

# Concentration de la mesure

Thomas HUDE et Pierre PETIT  
Sujet proposé par M. LE GALL

## Table des matières

<b>Introduction</b>	<b>2</b>
<b>1 Approche géométrique</b>	<b>3</b>
1.1 Fonction de concentration . . . . .	3
1.2 Exemples . . . . .	4
<b>2 Martingales</b>	<b>5</b>
2.1 Motivation . . . . .	5
2.2 Théorème fondateur . . . . .	7
2.3 Convergence de somme de variables aléatoires . . . . .	8
2.4 Espace de longueur au plus $l$ . . . . .	9
<b>3 Distances de Hamming</b>	<b>12</b>
3.1 Définitions et théorème principal . . . . .	12
3.2 Fonctions Hamming-lipschitziennes . . . . .	16
3.3 Cube $[0, 1]^n$ . . . . .	18
<b>4 Contrôle de coordonnées</b>	<b>20</b>
4.1 Définition et théorème principal . . . . .	20
4.2 Fonctions croissantes et sous-additives . . . . .	21
<b>5 Convolution par infimum</b>	<b>23</b>
5.1 Définitions et propriétés . . . . .	23
5.2 Comportement dans les espaces produits . . . . .	25
5.3 Conséquences . . . . .	26
<b>Références</b>	<b>28</b>

## Introduction

Etant donné un espace métrique mesuré  $(X, d, \mu)$ , en général, on ne « regarde » pas directement cet espace, mais à travers certaines applications : autrement dit, on ne s'intéresse pas à la mesure  $\mu$  elle-même, mais à des mesures images  $\mu_F$ , pour  $F$  mesure (au sens empirique) effectuée sur  $X$ . Focalisons-nous sur les applications lipschitziennes de  $F : X \rightarrow \mathbf{R}$ . Si, par exemple,  $X$  est borné, de diamètre fini  $\Delta$ , on peut s'attendre à ce qu'une application 1-lipschitzienne puisse se répartir assez uniformément sur un segment de  $\mathbf{R}$  de longueur  $\Delta$ , comme c'est le cas pour  $X = [0, 1]$  muni de la mesure uniforme et  $F = Id_{[0,1]}$ . Ce n'est pas possible sur tous les espaces : on parle alors de concentration de la mesure.

Plus précisément, on dira que  $X$  présente une concentration  $\alpha$ , où  $\alpha : X \rightarrow \mathbf{R}_+$ , si, pour toute fonction 1-lipschitzienne  $F : X \rightarrow \mathbf{R}$ , il existe  $a_F \in \mathbf{R}$  tel que

$$\forall r \geq 0 \quad P(|F - a_F| \geq r) \leq 2\alpha(r)$$

Bien entendu, il n'y a réelle concentration que si  $\alpha$  décroît assez vite. De plus, pour vérifier qu'un espace a une certaine concentration, prendre, pour  $a_F$ , une médiane ou la moyenne de  $F$  est souvent un bon choix.

Par exemple, considérons  $X = \{0, 1\}^n$ , muni de la distance de Hamming normalisée

$$d(x, y) = \frac{1}{n} \#\{i \in [1, n]; x_i \neq y_i\}$$

et de la probabilité uniforme  $P$ . On notera que le diamètre de  $X$  est  $\Delta = 1$ . L'application  $F : x \in \{0, 1\}^n \mapsto \frac{x_1 + \dots + x_n}{n}$  est 1-lipschitzienne. Or, le théorème central limite, appliqué à des v.a.  $X_k$  i.d.d. de loi uniforme sur  $\{0, 1\}$ , montre que  $F$  se concentre autour de la valeur  $\frac{1}{2}$ , dans un intervalle de longueur de l'ordre de  $\frac{1}{\sqrt{n}}$ . D'un point de vue géométrique, on peut voir  $F$  comme la première coordonnée de la projection orthogonale sur la diagonale  $x_1 = \dots = x_n$ . On voit immédiatement que

$$P\left(F = \frac{k}{n}\right) = \frac{C_n^k}{2^n}$$

On obtient ainsi un moyen d'évaluer la répartition des coefficients binomiaux.

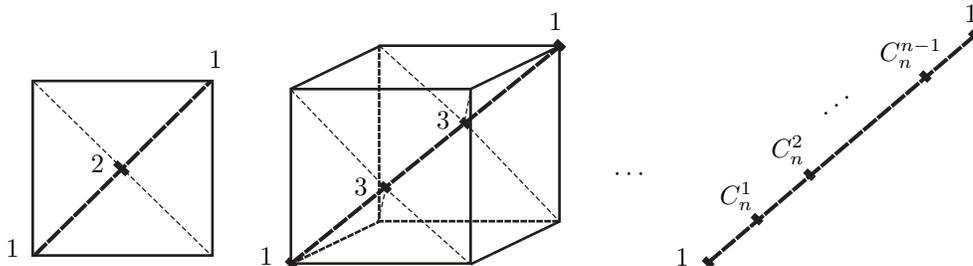


FIG. 1 – Concentration de la moyenne des coordonnées sur  $\{0, 1\}^n$

Comme nous le détaillerons par la suite, le résultat obtenu ici se généralise à toutes les fonctions  $F$  1-lipschitziennes sur  $\{0, 1\}^n$  :

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-nr^2/2}$$

On parle ici de concentration normale. On notera de plus que, dans ce cas, la concentration devient de plus en plus forte lorsque la dimension  $n$  augmente : c'est un phénomène général dans les espaces produits. L'objet de cet exposé est justement de ces espaces, travail effectué d'après le chapitre 4 du livre de M. Ledoux, *The Concentration of Measure Phenomenon* ([Led01]).

Nous introduirons avant tout la notion de fonction de concentration et l'explicitons géométriquement en donnant des exemples classiques qui relient concentration et isopérimétrie. Puis, nous nous focaliserons sur les espaces produits : nous développerons quatre approches différentes menant à des inégalités de concentration, en comparant les résultats obtenus : une première utilise les résultats sur les espérances conditionnelles pour affiner des inégalités obtenues dans un cadre plus général. La deuxième donne une concentration uniforme sur toutes les distances de Hamming d'un espace produit. Dans la troisième approche, on généralise les résultats sur la distance de Hamming en introduisant un contrôle de coordonnées qui donne lieu à de la concentration dans des espaces non métriques. La dernière, basée sur la convolution par infimum, introduit une condition suffisante de concentration normale (ou exponentielle) qui se conserve en passant au produit et qui est donc très adaptée ici.

## 1 Approche géométrique

### 1.1 Fonction de concentration

Etant donné un espace métrique mesuré  $(X, d, \mathcal{B}, \mu)$  (muni des boréliens), introduisons une fonction qui caractérise la concentration des fonctions autour de leur médiane.

**Définition 1.1.1** (Fonction de concentration). On définit, pour  $r \geq 0$ ,

$$\alpha_\mu(r) = \sup \left\{ 1 - \mu(A_r), A \in \mathcal{B}, \mu(A) \geq \frac{1}{2} \right\}$$

en notant  $A_r = \{x \in X; d(x, A) < r\}$  le  $r$ -voisinage ouvert de  $A$ .

Cette définition est géométrique et intrinsèque à  $(X, d, \mu)$ . Les deux résultats suivants établissent le lien avec la concentration des fonctions 1-lipschitziennes sur  $X$ .

**Théorème 1.1.2** (Inégalité de déviation - de concentration). Soit  $F : (X, d) \rightarrow \mathbf{R}$  1-lipschitzienne et  $m_F$  une médiane de  $F$ . On a les inégalités de déviation

$$\mu(\{F \geq m_F + r\}) \leq \alpha_\mu(r)$$

et de concentration

$$\mu(\{|F - m_F| \geq r\}) \leq 2\alpha_\mu(r)$$

*Démonstration.* En notant  $A = \{F \leq m_F\}$ , comme  $F$  est 1-lipschitzienne,

$$A_r \subset \{F < m_F + r\}$$

D'où, étant donné que  $A$  est de mesure  $\geq \frac{1}{2}$ ,

$$\mu(\{F \geq m_F + r\}) = 1 - \mu(\{F < m_F + r\}) \leq 1 - \mu(A_r) \leq \alpha_\mu(r)$$

Pour la seconde inégalité, il suffit de remarquer que

$$\{|F - m_F| \geq r\} = \{F \geq m_F + r\} \cup \{F \leq m_F - r\}$$

et d'appliquer la première inégalité à  $-F$ . □

Réciproquement, on a le résultat suivant :

**Théorème 1.1.3.** *On suppose qu'il existe  $\alpha : X \rightarrow \mathbf{R}_+$  telle que, pour toute fonction 1-lipschitzienne  $F : X \rightarrow \mathbf{R}$ ,*

$$\forall r \geq 0 \quad P(F \geq m_F + r) \leq \alpha(r)$$

Alors,  $\alpha_\mu \leq \alpha$

*Démonstration.* Il suffit d'utiliser, pour  $A \in \mathcal{B}$  de mesure  $\geq \frac{1}{2}$ , la fonction 1-lipschitzienne  $F : x \in X \mapsto d(x, A)$ , dont  $m_F = 0$  est une médiane ; on trouve alors

$$1 - \mu(A_r) = \mu(\{F \geq r\}) \leq \alpha(r)$$

D'où le résultat en passant à la borne inférieure. □

## 1.2 Exemples

L'introduction de la fonction  $\alpha_\mu$  relie la concentration aux caractéristiques géométriques de l'espace  $X$ , en particulier à l'inégalité isopérimétrique. Par exemple,  $\mathbf{R}$ , muni de la probabilité gaussienne standard, admet pour fonction de concentration

$$\alpha_\mu(r) = \int_r^{+\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \leq e^{-r^2/2} \int_0^{+\infty} \frac{e^{-ur} e^{-u^2/2}}{\sqrt{2\pi}} dt \leq \frac{e^{-r^2/2}}{2}$$

la première égalité relevant de l'inégalité isopérimétrique sur  $\mathbf{R}$  muni de la loi normale standard.

Illustrons aussi cette approche par le très classique — et non moins joli — exemple de la sphère  $\mathbb{S}^n$  munie de sa probabilité uniforme  $P$  et de sa métrique intrinsèque  $d$ . Elle présente une concentration normale ainsi que l'énonce le résultat suivant :

**Théorème 1.2.1** (Concentration sur  $\mathbb{S}^n$ ). *Soit  $F : \mathbb{S}^n \rightarrow \mathbf{R}$  1-lipschitzienne et  $m_F$  une médiane de  $F$ . Alors,*

$$\forall r \geq 0 \quad P(|F - m_F| \geq r) \leq 2e^{-(n-1)r^2/2}$$

*Remarque.* En particulier, en voyant  $\mathbb{S}^n$  comme sous-ensemble de  $\mathbf{R}^{n+1}$ , la fonction  $F : \mathbb{S}^n \rightarrow \mathbf{R}$  première coordonnée est 1-lipschitzienne et de médiane  $m_F = 0$ . L'inégalité ci-dessus pour  $F$  traduit donc que l'essentiel de la mesure sur  $\mathbb{S}^n$  est concentré dans un  $\frac{1}{\sqrt{n}}$ -voisinage de l'équateur  $\mathbb{S}^{n-1} = \mathbb{S}^n \cap \{0\} \times \mathbf{R}^n$ . Comme  $\{0, 1\}^n$  (et les espaces produits),  $\mathbb{S}^n$  est de plus en plus concentrée lorsque  $n$  augmente.

*Idée de démonstration.* Nous admettrons l'inégalité isopérimétrique sur la sphère. Il suffit de montrer que  $\alpha_P(r) \leq e^{-2(n-1)r^2/2}$ . Or, l'inégalité isopérimétrique donne

$$\alpha_P(r) = 1 - P((\mathbb{S}_+^n)_r)$$

où  $\mathbb{S}_+^n$  désigne une demi-sphère. Notons  $N = (1, 0, \dots, 0) \in \mathbb{S}^n \subset \mathbf{R}^{n+1}$  et définissons  $\delta(x) = d(N, x)$ . Alors

$$(\mathbb{S}_+^n)_r = \delta^{-1}\left(\left[0, \frac{\pi}{2} + r\right]\right)$$

Or,  $P_\delta$  (mesure image de  $P$  par  $\delta$ ) a une densité  $p$  proportionnelle à  $\text{vol}_{n-1}((\sin \theta)\mathbb{S}^{n-1})$  (cf. figure 2), donc à  $(\sin \theta)^{n-1}$ .

