



Inégalités de Poincaré et de Gross pour les mesures de Bernoulli, de Poisson, et de Gauss

Djalil Chafai

► **To cite this version:**

Djalil Chafai. Inégalités de Poincaré et de Gross pour les mesures de Bernoulli, de Poisson, et de Gauss. 2005.

HAL Id: hal-00012428

<https://hal.archives-ouvertes.fr/hal-00012428v3>

Submitted on 29 Oct 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inégalités de Poincaré et de Gross pour les mesures de Bernoulli, de Poisson, et de Gauss

Djalil CHAFAÏ

28 octobre 2005

Résumé

Les inégalités de Sobolev logarithmiques doivent leur nom à un article célèbre de Gross paru en 1975. Ces inégalités fonctionnelles apparaissent à la fois comme une expression de la propriété d'hypercontractivité de semi-groupes markoviens, comme une traduction de la décroissance exponentielle de l'entropie le long de ces semi-groupes, comme une information de concentration gaussienne de la mesure, et enfin comme une façon de renforcer les inégalités de Poincaré, plus traditionnelles en analyse.

Dans cet article, des inégalités fonctionnelles optimales de ce type sont obtenues pour les mesures de Poisson et de Gauss, par tensorisation infinie, à partir d'inégalités optimales pour des mesures de Bernoulli sur l'espace à deux points. Les comparaisons variance-entropie ainsi que certaines propriétés fondamentales de stabilité des inégalités associées sont également abordées. Des preuves par interpolation au moyen de semi-groupes de Markov complètent l'exposé. Le texte se termine par quelques notes historiques et bibliographiques.

Cet article est une refonte complète du premier chapitre du livre [ABC⁺00], écrit par l'auteur plus de cinq ans auparavant. Il est toujours conçu dans le souci de rester abordable au plus grand nombre, et nécessite assez peu de connaissances préalables en analyse et en probabilités. Les notions introduites possèdent des généralisations, ramifications et applications qui font l'objet de volumineux travaux.

1 Introduction

Soit μ une mesure de probabilité sur un espace mesurable E . On note $L^0(\mu)$ l'espace vectoriel des fonctions de E dans \mathbb{R} qui sont mesurables. Pour tout réel $p \geq 1$, on note $L^p(\mu)$ l'espace de Lebesgue des fonctions $f \in L^0(\mu)$ telles que $|f|^p$ est μ -intégrable. Pour toute fonction $f \in L^1(\mu)$, on note $\mathbf{E}_\mu(f)$ ou encore $\mathbf{E}_\mu f$ la moyenne de f sous μ définie par

$$\mathbf{E}_\mu(f) := \int_E f d\mu.$$

La notation \mathbf{E} est due au fait que $\mathbf{E}_\mu(f)$ est l'espérance de $f : (E, \mu) \rightarrow \mathbb{R}$ vue comme une variable aléatoire réelle. Dans tout ce qui suit, le terme « μ -presque-sûrement

équivalent au terme « μ -presque-partout ». L'ensemble \mathbb{R}^n est implicitement muni de sa tribu borélienne, tandis que les sous-ensembles de \mathbb{N} sont implicitement munis de la tribu discrète constituée de l'ensemble de leurs parties.

Définition 1.1 (Variance). La *variance* d'une fonction $f \in L^2(\mu)$ est définie par

$$\mathbf{Var}_\mu(f) := \mathbf{E}_\mu(f^2) - (\mathbf{E}_\mu f)^2.$$

La quantité $\mathbf{Var}_\mu(f)$ est la distance de f dans $L^2(\mu)$ au sous-espace vectoriel des fonctions constantes, puisque :

$$\mathbf{Var}_\mu(f) = \inf_{a \in \mathbb{R}} \{ \mathbf{E}_\mu((f - a)^2) \}. \quad (1)$$

Cet infimum est atteint pour $a = \mathbf{E}_\mu(f)$ et $\mathbf{Var}_\mu(f) = \mathbf{E}_\mu((f - \mathbf{E}_\mu f)^2)$. La fonctionnelle $f \mapsto \mathbf{Var}_\mu(f)$ est invariante par translation.

On désigne par $L^{1+}(\mu)$ l'ensemble des fonctions $f \in L^0(\mu)$ positives ou nulles telles que $f \log(f) \in L^1(\mu)$, avec la convention $0 \log(0) = 0$. Cet ensemble contient toute fonction constante positive ou nulle. La convexité de la fonction $u \in \mathbb{R}_+ \mapsto u \log(u)$ entraîne que $L^{1+}(\mu)$ est convexe. L'inégalité $u - 1 \leq u \log(u)$ pour $u \geq 0$ montre de plus que $L^{1+}(\mu) \subset L^1(\mu)$.

Définition 1.2 (Entropie). L'*entropie* d'une fonction $f \in L^{1+}(\mu)$ est définie par

$$\mathbf{Ent}_\mu(f) := \mathbf{E}_\mu(f \log(f)) - \mathbf{E}_\mu(f) \log \mathbf{E}_\mu(f).$$

On a $\mathbf{Ent}_\mu(f) = \mathbf{E}_\mu(f \log(f/\mathbf{E}_\mu f))$ dès que $\mathbf{E}_\mu f > 0$. Si $\mathbf{E}_\mu(f) = 0$, alors f est μ -presque-sûrement égale à une constante et $\mathbf{Ent}_\mu(f) = 0$. Enfin, une formule de Taylor montre que l'entropie s'écrit également

$$\mathbf{Ent}_\mu(f) = \inf_{a > 0} \{ \mathbf{E}_\mu((\log(f) - \log(a))f - a + f) \}, \quad (2)$$

et l'infimum est atteint ici encore pour $a = \mathbf{E}_\mu(f)$. Contrairement à la variance, l'entropie n'est pas invariante par translation. La variance et l'entropie sont des fonctionnelles de la forme

$$f \mapsto \mathbf{E}_\mu^\Phi(f) := \mathbf{E}_\mu(\Phi(f)) - \Phi(\mathbf{E}_\mu f), \quad (3)$$

où $\Phi : \mathcal{I} \rightarrow \mathbb{R}$ est une fonction strictement convexe définie sur un intervalle $\mathcal{I} \subset \mathbb{R}$. Pour la variance, $\mathcal{I} = \mathbb{R}$ et $\Phi : u \mapsto u^2$, tandis que pour l'entropie, $\mathcal{I} = \mathbb{R}_+$ et $\Phi : u \mapsto u \log(u)$. L'usage de l'inégalité de Jensen pour la fonction convexe Φ et la mesure de probabilité μ montre que $\mathbf{E}_\mu^\Phi(f)$ n'est jamais négative. De plus, la convexité stricte de Φ entraîne que $\mathbf{E}_\mu^\Phi(f)$ est nulle si et seulement si f est μ -presque-sûrement égale à une constante.

La fonctionnelle définie en (3) n'est pas sensible aux perturbations affines additives de Φ . En particulier, pour l'entropie, il est parfois commode de considérer la fonction $u \in \mathbb{R}_+ \mapsto u \log(u) - u + 1$ en lieu et place de $u \in \mathbb{R}_+ \mapsto u \log(u)$. Enfin, on peut observer que les formules (1) et (2) sont unifiées par la formule suivante qui découle d'une formule de Taylor pour la fonction convexe Φ :

$$\mathbf{E}_\mu^\Phi(f) = \sup_{a \in \text{Int}(\mathcal{I})} \underbrace{\mathbf{E}_\mu(\Phi(f) - \Phi(a) - \Phi'(a)(f - a))}_{\geq 0}.$$

La variance est homogène d'ordre 2 tandis que l'entropie est homogène d'ordre 1. Cependant, on a coutume de considérer l'entropie du carré $\mathbf{Ent}_\mu(f^2)$, qui est homogène d'ordre 2, et qui n'impose plus à f de prendre des valeurs positives ou nulles. La comparaison de la variance ou de l'entropie du carré à une forme quadratique conduit aux notions d'inégalités de Poincaré et de Sobolev logarithmique.

Définition 1.3 (Énergie). Soit μ une mesure de probabilité sur un espace mesurable E , et $Q : \mathcal{D} \subset L^0(\mu) \mapsto L^0(\mu)$ une fonctionnelle vérifiant

- \mathcal{D} est un sous espace de $L^0(\mu)$ contenant les fonctions constantes μ -p.-s.;
- $\mathbf{E}_\mu(Q(f))$ a un sens dans $\mathbb{R}_+ \cup \{+\infty\}$ pour tout $f \in \mathcal{D}$;
- Q est une forme quadratique;
- Q s'annule pour les fonctions μ -presque-sûrement constantes, c'est-à-dire que si B est une forme bilinéaire telle que $Q(f) = B(f, f)$ pour tout $f \in \mathcal{D}$, alors $B(f, g) = 0$ dès que f ou g est égale μ -presque-sûrement à une constante. En particulier, $Q(1) = 0$ et $Q(1 + f) = Q(f)$.

Alors on appelle *énergie*, ou encore *forme que Dirichlet*, la fonctionnelle

$$\mathcal{D} \ni f \mapsto \mathbf{E}_\mu(Q(f)) \in \mathbb{R}_+ \cup \{+\infty\}.$$

Lorsque E est inclus dans \mathbb{N} , on considère la forme quadratique $Q(f) = (Df)^2$ où $(Df)(n) := f(n+1) - f(n)$ pour tout $n \in E$, tandis que sur l'espace euclidien \mathbb{R}^n , on considère la forme quadratique $Q(f) = |\nabla f|^2$. Enfin, lorsque μ est la mesure invariante d'un processus de Markov de générateur \mathbf{L} , on a coutume de considérer la forme quadratique $Q(f) = -f\mathbf{L}f$. Dans toute la suite, on note $|x|$ la norme euclidienne d'un vecteur $x \in \mathbb{R}^n$, et $x \cdot y$ le produit scalaire des vecteurs x et y de \mathbb{R}^n . Ainsi, $|\nabla f|$ désigne la norme euclidienne du gradient ∇f de la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Définition 1.4 (Inégalité de Poincaré). Soit μ une mesure de probabilité sur un espace mesurable E . Soit $f \in \mathcal{D} \mapsto \mathbf{E}_\mu(Q(f))$ une énergie. Soit $\mathcal{F} \subset \mathcal{D} \cap L^2(\mu)$. On dit que μ vérifie une *inégalité de Poincaré* de constante $c > 0$ sur la classe \mathcal{F} lorsque

$$\forall f \in \mathcal{F}, \quad \mathbf{Var}_\mu(f) \leq c \mathbf{E}_\mu(Q(f)). \quad (4)$$

Définition 1.5 (Inégalité de Sobolev logarithmique). Soit μ une mesure de probabilité sur un espace mesurable E . Soit $f \in \mathcal{D} \mapsto \mathbf{E}_\mu(Q(f))$ une énergie. Soit $\mathcal{F} \subset \mathcal{D}$ telle que $f^2 \in L^{1+}(\mu)$ pour tout $f \in \mathcal{F}$. On dit que μ vérifie une *inégalité de Sobolev logarithmique*¹ de constante $c > 0$ sur la classe \mathcal{F} lorsque

$$\forall f \in \mathcal{F}, \quad \mathbf{Ent}_\mu(f^2) \leq c \mathbf{E}_\mu(Q(f)). \quad (5)$$

Les propriétés de l'énergie assurent que le membre de droite de (4) et de (5) n'est jamais négatif, et que son annulation entraîne l'annulation du membre de gauche. Comme Q est une forme quadratique, les deux membres de ces inégalités sont homogènes d'ordre 2. Lorsque la classe \mathcal{F} et l'énergie sont fixées, chacune des deux inégalités est d'autant plus forte que la constante c est petite. La constante « optimale » est donc la plus petite constante possible, à classe et énergie fixées. En utilisant la convention $0^{-1} = +\infty$, cela conduit à la définition suivante.

¹Dans la littérature, on parle également d'*inégalité de Gross*.

Définition 1.6 (Constantes optimales). La constante optimale de Poincaré est définie par

$$c_P(\mu) = \left(\inf \left\{ \frac{\mathbf{E}_\mu(Q(f))}{\mathbf{Var}_\mu(f)}; f \in \mathcal{F}; f \text{ non constante } \mu\text{-p.s.} \right\} \right)^{-1}.$$

De même, la constante optimale de Sobolev logarithmique est définie par

$$c_{SL}(\mu) = \left(\inf \left\{ \frac{\mathbf{E}_\mu(Q(f))}{\mathbf{Ent}_\mu(f^2)}; f \in \mathcal{F}; f \text{ non constante } \mu\text{-p.s.} \right\} \right)^{-1}.$$

Une constante optimale infinie signifie qu'il n'y a pas d'inégalité du type considéré. Le lecteur familier avec les plongements de Sobolev remarquera à juste titre que de telles inégalités peuvent être comprises comme une formulation de la continuité de l'injection entre deux espaces fonctionnels, et la constante optimale apparaît alors comme une norme d'opérateur.

Sur tout espace mesurable E , et pour tout $a \in E$, on note δ_a la mesure de Dirac en a , qui affecte la masse 1 aux ensembles mesurables contenant a et, et la masse 0 aux autres. C'est une mesure de probabilité. On note également $*$ le produit de convolution, et \otimes le produit tensoriel. Par la suite, nous considérerons les mesures de probabilité classiques suivantes.

Mesure de Bernoulli sur $E = \{0, 1\}$. La *mesure de Bernoulli* $\mathcal{B}(1, p)$ de paramètre $p \in [0, 1]$ est définie par

$$\mathcal{B}(1, p) := q\delta_0 + p\delta_1,$$

où $q := 1 - p$. Elle a pour support l'espace à deux points $\{0, 1\}$. Elle est qualifiée de *symétrique* lorsque $p = q = 1/2$. On considère dans ce cadre la classe \mathcal{F} de toutes les fonctions de $\{0, 1\}$ dans \mathbb{R} . La forme quadratique $Q : \mathcal{F} \rightarrow \mathcal{F}$ considérée est définie pour toute fonction $f \in \mathcal{F}$ par $Q(f) = (Df)^2$, où

$$(Df)(0) = -(Df)(1) = f(1) - f(0).$$

En identifiant l'espace à deux points $\{0, 1\}$ au cercle $\mathbb{Z}/2\mathbb{Z}$ pour lequel $1 + 1 = 0$, on constate que $(Df)(n) = f(n + 1) - f(n)$ pour tout $n \in \{0, 1\}$, et D apparaît comme l'opérateur aux différences classique.

Mesure binomiale sur $E = \{0, 1, \dots, n\}$. La *mesure binomiale* de taille $n \in \mathbb{N}^*$ et de paramètre $p \in [0, 1]$ est définie par

$$\mathcal{B}(n, p) := \mathcal{B}(1, p)^{*n} = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta_k,$$

où $\binom{n}{k} := k!/(n!(n-k)!)$ est le coefficient binomial classique. Son support est l'ensemble fini $\{0, 1, \dots, n\}$. La mesure de Bernoulli $\mathcal{B}(1, p)$ en est un cas particulier, ce qui explique sa notation. Dans ce cadre, la classe \mathcal{F} choisie est celle de toutes les fonctions de $\{0, 1, \dots, n\}$ dans \mathbb{R} . La forme quadratique $Q : \mathcal{F} \rightarrow \mathcal{F}$ considérée est donnée par $Q(f) = (Df)^2$, où Df est défini par $(Df)(k) = f(k+1) - f(k)$ pour tout $k \in \{0, 1, \dots, n\}$.

Ici encore, on a identifié l'ensemble fini $\{0, 1, \dots, n\}$ au cercle $\mathbb{Z}/(n+1)\mathbb{Z}$, de sorte que $n+1 = 0$. Comme $\mathcal{B}(n, p) = \mathcal{B}(1, p)^{*n}$, la mesure $\mathcal{B}(n, p)$ est la mesure image de la mesure produit $\mathcal{B}(1, p)^{\otimes n}$ par l'application qui somme les coordonnées de $\{0, 1\}^n$.

Mesure de Poisson sur $E = \mathbb{N}$. La *mesure de Poisson* de paramètre $\lambda \in \mathbb{R}_+$ est définie par

$$\mathcal{P}(\lambda) := e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \delta_k.$$

Elle a pour support l'ensemble \mathbb{N} des entiers naturels. Dans ce cadre, la classe \mathcal{F} choisie est $L^{1+}(\mathcal{P}(\lambda))$. La forme quadratique $Q : \mathcal{F} \rightarrow \mathcal{F}$ choisie est ici encore définie par $Q(f) = (Df)^2$ où $(Df)(n) = f(n+1) - f(n)$ pour tout $n \in \mathbb{N}$. Notons qu'a priori, $Q(f)$ peut valoir $+\infty$ pour certains $f \in L^{1+}(\mathcal{P}(\lambda))$. La mesure de Poisson se trouve à la « frontière » de l'ensemble des mesures binomiales puisque

$$\lim_{n \rightarrow \infty} B(n, p_n) = \mathcal{P}(\lambda) \quad (6)$$

dès que $\lim_{n \rightarrow \infty} np_n = \lambda$. Cette propriété correspond à une convergence de coefficients binomiaux : $\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \lambda^k / k!$.

