de C et c. Mais il faut bien faire attention qu'on demande que* (3.5) *soit vérifiée pour les fonctions 1-lipschitziennes. Or, dans les résultats vu précédemment, il n'y avait aucune condition sur la fonction f (nul besoin que f soit lipschitzienne, f mesurable suffit). En particulier, on peut appliquer les résultats à une fonction f lipschitzienne quelconque. Mais alors, il faudrait imaginer que la constante c est en fait de la forme $c_1/\|f\|_{\mathrm{lip}}^p$ avec c_1 uniforme, puisque la concentration sur les fonctions lipschitzienne s'écrit : pour toute f lipschitzienne sur (X, d)*

$$\forall r > 0, \qquad \mu\left(\left\{|f - a_f| \geq r\right\}\right) \leq Ce^{-c_1\, r^p/\|f\|_{\mathrm{lip}}^p}.$$

Or, changer le c est évidemment une mauvaise idée car les nouvelles constantes dépendraient alors de manière désagréable de $\|f\|_{\mathrm{lip}}$, ce qu'on ne souhaite pas. La bonne manière de faire est de se rappeler que les résultats précédents (« pour une fonction f il existe des constantes *tel que... ») sont à considérer pour des fonctions normalisées. Par exemple, dans le cas de fonctions lipschitziennes générales, on va appliquer les résultats à*

$$\frac{f}{\|f\|_{\mathrm{lip}}}$$

qui est bien 1*-lipschitzienne (et qui, dans le cas où on à une concentration, vérifiera* (3.5) *avec* c, C *indépendants de* f*). Notez évidemment que, dans le cas où* a_f *est une médiane ou la moyenne de* f *pour* μ*, alors*

$$\frac{a_f}{\|f\|_{\mathrm{lip}}}$$

est une médiane ou la moyenne de $\dfrac{f}{\|f\|_{\mathrm{lip}}}$ *pour* μ*.*

En d'autres termes, les dépendances des constantes les unes par rapport aux autres peuvent être compliquées et non homogènes. Mais les inégalités sur lesquelles on travaille sont elles homogènes ; il suffit (et il faut) donc de les établir pour des fonctions normalisées.

3.3. Transformée de Laplace et concentration

Dans la suite, on se donne un **e.m.p.** (X, d, μ). Nous avons déjà utilisé l'inégalité (triviale) de Markov disant que pour $f : X \longrightarrow \mathbb{R}$, $a \in \mathbb{R}$ et $\lambda \geq 0$,

$$\mu(\{f \geq a\}) \leq e^{-\lambda a} \int e^{\lambda f}\, d\mu \quad \text{et} \quad \mu(\{f \leq a\}) \leq e^{\lambda a} \int e^{-\lambda f}\, d\mu. \tag{3.6}$$

On en tire en particulier le fait suivant :

Fait 3.12. — *Soit* $f : X \longrightarrow \mathbb{R}$ *tel qu'il existe* λ_0 *pour lequel* $\int e^{\lambda_0 f} < +\infty$*. Alors, pour tout* $a \in \mathbb{R}$*,*

$$\mu(\{f \geq a\}) \leq \inf_{\lambda \geq 0} e^{-\lambda a} \int e^{\lambda f}\, d\mu.$$

Cela nous amène à introduire la fonctionnelle suivante : pour tout $\lambda \geq 0$,

$$\begin{aligned} \mathrm{E}_{(X,d,\mu)}(\lambda) &:= \sup\left\{\int e^{\lambda f}\, d\mu\ ;\ f : X \to \mathbb{R} \text{ 1-lipschitzienne sur } (X,d) \text{ avec } \int f\, d\mu = 0\right\} \\ &= \sup\left\{\int e^{\lambda[f - \int f\, d\mu]}\, d\mu\ ;\ f : X \to \mathbb{R} \text{ 1-lipschitzienne sur } (X,d)\right\}, \end{aligned}$$

que nous abrégerons parfois en $\mathrm{E}_\mu(\lambda)$. Cette fonctionnelle est la meilleure fonction $E : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \cap \{+\infty\}$ pour laquelle on a : pour toute fonction f lipschitzienne sur (X, d),

$$\forall \lambda \geq 0, \qquad \int e^{\lambda[f - \int f\, d\mu]}\, d\mu \leq E(\lambda \|f\|_{\mathrm{lip}}).$$

On tire du fait précédent que : pour toute fonction 1-lipschitzienne sur (X, d),

$$\forall r > 0, \qquad \mu\left(\left\{f \geq \int f\, d\mu + r\right\}\right) \leq \mathcal{E}(r) := \inf_{\lambda \geq 0} e^{-\lambda r} \mathrm{E}_{(X,d,\mu)}(\lambda) = e^{-\sup_{\lambda \geq 0} \left[\lambda r - \log \mathrm{E}_{(X,d,\mu)}(\lambda)\right]}.$$

et des résultats vus dans la partie précédente que

$$\forall r > 0, \qquad \alpha_{(X,d,\mu)}(r) \leq \mathcal{E}\left(\frac{r}{2}\right)$$

Le résultat suivant est par conséquent immédiat.

