Corrigé

Jean Diebolt

jean.dieb@wanadoo.fr

Introduction & préliminaires

Ecarts d_{VT} et d_2

On introduit dans ce problème plusieurs notions d'écart entre densités (q et r). L'écart entropique $\mathrm{Ent}_r\,(q/r)$ est lié en particulier à la théorie de l'information. Deux autres écarts, celui de la variation totale $d_{\mathrm{VT}}(q,\,r)$ (introduit en I.3) et $d_2(q|\,r)$ (introduit en II.4 pour $n=1$ et en IV pour $n \geq 1$), sont attachés à une intuition probabiliste qu'il n'est pas du tout nécessaire de maîtriser pour traiter le problème.

Notons tout d'abord que d'après I.3, $d_{\mathrm{VT}}(q,\,r)$ est, à un coefficient multiplicatif près, la distance L^1 entre q et r.

D'autre part, à toute mesure de probabilité γ sur $[0,\,1]^n$ correspond une forme linéaire positive Γ sur l'espace $\mathbb{B}_n$, telle que $\Gamma(\mathbf{1}) = 1$, avec

$$\forall f \in \mathbb{B}_n, \quad \Gamma(f) \,=\, \int_{[0,\,1]^n} f\,d\gamma\,.$$

De même, l'ensemble des mesures de probabilité sur $[0,\,1]^n \times [0,\,1]^n$ correspond à l'ensemble $\mathcal{L}_{n\times n}$. Si π est une mesure de probabilité sur $[0,\,1]^n \times [0,\,1]^n$, Π associée appartient à $\mathcal{L}_{n\times n}$, et

$$\forall h \in \mathbb{B}_{n\times n}, \quad \Pi(f) \,=\, \int_{[0,\,1]^n\times[0,\,1]^n} h\,d\pi\,.$$

Les formes Π_1 et Π_2 correspondent alors aux mesures de probabilité marginales de π.

Exprimons maintenant cela en termes de variables aléatoires. Soit X une variable aléatoire réelle de loi γ : l'espérance de $f(X)$ vérifie

$$\mathbb{E}(f(X)) \,=\, \int_{[0,\,1]^n} f\,d\gamma \,=\, \Gamma(f)\,.$$

De même, si $(X,\,Y)$ est un couple de variables aléatoires réelles de loi π, on a

$$\mathbb{E}(h(X,\,Y)) \,=\, \int_{[0,\,1]^n\times[0,\,1]^n} h\,d\pi \,=\, \Pi(f)\,.$$

La mesure de probabilité λ associée à la forme linéaire Λ charge la diagonale de $[0,\,1]^n \times [0,\,1]^n$ dès que la fonction ϕ n'est pas la fonction nulle. La mesure de la diagonale, c'est-à-dire la probabilité que $X = Y$, est égale à

$$\int_{[0,\,1]^n} \phi(x)\,dx\ .$$

Pour tout couple $(X,\,Y)$ de variables aléatoires dont la loi appartient à $\mathcal{L}(q,\,r)$, $d_{\mathrm{VT}}(q,\,r)$ est la borne supérieure de l'espérance de $|f(X) - f(Y)|$ sur un ensemble de fonctions f “pas trop grandes”.

La question II.3 montre que pour toute mesure de probabilité π de $\mathcal{L}(q, r)$, si (X, Y) désigne un couple de variables aléatoires de loi π on a

$$d_{\mathrm{VT}}(q, r) \leq \mathbb{E}(\mathbf{1}_{X \neq Y}) = \mathbb{P}\{X \neq Y\} .$$

Il existe une mesure de probabilité explicite λ de $\mathcal{L}(q, r)$, définie par (13), pour laquelle $d_{\mathrm{VT}}(q, r) = \mathbb{E}(\mathbf{1}_{X \neq Y}) = 1 - \mathbb{P}\{\mathbf{1}_{X=Y}\}$.

Autrement dit, étant données deux densités fixées q et r sur $[0, 1]^n$, si on condidère des couples de variables aléatoires réelles (X, Y) non indépendantes en général, la densité de la loi de X étant q et celle de la loi de Y étant r, $d_{\mathrm{VT}}(q, r)$ est la borne inférieure parmi tous les couples (X, Y) considérés ci-dessus de la probabilité que $X \neq Y$, c'est-à-dire la probabilité que $X \neq Y$ lorsque X et Y sont "le plus proches" possible.

Quant à l'écart $d_2(q|\, r)$, il a été construit par analogie avec ce qu'on vient d'expliquer pour d_{VT}. Dans le cas particulier $n = 1$, c'est la borne inférieure, parmi tous les couples (X, Y) considérés ci-dessus, de la borne supérieure de l'espérance de $\alpha(Y)\mathbf{1}_{X \neq Y}$ sur un ensemble de fonctions α "pas trop grandes". Pour $d_2(r|\, q)$, on remplace $\alpha(Y)\mathbf{1}_{X \neq Y}$ par $\beta(X)\mathbf{1}_{X \neq Y}$. Cette interprétation s'adapte au cas général $n \geq 1$, moyennant l'introduction de davantage de notations.

Inégalités de Sobolev logarithmiques de IV

Il n'est pas possible d'expliquer ici ce que sont en général des inégalités du type de Sobolev. Nous nous contenterons de remarquer que comme r est une densité, l'hypothèse de continuité de f sur tout $[0, 1]^n$ n'est en fait pas nécessaire. De même, il n'est pas nécessaire que ∇f soit continue sur $]0, 1[^n$, ni (sauf exception) même bornée sur $[0, 1]^n$. Il suffit que les intégrales apparaissant dans les seconds membres de ces inégalités soient finies. En particulier, les fonctions $\exp(f)\, r$ et $\|\nabla f\|^2\, r$ doivent être intégrables.