Mesure de Gauss sur $E = \mathbb{R}$. La *mesure de Gauss* $\mathcal{N}(m, v)$ de moyenne $m \in \mathbb{R}$ et de variance $v \in \mathbb{R}_+^*$ est la mesure absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} dont la densité de Radon-Nikodym est donnée par

$$x \in \mathbb{R} \mapsto (2\pi v)^{-\frac{1}{2}} e^{-\frac{(x-m)^2}{2v}}.$$

Elle est qualifiée de *mesure de Gauss standard* lorsque $(m, \sigma) = (0, 1)$. Son support est \mathbb{R} . Dans ce cadre, la classe \mathcal{F} choisie est l'espace de Sobolev² $H^1(\mathcal{N}(m, v))$. La forme quadratique utilisée $Q : \mathcal{F} \rightarrow L^1(\mathcal{N}(m, v))$ est donnée par $Q(f) = f'^2$. Le «théorème central limite» fait apparaître la mesure de Gauss comme limite de puissances de convolution, recentrées, et renormalisées :

Théorème 1.7 (Théorème central limite dans \mathbb{R}). Soit μ une mesure de probabilité sur \mathbb{R} telle que $\xi \in L^2(\mu)$, où $\xi : \mathbb{R} \rightarrow \mathbb{R}$ est l'application identité définie par $\xi(x) = x$ pour tout $x \in \mathbb{R}$. Alors pour toute fonction continue et bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbf{E}_{\mu^{\otimes n}}(f \circ \kappa_n \circ s_n) = \lim_{n \rightarrow \infty} \mathbf{E}_{\mu^{*n}}(f \circ \kappa_n) = \mathbf{E}_{\mathcal{N}(0,1)}(f), \quad (7)$$

où $\kappa_n : \mathbb{R} \rightarrow \mathbb{R}$ et $s_n : \mathbb{R}^n \rightarrow \mathbb{R}$ sont définies par

$$\kappa_n(s) := \frac{s - n\mathbf{E}_{\mu}(\xi)}{\sqrt{n\mathbf{Var}_{\mu}(\xi)}} \quad \text{et} \quad s_n(x) := x_1 + \dots + x_n.$$

² L'espace de Sobolev $H^1(\mu)$, noté également $W^{1,2}(\mu)$, est constitué des fonctions de $L^2(\mu)$ dont la dérivée au sens des distributions est dans $L^2(\mu)$. Le lecteur peu familier avec les espaces de Sobolev ou avec la dérivation faible pourra plutôt considérer la classe des fonctions \mathcal{C}^1 dont les dérivées d'ordre 0 et 1 sont de carré intégrable pour $\mu = \mathcal{N}(m, v)$.

Ce théorème peut être établi en utilisant la transformée de Fourier des mesures de probabilité. Lorsque $\mu = \mathcal{B}(1, p)$, on a $\mathbf{E}_\mu(\xi) = p$ et $\mathbf{Var}_\mu(\xi) = pq$. Il est utile de remarquer que la mesure de Gauss $\mathcal{N}(0, 1)$ est un point fixe de la transformation qui associe à μ la mesure image de $\mu^{\otimes n}$ par l'application $\kappa_n \circ s_n$.

Mesure de Gauss sur \mathbb{R}^n . Soit Σ une matrice $n \times n$ symétrique semi-définie positive. Il est toujours possible d'écrire Σ sous la forme AA^\top où $A \in \mathcal{M}_n(\mathbb{R})$. La mesure image par l'application affine $x \mapsto Ax + m$ de la mesure de probabilité $\mathcal{N}(0, 1)^{\otimes n}$ sur \mathbb{R}^n ne dépend que de m et Σ . Elle est appelée *mesure de Gauss $\mathcal{N}(m, \Sigma)$* , de moyenne $m \in \mathbb{R}^n$ et de matrice de covariance Σ . Elle est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^n si et seulement si Σ est inversible, et sa densité de Radon-Nikodym est alors donnée par

$$x \in \mathbb{R}^n \mapsto (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(x - m)^\top \Sigma^{-1}(x - m)\right).$$

En particulier, lorsque $m = 0$ et $\Sigma = I_n$, on retrouve bien la densité $x \in \mathbb{R}^n \mapsto (2\pi)^{-n/2} \exp(-\frac{1}{2}|x|^2)$ de la mesure de Gauss produit $\mathcal{N}(0, 1)^{\otimes n}$. Dans ce cadre, la classe \mathcal{F} choisie est l'espace de Sobolev³ $H^1(\mathcal{N}(m, \Sigma))$. La forme quadratique choisie $Q : \mathcal{F} \rightarrow L^1(\mathcal{N}(m, \Sigma))$ est donnée par $Q(f) = |\nabla f|^2$.

2 Formules variationnelles

La variance et l'entropie possèdent des formulations variationnelles qui s'avèrent très utiles pour démontrer des inégalités de Poincaré et de Sobolev logarithmiques.

Théorème 2.1 (Formules variationnelles). *Soit μ une mesure de probabilité sur un espace mesurable E . Pour toute fonction $f \in L^2(\mu)$,*

$$\mathbf{Var}_\mu(f) = \sup \{2\mathbf{E}_\mu(fg) - \mathbf{E}_\mu(g^2) ; g \in L^2(\mu), \mathbf{E}_\mu(g) = 0\}, \quad (8)$$

et le supremum est atteint pour $g = f - \mathbf{E}_\mu(f)$. De même, pour toute fonction $f \in L^1+(\mu)$,

$$\mathbf{Ent}_\mu(f) = \sup \{\mathbf{E}_\mu(fg) ; e^g \in L^1(\mu), \mathbf{E}_\mu(e^g) = 1\}, \quad (9)$$

et le supremum est atteint pour $g = \log(f) - \log(\mathbf{E}_\mu f)$.

Démonstration. Pour établir (8), on se ramène au cas où $\mathbf{E}_\mu(f) = 0$. On a alors $\mathbf{E}_\mu(f^2) \geq 2\mathbf{E}_\mu(fg) - \mathbf{E}_\mu(g^2)$, puisque $\mathbf{E}_\mu((f - g)^2) \geq 0$. Pour établir (9), on se ramène au cas où $\mathbf{E}_\mu(f) = 1$. Ensuite, comme $-\mathbf{E}_\mu(f) + \mathbf{E}_\mu(e^g) = 0$, on obtient $\mathbf{E}_\mu(fg) \leq \mathbf{E}_\mu(f \log(f))$ en utilisant l'inégalité $uv \leq u \log(u) - u + e^v$, valable pour tout $(u, v) \in \mathbb{R}_+ \times \mathbb{R}$. \square

La formule variationnelle de l'entropie est équivalente à l'inégalité suivante, connue sous le nom d'*inégalité entropique* : pour toute fonction f positive ou nulle de carré intégrable, g de carré intégrable et $t > 0$:

$$\mathbf{E}_\mu(fg) \leq \frac{\mathbf{E}_\mu(f)}{t} \log \mathbf{E}_\mu(e^{tg}) + \frac{\mathbf{Ent}_\mu(f)}{t}.$$

³Cf. note de bas de page n°2, en remplaçant $\mathcal{N}(m, v)$ par $\mathcal{N}(m, \Sigma)$.

En mécanique statistique, cette reformulation s'avère très utile pour contrôler le terme de covariance $\mathbf{E}_\mu(fg)$ à partir d'un contrôle de $\mathbf{Ent}_\mu(f)$ et de la transformée de Laplace $\mathbf{E}_\mu(e^{tg})$.

2.1 Convexité de la variance et de l'entropie

Les formules variationnelles (8) et (9) sont de même nature. En effet, (8) se réécrit sous la forme⁴

$$\mathbf{Var}_\mu(f) = \sup_{g \in L^2(\mu)} \{ \mathbf{Var}_\mu(g) - 2\mathbf{E}_\mu((g - \mathbf{E}_\mu g)(f - g)) \},$$

et le supremum est atteint pour $g = f$. De la même manière, (9) se réécrit

$$\mathbf{Ent}_\mu(f) = \sup_{g \in L^1+(\mu)} \{ \mathbf{Ent}_\mu(g) - \mathbf{E}_\mu((\log g - \log(\mathbf{E}_\mu g))(f - g)) \},$$

et le supremum est atteint pour $g = f$. Montrons à présent, de manière formelle, comment ces deux écritures variationnelles sont liées à une propriété de convexité supplémentaire de la fonction convexe $\Phi : \mathcal{I} \subset \mathbb{R} \rightarrow \mathbb{R}$ associée à la variance ou à l'entropie. La différence $\mathbf{E}_\mu^\Phi(f) - \mathbf{E}_\mu^\Phi(g)$ s'écrit

$$\mathbf{E}_\mu(A(g, f - g)) - A(\mathbf{E}_\mu g, \mathbf{E}_\mu(f - g)) + \mathbf{E}_\mu((\Phi'(g) - \Phi'(\mathbf{E}_\mu g))(f - g)),$$

où A est définie sur l'ensemble convexe $\{(u, v) \in \mathbb{R}^2; (u, u + v) \in \mathcal{I} \times \mathcal{I}\}$ par

$$A(u, v) := \Phi(u + v) - \Phi(u) - \Phi'(u)v. \quad (10)$$

Notons que la quantité $\mathbf{E}_\mu(\Phi'(g)(f - g))$ a bien un sens dans $\mathbb{R} \cup \{-\infty\}$ car elle est majorée par la quantité $\Phi(f) - \Phi(g)$ en vertu de la convexité de Φ . Dans l'exemple de la variance et de l'entropie, il se trouve que A est convexe sur son ensemble de définition, et l'inégalité de Jensen assure alors que

$$\mathbf{E}_\mu(A(g, f - g)) - A(\mathbf{E}_\mu g, \mathbf{E}_\mu(f - g)) \geq 0.$$

Cela donne finalement, en notant $L^{1,\Phi} := \{f \in L^0(\mu); \Phi(f) \in L^1(\mu)\}$,

$$\mathbf{E}_\mu^\Phi(f) = \sup_{g \in L^{1,\Phi}(\mu)} \{ \mathbf{E}_\mu^\Phi(g) + \mathbf{E}_\mu((\Phi'(g) - \Phi'(\mathbf{E}_\mu g))(f - g)) \}, \quad (11)$$

et le supremum est atteint pour $g = f$. C'est exactement la formule recherchée. La formule variationnelle (11) ci-dessus permet de concevoir la variance et l'entropie comme enveloppe de fonctionnelles affines tangentes. Cela correspond au fait que la variance et l'entropie sont elles même des fonctionnelles convexes, comme l'exprime le théorème suivant.

⁴On peut observer que l'argument du supremum s'écrit également $2\mathbf{Cov}_\mu(f, g) - \mathbf{Var}_\mu(g)$.

Théorème 2.2 (Convexité de la variance et de l'entropie). *Soit μ une mesure de probabilité sur un espace mesurable E . Les fonctionnelles $f \in L^2(\mu) \mapsto \mathbf{Var}_\mu(f) \in \mathbb{R}$ et $f \in L^{1+}(\mu) \mapsto \mathbf{Ent}_\mu(f) \in \mathbb{R}$ sont convexes.*

Démonstration. Soit $\Phi : \mathcal{I} \subset \mathbb{R} \rightarrow \mathbb{R}$ la fonction convexe associée à la variance ou à l'entropie. Pour tout $t \in [0, 1]$ et f et g raisonnables, on pose $h_t := tf + (1 - t)g$ et $\alpha(t) := \mathbf{E}_\mu^\Phi(h_t)$. La fonction α est continue sur $[0, 1]$ et lisse sur $]0, 1[$. Elle vérifie $\alpha(0) = \mathbf{E}_\mu^\Phi(g)$ et $\alpha(1) = \mathbf{E}_\mu^\Phi(f)$. De plus,

$$\alpha''(t) = \mathbf{E}_\mu(C(h_t, f - g)) - C(\mathbf{E}_\mu h_t, \mathbf{E}_\mu(f - g)),$$

où C est la fonction définie sur l'ensemble convexe $\{(u, v); (u, u + v) \in \mathcal{I} \times \mathcal{I}\}$ par

$$C(u, v) := \Phi''(u)v^2. \quad (12)$$

Or il se trouve que C est convexe sur son ensemble de définition. L'inégalité de Jensen entraîne que $\alpha''(t) \geq 0$, et α est donc convexe sur $[0, 1]$. En particulier, l'inégalité $\alpha(\lambda) \leq \lambda\alpha(1) + (1 - \lambda)\alpha(0)$ pour tout $\lambda \in [0, 1]$ se réécrit

$$\mathbf{E}_\mu^\Phi(\lambda f + (1 - \lambda)g) \leq \lambda\mathbf{E}_\mu^\Phi(f) + (1 - \lambda)\mathbf{E}_\mu^\Phi(g).$$

La convexité de la fonctionnelle $f \mapsto \mathbf{E}_\mu^\Phi(f)$ est donc établie. De plus, l'écriture $\alpha(1) = \sup_{t \in [0, 1]} \{\alpha(t) + \alpha'(t)(1 - t)\}$ correspond exactement à (11). \square

3 Comparaisons variance-entropie

Soit μ une mesure de probabilité sur un espace mesurable E . L'inégalité élémentaire $u \log(u) + u - 1 \leq u^2 - 1$ pour tout $u \geq 0$ entraîne que pour toute fonction $f \in L^2(\mu)$ positive ou nulle,

$$\mathbf{E}_\mu(f) \mathbf{Ent}_\mu(f) \leq \mathbf{Var}_\mu(f), \quad (13)$$

avec égalité si et seulement si f est égale à une constante μ -presque-sûrement. Comme $\mathbf{Var}_\mu(|f|) \leq \mathbf{Var}_\mu(f)$ pour tout $f \in L^2(\mu)$, cela montre en particulier que

$$L^2(\mu) \subset \{f \in L^0(\mu); |f| \in L^{1+}(\mu)\}.$$

D'autre part, l'inégalité élémentaire $(\sqrt{u} - 1)^2 \leq u \log(u) + 1 - u$ pour tout $u \geq 0$ entraîne que pour toute fonction f positive ou nulle telle que $f^2 \in L^{1+}(\mu)$,

$$\mathbf{Var}_\mu(f) \leq \mathbf{Ent}_\mu(f^2), \quad (14)$$

avec égalité si et seulement si f est égale à une constante μ -presque-sûrement. L'inégalité (14) est fautive en toute généralité pour f de signe quelconque comme on peut le vérifier sur l'exemple de la mesure de Bernoulli $\mathcal{B}(1, 1/2)$ et de la fonction valant $-a$ en 0 et a en 1. En effet, dans ce cas, la fonction est de carré constant et l'entropie du carré est donc

nulle tandis que la variance vaut a^2 . Ainsi, il n'existe pas de constante c qui permette une comparaison générique de la forme

$$\mathbf{Var}_\mu(f) \leq c \mathbf{Ent}_\mu(f^2).$$

On ne peut donc pas, malheureusement, utiliser (14) pour montrer que l'inégalité de Poincaré découle de l'inégalité de Sobolev logarithmique. Cela est cependant possible par linéarisation.

Théorème 3.1. *Soit μ une mesure de probabilité sur un espace mesurable E . Supposons que la classe \mathcal{F} est constituée de fonctions bornées. Alors, si μ vérifie une inégalité de Sobolev logarithmique de constante c , alors elle vérifie une inégalité de Poincaré de constante $c/2$. En d'autres termes, on a dans $\mathbb{R}_+ \cup \{+\infty\}$,*

$$c_P(\mu) \leq \frac{1}{2} c_{SL}(\mu). \quad (15)$$

En général, la classe de fonctions concernées peut ensuite être considérablement étendue par des techniques d'approximation, par exemple par régularisation et troncature. D'autre part, le lemme de Fatou est souvent utilisé pour se ramener au cas où $f^2 > 0$ en remplaçant f par $f + \varepsilon$ après le changement de fonction de f^2 vers f dans l'inégalité de Sobolev logarithmique.

Démonstration. Un développement limité de $u \mapsto (1+u) \log(1+u)$ en $u = 0$ donne pour toute fonction $f \in \mathcal{F}$:

$$\mathbf{Ent}_\mu((1 + \varepsilon f)^2) = 2\varepsilon^2 \mathbf{Var}_\mu(f) + O(\varepsilon^3).$$

D'autre part, comme Q est quadratique et s'annule sur les constantes, on a :

$$\mathbf{E}_\mu(Q(1 + \varepsilon f)) = \varepsilon^2 \mathbf{E}_\mu(Q(f)).$$

Il ne reste donc plus qu'à faire tendre ε vers 0^+ pour obtenir le résultat souhaité. \square

La mesure de Gauss abordée plus loin fournit un exemple de mesures de probabilité pour laquelle l'inégalité (15) entre constantes optimales est une égalité. Cela montre que cette comparaison des constantes optimales est elle-même optimale en toute généralité. Certaines mesures de probabilité satisfont à une inégalité de Poincaré mais pas à une inégalité de Sobolev logarithmique. C'est par exemple le cas de la *mesure exponentielle*, de densité $x \mapsto e^{-x}$ par rapport à la mesure de Lebesgue sur \mathbb{R}_+ , associé à la forme quadratique $Q(f) := f'^2$.