Proposition 3.13. — 1. *S'il existe $\lambda_0 > 0$ tel que $C := \mathrm{E}_{(X,d,\mu)}(\lambda_0) < +\infty$, alors on a la concentration exponentielle suivante :*

$$\forall r > 0, \qquad \alpha_{(X,d,\mu)}(r) \le C\, e^{-\lambda_0 r}.$$

2. *On suppose qu'il existe $q > 1$ et des constantes $c, C > 0$ tel que*

$$\forall \lambda \ge 0, \qquad \mathrm{E}_{(X,d,\mu)}(\lambda) \le C\, e^{c\lambda^q}.$$

Alors, si $p > 1$ est tel que $\frac{1}{p} + \frac{1}{q} = 1$, on a la concentration suivante :

$$\forall r > 0, \qquad \alpha_{(X,d,\mu)}(r) \le C\, e^{-\tilde{c}\, r^p},$$

où $\tilde{c}$ est une constante (calculable !) dépendant seulement de c et p.
Dans le cas particulier où

$$\forall \lambda \ge 0, \qquad \mathrm{E}_{(X,d,\mu)}(\lambda) \le C\, e^{c\,\lambda^2},$$

on a la concentration gaussienne suivante : pour toute fonction f 1-lipschitzienne sur (X, d),

$$\forall r \ge 0, \qquad \mu\left(\left\{f \ge \int f\, d\mu + r\right\}\right) \le C\, e^{-r^2/(4c)}$$

et

$$\forall r \ge 0, \qquad \alpha_{(X,d,\mu)}(r) \le C\, e^{-r^2/(16c)},$$

Encore une fois, attention à la normalisation. Rappelons que pour une fonction lipschitzienne $f : X \to \mathbb{R}$ de moyenne nulle, on a

$$\forall \lambda > 0, \qquad \int e^{\frac{\lambda f}{\|f\|_{\mathrm{lip}}}}\, d\mu \le \mathrm{E}_{(X,d,\mu)}(\lambda),$$

ou encore, de manière équivalente : $\forall \lambda > 0, \quad \displaystyle\int e^{\lambda f}\, d\mu \le \mathrm{E}_{(X,d,\mu)}\big(\|f\|_{\mathrm{lip}}\, \lambda\big)$.

Pour être complet, signalons que la réciproque de la proposition précédente est également vraie : une borne sur la fonction de concentration implique une borne sur $\mathrm{E}_{(X,d,\mu)}$. Nous allons nous contenter de le vérifier dans le cas de la concentration gaussienne (l'adaptation immédiate à une borne en e^{cr^q} est laissée en exercice).

Proposition 3.14. — *Supposons que*

$$\alpha_{(X,d,\mu)}(r) \le C e^{-c\, r^2}$$

pour certaines constantes $c, C > 0$. Alors,

$$\forall r > 0, \qquad \mathrm{E}_{(X,d,\mu)}(r) \le \tilde{C}\, e^{\tilde{c} r^2}$$

pour des constantes $\tilde{c}, \tilde{C} > 0$ dépendant seulement de c et C.

Démonstration. — On sait qu'il existe des constantes $c_1, C_1 > 0$ dépendant seulement de c et C telles que, pour toute fonction f 1-lipschitzienne sur (X,d) de moyenne nulle ,

$$\forall r > 0, \qquad \mu\left(\{f \geq r\}\right) \leq C_1 e^{-c_1 r^2}$$

On a alors, pour une telle fonction f, en utilisant le Fait 3.8, que pour tout $\lambda > 0$

$$\begin{aligned} \int e^{\lambda f}\, d\mu &= 1 + \lambda \int_0^{+\infty} \mu\left(\{f > r\}\right) e^{\lambda r}\, dr \\ &\leq 1 + C_1 \lambda \int_0^{+\infty} e^{\lambda r} e^{-c_1 r^2}\, dr \\ &= 1 + C_2 \lambda e^{c_2 \lambda^2} \\ &\leq 1 + C_2 e^{(c_2+1)\lambda^2} \\ &\leq (C_2+1) e^{(c_2+1)\lambda^2} \end{aligned}$$

□

La transformée de Laplace est un outil très efficace pour montrer des inégalités de concentration et de déviation, même au niveau d'une fonction particulière f et , et cela simplement grâce à l'inégalité de Markov (sous la forme du Fait 3.12).

Voici maintenant une observation élémentaire qui permet parfois d'estimer la fonction $\mathbb{E}_\mu$.

Lemme 3.15. — *Soit (X,μ) un espace de probabilité. Pour toute fonction $f : X \longrightarrow \mathbb{R}$ et $\lambda \geq 0$ on a*

$$\int e^{\lambda[f - \int f\, d\mu]}\, d\mu \leq \iint e^{\lambda[f(x)-f(y)]}\, d\mu(x)\, d\mu(y) = \iint \cosh\big(\lambda[f(x)-f(y)]\big)\, d\mu(x)\, d\mu(y).$$

En particulier, pour tout $k > 0$ on a

$$\Big[\forall x, y \in X, \quad |f(x)-f(y)| \leq k\Big] \Longrightarrow \quad \int e^{\lambda\left[f - \int f\, d\mu\right]}\, d\mu \leq e^{\lambda^2 k^2/2} \tag{3.7}$$

Démonstration. — La première découle de l'inégalité de Jensen $e^{-\int \lambda f\, d\mu} \leq \int e^{-\lambda f}\, d\mu$. Si on a $|f(x)-f(y)| \leq k$, alors

$$\int e^{\lambda[f - \int f\, d\mu]}\, d\mu \leq \cosh(\lambda k) \leq e^{k^2\lambda^2/2}.$$

□

La première application élémentaire du Lemme consiste à utiliser le fait que, pour une fonction 1-lipschitzienne sur un espace métrique (X,d) on a $e^{\lambda[f(x)-f(y)]} \leq e^{\lambda d(x,y)}$. On a donc toujours

$$\forall \lambda > 0, \qquad \mathrm{E}_{(X,d,\mu)}(\lambda) \leq \iint e^{\lambda d(x,y)}\, d\mu(x)\, d\mu(y).$$

En général, cette majoration est trop violente et ne permet pas de récupérer le bon ordre de concentration.

