Sur certaines mesures produit conditionnées

Djalil CHAFAÏ

`http://www.lsp.ups-tlse.fr/Chafai/`

Notes pour l'exposé du lundi 14 novemvre 2005 au groupe de travail
Probabilités et Modélisation Aléatoire de l'Institut de Mathématiques de Toulouse

Pourquoi cet exposé ? Il se trouve que j'ai travaillé dur sur le sujet pendant l'année 2001, comme en témoigne maigrement [Cha03], dans le sillage de [LPY02], parallèlement à [Cap03]. Depuis, je n'ai plus vraiment regardé ces questions, d'abord par lassitude, puis par manque de temps : je trouve le problème difficile. Cependant, la prépublication récente [GORV05] pourrait intéresser des membres du laboratoire, et un exposé sur le thème semble être une bonne idée.

Table des matières

Notations. On note $x \cdot y$ le produit scalaire des vecteurs x et y de $\mathbb{R}^d$, et $|x| := \sqrt{x \cdot x}$ la norme euclidienne de x. Pour tout réel $p > 0$, on désigne par $\mathrm{L}^p(\mu)$ l'espace de Lebesgue des fonctions mesurables f telles que $|f|^p$ est μ-intégrable. Pour toute mesure de probabilité μ, la *moyenne* d'une fonction f dans $\in \mathrm{L}^1(\mu)$, ou alternativement d'une fonction f mesurable à valeurs positives ou nulles, est définie par $\mathbf{E}_\mu(f) := \mathbf{E}_\mu f := \int f\,d\mu$. De même, la *variance* de f pour μ est définie par $\mathbf{Var}_\mu(f) := \mathbf{E}_\mu((f - \mathbf{E}_\mu f)^2)$. Si μ et ν sont deux mesures de probabilité sur le même espace mesurable, la *divergence de Kullback-Leibler* de ν par rapport à μ est définie par

$$\mathbf{Ent}(\nu \,|\, \mu) := \begin{cases} \mathbf{E}_\mu\Big(\frac{d\nu}{d\mu} \log\Big(\frac{d\nu}{d\mu}\Big)\Big) = \mathbf{E}_\nu\Big(\log\Big(\frac{d\nu}{d\mu}\Big)\Big) & \text{si } \nu \ll \mu \\ +\infty & \text{sinon} \end{cases}.$$

La notation **Ent** provient de la dénomination alternative *entropie relative*. Si $\nu \ll \mu$, alors $\mathbf{Ent}(\nu \,|\, \mu) < \infty$ si et seulement si $\log(d\nu/d\mu) \in \mathrm{L}^1(\nu)$. La convexité stricte de $u \in \mathbb{R}_+ \mapsto u \log(u)$ et l'inégalité de Jensen entraînent que $\mathbf{Ent}(\nu \,|\, \mu) \geqslant 0$, avec égalité si et seulement si $\mu = \nu$. Si μ et ν sont toutes deux absolument continues par rapport à une mesure de Borel de référence, avec pour densités respectives f_μ et f_ν, alors on note parfois $\mathbf{Ent}(f_\nu \,|\, f_\mu)$ la divergence de Kullback-Leibler $\mathbf{Ent}(\nu \,|\, \mu)$. La fonctionnelle $\mathbf{Ent}(\cdot \,|\, \mu)$ est la fonction de taux du théorème de grandes déviations de Sanov pour la mesure empirique associée à la loi des grands nombres, cf. [DZ98].

1 Excursion culturelle préliminaire

L'objectif ultime de cet exposé est d'aborder des inégalités fonctionnelles pour des mesures produit conditionnées qui apparaissent en mécanique statistique. La mécanique statistique fait partie de la physique statistique. Le terme « statistique » est trompeur ici car il n'y a pas d'estimateur ni d'échantillon. Ce qualificatif date d'une époque, le dix-neuvième siècle, où tout ce qui était aléatoire était qualifié de statistique, car la théorie moderne des probabilités n'était pas encore née. Aujourd'hui, on parlerait sans doute de « physique stochastique » et de « mécanique stochastique ». La physique statistique est l'une des branches les plus actives de la physique d'aujourd'hui. Nous avons tous entendu parler par exemple de condensats de Bose-Einstein. Savez-vous que 2005 est « l'année de la physique » ? Je vous renvoie aux multiples publications consacrées à cet évenement, comme par exemple [SPM05]. Parmis les pères fondateurs de la physique statistique, on trouve les grands physiciens Maxwell (1831-1879), Gibbs (1839-1903), et Boltzmann (1844-1906).

1.1 L'entropie selon Boltzmann

L'ambition de Boltzmann est d'établir le second principe de la thermondynamique de Carnot et Clausius : l'entropie d'un système isolé augmente, cf. [Car24]. Ce faisant, il construit une « théorie cinétique des gaz », dans la continuité de certains travaux de Maxwell. Sur l'entropie et ses liens avec la thermodynamique, la théorie de l'information, et la théorie ergodique, on lira avec profit le joli petit livre [Zin96], ainsi que certains chapitres du livre [CT91].

1.1.1 Approche combinatoire et mesure du désordre

Soit Σ_r un système « macroscopique » constitué de r particules « microscopiques » *indiscernables* pouvant être chacune dans l'un des n états possibles. L'état macroscopique du système est donnée par le nombre de particules dans chaque état, autrement dit, par le vecteur $(r_1, \ldots, r_n)$ où r_i est le nombre de particules dans l'état i. Pour un état macroscopique donné, le nombre de « degrés de liberté » du système Σ_r est donné naturellement par le nombre d'états microscopiques compatibles avec l'état macroscopique spécifié. Malheureusement, cette grandeur n'est pas *extensive* car le nombre de degrès de liberté de la juxtaposition de deux systèmes est le produit des degrès de liberté de chacun des deux systèmes, et pas leur somme. Il est donc plus commode de considérer le logarithme du nombre de degrés de liberté. Cette quantité va jouer le rôle d'*entropie*[1]. Ainsi, par définition, l'entropie $\mathbf{S}(\Sigma_r)$ du système Σ_r est le logarithme du nombre d'états microscopiques compatibles avec l'état macroscopique donné. Or il y a exactement

$$C_r^{r_1,\ldots,r_n} := \frac{r!}{r_1! \cdots r_n!}$$

états microscopiques possibles pour le système Σ_r lorsque l'état macroscopique $(r_1, \ldots, r_n)$ est fixé. En vertu de de qui précède, l'entropie moyenne du système par particule est alors donnée par

$$\mathbf{S}_{\mathrm{moy}}(\Sigma_r) := \frac{1}{r} \log C_r^{r_1,\ldots,r_n}.$$

[1]On retiendra la formule entropie = constante × log(degrés de libertés énergétiques). Cette formule, écrite sous la forme $\mathbf{S} = \kappa \log(\mathcal{W})$, est gravée sur la tombe de Boltzmann à Vienne.

Le vecteur $r^{-1}(r_1,\dots,r_n)$ est une loi de probabilité discrète puisque par définition $r = r_1 + \cdots + r_n$. Supposons que les fréquences $r^{-1}(r_1,\dots,r_n)$ convergent vers la loi de probabilité $(p_1,\dots,p_n)$ lorsque le nombre de particules r tend vers l'infini. La formule de Striling indique alors que l'entropie moyenne par particule pour le système infini Σ_∞ est

$$\mathbf{S}_{\mathrm{moy}}(\Sigma_\infty) := \lim_{r\to+\infty} \frac{1}{r} \log C_r^{r_1,\dots,r_n} = -\sum_{i=1}^{n} p_i \log(p_i),$$

Ce raisonnement combinatoire sera également utilisé par Shannon vers 1948 lorsqu'il forgea la théorie de l'information, cf. [ABC$^+$00, Chap. 10] ou [CT91]. La fonctionnelle de Boltzmann $(p_1,\dots,p_n) \mapsto \mathbf{S}(p_1,\dots,p_n) := -\sum_{i=1}^{n} p_i \log(p_i)$ est définie sur le simplexe des lois de probabilité sur $\{1,\dots,n\}$. La fonctionnelle $\mathbf{S}$ est l'asymptotique du nombre moyen d'états microscopiques compatibles avec un état macroscopique. Elle possède de très belles caractérisations par trois ou quatre axiomes, dont celui de l'extensivité, cf. [CT91]. Chaque élément du simplexe est une loi sur l'ensemble $\{1,\dots,n\}$ des états d'énergie. Lorsque l'ensemble des états d'énergie est[2] $\mathbb{R}^d$, la fonctionnelle naturelle à considérer est donnée par

$$f \mapsto \mathbf{S}(f) := -\int_{\mathbb{R}^d} f(u)\log(f(u))\,du, \tag{1}$$

définie pour toute densité de probabilité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. L'extensivité correspond à la propriété $\mathbf{S}(f\otimes g) = \mathbf{S}(f) + \mathbf{S}(g)$, où f et g sont des densités sur $\mathbb{R}^d$ et $\mathbb{R}^{d'}$ respectivement, et où $f\otimes g$ est la densité produit sur l'espace produit $\mathbb{R}^d \otimes \mathbb{R}^{d'} = \mathbb{R}^{d+d'}$.

Exercice 1.1. Calculer l'entropie de la mesure de Gauss $\mathcal{N}(m,\Sigma)$, de moyenne m et de matrice de covariance non-singulière Σ. Dépend-elle de m ? Que se passe-t-il lorsque Σ est singulière ? Montrer que l'entropie $\mathbf{S}$ prend ses valeurs dans tout $\mathbb{R}$. Est-ce le cas pour l'entropie discrète définie sur les lois atomiques ? On traitera séparément le cas à support fini et le cas à support infini dénombrable.