Inégalité de Poincaré de IV.3

Cette inégalité exprime que sous certaines conditions sur f, si X désigne une variable aléatoire à valeurs dans $[0, 1]^n$ dont la loi admet la densité r, alors la variance de la variable aléatoire réelle $f(X)$ est majorée par C fois l'espérance du carré de la norme euclidienne de $\nabla f(X)$. Notons l'homogénéité des deux membres de l'inégalité. Dans le second membre, l'intégrale $\displaystyle\int \|\nabla f(x)\|^2\, r(x)\, dx$ peut être interprétée comme une mesure différentielle de dispersion quadratique de $f(X)$. Des inégalités de ce type sont, entre autres choses, des outils pour établir des résultats dits "isopérimétriques".

Inégalités de concentration de V

Elles sont entièrement explicites, elles sont vraies pour chaque $n \geq 1$, et elles sont beaucoup plus précises que les inégalités usuelles comme celle de Chebyshev. Ce sont des "inégalités de moyenne déviation".

Liens

Des mathématiciens sont actuellement en train de dégager des relations étroites entre inégalités de Sobolev logarithmiques d'une part, inégalités de Poincaré et résultats isopérimétriques d'autre part, et enfin inégalités de concentration. L'énoncé de ce problème concerne des variables aléatoires indépendantes bornées, prenant toutes leurs valeurs dans le même intervalle compact. Il existe d'autres cadres possibles pour obtenir des résultats analogues, en particulier celui des variables aléatoires gaussiennes, ou au moins admettant des propriétés "proches" de celles des variables aléatoires gaussiennes.

I

I.1 Entropie positive

a) et *b)* Inégalité de Jensen (voir Préliminaires) appliquée à la fonction convexe $t > 0 \mapsto t \ln t$. Cette fonction est convexe car deux fois dérivable de dérivée seconde positive sur $]0, +\infty[$.

I.2 Inégalités auxiliaires

a) b) c) et *d)* Par exemple, on montre en dérivant deux fois que $0 \leq t < 1 \mapsto d(t) = J(1-t) - t^2/2$ est convexe. Comme $d'(0) = 0$, il en résulte que $d'(t) \geq 0$, donc d est croissante à partir de 0, donc ≥ 0. Pour K, c'est analogue ; on peut poser $u = 1 + t$, $t \geq 0$, puis $s = t/(1+t)$ pour se ramener à $[0, 1]$. On peut aussi utiliser la formule de Taylor-Lagrange à l'ordre 2. Pour I.2.c, application directe, car (avec la convention en 0)

$$u \ln u - u + 1 = J(u)\mathbf{1}_{0 \leq u \leq 1} + uK(u)\mathbf{1}_{u>1}\,.$$

La question I.2.d est immédiate (destinée en particulier à obliger les candidats à *lire* l'énoncé et à préparer la suite). On trouve que $\mathrm{Ent}_g(t\,m) = t\,\mathrm{Ent}_g(m)$, ce qui servira plus loin.

Commentaire. *Pour les questions a) et b), un tableau de variations peut êtreconseillé.*

e) Appliquer I.2.c avec $u = q(x)/r(x)$ et intégrer par rapport à $r(x)\,dx$.

I.3 Formule pour d_{VT}

a) et *b)* On peut intégrer sur les ensembles $\{x \ : q(x) > r(x)\}$ et $\{x \ : q(x) < r(x)\}$, que nous noterons simplement $\{q > r\}$ et $\{q < r\}$, puis utiliser le fait que $\displaystyle\int q(x)\,dx = \int r(x)\,dx$, d'où $\displaystyle\int \mathbf{1}_{\{q>r\}}\,(q(x) - r(x))\,dx + \int \mathbf{1}_{\{q<r\}}\,(q(x) - r(x))\,dx = 0$, d'où a) par définition de $[q - r]_+$.

Ensuite, il faut d'abord se débarrasser de la valeur absolue. On doit montrer que

$$\sup_{0 \leq f \leq 1} \left| \int f(x)q(x)\,dx - \int f(y)r(y)\,dy \right| = \sup_{0 \leq f \leq 1} \left\{ \int f(x)q(x)\,dx - \int f(y)r(y)\,dy \right\}.$$

En effet, par exemple (entre autres possibilités), pour toute fonction f telle que $0 \leq f \leq 1$, on a

$$\int f(x)(q(x) - r(x))\,dx = -\int (1 - f(x))(q(x) - r(x))\,dx$$

puisque $\displaystyle\int (q(x) - r(x))\,dx = 0$. On a $0 \leq 1 - f \leq 1$ et $f \mapsto 1 - f$ est une involution pour les fonctions telles que $0 \leq f \leq 1$.

Puis on montre que la borne supérieure est atteinte pour $f = \mathbf{1}_{q \geq r}$, d'où b) en revenant à la définition de $[q - r]_+$.

II

II.1 Unicité ?

Étant des densités, les fonctions ℓ_1 et ℓ_2, k_1 et k_2, ne sont définies qu'à un ensemble de mesure nulle près. Le plus simple ici est peut-être de déduire des définitions que $d_{\mathrm{VT}}(\ell_1, k_1) = d_{\mathrm{VT}}(\ell_2, k_2) = 0$. Par conséquent, $\displaystyle\int |\ell_1(x) - k_1(x)|\,dx = \int |\ell_2(x) - k_2(x)|\,dx = 0$, d'où la conclusion.

On peut aussi, dans le cas $n = 1$ ici, considérer des fonctions f et g de la forme $\mathbf{1}_{[0,\,t]}$ pour $t \in [0,\,1]$. On utilise alors le résultat suivant : si ℓ est intégrable sur $[0,\,1]$, alors la fonction $t \in [0,\,1] \mapsto \displaystyle\int_0^t \ell(x)\,dx$ est dérivable presque partout et chacune des dérivées possibles coïncide presque partout avec $\ell(t)$.

II.2 Étude d'un cas particulier

a) et *b)*

Conditions : $\phi(x)$ et $\psi(x,\,y) \geq 0$ presque partout, et $\displaystyle\int \phi(x)\,dx + \int \psi(x,\,y)\,dxdy = 1$. (Il faut démontrer les implications dans les deux sens.)