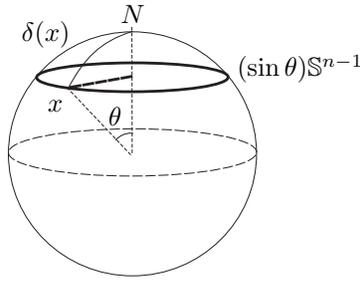


FIG. 2 – Densité de  $P_\delta$

Ainsi :

$$\alpha_P(r) = 1 - P((\mathbb{S}_+^n)_r) = \frac{\int_{\pi/2+r}^{\pi} \sin^{n-1} \theta \, d\theta}{\int_0^{\pi} \sin^{n-1} \theta \, d\theta} \leq e^{-(n-1)r^2/2}$$

la dernière inégalité demandant un peu de calcul. □

En particulier, la projection  $F$  sur la première coordonnée est 1-lipschitzienne, de médiane 0, et  $F^{-1}(\{0\})$  est un équateur de  $\mathbb{S}^n$ . Ainsi, pour avoir un ordre de grandeur, un  $\frac{1}{\sqrt{n}}$ -voisinage de l'équateur  $\mathbb{S}^{n-1}$  contient une grande partie de la mesure  $P$  sur  $\mathbb{S}^n$ . On trouvera la concentration de  $F$  obtenue pour quatre valeurs de  $n$  sur la figure 1.2.

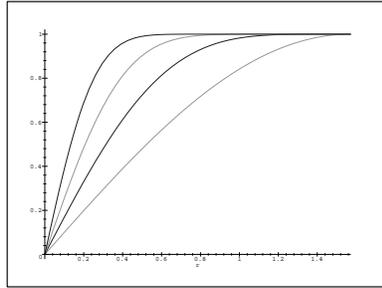


FIG. 3 –  $P((\mathbb{S}^{n-1})_r)$ , pour  $n = 1, 4, 10$  et  $25$

Après ces quelques exemples de nature géométrique, passons au centre de cet exposé qui consiste à l'étude plus particulière des espaces produits.

## 2 Martingales

### 2.1 Motivation

Au cours de cette première approche de la concentration dans les espaces produits, nous nous intéresserons à la concentration des fonctions autour de leur moyenne. Nous nous placerons

sur un espace métrique  $(\Omega, d, \mathcal{F}, P)$ , muni d'une probabilité sur les boréliens. Le but est de contrôler  $|F - E[F]|$  pour  $F : \Omega \rightarrow \mathbf{R}$  1-lipschitzienne. On peut déjà remarquer que, si  $\Omega$  est borné, de diamètre  $\Delta$ , alors

$$\|F - E[F]\|_\infty \leq \Delta$$

puisque  $E[F] \in [\inf F, \sup F]$ . Ainsi, on peut énoncer un premier résultat de concentration :

**Lemme 2.1.1** (Concentration dans un espace borné). *Soit  $(\Omega, d, \mathcal{F}, P)$  un espace de diamètre  $\Delta$ . Alors, pour toute fonction  $F : \Omega \rightarrow \mathbf{R}$  1-lipschitzienne,*

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-r^2/2\Delta^2}$$

*Démonstration.* Il suffit de montrer l'inégalité de déviation

$$P(\{F \geq E[F] + r\}) \leq e^{-r^2/2\Delta^2}$$

Pour ce faire, fixons  $\lambda \geq 0$ ; sachant que  $x \in \mathbf{R} \mapsto e^{\lambda x}$  est croissante et utilisant l'inégalité de Markov, on obtient

$$P(\{F \geq E[F] + r\}) = P(\{e^{\lambda(F-E[F])} \geq e^{\lambda r}\}) \leq e^{-\lambda r} E[F - E[F]]$$

Or, pour  $\frac{F-E[F]}{\Delta} = u \in [-1, 1]$  et  $\lambda' \in \mathbf{R}$ , on a, par convexité de  $x \in \mathbf{R} \mapsto e^{\lambda' x}$ ,

$$e^{\lambda' u} = \exp\left(\frac{1+u}{2}\lambda' + \frac{1-u}{2}(-\lambda')\right) \leq \frac{1+u}{2}e^{\lambda'} + \frac{1-u}{2}e^{-\lambda'} \quad (1)$$

Comme  $E\left[\frac{F-E[F]}{\Delta}\right] = 0$ , on en déduit

$$E[F - E[F]] \leq \cosh(\lambda\Delta) \leq \exp\left(\frac{(\lambda\Delta)^2}{2}\right)$$

d'où

$$P(\{F \geq E[F] + r\}) \leq e^{-\lambda r} e^{\lambda^2 \Delta^2 / 2}$$

Pour  $\lambda = \frac{r}{\Delta^2}$ , on obtient l'inégalité du lemme.  $\square$

Le résultat se généralise facilement dans un espace produit :

**Lemme 2.1.2.** *Si  $(\Omega_i, d_i, \mathcal{F}_i, P_i)$ , pour  $i \in \llbracket 1, n \rrbracket$ , est de diamètre  $\Delta_i$ , en notant  $d = d_1 + \dots + d_n$  (distance  $\ell^1$ ),  $P = P_1 \otimes \dots \otimes P_n$  et  $\Delta = \sqrt{\Delta_1^2 + \dots + \Delta_n^2}$ , on a, pour  $F : (\Omega, d) \rightarrow \mathbf{R}$  1-lipschitzienne,*

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-r^2/2\Delta^2}$$

A titre d'exemple, donnons une version de la loi des grands nombres : si  $X_1, \dots, X_n$  sont des variables aléatoires i.i.d :  $(\Omega', \mathbb{P}) \rightarrow [0, 1]$ , en notant  $S_n = \sum_{k=1}^n X_k$ , on a

$$\forall \epsilon > 0 \quad \mathbb{P}\left(\left\{\left|\frac{S_n}{n} - E[X_1]\right| \geq \epsilon\right\}\right) \leq 2e^{-n\epsilon^2/2} \quad (2)$$

En effet, il suffit d'appliquer le résultat ci-dessus pour  $\Omega_i = [0, 1]$ ,  $d_i(x, y) = \frac{1}{n}|x - y|$ ,  $P_i = \mathbb{P}_{Y_i}$  et  $P = P_1 \otimes \dots \otimes P_n = \mathbb{P}_{(Y_1, \dots, Y_n)}$ . On notera que  $F : (x_1, \dots, x_n) \in [0, 1]^n \mapsto \frac{x_1 + \dots + x_n}{n}$  est bien 1-lipschitzienne pour  $d = d_1 + \dots + d_n$ .

L'intérêt n'est pas d'apprendre que  $\frac{S_n}{n}$  converge en probabilité vers  $E[X_1]$  (la loi faible des grands nombres donne déjà ce résultat, avec même une convergence dans  $L^2$ ), mais d'avoir, d'une part, une estimation de la rapidité de cette convergence, qui est exponentielle, donc foudroyante, et, d'autre part, de retrouver la concentration autour de  $E[X_1]$  en  $\frac{1}{\sqrt{n}}$  donnée par le théorème central limite. Toutefois, les constantes obtenues ne sont pas optimales : l'approche qui suit permet d'obtenir un meilleur résultat.

## 2.2 Théorème fondateur

Une façon naturelle de passer de  $E[F]$  à  $F$  est d'introduire une suite (finie) croissante  $(\mathcal{F}_k)_{k \in \llbracket 0, n \rrbracket}$  de sous-tribus, avec  $\mathcal{F}_0 = \{0, \Omega\}$  et  $\mathcal{F}_n = \mathcal{F}$ , de sorte que

$$F - E[F] = \sum_{k=1}^n (E[F|\mathcal{F}_k] - E[F|\mathcal{F}_{k-1}])$$

sachant que  $E[F|\mathcal{F}_0] = E[F]$  et  $E[F|\mathcal{F}_n] = F$ . Nous noterons  $d_k(F) = E[F|\mathcal{F}_k] - E[F|\mathcal{F}_{k-1}]$ , qui correspond aux oscillations de  $E[F|\mathcal{F}_k]$  autour de  $E[F|\mathcal{F}_{k-1}]$ . On notera que la suite

$$\left( \sum_{i=1}^k d_i(F) \right)_{k \in \llbracket 0, n \rrbracket}$$

est une martingale de valeur finale  $F - E[F]$ , ce qui justifie l'appellation de la méthode. Cette démarche est, comme nous le verrons dans les exemples qui suivent, très adaptée aux espaces produits : si les  $(\Omega_i, \mathcal{A}_i)$  sont des espaces mesurés, une manière naturelle de choisir la suite  $(\mathcal{F}_k)$  sur  $\Omega_1 \times \cdots \times \Omega_n$  est de prendre

$$\mathcal{F}_k = \left\{ A_k \times \Omega_{k+1} \times \cdots \times \Omega_n, A_k \in \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_k \right\}$$

On choisit alors une distance sur  $\Omega_1 \times \cdots \times \Omega_n$  de façon à contrôler les  $d_k(F)$ . Plus généralement, l'idée est d'optimiser le choix de la suite  $(\mathcal{F}_k)$  : il en faut suffisamment pour contrôler chaque terme, sans en avoir un nombre conduisant à une inégalité grossière. La mise en forme de cette méthode est l'objet du théorème fondamental suivant :

**Théorème 2.2.1** (Concentration et espérance conditionnelle). *Soit  $F \in L^1(\Omega, \mathcal{F}, P)$ . En notant*

$$D(F) = \left( \sum_{k=1}^n \|d_k(F)\|_\infty^2 \right)^{1/2}$$

*on a l'inégalité de concentration*

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-r^2/2D(F)^2}$$

*Démonstration.* Pour alléger les notations, écrivons  $d_k = d_k(F)$  et  $D = D(F)$ . Montrons l'inégalité de déviation

$$P(\{F \geq E[F] + r\}) \leq e^{-r^2/2D^2}$$

De même que dans la démonstration du lemme supra, en fixant  $\lambda \geq 0$ , l'inégalité de Markov donne

$$P(\{F \geq E[F] + r\}) \leq e^{-\lambda r} E \left[ \exp \left( \lambda \sum_{k=1}^n d_k \right) \right]$$

Majorons maintenant l'espérance du membre de droite. Par récurrence, montrons que

$$E \left[ \exp \left( \lambda \sum_{k=1}^n d_k \right) \right] \leq \exp \left( \lambda^2 \sum_{k=n-i+1}^n \|d_k\|_\infty^2 \right) E \left[ \exp \left( \lambda \sum_{k=1}^{n-i} d_k \right) \right]$$

L'inégalité est trivialement vraie pour  $i = 0$ . Puis, si elle est vraie pour  $i - 1 \geq 0$ , étant donné que  $\sum_{k=1}^{n-i-1} d_k$  est  $\mathcal{F}_{n-i-1}$ -mesurable,

$$E \left[ \exp \left( \lambda \sum_{k=1}^{n-i} d_k \right) \right] = E \left[ \exp \left( \lambda \sum_{k=1}^{n-i-1} d_k \right) E \left[ e^{\lambda d_{n-i}} | \mathcal{F}_{n-i-1} \right] \right]$$

Comme  $E \left[ \frac{d_{n-i}}{\|d_{n-i}\|_\infty} | \mathcal{F}_{n-i-1} \right] = 0$ , la majoration (1) donne

$$E \left[ e^{\lambda d_{n-i}} | \mathcal{F}_{n-i-1} \right] \leq \exp \left( \frac{(\lambda \|d_{n-i}\|_\infty)^2}{2} \right)$$

D'où l'inégalité voulue au rang  $i$ . Pour  $i = n$ , il vient

$$P(\{F \geq E[F] + r\}) \leq e^{-\lambda r} e^{\lambda^2 D^2 / 2}$$

Pour  $\lambda = \frac{r}{D^2}$ , on obtient l'inégalité du théorème.  $\square$

Le résultat que nous venons d'énoncer permet d'obtenir des inégalités de concentration dans des espaces  $(\Omega, d, (\mathcal{F}_k), P)$  ayant les deux propriétés suivantes :

- toute fonction  $F : \Omega \rightarrow \mathbf{R}$  1-lipschitzienne est intégrable
- $D(F)$  est majoré indépendamment de  $F$  1-lipschitzienne

La première condition est automatiquement vérifiée si  $\Omega$  est de diamètre fini, et, plus généralement s'il existe  $\alpha : \mathbf{R}_+ \rightarrow \mathbf{R}_+$  prenant une valeur  $< \frac{1}{2}$  et telle que, pour toute fonction  $G : X \rightarrow \mathbf{R}$  1-lipschitzienne bornée,

$$P(G \geq \int G dP + r) \leq \alpha(r)$$

En effet, si  $F$  est 1-lipschitzienne, notons  $F_n = \min(n, |F|)$ . L'hypothèse appliquée à  $-F_n$  donne  $P(\{F_n \leq \int F_n dP - r\}) \leq \alpha(r)$ , i.e.  $P(\{F_n > \int F_n dP - r\}) \geq 1 - \alpha(r)$ . On choisit alors  $m$  tel que  $\mu(\{|F| \leq m\}) \geq \frac{1}{2}$  (ce qui est possible car  $F$  est partout finie) et  $r_0$  tel que  $\alpha(r_0) < \frac{1}{2}$ ; dans ces conditions,  $\{F_n > \int F_n dP - r_0\} \cap \{|F| \leq m\} \neq \emptyset$ , donc  $\int F_n dP < m + r_0$ . Par convergence monotone,  $F \in L^1(\mu)$  et  $\int |F| d\mu \leq m + r_0$ .