1.1.2 Le fameux théorème-H

Considérons à présent un gaz dilué constitué d'un grand nombre de molécules qui bougent et s'entre-choquent. Il est impossible de connaitre la position et la vitesse de chacune d'entre elle, et quand bien même cela était possible, l'écriture des lois de Newton conduirait alors à un système gigantesque impossible à étudier. Boltzman adopte alors une approche macroscopique « statistique ». Il considère

[2] Penser à l'énergie cinétique d'une partitule de vitesse v par exemple, ou encore à l'énergie potentielle dans un champ de force.

la densité $f_t(x,v)$ sur $\mathbb{R}^6$ des couples position-vitesse (x,v) des molécules du gaz au temps t. Moyennant certaines hypothèses sur le gaz, il propose ensuite une équation aux dérivées partielles de la forme

$$\partial_t f_t(x,v) + v \cdot \nabla_x f_t(x,v) = Q(f_t,f_t)(x,v)$$

qui exprime l'évolution de cette densité au cours du temps. Ici, Q est typiquement une forme quadratique intégrale, appelée *noyau de collision*, qui tient compte des chocs élastiques, cf. [EE90, Vil02, Cer02] et [Bol]. Supposons pour simplifier que f_t ne dépende pas de x, ce qui a pour effet de faire disparaître le terme de transport $v \cdot \nabla_x f_t$. Le célèbre théorème-**H** de Boltzmann affirme que la fonction

$$t \mapsto \mathbf{H}(f_t) := \int_{\mathbb{R}^3} f_t(v)\log(f_t(v))\,dv$$

est décroissante. Autrement dit, la fonction $t \mapsto \mathbf{S}(f_t) = -\mathbf{H}(f_t)$ est croissante. Ce faisant, Boltzmann donne une interprétation « statistique » du second principe de la thermodynamique de Clausius et Carnot qui affirme que l'entropie d'un système isolé augmente. La fonctionnelle **S** joue le rôle d'*entropie*. La croissance de **S** (i.e. la croissance de **H**) s'écrit également

$$\partial_t \mathbf{H}(f_t) = \int_{\mathbb{R}^3} Q(f_t,f_t)(1+\log(f_t))\,dv \leqslant 0.$$

La monotonie de **S** le long du processus d'évolution suggère que l'état d'équilibre du système correspond à la densité de probabilité f qui maximise **S** sous les contraintes de conservation d'énergie (le système est isolé). Il se trouve que la densité f sur $\mathbb{R}^3$ qui maximise la fonctionnelle **S** sous contrainte de variance (énergie cinétique moyenne fixée) correspond à une loi gaussienne[3].

1.2 Maximum d'entropie et minimum d'énergie libre de Helmholtz

Soit $\mathcal{F}$ l'ensemble des densités de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Considérons une fonction lisse $W : \mathbb{R}^d \to \mathbb{R}$ vérifiant $\int_{\mathbb{R}^d} f_W(u)\,du = 1$, où $f_W := e^{-W}$. Si $W(u)$ est interprété comme l'énergie de la configuration $u \in \mathbb{R}^d$, alors la densité f_W favorise les états de moindre énergie (DESSIN). La mesure de probabilité sur $\mathbb{R}^d$ de densité f_W par rapport à la mesure de Lebesgue est appelée *mesure de Boltzmann-Gibbs* associée à la *fonction d'énergie* W. Lorsque W est une forme quadratique, f_W est une densité gaussienne. La densité f_W est invariante par translation sur W, en ce sens que $f_{w+W} = f_W$ pour toute constante

[3]On parle de distribution de Maxwell en physique dans ce contexte.

w. Cette propriété est tout à fait naturelle lorsque W contient une énergie potentielle. On qualifie parfois W de *fonction potentiel* ou *hamiltonien*[4]. Soit à présent $\mathcal{F}_W$ l'ensemble des $f \in \mathcal{F}$ vérifiant

$$\int_{\mathbb{R}^d} W(u)f(u)\,du = \int_{\mathbb{R}^d} W(u)f_W(u)\,du.$$

En d'autres termes, les éléments de $\mathcal{F}_W$ ont la même énergie moyenne que f_W. Comme $\log(f_W) = -W$, on a pour tout $f \in \mathcal{F}_W$,

$$\mathbf{S}(f_W) = \int_{\mathbb{R}^d} W f_W\,du = \int_{\mathbb{R}^d} W f\,du = -\int_{\mathbb{R}^d} \log(f_W) f\,du.$$

Par conséquent,

$$\mathbf{S}(f_W) - \mathbf{S}(f) = \int_{\mathbb{R}^d} \frac{f}{f_W} \log\left(\frac{f}{f_W}\right) f_W\,du = \mathbf{Ent}(f \mid f_W)\,.$$

On a donc $\mathbf{S}(f_W) - \mathbf{S}(f) \geqslant 0$, avec égalité si et seulement si $f = f_W$ presque-partout. Ainsi, l'entropie $\mathbf{S}$ est maximale en f_W sur l'ensemble des densités de même énergie moyenne.

Soit à présent $f \in \mathcal{F}$, sans contrainte d'énergie moyenne. L'*énergie libre de Helmholtz* $\mathbf{F}(f)$ de f désigne la quantité

$$\int_{\mathbb{R}^d} W(u)f(u)\,du - \mathbf{S}(f).$$

L'énergie libre $\mathbf{F}(f)$ de f est donc égale à son énergie moyenne $\int_{\mathbb{R}^d} W(u)f(u)\,du$ diminuée de son entropie $\mathbf{S}(f)$. Comme $W = -\log(f_W)$, on a $\mathbf{F}(f_W) = 0$, et

$$\mathbf{F}(f) = \int_{\mathbb{R}^d} \frac{f}{f_W} \log\left(\frac{f}{f_W}\right) f_W\,du = \mathbf{Ent}(f \mid f_W)\,.$$

Pour un mathématicien, l'énergie libre est la divergence de Kullback-Leibler de f par rapport à la densité de probabilité de Gibbs-Boltzmann f_W. Ainsi, la fonctionnelle énergie libre $\mathbf{F}$ est minimisée par la densité de Gibbs-Boltzmann et par elle seulement (à un ensemble de mesure de Lebesgue nulle près). Enfin, $\mathbf{F}(f) - \mathbf{F}(f_W) = \mathbf{S}(f_W) - \mathbf{S}(f)$ lorsque $f \in \mathcal{F}_W$.

[4]En mécanique, l'hamiltonien d'un système est la fonctionnelle qui exprime son énergie en fonction de ses variables position et impulsion, cf. [GJ87]. L'ensemble des couples positions-impulsion forment l'espace des phases. Considérons par exemple une particule de masse m, sujette à un champ de force issu d'un potentiel de position W (gravitation, électrostatique,...). Son hamiltonien est donné par $H(u,v) = W(u) + mv^2/2$, où u est la position de la particule et v sa vitesse. La quantité $W(u)$ représente l'énergie potentielle de la particule tandis que la quantité $mv^2/2$ est son énergie cinétique.

1.2.1 Introduction de la tempétature

Considérons une fonction lisse $W : \mathbb{R}^d \to \mathbb{R}$ telle que pour tout $\beta \in \mathbb{R}_+^*$,

$$Z_{\beta W} := \int_{\mathbb{R}^d} e^{-\beta W}\, du < \infty.$$

La quantité $Z_{\beta W}$ est appelée *fonction de partition de Gibbs*. Soit $\mu_{\beta W}$ la loi de probabilité sur $\mathbb{R}^d$ absolument continue par rapport à la mesure de Lebesgue, et de densité de probabilité $f_{\beta W}$ définie par

$$f_{\beta W} := (Z_{\beta W})^{-1} e^{-\beta W} = e^{-\beta W - \log(Z_{\beta W})}.$$

On suppose également que $\mathbf{Var}_{\mu_{\beta W}}(W) < \infty$ pour tout $\beta > 0$. Si $W(u)$ représente l'énergie de la configuration u, la loi $\mu_{\beta W}$ favorise les configuration de moindre énergie, d'autant plus que β est grand. La quantité $T := 1/\beta$ joue le rôle de la *température absolue*[5]. Un petit calcul montre que $\partial_\beta \log(Z_\beta) = -\mathbf{E}_{\mu_{\beta W}}(W)$. Par suite, l'énergie moyenne $\Psi(\beta) := \mathbf{E}_{\mu_{\beta W}}(W)$ vérifie $\Psi'(u) = -\mathbf{Var}_{\mu_{\beta W}}(W) < 0$. De plus, $\lim_{\beta\to 0^+} \Psi(\beta) = +\infty$ et $\lim_{\beta\to\infty} = \inf_{\mathbb{R}^d} W$. En vertu du théorème des valeurs intermédiaires, pour tout réel $w > \inf_{\mathbb{R}^d} W$, il existe un unique réel $\beta \in \mathbb{R}_+^*$ tel que $\Psi(\beta) = w$.

Soit à présent $w > \inf_{\mathbb{R}^d} W$, et soit μ une loi de probabilité sur $\mathbb{R}^d$ de densité f par rapport à la mesure de Lebesgue. Si μ vérifie la contrainte d'énergie moyenne $\mathbf{E}_\mu(W) = w$, alors

$$\mathbf{S}(f_{\beta W}) - \mathbf{S}(f) = \int_{\mathbb{R}^d} \frac{f}{f_{\beta W}} \log\left(\frac{f}{f_{\beta W}}\right) f_{\beta W}\, du = \mathbf{Ent}(f \mid f_{\beta W}),$$

où β est l'unique réel positif tel que $\Psi(\beta) = \mathbf{E}_{\mu_{\beta W}}(W) = w$. Par conséquent $\mathbf{S}(f) \leqslant \mathbf{S}(f_{\beta W})$, avec égalité si et seulement si $\mu = \mu_{\beta W}$. De plus, $\mathbf{S}(f_{\beta W}) = \beta w = \beta\Psi(\beta) = w\Psi^{-1}(w)$. L'entropie est donc maximisée, à énergie moyenne fixée, par la mesure de Boltzmann-Gibbs $\mu_{\beta W}$, et par elle seulement.

Soit à présent $\beta \in \mathbb{R}_+^*$, et soit μ une loi de probabilité sur $\mathbb{R}^d$ de densité f par rapport à la mesure de Lebesgue. On définit l'*énergie libre de Helmholtz de* μ par

$$\mathbf{F}(f) := \mathbf{E}_\mu(W) - \frac{1}{\beta}\mathbf{S}(f).$$

[5]Ici, la température est donc un paramètre qui correspond à l'« agitation ». En physique, les quantités β et T sont plutôt reliées par la formule $\kappa\beta = 1/T$ où κ est la constante de Boltzmann. Cette constante vaut $1,3806503 \times 10^{-23}\, m^2\, kg\, s^{-2}\, K^{-1}$, de sorte que la quantité κT représente une énergie, et que la quantité βW soit sans dimension. On a également $N\kappa = R$, où N est le nombre d'Avogadro et R la constante des gaz parfaits. Enfin, les formules de Boltzmann-Planck-Einstein pour l'énergie s'écrivent $\kappa T = mc^2 = \hbar\nu$, où c est la vitesse de la lumière dans le vide, $\hbar$ la constante de Planck, m la masse, et ν la fréquence. Dans notre monde mathématique, $\kappa = 1$.