Densités marginales de Λ : dans ce cas, en prenant $h(x,\,y) = f(x)$ on trouve $\ell_1(x) = \phi(x) + \displaystyle\int \psi(x,\,y)\,dy$ presque partout. Formule analogue pour ℓ_2.

II.3 Interprétation variationnelle de d_{VT} (pts)

a) Soit g une fonction bornée. Soient $h(x,\,y) = g(x) - g(y)$ et $k(x,\,y) = |g(x) - g(y)|$.

Par hypothèse, $\Pi(h) = \displaystyle\int g(x)q(x)\,dx - \int g(y)r(y)\,dy$.

Comme Π est une forme positive, on a l'inégalité

$$\int g(x)q(x)\,dx - \int g(y)r(y)\,dy \leq \Pi(k) = \Pi\left(k\mathbf{1}_{x\neq y}\right)\,,$$

et on sait que $\forall x,\,y,\,|g(x) - g(y)| \leq 1$ lorsque $0 \leq g \leq 1$. Ceci étant vrai pour toute fonction g telle que $0 \leq g \leq 1$, l'inégalité voulue en résulte par passage au sup dans les inégalités.

b) On suppose $d_{\mathrm{VT}}(q,\,r) > 0$. Il faut d'abord vérifier que Λ_0 est une forme linéaire positive telle que $\Lambda_0(\mathbf{1}) = 1$. On utilise pour cela l'identité $\min(q,\,r) + [q-r]_+ = q$. Finalement, d'après II.2 on obtient :

$$\ell_1 = \min(q,\,r) + [q-r]_+ = q \quad \text{et} \quad \ell_2 = \min(q,\,r) + [r-q]_+ = r\,.$$

c) On suppose toujours que $d_{\mathrm{VT}}(q,\,r) > 0$. On vérifie que $\Lambda_0(\mathbf{1}_{x\neq y}) = d_{\mathrm{VT}}(q,\,r)$. Ainsi $d_{\mathrm{VT}}(q,\,r)$ est la borne inférieure (atteinte) des $\Lambda(\mathbf{1}_{x\neq y})$ pour $\Pi \in \mathcal{L}(q,\,r)$.

COMMENTAIRE. *On dit que la mesure de probabilité Λ_0 est un couplage des mesures de probabilité de densités $\ell_1 = q$ et $\ell_2 = r$. Il existe un couplage optimal, défini ci-dessus.*

II.4 Démonstration, cas $n = 1$

Revenons à la définition de d_2. On part du fait que pour majorer par A la borne inférieure d'une partie non vide de $\mathbb{R}$, il suffit de majorer par A un élément de cette partie. Ici, on choisit Λ_0.

Avec un abus de notation,

$$\Lambda_0\left(\alpha(y)\,\mathbf{1}_{x\neq y}\right) = \int \alpha(y)\ [r(y)\ -\ q(y)]_+\ dy = \int \alpha(y)\,\sqrt{r(y)}\left[1\ -\ \frac{q(y)}{r(y)}\right]_+ \sqrt{r(y)}\,dy\,.$$

On utilise alors l'inégalité de Cauchy-Schwarz et l'hypothèse $\displaystyle\int \alpha^2(y)r(y)\,dy \leq 1$, puis on revient à la question I.2.e. Enfin, on passe à la borne supérieure en les fonctions α telles que $\displaystyle\int \alpha^2(y)r(y)\,dy \leq 1$.

Démarche analogue avec $\beta(x)$. Remarquons qu'on aurait aussi pu utiliser l'inégalité de Cauchy-Schwarz directement dans l'espace $L^2([0,1],\ \sqrt{r(y)}\,dy)$.

COMMENTAIRE. *Le couplage optimal de II.3 est utilisé ici comme un "bon" couplage (on ne sait pas s'il est optimal ici) qui donne une bonne majoration de* d_2.

III

III.1 Une inégalité pour des sommes de variables aléatoires

a) On développe $\exp(\lambda^2/2)$ et $\mathrm{ch}\lambda$ en séries entières. On montre, directement ou par récurrence, que les coefficients du développement de $\exp(\lambda^2/2)$ sont $\geq$ ceux de $\mathrm{ch}\lambda$.

b) Ici encore, plusieurs possibilités. Nous en proposons trois.

Possibilité 1 : Utilisation directe de III.1.a et de la convexité de la fonction exponentielle afin de montrer que $e^{\lambda x} \leq \mathrm{ch}\lambda + x\,\mathrm{sh}\lambda$ pour les valeurs de x considérées, $x \in [-1,\ 1]$. En effet,

$$\mathrm{ch}\lambda\ +\ x\,\mathrm{sh}\lambda\ =\ \left(\frac{1+x}{2}\right)e^{\lambda}\ +\ \left(\frac{1-x}{2}\right)e^{-\lambda},$$

avec

$$\frac{1+x}{2} \geq 0,\quad \frac{1-x}{2} \geq 0,\quad \frac{1+x}{2} + \frac{1-x}{2} = 1\,.$$

Par convexité, ceci est $\geq \exp\left((1+x)\lambda/2\ -\ (1-x)\lambda/2\right)$, qui est égal à $\exp(\lambda x)$.

COMMENTAIRE. *Une figure est conseillée.*

Possibilité 2 : On fixe $\lambda > 0$ et on étudie la fonction $\phi(x) = \mathrm{ch}\lambda + x\,\mathrm{sh}\lambda - \exp(\lambda x)$ pour $-1 \leq x \leq 1$. Elle est concave et s'annule en $x = \pm 1$, d'où son signe sur $[-1,\ 1]$.