Pour la seconde, il faut introduire des hypothèses sur  $\Omega$ .

### 2.3 Convergence de somme de variables aléatoires

Commençons par reprendre l'exemple de la loi des grands nombres, en affinant la majoration obtenue.

**Corollaire 2.3.1** (Loi des grands nombres). *Soit  $X_1, \dots, X_n$  des variables aléatoires i.i.d à valeurs dans  $[0, 1]$ . En notant  $S_n = \sum_{k=1}^n X_k$  et  $m = \max(E[X_1], 1 - E[X_1])$ , on a*

$$\forall \epsilon > 0 \quad P \left( \left\{ \left| \frac{S_n}{n} - E[X_1] \right| \geq \epsilon \right\} \right) \leq 2e^{-n\epsilon^2/2m^2}$$

*Remarque.* En particulier, si  $E[X_1] = \frac{1}{2}$ ,  $m = \frac{1}{2}$  et

$$\forall \epsilon > 0 \quad P \left( \left\{ \left| \frac{S_n}{n} - E[X_1] \right| \geq \epsilon \right\} \right) \leq 2e^{-2n\epsilon^2}$$

qui est bien une majoration meilleure que celle obtenue en (2).

*Démonstration.* On applique le théorème 2.2.1 à  $F = \frac{S_n}{n}$ ,  $E[F] = E[X_1]$ ,  $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$ . On notera que

$$\|d_k(F)\|_\infty = \|X_k - E[X_k]\|_\infty \leq m$$

donc  $D(F)^2 = nm^2$ . □

Le résultat suivant donne une inégalité du même type.

**Corollaire 2.3.2** (Concentration normale pour des sommes de v.a. indépendantes). *Soit  $(E, \|\cdot\|)$  un espace vectoriel de Banach et  $Y_1, \dots, Y_n$  des v.a. à valeurs dans  $E$ , indépendantes et bornées. En notant  $S_n = \sum_{k=1}^n Y_k$  et  $D = \left( \sum_{k=1}^n \|Y_k\|_\infty^2 \right)^{1/2}$ ,*

$$\forall r \geq 0 \quad P(\{ \|S_n\| - E[\|S_n\|] \geq r \}) \leq 2e^{-r^2/8D^2}$$

*Démonstration.* Comme ci-dessus, on applique le théorème 2.2.1 à  $F = \|S_n\|$ ,  $\mathcal{F}_k = \sigma(Y_1, \dots, Y_k)$ . On note que

$$E[\|S_n - Y_k\| | \mathcal{F}_k] = E[\|S_n - Y_k\| | \mathcal{F}_{k-1}]$$

car

$$E[\|S_n - Y_k\| | \mathcal{F}_k] = \int \|Y_1 + \dots + Y_{k-1} + y_{k+1} + \dots + y_n\| dP_{(Y_{k+1}, \dots, Y_n)}(y_{k+1}, \dots, y_n)$$

est  $\mathcal{F}_{k-1}$ -mesurable. Alors

$$\begin{aligned} |d_k(F)| &= \left| E[\|S_n\| | \mathcal{F}_k] - E[\|S_n\| | \mathcal{F}_{k-1}] \right| \\ &= \left| E[\|S_n\| - \|S_n - Y_k\| | \mathcal{F}_k] - E[\|S_n\| - \|S_n - Y_k\| | \mathcal{F}_{k-1}] \right| \\ &\leq \left| E[\|Y_k\| | \mathcal{F}_k] \right| + \left| E[\|Y_k\| | \mathcal{F}_{k-1}] \right| \\ &= \|Y_k\| + E[\|Y_k\|] \end{aligned}$$

donc  $D(F)^2 \leq 4 \sum_{k=1}^n \|Y_k\|_\infty^2$ . □

## 2.4 Espace de longueur au plus $l$

Passons maintenant à un cas plus particulier : celui où  $\Omega$  est fini (et, *a fortiori*, de diamètre fini  $\Delta$ ). Nous savons déjà que, si  $F$  est 1-lipschitzienne,  $|F - E[F]| \leq \Delta$ . Mais le cadre des espaces de longueur au plus  $l$  permet d'affiner cette majoration.

**Définition 2.4.1** (Découpage progressif). Soit  $X$  un ensemble. On appelle découpage progressif de  $X$  toute suite finie de partitions

$$\{X\} = \mathcal{P}^0, \dots, \mathcal{P}^n = \{\{x\}, x \in X\}$$

telle que

$$\forall k \in [1, n] \quad \forall A^k \in \mathcal{P}^k \quad \exists A^{k-1} \in \mathcal{P}^{k-1} \quad A^k \subset A^{k-1}$$

( $\mathcal{P}^k$  est un raffinement, au sens des partitions, de  $\mathcal{P}^{k-1}$ ).

**Définition 2.4.2** (Espace de longueur au plus  $l$ ). Soit  $(X, d)$  un espace métrique fini. On dit que  $X$  est de longueur au plus  $l$  s'il existe

- un découpage progressif  $\mathcal{P}^0, \dots, \mathcal{P}^n$  de  $X$  (on notera  $\mathcal{P}^k = \{A_i^k, i \in [1, m_k]\}$ ),
- des réels  $a_1, \dots, a_n$  avec  $l^2 = \sum_{k=1}^n a_k^2$
- et, enfin, pour  $A_i^k, A_j^k \subset A_p^{k-1}$ , une bijection  $\phi_{i,j}^k : A_i^k \rightarrow A_j^k$

tels que

$$\forall x \in A_i^k \quad d(x, \phi_{i,j}^k(x)) \leq a_k$$

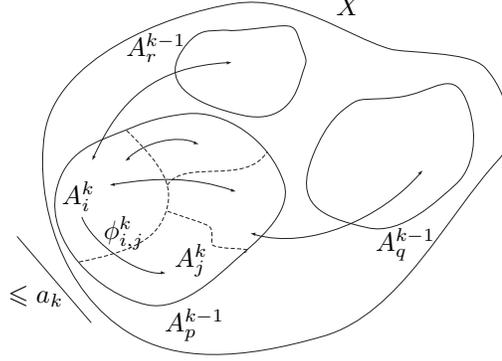


FIG. 4 – Espace de longueur au plus  $l$

*Remarque.* Notons que l'on peut toujours choisir  $l \leq \Delta$ , où  $\Delta$  désigne le diamètre de  $X$  (avec  $(\mathcal{P}^k) = (\mathcal{P}^0)$ ). On peut donc espérer des concentration plus fortes qu'avec le simple diamètre.

Voici le théorème que l'on peut énoncer dans ces espaces :

**Corollaire 2.4.3** (Concentration normale dans un espace de longueur au plus  $l$ ). Soit  $(\Omega, d)$  un espace fini de longueur au plus  $l$  et  $P$  la probabilité uniforme sur  $\Omega$ . Alors, pour toute fonction  $F : \Omega \rightarrow \mathbf{R}$  1-lipschitzienne,

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-r^2/2l^2}$$

*Démonstration.* Notons  $(\mathcal{P}^k)$  un découpage progressif de  $\Omega$  adapté et  $\mathcal{F}_k = \sigma(\mathcal{P}^k)$  (qui donne bien une suite croissante de tribus). On notera que  $\mathcal{P}^k$  est l'ensemble des atomes de  $\mathcal{F}_k$ ; en particulier, si  $F : \Omega \rightarrow \mathbf{R}$  est intégrable, pour tout  $A^k \in \mathcal{P}^k$ ,  $E[F|\mathcal{F}_k]|_{A^k}$  est une constante, que nous noterons  $F_{A^k}^k$ .

Soit  $F : \Omega \rightarrow \mathbf{R}$  1-lipschitzienne : vérifions que  $\|d_k(F)\|_\infty \leq a_k$ , ce qui permet de conclure en vertu du théorème 2.2.1. Soit  $\omega_0 \in \Omega$ . Il existe un unique couple  $(A^{k-1}, B^k) \in \mathcal{P}^{k-1} \times \mathcal{P}^k$  tel que  $\omega_0 \in B^k \subset A^{k-1}$ . Notons  $\mathcal{A}^k = \{C^k \in \mathcal{P}^k; C^k \subset A^{k-1}\}$  et  $N = \#\mathcal{A}^k$ . Alors, en remarquant que tout les  $C^k \subset A^{k-1}$  ont même cardinal, qui est celui de  $B^k$ ,

$$E[F|\mathcal{F}_{k-1}](\omega_0) = F_{A^{k-1}}^{k-1} = \frac{1}{N} \sum_{C^k \in \mathcal{A}^k} F_{C^k}^k$$

Puis

$$\begin{aligned}
\left| E[F|\mathcal{F}_k](\omega_0) - E[F|\mathcal{F}_{k-1}](\omega_0) \right| &= \left| F_{B_k}^k - F_{A_{k-1}}^{k-1} \right| \\
&= \left| \frac{1}{N} \sum_{C^k \in \mathcal{A}^k} (F_{B_k}^k - F_{C^k}^k) \right| \\
&\leq \frac{1}{N} \sum_{C^k \in \mathcal{A}^k} |F_{B_k}^k - F_{C^k}^k|
\end{aligned}$$

On conclut avec l'hypothèse de lipschitzianité :

$$|F_{B_k}^k - F_{C^k}^k| \leq \frac{1}{\#B^k} \sum_{\omega \in B^k} |F(\omega) - F(\phi^k(\omega))| \leq a_k$$

en notant  $\phi^k : B^k \rightarrow C^k$  la bijection associée.  $\square$

Ce résultat s'applique immédiatement dans le produit  $\Omega_1 \times \cdots \times \Omega_n$ , où les  $\Omega_i$  sont des espaces finis, muni d'une distance de Hamming

$$d_a(x, y) = \sum_{i=1}^n a_i 1_{x_i \neq y_i}$$

où  $a \in (\mathbf{R}_+)^n$ . En choisissant

$$\mathcal{P}^k = \{\{\omega_1\} \times \cdots \times \{\omega_k\} \times \Omega_{k+1} \times \cdots \times \Omega_n, (\omega_1, \dots, \omega_k) \in \Omega_1 \times \cdots \times \Omega_k\}$$

(on notera que la suite de tribus  $(\mathcal{F}_k)$  associée à ce découpage progressif est celle mentionnée page 7), on obtient la concentration

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-r^2/2\|a\|_2^2}$$

L'intérêt est de pouvoir choisir  $a$  de sorte que la concentration soit indépendante de la dimension  $n$ , voire de plus en plus forte.

Une autre illustration, plus spécifique puisque le lemme 2.1.2 ne s'applique pas directement, est la concentration de  $\mathfrak{S}_n$  muni de la distance

$$d(\sigma, \tau) = \frac{1}{n} \#\{i \in \llbracket 1, n \rrbracket; \sigma(i) \neq \tau(i)\}$$

En considérant  $\mathcal{P}^k$  la partition de  $\mathfrak{S}_n$  formée des ensembles

$$A_{i_1, \dots, i_k} = \{\sigma \in \mathfrak{S}_n; \forall j \in \llbracket 1, k \rrbracket \quad \sigma(j) = i_j\}$$

et, pour  $A_{i_1, \dots, i_{k+1}}, A_{i_1, \dots, i'_{k+1}} \subset A_{i_1, \dots, i_k}$ ,

$$\phi^k : \sigma \in A_{i_1, \dots, i_{k+1}} \mapsto \sigma \circ (i_{k+1}, i'_{k+1}) \in A_{i_1, \dots, i'_{k+1}}$$

on voit que  $d(\sigma, \phi^k(\sigma)) = \frac{2}{n}$ , sauf si  $i_{k+1} = i'_{k+1}$ , auquel cas  $d(\sigma, \phi^k(\sigma)) = 0$ . D'où la concentration

$$\forall r \geq 0 \quad P(\{|F - E[F]| \geq r\}) \leq 2e^{-nr^2/8}$$

pour toute fonction  $F : \mathfrak{S}_n \rightarrow \mathbf{R}$  1-lipschitzienne.