Par contre, la transformée de Laplace permet, via le Lemme précédent, d'obtenir immédiatement une inégalité de concentration dépendant du diamètre en utilisant qu'une fonction 1-lipschitzienne sur (X,d) vérifie $\forall x,y \in X, \quad |f(x)-f(y)| \leq \mathrm{Diam}(X,d)$.

Proposition 3.16. — *Soit* (X,d) *un espace métrique ayant un diamètre fini* $D := \mathrm{Diam}(X,d) < +\infty$*, et* μ *une probabilité borélienne sur* X*. On a, pour toute fonction* $f : X \to \mathbb{R}$ *lipschitzienne sur* X *:*

$$\forall \lambda > 0, \qquad \int e^{\lambda[f - \int f\,d\mu]}\,d\mu \leq e^{\lambda^2 D^2 \|f\|_{\mathrm{lip}}^2/2}$$

On a donc

$$\forall \lambda > 0, \qquad \mathrm{E}_{(X,d,\mu)}(\lambda) \leq e^{D^2\lambda^2/2},$$

et l'inégalité de concentration : $\forall r > 0, \quad \alpha_{(X,d,\mu)}(r) \leq e^{-r^2/(8D^2)}$.

Il semble à première vue que le Lemme 3.15 donne un résultat plus général que cette proposition puisqu'on a pas besoin de structure métrique. En fait, c'est vraiment équivalent, car une fonction $f : X \to \mathbb{R}$ sur un ensemble X, vérifiant $\forall x,y \in X, \quad |f(x)-f(y)| \leq k$, est lipschitzienne par rapport à la distance triviale sur X. En effet, on peut aussi bien écrire

$$\forall x,y \in X, \quad |f(x)-f(y)| \leq k\mathbf{1}_{x\neq y}$$

ce qui traduit que f est 1-lipschitzienne par rapport à la distance triviale $k\mathbf{1}_{x\neq y}$, distance pour laquelle X est de diamètre k. Ainsi, la proposition entraîne bien le Lemme. Il est donc utile, surtout dans la partie qui suit, de se souvenir que les résultats pour les fonctions lipschitziennes sont intéressants lorsqu'on met sur un ensemble abstrait une métrique triviale.

Remarque 3.17. — *On pourrait penser que la borne (découlant aussi du Lemme)*

$$\mathrm{E}_{(X,d,\mu)}(\lambda) \leq \cosh(D\lambda) = e^{\log\cosh(D\lambda)}$$

permettrait d'obtenir une meilleure concentration. Il n'en est rien car la transformée de Legendre

$$\sup_{\lambda \geq 0}[r\lambda - \log\cosh(D\lambda)]$$

se comporte bien comme un multiple de r^2 *pour* $r \leq D$ *(et vaut* $+\infty$ *pour* $r \geq D$*).*

En général, la proposition précédente ne donne pas la « bonne » inégalité de concentration, et tout d'abord, elle ne s'applique pas aux espaces de diamètre infini, comme l'espace gaussien. Le problème principal est que l'estimation ne dépend pas de la mesure μ, mais seulement de la structure métrique.

Cependant, nous allons voir que, combiné avec une structure produit, cette estimation permet d'obtenir de très bonnes inégalités de concentration.

3.4. Transformée de Laplace et produit. Application au cube discret

La transformée de Laplace est très utile pour obtenir des inégalités de concentration gaussiennes sur les espace produit avec la distance produit dite ℓ_1. Nous allons l'appliquer ici à l'estimation de $\mathrm{E}_{(X,d,\mu)}$, c'est-à-dire au cas de fonctions lipschitziennes. Le plus important est de retenir la méthode, car elle s'applique à diverses classes de fonctions, et même individuellement, pour une fonction f donnée (voir « méthode de martingale »).

Proposition 3.18 **(Transformée de Laplace et produit)**. — *Soit (X,d) et (Y,δ) deux espaces métriques, muni de probabilités boréliennes μ et ν, respectivement.*

1. *Pour $q \geq 1$, on définit sur $X \times Y$ la métrique $d \oplus_1 \delta$ par*
$$(d \oplus_1 \delta)\big((x,y),(x',y')\big) = d(x,x') + \delta(y,y'), \qquad \forall (x,y),(x',y') \in X \times Y.$$
Alors, pour tout $\lambda \geq 0$, on a :
$$\mathrm{E}_{(X\times Y, d\oplus_q \delta, \mu\otimes\nu)}(\lambda) \leq \mathrm{E}_{(X,d,\mu)}(\lambda)\, \mathrm{E}_{(X,d,\mu)}(\lambda).$$
2. *Supposons que X et Y soient de diamètres finis D_1 et D_2 respectivement. Si $F : X \times Y \longrightarrow \mathbb{R}$ est une fonction sur le produit et $\alpha_1, \alpha_2 \geq 0$ des constantes tels que*
$$\forall (x,y),(x',y') \in X \times Y \quad \big|F(x,y) - F(x',y')\big| \leq \alpha_1 d(x,x') + \alpha_2 \delta(y,y'),$$
alors on a :
$$\forall \lambda \geq 0, \qquad \int_{X\times Y} e^{\lambda\left[F - \int F d(\mu\otimes\nu)\right]}\, d(\mu\otimes\nu) \;\leq\; e^{(\alpha_1^2 D_1^2 + \alpha_2^2 D_2^2)\lambda^2/2}.$$