COMMENTAIRE. *Une figure est conseillée.*

Possibilité 3 : On cherche à démontrer directement que

$$e^{\lambda x}\ \leq\ \exp\left(\frac{\lambda^2}{2}\right)\ +\ x\,\mathrm{sh}\lambda\,.$$

Considérons la fonction $\phi(x,\ \lambda) = \exp(\lambda^2/2) + x\,\mathrm{sh}\lambda - \exp(\lambda x)$ pour $\lambda \geq 0$. On a $\partial\phi/\partial\lambda = \lambda\exp(\lambda^2/2) + x\,\mathrm{ch}\lambda - x\exp(\lambda x)$. Comme $\forall x,\ \phi(x,\ 0) = 0$, il résulte d'inégalités analogues à celles de a) que $\forall\lambda \geq 0$, $\phi(-1,\ \lambda)$ et $\phi(1,\ \lambda)$ sont positifs. Enfin, la dérivée partielle seconde $\partial^2\phi/\partial x^2 = -\lambda^2\exp(\lambda x)$ est strictement négative pour $\lambda > 0$, donc à x fixé $\lambda > 0 \mapsto \phi(x,\ \lambda)$ est strictement concave. Comme elle prend des valeurs positives aux extrémités de l'intervalle $[-1,\ 1]$, il en résulte qu'elle prend des valeurs positives sur tout $[-1,\ 1]$, d'où le résultat.

COMMENTAIRE. *Une figure est conseillée.*

c), d) et *e)* Le synopsis est le suivant. Ici, on ne suppose pas que la loi de X_i admet une densité par rapport à la mesure de Lebesgue. Parce que $\exp(u+v) = \exp(u)\exp(v)$ et par indépendance, tout revient à montrer que $E(\exp(\lambda X)) \leq \exp(\lambda^2/2)$ pour tout $\lambda > 0$. On utilise l'inégalité admise et les hypothèses. On établit ensuite la même inégalité pour $-X$. On conclut en considérant la probabilité de la réunion de deux événements, d'où le facteur 2. Entrons dans les détails.

c) Par linéarité et monotonie, on déduit de III.1.b que

$$\mathbb{E}\left(e^{\lambda X}\right)\ \leq\ \exp\left(\frac{\lambda^2}{2}\right),$$

puisque $\mathbb{E}(X) = 0$. De même pour $-X$.

COMMENTAIRE. *Dans le cas où la loi de X admet une densité r, on a* $\mathbb{E}(X) = \int_{-1}^{1} t\, r(t)\, dt$ *et* $\mathbb{E}\left(e^{\lambda X}\right) = \int_{-1}^{1} e^{\lambda t}\, r(t)\, dt$.

d) Montrons d'abord que

$$\mathbb{P}\{X \geq a\} \leq \exp\left(-\frac{a^2}{2}\right) \quad \text{pour tout } a > 0\,.$$

D'après l'inégalité de Chernoff (voir le début de la partie III) combinée avec III.2.c, on a, pour tout $\lambda > 0$,

$$\mathbb{P}\{X \geq a\} \leq \exp\left(-\left(\lambda a - \frac{\lambda^2}{2}\right)\right).$$

Le trinôme du second degré $-\lambda^2/2 + \lambda a$ atteint son maximum en $\lambda = a$, où il vaut $a^2/2$. On en déduit l'inégalité ci-dessus.

De plus, $\mathbb{P}\{-X \geq a\}$ vérifie la même inégalité.

La probabilité de la réunion de deux événements est inférieure ou égale à la somme des probabilités de ces deux événements. (Lorsque l'intersection de ces deux événements est de probabilité nulle, il y a égalité.) Puisque pour $a > 0$,

$$\{|X| \geq a\} = \{X \geq a\} \cup \{-X \geq a\}\,,$$

l'inégalité cherchée en résulte, avec son facteur 2.

REMARQUE. *Si $a > 1$, alors $\mathbb{P}(|X| \geq a) = 0$ et si $a \leq 1$, alors le terme de droite de l'inégalité (20) est $\geq 2\exp(-1/2) > 1$. Ceci permet de démontrer l'inégalité (20) directement, sans l'hypothèse que X est centrée ! Par contre, on ne peut pas démontrer ainsi que*

$$\mathbb{P}\{X \geq a\} \leq \exp\left(-\frac{a^2}{2}\right) \quad \textit{pour tout } a \geq 0\,.$$

En effet, $\exp(-1/2) < 1$. Cependant, si on supposait de plus que X est symétrique (donc que $\forall a$, $\mathbb{P}\{X \geq a\} = \mathbb{P}\{-X \geq a\}$) alors on obtiendrait dans ce cas que $\mathbb{P}\{X \geq a\} \leq \exp(-1/2)\exp(-a^2/2)$ pour tout $a \geq 0$.

e) On pourrait songer à appliquer l'inégalité que nous venons d'établir à la variable aléatoire centrée $Z = n^{-1/2} \sum_{i=1}^{n} X_i$, mais elle prend ses valeurs dans l'intervalle $[-\sqrt{n}, \sqrt{n}]$ et ne vérifie donc pas les hypothèses de III.1.d. Il faut s'y prendre un peu autrement.

En vue d'appliquer à nouveau l'inégalité de Chernoff, calculons $\mathbb{E}\left(e^{\lambda Z}\right)$. Par indépendance des variables aléatoires X_i et d'après III.1.c, on obtient

$$\mathbb{E}\left(e^{\lambda Z}\right) = \mathbb{E}\left(\prod_{i=1}^{n} e^{\lambda n^{-1/2} X_i}\right) = \prod_{i=1}^{n} \mathbb{E}\left(e^{\lambda n^{-1/2} X_i}\right) \leq \prod_{i=1}^{n} \exp\left(\frac{\lambda^2}{2n}\right) = \exp\left(\frac{\lambda^2}{2}\right).$$

On peut alors utiliser le même raisonnement que dans la question III.1.d, d'où la conclusion.