Plus généralement, on a le résultat suivant dans les groupes compacts (qui donne le même résultat dans  $\mathfrak{S}_n$ ) :

**Corollaire 2.4.4** (Concentration normale dans un groupe compact). *Soit  $(G, d, \mu)$  un groupe topologique compact muni d'une distance  $d$  invariante par translation et de sa probabilité de Haar<sup>1</sup>  $P$ . Si*

$$G = G_0 \supset \cdots \supset G_n = \{e\}$$

*est une suite de sous-groupe fermés de  $G$ ,  $\Delta_k = \text{Diam}(G_{k-1}/G_k)$  (pour la distance  $\delta_k(gG_k, hG_k) = d(g, hG_k)$ ) et  $\Delta^2 = \sum_{k=1}^n \Delta_k^2$ , alors, pour toute fonction 1-lip.  $F : G \rightarrow \mathbf{R}$ ,*

$$\forall r \geq 0 \quad P(\{|F - E[F]|\} \geq r) \leq 2e^{-r^2/2\Delta^2}$$

*Démonstration.* La démonstration est analogue à celle dans les espaces de longueur au plus  $l$ , en partitionnant  $G$ , sauf que  $G$  n'est pas fini, *a priori* : on considère ici  $\mathcal{F}_k = \sigma((gG_k)_{g \in G})$  (en fait,  $\mathcal{P}^k = \{gG_k, g \in G\}$  est la partition subordonnée). On peut alors majorer  $\|d_k(F)\|_\infty$  par  $\Delta_k$  : en effet, si  $g_0 \in G$ , on choisit  $h_k, h_{k-1} \in G$  tels que  $g_0 \in h_k G_k \subset h_{k-1} G_{k-1}$ . Puis, si  $h'_k G_k \subset h_{k-1} G_{k-1}$ , sachant que  $G_k$  est compact et que  $h_{k-1}^{-1} h_k, h_{k-1}^{-1} h'_k \in G_{k-1}$ , il existe  $s_k \in G_k$  tel que

$$d(h_{k-1}^{-1} h_k, h_{k-1}^{-1} h'_k s_k) = d(h_{k-1}^{-1} h_k, h_{k-1}^{-1} h'_k G_k) = \delta_k(h_{k-1}^{-1} h_k G_k, h_{k-1}^{-1} h'_k G_k) \leq \Delta_k$$

Alors la translation (bijective)  $\phi_k : h_k r_k \in h_k G_k \mapsto h'_k s_k r_k \in h'_k G_k$  vérifie

$$d(h_k r_k, \phi_k(h_k r_k)) \leq \Delta_k$$

et

$$\int_{h'_k G_k} F dP = \int_{h_k G_k} F \circ \phi_k dP$$

d'où la conclusion par un raisonnement analogue à celui effectué pour démontrer le corollaire précédent.  $\square$

## 3 Distances de Hamming

### 3.1 Définitions et théorème principal

Dans cette partie, on étudiera la concentration de mesures produits et certaines propriétés de convexité.

Soient  $(\Omega_i, \Sigma_i, \mu_i)$  des espaces arbitraires de probabilité,  $P = \mu_1 \otimes \cdots \otimes \mu_n$  la mesure produit sur  $X = \Omega_1 \times \cdots \times \Omega_n$ . On notera  $|\cdot|$  la norme euclidienne lorsqu'il n'y aura pas d'ambiguïté et on rappelle que

$$d_a(x, y) = \sum_{i=1}^n a_i 1_{x_i \neq y_i}$$

On a alors :

---

<sup>1</sup>On appelle mesure de Haar sur un groupe topologique localement compact  $G$  une mesure  $\mu$  régulière sur la tribu borélienne de  $G$ , non triviale, et telle que, si  $B \subset G$  et  $g \in G$ ,  $\mu(gB) = \mu(B)$  (invariance par translation à gauche). En particulier

$$\int f(gh) d\mu(h) = \int f(h) d\mu(h)$$

De plus, on montre (cf. [Hal50]) que cette mesure existe et est unique, à une constante multiplicative près. Enfin, si  $G$  est compact, elle est finie (donc il existe une probabilité de Haar) et invariante par translation à gauche et à droite.

**Définition 3.1.1** (« Distance »  $\mathcal{D}_A^c(x)$ ). Soit  $X = \Omega_1 \times \cdots \times \Omega_n$ . Si  $x \in X$  et  $A \subset X$ , on définit la « distance »

$$\mathcal{D}_A^c(x) = \sup_{\substack{a \in \mathbf{R}_+^n \\ \|a\|_2=1}} d_a(x, A)$$

**Lemme 3.1.2.** En notant  $U_A(x) = \{s = (1_{x_i \neq y_i}), y \in A\}$  et  $V_A(x)$  son enveloppe convexe,

$$\mathcal{D}_A^c(x) = d(0, V_A(x))$$

(ces ensembles sont contenus dans  $[0, 1]^n$ ).

*Démonstration.* On a, pour  $\|a\|_2 = 1$ ,

$$d_a(x, A) = \inf_{y \in A} \sum_{i=1}^{i=n} a_i 1_{x_i \neq y_i} = \inf_{y \in U_A(x)} \langle a|y \rangle$$

Montrons maintenant que :

$$\inf_{y \in U_A(x)} \langle a|y \rangle = \inf_{y \in V_A(x)} \langle a|y \rangle$$

Déjà, puisque  $U_A(x) \subset V_A(x)$ ,

$$\inf_{y \in U_A(x)} \langle a|y \rangle \geq \inf_{y \in V_A(x)} \langle a|y \rangle$$

Puis, si  $y \in V_A(x)$ , il existe  $(\lambda_1, \dots, \lambda_n)$ , avec  $\sum_{i=1}^n \lambda_i = 1$ , et  $(y_1, \dots, y_n) \in U_A(x)^n$  tels que

$y = \sum_{i=1}^n \lambda_i y_i$ . On a alors

$$\langle a|y \rangle = \sum_{i=1}^n \lambda_i \langle a|y_i \rangle \geq \sum_{i=1}^n \lambda_i \inf_i \langle a|y_i \rangle \geq \inf_{y \in U_A(x)} \langle a|y \rangle.$$

□

*Remarque.* En notant  $d$  la dimension minimale d'un sous-espace affine passant par  $x$  et coupant  $A$ , et défini comme noyau de formes linéaires coordonnées dans la base canonique, si  $d$  est atteinte une seule fois sur les  $2^n$  sous-espaces affines considérés, alors  $\mathcal{D}_A^c(x) = \sqrt{n-d}$ . Sinon, c'est plus subtil (cf. déjà  $n=3$ ).

Pour  $X = [0, 1]^n$ , si  $C$  désigne l'enveloppe convexe de  $A$ , pour tout  $x \in X$ ,  $d(x, C) \leq \mathcal{D}_A^c(x)$  ( $d$  étant la distance euclidienne). En effet, soit  $x \in X$ . Si  $y \in C$ ,  $y = \sum_{k=1}^p \lambda_k y^{(k)}$ , avec  $\lambda_k \geq 0$  et  $\sum \lambda_k = 1$ . Alors

$$\begin{aligned} d(x, C)^2 &\leq \|x - y\|_2^2 \\ &= \sum_{i=1}^n \left( \sum_{k=1}^p \lambda_k (x_i - y_i^{(k)}) \right)^2 \\ &\leq \sum_{i=1}^n \left( \sum_{k=1}^p \lambda_k 1_{x_i \neq y_i^{(k)}} \right)^2 \\ &= d \left( 0, \sum_{k=1}^p \lambda_k \left( 1_{x_i \neq y_i^{(k)}} \right)_{i \in [1, n]} \right)^2 \end{aligned}$$

Il reste à passer à la borne inférieure sur les  $\lambda_k$  et à utiliser le lemme ci-dessus.  
Le théorème principal de cette partie est le suivant :

**Théorème 3.1.3** (Concentration normale de  $\mathcal{D}_A^c$ ). *Soit  $(X = \Omega^n, P = \mu^{\otimes n})$  un espace de probabilité et  $A \subset X$  mesurable non vide. Alors*

$$\int e^{(\mathcal{D}_A^c)^2/4} dP \leq \frac{1}{P(A)} \quad (3)$$

En particulier,

$$\forall r \geq 0 \quad P(\mathcal{D}_A^c \geq r) \leq \frac{1}{P(A)} e^{-r^2/4}$$

*Remarque.* Le théorème est plus fort qu'une simple inégalité de concentration sur  $d_a$ , pour  $a \in (\mathbf{R}_+)^n$ . En effet, il implique que

$$\forall r \geq 0 \quad 1 - P(A_r) \leq P(\mathcal{D}_A^c \geq r) \leq \frac{1}{P(A)} e^{-r^2/4}$$

où  $A_r$  est le  $r$ -voisinage ouvert de  $A$  au sens de  $d_a$ . On a donc une concentration normale uniforme sur tous les  $d_a$  pour  $|a| = 1$ .

*Démonstration.* Par récurrence sur  $n$ . Pour  $n = 1$ ,

$$\int e^{(\mathcal{D}_A^c)^2/4} dP = P(A) + (1 - P(A))e^{1/4} \leq \frac{1}{P(A)}$$

Supposons le résultat vrai pour  $n$ . Soit  $A \subset \Omega^{n+1}$  et  $B$  sa projection sur  $\Omega^n$ . Soit  $A(\omega)$  la section de  $A$  selon  $\omega$ . Pour  $x \in \Omega^n, \omega \in \Omega$ , posons  $z = (x, \omega)$ . Si  $s \in U_{A(\omega)}(x)$ , alors  $(s, 0) \in U_A(z)$ , et si  $t \in U_B(x)$ , alors  $(t, 1) \in U_A(z)$ . Il en découle que si  $\xi \in V_{A(\omega)}(x)$ ,  $\zeta \in V_B(x)$  et  $0 \leq \theta \leq 1$ , alors  $(\theta\xi + (1 - \theta)\zeta, 1 - \theta) \in V_A(z)$ . Puisque  $\mathcal{D}_A^c(x) = \|y\|_2$ , et par convexité de la fonction carré :

$$\mathcal{D}_A^c(z)^2 \leq (1 - \theta)^2 + |\theta\xi + (1 - \theta)\zeta|^2 \leq (1 - \theta)^2 + \theta|\xi|^2 + (1 - \theta)|\zeta|^2$$

D'où

$$\mathcal{D}_A^c(z)^2 \leq (1 - \theta)^2 + \theta\mathcal{D}_{A(\omega)}^c(x)^2 + (1 - \theta)\mathcal{D}_B^c(x)^2$$

En utilisant l'inégalité de Hölder et l'hypothèse de récurrence :

$$\begin{aligned} \int_{\Omega^n} e^{\mathcal{D}_A^c(x, \omega)^2/4} dP(x) &\leq e^{(1-\theta)^2/4} \left( \int_{\Omega^n} e^{\mathcal{D}_{A(\omega)}^c(x)^2/4} dP \right)^\theta \left( \int_{\Omega^n} e^{\mathcal{D}_B^c(x)^2/4} dP \right)^{1-\theta} \\ &\leq e^{(1-\theta)^2/4} \left( \frac{1}{P(A(\omega))} \right)^\theta \left( \frac{1}{P(B)} \right)^{1-\theta}. \end{aligned}$$

Aussi :

$$\int_{\Omega^n} e^{\mathcal{D}_A^c(x, \omega)^2/4} dP(x) \leq \frac{1}{P(B)} e^{(1-\theta)^2/4} \left( \frac{P(A(\omega))}{P(B)} \right)^{-\theta}$$

Choisissons maintenant un  $\theta$  convenable. On sait que pour tout  $u \in [0, 1]$ ,

$$\inf_{\theta \in [0, 1]} e^{(1-\theta)^2/4} u^{-\theta} \leq 2 - u$$

On prend  $\theta = 1 + 2 \ln u$  si  $u \geq e^{-1/2}$ ,  $\theta = 0$  sinon, on passe au logarithme, et cela suffit pour montrer que

$$\ln(2 - u) + \ln u + (\ln u)^2 \geq 0$$

On obtient donc :

$$\int_{\Omega^n} e^{\mathcal{D}_A^c(x, \omega)^2/4} dP(x) \leq \frac{1}{P(B)} \left( 2 - \frac{P(A(\omega))}{P(B)} \right)$$