Remarque 3.19. — *Le résultat précédent reste valable si l'on remplace la distance $d \otimes_1 \delta$ par la somme ℓ_q des distances :*
$$(d \oplus_q \delta)\big((x,y),(x',y')\big) := \big(d(x,x')^q + \delta(y,y')^q)\big)^{1/q}, \qquad \forall (x,y),(x',y') \in X \times Y.$$
Le résultat le plus fort est obtenu en prenant $q = 1$; les autres cas s'en déduisent. En effet, comme pour $a, b \geq 0$, $a^q + b^q \leq (a+b)^q$, on a l'inégalité suivante entre les distances sur $X \otimes Y$: pour $q \geq 1$
$$d \oplus_q \delta \leq d \oplus_1 \delta.$$
Or, si on a deux distances d et d' sur X avec $d' \leq d$, toute inégalité de concentration (pour les fonctions lipschitziennes) valable pour (X,d,μ) est automatiquement valable pour (X,d',μ). (On peut aussi dire qu'il existe plus de fonction 1-lipschitziennes pour d que pour d').

Ainsi, le cas $q = 1$ est celui qui permet d'obtenir la meilleure concentration sur $X \times Y$. Justement, le fait qu'on puisse obtenir en toute généralité quelque chose pour distance « aussi forte » que celle de la somme ℓ_1 suggère que, dans certains cas particuliers, ce résultat peut être amélioré.

Démonstration. — On peut remarquer que le deuxième point se déduit du premier et de l'estimée vue à la partie précédente, puisque F est 1-lipschitzienne par rapport à la somme ℓ_1 des distances $\alpha_1 d$ et $\alpha_2\delta$ qui sont de diamètre $\alpha_1 D_1$ et $\alpha_2 D_2$, respectivement. Cela dit, la preuve donne directement les deux résultats.

Soit F une fonction sur $X \times Y$ de $\mu \otimes \nu$-moyenne nulle. Supposons que

$$\forall (x,y),(x',y') \in X \times Y \quad \big|F(x,y) - F(x',y')\big| \leq \alpha_1 d(x,x') + \alpha_2 \delta(y,y').$$

(Si F est 1-lipschitzienne sur $(X \times Y, d \otimes_1 \delta)$ on prendra $\alpha_1 = \alpha_2 = 1$). Posons, pour $y \in Y$,

$$\forall x \in X, \quad F^y(x) := F(x,y) \quad \text{et} \quad G(y) := \int_X F^y \, d\mu = \int_X F(x,y)\, d\mu(x).$$

Notons que $\int G\, d\nu = 0$ et que F^y (pour chaque $y \in Y$) et G sont α_1 et α_2-lipschitzienne sur (X,d) et (Y,δ) respectivement. En effet, on a par hypothèse

$$|F^y(x) - F^y(x')| \leq \alpha_1 d(x,x')$$

et

$$|G(y) - G(y')| \leq \int |F(x,y) - F(x,y')|\, d\mu(x) \leq \int_X \alpha_2 \delta(y,y')\, d\mu(x) = \alpha_2 \delta(y,y').$$

Pour $\lambda > 0$, on a donc,

$$\begin{aligned}
\int_{X\times Y} e^{\lambda F}\, d(\mu \otimes \nu) &= \int_Y \left(\int_X e^{\lambda F^y(x)}\, d\mu(x) \right) d\nu(y) \\
&= \int_Y e^{\lambda G(y)} \left(\int_X e^{\lambda \left[F^y(x) - \int_X F^y \, d\mu \right]}\, d\mu(x) \right) d\nu(y) \\
&\leq \mathrm{E}_{(X,d,\mu)}(\alpha_1 \lambda) \int_Y e^{\lambda G}\, d\nu \\
&\leq \mathrm{E}_{(X,d,\mu)}(\alpha_1 \lambda)\, \mathrm{E}_{(Y,\delta,\nu)}(\alpha_2 \lambda).
\end{aligned}$$

□

On déduit des résultats précédents, par récurrence, le théorème suivant.

Théorème 3.20. — *Soit $(X_1,d_1),\ldots,(X_n,d_n)$ des espaces métriques de diamètres respectifs D_i pour $i \leq n$. On se donne sur le produit $X := X_1 \times \ldots \times X_n$ une probabilité produit (quelconque) $P_n = \mu_1 \otimes \ldots \otimes \mu_n$.*

Soit $F : X \longrightarrow \mathbb{R}$ et $\alpha_1,\ldots,\alpha_n \geq 0$ tel que

$$\forall x_1, z_1 \in X_1, \ldots, x_n, z_n \in X_n, \quad |F(x_1,\ldots,x_n) - F(z_1,\ldots,z_n)| \leq \sum_{i=1}^{n} \alpha_i d_i(x_i,z_i). \tag{3.8}$$

Alors on a

$$\forall r > 0, \qquad P_n\left(\left\{ F \geq \int F\, dP_n + r \right\} \right) \leq e^{-r^2/(2K^2)},$$

où $K^2 = \alpha_1^2 D_1^2 + \ldots + \alpha_n^2 D_n^2$.