III.2 Optimalité ?

a) Équation différentielle linéaire homogène d'ordre 1. Facteur intégrant $\exp(x^2/2)$.

b) On peut utiliser une intégration par parties (mais d'autres solutions sont existent). On a

$$1 - \Phi(x) = \int_x^{+\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt = -\int_x^{+\infty} \frac{-t\,e^{-t^2/2}}{t\sqrt{2\pi}}\, dt\,.$$

Or, la fonction $-t\,e^{-t^2/2}$ est la dérivée de la fonction $e^{-t^2/2}$. On obtient donc

$$-\int_x^{+\infty} \frac{-t\,e^{-t^2/2}}{t\sqrt{2\pi}}\, dt = \frac{e^{-x^2/2}}{x\sqrt{2\pi}} - \int_x^{+\infty} \frac{e^{-t^2/2}}{t^2\sqrt{2\pi}}\, dt\,.$$

Il reste à justifier le fait que l'intégrale dans le second membre est

$$o\left(\int_x^{+\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt\right)$$

quand $x \to +\infty$. Cela résulte d'un théorème du cours.

c) C'est impossible. En effet, considérons le cas de variables aléatoires X_i centrées, indépendantes et de même loi, prenant chacune les valeurs -1 et 1 avec probabilité $1/2$. Leur variance commune est 1.

D'après le théorème de la limite centrale qui s'applique ici (il faut vérifier les hypothèses) on a, pour tout $a \geq 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|n^{-1/2}\sum_{i=1}^{n} X_i\right| \geq a\right\} = 2\,(1 - \Phi(a))\,.$$

Si l'inégalité était vraie $\forall n$, alors $\forall a \geq 0$, la limite ci-dessus serait $\leq A\exp(-\kappa a^2)$ avec $\kappa > 1/2$, par passage à la limite dans les inégalités.

Or d'après b), $1 - \Phi(a) \sim \varphi(a)/a$ quand $a \to +\infty$. Puisque le coefficient de a^2 de l'exponentielle dans $\varphi(a)$ est $-1/2$, on obtiendrait une contradiction.

IV

IV.1 Inégalités pour des fonctions convexes ou concaves sur $[0,\,1]^n$

a) Traitons d'abord le cas où $n = 1$ et f est convexe. On peut, par exemple, exprimer $f(x) - f(y)$ en utilisant la formule de Taylor-Lagrange (ici, accroissements finis) ou celle de Taylor avec reste intégral.

Supposons d'abord que $x > y$. On a alors, puisque f' est croissante et que $0 < x - y \leq 1$,

$$f(x) - f(y) = \int_y^x f'(t)\, dt \leq \int_y^x f'(x)\, dt = (x - y)f'(x) \leq \left|f'(x)\right| \mathbf{1}_{x\neq y}\,.$$

De même, si $x < y$, on obtient $f(y) - f(x) \geq (y - x)f'(x)$, et on multiplie les deux membres par -1. Soulignons l'importance de l'hypothèse $0 \leq x,\, y \leq 1$.

b) Si $n \geq 1$, on applique le même raisonnement à $0 \leq t \leq 1 \mapsto f(tx + (1-t)y)$, qui est convexe (respectivement, concave) si f l'est, et dont la dérivée est

$$\frac{d}{dt} f\,(tx + (1-t)y) = \sum_{i=1}^{n} (x_i - y_i)\,\partial_i f\,(tx + (1-t)y)\,.$$

COMMENTAIRE. *Ces raisonnements reflètent directement le fait que si f est convexe et C^1, alors son graphe est situé au-dessus de tous ses espaces tangents. Un croquis est conseillé.*

IV.2 Résultats intermédiaires

a) Supposons que $\int \|\nabla f(x)\|^2 q(x)\,dx > 0$. Si $\Pi \in \mathcal{L}(q,\,r)$, alors, avec un abus de notation,

$$\begin{aligned}\int f(x)q(x)\,dx - \int f(y)r(y)\,dy &= \Pi\,(f(x) - f(y)) \le \Pi\left(\sum_{i=1}^n |\partial_i f(x)|\,\mathbf{1}_{x_i\neq y_i}\right)\\ &\le \left(\int \|\nabla f(x)\|^2\,q(x)\,dx\right)^{1/2} \Pi\left(\sum_{i=1}^n \beta_i(x)\,\mathbf{1}_{x_i\neq y_i}\right),\end{aligned}$$

où $\beta_i = |\partial_i f| / \left(\int \|\nabla f(x)\|^2 q(x)\,dx\right)^{1/2}$. On passe en $\sup\limits_{\beta}$ puis en $\inf\limits_{\Pi}$.

b) Même principe. Supposons $\int \|\nabla f(y)\|^2\,r(y)\,dy > 0$. Si $\Pi \in \mathcal{L}(q,\,r)$, on a

$$\begin{aligned}\int f(x)q(x)\,dx - \int f(y)r(y)\,dy &\le \Pi\left(\sum_{i=1}^n |\partial_i f(y)|\,\mathbf{1}_{x_i\neq y_i}\right)\\ &= \left(\int \|\nabla f(y)\|^2\,r(y)\,dy\right)^{1/2} \Pi\left(\sum_{i=1}^n \alpha_i(y)\,\mathbf{1}_{x_i\neq y_i}\right),\end{aligned}$$

où $\alpha_i = |\partial_i f| / \left(\int \|\nabla f(y)\|^2 r(y)\,dy\right)^{1/2}$.