C'est donc l'énergie moyenne à température $1/\beta$, diminuée du produit de l'entropie par la température. On a $\mathbf{F}(\mu_{\beta W}) = -\beta^{-1}\log(Z_{\beta W})$, de sorte que

$$\mathbf{F}(f) - \mathbf{F}(f_{\beta W}) = \frac{1}{\beta}\int_{\mathbb{R}^d} \frac{f}{f_{\beta W}} \log\left(\frac{f}{f_{\beta W}}\right) f_{\beta W}\, du = \mathbf{Ent}(f \mid f_{\beta W})\,.$$

Ainsi, $\mathbf{F}(f) \geqslant \mathbf{F}(f_{\beta W})$, avec égalité si et seulement si $\mu = \mu_{\beta W}$. L'énergie libre de Helmholtz est minimisée, à température fixée $1/\beta$, par la mesure de Gibbs-Boltzmann $\mu_{\beta W}$, et par elle seulement. Enfin, $\mathbf{F}(f) - \mathbf{F}(f_{\beta W}) = \mathbf{S}(f_{\beta W}) - \mathbf{S}(f)$ lorsque $\mathbf{E}_\mu(W) = \mathbf{E}_{\mu_{\beta W}}(W)$.

1.2.2 Lois discrètes et fonctionnelles discrètes

Considérons l'entropie discrète $\mathbf{S}$, définie sur le simplexe des lois de probabilité sur l'ensemble fini $\{1, \ldots, n\}$ par

$$\mu \mapsto \mathbf{S}(\mu) := -\sum_{i=1}^n \mu_i \log(\mu_i)$$

Le simplexe est un sous-ensemble convexe compact de $\mathbb{R}^n$, et $\mathbf{S}$ est continue et concave. Par conséquent, $\mathbf{S}$ atteint son minimum 0 sur les points extrémaux (masses de Dirac) et son maximum $\log(n)$ pour la mesure uniforme. La mesure uniforme sur $\{1, \ldots, n\}$ joue le même rôle que la mesure de Lebesgue sur $\mathbb{R}^d$. Cependant, la mesure uniforme est de masse finie, ce qui explique que l'entropie discrère soit positive et bornée.

Soit à présent $W : \{1, \ldots, n\} \to \mathbb{R}$ une fonction (un vecteur de $\mathbb{R}^n$ en fait). Pour tout $\beta \in \mathbb{R}_+$, on définit la loi μ_β par

$$\mu_\beta := \frac{1}{Z_\beta}(e^{-\beta W_1}, \ldots, e^{-\beta W_n})$$

où $Z_\beta := \sum_{i=1}^n e^{-\beta W_i}$. Un calcul simple montre que la fonction lisse $\Psi : \beta \in \mathbb{R}_+ \mapsto \mathbf{E}_{\mu_\beta}(W)$ vérifie $\Psi'(\beta) = -\mathbf{Var}_{\mu_\beta}(W) < 0$. Il s'agit donc d'une bijection de $\mathbb{R}_+$ dans $\mathbb{R}$. Il en découle que pour tout réel w, il existe un unique réel $\beta \in \mathbb{R}_+$ tel que $\mathbf{E}_{\mu_\beta}(W) = w$.

Soit à présent $w \in \mathbb{R}$ et μ un élément du simplexe, c'est-à-dire une loi de probabilité sur $\{1, \ldots, n\}$. Si μ vérifie la contrainte d'énergie moyenne $\mathbf{E}_\mu(W) = w$, alors

$$\mathbf{S}(\mu_\beta) - \mathbf{S}(\mu) = \sum_{i=1}^n \frac{\mu_i}{\mu_{\beta,i}} \log\left(\frac{\mu_i}{\mu_{\beta,i}}\right)\mu_{\beta,i}, = \mathbf{Ent}(\mu \mid \mu_\beta)\,,$$

où β est l'unique réel positif tel que $\mathbf{E}_{\mu_\beta}(W) = w$. Ainsi, $\mathbf{S}(\mu) \leqslant \mathbf{S}(\mu_\beta)$ avec égalité si et seulement si $\mu = \mu_\beta$. Nous avons obtenu qu'à énergie moyenne fixée

w, l'entropie $\mathbf{S}$ est maximisée de façon unique par la mesure de Gibbs-Boltzmann μ_β où β est déterminé de façon unique par W et w. On a $\mathbf{S}(\mu_\beta) = \log(Z_\beta) + \beta w$.

Soit à présent une température $1/\beta$ et soit μ un élément du simplexe, c'est-à-dire une loi de probabilité sur $\{1, \ldots, n\}$. L'*énergie libre de Helmholtz* $\mathbf{F}(\mu)$ de μ est définie par

$$\mathbf{F}(\mu) := \mathbf{E}_\mu(W) - \frac{1}{\beta}\mathbf{S}(\mu).$$

C'est donc l'énergie moyenne à température $1/\beta$, diminuée du produit de l'entropie par la température. On a $\mathbf{F}(\mu_\beta) = -\beta^{-1}\log(Z_\beta)$, et

$$\mathbf{F}(\mu) - \mathbf{F}(\mu_\beta) := \frac{1}{\beta}\sum_{i=1}^{n} \frac{\mu_i}{\mu_{\beta,i}} \log\left(\frac{\mu_i}{\mu_{\beta,i}}\right)\mu_{\beta,i} = \mathbf{Ent}(\mu \,|\, \mu_\beta)\,.$$

Ainsi, $\mathbf{F}(\mu) \geqslant \mathbf{F}(\mu_\beta)$, avec égalité si et seulement si $\mu = \mu_\beta$. L'énergie libre de Helmholtz est donc minimisée, à température fixée $1/\beta$, par la mesure de Gibbs-Boltzmann μ_β, et par elle seulement. Enfin, $\mathbf{F}(\mu) - \mathbf{F}(\mu_\beta) = \mathbf{S}(\mu_\beta) - \mathbf{S}(\mu)$ lorsque $\mathbf{E}_\mu(W) = \mathbf{E}_{\mu_\beta}(W)$.

Les maximum d'entropie discrète font toujours l'objet de recherches, en liaison avec la théorie de l'information (codage optimal, ...), l'optimisation discrète (équilibres de Nash, ...), la statistique mathématique (minimisation du risque, ...), etc. Le lecteur trouvera quelques points d'entrée bibliographiques dans [CT91], ainsi que dans les articles de Flemming Topsøe par exemple.

Exercice 1.2. Que se passe-t-il sur l'ensemble des lois de probabilité sur $\mathbb{N}$ ou $\mathbb{Z}$? Quelle est l'énergie qui donne naissance aux lois géométriques ? Idem pour la loi de Poisson ? La contrainte $\mathbf{E}_\mu(W) = w$ est linéaire en μ, et μ appartient à l'ensemble convexe des lois de probabilités. Donner une interprétation différentielle en termes d'optimum sous contrainte de la propriété de maximum d'entropie. On pourra penser à utiliser les conditions d'optimalité de Lagrange-Karush-Kuhn-Tucker, y compris dans le cas continu, cf. [HUL01, FLP01]. À quoi correspondent les multiplicateurs ? À quelle condition une famille $\{\mu_\alpha; \alpha \in \mathcal{I}\}$ de mesures de probabilité est associée à un maximum d'entropie pour une fonction énergie ? Comment le paramètre α est relié à la température ?

1.2.3 Famille exponentielle à température fixée et recentrage

Soit $W : \mathbb{R}^d \to \mathbb{R}$ un potentiel lisse tel que[6]

$$Z_{\beta,\lambda} := \int_{\mathbb{R}^d} e^{-\beta(W(u)-\lambda\cdot u)}\,du < \infty$$

pour tout $\beta > 0$ et tout $\lambda \in \mathbb{R}^d$. Soit $\mu_{\beta,\lambda}$ la mesure de probabilité de Boltzmann-Gibbs associée. Lorsque W est une forme quadratique, la mesure $\mu_{\beta,\lambda}$ est gaussienne[7]. Supposons que $u \mapsto \beta(W(u) + \lambda \cdot u) \in \mathrm{L}^2(\mu_{\beta,\lambda})$ pour tout λ. Un calcul simple donne

$$\nabla_\lambda \log(Z_{\beta,\lambda}) = \beta m_{\beta,\lambda},$$

où $m_{\beta,\lambda} := \mathbf{E}_{\mu_{\beta,\lambda}}(u \mapsto u)$. On a également pour tout i et j dans $\{1,\ldots,d\}$

$$\partial^2_{i,j} \log(Z_{\beta,\lambda}) = \partial_{\lambda_i} \log(m_{\beta,\lambda,j}) = \beta \mathbf{Cov}_{\mu_{\beta,\lambda}}(u \mapsto u_i, u \mapsto u_j)\,.$$

Ainsi, $\nabla^2_\lambda \log(Z_{\beta,\lambda}) = \beta\mathbf{Cov}_{\mu_{\beta,\lambda}}$. En particulier, $\Psi_\beta : \lambda \mapsto \log(Z_{\beta,\lambda})$ est convexe pour tout $\beta > 0$. Il s'agit d'une propriété classique des familles exponentielles[8], cf. [KS97, Let92, JL92].