Avant de clore cette section, signalons l'inégalité de Rothaus, valable pour toute mesure de probabilité μ et toute fonction f bornée :

$$\mathbf{Ent}_\mu(f^2) \leq 2 \mathbf{Var}_\mu(f) + \mathbf{Ent}_\mu((f - \mathbf{E}_\mu f)^2).$$

Elle peut être obtenue en considérant la dérivée à droite en $q = 2$ dans l'inégalité

$$\|f\|_q^2 \leq \|f\|_2 + (q-1) \|f - \mathbf{E}_\mu f\|_q^2,$$

valable pour tout $q > 2$. Ici, $\|f\|_p := \mathbf{E}_\mu(|f|^p)^{1/p}$ désigne la norme de f dans $L^p(\mu)$. Enfin, il est utile de retenir les deux identités suivantes, valables pour f positive :

$$\mathbf{Ent}_\mu(f) = \partial_{p=1+} \|f\|_p \quad \text{et} \quad \mathbf{Ent}_\mu(f) = \partial_{p=1+} \left[\|f\|_p^p - \|f\|_1^p \right].$$

Il s'agit du point de départ du théorème d'hypercontractivité de Gross.

4 Espace à deux points et mesures de Bernoulli

Cette section concerne exclusivement la mesure de Bernoulli $\mathcal{B}(1, p)$ sur l'espace à deux points $\{0, 1\}$, identifié au cercle $\mathbb{Z}/2\mathbb{Z}$. Dans toute cette partie, $p \in]0, 1[$. Pour toute fonction $f : \{0, 1\} \rightarrow \mathbb{R}$, la fonction $(Df)^2 : \{0, 1\} \rightarrow \mathbb{R}$ est constante et égale à $(f(1) - f(0))^2$. Or $\mathbf{Var}_{\mathcal{B}(1, p)}(f) = pq(f(1) - f(0))^2$, de sorte que $\mathcal{B}(1, p)$ vérifie une (in)égalité de Poincaré de constante optimale pq .

Théorème 4.1. *Pour toute fonction $f : \{0, 1\} \rightarrow \mathbb{R}$, on a*

$$\mathbf{Var}_{\mathcal{B}(1, p)}(f) = pq \mathbf{E}_{\mathcal{B}(1, p)}((Df)^2), \quad (16)$$

et en particulier, $c_P(\mathcal{B}(1, p)) = pq$.

L'inégalité de Sobolev logarithmique est plus subtile. Cependant, le cas symétrique – le plus important en pratique – reste très simple à démontrer.

Théorème 4.2. *On a $c_{\text{SL}}(\mathcal{B}(1, p)) = pq(\log(q) - \log(p))/(q - p)$, avec le prolongement par continuité valant $1/2$ lorsque $p = q = 1/2$. En d'autres termes, pour toute fonction $f : \{0, 1\} \rightarrow \mathbb{R}$, on a*

$$\mathbf{Ent}_{\mathcal{B}(1, p)}(f^2) \leq \frac{\log(q) - \log(p)}{q - p} pq \mathbf{E}_{\mathcal{B}(1, p)}((Df)^2), \quad (17)$$

et la constante est optimale.

Vues comme des fonctions de p , les constantes optimales c_P et c_{SL} sont concaves, symétriques d'axe $p = 1/2$, et atteignent leur maximum $1/2$ pour $p = 1/2$. Elles tendent vers 0 lorsque p tend vers 0^+ ou 1^- . Vu comme une fonction de p , le rapport c_{SL}/c_L est convexe, symétrique d'axe $p = 1/2$, et atteint son minimum 2 pour $p = 1/2$. Il tend vers $+\infty$ quand p tend vers 0^+ ou 1^- .

Démonstration. Comme $|D|f|| \leq |Df|$, il suffit d'établir (17) pour f positive ou nulle. Donnons une preuve rapide de (17) lorsque $p = q = 1/2$. En posant, $(a, b) := (f(0)^2, f(1)^2)$, l'inégalité (17) pour f positive ou nulle s'écrit :

$$\forall a \geq 0, \forall b \geq 0, (a \log a + b \log b) - (a + b) \log \left(\frac{a + b}{2} \right) \leq (\sqrt{a} - \sqrt{b})^2.$$

On peut supposer par homogénéité que $b = 1$ et poser $a = 1 + x$ avec $x \geq -1$, ce qui donne l'inégalité élémentaire suivante, qui a bien lieu :

$$\forall x \geq -1, (1 + x) \log(1 + x) - (2 + x) \log \left(\frac{2 + x}{2} \right) \leq (\sqrt{1 + x} - 1)^2.$$

On peut remarquer que $(1+x)\log(1+x) - (2+x)\log(\frac{2+x}{2}) = h(x) - 2h(x/2)$ où $h(x) := (1+x)\log(1+x) - x$ pour tout $x \geq -1$. Nous avons déjà utilisé l'inégalité $h(x) \geq (\sqrt{1+x} - 1)^2$ lors de la comparaison variance-entropie.

La preuve de (17) quand $p \neq q$ est également élémentaire mais beaucoup plus laborieuse. Elle consiste à calculer directement la constante optimale à partir de sa définition comme supremum. La formulation variationnelle (9) de l'entropie conduit à la formule $c_{\text{SL}}(\mu) = \sup \{c(g), \mathbf{E}_\mu(e^g) = 1\}$ où

$$c(g) := \sup \left\{ \frac{\mathbf{E}_\mu(f^2 g)}{\mathbf{E}_\mu(Q(f))}; f \in \mathcal{F}, \mathbf{E}_\mu(Q(f)) > 0 \right\}.$$

Soit à présent $\mu = \mathcal{B}(1, p)$ et $Q(f) = (Df)^2$. Par symétrie, on peut se restreindre au cas $0 < p \leq 1/2$. Soit $g : \{0, 1\} \rightarrow \mathbb{R}$ telle que $\mathbf{E}_\mu(e^g) = 1$. Il est clair que l'on peut supposer g non nulle. En posant $a = g(0)$ et $b = g(1)$, on voit que l'on doit avoir $pe^a + qb = 1 = e^0$, ce qui entraîne $ab < 0$.

Rappelons que l'on peut supposer que f est positive ou nulle car $|D|f|| \leq |Df|$. D'autre part, on voit en remplaçant f par $f + \varepsilon$ qu'on peut de plus supposer que f est strictement positive. Par homogénéité, on se ramène alors à $f(1) = 1$. Posons $x = f(0)$ avec $x > 0$. La fonction f n'étant pas constante, on a $x \neq 1$. On peut donc écrire :

$$pq c(g) = \sup \left\{ \frac{pax^2 + qb}{(x-1)^2}, x > 0 \text{ et } x \neq 1 \right\}.$$

Ce supremum est atteint en $x = -qb/pa$ et vaut $c(g) = (p/b + q/a)^{-1}$. Ainsi,

$$c_{\text{SL}}(\mathcal{B}(1, p)) = \left(\inf \left\{ \frac{p}{b} + \frac{q}{a}, pe^a + qb = 1 \right\} \right)^{-1}.$$

Posons alors $t = e^a$, $s = e^b = (1-pt)/q$ et $\varphi(t) = p/\log(s) + q/\log(t)$. Comme $ab < 0$, on a

$$c_{\text{SL}}(\mathcal{B}(1, p)) = \left(\inf \{ \varphi(t), t \in]0, 1[\cup]1, p^{-1}[\} \right)^{-1}.$$

Un développement limité en $t = 1$ à l'ordre 2 fournit :

$$\varphi(t) = \frac{1}{2} + \frac{p-q}{12q}(t-1) + \frac{1-3p+3p^2}{24q^2}(t-1)^2 + O((t-1)^3).$$

Nous pouvons étendre φ par continuité sur $[0, p^{-1}]$ en posant $\varphi(0) = -p/\log(q)$, $\varphi(1) = 1/2$ et $\varphi(p^{-1}) = -q/\log(p)$. On remarque que 1 est minimum local si et seulement si $p = 1/2$. Calculons la dérivée de φ en dehors de 1 :

$$\varphi'(t) = \frac{p^2}{qs(\log(s))^2} - \frac{q}{t(\log(t))^2}.$$

On voit alors que cette dérivée tend vers $+\infty$ en $1/p$ et $-\infty$ en 0, donc 0 et $1/p$ ne sont pas des minima locaux. Nous pouvons donc nous intéresser aux zéros de φ' . Comme $\log(t)\log(s) = ab < 0$, $\varphi'(t) = 0$ si et seulement si le système

$$\begin{cases} p\sqrt{t}\log(t) + q\sqrt{s}\log(s) = 0 \\ p(1-t) + q(1-s) = 0 \end{cases}$$

admet une solution non nulle. Ceci entraîne $s = t = 1$, ou bien

$$\frac{\sqrt{t} \log(t)}{1-t} = \frac{\sqrt{s} \log(s)}{1-s}.$$

La fonction $u(x) = \sqrt{x} \log(x)/(1-x)$ est nulle en 0 et $+\infty$. Elle vaut -1 en 1 et vérifie $u(x) = u(x^{-1})$. Elle décroît sur $]0, 1[$ et croît sur $]1, +\infty[$. Il en résulte que les seules valeurs de t pour lesquelles $u(t) = u((1-pt)/q)$ sont 1 et q/p .

Si $p \neq 1/2$, $\varphi'(1) \neq 0$ et φ' s'annule seulement en q/p . La valeur du minimum de φ est alors $(q-p)/(\log(q) - \log(p))$. Si $p = 1/2$, alors $\varphi'(1) = 0$ et φ' s'annule seulement en 1. Le minimum de φ vaut alors $1/2$. \square

Les théorèmes 4.1 et 4.2 impliquent que pour toute fonction $f : \{0, 1\} \rightarrow \mathbb{R}$,

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f^2) \leq \frac{\log(q) - \log(p)}{q-p} \mathbf{Var}_{\mathcal{B}(1,p)}(f).$$

De plus, la constante $(\log(q) - \log(p))/(q-p)$ est optimale. Vue comme une fonction de p , cette constante est convexe, symétrique d'axe $p = 1/2$, et atteint son minimum 2 pour $p = 1/2$. Elle tend vers $+\infty$ lorsque p tend vers 0^+ ou 1^- , ce qui montre qu'il n'existe pas de constante c qui permette une comparaison générique de la forme

$$\mathbf{Ent}_{\mu}(f^2) \leq c \mathbf{Var}_{\mu}(f).$$

Cela complète utilement la discussion sur les comparaisons variance-entropie.

4.1 Inégalités avec énergie modifiée

La majoration $|Df| \leq |Df|$ permet de réduire l'inégalité (17) au cas où f est positive ou nulle. Ainsi, cela revient à dire que pour tout $f : \{0, 1\} \rightarrow \mathbb{R}_+$,

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f) \leq pq \frac{\log(q) - \log(p)}{q-p} \mathbf{E}_{\mathcal{B}(1,p)}\left((D\sqrt{f})^2\right). \quad (18)$$

Les deux membres de cette inégalité sont homogènes d'ordre 1. Le théorème suivant montre que la mesure de Bernoulli $\mathcal{B}(1,p)$ satisfait à d'autres inégalités de ce type, qui font intervenir l'entropie plutôt que l'entropie du carré. On les qualifie parfois d'*inégalités* L^1 , ou encore d'*inégalités modifiées*.

Théorème 4.3. *Pour toute fonction $f : \{0, 1\} \rightarrow \mathbb{R}_+^*$, on a*

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f) \leq pq \mathbf{E}_{\mathcal{B}(1,p)}\left(\frac{(Df)^2}{f}\right), \quad (19)$$

et

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f) \leq pq \mathbf{E}_{\mathcal{B}(1,p)}(Df D \log(f)), \quad (20)$$

et

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f) \leq pq \mathbf{E}_{\mathcal{B}(1,p)}((f + Df)D \log(f) - Df), \quad (21)$$

et

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f) \leq pq \left(1 - \frac{1}{2} \min \left\{ \frac{f(1)}{f(0)}, \frac{f(0)}{f(1)} \right\}\right) \mathbf{E}_{\mathcal{B}(1,p)}\left(\frac{(Df)^2}{f}\right). \quad (22)$$

La détermination des constantes optimales dans les inégalités ci-dessus constitue un excellent exercice pour le lecteur oisif et curieux.

Démonstration. L'inégalité (21) est plus forte que (19) et que (20). En effet, pour tout $u \in \mathbb{R}_+$ et tout $v \in \mathbb{R}$ tel que $u + v \in \mathbb{R}_+$, on a

$$(u + v)(\log(u + v) - \log(u)) - v \leq v(\log(u + v) - \log(u)), \quad (23)$$

et ceci permet d'obtenir (20) à partir de (21) en considérant le couple $(u, v) = (f, Df)$. De même, l'inégalité

$$(u + v)(\log(u + v) - \log(u)) - v \leq \frac{v^2}{u} \quad (24)$$

permet d'obtenir de la même manière (19) à partir de (21). Démontrons à présent (21). Fixons une fonction $f : \{0, 1\} \rightarrow \mathbb{R}_+$, et considérons la fonction $U : [0, 1] \rightarrow \mathbb{R}$ définie pour tout $p \in [0, 1]$ par

$$U(p) := \mathbf{Ent}_{\mathcal{B}(p,1)}(f) - pq\mathbf{E}_{\mathcal{B}(1,p)}(R(f)),$$

où $R(f) := (f + Df)D \log(f) - Df$. En posant $(a, b) := (f(0), f(1))$, on obtient

$$U(p) = q\Phi(a) + p\Phi(b) - \Phi(qa + pb) - pq(qR(0) + pR(1)),$$

où $\Phi(x) := x \log(x)$. Il nous faut établir que $U \leq 0$ sur $[0, 1]$. La dérivée quatrième de U en $p \in]0, 1[$ s'écrit

$$U''''(p) = -(b - a)^4 \Phi''''(qa + pb).$$

Comme Φ'' est convexe, on a $U'''' \leq 0$ sur $]0, 1[$ et donc U'' est concave. Par conséquent, il existe $0 \leq p_0 \leq p_1 \leq 1$ tels que $U'' \leq 0$ sur $[0, p_0] \cup [p_1, 1]$ et $U'' \geq 0$ on $[p_0, p_1]$. Ainsi, U est concave sur $[0, p_0]$. Or $U(0) = 0$ et $U'(0) \leq 0$, et donc $U \leq 0$ sur $[0, p_0]$ par concavité. Il en découle que $U(p_0) \leq 0$. Par symétrie, on obtient que $U \leq 0$ sur $[p_1, 1]$ et que $U(p_1) \leq 0$. Maintenant, comme U est convexe sur $[p_0, p_1]$ et négative ou nulle aux bords, elle est négative ou nulle sur tout l'intervalle $[p_0, p_1]$. Il en découle finalement que $U \leq 0$ sur $[0, 1]$, et la démonstration est achevée. Il est intéressant de remarquer ici que $R(f) = A(f, Df)$, où A est définie par (10).

Il reste à établir (22). En considérant une fonction U comme précédemment, on montre en exprimant $U'(0)$ et $U'(1)$ en fonction de A que l'inégalité

$$\mathbf{Ent}_{\mathcal{B}(1,p)}(f) \leq pqc\mathbf{E}_{\mathcal{B}(1,p)}\left(\frac{1}{f}\right)$$

a lieu avec $c = \max\{bA(b, a - b), aA(a, b - a)\}$, où $(a, b) := (f(0), f(1))$. Par symétrie, il suffit de considérer le cas où $0 < a < b$, ce qui donne $\min\{a/b, b/a\} = a/b$ et $c = bA(b, a - b)$. Une étude de fonction montre alors finalement que

$$bA(b, b - a) \leq \left(1 - \frac{a}{2b}\right)(b - a)^2.$$

Le résultat désiré en découle puisque $(Df)(a) = -(Df)(b) = b - a$. □

En utilisant l'inégalité élémentaire suivante valable pour tout $u, v \in \mathbb{R}_+$,

$$4(\sqrt{u} - \sqrt{v})^2 \leq (u - v)(\log(u) - \log(v)),$$

on obtient immédiatement, à partir de (18), une inégalité du même type que (20), avec la constante $pq(\log(q) - \log(p))/(4(q - p))$ au lieu de pq . La constante pq est meilleure, i.e. plus petite, si et seulement si p est en dehors de l'intervalle $[p^*, 1 - p^*]$, où p^* est solution de l'équation

$$\log\left(\frac{1}{p^*} - 1\right) = 4(1 - 2p^*).$$

Un calcul numérique sur ordinateur montre que $p^* \approx 0.02124798796$. Par la suite, l'inégalité de Sobolev logarithmique optimale « L^2 » pour $\mathcal{B}(1, p)$ est utilisée avec $p = 1/2$ pour obtenir l'inégalité optimale pour la mesure de Gauss. D'autre part, les inégalités de Sobolev logarithmiques « L^1 » pour $\mathcal{B}(1, p)$, de constante pq , sont utilisées avec $p \rightarrow 0$ pour obtenir les inégalités optimales pour la mesure de Poisson. Cela correspond à l'exploitation de deux zones d'optimalité.

Remarque 4.4 (Invariance des inégalités par changement de support). Les inégalités fournies par les théorèmes 4.1, 4.2, et 4.3 restent valables avec les mêmes constantes pour la mesure de Bernoulli $(1 - p)\delta_a + p\delta_b$ où $a \neq b$. L'opérateur D est alors défini pour toute fonction $f : \{a, b\} \rightarrow \mathbb{R}$ par $(Df)(a) = -(Df)(b) = f(b) - f(a)$. En d'autres termes, ces inégalités, ainsi que leur constantes optimales, dépendent du paramètre p de la mesure de Bernoulli, mais pas de son support $\{a, b\}$.

5 Propriété de tensorisation

La variance et l'entropie pour une mesure produit peuvent être majorées par la somme des moyennes de fonctionnelles du même type sur chaque espace facteur. Cette propriété fondamentale de « tensorisation » découle des formules variationnelles associées.