En particulier, si on muni le produit de la métrique produit ℓ_1 donnée par $d = d_1 \oplus \ldots \oplus d_n$, alors, pour toute fonction F 1-lipschitzienne sur (X,d) on a

$$\forall r > 0, \qquad P_n\left(\left\{ F - \int F\, dP_n \geq r \right\} \right) \leq e^{-r^2/(2D^2)},$$

où $D^2 := D_1^2 + \ldots + D_n^2$. On a donc aussi l'inégalité de concentration : $\forall r > 0, \quad \alpha_{(X,d,P_n)}(r) \leq e^{-r^2/(8D^2)}$.

Le résultat est valable tel quel si on prend une distance ℓ_q sur le produit : $d = d_1 \oplus_q \ldots \oplus_q d_n$, mais c'est le cas $q = 1$ qui donne le résultat le plus fort.

***Remarque 3.21* (Où l'on voit que l'on a gagné un facteur n)**
Dans les inégalités sur des espaces produit, la dépendance en n est très importante. Notez qu'on a toujours, pour la distance ℓ_1, $d = d_1 \oplus \ldots \oplus d_n$,

$$Diam(X, d) = Diam(X_1, d_1) + \ldots + Diam(X_n, d_n)$$

et que donc $D^2 \leq Diam(X,d)^2$. Cependant, la différence entre ces termes peut être très grande, comme le montre le cas particulier important suivant. Supposons que l'on prenne le produit d'un même espace métrique X_1, c'est-à-dire $X = X_1^n$. Alors, si D_1 est le diamètre de X_1, on a

$$D^2 = nD_1^2 \quad \text{et} \quad Diam(X, d)^2 = n^2 D_1^2.$$

On a donc une différence d'un facteur multiplicatif n. Une application directe de la Proposition 3.16 aurait donné une bien moins bonne inégalité que celle du théorème. On voit ici un premier exemple où le fait d'avoir une structure produit donne automatiquement une amélioration de la concentration. Ce phénomène a été étudié en détail par M. Talagrand. La concentration donné par le théorème peut, dans certains cas, être sensiblement améliorée.

***Remarque 3.22* (Reformulation de la condition (3.8)).** — *Il est utile de remarquer qu'on a l'équivalence entre les assertions suivantes :*

1. *F vérifie* (3.8)
2. *Pour tout $i \leq n$, tout $(x_1, \ldots, x_{i-1}, x_{i+1}, x_n) \in \prod_{j \neq i} X_j$, on a : $\forall x_i, x_i' \in X_i$,*

$$|F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, x_n) - F(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, x_n)| \leq \alpha_i \, d_i(x_i, x_i').$$

3. *Pour tout $i \leq n$, tout $(x_1, \ldots, x_{i-1}, x_{i+1}, x_n) \in \prod_{j \neq i} X_j$, on a que la fonction $t \longrightarrow F(x_1, \ldots, x_{i-1}, t, x_{i+1}, x_n)$ est α_i-lipschitzienne sut (X_i, d_i).*

On peut déduire du résultat précédent des inégalités sous-gaussiennes pour des sommes de variables aléatoire indépendantes bornées.

***Corollaire 3.23* (Inégalité de Hoeffding).** — *Soit $Y_1, \ldots, Y_n$ des variables aléatoires indépendantes (sur un certain espace de probabilité $(\Omega, \mathbb{P})$). On suppose que, pour chaque $i \leq n$, il existe des constantes $\alpha_i, \beta_i \in \mathbb{R}$ tel que $Y_i \in [\alpha_i, \beta_i]$. Alors, si on pose $S = Y_1 + \ldots + Y_n$ on a*

$$\forall r > 0, \qquad \mathbb{P}\big(\{S \geq \mathbb{E}(S) + r\}\big) \leq e^{-r^2/(2D^2)}$$

avec $D^2 := \displaystyle\sum_{i=1}^{n} (\beta_i - \alpha_i)^2$.

On peut aussi appliquer le résultat à $-S$ et obtenir, avec les mêmes notations,

$$\forall r > 0, \qquad \mathbb{P}\big(\{|S - \mathbb{E}(S)| \geq r\}\big) \leq 2e^{-r^2/(2D^2)}$$

Démonstration. — Soit μ_i la loi (sur $\mathbb{R}$) de de Y_i. Elle est a support dans $[\alpha_i, \beta_i]$. On considère donc l'espace métrique mesuré (de probabilité) $([\alpha_i, \beta_i], |\cdot|, \mu_i)$ de diamètre $D_i = |\beta_i - \alpha_i|$. Alors, par indépendance, la loi de $(Y_1, \ldots, Y_n)$ est la probabilité $\mu_1 \otimes \ldots \otimes \mu_n =: P_n$ sur $X = [\alpha_1, \beta_1] \times \ldots \times [\alpha_n, \beta_n] \subset \mathbb{R}^n$. La métrique ℓ_1 sur X n'est alors rien d'autre que la distance associé à la norme $\|\cdot\|_1$ sur $\mathbb{R}^n$. On considère la fonction

$$F(x_1, \ldots, x_n) = x_1 + \ldots + x_n.$$

Cette fonction est 1-lipschitzienne par rapport à la norme $\|\cdot\|_1$ (notez qu'elle n'est pas 1-lipschitzienne par rapport à la norme euclidienne $\|\cdot\|_2$). On a donc, en utilisant le Théorème 3.20,

$$\mathbb{P}\big(\big\{S \geq \mathbb{E}(S) + r\big\}\big) = P_n\left(\left\{x \in X\ ;\ F(x_1, \ldots, x_n) \geq \int F\, dP_n + r\right\}\right) \leq e^{-r^2/(2D^2)}.$$