On intègre selon  $\omega$  et, par le théorème de Fubini, on a :

$$\begin{aligned} \int_{\Omega^n} e^{\mathcal{D}_A^c(x, \omega)^2/4} dP(x) &\leq \frac{1}{P(B)} \left( 2 - \frac{P \otimes \mu(A)}{P(B)} \right) \\ &\leq \frac{1}{P \otimes \mu(A)} \end{aligned}$$

puisque  $u(2 - u) \leq 1$  pour tout réel  $u$ , ce qui achève la preuve.  $\square$

*Remarque.* Le résultat s'étend à  $X = \Omega_1 \times \dots \times \Omega_n$ . En effet, l'idée est d'appliquer le théorème à  $\Omega = X$ , muni de  $P = P_1 \otimes \dots \otimes P_n$ . Montrons l'inégalité 3 pour  $A = A_1 \times \dots \times A_n \subset \Omega_1 \times \dots \times \Omega_n$ . Posons

$$B = (A_1 \times \Omega_2 \times \dots \times \Omega_n) \times (\Omega_1 \times A_2 \times \dots \times \Omega_n) \times \dots \times (\Omega_1 \times \dots \times A_n) \subset \Omega^n$$

Déjà,  $P^n(B) = P(A)$ . De plus, pour tout  $a \in (\mathbf{R}_+)^n$  et  $x = (x_i^j) \in \Omega^n$ , en notant  $d_a$  la distance de Hamming sur  $X = \Omega_1 \times \dots \times \Omega_n$  et  $\tilde{d}_a$  celle sur  $\Omega^n$ ,

$$\begin{aligned} \tilde{d}_a(x, B) &= \inf_{y \in B} \sum_{i=1}^n a_i 1_{x_i \neq y_i} \\ &= \inf_{z \in A_1 \times \dots \times A_n} \sum_{i=1}^n a_i 1_{x_i^i \neq z_i} \\ &= d_a((x_i^i), A) \end{aligned}$$

la deuxième égalité venant de la définition de  $B$ . On en déduit

$$\tilde{\mathcal{D}}_B(x) = \mathcal{D}_A((x_i^i))$$

Puis, le théorème de Fubini-Tonnelli donne

$$\begin{aligned} \int_{\Omega^n} e^{(\tilde{\mathcal{D}}_B(x))^2} &= \int_{\Omega_n} \dots \int_{\Omega_1} \dots \int_{\Omega_n} \int_{\Omega_1} e^{[\mathcal{D}_A((x_i^i))]^2} dP_1(x_1^1) \dots dP_n(x_1^n) \dots dP_1(x_n^1) \dots dP_n(x_n^n) \\ &= \int_X e^{(\mathcal{D}_A(z))^2} dP(z) \end{aligned}$$

L'intérêt de l'opérateur  $\mathcal{D}_A^c(x)$ , est que l'on peut choisir le  $a$  (dépendant de  $x$ ) qui convient. On va illustrer cela par un exemple. Soit  $\Omega = \{0, 1\}$ ,  $\mu_p$  la mesure de probabilité telle que  $\mu_p(1) = p$ . Soit  $A \subset \Omega^n$  vérifiant : si  $x = (x_1, \dots, x_n) \in A$ , et si  $y = (y_1, \dots, y_n)$  vérifie pour tout  $0 \leq i \leq n$  :  $y_i \leq x_i$ , alors  $y \in A$ . Pour  $x \in \Omega^n$ , on pose  $J = \{i \in \llbracket 1, n \rrbracket; x_i = 1\}$ , et  $N(x) = \#J$ . Le choix de  $a$  vérifiant :  $a_i = 1$  si  $x \in J$ , et 0 sinon, montre que :

$$d(x, A) \leq \mathcal{D}_A^c(x) \sqrt{N(x)}$$

où  $d(x, A)$  est la distance de Hamming de  $x$  à  $A$ . En effet, on a

$$d(x, A) = \inf_{y \in A} \sum_{i=1}^n 1_{x_i \neq y_i} \leq \inf_{y \in A} \sum_{i=1}^n a_i 1_{x_i \neq y_i}$$

Deux cas se présentent :

- si  $i \in J$ ,  $a_i = 1$  et  $1_{x_i \neq y_i} = a_i 1_{x_i \neq y_i}$ ,
- si  $i$  n'est pas dans  $J$ ,  $a_i = 0$ , donc  $x_i = 0$ . Si  $x_i = y_i$ , on a encore  $1_{x_i \neq y_i} = a_i 1_{x_i \neq y_i}$ ; si  $x_i \neq y_i$ , alors  $y_i = 1$ . Mais puisque l'élément  $z = (y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_n)$  est encore

$$\text{dans } A, \text{ on a bien encore } \inf_{y \in A} \sum_{i=1}^n 1_{x_i \neq y_i} \leq \inf_{y \in A} \sum_{i=1}^n a_i 1_{x_i \neq y_i}$$

Finalement, on a  $d(x, A) \leq \inf_{y \in A} d_a(x, y) \sqrt{N(x)} \leq \mathcal{D}_A^c(x) \sqrt{N(x)}$ . Ainsi, pour tout  $r, s > 0$ ,

$$\begin{aligned} P(d(\cdot, A) \geq r) &\leq P(\mathcal{D}_A^c \geq rs^{1/2}) + P(y, N(y) > s) \\ &\leq \frac{1}{P(A)} e^{-r^2/4} + P(y, N(y) > s) \end{aligned}$$

puisque le dernier terme est très petit quand  $s > pn$ , on a bien :

$$P(d(\cdot, A) \geq r) \leq \frac{1}{P(A)} e^{-r^2/(4pn)}$$

Voyons ce qui se passe dans les espaces mesurés par la distance de Hamming introduite plus haut.

### 3.2 Fonctions Hamming-lipschitziennes

**Définition 3.2.1** (Application Hamming-lipschitzienne). Soit  $F : X = \Omega_1 \times \dots \times \Omega_n \rightarrow \mathbf{R}$ . On dira que  $F$  est Hamming-lipschitzienne si

$$\forall x \in X \quad \exists a(x) \in B_2(0, 1) \cap \mathbf{R}_+^n \quad \forall y \in X \quad F(x) \leq F(y) + d_{a(x)}(x, y)$$

**Corollaire 3.2.2** (Concentration normale des fonctions Hamming-lipschitziennes). Soit  $(X = \Omega^n, P = \mu^{\otimes n})$  un espace de probabilité et  $F : X \rightarrow \mathbf{R}$  Hamming-lipschitzienne. Alors, si  $m_F$  est une médiane de  $F$  pour  $P$ ,

$$\forall r \geq 0 \quad P(|F - m_F| \geq r) \leq 4e^{-r^2/4}$$

*Démonstration.* Soit  $A_m = \{F \leq m\}$ . Pour tout  $y \in A$ , il existe  $a(x)$  tel que

$$F(x) \leq F(y) + d_{a(x)}(x, y) \leq m + d_{a(x)}(x, y)$$

En passant à la borne inférieure sur  $y \in A$ ,

$$F(x) \leq m + d_{a(x)}(x, A) \leq m + \mathcal{D}_A^c(x)$$

D'où,

$$P(F \geq m + r) \leq P(\mathcal{D}_A^c \geq r) \leq \frac{1}{P(F \leq m)} e^{-r^2/4}$$

en utilisant le théorème; en appliquant cette inégalité à  $m = m_F$  et  $m = m_F - r$ , sachant que  $P(F \geq m_F) = P(F \leq m_F) = \frac{1}{2}$ , on obtient le résultat.  $\square$

**Théorème 3.2.3** (Concentration normale d'un supremum de variables aléatoires). Soit  $Y_k : \Omega \rightarrow [u_k, v_k]$  ( $k \in \llbracket 1, n \rrbracket$ ) des variables aléatoires indépendantes et  $\mathcal{T} \subset \mathbf{R}^n$  dénombrable tel que  $\left( \sum_{k=1}^n t_k^2 (v_k - u_k)^2 \right)_{t \in \mathcal{T}}$  soit bornée. Alors, en notant  $\sigma^2 = \sup_{t \in \mathcal{T}} \sum_{k=1}^n t_k^2 (v_k - u_k)^2$  (qu'on supposera  $> 0$ ) et  $Z = \sup_{t \in \mathcal{T}} \sum_{k=1}^n t_k Y_k$ ,

$$\forall r \geq 0 \quad P(|Z - m_Z| \geq r) \leq 4e^{-r^2/4\sigma^2}$$

De plus,  $|E[Z] - m_Z| \leq 4\sigma\sqrt{\pi}$  et  $\text{var}(Z) \leq 16\sigma^2$ .

*Démonstration.* On considère

$$F : X = \prod_{k=1}^n [u_k, v_k] \longrightarrow \mathbf{R}$$

$$x \longmapsto \sup_{t \in \mathcal{T}} \sum_{k=1}^n t_k x_k$$

Soit  $x \in X$  et  $t \in \text{Adh}_{\mathbf{R}^n}(\mathcal{T})$  tel que  $F(x) = \sum_{k=1}^n t_k x_k$  (on notera que  $t_k = \pm\infty$  entraîne  $u_k = v_k$ ; cf. remarque infra); alors, pour tout  $y \in X$ ,

$$F(x) \leq \sum_{k=1}^n t_k y_k + \sum_{k=1}^n |t_k(x_k - y_k)| \leq F(y) + \sigma \sum_{k=1}^n \frac{|t_k(v_k - u_k)|}{\sigma} 1_{x_k \neq y_k} \leq F(y) + \sigma d_a(x, y)$$

avec  $a = \left( \frac{|t_k(v_k - u_k)|}{N} \right)_{k \in \llbracket 1, n \rrbracket}$  où  $N = (\sum |t_i(v_i - u_i)|^2)^{1/2}$ . D'où le résultat en appliquant le théorème ci-dessus à  $\frac{F}{\sigma}$  et  $P_Y = P_{Y_1} \otimes \dots \otimes P_{Y_n}$  (on notera que  $m_Z$  est une médiane de  $F$  car  $P(Z \geq m_Z) = P(F(Y) \geq m_Z) = P_Y(F \geq m_Z)$ ).

On en déduit :

$$\begin{aligned} |E[Z] - m_Z| &= \left| \int (Z - m_Z) dP \right| \\ &\leq \int |Z - m_Z| dP \\ &= \int_0^{+\infty} P(|Z - m_Z| \geq r) dr \\ &\leq 4 \int_0^{+\infty} e^{-r^2/4\sigma^2} \\ &= 4\sigma\sqrt{\pi} \end{aligned}$$

et :

$$\text{var}(Z) \leq E[(Z - m_Z)^2] = \int_0^{+\infty} P(|Z - m_Z| \geq \sqrt{r}) dr \leq 4 \int_0^{+\infty} e^{-r/4\sigma^2} = 16\sigma^2$$

□

*Remarque.* La condition  $\left( \sum_{k=1}^n t_k^2 (v_k - u_k)^2 \right)_{t \in \mathcal{T}}$  bornée est équivalente à : pour tout  $k \in \llbracket 1, n \rrbracket$  tel que  $u_k < v_k$ ,  $(t_k)_{t \in \mathcal{T}}$  est bornée (équivalence des normes en dimension finie).

### 3.3 Cube $[0, 1]^n$

Voici une autre conséquence du théorème 3.1.3, sachant, comme nous l'avons remarqué plus haut que : si  $C$  est l'enveloppe convexe de  $A$ , pour tout  $x \in X$ ,  $d(x, C) \leq \mathcal{D}_A^c(x)$ .

**Théorème 3.3.1** (Concentration normale de  $d(\cdot, C)$ ). *Soit  $C \subset [0, 1]^n$  convexe. Alors (pour  $P = \mu^{\otimes n}$ ),*

$$\int e^{d(x, C)^2/4} dP(x) \leq \frac{1}{P(C)}$$

*En particulier, si  $r \geq 0$ ,  $P(d(\cdot, C) \geq r) \leq \frac{1}{P(C)} e^{-r^2/4}$ .*

*Démonstration.*  $d(x, C) \leq \mathcal{D}_C^c$  (cf. remarque page 14). □

*Remarque.* Si  $A$  est quelconque, on obtient

$$\int e^{d(x, C)^2/4} dP(x) \leq \frac{1}{P(A)}$$

où  $C$  est l'enveloppe convexe de  $A$ .