Soit à présent $\beta > 0$ tel que $\mathbf{Cov}_{\mu_{\beta,\lambda}}$ soit inversible pour tout $\lambda \in \mathbb{R}^d$. Cela revient à dire que la loi $\mu_{\beta,\lambda}$ n'est pas portée par un hyperplan. La fonction $\lambda \mapsto \log(Z_{\beta,\lambda})$ est dans ce cas strictement convexe. Considérons pour tout $\theta \in \mathbb{R}^d$ l'équation $m_{\beta,\lambda} = \theta$ en la variable λ. Cela correspond à rechercher le paramètre λ tel que $\mu_{\beta,\lambda}$ ait pour moyenne θ. Comme $\lambda \mapsto m_{\beta,\lambda}$ est le gradient de la fonction convexe Ψ_β, cette équation à une solution unique dans chaque direction pour λ (DESSIN, le gradient croit radialement). La transformée de Legendre Ψ^*_β de Ψ_β est définie pour tout $\theta \in \mathbb{R}^d$ par :

$$\Psi^*_\beta(\theta) := \sup_{\lambda\in\mathbb{R}^d} (\lambda \cdot \theta - \Psi_\beta(\lambda)).$$

Le supremum est atteint aux points λ vérifiant $\theta = \nabla\Psi_\beta(\lambda)$, et dans ce cas, $\Psi^*_\beta(\theta) = \lambda \cdot w - \Psi_\beta(\lambda)$. En conclusion, pour tout $\theta \in \mathbb{R}^d$, et tout $\lambda \in \mathbb{R}^d$, il y a

[6] Les variables β et λ sont parfois qualifiées de *paramètre d'ordre*. En physique, la partie linéaire $u \mapsto \lambda \cdot u$ correspond par exemple à champ gravitationnel ou électrostatique si u représente la position, ou à un champ magnétique si u représente la vitesse. Rappelons qu'une particule de charge q de position u et de vitesse v, plongée dans un champ électrique E et dans un champ magnétique B, subit une force de Laplace-Lorentz de la forme $q(E_u + v \wedge B_u)$, où $\wedge$ représente le produit vectoriel dans $\mathbb{R}^3$. Lorsque les champs E et B sont constants, la trajectoire d'une telle particule est une hélice. Les cyclotrons sont basés sur ce phénomène.

[7] En particulier, quand $W(u) = |u|^2/2$ pour tout $u \in \mathbb{R}^d$, alors $\mu_{\beta,\lambda}$ est de moyenne λ et de matrice de covariance $\beta^{-1}\mathrm{I}_d$. Dans ce cas, $Z_{\beta,\lambda} = (2\pi/\beta)^{n/2}$

[8] Pour tout $\beta > 0$, la famille de lois $(\mu_{\beta,\beta^{-1}\lambda})_{\lambda\in\mathbb{R}^d}$ est la famille exponentielle associée à la loi de Boltzmann-Gibbs $\mu_{\beta,0}$. La quantité $Z_{\beta,\lambda}/Z_{\beta,0}$ est la transformée de Laplace en $\beta\lambda$ de $\mu_{\beta,0}$.

équivalence entre les équations

$$m_{\beta,\lambda} = \theta \quad \text{et} \quad \nabla\Psi_\beta(\lambda) = \theta,$$

et de plus, θ et λ sont reliés par l'équation

$$\Psi^*_\beta(\theta) - \Psi_\beta(\lambda) = \lambda \cdot \theta.$$

Ces solutions λ dépendent de W, de β, et de θ. Elle décrivent un cercle horizontal. Il y a unicité de la solution le long de chaque demi-droite issue de l'origine.

Fixons $\lambda \in \mathbb{R}^d$. Pour toute loi de probabilité μ sur $\mathbb{R}^d$, l'énergie moyenne $\mathbf{E}_\mu(u \mapsto W(u) - \lambda \cdot u)$ vaut $\mathbf{E}_\mu(W) - \lambda \cdot \mathbf{E}_\mu(u \mapsto u)$. La mesure de probabilité de Gibbs-Boltzmann $\mu_{\beta,\lambda}$ réalise le maximum de la fonctionnelle entropie $\mu \mapsto \mathbf{S}(\mu)$ sous la contrainte linéaire $\mathbf{E}_\mu(W) - \lambda \cdot \mathbf{E}_\mu(u \mapsto u) = w$. La température $1/\beta$ est déterminée de manière unique par w, et $\mathbf{E}_{\mu_{\beta,\lambda}}(W) = w + \lambda \cdot m_{\beta,\lambda}$.

1.3 Mécanique statistique et mesures de Gibbs

FIXME:

1.4 Statistique bayésienne

Les maxima d'entropie sont utilisés dans des contextes très variés. Ils jouent par exemple un rôle important en statistique bayésienne, pour le choix de loi a priori dites « non-informatives », cf. [Rob01] et [CT91]. Ce principe a été formulé par Jaynes dans les années 1950, cf. [Jay89]. Chaque contrainte intervenant dans la maximisation de l'entropie représente une information dont on dispose a priori sur la loi recherchée (moments, support, ...). Toute loi à densité sur $\mathbb{R}^d$ est un maximum d'entropie à énergie moyenne fixée, l'énergie à utiliser étant l'opposé en signe du logarithme de sa densité. Ainsi par exemple, sur l'ensemble des lois à densité à support compact fixé, la mesure uniforme maximise l'entropie. De même, sur l'ensemble des lois à densité à support dans $\mathbb{R}_+$ et à moyenne fixée, loi exponentielle maximise l'entropie. Enfin, et comme nous l'avons déjà mentionné, sur l'ensemble des lois à densité à variance fixée, la mesure de Gauss maximise l'entropie.

Exercice 1.3. Quelle est la contrainte qui donne naissance à la loi de Wishart en théorie des matrices aléatoires ?

1.5 Conclusion

Pour une fonction potentiel donnée, l'entropie est maximisée, à énergie moyenne fixée, par la mesure de Boltzmann-Gibbs dont la température est déterminée de façon unique par le niveau moyen d'énergie spécifié. Dualement, l'énergie libre de

Helmholtz est minimisée, à température fixée, par la mesure de Boltzmann-Gibbs associée à cette température.

Comme $x \in \mathbb{R}_+ \mapsto x \log(x)$ est convexe, la fonctionnelle $f \mapsto \mathbf{S}(f)$ est convexe sur l'ensemble convexe des densités de probabilité. La grande idée qui sera reprise sous différentes formes plus tard est l'existence de fonctionnelles convexes qui sont monotones le long d'un processus d'évolution. La mesure de probabilité d'équilibre du système correspond alors à un optimum de ces fonctionnelles. La fonctionnelle $\mathbf{H}$ pour l'équation de Boltzman en est un exemple parmis d'autres. Pour un processus d'évolution donné, il est commode de désigner par le mot entropie toute fonctionnelle convexe et croissante (resp. concave et décroissante) le long du processus. Dans la littérature, on parle de « dissipation d'entropie » ou de « production d'entropie », selon que la fonctionnelle est convexe ou concave, et cela ne tient qu'à un signe ! L'énergie libre est par exemple une fonctionnelle concave et typiquement décroissante le long du processus de Markov dont la loi invariante est la mesure de Gibbs-Boltzmann associée à l'énergie libre considérée. De nombreuses équations aux dérivées partielles ont des solutions auxquelles il est possible d'associer ce type de fonctionnelles. On en trouvera quelques exemples dans le livre de Villani [Vil03] sur la théorie du transport optimal de mesure. De telles fonctionnelles apparaissent également dans la preuve de Hamilton et Perelman de la conjecture de Poincaré, l'équation d'évolution étant liée au flot de Ricci, cf. [Bes05].

2 Processus de Markov et énergie libre

La théorie moderne des probabilités nait pendant le vingtième siècle, suite notamment aux travaux de l'école russe de Kolmogorov. La théorie mathématique des processus de Markov et des diffusions est crée à ce moment là. Un processus de Markov est associé à une équation d'évolution de mesures de probabilité, qui est linéaire. Les diffusions correspondent au cas où cette équation est donnée par un opérateur différentiel linéaire du second ordre. Toute mesure de probabilité de Gibbs-Boltzmann peut être vue comme la mesure invariante et réversible d'un processus de Markov diffusif, pour lequel elle représente l'état d'équilibre. Ce processus d'évolution transforme une densité de probabilité en la mesure de Gibbs-Boltzmann, par ergodicité. Cette évolution ne conserve pas l'énergie moyenne, et l'entropie n'a aucune raison de croitre. En revanche, l'énergie libre de Helmholtz est décroissante. De plus, sa décroissance exponentielle est équivalente à l'existence d'une inégalité de Sobolev logarithmique pour la mesure de Gibbs-Boltzmann.

Soit $W : \mathbb{R}^d \to \mathbb{R}$ une fonction lisse telle que $\int_{\mathbb{R}^d} e^{-W} \, du = 1$. On note μ la loi de probabilité de densité e^{-W} par rapport à la mesure de Lebesgue. On note également $\mathbf{E}_\mu(f) := \int_{\mathbb{R}^d} f \, d\mu$. Soit $(X_t)_{t \geqslant 0}$ le processus solution de l'équation différentielle stochastique[9]

$$dX_t = \sqrt{2}\, dB_t - \nabla W(X_t)dt,$$

où $(B_t)_{t\geqslant 0}$ est un mouvement brownien standard sur $\mathbb{R}^d$ (DESSIN). Le processus $(X_t)_{t\geqslant 0}$ est une diffusion de Kolmogorov, associée au potentiel W. C'est également un processus de Markov. Lorsque W est une forme quadratique, il s'agit d'un processus d'Ornstein-Uhlenbeck. Il est important de rappeler que la solution $(X_t)_{t\geqslant 0}$ de l'É.D.S. n'est définie en réalité que sur un intervalle de temps maximal $[0, T[$ où T est le *temps d'explosion du processus*, cf. [Roy99, p. 26]. Le temps d'explosion T est presque-sûrement infini lorsque le potentiel W vérifie l'une des deux conditions suffisantes suivantes :

1. $\lim_{|u|\to\infty} W(u) = \infty$ et $|\nabla W|^2 - \mathbf{\Delta} W$ est bornée inférieurement sur $\mathbb{R}^d$;
2. $\exists(\alpha, \beta) \in \mathbb{R}^2;\ \forall u \in \mathbb{R}^d;\ u \cdot \nabla W(u) \geqslant a|u|^2 + b.$

En vertu du théorème des accroissements finis, les conditions de croissances à l'infini ci-dessus sont par exemple assurées simultanément lorsque $\lim_{|u|\to\infty} W(u) = \infty$ et que le spectre de la matrice hessienne de W est uniformément minoré sur $\mathbb{R}^d$ par une constante réelle finie. Dans toute la suite, Nous supposerons toujours que le temps d'explosion est infini presque-sûrement. Pour tout $t \geqslant 0$, tout $x \in \mathbb{R}^d$, et toute fonction continue et bornée $f : \mathbb{R}^d \to \mathbb{R}$, on pose

$$\mathbf{P}_t(f)\,(x) := \mathbb{E}(f(X_t)\,|\,X_0 = x).$$

La famille d'opérateurs linéaires $(\mathbf{P}_t)_{t\geqslant 0}$ est un semi-groupe de Markov : $\mathbf{P}_t(\mathbf{P}_s(f)) = \mathbf{P}_{t+s}(f)$, $\mathbf{P}_0(f) = f$, $\mathbf{P}_t(f) \geqslant 0$ si $f \geqslant 0$ sur $\mathbb{R}^d$, et enfin, $\mathbf{P}_t(1) = 1$. De plus, si f est lisse,

$$\lim_{t\to 0^+} \frac{\mathbf{P}_t(f)\,(x) - \mathbf{P}_0(f)\,(x)}{t} = (\mathbf{L}f)(x),$$

où $\mathbf{L}f := \mathbf{\Delta} f - \nabla W \cdot \nabla f$. On dit que l'opérateur différentiel linéaire du second ordre $\mathbf{L}$ est le générateur infinitésimal de $(X_t)_{t\geqslant 0}$. On a également $\partial_t \mathbf{P}_t(f) = \mathbf{P}_t(\mathbf{L}f) = \mathbf{L}\mathbf{P}_t(f)$.