□

Voici un cas particulier bien connu de ce résultat (inégalités sous-gaussiennes pour les processus de Bernouilli) :

Corollaire 3.24. — *Soit $\varepsilon_1, \ldots \varepsilon_n$ des variables de Bernoulli± 1 indépendantes. Alors, pour $(a_1, \ldots, a_n) \in \mathbb{R}^n$ on a*

$$\forall r > 0, \qquad \mathbb{P}\big(\big\{a_1\varepsilon_1 + \ldots a_n\varepsilon_n \geq r\big\}\big) \leq e^{-r^2/(8|a|^2)}.$$

Démonstration. — On peut appliquer le corollaire précédent à $Y_i = a_i\varepsilon_i \in \{-a_i, a_i\} \subset [-|a_i|, |a_i|]$ qui est d'espérance nulle. Mais en fait, le résultat découle encore plus directement du Théorème 3.20 en prenant $X_i = \{-1, +1\}$ muni de la métrique triviale et avec $\alpha_i = a_i$. Nous allons détailler cela dans la suite. □

On préfère parfois énoncer le résultat sous la forme,

$$\forall r > 0, \qquad \mathbb{P}\big(\big\{a_1\varepsilon_1 + \ldots a_n\varepsilon_n \geq\ r\,|a|\big\}\big) \leq e^{-r^2/8}. \tag{3.9}$$

Il est intéressant d'avoir sur ce corollaire un point de vue légèrement différent. Cela correspond à prendre, dans la Proposition 3.20, la métrique triviale sur X_i, à savoir : pour tout x_i, $y_i \in X_i$, $d_i(x_i, y_i) = 1$ si $x_o \neq y_i$, ce qu'on résume en écrivant

$$d_i(x_i, y_i) = \mathbf{1}_{x_i \neq y_i}$$

Le résultat suivant se déduit du Théorème 3.20 puisque la métrique précédente a toujours un diamètre 1. Mais en fait, on aurait pu commencer par démontrer le résultat ci-dessous et déduire le théorème 3.20 (voir le commentaire après la proposition 3.16).

Théorème 3.25. — *Soit $X_1, \ldots, X_n$ des espaces munis de probabilités $\mu_1, \ldots, \mu_n$, respectivement. On note $X = X_1 \times \ldots \times X_n$ et $P_n = \mu_1 \otimes \ldots \otimes \mu_n$. Si $c_1, \ldots, c_n \geq 0$ et $F : X \to \mathbb{R}$ sont tels que*

$$\forall x, y \in X, \qquad |F(x) - F(y)|\ \leq\ \sum_{i=1}^{n} c_i\, \mathbf{1}_{x_i \neq y_i},$$

alors on a :

$$\forall r > 0, \qquad P_n\left(\left\{F \geq \int F\, dP_n + r\right\}\right) \leq e^{-r^2/(2C^2)},$$

où $C^2 := c_1^2 + \ldots + c_n^2$.

Comme toujours, on a, sous les mêmes hypothèses sur F (qui sont identiques pour $-F$) :

$$\forall r > 0, \qquad P_n\left(\left\{\left|F - \int F\,dP_n\right| \geq r\right\}\right) \leq 2e^{-r^2/(2C^2)}.$$

La condition sur F peut de manière immédiate se reformuler comme suit :

Exercice 3.2. — *Soit* $X = X_1 \times \ldots \times X_n$ *un produit de* n *ensembles et* $c_1, \ldots, c_n \geq 0$. *Montrer que pour* $F : X \to \mathbb{R}$ *donnée, les propriétés suivantes sont équivalentes :*

i) $\forall x, y \in X, \qquad |F(x) - F(y)| \ \leq \ \sum_{i=1}^{n} c_i\, \mathbf{1}_{x_i \neq y_i}$.

ii) $|F(x) - F(y)| \leq c_i$ *pour tout* $i \leq$ *et tout couple* $x, y \in X$ *ne différant que de la* i*-ème coordonnée.*

Dans le cas où tous les X_i sont $\{-1, 1\}$, on peut alors retrouver le corrolaire précédent (OK ?).