Soit $\mu := \mu_1 \otimes \cdots \otimes \mu_n$ une mesure de probabilité produit sur un espace mesurable produit $E := E_1 \times \cdots \times E_n$. Pour toute fonction $f : E \rightarrow \mathbb{R}$ appartenant à $L^1(\mu)$ et tout $i \in \{1, \dots, n\}$, on désigne par $\mathbf{E}_{\mu_i} f$ la fonction de E dans \mathbb{R} définie pour tout $x \in E$ par

$$(\mathbf{E}_{\mu_i} f)(x) := \int_{E_i} f(x_1, \dots, x_n) d\mu_i(x_i).$$

Elle ne dépend pas de la composante x_i . On définit de la même manière les fonctions $\mathbf{Var}_{\mu_i}(f)$ et $\mathbf{Ent}_{\mu_i}(f)$. Considérons à présent une fonction $f : E \rightarrow \mathbb{R}$ de la forme $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ où $f_i : E_i \rightarrow \mathbb{R}$ avec $f_i \in L^2(\mu_i)$ pour tout $i \in \{1, \dots, n\}$. On a d'une part

$$\sum_{i=1}^n \mathbf{E}_{\mu}(\mathbf{Var}_{\mu_i}(f)) = \sum_{i=1}^n \mathbf{Var}_{\mu_i}(f) \prod_{j \neq i} \mathbf{E}_{\mu_j}(f_j^2),$$

et d'autre part, en utilisant une somme télescopique, il vient

$$\begin{aligned}\mathbf{Var}_\mu(f) &= \prod_{i=1}^n \mathbf{E}_{\mu_i}(f_i^2) - \prod_{i=1}^n \mathbf{E}_{\mu_i}(f_i)^2 \\ &= \sum_{i=1}^n \mathbf{Var}_{\mu_i}(f_i) \prod_{j=1}^{i-1} \mathbf{E}_{\mu_j}(f_j)^2 \prod_{k=i+1}^n \mathbf{E}_{\mu_k}(f_k^2).\end{aligned}$$

Par conséquent, l'inégalité $\mathbf{E}_{\mu_k}(f_k)^2 \leq \mathbf{E}_{\mu_k}(f_k^2)$ entraîne que $\mathbf{Var}_\mu(f)$ est majoré par $\sum_{i=1}^n \mathbf{E}_\mu(\mathbf{Var}_{\mu_i}(f))$. Le théorème suivant exprime le fait que cette majoration subsiste lorsque f n'est plus une fonction tensorielle, et qu'elle a également lieu pour l'entropie.

Théorème 5.1 (Formules de tensorisation). *Soit $\mu := \mu_1 \otimes \cdots \otimes \mu_n$ une mesure de probabilité sur un espace mesurable produit $E_1 \times \cdots \times E_n$. Pour toute fonction $f : E \rightarrow \mathbb{R}$ telle que $f \in L^2(\mu)$,*

$$\mathbf{Var}_\mu(f) \leq \mathbf{E}_\mu \left(\sum_{i=1}^n \mathbf{Var}_{\mu_i}(f) \right). \quad (25)$$

De même, pour toute fonction $f : E \rightarrow \mathbb{R}_+$ telle que $f \log f \in L^1(\mu)$,

$$\mathbf{Ent}_\mu(f) \leq \mathbf{E}_\mu \left(\sum_{i=1}^n \mathbf{Ent}_{\mu_i}(f) \right). \quad (26)$$

Démonstration. Soit $\Phi : \mathcal{I} \subset \mathbb{R} \rightarrow \mathbb{R}$ la fonction convexe associée à la variance ou à l'entropie. Soit $f : E \rightarrow \mathcal{I}$ telle que $\Phi(f) \in L^1(\mu)$. En vertu de la formule (11) pour le triplet (μ, f, g) , il suffit d'établir que pour toute fonction $g : E \rightarrow \mathcal{I}$ telle que $\Phi(g) \in L^1(\mu)$,

$$\mathbf{E}_\mu^\Phi(g) + \mathbf{E}_\mu((\Phi'(g) - \Phi'(\mathbf{E}_\mu g))(f - g)) \leq \sum_{i=1}^n \mathbf{E}_\mu(\mathbf{E}_{\mu_i}^\Phi f).$$

Nous procédons par minoration de chacun des n termes du membre de droite de l'inégalité ci-dessus. On pose $g_0 := g$ et $g_i := \mathbf{E}_{\mu_1 \otimes \cdots \otimes \mu_i}(g)$ pour tout $i \in \{1, \dots, n\}$. La formule (11) pour le triplet (μ_i, f, g_{i-1}) et l'identité $\mathbf{E}_{\mu_i}(g_{i-1}) = g_i$ donnent

$$\mathbf{E}_{\mu_i}^\Phi(f) \geq \mathbf{E}_{\mu_i}((\Phi'(g_{i-1}) - \Phi'(g_i))(f - g_{i-1})) + \mathbf{E}_{\mu_i}^\Phi(g_{i-1}).$$

L'intégration des deux membres pour μ et l'usage de l'identité $\mathbf{E}_\mu \mathbf{E}_{\mu_i} = \mathbf{E}_\mu$ donne

$$\mathbf{E}_\mu(\mathbf{E}_{\mu_i}^\Phi(f)) \geq \mathbf{E}_\mu((\Phi'(g_{i-1}) - \Phi'(g_i))(f - g_{i-1})) + \mathbf{E}_\mu(\mathbf{E}_{\mu_i}^\Phi g_{i-1}).$$

Comme $g_0 = g$ et $g_n = \mathbf{E}_\mu g$, les identités $\mathbf{E}_\mu \mathbf{E}_{\mu_i} = \mathbf{E}_\mu$ et $\mathbf{E}_{\mu_i}(g_{i-1}) = g_i$ entraînent par simplification télescopique que

$$\sum_{i=1}^n \mathbf{E}_\mu(\mathbf{E}_{\mu_i}^\Phi(g_{i-1})) = \sum_{i=1}^n [\mathbf{E}_\mu(\Phi(g_{i-1})) - \mathbf{E}_\mu(\Phi(g_i))] = \mathbf{E}_\mu^\Phi(g).$$

De la même manière, on obtient par simplification télescopique,

$$\sum_{i=1}^n \mathbf{E}_\mu((\Phi'(g_{i-1}) - \Phi'(g_i))f) = \mathbf{E}_\mu((\Phi'(g) - \Phi'(\mathbf{E}_\mu g))f).$$

D'autre part, comme g_{i-1} et g_i ne dépendent pas des coordonnées d'indice inférieur ou égal à $i-1$, on obtient $\mathbf{E}_\mu(F(g_{i-1}, g_i)g_{i-1}) = \mathbf{E}_\mu(F(g_{i-1}, g_i)g)$ où $F : \mathbb{R}^2 \rightarrow \mathbb{R}$. Cela donne par simplification télescopique

$$-\sum_{i=1}^n \mathbf{E}_\mu((\Phi'(g_{i-1}) - \Phi'(g_i))g_{i-1}) = -\mathbf{E}_\mu((\Phi'(g) - \Phi'(\mathbf{E}_\mu g))g).$$

□

5.1 Au-delà des mesures de probabilité produit

Les preuves données ci-dessus des formules de tensorisations n'ont plus lieu lorsque la mesure de probabilité considérée n'est pas une mesure produit. Le lecteur probabiliste ne sera pas étonné d'apprendre qu'il est cependant possible d'adopter une approche par conditionnement. Soit par exemple μ une mesure de probabilité sur un espace mesurable produit $E_1 \times \cdots \times E_n$, absolument continue par rapport à une mesure de Borel produit $\nu^{\otimes n}$, et de densité Ψ positive. Pour tout $i \in \{1, \dots, n-1\}$, on note $\mu_{(i)}$ l'application à valeurs mesures qui à tout $(x_1, \dots, x_i) \in E_1 \times \cdots \times E_i$ associe la mesure de probabilité sur $E_{i+1} \times \cdots \times E_n$, absolument continue par rapport à $\nu^{\otimes(n-i)}$, et de densité

$$(x_{i+1}, \dots, x_n) \mapsto \frac{1}{Z_{(i)}(x_1, \dots, x_i)} \Psi(x_1, \dots, x_n),$$

où

$$Z_{(i)}(x_1, \dots, x_i) := \int_{E_{i+1} \times \cdots \times E_n} \Psi(x_1, \dots, x_n) d\nu(x_{i+1}) \cdots d\nu(x_n).$$

On définit également $\mu_{(0)}$ comme étant l'application constante et égale à μ , et $\mu_{(n)}$ comme étant l'application qui associe $\delta_{(x_1, \dots, x_n)}$ à (x_1, \dots, x_n) . Maintenant, pour toute fonction mesurable et bornée $f : E_1 \times \cdots \times E_n \rightarrow \mathbb{R}$, on pose $f_i := \mathbf{E}_{\mu_{(i)}}(f)$ pour tout $i \in \{0, \dots, n\}$. En particulier, $f_0 = \mathbf{E}_\mu(f)$ et $f_n = f$. Pour $i \in \{1, \dots, n-1\}$, la fonction f_i ne dépend que des coordonnées dont l'indice est dans $\{1, \dots, i\}$. Une sommation télescopique donne alors

$$\mathbf{E}_\mu^\Phi(f) = \sum_{i=1}^n \mathbf{E}_\mu(\Phi(f_i) - \Phi(f_{i-1})).$$

Comme $f_{i-1} = \mathbf{E}_{\mu_{i-1}}(f_i)$ et $\mathbf{E}_\mu \mathbf{E}_{\mu_{(i-1)}} = \mathbf{E}_\mu$, on obtient finalement

$$\mathbf{E}_\mu^\Phi(f) = \sum_{i=1}^n \mathbf{E}_\mu(\mathbf{E}_{\mu_{(i-1)}}^\Phi(f_i)).$$

On est donc ramené à contrôler la fonction $\mathbf{E}_{\mu_{(i-1)}}^{\Phi}(f_i)$ pour tout i . Cette approche joue un grand rôle en mécanique statistique, où les mesures non produit sont des objet d'étude naturels. On retiendra que «l'entropie est la somme des moyennes des entropies des espérances conditionnelles ».

Lorsque $\mu = \mu_1 \otimes \cdots \otimes \mu_n$, l'application à valeur mesures $\mu_{(i)}$ est constante et égale à $\mu_{i+1} \otimes \cdots \otimes \mu_n$ pour tout $i \in \{1, \dots, n-1\}$. La formule s'écrit alors

$$\mathbf{E}_{\mu}^{\Phi}(f) = \sum_{i=1}^n \mathbf{E}_{\mu}(\mathbf{E}_{\mu_i \otimes \cdots \otimes \mu_n}^{\Phi}(\mathbf{E}_{\mu_{i+1} \otimes \cdots \otimes \mu_n} f)).$$

Remarquons que l'usage direct de l'inégalité de Jensen dans l'identité ci-dessus pour la fonction convexe Φ ne conduit pas à la formule de tensorisation, qui s'écrit :

$$\mathbf{E}_{\mu}^{\Phi}(f) \leq \sum_{i=1}^n \mathbf{E}_{\mu}(\mathbf{E}_{\mu_i}^{\Phi}(f)).$$

6 Mesures de Gauss sur \mathbb{R}^n

Dans cette section, nous établissons les inégalités de Poincaré et de Sobolev logarithmiques optimales pour la mesure de Gauss standard sur \mathbb{R} en tensorisant les inégalités optimales pour la mesure de Bernoulli symétrique, puis en faisant tendre la dimension de l'espace produit ainsi construit vers l'infini. Le théorème central limite (7) fait naturellement apparaître la mesure de Gauss standard par ce procédé. À leur tour, les constantes optimales pour la mesure de Gauss sur \mathbb{R}^n se déduisent de celles pour la mesure de Gauss standard sur \mathbb{R} par tensorisation, translation, et dilatation.

Théorème 6.1. *On a $c_{\mathcal{P}}(\mathcal{N}(0,1)) = 1$ et $c_{\text{SL}}(\mathcal{N}(0,1)) = 2$. En d'autres termes, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ dans $\mathbf{H}^1(\mathcal{N}(0,1))$,*

$$\mathbf{Var}_{\mathcal{N}(0,1)}(f) \leq \mathbf{E}_{\mathcal{N}(0,1)}(f'^2) \tag{27}$$

et

$$\mathbf{Ent}_{\mathcal{N}(0,1)}(f^2) \leq 2\mathbf{E}_{\mathcal{N}(0,1)}(f'^2), \tag{28}$$

et les constantes sont optimales.

Démonstration. Nous donnons ici une preuve de l'inégalité de Sobolev logarithmique (28), celle de l'inégalité de Poincaré (27) étant en tout point similaire. Nous établissons (28), avec la constante 2, sur la classe des fonctions de \mathbb{R} and \mathbb{R} , de classe \mathcal{C}^2 et à support compact. Cette classe s'étend par régularisation et troncature à l'espace de Sobolev $\mathbf{H}^1(\mathcal{N}(0,1))$, contenant la fonction $f(x) = \exp(x)$, pour laquelle l'égalité est atteinte. L'optimalité de la constante 1 dans (27) se vérifie de la même manière sur la fonction $f(x) = x$.

Soit $p \in]0, 1[$, et $c_p := c_{\text{SL}}(\mathcal{B}(1,p))$. L'usage du théorème 5.1 pour la mesure produit $\mathcal{B}(1,p)^{\otimes n}$ sur l'espace produit $\{0,1\}^n$, suivi du théorème 4.2, entraîne que pour toute

fonction $F : \{0, 1\}^n \rightarrow \mathbb{R}$:

$$\mathbf{Ent}_{\mathcal{B}(1,p)^{\otimes n}}(F^2) \leq c_p \sum_{i=1}^n \mathbf{E}_{\mathcal{B}(1,p)^{\otimes n}} \left(|F_i(1) - F_i(0)|^2 \right),$$

où l'indice i indique que seule la composante d'indice i est concernée. Appliquons à présent l'inégalité précédente à une fonction F de la forme $f \circ \kappa_n$, où $f : \mathbb{R} \rightarrow \mathbb{R}$ est de classe \mathcal{C}^2 et où κ_n est comme en (7). Le membre de gauche est facile à étudier. En effet, en vertu de (7), la mesure image de $\mathcal{B}(1,p)^{\otimes n}$ par κ_n converge étroitement vers la mesure de Gauss standard $\mathcal{N}(0,1)$. Comme f est \mathcal{C}^2 et à support compact, elle est continue et bornée, et donc

$$\lim_{n \rightarrow \infty} \mathbf{Ent}_{\mathcal{B}(1,p)^{\otimes n}}((f \circ \kappa_n)^2) = \mathbf{Ent}_{\mathcal{N}(0,1)}(f^2).$$

Il reste à étudier la convergence du membre de droite. La fonction f étant de classe \mathcal{C}^2 et à support compact, ses deux premières dérivées sont bornées par une constante K . En particulier, par une formule de Taylor,

$$|(f \circ \kappa_n)_i(1) - (f \circ \kappa_n)_i(0)| \leq \frac{1}{\sqrt{npq}} |f'(\kappa_{n,i}(x))| + \frac{K}{npq},$$

où $\kappa_{n,i}(x) := \kappa_n(x) - (npq)^{-1/2}x_i$. Or la mesure image de $\mathcal{B}(1,p)^{\otimes n}$ par $\kappa_{n,i}$ ne dépend pas de i . Baptisons-la ν_n . Cela donne la majoration suivante

$$\sum_{i=1}^n \mathbf{E}_{\mathcal{B}(1,p)^{\otimes n}} \left(|(f \circ \kappa_n)_i(1) - (f \circ \kappa_n)_i(0)|^2 \right) \leq \frac{1}{pq} \mathbf{E}_{\nu_n} \left(|f'|^2 \right) + O\left(\frac{1}{\sqrt{n}}\right).$$

Maintenant, il est facile de voir qu'ici encore, en vertu de (7), la mesure de probabilité ν_n converge étroitement vers $\mathcal{N}(0,1)$. Il en découle que

$$\limsup_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{E}_{\mathcal{B}(1,p)^{\otimes n}} \left(|(f \circ \kappa_n)_i(1) - (f \circ \kappa_n)_i(0)|^2 \right) \leq \frac{1}{pq} \mathbf{E}_{\mathcal{N}(0,1)} \left(|f'|^2 \right).$$

Ainsi, pour $p = 1/2$, on obtient $\mathbf{Ent}_{\mathcal{N}(0,1)}(f^2) \leq 2\mathbf{E}_{\mathcal{N}(0,1)}(f'^2)$. \square

Corollaire 6.2. *Pour tout $m \in \mathbb{R}$ et toute matrice Σ carrée $n \times n$ symétrique semi-définie positive et non identiquement nulle, on a $c_{\mathbb{P}}(\mathcal{N}(m, \Sigma)) = \rho(\Sigma)$ et $c_{\text{SL}}(\mathcal{N}(m, \Sigma)) = 2\rho(\Sigma)$, où $\rho(\Sigma)$ est le rayon spectral de Σ . En d'autres termes, pour toute fonction $f \in \mathbf{H}^1(\mathcal{N}(m, \Sigma))$,*

$$\mathbf{Var}_{\mathcal{N}(m, \Sigma)}(f) \leq \rho(\Sigma) \mathbf{E}_{\mathcal{N}(m, \Sigma)} \left(|\nabla f|^2 \right) \quad (29)$$

et

$$\mathbf{Ent}_{\mathcal{N}(m, \Sigma)}(f^2) \leq 2\rho(\Sigma) \mathbf{E}_{\mathcal{N}(m, \Sigma)} \left(|\nabla f|^2 \right), \quad (30)$$

et les constantes sont optimales.