La mesure de probabilité μ est invariante pour le processus : si $\mathcal{L}(X_0) = \mu$, alors $\mathcal{L}(X_t) = \mu$ pour tout $t \geqslant 0$. Cela s'écrit également $\mathbf{E}_\mu(\mathbf{P}_t(f)) = \mathbf{E}_\mu(f)$ pour tout $t \geqslant 0$ et toute fonction continue bornée f, ou encore $\mathbf{E}_\mu(\mathbf{L}f) = 0$ pour toute

[9] Il s'agit par exemple de l'écriture de la relation fondamentale de la dynamique de Newton : somme des forces = masse $\times$ accélération. Ici, la masse vaut 1, la quantité X_t correspond à la vitesse d'une particule, dX_t/dt à son accélération, dB_t/dt à une force aléatoire gaussienne homogène due aux chocs (bruit blanc), et enfin $-\nabla W(X_t)$ à la force associée au potentiel de vitesse W (résistance à l'avancement par exemple). Ce modèle a été utilisé par Einstein et Langevin au début de vingtième siècle pour modéliser les diffusions de particules dans un fluide.

fonction lisse f. Il en découle en particulier que si f est une densité de probabilité par rapport à μ, alors $\mathbf{P}_t(f)$ est encore une densité de probabilité par rapport à μ.

La mesure de probabilité μ est symétrique : pour tout f et g lisses et à support compact, on dispose de la formule d'intégration par parties suivante :

$$\mathbf{E}_\mu(f\mathbf{L}g) = \mathbf{E}_\mu(g\mathbf{L}f) \quad \text{et} \quad \mathbf{E}_\mu(f\mathbf{L}f) = -\mathbf{E}_\mu(|\nabla f|^2).$$

Un célèbre résultat de Birkoff et von Neumann en théorie ergodique affirme que pour toute fonction $f \in \mathrm{L}^1(\mu)$ et pour presque-tout $x \in \mathbb{R}^d$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s)\, ds = \mathbf{E}_\mu(f)\,.$$

Ainsi, le temps d'occupation de x converge vers la moyenne en espace pour la mesure invariante μ. On a également que $\lim_{t\to\infty} \mathbf{P}_t(f) = \mathbf{E}_\mu(f)$ dans $\mathrm{L}^2(\mu)$. Cette dernière propriété peut être déduite d'une décomposition spectrale de $\mathbf{L}$ et du calcul fonctionnel, cf. [Bak94]. La propriété d'invariance donne

$$\|\mathbf{P}_t(f) - \mathbf{E}_\mu f\|_2^2 = \mathbf{E}_\mu\big((\mathbf{P}_t(f) - \mathbf{E}_\mu f)^2\big) = \mathbf{Var}_\mu(\mathbf{P}_t(f))\,.$$

D'autre part, en vertu de la propriété de semi-groupe et de l'inégalité de Jensen,

$$\|\mathbf{P}_{t+s}(f) - \mathbf{E}_\mu f\|_2^2 = \|\mathbf{P}_{t+s}(f - \mathbf{E}_\mu f)\|_2^2 \leqslant \|\mathbf{P}_s(f - \mathbf{E}_\mu f)\|_2^2.$$

Ainsi, la fonctionnelle variance décroit le long de $\mathbf{P}_t(f)$. L'intégration par parties entraîne lorsque f est lisse :

$$\partial_t \mathbf{E}_\mu\big((\mathbf{P}_t(f) - \mathbf{E}_\mu f)^2\big) = -2\mathbf{E}_\mu\big(|\nabla \mathbf{P}_t(f)|^2\big) \leqslant 0.$$

Nous allons à présent faire de même pour la fonctionnelle énergie libre.

Remarque 2.1 (Énergie moyenne le long du semi-groupe). Pour une densité f par rapport à μ, à support compact, on peut écrire formellement en vertu de la formule d'intégration par parties :

$$\partial_t \mathbf{E}_\mu(W\mathbf{P}_t(f)) = \mathbf{E}_\mu(W\mathbf{L}\mathbf{P}_t(f)) = \mathbf{E}_\mu(\mathbf{P}_t(f)\,\mathbf{L}W) = \mathbf{E}_\mu\big(\mathbf{P}_t(f)\,(\mathbf{\Delta} W - |\nabla W|^2)\big)\,.$$

L'énergie moyenne n'a aucune raison d'être monotone. Cependant, des bornes uniformes sur $\mathbf{\Delta} W - |\nabla W|^2$ peuvent conduire à des bornes de tendances. On pense tout naturellement à l'exemple explicite du processus d'Ornstein-Uhlenbeck…

2.1 Décroissance de l'énergie libre

Pour toute densité f par rapport à μ, on considère la quantité

$$\mathbf{Ent}_\mu(f) := \mathbf{E}_\mu(f\log(f)) := \int_{\mathbb{R}^d} f\log(f)\,d\mu.$$

C'est donc la divergence de Kullback-Leibler de $fd\mu$ par rapport à μ. On a toujours $\mathbf{Ent}_\mu(f) \geqslant 0$, avec égalité si et seulement si f est égale μ-presque sûrement à 1. La fonctionnelle $f \mapsto \mathbf{Ent}_\mu(f)$ est convexe. La quantité $\mathbf{Ent}_\mu(f)$ est également l'énergie libre de Helmholtz de fe^{-W} pour la fonction d'énergie W à température $\beta = 1$:

$$\mathbf{Ent}_\mu(f) = \int_{\mathbb{R}^d} Wfe^{-W}\,du + \mathbf{H}(fe^{-W}).$$

Ainsi, $\mathbf{Ent}_\mu(f)$ est égale à l'énergie moyenne de fe^{-W} diminuée de son entropie[10]. Nous savons que l'énergie libre est minimisée par la mesure de Gibbs-Boltzmann μ, et cela correspond à $f \equiv 1$. Lorsque $f : \mathbb{R}^d \to \mathbb{R}_+$ n'est pas une densité par rapport à μ, on note encore $\mathbf{Ent}_\mu(f)$ la quantité

$$\mathbf{E}_\mu(f\log(f)) - \mathbf{E}_\mu(f)\log(\mathbf{E}_\mu(f)) = \mathbf{E}_\mu\left(f\log\frac{f}{\mathbf{E}_\mu(f)}\right).$$

On observe donc que $\mathbf{Ent}_\mu(f) = \mathbf{E}_\mu(f)\,\mathbf{Ent}_\mu(f/\mathbf{E}_\mu(f))$, où $\mathbf{Ent}_\mu(f/\mathbf{E}_\mu(f))$ est l'énergie libre de la densité $f/\mathbf{E}_\mu(f)$. L'inégalité de Jensen montre que $\mathbf{Ent}_\mu(f)$ n'est jamais négative.

Considérons à présent les propriétés de monotonie de l'énergie libre. Soit f une densité par rapport à μ. La propriété de semi-groupe, suivie de l'inégalité de Jensen pour la fonction convexe $u \mapsto u\log(u)$ et l'espérence $\mathbf{P}_t(\cdot)$ donnent $\mathbf{P}_{t+s}(f)\log(\mathbf{P}_{t+s}(f)) \leqslant \mathbf{P}_t(\mathbf{P}_s(f)\log(\mathbf{P}_s(f)))$. Par invariance de μ, on obtient donc

$$\mathbf{Ent}_\mu(\mathbf{P}_{t+s}(f)) \leqslant \mathbf{Ent}_\mu(\mathbf{P}_s(f)).$$

Seule la convexité de la fonction $x \mapsto x\log(x)$ intervient pour l'établissement de cette propriété de décroissance.

Théorème 2.2 (Décroissance de l'énergie libre). *Pour toute densité lisse f par rapport à μ,*

$$\partial_t \mathbf{Ent}_\mu(\mathbf{P}_t(f)) = -\mathbf{E}_\mu\left(\frac{|\nabla \mathbf{P}_t(f)|^2}{\mathbf{P}_t(f)}\right) \leqslant 0. \tag{2}$$

« Il est amusant de constater que **Ent** est à la fois le début du mot *Entropie* et du mot *Enthalpie* ! » (boutade de D. Bakry, en aparté à M. Ledoux). Pourtant, la quantité $\mathbf{Ent}_\mu(f)$ désigne à la fois une *entropie relative de Kulback-Leibler*, et une *énergie libre de Helmholtz.*

[10]En toute rigueur, il faudrait inclure la température dans toute cette partie.