IV.3 Deux inégalités de Sobolev logarithmiques

a) On explicite l'entropie dans ce cas, en utilisant le fait que pour tout réel positif t, on a la relation $\mathrm{Ent}_g(t\,m) = t\,\mathrm{Ent}_g(m)$ (voir question I.2.d) :

$$\begin{aligned}\mathrm{Ent}_r\left(m^f\right) &= \mathrm{Ent}_r\left(\frac{e^f}{R\left(e^f\right)}\right) = \frac{1}{R\left(e^f\right)}\,\mathrm{Ent}_r\left(e^f\right)\\ &= \frac{1}{R\left(e^f\right)}\left[\int f(y)\,e^{f(y)}\,r(y)\,dy - \left(\int e^{f(y)}\,r(y)\,dy\right)\left(\ln\int e^{f(y)}\,r(y)\,dy\right)\right]\\ &= \int f(y)\,r^f(y)\,dy - \ln\int e^{f(y)}\,r(y)\,dy,\end{aligned}$$

puisque $R\left(e^f\right) = \int e^{f(y)}\,r(y)\,dy$. Or, l'inégalité de Jensen appliquée à la fonction logarithme (ou exponentielle) implique que

$$\ln\int e^{f(y)}\,r(y)\,dy \ge \int f(y)r(y)\,dy,$$

d'où la conclusion.

b) On utilise le résultat de IV.2.a, avec $q = r^f$:

$$\int f(x)\,r^f(x)\,dx - \int f(y)\,r(y)\,dy \le \sqrt{2\,I_{r^f}(f)\,\mathrm{Ent}_{r^f}\left(\frac{r^f}{r}\right)}.$$

Or

$$I_{r^f}(f) \;=\; \int \|\nabla f(y)\|^2\, r^f(y)\; dy \;=\; \frac{\int \|\nabla f(y)\|^2\, e^{f(y)}\, r(y)\; dy}{\int e^{f(y)}\, r(y)\; dy}\,.$$

On obtient ainsi

$$\mathrm{Ent}_r\left(m^f\right) \;\leq\; \sqrt{2\, I_{r^f}(f)\ \mathrm{Ent}_r\left(\frac{r^f}{r}\right)}\,,$$

avec

$$\mathrm{Ent}_r\left(\frac{r^f}{r}\right) \;=\; \mathrm{Ent}_r\left(m^f\right)\,.$$

Finalement, en élevant au carré les deux membres de l'inégalité ci-dessus, puis en les divisant par la quantité positive $\mathrm{Ent}_r\left(m^f\right)$, on parvient à

$$\mathrm{Ent}_r\left(m^f\right) \;\leq\; 2\,\frac{\int \|\nabla f(y)\|^2\, e^{f(y)}\, r(y)\; dy}{R\left(e^f\right)}\,,$$

d'où le résultat en multipliant les deux membres de l'inégalité par la quantité positive $R\left(e^f\right)$, et en exploitant la relation $R\left(e^f\right)\mathrm{Ent}_r\left(m^f\right) = \mathrm{Ent}_r\left(e^f\right)$.

c) Même principe, avec $d_2(r^f\,|\,r) \leq \sqrt{2\,\mathrm{Ent}_r(m^f)}$.

IV.4 Une inégalité de Poincaré

Supposons f convexe, sinon on remplace par exemple f par $-f$ (ou bien on reprend le raisonnement à partir de l'inégalité relative au cas concave).

1. Il suffit de se ramener au cas où $R(f) = 0$, quitte à remplacer f par $f - R(f)$. En effet (formule de Huyghens)

$$\int (f(x) \;-\; R(f))^2\, r(x)\, dx \;=\; \int f(x)^2\, r(x)\, dx \;-\; (R(f))^2\,.$$

2. L'idée générale est de remplacer f par εf ($\varepsilon > 0$ petit), puis de faire des développements limités en ε quand $\varepsilon \to 0$. Il s'agit alors de justifier les convergences d'intégrales dépendant du paramètre ε, quand $\varepsilon \to 0$. Pour cela, on utilise (par exemple) le théorème de convergence dominée de Lebesgue (TCD), ce qui est rendu possible du fait qu'en raison des hypothèses, f est continue sur le compact $[0,1]^n$, donc bornée, et que $\|\nabla f\|$ est bornée.

3. Remplaçons $\mathrm{Ent}_r(e^{\varepsilon f})$ par son expression. Après simplification,

$$\varepsilon \int f(x)\, e^{\varepsilon f(x)} r(x)\, dx \;\leq\; 2\varepsilon^2 \int \|\nabla f(x)\|^2 e^{\varepsilon f(x)} r(x)\, dx + \left(\int e^{\varepsilon f(x)} r(x)\, dx\right) \ln \int e^{\varepsilon f(x)} r(x)\, dx\,.$$

4. Comme expliqué ci-dessus, $\lim\limits_{\varepsilon\to 0} \int e^{\varepsilon f}\, r = 1$, donc $\lim\limits_{\varepsilon\to 0} \ln \int e^{\varepsilon f}\, r = 0$, tandis que $\lim\limits_{\varepsilon\to 0} \int \|\nabla f\|^2 e^{\varepsilon f}\, r = \int \|\nabla f\|^2\, r$ et $\lim\limits_{\varepsilon\to 0} \int f\, e^{\varepsilon f}\, r = \int f\, r = R(f) = 0$.

5. On doit examiner en détail $\int f\, e^{\varepsilon f}\, r$ et $\ln \int e^{\varepsilon f}\, r$.

6. Dans le premier cas, on écrit par exemple

$$\int f(x)\, e^{\varepsilon f(x)} r(x)\, dx = \varepsilon \int f(x) \left(\frac{e^{\varepsilon f(x)} - 1}{\varepsilon} \right) r(x)\, dx\,,$$

avec $|e^{\varepsilon f} - 1|/\varepsilon \le |f| e^{|f|}$, ce qui permet (par exemple) d'utiliser le TCD.

7. Dans le second cas, on développe de la même manière à l'ordre 3, avec

$$6 \left| e^{\varepsilon f} - 1 - \varepsilon f - \varepsilon^2 f^2/2 \right| / \varepsilon^3 \le |f|^3 e^{|f|}\,.$$

On obtient, puisqu'on a supposé que $R(f) = 0$,

$$\ln \int e^{\varepsilon f(x)} r(x)\, dx = \frac{\varepsilon^2}{2} \int f^2(x)\, r(x)\, dx + o\left(\varepsilon^2\right) \quad \text{quand } \varepsilon \to 0\,.$$

8. On divise les deux membres de cette inégalité par ε^2, puis on fait tendre ε vers 0. Par passage à la limite dans les inégalités, on trouve finalement

$$\int f^2(x)\, r(x)\, dx \le 2 \int \|\nabla f(x)\|^2 r(x)\, dx + \frac{1}{2} \int f^2(x)\, r(x)\, dx\,.$$

Cela conduit ici à $C = 4$.