Démonstration. Soit A une matrice carrée $n \times n$ telle que $AA^\top = \Sigma$. Considérons la transformation affine $T : x \in \mathbb{R}^n \mapsto Ax + m$. La mesure image de $\mathcal{N}(0, I_n)$ par T est la mesure $\mathcal{N}(m, \Sigma)$. De plus,

$$|\nabla(f \circ T)|^2 = \nabla f \cdot \Sigma \nabla f \leq \rho(\Sigma) |\nabla f|^2$$

pour tout $f \in H^1(\mathcal{N}(0, I_n))$. Enfin, $\{f \circ T; f \in H^1(\mathcal{N}(0, I_n))\} = H^1(\mathcal{N}(m, \Sigma))$. De cette manière, il suffit de démontrer les inégalités pour $\mathcal{N}(0, I_n)$. Ce dernier cas se traite par tensorisation, puisque $\mathcal{N}(0, I_n) = \mathcal{N}(0, 1)^{\otimes n}$. Pour l'entropie, par exemple, le théorème 5.1 appliqué à $\mu_i = \mathcal{N}(0, 1)$ sur $E_i = \mathbb{R}$, suivi du théorème 6.1, donne pour tout $f \in H^1(\mathcal{N}(0, I_n))$,

$$\mathbf{Ent}_{\mathcal{N}(0, I_n)}(f^2) \leq \sum_{i=1}^n \mathbf{E}_{\mathcal{N}(0, I_n)}((\partial_i f)^2) = \mathbf{E}_{\mathcal{N}(0, I_n)}(|\nabla f|^2).$$

□

L'inégalité (29) peut être obtenue directement en utilisant les polynômes orthogonaux de Hermite. On peut également la voir comme une conséquence de l'inégalité de Sobolev logarithmique (30) au moyen du théorème 3.1.

Les constantes optimales de Poincaré et de Sobolev logarithmique ne dépendent pas de la moyenne m , qui s'interprète comme une « localisation » de la mesure de Gauss. La moyenne pour la mesure de Gauss joue un rôle similaire à celui du support pour la mesure de Bernoulli.

Pour une fonction lisse $f : \mathbb{R}^n \rightarrow \mathbb{R}$, on a $|\nabla|f|| \leq |\nabla f|$ sur l'ensemble des $x \in \mathbb{R}^n$ tels que $f(x) \neq 0$. Cette formule est l'analogie de la formule $|D|f|| \leq |Df|$ sur l'espace à deux points $\{0, 1\}$. On dispose de plus, sur l'espace euclidien \mathbb{R}^n , de la *règle de la chaîne*

$$\nabla \Psi(f) = \Psi'(f) \nabla f,$$

valable pour toutes fonctions lisses $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $\Psi : \mathbb{R} \rightarrow \mathbb{R}$. Cela permet de transformer les formes « L^1 » des inégalités de Sobolev logarithmiques en forme « L^2 » et réciproquement. On écrit pour $f > 0$:

$$4|\nabla \sqrt{f}|^2 = \frac{|\nabla f|^2}{f} = \nabla f \cdot \nabla \log(f).$$

De même, on dispose des formules inverses $4|\nabla f|^2 = f^{-2}|\nabla f^2|^2 = \nabla f^2 \cdot \nabla \log(f^2)$. Ces formules font défaut sur l'espace à deux points, pour lequel la règle de la chaîne n'a pas lieu. Les comparaisons palliatives du type $4(D\sqrt{f})^2 \leq DfD \log(f)$ permettent seulement le passage de formes « L^2 » à formes « L^1 ». Cela est directement relié à l'absence de véritables processus de diffusion sur les espaces discrets.

6.1 Mesures à densité log-concave

Une lecture attentive de la preuve des inégalités (29) et (30) montre qu'il est possible d'en donner les expressions équivalentes suivantes, par translation et dilatation.

Corollaire 6.3 (Formulation matricielle). *Pour tout $m \in \mathbb{R}$, toute matrice Σ carrée $n \times n$ symétrique semi-définie positive, et tout $f \in \mathcal{H}^1(\mathcal{N}(m, \Sigma))$,*

$$\mathbf{Var}_{\mathcal{N}(m, \Sigma)}(f) \leq \mathbf{E}_{\mathcal{N}(m, \Sigma)}(\Sigma \nabla f \cdot \nabla f), \quad (31)$$

et

$$\mathbf{Ent}_{\mathcal{N}(m, \Sigma)}(f^2) \leq 2\mathbf{E}_{\mathcal{N}(m, \Sigma)}(\Sigma \nabla f \cdot \nabla f). \quad (32)$$

La mesure de Gauss $\mathcal{N}(m, \Sigma)$ sur \mathbb{R}^n , de matrice de covariance inversible Σ , admet pour densité la fonction e^H où $H : \mathbb{R}^n \rightarrow \mathbb{R}$ est définie pour tout $x \in \mathbb{R}^n$ par

$$H(x) = \frac{1}{2}(x - m) \cdot \Sigma^{-1}(x - m) - \frac{n}{2} \log(2\pi).$$

En particulier, le logarithme de la densité de probabilité de la mesure de Gauss $\mathcal{N}(m, \Sigma)$ est concave. Cela mène, par comparaison, à la définition suivante.

Définition 6.4. Une mesure de probabilité μ sur \mathbb{R}^n est *log-concave* de courbure $\kappa \in \mathbb{R}_+$ lorsque qu'elle est absolument continue par rapport à la mesure de Lebesgue et qu'elle admet une densité $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 s'écrivant $f = e^{-H}$ où $H : \mathbb{R}^n \rightarrow \mathbb{R}$ vérifie pour tout x, y dans \mathbb{R}^n

$$y \cdot (\nabla^2 H(x))y \geq \kappa |y|^2.$$

La notation $\nabla^2 H(x)$ désigne la matrice hessienne de H au point x .

La condition s'écrit au sens des formes quadratiques $\nabla^2 H(x) \geq \kappa \mathbf{I}_n$ pour tout $x \in \mathbb{R}^n$. Elle est équivalente à dire que la plus petite valeur propre de la matrice symétrique $\nabla^2 H(x)$ est minorée par κ pour tout $x \in \mathbb{R}^n$. Pour la mesure de Gauss $\mathcal{N}(m, \Sigma)$ sur \mathbb{R}^n de matrice de covariance Σ inversible, on a $\nabla^2 H(x) = \Sigma^{-1}$ pour tout $x \in \mathbb{R}^n$. Ainsi, $\mathcal{N}(m, \Sigma)$ est log-concave de courbure $\kappa = \rho(\Sigma)^{-1}$, où $\rho(\Sigma)$ désigne le rayon spectral de Σ . Comme Σ est symétrique positive, $\rho(\Sigma)$ est la plus grande valeur propre de Σ . L'inégalité (31) est un cas particulier du théorème suivant.

Théorème 6.5 (Brascamp-Lieb). *Soit μ une mesure de probabilité log-concave sur \mathbb{R}^n de densité $f = e^{-H}$ et de courbure positive. Alors pour toute fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 ,*

$$\mathbf{Var}_{\mu}(f) \leq \mathbf{E}_{\mu}(\nabla f \cdot (\nabla^2 H)^{-1} \nabla f).$$

Démonstration. Considérons le cas $n = 1$, et notons h la fonction H . Comme μ est log-concave de courbure positive, la fonction h possède un unique minimum en un point $m \in \mathbb{R}$. Soit $k : \mathbb{R} \rightarrow \mathbb{R}$ la fonction définie par $k(x) := (f(x) - f(m))/h'(x)$ si $x \neq m$, et $k(m) := f'(m)/h''(m)$. Cette fonction est continue en m . En dehors de m , la fonction k

est de classe \mathcal{C}^1 , et $k'h' + kh'' = f'$. On écrit à présent :

$$\begin{aligned}
\mathbf{E}_\mu(h''^{-1}f'^2) &= \int \left(\frac{(k'h')^2}{h''} + 2kk'h' + k^2h'' \right) e^{-h} dx \\
&= \int \left(\frac{(k'h')^2}{h''} + (kh')^2 \right) d\mu + [k^2h'e^{-h}]_{-\infty}^m + [k^2h'e^{-h}]_m^{+\infty} \\
&\geq \int (kh')^2 d\mu \\
&= \int (f(x) - f(m))^2 d\mu(x) \\
&\geq \mathbf{Var}_\mu(f).
\end{aligned}$$

La seconde égalité découle d'une intégration par parties, tandis que la dernière inégalité découle de l'identité (1). La preuve du cas $n > 1$ peut être menée par récurrence sur n . Nous l'omettons ici car elle est assez laborieuse. Elle fait appel en particulier à la log-concavité des mesures partielles obtenues en intégrant la densité de μ selon une seule coordonnée. \square

Malheureusement, et comme peut le montrer un contre exemple judicieux, ce théorème de Brascamp-Lieb ne possède pas d'équivalent pour l'entropie. Il n'est pas possible de remplacer la matrice constante Σ par l'inverse de la matrice hessienne $(\nabla^2 H)^{-1}$ dans (32), lorsque $\mathcal{N}(m, \Sigma)$ est remplacée par une mesure de probabilité log-concave de densité e^{-H} . Cependant, les inégalités de Poincaré (29) et de Sobolev logarithmique (30) pour la mesure de Gauss sur \mathbb{R}^n peuvent se généraliser aux mesures de probabilité log-concave par déformation.

Théorème 6.6 (Déformation lipschitzienne). *Soit $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction différentiable vérifiant $\kappa|F(x) - F(y)|^2 \leq |x - y|^2$ pour tout x, y dans \mathbb{R}^n , où κ est une constante positive. Soit μ la mesure image de $\mathcal{N}(0, I_n)$ par F . Alors, pour toute fonction $g \in \mathbf{H}^1(\mu)$,*

$$\mathbf{Var}_\mu(g) \leq \kappa^{-1} \mathbf{E}_\mu(|\nabla g|^2)$$

et

$$\mathbf{Ent}_\mu(g^2) \leq 2\kappa^{-1} \mathbf{E}_\mu(|\nabla g|^2).$$

Démonstration. On se ramène aux fonctions de classe \mathcal{C}^2 et à support compact par régularisation et troncature. Soit donc g de ce type. La fonction $f := F \circ g$ vérifie $\mathbf{Ent}_{\mathcal{N}(0, I_n)}(f^2) = \mathbf{Ent}_\mu(g^2)$ et $\mathbf{Var}_{\mathcal{N}(0, I_n)}(f) = \mathbf{Var}_\mu(g)$. D'autre part,

$$|\nabla f|^2 \leq \|\text{Jac}(F)\|^2 |\nabla g|^2,$$

où $\|\text{Jac}(F)\|$ est la norme d'opérateur euclidienne de la matrice jacobienne de F . Or $\|\text{Jac}(F)\|^2 \leq \kappa^{-1}$, et donc $\mathbf{E}_{\mathcal{N}(0, I_n)}(|\nabla f|^2) \leq \kappa^{-1} \mathbf{E}_\mu(|\nabla g|^2)$. Les inégalités désirées découlent alors des inégalités du corollaire 6.2 concernant la mesure de Gauss $\mathcal{N}(0, I_n)$. \square

Le théorème 6.6 fait poindre une stabilité par transformation lipschitzienne de la mesure des inégalités de Poincaré et de Sobolev logarithmiques sur \mathbb{R}^n avec $Q(f) = |\nabla f|^2$. Nous avons déjà utilisé cette stabilité dans la preuve du corollaire 6.6. Il s'agit d'une conséquence directe d'une propriété de l'énergie utilisée. Il se trouve que les mesures log-concaves sont des déformations lipschitziennes de la mesure de Gauss. C'est le fameux théorème de Caffarelli. Nous en omettons la preuve, car elle dépasse le cadre de notre exposé élémentaire. Elle fait intervenir le théorème de Brenier pour le transport optimal de la mesure et l'équation de Monge-Ampère associée.

Théorème 6.7 (Caffarelli). *Soit μ une mesure de probabilité log-concave sur \mathbb{R}^n , de courbure κ . Alors μ est la mesure image de la mesure de Gauss $\mathcal{N}(0, I_n)$ par une fonction $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ différentiable vérifiant $\kappa|F(x) - F(y)|^2 \leq |x - y|^2$ pour tout x, y dans \mathbb{R}^n .*

La combinaison des théorèmes 6.6 et 6.7 conduit aux inégalités de Poincaré et de Sobolev logarithmiques pour les mesures log-concaves.

Corollaire 6.8 (Mesures log-concaves). *Soit μ une mesure de probabilité log-concave sur \mathbb{R}^n , de courbure $\kappa > 0$. Alors, pour toute fonction $g \in H^1(\mu)$,*

$$\mathbf{Var}_\mu(g) \leq \kappa^{-1} \mathbf{E}_\mu(|\nabla g|^2)$$

et

$$\mathbf{Ent}_\mu(g^2) \leq 2\kappa^{-1} \mathbf{E}_\mu(|\nabla g|^2).$$

Les constantes sont de plus optimales pour les mesures de Gauss.

6.2 Formulation euclidienne et mesure de Lebesgue

L'inégalité de Sobolev logarithmique optimale (30) pour la mesure de Gauss $\mathcal{N}(0, I_n)$ possède une formulation équivalente sur la mesure de Lebesgue sur \mathbb{R}^n , comme l'exprime le corollaire suivant.

Corollaire 6.9 (Stam-Carlen). *Pour toute fonction à décroissance rapide $g : \mathbb{R}^n \rightarrow \mathbb{R}$ vérifiant $\int_{\mathbb{R}^n} g^2 dx = 1$, on a*

$$\int_{\mathbb{R}^n} g^2 \log(g^2) dx \leq \frac{n}{2} \log \left(\frac{2}{\pi n e} \int_{\mathbb{R}^n} |\nabla g|^2 dx \right). \quad (33)$$

De plus, cette inégalité devient une égalité pour des fonctions g telles que $\log(g)$ est quadratique, et pour elles seulement.

Démonstration. Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à décroissance rapide vérifiant $\int_{\mathbb{R}^n} h^2 dx = 1$. L'inégalité de Sobolev logarithmique (30) pour la mesure de Gauss $\mathcal{N}(0, I_n)$ et une fonction f vérifiant $f^2(x) = h^2(x)(2\pi)^{n/2} e^{-|x|^2/2}$ fournit, après intégration par parties, l'inégalité suivante :

$$\int_{\mathbb{R}^n} h^2 \log(h^2) dx \leq 2 \int_{\mathbb{R}^n} |\nabla h|^2 dx - n - \frac{n}{2} \log(2\pi).$$

Le résultat désiré en découle en écrivant l'inégalité ci-dessus pour $h(x) := \lambda^{n/2}g(\lambda x)$, puis en optimisant en $\lambda > 0$. Le procédé décrit ici pour passer de (30) à (33) est réversible au moyen d'un changement de fonction adéquat. Le fait que les fonctions log-quadratiques sont les seules fonctions extrémales est plus délicat, et nous en omettons la preuve ici. \square

L'inégalité (33) apparaît, sous une forme différente, en théorie de l'information. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ une densité de probabilité par rapport à la mesure de Lebesgue, telle que $f \log(f)$ soit Lebesgue intégrable. Soit $\mathbf{H}(f)$ son entropie de Shannon⁵ définie par

$$\mathbf{H}(f) := - \int_{\mathbb{R}^n} f \log(f) dx. \quad (34)$$

L'entropie exponentielle de Shannon $\mathbf{N}(f)$ et l'information de Fisher $\mathbf{J}(f)$ sont définies à leur tour, dans $\mathbb{R}_+ \cup \{+\infty\}$, par

$$\mathbf{N}(f) := (2\pi e)^{-1} \exp\left(\frac{2}{n}\mathbf{H}(f)\right) \quad \text{et} \quad \mathbf{J}(f) := \int_{\mathbb{R}^n} |\nabla \log(f(x))|^2 dx.$$

L'inégalité (33) se réécrit alors simplement $\mathbf{N}(X)\mathbf{J}(X) \geq n$, avec égalité lorsque $\mathcal{L}(X)$ est une mesure de Gauss. L'information de Fisher \mathbf{J} apparaît également, sous une forme différente, dans une formulation en terme de mesures des inégalités de Sobolev logarithmiques. Elle s'énonce de la façon suivante : pour toute mesure de probabilité ν absolument continue par rapport à une mesure de probabilité de référence μ , on a :

$$\mathbf{Ent}_\mu\left(\frac{d\nu}{d\mu}\right) \leq c \mathbf{E}_\mu\left(\left|\nabla \log\left(\frac{d\nu}{d\mu}\right)\right|^2\right).$$

Le membre de gauche est connu sous le nom d'*entropie relative* de μ par rapport à ν , ou encore d'information de Kullback-Leibler. Cette formulation intervient en théorie du transport optimal de la mesure et en théorie des grandes déviations. La fonctionnelle convexe entropie relative apparaît par exemple comme la fonction de taux du principe de grandes déviations de Sanov.