On en déduit le résultat :

**Théorème 3.3.2** (Concentration normale d'une fonction 1-lip. convexe). *Soit*

$$F : ([0, 1]^n, P = \mu^{\otimes n}) \rightarrow \mathbf{R}^n$$

*1-lipschitzienne et convexe. Alors*

$$\forall r \geq 0 \quad P(\{|F - m_F| \geq r\}) \leq 4e^{-r^2/4}$$

*Démonstration.* Soit  $m \in \mathbf{R}$  et  $C = \{F \leq m\}$ , qui est convexe, car  $F$  l'est. De plus, comme  $F$  est 1-lip.,  $\{d(\cdot, C) < r\} \subset \{F < m\}$ ; d'où, en utilisant le théorème ci-dessus,

$$P(F \leq m)P(F \geq m + r) \leq P(C)P(d(\cdot, C) \geq r) \leq e^{-r^2/4}$$

On conclut<sup>2</sup>, comme pour le théorème 3.2.2, en appliquant cette inégalité à  $m = m_F$  et à  $m = m_F - r$ . □

*Remarque.* Plus généralement, si  $F : \prod_{i=1}^n [u_i, v_i] \rightarrow \mathbf{R}$  est convexe et vérifie

$$\sigma^2 = \sup \left( \sum_{i=1}^n (v_i - u_i)^2 \left( \frac{\partial F}{\partial x_i} \right)^2 \right) < \infty$$

alors, comme pour les suprema de variables aléatoires indépendantes,

$$P(|F - m_F| \geq r) \leq 4e^{-r^2/4\sigma^2}$$

---

<sup>2</sup>Une autre démonstration utilise le théorème de Rademacher, qui dit que toute fonction convexe  $F : \mathbf{R}^n \rightarrow \mathbf{R}$  admet des dérivées partielles  $\lambda$ -presque partout. On en déduit

$$F(y) \geq F(x) + \sum_{i=1}^n (y_i - x_i) \frac{\partial F}{\partial x_i}$$

d'où

$$F(x) \leq F(y) + \sum_{i=1}^n 1_{x_i \neq y_i} \frac{dF}{dx_i} \leq F(y) + d_a(x, y)$$

où  $a = \alpha \left( \frac{\partial F}{\partial x_i} \right)_{i \in [1, n]}$  avec  $\alpha \geq 1$  de sorte que  $\|a\|_2 = 1$  ( $F$  étant 1-lip.,  $\|\nabla F\|_2 \leq 1$ ).

L'hypothèse de convexité est nécessaire, comme le montre le contre-exemple suivant :  
Soit  $P$  la mesure produit uniforme sur  $\{0, 1\}^n \subset [0, 1]^n$ , et prenons  $n = 2k$  pair. Soit la boule de Hamming,

$$A = \left\{ x \in \{0, 1\}^n; \sum_{i=1}^n x_i \leq k \right\}$$

et  $F(x) = d(x, A)$ , où  $d$  est la distance euclidienne. Il est clair que  $F$  est lipschtizienne avec  $\|F\|_{Lip} = 1$ , et que 0 est une médiane de  $F$ ; en effet,

$$P(F \geq 0) = 1 \geq 1/2,$$

et

$$\begin{aligned} P(F \leq 0) = P(F = 0) = P(A) &= P\left(\sum_{i=1}^n x_i \leq k\right) = \sum_{p=1}^k P\left(\sum_{i=1}^n x_i = p\right) \\ &= \sum_{p=1}^k C_n^p \frac{1}{2^n} = 1/2 \end{aligned}$$

Si  $x$  est tel que  $\sum_{i=1}^n (2x_i - 1) \geq \rho\sqrt{k}$  pour un  $\rho > 0$  à préciser ultérieurement, on montre que pour tout  $y \in A$  :

$$\frac{\rho\sqrt{k}}{2} \leq \frac{1}{2} \sum_{i=1}^n (2x_i - 1) \leq \frac{1}{2} \sum_{i=1}^n (2x_i - 2y_i) \leq \sum_{i=1}^n |x_i - y_i| = \sum_{i=1}^n (x_i - y_i)^2,$$

puisque les  $(x_i - y_i)$  sont dans  $\{-1, 0, 1\}$ .

Ainsi,  $d(x, A) \geq \sqrt{\frac{\rho}{2}} \left(\frac{n}{2}\right)^{1/4}$ . En appliquant le théorème central limite, on obtient que :

$$\frac{1}{\sqrt{n}} \left( \sum_{i=1}^n (x_i - E[x_1]) \right) = \frac{1}{\sqrt{n}} \left( \sum_{i=1}^n (x_i - 1/2) \right)$$

converge en loi vers  $\mathcal{N}(0, 1/4)$ . Ainsi :

$$P\left(\left\{x; d(x, A) \geq \sqrt{\frac{\rho}{2}} \left(\frac{n}{2}\right)^{1/4}\right\}\right) \geq P\left(\left\{x; \frac{1}{\sqrt{n}} \sum_{i=1}^n (2x_i - 1) \geq \rho/\sqrt{2}\right\}\right) \geq 1/4$$

pour  $\rho > 0$  suffisamment petit (il ne dépend pas de  $n$ ). On voit que la concentration ne s'applique pas.

Il existe aussi des théorèmes similaires pour des distances non quadratiques. Notons pour  $\beta \geq 0$  :

$$\mathcal{D}_{\beta, A}(x) = \inf_{y \in V_a(x)} \sum_{i=1}^n \tau_{\beta}(y_i)$$

avec

$$\tau_{\beta}(1 - u) = \beta u \ln u - (1 + \beta u) \ln \left( \frac{1 + \beta u}{1 + \beta} \right)$$

pour  $u \in [0, 1]$ .

On a le théorème :

**Théorème 3.3.3.** *Pour tout sous-ensemble  $A$  non vide de  $X = \omega^1 \times \dots \times \omega^n$ , et pour toute mesure de probabilité  $P$ , alors*

$$\int e^{\mathcal{D}_{\beta,A}} dP \leq \frac{1}{P(A)^\beta}$$

En particulier,

$$\forall r \geq 0 \quad P(\mathcal{D}_{\beta,A} \geq r) \leq \frac{1}{P(A)^\beta} e^{-r}$$

## 4 Contrôle de coordonnées

### 4.1 Définition et théorème principal

La partie suivante traite d'un contrôle non métrique par un nombre fini de points.

**Définition 4.1.1** (« Distance »  $\mathcal{D}^q(x)$ ). Soit  $X = \Omega_1 \times \dots \times \Omega_n$ . Si  $x \in X$  et  $A^1, \dots, A^q \subset X$ , on définit la "distance"

$$\mathcal{D}^q(x) = \inf \left\{ k = \# \{ i \in [1, n] \mid x \notin \{y_i^1, \dots, y_i^q\} \}, (y^1, \dots, y^q) \in A^1 \times \dots \times A^q \right\}$$

et  $\mathcal{D}^q = \infty$  si l'un des  $A^j$  est vide.

Voici le théorème principal de cette partie :

**Théorème 4.1.2** (Concentration exponentielle de  $\mathcal{D}^q$ ). Soit  $(X = \Omega^n, P = \mu^{\otimes n})$  un espace de probabilité et  $A^1, \dots, A^q \subset X$  mesurables non vides. Alors

$$\int q^{\mathcal{D}^q} dP \leq \prod_{i=1}^q \frac{1}{P(A^i)}$$

En particulier,

$$\forall k \geq 0 \quad P(\mathcal{D}^q \geq k) \leq q^{-k} \prod_{i=1}^q \frac{1}{P(A^i)}$$

*Démonstration.* On démontre le théorème par récurrence sur  $n$ . On a besoin du lemme suivant :

**Lemme 4.1.3.** Soient  $g_i, i \in [1, q]$ , des fonctions définies sur  $\Omega$  vérifiant  $0 \leq g_i \leq 1$ . On a :

$$\int \min \left( q, \min_{1 \leq i \leq q} \frac{1}{g_i} \right) d\mu \left( \prod_{i=1}^q \int g_i d\mu \right) \leq 1$$

*Démonstration.* Déjà, si  $g$  est une fonction sur  $\Omega$  telle que  $\frac{1}{q} \leq g \leq 1$ , alors :

$$\int \frac{1}{g} d\mu \left( \int g d\mu \right)^q \leq 1.$$

En effet, comme  $\ln u \leq u - 1$  pour  $u \geq 0$ , il suffit de prouver que

$$\int \frac{1}{g} d\mu + q \left( \int g d\mu \right) = \int \left( \frac{1}{g} + qg \right) d\mu \leq q + 1$$

ce qui est manifestement vrai car  $\frac{1}{u} + qu \leq q + 1$  pour  $u \in [0, 1]$ .

On applique alors cette inégalité à  $g$ , où  $\frac{1}{g} = \min \left( q, \min_{1 \leq i \leq q} \frac{1}{g_i} \right)$ , ce qui donne le résultat, car  $g_i \leq g$  pour tout  $i$ .  $\square$

Reprenons la démonstration du théorème. Si  $n = 1$ , on applique l'inégalité précédente avec  $g_i = 1_{A^i}$ . On suppose le résultat vrai pour  $n$ , montrons-le pour  $n + 1$ . Considérons les sous-ensembles  $A^1, \dots, A^q$  de  $\Omega^{n+1}$ . Pour  $\omega \in \Omega$ , notons  $A^i(\omega)$  les sections des  $A^i$ ,  $i \in \llbracket 1, q \rrbracket$ , et  $B^i$  leur projection sur  $\Omega^n$ . Si l'on applique l'inégalité précédente avec  $g_i = \frac{P(A^i(\omega))}{P(B^i)}$ , on obtient, en appliquant Fubini,

$$\int \min \left( q \prod_{i=1}^q \frac{1}{P(B^i)}, \min_{1 \leq j \leq q} \frac{1}{P(C^{ij})} \right) d\mu \leq \prod_{i=1}^q \frac{1}{P \otimes \mu(A^i)}$$

où  $C^{ij} = B^i$  si  $i \neq j$ , et  $C^{ii} = A^i(\omega)$ . La remarque importante est la suivante : pour  $(x, \omega) \in \Omega \times \Omega$ ,

$$\mathcal{D}_{A^1, \dots, A^q}^q(x, \omega) \leq 1 + \mathcal{D}_{B^1, \dots, B^q}^q(x)$$

et pour tout  $1 \leq j \leq q$ ,

$$\mathcal{D}_{A^1, \dots, A^q}^q(x, \omega) \leq \mathcal{D}_{C^{1j}, \dots, C^{qj}}^q(x).$$

Il s'ensuit que

$$\begin{aligned} & \int_{\Omega \times \Omega^n} q^{\mathcal{D}_{A^1, \dots, A^q}^q(x, \omega)} dP(x) d\mu(\omega) \\ & \leq \int_{\Omega \times \Omega^n} \min \left( q \cdot q^{\mathcal{D}_{B^1, \dots, B^q}^q(x)}, \min_{1 \leq j \leq q} q^{\mathcal{D}_{C^{1j}, \dots, C^{qj}}^q(x)} \right) dP(x) d\mu(\omega) \\ & \leq \int_{\Omega} \min \left( q \int_{\Omega^n} q^{\mathcal{D}_{B^1, \dots, B^q}^q(x)} dP(x), \min_{1 \leq j \leq q} \int_{\Omega^n} q^{\mathcal{D}_{C^{1j}, \dots, C^{qj}}^q(x)} dP(x) \right) d\mu(\omega) \\ & \leq \int \min \left( q \prod_{i=1}^q \frac{1}{P(B^i)}, \min_{1 \leq j \leq q} \frac{1}{P(C^{ij})} \right) d\mu \\ & \leq \prod_{i=1}^q \frac{1}{P \otimes \mu(A^i)} \end{aligned}$$

par l'hypothèse de récurrence, et donc, d'après ce qui précède,

$$\int_{\Omega \times \Omega^n} q^{\mathcal{D}_{A^1, \dots, A^q}^q(x, \omega)} dP(x) d\mu(\omega) \leq \prod_{i=1}^q \frac{1}{P \otimes \mu(A^i)}$$

ce qui achève la récurrence.  $\square$

Ce théorème donne un moyen de contrôler, par un sous ensemble fixé  $A$ , des éléments arbitraires dont on néglige un nombre de coordonnées décroissant exponentiellement. C'est le sens de la seconde inégalité. Le théorème trouve une application dans le cadre suivant.

## 4.2 Fonctions croissantes et sous-additives

**Définition 4.2.1** (Fonction croissante et sous-additive). Soit  $X = \Omega_1 \times \dots \times \Omega_n$ .

On note  $\Omega^* = \left\{ \Omega_I = \prod_{i \in I} \Omega_i, I \subset \llbracket 1, n \rrbracket \right\}$ . Alors, si  $F : \Omega^* \rightarrow \mathbf{R}_+$ , on dit que  $F$  est

- croissante si :  $I \subset J \Rightarrow F(x_I) \leq F(x_J)$
- sous-additive si :  $I \cap J = \emptyset \Rightarrow F(x_{I \sqcup J}) \leq F(x_I) + F(x_J)$

Par exemple,  $F(x_I) = \sum_{i \in I} x_i$  sur  $\Omega_i = [0, +\infty[$  (ou  $\Omega_i = [0, +\infty]$ ).