V

V.1 Cas concave

a) On se ramène au cas $R(f) = 0$ pour simplifier, et on utilise l'indication. Rappelons que $\int f(x) m(x) r(x)\, dx = \int f(x) q(x)\, dx$, puisque $m = q/r$.

C'est un trinôme du second degré en λ.

Variante 1 : Mise sous forme canonique du trinôme : pour tout λ, on a

$$\int \left(\lambda f(x) - \frac{\tau^2 \lambda^2}{2} \right) m(x)\, r(x)\, dx = \frac{\left(\int f(x)\, q(x)\, dx\right)^2}{2\tau^2} - \frac{\tau^2}{2} \left(\lambda - \frac{\int f(x) q(x)\, dx}{\tau^2} \right)^2 .$$

Or, on a choisi $R(f) = 0$, c'est-à-dire $\int f(x) r(x)\, dx = 0$. D'après IV.2.b, on a

$$\int f(x)\, q(x)\, dx = \int f(x)\, q(x)\, dx - \int f(y)\, r(y)\, dy \le \sqrt{2\, \tau^2\, \mathrm{Ent}_r\,(m)}\,.$$

Maintenant, ou bien $\int f(x) q(x)\, dx \le 0$, ou bien $\int f(x) q(x)\, dx > 0$.

— Si $\int f(x) q(x)\, dx > 0$, l'inégalité cherchée se déduit alors directement par élévation au carré des deux membres de l'inégalité ci-dessus.

— Si $\int f(x) q(x)\, dx \le 0$: dans ce cas, la quantité

$$\lambda \int f(x)\, q(x)\, dx - \frac{\tau^2 \lambda^2}{2}$$

ne prend que des valeurs $\leq 0 \leq \mathrm{Ent}_r(m)$ lorsque λ reste ≥ 0.

Variante 2 : Considérons la fonction $\psi(\lambda) = \lambda \int fq - \tau^2\lambda^2/2$ pour $\lambda \geq 0$, et cherchons sa valeur maximale. Ou bien $\int fq \leq 0$, auquel cas $\forall\lambda \geq 0$, $\psi(\lambda) \leq 0 \leq \mathrm{Ent}_r(m)$; ou bien $\int fq > 0$. Dans ce cas, la valeur maximale de $\psi(\lambda)$ pour $\lambda \geq 0$ est $(\int fq)^2/2\tau^2$. Or, d'après IV.2.b on a

$$\int f(x)\,q(x)\;dx \;-\; \int f(y)\,r(y)\;dy \;\leq\; \sqrt{2\,\tau^2\,\mathrm{Ent}_r\left(\frac{q}{r}\right)}\,.$$

Puisqu'on a supposé que $R(f) = \int fr = 0$ et que $\int fq > 0$, on a $0 < \int fq \leq \sqrt{2\tau^2\mathrm{Ent}_r(m)}$, d'où le résultat cherché en élevant les deux membres au carré.

Variante 3 : Une brève étude de la fonction convexe $\lambda > 0 \mapsto b\lambda/2 + c/\lambda$ (avec $b > 0$ et $c > 0$) montre qu'elle admet un seul minimum, qu'on peut déterminer en annulant sa dérivée. Pour $b = \tau^2$ et $c = \mathrm{Ent}_r(m)$,

$$\int f(x)q(x)\,dx \;-\; \int f(y)r(y)\,dy \;\leq\; \sqrt{2bc} \;\leq\; \frac{b\lambda}{2} \;+\; \frac{c}{\lambda}$$

pour tout $\lambda > 0$. (Rappelons que $m = q/r$, $\int m\,r = 1$, $R(f) = \int f\,r = 0$ et $\lambda > 0$.) On multiplie les deux membres de l'inégalité ci-dessus par λ, puis on retranche $b\lambda^2/2$ des deux côtés.

b) Supposons encore que $R(f) = 0$. Notant ℓ la fonction $\lambda f - \lambda^2\tau^2/2$, le résultat de V.1.a s'écrit $\int \ell(x)\,m(x)\,r(x)\,dx \leq \mathrm{Ent}_r(m)$. Choisissons m de la forme exponentielle $m = e^\ell/R(e^\ell)$, ce qui revient à choisir $q = e^\ell r/R(e^\ell)$. L'entropie $\mathrm{Ent}_r(m)$ sécrit alors, comme dans IV.3,

$$\mathrm{Ent}_r(m) = \int \frac{\ell(x)e^{\ell(x)}}{R\left(e^\ell\right)}r(x)\,dx \;-\; \ln\int e^{\ell(x)}r(x)\,dx = \int \ell(x)m(x)r(x)\,dx \;-\; \ln\int e^{\ell(x)}r(x)\,dx\,.$$

Avec ce choix, l'inégalité $\int \ell\,m\,r\,dx \leq \mathrm{Ent}_r(m)$ se réécrit $\ln\int e^\ell r\,dx \leq 0$, et on peut alors conclure, puisque cela s'écrit aussi

$$\int e^{\lambda f(x) \,-\, \lambda^2\tau^2/2}\,r(x)\;dx \;\leq\; 1\,.$$

c) Supposons pour simplifier que $\tau = 1$. Le résultat ci-dessus se traduit par la formule de transfert

$$\mathbb{E}(\phi\,(X_1,\ldots,X_n)) = \int_{[0,\,1]^n} \phi\,(x_1,\ldots,x_n)\;r_1(x_1)\ldots r_n(x_n)\,dx_1\ldots dx_n\,,$$

r_i désignant la densité de la loi de la variable aléatoire X_i $(1 \leq i \leq n)$. On obtient

$$\mathbb{E}\{\exp\left(\lambda\left(f\left(X_1,\ldots,X_n\right) \;-\; \mathbb{E}[f\left(X_1,\ldots,X_n\right)]\right)\right)\} \;\leq\; \exp\left(\frac{\lambda^2}{2}\right)$$

pour tout $\lambda > 0$. On applique l'inégalité de Chernoff à $f(X_1,\ldots,X_n) - \mathbb{E}(f(X_1,\ldots,X_n))$, puis on fait apparaître le trinôme $\lambda^2/2 \,-\, \lambda a \,+\, a^2/2$ dans l'exposant, et on choisit $\lambda = a$.