6.3 Mesure uniforme sur $[0, 1]$

Il est naturel de s'intéresser à présent à la mesure uniforme \mathcal{U} , trace de la mesure de Lebesgue sur l'intervalle $[0, 1] \subset \mathbb{R}$. Sa densité par rapport à la mesure de Lebesgue sur \mathbb{R} est l'indicatrice de l'intervalle $[0, 1]$. Pour tout $x \in \mathbb{R}$, on note $F(x)$ la masse affectée à l'intervalle $] -\infty, x]$ par $\mathcal{N}(0, 1)$. Alors \mathcal{U} est la mesure image de $\mathcal{N}(0, 1)$ par $F : \mathbb{R} \rightarrow [0, 1]$. Comme $F'^2 \leq (2\pi)^{-1}$, le théorème 6.6 entraîne que pour toute fonction lisse $g : [0, 1] \rightarrow \mathbb{R}$,

$$\mathbf{Var}_\mathcal{U}(g) \leq \frac{1}{2\pi} \mathbf{E}_\mathcal{U}(g'^2) \quad \text{et} \quad \mathbf{Ent}_\mathcal{U}(g^2) \leq \frac{1}{\pi} \mathbf{E}_\mathcal{U}(g'^2).$$

⁵La quantité $-\mathbf{H}(f)$ sur \mathbb{R}^6 est l'entropie de Boltzmann en théorie cinétique des gaz.

Malheureusement, ces constantes ne sont pas optimales. On peut montrer par d'autres méthodes que les constantes optimales sur la classe des fonctions \mathcal{C}^1 est $\frac{1}{\pi^2}$ pour l'inégalité de Poincaré et $\frac{2}{\pi^2}$ pour l'inégalité de Sobolev logarithmique. En revanche, lorsqu'on restreint la classe aux fonctions qui prennent la même valeur en 0 et 1, les constantes optimales sont $\frac{1}{4\pi^2}$ et $\frac{1}{2\pi^2}$ respectivement. Ce phénomène fournit un exemple simple pour lequel les constantes optimales dépendent de la classe de fonctions choisie.

6.4 Renforcement de Carlen par transformée de Wiener

Il est possible d'insérer un terme supplémentaire dans le membre de gauche de l'inégalité de Sobolev logarithmique optimale (30) pour la mesure de Gauss. En effet, si Γ désigne par exemple la mesure de Gauss $\mathcal{N}(0, (4\pi)^{-1}\mathbf{I}_n)$, alors

$$\mathbf{Ent}_\Gamma(f^2) + \mathbf{Ent}_\Gamma(\mathcal{W}(f)^2) \leq \frac{1}{2\pi} \mathbf{E}_\Gamma(|\nabla f|^2),$$

où $\mathcal{W}(f)$ désigne la transformée de Wiener de f définie par $\mathcal{W}(f) := \widehat{(f/\gamma)}\gamma$, où $\gamma(x) := e^{-\pi|x|^2}$, et où le chapeau désigne la transformée de Fourier. Cette inégalité optimale renforcée est la réécriture d'un principe d'incertitude célèbre de Beckner-Hirschman, qui affirme que pour toute fonction $\varphi : \mathbb{R}^n \rightarrow \mathbb{C}$ telle que $|\varphi|^2$ est une densité de probabilité par rapport à la mesure de Lebesgue sur \mathbb{R}^n ,

$$\mathbf{H}(|\varphi|^2) + \mathbf{H}(|\widehat{\varphi}|^2) \geq n(1 - \log(2)),$$

où \mathbf{H} est l'entropie de Shannon définie par (34). Le changement de fonction à utiliser pour la réécriture est donné par $\varphi(x) = 2^{n/4}\gamma(x)f(x)$. Ce principe d'incertitude découle de l'inégalité de Young optimale sur la norme L^p du produit de convolution.

7 Mesure de Poisson

Dans toute cette section, $\lambda > 0$. Supposons que la mesure de probabilité $\mathcal{P}(\lambda)$ vérifie une inégalité de Sobolev logarithmique de constante $c > 0$, de la forme :

$$\mathbf{Ent}_{\mathcal{P}(\lambda)}(f^2) \leq c \mathbf{E}_{\mathcal{P}(\lambda)}(|Df|^2),$$

pour toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ à support fini. Par troncature, on étend cette inégalité à toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ telle que $f^2 \in L^1+(\mathcal{P}(\lambda))$. En appliquant l'inégalité obtenue aux fonctions f_k , indicatrices des ensembles infinis $A_k := \{k+1, \dots\}$, on obtient :

$$-\mathcal{P}(\lambda)(A_k) \log \mathcal{P}(\lambda)(A_k) \leq c \mathcal{P}(\lambda)(\{k\}),$$

ce qui contredit la finitude de c lorsque k tend vers l'infini. Ainsi, la mesure de Poisson ne satisfait pas d'inégalité de Sobolev logarithmique sur la classe des fonctions $f : \mathbb{N} \rightarrow \mathbb{R}$ à support fini. Elle satisfait cependant à une inégalité de Poincaré et à des inégalités de Sobolev logarithmiques modifiées similaires à celles établies pour la mesure de Bernoulli.

Théorème 7.1. *Pour tout $\lambda > 0$, on a $c_{\mathcal{P}(\lambda)} = \lambda$. En d'autres termes, pour toute fonction $f \in L^2(\mathcal{P}(\lambda))$,*

$$\mathbf{Var}_{\mathcal{P}(\lambda)}(f) \leq \lambda \mathbf{E}_{\mathcal{P}(\lambda)}((Df)^2), \quad (35)$$

et la constante est optimale. De plus, pour tout $f \in L^{1+}(\mathcal{P}(\lambda))$,

$$\mathbf{Ent}_{\mathcal{P}(\lambda)}(f) \leq \lambda \mathbf{E}_{\mathcal{P}(\lambda)}\left(\frac{(Df)^2}{f}\right), \quad (36)$$

et

$$\mathbf{Ent}_{\mathcal{P}(\lambda)}(f) \leq \lambda \mathbf{E}_{\mathcal{P}(\lambda)}(Df D \log f), \quad (37)$$

et

$$\mathbf{Ent}_{\mathcal{P}(\lambda)}(f) \leq \lambda \mathbf{E}_{\mathcal{P}(\lambda)}((f + Df)D \log f - Df), \quad (38)$$

et les constantes sont optimales.

Si la règle de la chaîne était valable sur \mathbb{N} , le rapport entre la constante optimale dans (35) et la constante optimale dans (37) (ou (36)) serait de 2. Cela n'est bien entendu pas le cas ici.

Démonstration. Pour établir les inégalités désirées, on se ramène tout d'abord aux fonctions f à support fini par troncature. Les preuves des quatre inégalités sont similaires. Établissons par exemple (37). Le théorème 5.1 appliqué à μ_i sur $E_i = \{0, 1\}$, suit de l'inégalité (20) fournit pour toute fonction $F : \{0, 1\}^n \rightarrow \mathbb{R}_+$,

$$\mathbf{Ent}_{\mathcal{B}(1,p)^{\otimes n}}(F) \leq pq \mathbf{E}_{\mathcal{B}(1,p)^{\otimes n}}\left(\sum_{i=1}^n D_i F D_i \log(F)\right),$$

où D_i opère sur la composante d'indice i . Si $f : \mathbb{N} \rightarrow \mathbb{R}_+$ est à support fini, on pose $F = f \circ s_n$ où s_n est définie comme dans le théorème (1.7). Or il s'avère que pour tout $x \in \{0, 1\}^n$,

$$\sum_{i=1}^n (D_i F D_i \log(F))(x) = (n - s_n(x))R(s_n(x)) + s_n(x)R^*(s_n(x)),$$

$R := Df D \log(f)$, et $R^* := D^* f D^* \log(f)$, avec pour tout $g : \mathbb{Z} \rightarrow \mathbb{R}$ et tout $n \in \mathbb{Z}$,

$$(D^* g)(n) := g(n-1) - g(n). \quad (39)$$

Pour n assez grand, on considère $p := p_n := \lambda/n$ et $q_n := 1 - p_n$, de sorte que $\lim_{n \rightarrow \infty} (np_n, p_n, q_n, p_n q_n, np_n q_n) = (\lambda, 0, 1, 0, \lambda)$. La mesure image de $\mathcal{B}(1, p_n)^{\otimes n}$ par s_n est $\mathcal{B}(n, p_n)$, et converge étroitement vers $\mathcal{P}(\lambda)$ lorsque $n \rightarrow \infty$ en vertu de (6). Par conséquent

$$\lim_{n \rightarrow \infty} \mathbf{Ent}_{\mathcal{B}(1,p_n)^{\otimes n}}(F) = \mathbf{Ent}_{\mathcal{P}(\lambda)}(f),$$

et

$$\lim_{n \rightarrow \infty} p_n q_n \mathbf{E}_{\mathcal{B}(1,p)^{\otimes n}}\left(\sum_{i=1}^n D_i F D_i \log(F)\right) = \lambda \mathbf{E}_{\mathcal{P}(\lambda)}(Df D \log(f)).$$

Les termes multipliés par $s_n(x)$ ont été écrasés par $\lim_{n \rightarrow \infty} p_n q_n = 0$, tandis que les termes multipliés par n ont survécu grâce à $\lim_{n \rightarrow \infty} n p_n q_n = \lambda$. L'inégalité (37) est donc démontrée. Les inégalités (16), (19), et (21) fournissent par le même procédé les inégalités désirées (35), (36), et (38).

Il est également possible de déduire (36) et (37) directement à partir de (38), en procédant comme dans le théorème 4.3. Ceci montre en particulier que l'optimalité de la constante de (38) découle de l'optimalité de la constante dans (36) ou (37). L'optimalité de la constante λ dans (36) et (37) se vérifie sur la famille de fonctions f de la forme $f(n) = \exp(-\alpha n)$ en faisant tendre α vers 0. L'optimalité de la constante λ dans (35) se vérifie sur les fonctions f linéaires. \square

Il est possible de déduire l'inégalité de Poincaré optimale (27) pour la mesure de Gauss à partir de l'inégalité de Poincaré (35) pour la mesure de Poisson, en observant que $\mathcal{P}(n\lambda) = \mathcal{P}(\lambda)^{*n}$ pour tout $n \in \mathbb{N}^*$ et en utilisant le théorème central limite 1.7 avec $\mu = \mathcal{P}(\lambda)$. Ce procédé est un cas très particulier des «limites fluides» utilisées en théorie des files d'attente, et des «limites hydrodynamiques» utilisées en mécanique statistique. La même méthode, appliquée à l'inégalité (38), conduit à l'inégalité de Sobolev logarithmique optimale pour la mesure de Gauss (28). Les inégalités (37) et (36) quant à elles fournissent une constante non-optimale : 4 au lieu de 2. Cela s'explique par le fait que les comparaisons (23) et (24) s'améliorent d'un facteur multiplicatif 2 lorsque v tend vers 0. Enfin, la tensorisation de l'inégalité (22) pour la mesure de Bernoulli et l'usage de (7) conduisent à la meilleure constante 2. Ces phénomènes montrent encore une fois les conséquences de l'absence de règle de la chaîne sur les espaces discrets, ainsi que le manque de finesse des inégalités (20), (19) et (37), (36).

Contrairement à la famille des mesures de Gauss $\{\mathcal{N}(m, v); (m, v) \in \mathbb{R} \times \mathbb{R}_+\}$, la famille des mesures de Poisson $\{\delta_n * \mathcal{P}(\lambda); (n, \lambda) \in \mathbb{N} \times \mathbb{R}_+\}$ n'est pas stable par dilatation. En revanche, ces deux familles sont stables par convolution. Cette propriété permet une approche par interpolation, que nous présentons maintenant.

8 Interpolation par semi-groupes de convolution

Cette section est consacrée à l'établissement des inégalités de Poincaré et de Sobolev logarithmiques pour les mesures de Gauss et de Poisson par interpolation, à partir de la mesure de Dirac. Les fonctionnelles variance et entropie sont identiquement nulles pour une mesure de Dirac. L'interpolation est fournie par des semi-groupes de Markov bien choisis. Ces interpolations ont lieu sur l'intervalle de temps $[0, t]$. La mesure de Dirac correspond au temps 0 et la mesure cible au temps t . Cette technique fournit les versions « L^1 » de l'inégalité de Sobolev logarithmique.

8.1 Semi-groupe de la chaleur

Pour toute fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continue et bornée et tout réel $t > 0$, on note $\mathbf{P}_t(f) : \mathbb{R}^n \rightarrow \mathbb{R}$ la fonction définie pour tout $x \in \mathbb{R}^n$ par

$$\mathbf{P}_t(f)(x) := (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(x+y) e^{-\frac{1}{2t}\|y-x\|_2^2} dx.$$

On définit $\mathbf{P}_0(f)(x)$ par $\mathbf{P}_0(f)(x) := f(x) = \lim_{t \rightarrow 0^+} \mathbf{P}_t(f)(x)$. On a donc

$$\mathbf{P}_0(f)(x) = \mathbf{E}_{\delta_x}(f) \quad \text{et} \quad \mathbf{P}_t(f)(x) = \mathbf{E}_{\mathcal{N}(x, t\mathbf{I}_n)}(f).$$

Ainsi la famille $(\mathbf{P}_s)_{s \in [0,1]}$ est une interpolation entre la mesure de Dirac $\delta_x = \mathbf{P}_0(\cdot)(x)$ et la mesure de Gauss $\mathcal{N}(x, t\mathbf{I}_n) = \mathbf{P}_t(\cdot)(x)$. Vue comme une famille d'opérateurs linéaires agissant sur les fonctions continues et bornées, la famille $(\mathbf{P}_t)_{t \geq 0}$ forme un semi-groupe car $\mathbf{P}_0 = \text{Id}$ et $\mathbf{P}_t \circ \mathbf{P}_s = \mathbf{P}_{t+s}$ pour tous $s, t \in \mathbb{R}_+$. Cette formule correspond ici à la convolution

$$\mathcal{N}(x, t\mathbf{I}_n) * \mathcal{N}(y, s\mathbf{I}_n) = \mathcal{N}(x+y, (t+s)\mathbf{I}_n).$$

La fonction $u : (x, t) \in \mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbf{P}_t(f)(x)$ est continue, de classe \mathcal{C}^∞ en x pour tout $t > 0$. Elle est solution de l'équation de la chaleur

$$\begin{cases} \partial_t u(t, x) = \Delta u(t, x) & \text{pour tout } x \in \mathbb{R}^n \text{ et } t > 0 \\ u(0, x) = f(x) & \text{pour tout } x \in \mathbb{R}^n \end{cases},$$

où Δ est l'opérateur différentiel de Laplace usuel défini par $\Delta := \sum_{i=1}^n \partial_{x_i}^2$. En particulier, $\partial_t \mathbf{P}_t(f)(x) = \Delta \mathbf{P}_t(f)(x)$ pour tout $x \in \mathbb{R}^n$ et tout $t > 0$. De plus, on a également $\Delta \mathbf{P}_t(f)(x) = \mathbf{P}_t(\Delta f)(x)$ lorsque la fonction f est lisse.

Démontrons à présent l'inégalité de Poincaré et l'inégalité de Sobolev logarithmique pour la mesure de Gauss $\mathcal{N}(x, t\mathbf{I}_n)$. Soit donc $\Phi : \mathcal{I} \subset \mathbb{R} \rightarrow \mathbb{R}$ la fonction convexe associée à la variance ou à l'entropie. Soit $f : \mathbb{R}^n \rightarrow \mathcal{I}$ une fonction lisse qui n'atteint pas les bornes de \mathcal{I} . On a

$$\mathbf{E}_{\mathcal{N}(x, t\mathbf{I}_n)}^\Phi(f) = \int_0^t \frac{d}{ds} \mathbf{P}_s(\Phi(\mathbf{P}_{t-s}(f))) ds,$$

où nous avons omis la dépendance en x du semi-groupe pour alléger les notations. Un simple calcul donne, en notant $g := \mathbf{P}_{t-s}(f)$,

$$\frac{d}{ds} \mathbf{P}_s(\Phi(\mathbf{P}_{t-s}(f))) = \mathbf{P}_s(\Delta \Phi(g) - \Phi'(g)\Delta g).$$

Or pour toute fonction lisse $h : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\Delta \Phi(h) = \Phi'(h)\Delta h + \Phi''(h)|\nabla h|^2.$$

D'autre part, la formulation intégrale du semi-groupe donne

$$|\nabla g| = |\nabla \mathbf{P}_{t-s}(f)| = |\mathbf{P}_{t-s}(\nabla f)| \leq \mathbf{P}_{t-s}(|\nabla f|).$$

Par conséquent, comme Φ'' est positive ou nulle, on obtient

$$\Delta\Phi(g) - \Phi'(g)\Delta g \leq \Phi''(\mathbf{P}_{t-s}(f))\mathbf{P}_{t-s}\left(|\nabla f|^2\right) = C(\mathbf{P}_{t-s}(f), \mathbf{P}_{t-s}(|\nabla f|)),$$

où C est la fonction convexe définie par (12). L'inégalité de Jensen pour la mesure de probabilité $\mathbf{P}_{t-s}(\cdot)(x)$ entraîne

$$C(\mathbf{P}_{t-s}(f), \mathbf{P}_{t-s}(|\nabla f|)) \leq \mathbf{P}_{t-s}(C(f, |\nabla f|)).$$

La formule de semi-groupe $\mathbf{P}_s \circ \mathbf{P}_{t-s} = \mathbf{P}_t$ conduit enfin à la majoration

$$\frac{d}{ds}\mathbf{P}_s(\Phi(\mathbf{P}_{t-s}(f))) \leq \mathbf{P}_t(C(f, |\nabla f|)).$$

Le membre de droite ne dépend pas de s , et il en découle l'inégalité souhaitée

$$\mathbf{E}_{\mathcal{N}(x,tI_n)}^\Phi(f) \leq t \mathbf{E}_{\mathcal{N}(x,tI_n)}(C(f, |\nabla f|)).$$

Pour la variance, $\Phi''(f) = 1$ et $C(f, |\nabla f|) = |\nabla f|^2$, tandis que pour l'entropie, $\Phi''(f) = 1/f$ et $C(f, |\nabla f|) = |\nabla f|^2/f$. Le processus de Markov associé au semi-groupe de la chaleur est le mouvement brownien.