Ou encore, soit  $\Omega_i \subset E$ , où  $(E, \|\cdot\|)$  est un evn ; alors, si  $\epsilon_i : \Omega_i \rightarrow \{\pm 1\}$  sont des variables aléatoires i.i.d. de loi de Bernoulli de paramètre  $p = \frac{1}{2}$ ,

$$F(x_I) = E \left[ \left\| \sum_{i \in I} \epsilon_i(x_i) x_i \right\| \right]$$

est

– croissante : si  $I \subsetneq \llbracket 1, n \rrbracket$  et  $j \notin I$ ,

$$\begin{aligned} E \left[ \left\| \sum_{i \in I} \epsilon_i(x_i) x_i + \epsilon_j(x_j) x_j \right\| \right] &= E \left[ 1_{\{\epsilon_j=1\}} \left\| \sum_{i \in I} \epsilon_i(x_i) x_i + x_j \right\| \right] + E \left[ 1_{\{\epsilon_j=-1\}} \left\| \sum_{i \in I} \epsilon_i(x_i) x_i - x_j \right\| \right] \\ &= \frac{1}{2} E \left[ \left\| \sum_{i \in I} \epsilon_i(x_i) x_i + x_j \right\| + \left\| \sum_{i \in I} \epsilon_i(x_i) x_i - x_j \right\| \right] \\ &\geq F(x_I) \end{aligned}$$

en remarquant que

$$\|m\| = \left\| \frac{1}{2} ((m+x) + (m-x)) \right\| \leq \frac{1}{2} (\|m+x\| + \|m-x\|)$$

– sous-additive : par inégalité triangulaire.

**Théorème 4.2.2** (Déviation d'une fonction croissante et sous-additive). *Soit  $(X = \Omega_1 \times \dots \times \Omega_n, P = \mu^{\otimes n})$  un espace de probabilité et  $F : \Omega^* \rightarrow \mathbf{R}_+$  croissante et sous-additive. Si  $P(F|_X \leq m) \geq \frac{1}{M}$ , alors, pour tous  $k, q \in \mathbf{N}^*$ ,*

$$\forall r \geq 0 \quad P(F|_X \geq qm + r) \leq M^q q^{-k-1} + P(\{x \in X; \exists I \subset \llbracket 1, n \rrbracket, \#I \leq k \text{ et } F(x_I) \geq r\})$$

*Démonstration.* Soit  $A = \{F \leq m\}$  et  $\mathcal{D}^q$  pour  $A_1 = \dots = A^q = A$ . On a :

$$\{F \geq qm + r\} \subset (\{F \geq qm + r\} \cap \{\mathcal{D}^q \leq k\}) \sqcup \{\mathcal{D}^q \geq k+1\}$$

Déjà, avec le théorème 4.1.2,  $P(\mathcal{D}^q \geq k+1) \leq M^q q^{-k-1}$ . D'autre part, pour tout  $x \in \{\mathcal{D}^q \leq k\}$ , il existe  $y^1, \dots, y^q \in A$  tels que, en notant  $I = \{i \in \llbracket 1, n \rrbracket \mid x_i \notin \{y_i^1, \dots, y_i^q\}\}$ ,  $\#I \leq k$ . Alors, si  $i \in I^c$ , il existe  $\tilde{j}(i) \in [1, q]$  tel que  $x_i = y_i^{\tilde{j}(i)}$  ; on obtient ainsi une partition  $(J_j)_{j \in [1, q]}$  de  $I^c$  en définissant  $J_j = \tilde{j}^{-1}(j)$  ; puis

$$\begin{aligned} F(x) &\leq F(x_I) + F(x_{I^c}) \leq F(x_I) + \sum_{j=1}^q F(x_{J_j}) \leq F(x_I) + \sum_{j=1}^q F(y_{J_j}^j) \\ &\leq F(x_I) + \sum_{j=1}^q F(y^j) \leq F(x_I) + qm \end{aligned}$$

en notant que  $F(y^j) \leq m$ . En particulier, si  $F(x) \geq qm + r$ ,  $F(x_I) \geq r$ . D'où le résultat.  $\square$

Voici un cas particulier d'application à des variables aléatoires :

**Théorème 4.2.3** (Déviation d'un supremum de variables aléatoires  $\geq 0$ ). Soit  $Y_1, \dots, Y_n$  des variables aléatoires indépendantes à valeurs dans  $\mathbf{R}_+$ ,  $\mathcal{T} \subset \mathbf{R}_+^n$ ,  $S = \sup_{t \in \mathcal{T}} \sum_{i=1}^n t_i Y_i$  et  $\sigma = \sup_{t \in \mathcal{T}} \|t\|_\infty$ . Quitte à redéfinir<sup>3</sup> les  $Y_i$ , on peut supposer  $Y_1 \geq \dots \geq Y_n$ . Si  $P(S \leq m) \geq \frac{1}{M}$ , alors

$$\forall k, q \in \mathbf{N}^* \quad \forall r \geq 0 \quad P(S \geq qm + r) \leq M^q q^{-k-1} + P\left(\sigma \sum_{i=1}^k Y_i \geq r\right)$$

*Démonstration.* On applique le théorème précédent à  $F : x \in \Omega^n \mapsto \sup_{t \in \mathcal{T}} \sum_{i=1}^n t_i Y_i(x_i)$  (qui est bien croissante et sous-additive).  $\square$

En particulier, pour  $\mathcal{T} = \{(1, \dots, 1)\}$ ,  $S = \sum_{i=1}^n Y_i$  et on obtient

$$\forall k, q \in \mathbf{N}^* \quad \forall r \geq 0 \quad P(S \geq qm + r) \leq M^q q^{-k-1} + P\left(\sum_{i=1}^k Y_i \geq r\right)$$

De plus, si, pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $|Y_i| \leq C$ , il vient, pour  $k \geq Cr$ ,  $P(S \geq qm + r) \leq M^q q^{-k-1}$ .

## 5 Convolution par infimum

### 5.1 Définitions et propriétés

Nous introduisons ici une condition suffisante pour avoir de la concentration dans un espace, condition qui est très agréable dans le cas espaces produits car il suffit de la vérifier sur chacun des facteurs. Commençons par introduire la notion de convolution par infimum.

**Définition 5.1.1** (Convolution par infimum). Soit  $X$  un ensemble et  $c : X \times X \rightarrow \mathbf{R}_+$  une application. Si  $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$  est minorée, on définit la convolution de  $f$  avec le coût  $c$

$$Q_c f(x) = \inf_{y \in X} (f(y) + c(x, y))$$

*Remarque.* Etant donné que  $f$  et  $c$  sont minorées, la fonction  $Q_c f$  est minorée elle aussi (par  $\inf f$ ). De plus,  $Q_c f$  prend la valeur  $+\infty$  si et seulement si  $f = +\infty$  et, dans les autres cas,  $Q_c f$  est à valeurs dans  $\mathbf{R}$ .

On notera que, même si  $(X, \mathcal{A})$  est un espace mesurable et si  $c$  et  $f$  sont mesurables, la fonction  $Q_c f$  ainsi définie n'a pas de raison de l'être. Toutefois, en introduisant des hypothèses de convexité, on peut se placer dans un cadre où tout se passe bien et qui permet de généraliser le résultat 3.3.1.

<sup>3</sup>Par récurrence, on définit, pour  $\omega \in \Omega$ ,

$$Y_1^*(\omega) = \max_{i \in \llbracket 1, n \rrbracket} Y_i(\omega) = Y_{i_1}(\omega)$$

puis

$$Y_2^*(\omega) = \max_{i \in \llbracket 1, n \rrbracket \setminus \{i_1\}} Y_i(\omega) = Y_{i_2}(\omega)$$

et ainsi de suite. On notera que  $S = \sum_{i=1}^n Y_i = \sum_{i=1}^n Y_i^*$ .

Désormais, nous supposons donc que  $X$  est un espace vectoriel normé, muni de sa tribu borélienne  $\mathcal{B}$  (nous nous placerons dans  $\mathbf{R}^n$  pour les applications). On peut montrer que, si  $g : X \rightarrow \mathbf{R}$  est convexe et minorée sur un ouvert  $U \subset X$ , alors  $g$  est continue sur  $X$  (cf. [Rob73]) et, en particulier, mesurable. Alors, en supposant  $f$  convexe minorée et  $c$  convexe, montrons que la fonction  $Q_c f$  est convexe<sup>4</sup> (donc mesurable).

En effet,  $F : (x, y) \in X^2 \mapsto f(y) + c(x, y)$  est convexe et minorée et  $Q_c f(x) = \inf\{F(x, y), y \in X\}$ . Soit  $x_1, x_2 \in X$ ,  $\lambda \in [0, 1]$  et  $\epsilon > 0$ . Il existe  $y_1, y_2 \in X$  tels que, pour  $i \in \{1, 2\}$ ,  $F(x_i, y_i) \leq Q_c f(x_i) + \epsilon$ . Alors

$$\begin{aligned} Q_c f((1 - \lambda)x_1 + \lambda x_2) &\leq F((1 - \lambda)x_1 + \lambda x_2, (1 - \lambda)y_1 + \lambda y_2) \\ &\leq (1 - \lambda)F(x_1, y_1) + \lambda F(x_2, y_2) \\ &\leq (1 - \lambda)Q_c f(x_1) + \lambda Q_c f(x_2) + \epsilon \end{aligned}$$

d'où le résultat, étant donné que  $\epsilon$  peut être choisi arbitrairement petit<sup>5</sup>. Avec ces hypothèses, introduisons la notion d'inégalité de convolution.

**Définition 5.1.2** (Inégalité de convolution convexe). Soit  $(X, \mathcal{A}, P)$  un espace de probabilité et  $c : X \times X \rightarrow \mathbf{R}_+$  une fonction de coût mesurable et convexe. On dira que  $P$  vérifie une *inégalité de convolution convexe par rapport à  $c$  (ICC)* si, pour toute fonction  $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$  mesurable convexe et minorée,

$$\left( \int e^{Q_c f} dP \right) \left( \int e^{-f} dP \right) \leq 1$$

avec, par convention,  $\infty \cdot 0 = 0$ .

La propriété qui suit justifie l'introduction de ces définitions en faisant le lien avec la concentration.

**Propriété 5.1.3** (Conséquence de l'ICC). Si  $P$  vérifie une ICC, alors, pour tout  $C \in \mathcal{B}$  convexe,

$$\forall r > 0 \quad P\left(\left\{ \inf_{y \in C} c(\cdot, y) \geq r \right\}\right) \leq \frac{1}{P(C)} e^{-r}$$

En particulier, si  $c(x, y) = \tilde{c}(x - y)$ ,

$$1 - P(\{C + \{\tilde{c} < r\}\}) \leq \frac{1}{P(C)} e^{-r}$$

*Démonstration.* L'inégalité est conséquence directe de la définition de l'ICC avec la fonction  $f = \infty \cdot 1_C$ , qui est mesurable minorée et convexe (car  $C$  est convexe).  $\square$

On obtient ainsi des résultats de concentration si  $\tilde{c}$  est de la forme :  $\tilde{c}(x) = \alpha \|x\|^\beta$ . En effet, l'inégalité ci-dessus se réécrit alors, pour  $C$  convexe, avec  $P(C) \geq 2$ ,

$$1 - P(C_r) \leq \frac{1}{P(C)} e^{-\alpha r^\beta} \leq 2e^{-\alpha r^\beta}$$

<sup>4</sup>En vertu de la remarque précédente, ou bien  $f = +\infty$  et alors  $Q_c f = +\infty$  est mesurable; ou bien,  $f \neq +\infty$  et alors  $Q_c f : X \rightarrow \mathbf{R}$ , et le résultat de [Rob73] s'applique.

<sup>5</sup>Plus généralement, on vient de montrer que, si  $F : X \times Y \rightarrow \mathbf{R}$  est convexe minorée,  $x \in X \mapsto \inf_{y \in Y} F(x, y)$  est convexe.