Démonstration. Si $f_t := \mathbf{P}_t(f)$, on a $\partial_t \mathbf{Ent}_\mu(f_t) = \mathbf{E}_\mu((1+\log(f_t))\mathbf{L}f_t)$. Comme μ est symétrique, on obtient $\partial_t \mathbf{Ent}_\mu(f_t) = \mathbf{E}_\mu(f_t \mathbf{L}\log(f_t))$. Or pour toute fonction lisse $\Phi : \mathbb{R} \to \mathbb{R}$, $\mathbf{L}\Phi(f) = \Phi'(f)\mathbf{L}f + \Phi''(f)|\nabla f|^2$. La formule requise s'en déduit avec $\Phi = \log$. □

Théorème 2.3 (Décroissance exponentielle de l'énergie libre). *Soit $c > 0$ une constante et $\mathcal{F}$ l'ensemble des densités lisses par rapport à μ. Il y a équivalence entre*

$$\forall f \in \mathcal{F},\ \forall t \geqslant 0, \quad \mathbf{Ent}_\mu(\mathbf{P}_t(f)) \leqslant e^{-t/c}\mathbf{Ent}_\mu(f) \tag{3}$$

et

$$\forall f \in \mathcal{F}, \quad \mathbf{Ent}_\mu(f) \leqslant c\,\mathbf{E}_\mu\left(\frac{|\nabla f|^2}{f}\right). \tag{4}$$

Démonstration. Soit $\alpha(t) := \mathbf{Ent}_\mu(\mathbf{P}_t(f))$. En combinant (2) avec (4), il vient $c\alpha'(t) \leqslant \alpha(t)$. Le lemme de Gronwall fournit alors (3). Réciproquement, soit $\beta(t) := \mathbf{Ent}_\mu(\mathbf{P}_t(f)) - e^{-t/c}\mathbf{Ent}_\mu(f)$. La propriété (3) s'écrit $\beta \leqslant 0$. En particulier, $\beta'(0) \leqslant 0$, qui donne (4) après usage de (2). □

On dit que μ vérifie une *inégalité de Gross* ou de *Sobolev logarithmique* de constante c lorsque (4) a lieu. On remarquera que

$$\mathbf{Ent}_\mu(f) = \mathbf{E}_\nu\left(\log\frac{d\nu}{d\mu}\right),$$

où $d\nu := f d\mu$. On appelle parfois *information de Fisher* la quantité

$$\mathbf{E}_\mu\left(\frac{|\nabla f|^2}{f}\right) = \mathbf{E}_\mu\big(|\nabla \log(f)|^2 f\big) = \mathbf{E}_\nu\left(\left|\nabla \log\frac{d\nu}{d\mu}\right|^2\right).$$

Pour toute fonction lisse $g : \mathbb{R}^d \to \mathbb{R}_+$, l'écriture de (4) pour la densité $f = g/\mathbf{E}_\mu(g)$ donne :

$$\mathbf{E}_\mu(g\log(g)) - \mathbf{E}_\mu(g)\log\mathbf{E}_\mu(g) \leqslant c\,\mathbf{E}_\mu\left(\frac{|\nabla g|^2}{g}\right).$$

De même, pour toute fonction lisse $h : \mathbb{R}^d \to \mathbb{R}$, cela donne pour $g = h^2$:

$$\mathbf{E}_\mu\big(h^2\log(h^2)\big) - \mathbf{E}_\mu\big(h^2\big)\log\mathbf{E}_\mu\big(h^2\big) \leqslant 4c\,\mathbf{E}_\mu\big(|\nabla h|^2\big).$$

2.2 Mesures de probabilité log-concaves

On dit que la mesure de probabilité μ est *log-concave de courbure* $\rho > 0$ lorsque pour tout x et y dans $\mathbb{R}^d$,

$$y^\top \nabla^2 W(x) y \geqslant \rho |y|^2.$$

(DESSIN) Cela signifie que le spectre de la matrice hessienne $\nabla^2 W(x)$ au point x est minoré par ρ, pour tout x. L'exemple emblématique est fourni par les lois gaussiennes pour lesquelles W est une forme quadratique non dégénée. En particulier, la gaussienne standard $\mathcal{N}(0, \mathrm{I}_d)$ est log-concave de courbure 1.

Théorème 2.4 (Bakry & Émery, $\approx$ 1984). *Si μ est log-concave de courbure ρ, alors elle vérifie une inégalité de Sobolev logarithmique de constante $1/(2\rho)$.*

Démonstration. Soit f une densité par rapport à μ, et $f_t := \mathbf{P}_t(f)$. La log-concavité entraîne que $[\nabla, \mathbf{L}] = \nabla^2 W \nabla \geqslant \rho |\nabla|^2$, qui s'avère[11] être la version infinitésimale de $|\nabla \mathbf{P}_t(f)| \leqslant e^{-\rho t} \mathbf{P}_t(|\nabla f|)$. Cela donne

$$\frac{|\nabla f_t|^2}{f_t} \leqslant e^{-2\rho t} \frac{\mathbf{P}_t(|\nabla f|)^2}{\mathbf{P}_t(f)}.$$

L'inégalité de Jensen pour l'espérence $\mathbf{P}_t(\cdot)$ et la fonction convexe $(u,v) \mapsto v^2/u$ donne[12]

$$\frac{\mathbf{P}_t(|\nabla f|)^2}{\mathbf{P}_t(f)} \leqslant \mathbf{P}_t\left(\frac{|\nabla f|^2}{f}\right).$$

L'invariance de μ conduit alors au résultat désiré à

$$\partial_t \mathbf{Ent}_\mu(f_t) = \mathbf{E}_\mu\left(\frac{|\nabla f_t|^2}{f_t}\right) \leqslant e^{-2\rho t} \mathbf{E}_\mu\left(\frac{|\nabla f|^2}{f}\right).$$

□

La constante obtenue est optimale pour les lois gaussiennes, car l'inégalité devient une égalité pour les fonctions f de la forme $x \mapsto e^{-\lambda x}$. L'inégalité de Sobolev logarithmique pour les lois gaussiennes peut être obtenue de manière élémentaire à partir de la mesure de Bernoulli sur l'espace à deux points, en utilisant le théorème central limite, comme l'a fait Gross dans [Gro75]. Il est alors possible d'en déduire les inégalités pour les mesures log-concaves à courbure positive par le biais du théorème suivant.

[11] Un calcul non trivial est à faire, cf. par exemple [ABC$^+$00, Chap. 5] ou [Led00]. D'autre part, il est intéressant d'observer que la condition de log-concavité entraîne en particulier la non-explosion, cf. [Roy99, p. 48].

[12] On peut ici utiliser l'inégalité de Cauchy-Schwarz en lieu et place de l'inégalité Jensen. Cependant, les généralisations à des processus de Markov à espace d'état discret doivent contourner l'absence de propriété de diffusion, et l'inégalité de Jensen reste la seule approche valable.

Théorème 2.5 (Caffarelli, ≈ 2000). *Soit μ une mesure de probabilité sur $\mathbb{R}^d$, log-concave de courbure $\rho > 0$. Alors elle est l'image de la mesure gaussienne standard $\mathcal{N}(0, \mathrm{I}_d)$ par une application lisse $F : \mathbb{R}^d \to \mathbb{R}^d$ qui vérifie $\rho |F(x) - F(y)|^2 \leqslant |x - y|^2$ pour tout x et y dans $\mathbb{R}^d$.*

La fonction F est donc lipschitz et vérifie $\rho \|\mathrm{D}F\|_{2,2}^2 \leqslant 1$. Ce théorème n'est pas facile. Il fait appel à l'équation de Monge-Ampère liée au problème de transport optimal de la mesure via le théorème de Brenier. Voici comment passer de $\mathcal{N}(0, \mathrm{I}_d)$ à μ pour le membre de droite de l'inégalité de Sobolev logarithmique :

$$\mathbf{E}_{\mathcal{N}(0,\mathrm{I}_d)}(|\nabla(f \circ F)|)^2 = \mathbf{E}_\mu\big(|(\mathrm{D}F)(\nabla f)|^2 \circ F\big) \leqslant \frac{1}{\rho} \mathbf{E}_\mu\big(|\nabla f|^2\big) .$$

Remarque 2.6. Question de D. Bakry : l'entropie décroit-elle exponentiellement vite dans l'équation de Boltzmann ? Question de F. Baudoin : existe-t-il un mécanisme général qui associe un processus d'évolution (une É.D.P.) à une inégalité fonctionnelle ? On trouvera quelques éléments de réponse dans [Vil02] et [Vil03] par exemple.

2.3 Tensorisation et perturbation

Les inégalités de Sobolev logarithmiques sont stables par translation et dilatation de la mesure. Il s'avère qu'elles sont également stables par perturbation et tensorisation de la mesure de probabilité sous-jacente.

Théorème 2.7 (Holley & Stroock, ≈ 1987). *Soit μ une mesure de probabilité sur $\mathbb{R}^n$ vérifiant une inégalité de Sobolev logarithmique de constante c. Pour toute fonction bornée $B : \mathbb{R}^n \to \mathbb{R}$, la mesure de probabilité μ_B définie par $d\mu_B := Z_B^{-1} e^B d\mu$ où $Z_B := \mathbf{E}_\mu\big(e^B\big)$ vérifie une inégalité de Sobolev logarithmique de constante $ce^{2\mathrm{osc}(B)}$, où $\mathrm{osc}(B) := \sup_{\mathbb{R}^n} B - \inf_{\mathbb{R}^n} B$.*

Démonstration. Pour toute mesure de probabilité μ sur $\mathbb{R}^n$ et toute fonction $f : \mathbb{R}^n \to \mathbb{R}_+$,

$$\mathbf{E}_\mu(f \log(f)) - \mathbf{E}_\mu(f) \log(\mathbf{E}_\mu(f)) = \inf_{a>0} \mathbf{E}_\mu(f \log(f/a) + a - f) .$$

L'infimum est atteint en $a = \mathbf{E}_\mu(f)$. Voir également [Roy99, 3.1.18] et [Cat05]. □

Soit $W_B := W + B$, où $d\mu = e^{-W} dx$. La condition $\mathrm{osc}(B) < \infty$ n'est pas suffisante pour garantir l'absence d'explosion du processus de générateur $\boldsymbol{\Delta} - \nabla W_B \cdot \nabla$ pour lequel μ_B est invariante et symétrique. Une condition suffisante pour l'absence d'explosion est $\mathrm{osc}(\nabla B) < \infty$ et $\mathrm{osc}(\nabla^2 B) < \infty$. Le théorème de Holley-Stroock donne une propriété de stabilité de l'inégalité de Sobolev logarithmique par perturbation de la mesure, qui n'assure pas la stabilité de l'absence d'explosion de la diffusion de Kolmogorov associée.