8.2 Semi-groupe de Poisson simple

Pour toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ à support fini et tout $t \geq 0$, on note $\mathbf{P}_t(f) : \mathbb{R}^n \rightarrow \mathbb{R}$ la fonction définie pour tout $n \in \mathbb{N}$ par

$$\mathbf{P}_t(f)(n) := e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} f(n+k).$$

On a donc $\mathbf{P}_0(f)(n) = \mathbf{E}_{\delta_n}(f)$ et $\mathbf{P}_t(f)(n) = \mathbf{E}_{\delta_n * \mathcal{P}(t)}(f)$. Ainsi, la famille $(\mathbf{P}_s)_{s \in [0,1]}$ est une interpolation entre la mesure de Dirac $\delta_n = \mathbf{P}_0(\cdot)(n)$ et la mesure de Poisson translatée $\delta_n * \mathcal{P}(t) = \mathbf{P}_t(\cdot)(n)$. La fonction $(n, t) \in \mathbb{N} \times \mathbb{R}_+ \mapsto \mathbf{P}_t(f)(n)$ vérifie

$$\frac{d}{dt}\mathbf{P}_t(f)(n) = \mathbf{P}_t(f)(n+1) - \mathbf{P}_t(f)(n) = D\mathbf{P}_t(f)(n).$$

On a également la formule de commutation suivante :

$$D\mathbf{P}_t(f) = \mathbf{P}_t(Df).$$

Vue comme une famille d'opérateurs linéaires agissant sur les fonctions continues et bornées, la famille $(\mathbf{P}_t)_{t \geq 0}$ forme un semi-groupe car $\mathbf{P}_0 = \text{Id}$ et $\mathbf{P}_t \circ \mathbf{P}_s = \mathbf{P}_{t+s}$ pour tous $s, t \in \mathbb{R}_+$. Cette formule correspond ici à la convolution

$$\delta_n * \mathcal{P}(t) * \delta_m * \mathcal{P}(s) = \delta_{n+m} * \mathcal{P}(t+s).$$

Démontrons à présent l'inégalité de Poincaré et l'inégalité de Sobolev logarithmique pour la mesure de Poisson translaturée $\delta_n * \mathcal{P}(t)$. Soit donc $\Phi : \mathcal{I} \subset \mathbb{R} \rightarrow \mathbb{R}$ la fonction convexe associée à la variance ou à l'entropie. Pour toute fonction $f : \mathbb{N} \rightarrow \mathcal{I}$ à support fini, on a

$$\mathbf{E}_{\delta_n * \mathcal{P}(t)}^\Phi(f) = \int_0^t \frac{d}{ds} \mathbf{P}_s(\Phi(\mathbf{P}_{t-s}(f))) ds,$$

où nous avons omis la dépendance en n du semi-groupe pour alléger les notations. Un simple calcul donne, en notant $g := \mathbf{P}_{t-s}(f)$,

$$\frac{d}{ds} \mathbf{P}_s(\Phi(\mathbf{P}_{t-s}(f))) = \mathbf{P}_s(\mathbf{D}\Phi(g) - \Phi'(g)\mathbf{D}g).$$

Or si A est la fonction convexe définie par (10), on a pour toute fonction $h : \mathbb{N} \rightarrow \mathbb{R}$,

$$\mathbf{D}\Phi(h) - \Phi'(h)\mathbf{D}g = A(h, \mathbf{D}h).$$

Comme $\mathbf{D}g = \mathbf{D}\mathbf{P}_{t-s}(f) = \mathbf{P}_{t-s}(\mathbf{D}f)$, l'inégalité de Jensen pour la fonction convexe A et la mesure de probabilité $\mathbf{P}_{t-s}(\cdot)(n)$ entraîne que

$$\mathbf{D}\Phi(g) - \Phi'(g)\mathbf{D}g = A(\mathbf{P}_{t-s}(f), \mathbf{P}_{t-s}(\mathbf{D}f)) \leq \mathbf{P}_{t-s}(A(f, \mathbf{D}f)).$$

La formule de semi-groupe $\mathbf{P}_s \circ \mathbf{P}_{t-s} = \mathbf{P}_t$ conduit alors à la majoration

$$\frac{d}{ds} \mathbf{P}_s(\Phi(\mathbf{P}_{t-s}(f))) \leq \mathbf{P}_t(A(f, \mathbf{D}f)).$$

Le membre de droite ne dépend pas de s . Il en découle l'inégalité souhaitée

$$\mathbf{E}_{\delta_n * \mathcal{P}(t)}^\Phi(f) \leq t \mathbf{E}_{\delta_n * \mathcal{P}(t)}(A(f, \mathbf{D}f)).$$

Pour la variance, $A(f, \mathbf{D}f) = |\mathbf{D}f|^2$, tandis que pour l'entropie,

$$A(f, \mathbf{D}f) = (f + \mathbf{D}f)\mathbf{D} \log(f) + \mathbf{D}f.$$

Nous avons retrouvé l'inégalité la plus forte du théorème 7.1. Le processus de Markov associé au semi-groupe utilisé ici est le processus de Poisson simple.

8.3 Autres interpolations

Les mesures de Gauss et de Poisson sont infiniment divisibles, et cette nature se traduit sur les interpolations utilisées ci-dessus par le fait que les semi-groupes sont de convolution. Le lecteur familier avec la notion de générateur de Markov retiendra que la propriété de convolution ici correspond à une commutation aux translations pour le générateur de Markov. Cette commutation correspond sur le semi-groupe aux identités $\nabla \mathbf{P}_t(f) = \mathbf{P}_t(\nabla f)$ et $\mathbf{D}\mathbf{P}_t(f) = \mathbf{P}_t(\mathbf{D}f)$ respectivement. Les processus de Markov associés aux semi-groupes de convolution sont les processus de Lévy. En géométrie, la

commutation aux translations peut être vue comme une absence de courbure propre au processus.

Les mesures de Gauss et de Poisson sont également les mesures invariantes réversibles de semi-groupes de Markov ergodiques, qui ne sont pas de convolution. Ce type de semi-groupes permet des interpolations pour des classes de mesures de probabilité plus générales, incluant essentiellement les mesures log-concaves à courbure positive. Ces interpolations ont lieu sur l'intervalle de temps $[0, +\infty[$. La mesure de Dirac correspond au temps 0 et la mesure cible à la limite quand $t \rightarrow +\infty$ par ergodicité.

La mesure de Gauss $\mathcal{N}(0, I_n)$ est invariante et réversible pour le semi-groupe d'Ornstein-Uhlenbeck, défini pour toute $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continue et bornée par

$$\mathbf{P}_t(f)(x) := \mathbf{E}_{\mathcal{N}(e^{-t}x, (1-e^{-2t})I_n)}(f).$$

Il vérifie $\nabla \mathbf{P}_t(f) = e^{-t} \mathbf{P}_t(\nabla f)$. Il est associé à l'opérateur différentiel linéaire du second ordre $\sum_{i=1}^n \partial_{x_i}^2 - x_i \partial_{x_i}$. Ce semi-groupe conduit par interpolation aux inégalités de Poincaré et de Sobolev logarithmique optimales pour la mesure de Gauss. De manière similaire, la mesure de Poisson $\mathcal{P}(\lambda)$ est invariante et réversible pour le semi-groupe de la file d'attente M/M/ ∞ , défini pour toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ bornée par

$$\mathbf{P}_t(f)(n) := \mathbf{E}_{\mathcal{B}(n, e^{-t}) * \mathcal{P}(\lambda(1-e^{-t}))}(f).$$

Il vérifie $D \mathbf{P}_t(f) = e^{-t} \mathbf{P}_t(Df)$. Il est associé à l'opérateur aux différences $\lambda Df + n D^* f$ où D^* est comme en (39). Ce semi-groupe conduit, par interpolation, aux inégalités de Poincaré et de Sobolev logarithmiques modifiées optimales pour la mesure de Poisson.

La propriété d'ergodicité entraîne que ces semi-groupes convergent vers la moyenne spatiale pour la mesure invariante. Cette limite a donc une variance et une entropie nulle. De plus, la dérivation de ces fonctionnelles le long du semi-groupe fournit une forme de Dirichlet naturelle liée au générateur infinitésimal du semi-groupe. Il s'avère que l'inégalité de Poincaré est équivalente à une décroissance exponentielle de la variance le long du semi-groupe. De même, l'inégalité de Sobolev logarithmique est équivalente à une décroissance exponentielle de l'entropie le long du semi-groupe. La décroissance exponentielle de l'entropie le long du semi-groupe de la file d'attente M/M/ ∞ correspond à l'inégalité de Sobolev logarithmique modifiée associée à la fonctionnelle $f \mapsto Df D \log(f)$.

9 Stabilité par tensorisation, convolution, perturbation

Nous savons déjà que les inégalités de Poincaré et de Sobolev logarithmique possèdent une propriété de stabilité par translation et dilatation, comme le montre l'exemple des mesures de Bernoulli et de Gauss. Cette propriété est due aux invariances de l'énergie. D'autre part, nous avons utilisé à plusieurs reprises une stabilité par tensorisation de ces inégalités, dues à la convexité des fonctionnelles variance et entropie. Nous résumons ici, de manière formelle, trois types de stabilité importants dans la pratique. Nous les écrivons pour l'inégalité de Sobolev logarithmique, mais elles ont également lieu de la même manière pour l'inégalité de Poincaré.

9.1 Tensorisation

Soit $\mu_1 \otimes \cdots \otimes \mu_n$ une mesure de probabilité produit sur un espace mesurable produit $E_1 \times \cdots \times E_n$. Supposons que chacune des mesures μ_i vérifie une inégalité du type $\mathbf{Ent}_{\mu_i}(f^2) \leq c_i \mathbf{E}_{\mu_i}(Q_i(f))$ pour tout $f \in \mathcal{F}_i$. Alors, le théorème 5.1 entraîne que pour toute $f : E_1 \times \cdots \times E_n \rightarrow \mathbb{R}$ raisonnable,

$$\mathbf{Ent}_{\mu_1 \otimes \cdots \otimes \mu_n}(f^2) \leq \max(c_1, \dots, c_n) \mathbf{E}_{\mu_1 \otimes \cdots \otimes \mu_n}(Q_1(f) + \cdots + Q_n(f)).$$

On pense naturellement à l'exemple $\mu_i = \mathcal{N}(0, 1)$ et $Q_i(f) = (\partial_{x_i} f)^2$, pour lequel $c_1 = \cdots = c_n$, et $\mu_1 \otimes \cdots \otimes \mu_n = \mathcal{N}(0, I_n)$, et $Q_1(f) + \cdots + Q_n(f) = |\nabla f|^2$.

9.2 Convolution

Soit μ_1, \dots, μ_n des mesures de probabilité sur un espace mesurable E sur lequel la convolution de mesures a un sens. Supposons que chacune des mesures μ_i vérifie une inégalité du type $\mathbf{E}_{\mu_i}^\Phi(f) \leq c_i \mathbf{E}_{\mu_i}(Q(f))$ pour tout $f \in \mathcal{F}$, où \mathcal{F} est une classe de fonctions invariante par translations. La mesure $\mu_1 * \cdots * \mu_n$ est la mesure image de la mesure produit $\mu_1 \otimes \cdots \otimes \mu_n$ par l'application $x \in E^n \mapsto x_1 + \cdots + x_n \in E$. Par conséquent, la stabilité par tensorisation précédente entraîne que pour tout $f \in \mathcal{F}$,

$$\mathbf{Ent}_{\mu_1 * \cdots * \mu_n}(f^2) \leq (c_1 + \cdots + c_n) \mathbf{E}_{\mu_1 * \cdots * \mu_n}(Q(f)).$$

On pense naturellement aux semi-groupes de convolution, par exemple à l'exemple gaussien où $\mu_i = \mathcal{N}(x_i, t_i)$ sur $E_i = \mathbb{R}$, pour lequel $Q(f) = f'^2$, et $c_1 = \cdots = c_n$, et $\mu_1 * \cdots * \mu_n = \mathcal{N}(x_1 + \cdots + x_n, t_1 + \cdots + t_n)$. Un autre exemple est fourni par les mesures de Poisson.

9.3 Perturbation

Soit μ une mesure de probabilité sur un espace mesurable E . Soit $B : E \rightarrow \mathbb{R}$ une fonction bornée et soit μ_B la mesure de probabilité absolument continue par rapport à μ et de densité $Z_B^{-1} e^B$ où $Z_B := \mathbf{E}_\mu e^B$ est la constante de normalisation. Supposons que μ vérifie une inégalité du type $\mathbf{E}_\mu^\Phi(f) \leq c \mathbf{E}_\mu(Q(f))$ pour tout $f \in \mathcal{F}$. La formule (2) entraîne alors que pour tout $f \in \mathcal{F}$,

$$\mathbf{Ent}_{\mu_B}(f^2) \leq c e^{2\text{osc}(B)} \mathbf{E}_{\mu_B}(Q(f)).$$

Un exemple élémentaire est donné par $\mu = \mathcal{N}(0, 1)$, et $B = \cos$ sur $E = \mathbb{R}$. Cette technique de perturbation, souvent attribuée à Holley & Stroock, reste encore valable pour les inégalités de Sobolev logarithmiques modifiées.

10 Notes historiques et bibliographiques

Comme nous l'avons évoqué dans le résumé, le concept d'inégalité de Sobolev logarithmique a été introduit par Gross dans son célèbre article [Gro75], soumis en 1973 et

paru en 1975. Ce type d'inégalité est alors présenté comme une nouvelle formulation de la propriété d'hypercontractivité. Gross montre que la mesure de Bernoulli symétrique satisfait à une inégalité de Sobolev logarithmique, ce qui lui permet ensuite de traiter le cas de la mesure de Gauss en utilisant la propriété de tensorisation de l'hypercontractivité, retrouvant ainsi le théorème d'hypercontractivité de Nelson [Nel66] issu de la théorie quantique des champs. Ces idées étaient dans l'air du temps à cette époque, comme en témoignent les travaux apparentés de Federbush [Fed69] et de Faris [Far75]. On trouvera une présentation élémentaire de la méthode de Gross dans [ABC⁺00, Chap. 2].

Bonami avait déjà traité le cas de la mesure de Bernoulli symétrique sous sa forme hypercontractive dans un article [Bon71] paru en 1971. Le cas non symétrique ne fut élucidé que vingt-cinq ans plus tard par Higuchi et Yoshida dans [HY95]. Ce résultat fut obtenu indépendamment en 1996 par Diaconis et Saloff-Coste dans [DSC96]. La preuve que nous avons présentée est due à Bobkov et figure dans le cours donné à l'école d'été de Saint-Flour par Saloff-Coste en 1996 [SC97]. La preuve rapide du cas symétrique est due à Gross [Gro75, p. 1068]. Les preuves des inégalités modifiées présentées ici sont librement inspirées de [BL98], [Wu00], ainsi que de travaux récents de l'auteur. Les inégalités de Sobolev logarithmiques modifiées ont été étudiées par de nombreux auteurs. On pourra consulter par exemple [BL98], [Wu00], [AL00], et [Cha04]. L'étude des inégalités de Poincaré et de Sobolev logarithmiques pour les mesures invariantes des chaînes de Markov finies, qui incluent la mesure binomiale et de Bernoulli, a été menée essentiellement par Diaconis et Saloff-Coste. On pourra consulter par exemple à ce sujet [DSC96] et [SC97]. Certains cas particuliers font toujours l'objet d'articles dédiés, comme par exemple [Goe04] ou [GQ03].

La propriété de tensorisation de la variance et de l'entropie que nous avons énoncée est parfois attribuée à Bobkov. On en trouvera diverses formes dans [Lie75], [CT91], [LO00], [Led01a], [Ces01], [Cha04] et [BBLM05] par exemple. La généralisation des inégalités de Poincaré et de Sobolev logarithmiques a été tentée de diverses manières, et l'on pourra consulter à ce sujet [Bec89], [LO00], [Cha04], [BZ05], [Wan05], [BCR04], et [GGM05]. L'étude des inégalités de Poincaré et de Sobolev logarithmiques pour des mesures discrètes ou continues en dimension 1 peut être menée au moyen d'inégalités de Hardy et Muckenhoupt. Cette approche fournit des critères d'existence ainsi qu'un contrôle des constantes. On pourra consulter par exemple [BG99], [Mic99], [ABC⁺00, Chap. 6].