## 5.2 Comportement dans les espaces produits

Sachant qu'une ICC donne lieu à de la concentration, intéressons-nous maintenant aux espaces produits. Etant donnés deux espaces  $(X_1, P_1)$  et  $(X_2, P_2)$  dans lesquels  $P_i$  ( $i \in \{1, 2\}$ ) vérifie une ICC par rapport à un coût  $c_i$ , on peut se demander ce qu'il en est de  $P_1 \otimes P_2$ . Il s'avère que la notion d'ICC se comporte très bien vis-à-vis du produit, comme le montre le résultat fondamental suivant.

**Théorème 5.2.1** (ICC dans un produit). *Si  $P_1$  vérifie une ICC sur  $X_1$  par rapport à  $c_1$ , et  $P_2$  une ICC sur  $X_2$  par rapport à  $c_2$ , alors  $P_1 \otimes P_2$  vérifie une ICC sur  $X_1 \times X_2$  par rapport à la fonction de coût convexe*

$$c_1 + c_2 : (x_1, x_2, y_1, y_2) \in (X_1 \times X_2)^2 \mapsto c_1(x_1, y_1) + c_2(x_2, y_2)$$

*Démonstration.* Soit  $f : X_1 \times X_2 \rightarrow \mathbf{R} \cup \{+\infty\}$  une fonction convexe et minorée. Introduisons deux fonctions, l'une sur  $X_1$  et l'autre sur  $X_2$ , auxquelles on appliquera les ICC de  $P_1$  et  $P_2$ . Tout d'abord, soit

$$f^{x_2} : x_1 \in X_1 \mapsto f(x_1, x_2)$$

qui est une fonction convexe et minorée par  $\inf f$ ; nous pouvons donc ensuite définir ( $Q_{c_1} f^{x_2}$  étant mesurable)

$$g : x_2 \in X_2 \mapsto \ln \left( \int e^{Q_{c_1} f^{x_2}(x_1)} dP_1(x_1) \right)$$

Comme  $Q_{c_1} f^{x_2}$  est minorée par  $\inf f$ ,  $g$  est aussi minorée (par  $\inf f$ ). Vérifions que  $g$  est convexe. Si  $x_2, y_2 \in X_2$  et  $\lambda \in [0, 1]$ , on a :

$$\begin{aligned} g((1-\lambda)x_2 + \lambda y_2) &= \ln \int \exp \left[ \inf_{y_1 \in X_1} \left( f(y_1, (1-\lambda)x_2 + \lambda y_2) + c_1(x_1, y_1) \right) \right] \\ &\leq \ln \int \exp \left[ (1-\lambda) \inf_{y_1 \in X_1} \left( f(y_1, x_2) + c_1(x_1, y_1) \right) + \lambda \inf_{y_1 \in X_1} \left( f(y_1, y_2) + c_1(x_1, y_1) \right) \right] \\ &= \ln \int \left( e^{Q_{c_1} f^{x_2}} \right)^{1-\lambda} \left( e^{Q_{c_1} f^{y_2}} \right)^\lambda \\ &\leq (1-\lambda)g(x_2) + \lambda g(y_2) \end{aligned}$$

La première inégalité ci-dessus est conséquence de la note 5 appliquée, pour tout  $x_1 \in X_1$ , à la fonction convexe minorée  $F(y_1, z_2) = f(y_1, z_2) + c_1(x_1, y_1)$ ; la seconde découle de l'inégalité de Hölder appliquée aux fonctions  $e^{Q_{c_1} f^{x_2}}$  et  $e^{Q_{c_1} f^{y_2}}$  et à  $\frac{1}{p} = 1 - \lambda$ ,  $\frac{1}{q} = \lambda$ .

Notons maintenant que, pour tout  $x_2 \in X_2$ ,

$$\begin{aligned} \int e^{Q_{c_1+c_2} f(x_1, x_2)} dP_1(x_1) &= \int \exp \left( \inf_{\substack{y_1 \in X_1 \\ y_2 \in X_2}} f(x_2, y_2) + c_1(x_1, y_1) + c_2(x_2, y_2) \right) dP_1(x_1) \\ &\leq \inf_{y_2 \in X_2} \left( \int \exp \left( \inf_{y_1 \in X_1} f(x_2, y_2) + c_1(x_1, y_1) + c_2(x_2, y_2) \right) dP_1(x_1) \right) \\ &= \inf_{y_2 \in X_2} \left( \int e^{Q_{c_1} f^{y_2}(x_1)} dP_1(x_1) \cdot e^{c_2(x_2, y_2)} \right) \\ &= e^{Q_{c_2} g(x_2)} \end{aligned}$$

D'où, en utilisant le théorème de Fubini-Tonnelli et les hypothèses d'ICC pour  $P_1$  (avec  $f^{x_2}$ ) puis  $P_2$  (avec  $g$ ), il vient

$$\begin{aligned} \int e^{Q_{c_1+c_2}f} dP_1 \otimes P_2 \int e^{-f} dP_1 \otimes P_2 &\leq \int e^{Q_{c_2}g} dP_2 \int \left( \int e^{-f^{x_2}} dP_1 \right) dP_2 \\ &\leq \int e^{Q_{c_2}g} dP_2 \int e^{-g} dP_2 \\ &\leq 1 \end{aligned}$$

ce qu'on voulait. □

Passons à des applications de ce résultat.

### 5.3 Conséquences

Tout d'abord, généralisons par ce biais l'inégalité 3.3.1.

**Corollaire 5.3.1.** *Soit  $(X_i, \|\cdot\|_i)$  des evn munis, chacun, d'une probabilité  $P_i$  dont le support est de diamètre  $\leq 1$ . Alors,  $P = P_1 \otimes \cdots \otimes P_n$  vérifie une ICC sur  $X = X_1 \times \cdots \times X_n$  par rapport au coût  $c$  associé à*

$$\tilde{c}(x) = \frac{1}{4} \sum_{i=1}^n \|x_i\|_i^2$$

*Démonstration.* Avec le lemme précédent, il suffit de montrer le résultat pour  $n = 1$ . Notons  $X_1 = X$ ,  $\|\cdot\|_1 = \|\cdot\|$ ,  $P_1 = P$  et  $A$  le support de  $P$ .

Soit  $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$  minorée et convexe. Quitte à retrancher  $\inf_A f$  (ce qui multiplie

l'ICC par  $e^{-\inf_A f}$  des deux côtés), supposons  $\inf_A f = 0$ . Il nous suffit de montrer que

$$\forall x \in A \quad e^{Q_c f(x)} \leq 2 - e^{-f(x)}$$

L'ICC pour  $f$  est alors obtenue en intégrant et grâce à l'inégalité  $y + \frac{1}{y} \geq 2$ , appliquée à  $y = \int e^{-f} dP > 0$ .

Fixons  $x \in A$ . Pour tout  $\epsilon > 0$ , il existe  $a_\epsilon \in A$  tel que  $f(a_\epsilon) \leq \epsilon$ ; alors, pour tout  $\lambda \in [0, 1]$  et  $y_\lambda = \lambda x + (1 - \lambda)a_\epsilon$ , on a, en utilisant l'hypothèse de convexité de  $f$  :

$$\begin{aligned} Q_c f(x) &\leq f(y_\lambda) + \frac{1}{4} \|x - y_\lambda\|^2 \\ &\leq \lambda f(x) + (1 - \lambda) f(a_\epsilon) + \frac{1}{4} (1 - \lambda)^2 \|x - a_\epsilon\|^2 \\ &\leq \lambda f(x) + (1 - \lambda) \epsilon + \frac{1}{4} (1 - \lambda)^2 \end{aligned}$$

Cela étant vrai pour tout  $\epsilon > 0$ , il vient

$$Q_c f(x) \leq \lambda f(x) + \frac{1}{4} (1 - \lambda)^2 = g(\lambda)$$

On majore alors ainsi :

- si  $f(x) = u \geq \frac{1}{2}$ ,  $Q_c f(x) \leq g(0) = \frac{1}{4}$ ; puis  $e^{1/4} \leq 2 - e^{-u}$  pour  $u \geq \frac{1}{2}$

– si  $f(x) = u \in [0, \frac{1}{2}]$ ,  $Q_c f(x) \leq g(1 - 2f(x)) = u - u^2$ ; puis

$$e^{u-u^2} + e^{-u} = e^{-u^2/2} \cdot 2 \cosh(u - \frac{u^2}{2}) \leq 2e^{-u^2/2} \cosh(u) \leq 2$$

(car  $0 \leq u - \frac{u^2}{2} \leq u$ )

D'où le résultat voulu.  $\square$

En appliquant le théorème à  $X_i = \mathbf{R}$ ,  $\|\cdot\|_i = |\cdot|$ ,  $P_i = \lambda_{[0,1]}$  et  $f = \infty \cdot 1_{C^c}$  (pour  $C \subset [0, 1]^n$  convexe), qui est minorée et convexe, on retrouve (cf. 3.3.1)

$$\int e^{d(x,C)^2/4} dP(x) = \int e^{Q_c f(x)} dP(x) \leq \left( \int e^{-f} dP \right)^{-1} = \frac{1}{P(C)}$$

La méthode des ICC donne aussi des résultats dans  $(\mathbf{R}^n, \mathcal{B})$ , muni de la mesure produit  $\nu_p^{\otimes n}$  ( $p \in [1, +\infty[$ ) où  $\nu_p$  est la probabilité sur  $\mathbf{R}$  de densité  $q_p(x) = \alpha_p e^{-|x|^p}$  (en particulier,  $\alpha_1 = \frac{1}{2}$ ). On a le résultat suivant :

**Corollaire 5.3.2** ("Concentration" exponentielle de  $\nu_p^n$ ). *Notons  $\mathbb{B}_i^n$  la boule unité de  $(\mathbf{R}^n, \|\cdot\|_i)$ . Alors, pour tout  $p \in [1, +\infty[$ , il existe  $K_p > 0$  tel que*

$$\forall A \in \mathcal{B} \quad \forall r > 0 \quad 1 - \nu_p^{\otimes n}(A + \sqrt{r}\mathbb{B}_2^n + r^{1/p}\mathbb{B}_p^n) \leq \frac{1}{\nu_p^{\otimes n}(A)} e^{-r/K_p}$$

Pour  $p = 1$ , on a, plus précisément :

$$\forall A \in \mathcal{B} \quad \forall r > 0 \quad 1 - \nu_1^{\otimes n}(A + 6\sqrt{r}\mathbb{B}_2^n + 9r\mathbb{B}_1^n) \leq \frac{1}{\nu_1^{\otimes n}(A)} e^{-r}$$

*Idée de démonstration.* On montre d'abord qu'on peut se ramener au cas  $p = 1$ . Puis, pour  $p = 1$ , il suffit de montrer la seconde inégalité (plus forte que la première, avec  $K_1 = 1$ ). On cherche alors un coût convexe  $c$  sur  $\mathbf{R}^n$  tel que

- $\nu_1^{\otimes n}$  vérifie une ICC par rapport à  $c$
- $\{c < r\} \subset 6\sqrt{r}\mathbb{B}_2^n + 9r\mathbb{B}_1^n$

Le résultat voulu sera alors conséquence de la propriété 5.1.3. En vertu du théorème 5.2.1, il n'y a qu'à trouver un coût  $c_1$  sur  $(\mathbf{R}, \mathcal{B}, \nu_1)$  tel que  $\tilde{c}(x) = \sum_{i=1}^n \tilde{c}_1(x_i)$  vérifie les hypothèses susmentionnées. On montre que

$$\tilde{c}_1(x) = \begin{cases} \frac{x^2}{18} & \text{si } |x| \leq 2 \\ \frac{2(|x|-1)}{9} & \text{sinon} \end{cases}$$

convient.  $\square$

Une telle inégalité n'est pas une inégalité de concentration au sens strict : elle fait intervenir deux normes. Ainsi, on obtient en fait une concentration des fonctions lipschitziennes pour ces deux normes, de la forme :

$$\forall r \geq 0 \quad \nu_p^{\otimes n}(F \geq m_F + r) \leq Ce \left( -\frac{1}{C} \min \left( \frac{r}{\|F\|_{Lip,1}}, \frac{r^p}{\|F\|_{Lip,p}^p} \right) \right)$$

où  $m_F$  désigne une médiane ou la moyenne de  $F$  et  $C$  une constante.

## Références

- [Led01] M. Ledoux, *The Concentration of Measure Phenomenon*, American Mathematical Society, 2001
- [Led02] M. Ledoux, M. Talagrand *Probability in Banach Spaces*, Springer-Verlag, 2002
- [Gro99] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhäuser, 1999
- [Rob73] A. Wayne Roberts, Dale E. Varberg, *Convex Functions*, Academic Press New York and London, 1973
- [Hal50] P. Halmos, *Measure theory*, D. Van Nostrand Company, Inc. New York, 1950