Théorème 2.8 (Gross & Segal, ≈ 1975). *Soit μ_1 et μ_2 des mesures de probabilité sur $\mathbb{R}^n$ et $\mathbb{R}^m$ vérifiant des inégalités de Sobolev logarithmiques de constantes c_1 et c_2 respectivement. Alors la mesure produit $\mu_1 \otimes \mu_2$ sur $\mathbb{R}^{n+m}$ vérifie une inégalité de Sobolev logarithmique de constante $\max(c_1, c_2)$. En particulier, si une mesure de probabilité μ vérifie une inégalité de Sobolev logarithmique de constante c sur $\mathbb{R}^d$, alors pour tout entier $n > 0$, la mesure de probabilité produit $\mu^{\otimes n}$ sur $\mathbb{R}^{nd}$ vérifie également une inégalité de Sobolev logarithmique de constante c.*

Démonstration. Soit $f : \mathbb{R}^{n+m} \to \mathbb{R}_+$ une densité pour $\mu_1 \otimes \mu_2$. Soient $f_1 := f/\mathbf{E}_{\mu_1}(f)$ et $f_2 := f/\mathbf{E}_{\mu_2}(f)$ ses densités conditionnelles marginales. On écrit

$$\mathbf{Ent}_{\mu_1\otimes\mu_2}(f) = \mathbf{E}_{\mu_1\otimes\mu_2}(f \log f_2) + \mathbf{E}_{\mu_1\otimes\mu_2}(f \log \mathbf{E}_{\mu_2}(f)).$$

Or en notant $g := \mathbf{E}_{\mu_1}(f)\,\mathbf{E}_{\mu_2}(f)$ et $d\mu_g := gd(\mu_1 \otimes \mu_2)$, il vient

$$\mathbf{E}_{\mu_1\otimes\mu_2}(f \log f_1) - \mathbf{E}_{\mu_1\otimes\mu_2}(f \log \mathbf{E}_{\mu_2}(f)) = \mathbf{E}_{\mu_g}\left(\frac{f}{g} \log \frac{f}{g}\right) \geqslant 0.$$

Ainsi, nous avons obtenu que

$$\mathbf{Ent}_{\mu_1\otimes\mu_2}(f) \leqslant \mathbf{E}_{\mu_1\otimes\mu_2}(f \log f_2) + \mathbf{E}_{\mu_1\otimes\mu_2}(f \log f_1).$$

L'expression est symétrique en les indices 1 et 2. Il ne reste plus qu'à observer que

$$\mathbf{E}_{\mu_1\otimes\mu_2}(f \log f_2) = \mathbf{E}_{\mu_1}(\mathbf{E}_{\mu_2}(f \log(f)) - \mathbf{E}_{\mu_2}(f) \log \mathbf{E}_{\mu_2}(f)).$$

$\square$

Il est possible de remplacer la fonction $u \mapsto u\log(u)$ par toute fonction convexe Φ telle que $(u, v) \mapsto \Phi''(u)v^2$ soit convexe. La fonctionnelle correspondante est alors parfois qualifiée de Φ-entropie. Elle est convexe, et monotone le long du processus de diffusion ergodique associé à la mesure de Gibbs-Boltzmann considérée. De plus, sa décroissance exponentielle est équivalente à une inégalité de Φ-Sobolev, qui a lieu pour les mesures à densité log-concaves. Enfin, elle vérifie les propriétés de tensorisation et de perturbation. Tous ces aspects sont considérés dans [Cha04]. Le cas $\Phi(u) = u^2$ correspond à la fonctionnelle variance, et aux inégalités de Poincaré.

3 Un problème de mesures produit conditionnées

On dit qu'une mesure de probabilité sur $\mathbb{R}^N$ est de potentiel W pour une mesure de référence ν lorsque $\mu \ll \nu$ avec $d\mu = Z^{-1}e^{-W}d\nu$ où $Z^{-1} := \int e^{-W}\,d\nu$. On note $\mathcal{L}^N$ la mesure de Lebesgue sur $\mathbb{R}^N$.

Soit μ une mesure de probabilité de potentiel lisse $V : \mathbb{R} \to \mathbb{R}$ par rapport à $\mathcal{L}$. La mesure produit $\mu_N := \mu^{\otimes N} = \mu \otimes \cdots \otimes \mu$ est de potentiel $\sum_{i=1}^N V(x_i)$ pour $\mathcal{L}^N$. On considère la fonction $\sigma_N : \mathbb{R}^N \to \mathbb{R}$ définie pour tout $x \in \mathbb{R}^N$ par $\sigma(x) := \frac{1}{N}\sum_{i=1}^N x_i$. Soit $m \in \mathbb{R}$. On considère la mesure conditionnelle

$$\nu_{N,m} := \mu_N(\cdot \,|\, \sigma_N = m).$$

De telles mesure modélisent naturellement des dynamiques conservatives, parfois qualifiées de dynamiques de Ginzburg-Landau, par exemple des interfaces SOS[13] (DESSIN). L'hyperplan affine

$$X_{N,m} := \left\{x \in \mathbb{R}^N ; \sigma_N(x) = m\right\},$$

supportant $\nu_{N,m}$, est de mesure nulle pour μ_N. On parle parfois de contrainte mince (DESSIN). La définition de $\nu_{N,m}$ mérite d'être précisée. Pour une fonction continue et bornée $f : \mathbb{R}^N \to \mathbb{R}$, la quantité $\mathbf{E}_{\nu_{N,m}}(f)$ peut être définie de différentes manières :

- **Approche de physicien.** $\mathbf{E}_{\nu_{N,m}}(f) = \lim_{\beta\to\infty} \mathbf{E}_{\mu_{N,m,\beta}}(f)$, où $\mu_{N,m,\beta}$ est de potentiel $\beta(\sigma_N(x) - m)^2$ pour μ_N. Cette approche, adoptée dans [Cap03], correspond à l'ajout d'un potentiel infini en dehors de l'hyperplan $X_{N,m}$;
- **Approche de géomètre.** $\mathbf{E}_{\nu_{N,m}}(f) = \int_{\mathbb{R}^{N-1}} f(x_1, \ldots, x_{N-1}, Nm - \sigma_{N-1}(x)) d\mu_{(N-1)}(x)$, où $\mu_{(N-1)}$ est de potentiel $\sum_{i=1}^{N-1} V(x_i) + V(Nm - \sigma_{N-1}(x))$ pour $\mathcal{L}^{N-1}$. Cette approche, adoptée dans [Cha03], revient à se placer dans une carte de la variété affine $X_{N,m}$.

Ces deux définitions sont équivalentes. Il est plus commode de définir $\nu_{N,m}$ comme étant de potentiel $\sum_{i=1}^N V(x_i)$ pour la mesure de Hausdorff $\mathcal{H}^{N-1}$ sur l'hyperplan affine $X_{N,m}$. C'est la définition choisie dans [GORV05]. La mesure $\mathcal{H}^{N-1}$ est la trace de $\mathcal{L}^N$ sur $X_{N,m}$.

La question que l'on se pose est la suivante : existe-t-il une constante $c > 0$ finie telle que pour tout N, tout m, et toute fonction lisse $f : \mathbb{R}^N \to \mathbb{R}_+$ vérifiant $\mathrm{div} f = 0$,

$$\mathbf{E}_{\nu_{N,m}}\left(f \log \frac{f}{\mathbf{E}_{\nu_{N,m}}(f)}\right) \leqslant c\, \mathbf{E}_{\nu_{N,m}}\left(\frac{|\nabla f|^2}{f}\right) ? \tag{5}$$

où $|\nabla f|^2 := \sum_{i=1}^N (\partial_i f)^2$. La condition $\mathrm{div} f := \sum_{i=1}^N \partial_i f = 0$ correspond à dire que le vecteur ∇f est orthogonal à $(1, \ldots, 1)$, et donc parallèle à $X_{N,m}$. Le point important ici est que la constante c ne dépend ni de N ni de m.

On observera que les mesures μ_N et $\nu_{N,m}$ sont échangeables, c'est-à-dire qu'elles sont invariantes par permutation des coordonnées.

[13]Voir également [BLO97].

3.1 Cas des potentiels convexes

Soit W le potentiel d'une mesure de probabilité sur $\mathbb{R}^N$. Comme $X_{N,m}$ est affine, le théorème de Bakry-Émery reste valable pour les mesures de probabilité de potentiel W pour la mesure $\mathcal{H}^{N-1}$ sur $X_{N,m}$, et conduit à des inégalités du même type que (5). On peut par exemple le vérifier sur les deux approches alternatives évoquées précédement pour la définition de $\nu_{N,m}$.

S'il existe $\rho > 0$ tel que $V''(x) \geqslant \rho > 0$ pour tout $x \in \mathbb{R}$, alors la mesure μ_N, de potentiel $\sum_{i=1}^N V(x_i)$ pour $\mathcal{L}^N$, est log-concave de courbure ρ. Le théorème de Bakry-Émery nous assure alors qu'elle vérifie une inégalité de Sobolev logarithmique de constante $1/(2\rho)$. De même, la mesure $\nu_{N,m}$, qui est de même potentiel, vérifie l'inégalité (5) avec une constante $1/(2\rho)$, qui ne dépend pas de N ni de m.

3.2 Cas des perturbations de potentiels convexes

Que se passe-t-il lorsque V n'est plus uniformément strictement convexe, mais vérifie $\|B\|_\infty < \infty$ où $B(x) := V(x) - x^2/2$ pour tout $x \in \mathbb{R}$? En vertu du théorème de Bakry-Émery pour la gaussienne standard sur $\mathbb{R}$, du théorème de Holley-Stroock, et du théorème de tensorisation de Gross-Segal, on obtient que la mesure produit μ_N satisfait à une inégalité de Sobolev logarithmique de constante $e^{2\mathrm{osc}(B)}/2$, qui ne dépend pas de N. Il est frappant de constater que l'application du théorème de Holley-Stroock après la tensorisation aurait donné une constante $e^{2N\mathrm{osc}(B)}/2$ qui dépend de N. Que se passe-t-il pour $\nu_{N,m}$? Nous ne pouvons qu'affirmer que $\nu_{N,m}$ vérifie (5) avec une constante $e^{2N\mathrm{osc}(B)}/2$ qui dépend de N.

Le but de ce qui suit est de montrer que si de plus $\|B''\|_\infty < \infty$, alors $\nu_{N,m}$ vérifie (5) avec une constante qui ne dépend ni N ni de m.