L'inégalité de Brascamp-Lieb a été publiée dans [BL76], et peut être déduite d'une représentation de Helffer-Sjöstrand de la covariance, en liaison avec une propriété du laplacien de Witten sur les formes différentielles, cf. [Hel02, HN05]. L'inégalité de Sobolev logarithmique pour la mesure de Gauss a été démontrée par de nombreuses méthodes. La preuve de Gross [Gro75] utilise la notion d'hypercontractivité, celle de Adams et Clarke [AC79] est fondée sur un argument variationnel, celle de Neveu [Nev76] fait appel au mouvement brownien, tandis que celle de Beckner [Bec99, Bec92] exploite l'inégalité de Sobolev. Citons également celle de Rothaus [Rot78], qui fait intervenir un problème de Sturm-Liouville en dimension 1. Un progrès important est dû à Bakry et

Émery [BE84, BE85]. Leur méthode, très générale, revient en quelque sorte à une comparaison à la mesure de Gauss par convexité. Elle reste valable pour une large classe de mesures liées à des diffusions sur les variétés riemanniennes, et inclue les mesures de probabilité à densité log-concave sur \mathbb{R}^n de courbure positive. On pourra consulter par exemple [Bak91a, Bak91b, Bak94], [Led00], et [ABC⁺00, Chap. 5] à ce sujet. Le théorème de Caffarelli sera établi plus de quinze ans plus tard dans [Caf00, Caf02], en liaison avec l'équation de Monge-Ampère. Dans [BL00], l'inégalité de Sobolev logarithmique ainsi que l'inégalité de Brascamp-Lieb sont déduites de l'inégalité de Brunn-Minkowski via l'inégalité de Prékopa-Leindler. Dans [OV00], Otto et Villani établissent une inégalité plus générale reliant l'entropie, la distance de Wasserstein, et l'information de Fisher. Cette inégalité est reliée dans [OV01, BGL02] à l'hypercontractivité des solutions d'équations d'Hamilton-Jacobi. Dans la même veine, on trouvera enfin dans [CE02] une preuve de l'inégalité de Sobolev logarithmique pour la mesure de Gauss basée sur le théorème de Brenier et l'équation de Monge-Ampère pour le transport optimal de la mesure. La démonstration élémentaire présentée ici est tirée de l'article fondateur de Gross [Gro75, p. 1072], voir également [Gro93, p. 60]. Les généralisations à des mesures sur des espaces de différentes natures ont été menées par de nombreux auteurs. On pourra consulter par exemple à ce sujet [Hsu99], [DM05], ainsi que les travaux de Fengyu Wang.

La version euclidienne de l'inégalité de Sobolev logarithmique optimale gaussienne, qui concerne la mesure de Lebesgue, remonte à la thèse de Stam dans les années 1950, en théorie de l'information, cf. [Sta59]. On en trouvera une version L^p optimale dans [Gen03]. Cette formulation euclidienne est liée aux travaux importants de Beckner [Bec95, Bec75b, Bec75a] et de Carlen [Car91] portant en particulier sur l'analyse de Fourier, l'inégalité de Young optimale, et les principes d'incertitude. Ces aspects sont abordés par exemple dans [DCT91, CT91] et [ABC⁺00, Chap. 10].

Les inégalités de Poincaré et de Sobolev logarithmiques ont été établies pour de nombreuses mesures complexes issues de la mécanique statistique, suite notamment aux travaux fondateurs de Stroock, Zegarliński, et Holley d'une part, et de Lu et Yau d'autre part. On pourra consulter par exemple à ce sujet les livres [Roy99], [Hel02], et les cours [Mar99] et [GZ03]. Le livre [DS89] donne quelques utilisations des inégalités de Sobolev logarithmiques en théorie des grandes déviations. Les multiples liens entre inégalités fonctionnelles et phénomène de concentration de la mesure sont abordés par exemple dans [ABC⁺00, Chap. 7] et dans le livre [Led01b], en liaison notamment avec l'isopérimétrie et la théorie du transport optimal de la mesure. On lira avec profit les fameux travaux de Barthe à ce sujet. La concentration de la mesure obtenue par ces approches est au cœur de méthodes de sélection de modèles en statistique adaptative. On pourra par exemple consulter le cours de Saint-Flour de Massart donné en 2003.

La monotonie de fonctionnelles convexes le long de processus d'évolution est un principe très général, dont la première apparition remonte sans doute au fameux théorème **H** de Boltzmann en théorie cinétique des gaz. Les fonctionnelles convexes associées aux solutions d'équations aux dérivées partielles ont été étudiées en liaison avec la théorie du transport optimal de la mesure. On pourra consulter à ce sujet les livres [Vil03] et

[ACB⁺03]. Le résultat de Caffarelli évoqué plus haut fait partie de cette nouvelle vague de développements. Des travaux récents – de Villani et Lott d’une part, et de Sturm et von Renesse d’autre part – montrent que la minoration de la courbure des diffusions sur les variétés riemanniennes, à la base de l’approche de Bakry-Émery, peut être reliée à une propriété de convexité de l’entropie sur l’espace des chemins. On pourra consulter à ce sujet [vRS05] ainsi que le cours donné par Villani à Saint-Flour en 2005.

Références

- [ABC⁺00] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO et G. SCHEFFER – *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux.
- [AC79] R. A. ADAMS et F. H. CLARKE – « Gross’s logarithmic Sobolev inequality : a simple proof », *Amer. J. Math.* **101** (1979), no. 6, p. 1265–1269.
- [ACB⁺03] L. AMBROSIO, L. A. CAFFARELLI, Y. BRENIER, G. BUTTAZZO et C. VILLANI – *Optimal transportation and applications*, Lecture Notes in Mathematics, vol. 1813, Springer-Verlag, Berlin, 2003, Lectures from the C.I.M.E. Summer School held in Martina Franca, September 2–8, 2001, Edited by Caffarelli and S. Salsa.
- [AL00] C. ANÉ et M. LEDOUX – « On logarithmic Sobolev inequalities for continuous time random walks on graphs. », *Probab. Theor. Relat. Fields* **116** (2000), no. 4, p. 573–602.
- [Bak91a] D. BAKRY – « Inégalités de Sobolev faibles : un critère Γ_2 », Séminaire de Probabilités, XXV, Lecture Notes in Math., vol. 1485, Springer, Berlin, 1991, p. 234–261.
- [Bak91b] — , « Weak Sobolev inequalities », Stochastic analysis and applications (Lisbon, 1989), Birkhäuser Boston, Boston, MA, 1991, p. 63–81.
- [Bak94] — , « L’hypercontractivité et son utilisation en théorie des semigroupes », Lectures on probability theory. École d’été de probabilités de St-Flour 1992, Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, p. 1–114.
- [BBLM05] S. BOUCHERON, O. BOUSQUET, G. LUGOSI et P. MASSART – « Moment inequalities for functions of independent random variables », *Ann. Probab.* **33** (2005), no. 2, p. 514–560.
- [BCR04] F. BARTHE, P. CATTIAUX et C. ROBERTO – « Interpolated inequalities between exponential and Gaussian. Orlicz hypercontractivity and application to isoperimetry », prépublication, 2004.
- [BE84] D. BAKRY et M. EMERY – « Hypercontractivité de semi-groupes de diffusion », *C. R. Acad. Sci. Paris Sér. I Math.* **299** (1984), no. 15, p. 775–778.

- [BE85] D. BAKRY et M. EMERY – « Diffusions hypercontractives », Séminaire de probabilités, XIX, 1983/84, Springer, Berlin, 1985, p. 177–206.
- [Bec75a] W. BECKNER – « Inequalities in Fourier analysis », *Ann. of Math. (2)* **102** (1975), no. 1, p. 159–182.
- [Bec75b] — , « Inequalities in Fourier analysis on R^n », *Proc. Nat. Acad. Sci. U.S.A.* **72** (1975), p. 638–641.
- [Bec89] — , « A generalized Poincaré inequality for Gaussian measures », *Proc. Amer. Math. Soc.* **105** (1989), no. 2, p. 397–400.
- [Bec92] — , « Sobolev inequalities, the Poisson semigroup, and analysis on the sphere S^n », *Proc. Nat. Acad. Sci. U.S.A.* **89** (1992), no. 11, p. 4816–4819.
- [Bec95] — , « Pitt’s inequality and the uncertainty principle », *Proc. Amer. Math. Soc.* **123** (1995), no. 6, p. 1897–1905.
- [Bec99] — , « Geometric asymptotics and the logarithmic Sobolev inequality », *Forum Math.* **11** (1999), no. 1, p. 105–137.
- [BG99] S. G. BOBKOV et F. GÖTZE – « Exponential integrability and transportation cost related to logarithmic Sobolev inequalities », *J. Funct. Anal.* **163** (1999), no. 1, p. 1–28.
- [BGL02] S. G. BOBKOV, I. GENTIL et M. LEDOUX – « Hypercontractivity of Hamilton-Jacobi equations », *Markov Process. Related Fields* **8** (2002), no. 2, p. 233–235, Inhomogeneous random systems (Cergy-Pontoise, 2001).
- [BL76] H. J. BRASCAMP et E. H. LIEB – « On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation », *J. Funct. Anal.* **22** (1976), no. 4, p. 366–389.
- [BL98] S. BOBKOV et M. LEDOUX – « On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. », *J. Funct. Anal.* **156** (1998), no. 2, p. 347–365 (English).
- [BL00] S. G. BOBKOV et M. LEDOUX – « From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities », *Geom. Funct. Anal.* **10** (2000), no. 5, p. 1028–1052.
- [Bon71] A. BONAMI – « Étude des coefficients de Fourier des fonctions de $L^p(G)$ », *Ann. Inst. Fourier (Grenoble)* **20** (1971), no. 2, p. 335–402.
- [BZ05] S. G. BOBKOV et B. ZEGARLINSKI – « Entropy bounds and isoperimetry », *Mem. Amer. Math. Soc.* **176** (2005), no. 829, p. x+69.
- [Caf00] L. A. CAFFARELLI – « Monotonicity properties of optimal transportation and the FKG and related inequalities », *Comm. Math. Phys.* **214** (2000), no. 3, p. 547–563.
- [Caf02] — , « Erratum : “Monotonicity of optimal transportation and the FKG and related inequalities” [Comm. Math. Phys. **214** (2000), no. 3, 547–563 ; MR1800860 (2002c :60029)] », *Comm. Math. Phys.* **225** (2002), no. 2, p. 449–450.

- [Car91] E. A. CARLEN – « Super-additivity of Fisher’s information and logarithmic Sobolev inequalities », *J. Funct. Anal.* **101** (1991), no. 1, p. 194–211.
- [CE02] D. CORDERO-ERAUSQUIN – « Some applications of mass transport to Gaussian-type inequalities », *Arch. Ration. Mech. Anal.* **161** (2002), no. 3, p. 257–269.
- [Ces01] F. CESI – « Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields », *Probab. Theory Related Fields* **120** (2001), no. 4, p. 569–584.
- [Cha04] D. CHAFAÏ – « Entropies, convexity, and functional inequalities : on Φ -entropies and Φ -Sobolev inequalities », *J. Math. Kyoto Univ.* **44** (2004), no. 2, p. 325–363.
- [CT91] T. M. COVER et J. A. THOMAS – *Elements of information theory*, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication.
- [DCT91] A. DEMBO, T. M. COVER et J. A. THOMAS – « Information-theoretic inequalities », *IEEE Trans. Inform. Theory* **37** (1991), no. 6, p. 1501–1518.
- [DM05] B. K. DRIVER et T. MELCHER – « Hypocoelliptic heat kernel inequalities on the Heisenberg group », *J. Funct. Anal.* **221** (2005), no. 2, p. 340–365.
- [DS89] J.-D. DEUSCHEL et D. W. STROOCK – *Large deviations*, Academic Press Inc., Boston, MA, 1989.
- [DSC96] P. DIACONIS et L. SALOFF-COSTE – « Logarithmic Sobolev inequalities for finite Markov chains », *Ann. Appl. Probab.* **6** (1996), no. 3, p. 695–750.
- [Far75] W. G. FARIS – « Product spaces and Nelson’s inequality », *Helv. Phys. Acta* **48** (1975), no. 5/6, p. 721–730 (1976).
- [Fed69] P. FEDERBUSH – « A partially alternate derivation of a result of Nelson », *J. Math. Phys.* **10** (1969), p. 50–52.
- [Gen03] I. GENTIL – « The general optimal L^p -Euclidean logarithmic Sobolev inequality by Hamilton-Jacobi equations », *J. Funct. Anal.* **202** (2003), no. 2, p. 591–599.
- [GGM05] I. GENTIL, A. GUILLIN et L. MICLO – « Modified logarithmic Sobolev inequalities and transportation inequalities », à paraître dans *Probability Theory and Related Fields*, 2005.
- [Goe04] S. GOEL – « Modified logarithmic Sobolev inequalities for some models of random walk », *Stochastic Process. Appl.* **114** (2004), no. 1, p. 51–79.
- [GQ03] F. GAO et J. QUASTEL – « Exponential decay of entropy in the random transposition and Bernoulli-Laplace models », *Ann. Appl. Probab.* **13** (2003), no. 4, p. 1591–1600.
- [Gro75] L. GROSS – « Logarithmic Sobolev inequalities », *Amer. J. Math.* **97** (1975), no. 4, p. 1061–1083.
- [Gro93] — , « Logarithmic Sobolev inequalities and contractivity properties of semi-groups », *Dirichlet forms (Varenna, 1992)*, Springer, Berlin, 1993, p. 54–88.

- [GZ03] A. GUIONNET et B. ZEGARLINSKI – « Lectures on logarithmic Sobolev inequalities », Séminaire de Probabilités, XXXVI, Lecture Notes in Math., vol. 1801, Springer, Berlin, 2003, p. 1–134.
- [Hel02] B. HELFFER – *Semiclassical analysis, Witten Laplacians, and statistical mechanics*, Series on Partial Differential Equations and Applications, vol. 1, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.
- [HN05] B. HELFFER et F. NIER – *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, Lecture Notes in Mathematics, vol. 1862, Springer-Verlag, Berlin, 2005.
- [Hsu99] E. P. HSU – « Analysis on path and loop spaces », Probability theory and applications (Princeton, NJ, 1996), Amer. Math. Soc., Providence, RI, 1999, p. 277–347.
- [HY95] Y. HIGUCHI et N. YOSHIDA – « Analytic conditions and phase transition for Ising models », notes en japonais, non publiées, 1995.
- [Led00] M. LEDOUX – « The geometry of Markov diffusion generators », *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000), no. 2, p. 305–366, Probability theory.
- [Led01a] M. LEDOUX – « Logarithmic Sobolev inequalities for unbounded spin systems revisited », Séminaire de Probabilités, XXXV, Lecture Notes in Math., vol. 1755, Springer, Berlin, 2001, p. 167–194.
- [Led01b] M. LEDOUX – *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.
- [Lie75] E. H. LIEB – « Some convexity and subadditivity properties of entropy », *Bull. Amer. Math. Soc.* **81** (1975), p. 1–13.
- [LO00] R. LATAŁA et K. OLESZKIEWICZ – « Between Sobolev and Poincaré », Geometric aspects of functional analysis, Springer, Berlin, 2000, p. 147–168.
- [Mar99] F. MARTINELLI – « Lectures on Glauber dynamics for discrete spin models », Lectures on probability theory and statistics (Saint-Flour, 1997), Lecture Notes in Math., vol. 1717, Springer, Berlin, 1999, p. 93–191.
- [Mic99] L. MICLO – « An example of application of discrete Hardy’s inequalities », *Markov Process. Related Fields* **5** (1999), no. 3, p. 319–330.
- [Nel66] E. NELSON – « A quartic interaction in two dimensions », Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965), M.I.T. Press, Cambridge, Mass., 1966, p. 69–73.
- [Nev76] J. NEVEU – « Sur l’espérance conditionnelle par rapport à un mouvement brownien », *Ann. Inst. H. Poincaré Sect. B (N.S.)* **12** (1976), no. 2, p. 105–109.
- [OV00] F. OTTO et C. VILLANI – « Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality », *J. Funct. Anal.* **173** (2000), no. 2, p. 361–400.

- [OV01] F. OTTO et C. VILLANI – « Comment on : “Hypercontractivity of Hamilton-Jacobi equations” [J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696 ; MR1846020 (2003b :47073)] by S. G. Bobkov, I. Gentil and M. Ledoux », *J. Math. Pures Appl. (9)* **80** (2001), no. 7, p. 697–700.
- [Rot78] O. S. ROTHHAUS – « Lower bounds for eigenvalues of regular Sturm-Liouville operators and the logarithmic Sobolev inequality », *Duke Math. J.* **45** (1978), no. 2, p. 351–362.
- [Roy99] G. ROYER – *Une initiation aux inégalités de Sobolev logarithmiques*, Société Mathématique de France, Paris, 1999.
- [SC97] L. SALOFF-COSTE – « Lectures on finite Markov chains », Lectures on probability theory and statistics. École d’été de probabilités de St-Flour 1996, Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, p. 301–413.
- [Sta59] A. J. STAM – « Some inequalities satisfied by the quantities of information of Fisher and Shannon », *Information and Control* **2** (1959), p. 101–112.
- [Vil03] C. VILLANI – *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
- [vRS05] M.-K. VON RENESSE et K.-T. STURM – « Transport inequalities, gradient estimates, entropy, and Ricci curvature », *Comm. Pure Appl. Math.* **58** (2005), no. 7, p. 923–940.
- [Wan05] F.-Y. WANG – « A generalization of Poincaré and log-Sobolev inequalities », *Potential Anal.* **22** (2005), no. 1, p. 1–15.
- [Wu00] L. WU – « A new modified logarithmic Sobolev inequality for Poisson point processes and several applications », *Probab. Theor. Relat. Fields* **118** (2000), no. 3, p. 427–438.