Il est, justement, intéressant de réécrire ces résultats dans le cas du cube discret $\Omega_n = \{-1, 1\}^n$ muni de la probabilité (produit) σ_n. Rappelons que la distance de Hamming sur le cube discret $\Omega_n = \{-1, 1\}^n$ est donnée par

$$d(x, y) = \sum_{i=1}^{n} \mathbf{1}_{\{x_i \neq y_i\}} = \big|\{i \leq n\ ;\ x_i \neq y_i\}\big|$$

ce qui correspond au produit de la distance triviale sur chaque $\{-1, 1\}$, pour laquelle on a, dans les estimées précédentes, $D^2 = 1 + \ldots + 1 = n$, c'est-à-dire $D = \sqrt{n}$ (alors que le diamètre de Ω_n vaut n). Notons, pour $i \in \{1, \ldots, n\}$, $\tau_i : \Omega_n \to \Omega_n$ le *flip* de la $i$ème coordonnée : pour $x \in \Omega_n$,

$$\tau_i(x) := (z_1, \ldots, z_n) \ \text{ où, pour } j \leq n\ , \quad z_j := \begin{cases} x_j & \text{si } j \neq i \\ -x_j & \text{pour } j = i \end{cases}$$

Remarquons que, pour $x, y \in \Omega_n$,

$$d(x, y) = 1 \quad \Longleftrightarrow \quad \exists i \leq n, \quad y = \tau_i(x).$$

Pour généralement, si $d(x, y) = k$ et $\{i_1, \ldots, i_k\} = \{i \leq n\ ;\ x_i \neq y_i\}$, alors

$$y = \tau_{i_k} \circ \ldots \circ \tau_{i_1}(x).$$

On a (d'après l'exercice précédent) : pour $f : \Omega_n \longrightarrow \mathbb{R}$ et $c_1, \ldots c_n \geq 0$ donnés, on a :

$$\forall x, y \in \Omega_n, \qquad |f(x) - f(y)| \leq \sum_{i=1}^{n} c_i \mathbf{1}_{x_i \neq y_i}$$

si, et seulement si,

$$\forall x \in \Omega_n,\ \forall i \leq n, \quad |F(\tau_i(x)) - F(x)| \leq c_i.$$

***Théorème 3.26* (Concentration sur cube discret).** — *Soit* Ω_n *le cube discret muni de la distance* d *de Hamming* (3.1) *et de la mesure uniforme* σ_n. *Alors, on a :*

1. *Soit* $f : \Omega_n \longrightarrow \mathbb{R}$ *et* $c_1, \dots c_n \geq 0$ *tel que*
$$\forall x \in \Omega_n, \ \forall i \leq n, \quad |F(\tau_i(x)) - F(x)| \leq c_i.$$
Alors, si on pose $C^2 := c_1^2 + \ldots + c_n^2$, *on a :*
$$\forall r > 0, \qquad \sigma_n \left(\left\{ f - \int f \, d\sigma_n \geq r \right\} \right) \leq e^{-r^2/(2C^2)}.$$
2. *Pour toute fonction* $f : \Omega_n \to \mathbb{R}$ *lipschitzienne sur* (Ω_n, d) *on a*
$$\forall r > 0, \qquad \sigma_n \left(\left\{ f - \int f \, d\sigma_n \geq r\sqrt{n} \|f\|_{\text{lip}} \right\} \right) \leq e^{-r^2/2}.$$
3. *Pour tout ensemble* $A \subset \Omega_n$ *avec* $\sigma_n(A) \geq \frac{1}{2}$,
$$\forall r > 0, \qquad \sigma_n \left(\{ x \in \Omega_n \ ; \ d(x, A) \geq r\sqrt{n} \} \right) \leq e^{-r^2/8}.$$

Question. Comparez ce que donnent les inégalités 1. (c'est-à-dire (3.9)) et 2. du Théorème précédent pour $f(x) = a_1 x_1 + \ldots + a_n x_n$.

3.5. Exercices supplémentaires

***Exercice 3.3* (Inf-convolution et produit)**. — *Étant donné un espace de probabilité* (X, μ) *et une fonction* $c : X \times X \longrightarrow \mathbb{R}^+$, *on dit que* (X, μ, c) *vérifie la propriété* τ *si pour toute fonction mesurable (bornée pour simplifier)* $\varphi : X \longrightarrow \mathbb{R}$ *on a*
$$\int e^{-\varphi} \, d\mu \int e^{Q_c \varphi} \, d\mu \leq 1$$
où, pour tout $y \in X$,
$$Q_c \varphi(y) = \inf_{x \in X} \big[f(x) + c(x, y) \big].$$
Montrer que si (X_1, μ_1, c_1) *et* (X_2, μ_2, c_2) *vérifient la propriété* τ, *alors il en va de même pour* $(X_1 \times X_2, \mu_1 \otimes \mu_2, c)$ *où le coût* $c : (X_1 \times X_2) \times (X_1 \times X_2) \longrightarrow \mathbb{R}^+$ *est la some* ℓ_1 *des coûts, c'est-à-dire , pour* $x_1, y_1 \in X_1$ *et* $x_2, y_2 \in X_2$ *:*
$$c((x_1, x_2), (y_1, y_2)) := c_1(x_1, y_1) + c_2(x_2, y_2).$$
Indication : On pourra poser $\varphi^{x_2}(x_1) = \varphi(x_1, x_2)$ et appliquer la propriété (τ) à la fonction $\psi(x_2) = \log \displaystyle\int_{X_1} e^{Q_{c_1} \varphi^{x_2}} \, d\mu(x_1)$.

***Exercice 3.4* (Propriété (τ)-convexe et concentration pour les fonctions convexes et lipschitziennes sur $[0,1]^n$)**

1. *Soit* $f : [0,1] \to \mathbb{R}$ *une fonction convexe,* $\omega(t) = t^2/4$ *et* $\psi := Q_\omega \varphi$.*On suppose que* $\min f = 0$.
   (a) *Montrer que pour tout* $x \in [0,1]$ *on a, pour tout* $\theta \in [0,1]$, $\psi(x) \leq (1-\theta) f(x) + \frac{\theta^2}{4}$.
   (b) *En déduire que* $\psi(x) \leq +k\big(f(x)\big)$ *où* $k(u) := u - u^2$ *si* $0 \leq u \leq \frac{1}{2}$ *et* $k(u) := \frac{1}{4}$ *si* $u \geq \frac{1}{2}$.
   (c) *Montrer que* $e^{k(u)} \leq 2 - e^{-u}$.