3.2.1 Recentrage et énergie libre

La mesure $\nu_{N,m}$ est insensible aux perturbations linéaires additive sur le potentiel V. Plus précisément, pour tout $\lambda \in \mathbb{R}$, on considère la mesure de probabilité μ_λ de potentiel $x \mapsto V(x) - \lambda x$ pour $\mathcal{L}$. On a donc $\mu_0 = \mu$. Soit $m \in \mathbb{R}$. La mesure conditionnelle $\mu_\lambda^{\otimes}(\cdot \,|\, \sigma_N = m)$ ne dépend pas de λ, et on la notera donc encore $\nu_{N,m}$. Sous $\nu_{N,m}$, les fonctions coordonnées ont pour moyenne commune m. Choisissons $\lambda = \lambda_m$ de sorte que μ soit de moyenne m. L'équation vérifier par λ_m s'écrit :

$$\int_{\mathbb{R}} x d\mu_\lambda(x) = m.$$

Cela donne $\partial_\lambda \psi(\lambda_m) = m$, où $\psi : \mathbb{R} \to \mathbb{R}$ est définie par

$$\psi(\lambda) := \log \int_{\mathbb{R}} e^{-V(x)+\lambda x}\, dx.$$

C'est le logarithme de la constante de normalisation (appelée fonction de partition) de μ_λ. Un calcul simple montre qu'il existe une constante $c > 0$ (dépendant de B) telle que pour tout $\lambda \in \mathbb{R}$,

$$\partial_\lambda^2 \psi(\lambda) \geqslant c.$$

Ainsi, ψ est uniformément strictement convexe. On a également $\lambda_m = \partial_m \psi^*(m)$ où ψ^* est la transformée de Legendre de ψ, définie par $\psi^*(m) = \sup_{\lambda\in\mathbb{R}}(\lambda m - \psi(\lambda))$ pour tout $m \in \mathbb{R}$. De plus, le couple (m, λ_m) vérifie $\psi(\lambda_m) + \psi^*(m) = m\lambda_m$.

Pour tout $x \in X_{N,m}$, on a $\sum_{i=1}^N \lambda x_i = Nm$. On dispose de l'écriture suivante pour la mesure $\nu_{N,m} = \mu_\lambda^{\otimes N}(\cdot \mid \sigma_N = m)$:

$$\frac{d\nu_{N,m}}{d\mathcal{H}^{N-1}}(x) = \exp\left(N\psi_N(m)\right)\exp\left(-\sum_{i=1}^N V(x_i)\right),$$

où

$$\psi_N(m) := -\frac{1}{N}\int_{X_{N,m}} \exp\left(-\sum_{i=1}^N V(x_i)\right) d\mathcal{H}^{N-1}(x).$$

On appelle *énergie libre* de $\nu_{N,m}$ la quantité suivante :

$$-\frac{1}{N}\int_{X_{N,m}} \exp\left(-\sum_{i=1}^N \left(V(x_i) - \lambda x_i\right)\right) d\mathcal{H}^{N-1}(x) = -\lambda m + \psi_N(m).$$

FIXME: Je finirai plus tard cette rédaction ! 1er décembre 2005, 19:14.

3.2.2 Décomposition en blocs et gros grainage

3.2.3 Estimées liées au théorème central limite local

[GORV05] et [KL99]

3.3 Le cas de la variance et des inégalités de Poincaré

[CCL01] et [Cap03]

3.4 Le cas des modèles à spins discrets

[DPP05], [BCDPP05], et [CP04].

Remarque 3.1. Question de F. Barthe : tu as dis que c'était vrai pour la variance (inégalité de Poincaré) lorsque V est de la forme convexe+borné. Cela repose sur la nature hilbertienne de la variance. Est-t-il possible d'utiliser une représentation de l'énergie libre en terme de supremum de formes quadratiques ? Cette approche est fructueuse pour montrer par exemple la décroissance de l'entropie le long du TCL.

4 Notes bibliographiques

FIXME: mettre ici une présentation de la littérature sur la question.

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.

[Bak94] D. Bakry – « L'hypercontractivité et son utilisation en théorie des semigroupes », Lectures on probability theory (Saint-Flour, 1992), Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, p. 1–114.

[BCDPP05] A.-S. Boudou, P. Caputo, P. Dai Pra et G. Posta – « Spectral gap estimates for interacting particle systems via a Bakry & Emery-type approach », Preprint arXiv math.PR/0505533, 2005.

[Bes05] L. Bessières – « Conjecture de Poincaré: la preuve de R. Hamilton et G. Perelman », *Gazette des Mathématiciens, Société Mathématique de France* (2005), no. 106.

[BLO97] L. Bertini, C. Landim et S. Olla – « Derivation of Cahn-Hilliard equations from Ginzburg-Landau models », *J. Statist. Phys.* **88** (1997), no. 1-2, p. 365–381.

[Bol] F. Bolley – « Autour de l'équation de Boltzmann », Mémoire de Magistère, `http://www.lsp.ups-tlse.fr/Fp/Bolley/`.

[Cap03] P. Caputo – « Uniform Poincaré inequalities for unbounded conservative spin systems: the non-interacting case », *Stochastic Process. Appl.* **106** (2003), no. 2, p. 223–244.

[Car24] S. Carnot – « Reflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance », 1824.

[Cat05] P. Cattiaux – « Hypercontractivity for perturbed diffusion semigroups », Preprint arXiv math.PR/0510258, 2005.

[CCL01] E. Carlen, M. C. Carvalho et M. Loss – « Many-body aspects of approach to equilibrium », Séminaire: Équations aux Dérivées Partielles, 2000–2001, Sémin. Équ. Dériv. Partielles, École Polytech., Palaiseau, 2001, p. Exp. No. XIX, 12.

[Cer02] C. Cercignani – « The Boltzmann equation and fluid dynamics », Handbook of mathematical fluid dynamics, Vol. I, North-Holland, Amsterdam, 2002, p. 1–69.

[Cha03] D. Chafaï – « Glauber versus Kawasaki for spectral gap and logarithmic Sobolev inequalities of some unbounded conservative spin systems », *Markov Process. Related Fields* **9** (2003), no. 3, p. 341–362.

[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.

[CP04] P. Caputo et G. Posta – « Entropy dissipation estimates in a Zero-Range dynamics », Preprint arXiv math.PR/0405455, 2004.

[CT91] T. M. Cover et J. A. Thomas – *Elements of information theory*, Wiley Series in Telecommunications, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication.

[DPP05] P. Dai Pra et G. Posta – « Logarithmic Sobolev inequality for zero-range dynamics: independence of the number of particles », *Electron. J. Probab.* **10** (2005), p. no. 15, 525–576 (electronic).

[DZ98] A. Dembo et O. Zeitouni – *Large deviations techniques and applications*, second éd., Applications of Mathematics (New York), vol. 38, Springer-Verlag, New York, 1998.

[EE90] P. Ehrenfest et T. Ehrenfest – *The conceptual foundations of the statistical approach in mechanics*, english éd., Dover Publications Inc., New York, 1990, Translated from the German by Michael J. Moravcsik, With a foreword by M. Kac and G. E. Uhlenbeck.

[FLP01] V. P. Fonf, J. Lindenstrauss et R. R. Phelps – « Infinite dimensional convexity », Handbook of the geometry of Banach spaces, Vol. I, North-Holland, Amsterdam, 2001, p. 599–670.

[GJ87] J. Glimm et A. Jaffe – *Quantum physics*, second éd., Springer-Verlag, New York, 1987, A functional integral point of view.

[GORV05] N. Grunewald, F. Otto, M. Reznikoff et C. Villani – « A two-scale proof of a logarithmic Sobolev inequality », prépublication, `http://www.math.princeton.edu/~mrezniko/`, 2005.

[Gro75] L. Gross – « Logarithmic Sobolev inequalities », *Amer. J. Math.* **97** (1975), no. 4, p. 1061–1083.

[HUL01] J.-B. Hiriart-Urruty et C. Lemaréchal – *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001, Abridged version of *Convex analysis and minimization algorithms. I* [Springer, Berlin, 1993; MR 95m:90001] and *II* [ibid.; MR 95m:90002].

[Jay89] E. T. Jaynes – *E. T. Jaynes: papers on probability, statistics and statistical physics*, Pallas Paperbacks, vol. 50, Kluwer Academic Publishers Group, Dordrecht, 1989, Reprint of the 1983 original, Edited and with an introduction by R. D. Rosenkrantz.

[JL92] B. Jørgensen et R. S. Labouriau – *Famílias exponenciais e inferência teórica*, Monografías de Matemática [Mathematical Monographs], vol. 52, Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 1992, With the collaboration of José Raúl Martínez.

[KL99] C. Kipnis et C. Landim – *Scaling limits of interacting particle systems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 320, Springer-Verlag, Berlin, 1999.

[KS97] U. Küchler et M. Sørensen – *Exponential families of stochastic processes*, Springer Series in Statistics, Springer-Verlag, New York, 1997.

[Led00] M. Ledoux – « The geometry of Markov diffusion generators », *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000), no. 2, p. 305–366, Probability theory.

[Let92] G. Letac – *Lectures on natural exponential families and their variance functions*, Monografías de Matemática [Mathematical Monographs], vol. 50, Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 1992.

[LNN05] C. Landim et J. Noronha Neto – « Poincaré and logarithmic Sobolev inequality for Ginzburg-Landau processes in random environment », *Probab. Theory Related Fields* **131** (2005), no. 2, p. 229–260.

[LPY02] C. Landim, G. Panizo et H. T. Yau – « Spectral gap and logarithmic Sobolev inequality for unbounded conservative spin systems », *Ann. Inst. H. Poincaré Probab. Statist.* **38** (2002), no. 5, p. 739–777.

[Rob01] C. P. Robert – *The Bayesian choice*, second éd., Springer Texts in Statistics, Springer-Verlag, New York, 2001, From decision-theoretic foundations to computational implementation, Translated and revised from the French original by the author.

[Roy99] G. Royer – *Une initiation aux inégalités de Sobolev logarithmiques*, Cours Spécialisés [Specialized Courses], vol. 5, Société Mathématique de France, Paris, 1999.

[SPM05] SPM – « Spécial année mondiale de la physique », Images de la physique, Publication annuelle du département SPM du Centre National de la Recherche Scientifique, 2005, `http://www.spm.cnrs-dir.fr/actions/publications/IdP.htm`.

[Vil02] C. Villani – « A review of mathematical topics in collisional kinetic theory », Handbook of mathematical fluid dynamics, Vol. I, North-Holland, Amsterdam, 2002, `http://www.umpa.ens-lyon.fr/~cvillani/`, p. 71–305.

[Vil03] — , *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

[Zin96] M. Zinsmeister – *Formalisme thermodynamique et systèmes dynamiques holomorphes*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 4, Société Mathématique de France, Paris, 1996.

Compilé le 1er décembre 2005 avec pdfLaTeX.