2. *Montrer que pour toute fonction* φ *convexe sur* $[0,1]$ *on a :* $\int_{[0,1]} e^{-\varphi} \int_{[0,1]} e^{Q_\omega \varphi} \leq 1.$

3. *Démontrer que si* φ *est une fonction convexe sur* $[0,1]^n$*, et si on note* $\omega(x) = |x|^2/4$ *pour la norme euclidienne sur* $\mathbb{R}^n$*, alors*
$$\int_{[0,1]^n} e^{-\varphi} \int_{[0,1]^n} e^{Q_\omega \varphi} \leq 1.$$

4. *Démontrer le théorème suivant :*

   ***Théorème 3.27* (Talagrand).** — *Soit* $f : [0,1]^n \to \mathbb{R}$ *une fonction convexe et* 1*-lipschitzienne par rapport à la norme euclidienne* $|\cdot|$*. Alors on a :*
$$\forall r \geq 0, \quad \left|\left\{x \in [0,1]^n \ ; \ \left|f(x) - \textstyle\int_{[0,1]^n} f\right| \geq r\right\}\right| \leq 2e^{-r^2/4}.$$

   Comparer ce résultat aux résultats du cours.

***Exercice 3.5* (Nombre Chromatique aléatoire).** — *Soit* $G = (V, E)$ *un graphe sur* $n \geq 1$ *sommets* $V = \{1, \ldots, n\}$ *(c'est-à-dire la donnée d'un sous-ensemble* E *de partie à* 2 *éléments de* $\{1, \ldots n\}$*.). On écrira* $e_{i,j} \in G$ *si* $\{i, j\}$ *est une arête. Un* N*-coloriage est une application* $\varphi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, N\}$ *tel que*
$$e_{i,j} \in G \Longrightarrow \varphi(i) \neq \varphi(j).$$
Le nombre chromatique de G*, noté* $\chi(G)$*, est le plus petit entier* N *tel qu'il existe un* N*-coloriage de* G*.*

Le modèle classique de graphe aléatoire G_n *est le suivant : Étant donné* n *sommets notés* $\{1, \ldots, n\}$ *pour chaque* $\{i, j\}$ $(i \neq j)$ *on décide indépendamment avec probabilité* $1/2$ *de le prendre comme arête. On va représenter ce modèle de la manière suivante. On travaille sur* Ω_n^n*. Un élément* $x \in \Omega_n^n$ *s'écrit* $x = (x_i)$ *avec* $x_i \in \Omega_n$ *pour* $i \leq n$*. On peut aussi faire l'identification* $\Omega_n^n = \Omega_{n^2}$ *et écrire* $x = (x_{ij})$ *avec la convention* $x_{ij} = (x_i)_j$*, c'est-à-dire* $x_i = (x_{i1}, \ldots, x_{in}) \in \Omega_n$*. Pour* $x \in \Omega_n^n$*, on considère le graphe* $G(x)$ *sur* n *sommets définit comme suit :*
$$e_{i,j} \in G(x) \quad si \quad \big(\, j < i \quad et \quad x_{ij} = 1\big)$$
et on ignore donc les x_{ij} *avec* $j \geq i$ *(i.e. la fonction* $G(x)$ *ne dépend que des coordonnées* $x_{ij} = (x_i)_j$ *avec* $j < i$*.*

1. *Montrer (se persuader) que la loi* G_n *est bien la loi de* $G(x)$ *sous* $\sigma_{n^2} = \sigma_n^{\otimes n}$*.*

2. *Montrer que* $x \longrightarrow G(x)$ *est* 1*-lipschitzienne par rapport à la distance de Hamming sur* Ω_{n^2} *et que pour le graphe aléatoire* G_n*, on a*
$$\forall t \geq 0, \quad \mathbb{P}\left(\left\{\left|\chi_{G_n} - \mathbb{E}\chi_{G_n}\right| \geq t\right\}\right) \leq 2e^{-t^2/2n^2}.$$

   Le but de l'exercice est d'améliorer sensiblement cette estimée en changeant de métrique.

3. *Pour* $x \in \Omega_n^n$ *donné, soit* $G^i(x)$ *le graphe obtenu en enlevant à* $G(x)$ *toutes les arêtes* e_{ik} *avec* $k < i$*. Montrer que*
$$\chi_{G^i(x)} \leq \chi_{G(x)} \quad et \quad \chi_{G(x)} \leq \chi_{G^i(x)} + 1.$$

4. *Montrer que si $x, y \in \Omega_n^n$ ne diffèrent que d'une de leur coordonnée $i \leq n$, alors*
$$|\chi_{G(x)} - \chi_{G(y)}| \leq 1.$$
5. *En déduire que, pour le graphe aléatoire G_n, on a*
$$\forall t \geq 0, \quad \mathbb{P}\left(\left\{\left|\chi_{G_n} - \mathbb{E}\chi_{G_n}\right| \geq t\right\}\right) \leq 2e^{-t^2/2n}.$$