V.2 Cas convexe

On suppose la fonction $\|\nabla f(x)\|^2$ majorée par une quantité finie ρ^2_{max} sur $]0,\,1[^n$. On obtient l'analogue de l'inégalité de concentration du cas concave, avec ρ^2_{max} à la place de τ^2. Il suffit de reprendre le raisonnement de V.1 point par point, en remplaçant IV.2.b par IV.2.a, et en majorant $I_q(f)$ par ρ^2_{max}.

V.3 Une inégalité dans les espaces de Banach

Question V.3 : Soit E un espace de Banach, de norme $\|\cdot\|_{\mathrm{E}}$. Soit E' son dual topologique (c'est-à-dire l'espace vectoriel des formes linéaires continues sur E), muni de la norme associée $\|\cdot\|_{\mathrm{E}'}$. Soient $b_1, \ldots, b_n$ des éléments de E.

Montrer que si les variables aléatoires indépendantes $X_1, \ldots, X_n$ prennent leurs valeurs dans $[-1,\, 1]$ et si leurs lois respectives admettent chacune une densité strictement positive sur $[-1,\, 1]$, alors pour tout $a \geq 0$,

$$\mathbb{P}\left\{ \left| \left\| \sum_{i=1}^{n} X_i\, b_i \right\|_{\mathrm{E}} - \mathbb{E}\left(\left\| \sum_{i=1}^{n} X_i\, b_i \right\|_{\mathrm{E}} \right) \right| \geq a \right\} \leq 2 \exp\left(- \frac{a^2}{8\sigma_n^2} \right), \tag{1}$$

où

$$\sigma_n^2 = \sup_{\xi \in E' \,:\, \|\xi\|_{\mathrm{E}'} \leq 1} \sum_{i=1}^{n} \left(\xi\,(b_i)\right)^2 , \tag{2}$$

la borne supérieure portant sur l'ensemble des formes linéaires continues ξ sur E de norme associée $\|\xi\|_{\mathrm{E}'} \leq 1$.

Éléments de réponse : Considérons la fonction f définie par $f(x_1, \ldots, x_n) = \| \sum_{i=1}^{n} x_i b_i \|_{\mathrm{E}}$, les x_i appartenant à l'intervalle $[-1,\, 1]$. Utilisons la formule variationnelle

$$\left\| \sum_{i=1}^{n} x_i\, b_i \right\|_{\mathrm{E}} = \sup_{\|\xi\|_{\mathrm{E}'} \leq 1} \xi \left(\sum_{i=1}^{n} x_i b_i \right) = \sup_{\|\xi\|_{\mathrm{E}'} \leq 1} \sum_{i=1}^{n} x_i \xi\,(b_i) \;.$$

(La dernière égalité pourrait permettre de se ramener au cas d'un espace vectoriel de dimension finie.) En vertu, par exemple, du théorème de Hahn-Banach (ou encore de Banach-Alaoglu selon lequel la boule unité de E' est compacte pour la topologie faible), il existe une forme linéaire continue $\xi = \xi^\star$ dépendant de $x_1, \ldots, x_n$ telle que $\|\xi^\star\|_{\mathrm{E}'} \leq 1$ et que $f(x_1, \ldots, x_n) = \xi^\star(\sum_{i=1}^{n} x_i b_i)$. On a alors

$$\left\| \sum_{i=1}^{n} y_i\, b_i \right\|_{\mathrm{E}} \geq \xi^\star \left(\sum_{i=1}^{n} y_i b_i \right)$$

pour tout n-uple $(y_1, \ldots, y_n)$, donc

$$\begin{aligned} f\,(x_1, \ldots, x_n) - f\,(y_1, \ldots, y_n) &\leq \xi^\star \left(\sum_{i=1}^{n} x_i b_i \right) - \xi^\star \left(\sum_{i=1}^{n} y_i b_i \right) \\ &= \sum_{i=1}^{n} (x_i - y_i)\, \xi^\star\,(b_i) \\ &\leq \sum_{i=1}^{n} |x_i - y_i|\, L_i\,(x_1, \ldots, x_n) \;, \end{aligned}$$

où $L_i(x) = |\xi^\star(b_i)|$, de sorte que $L^2(x) \leq L^2_{\max} = \sigma^2$. On conclut via $x_i \in [-1,\, 1]$.

VI Cas $n \geq 2$

n = 2 . — Introduisons la densité de probabilité q_1 définie par $q_1(x_1) = \displaystyle\int_0^1 q(x_1,\, x_2)\, dx_2$ sur $[0,\, 1]$. Il résulte de l'hypothèse $q(x_1,\, x_2) > 0$ pour tous x_1 et x_2 de $[0,\, 1]$ que $q_1(x_1) > 0$ pour tout x_1 de $[0,\, 1]$. Pour chaque x_1 de $[0,\, 1]$, on définit

$$q_2\,(x_2\,|x_1\,) \;=\; \frac{q\,(x_1,\, x_2)}{q_1\,(x_1)}\,. \tag{3}$$

Pour chaque x_1 de $[0,\, 1]$, la fonction $x_2 \mapsto q_2(x_2|\, x_1)$ est une densité de probabilité, que nous noterons $q_2(\bullet|\, x_1)$. C'est la densité conditionnelle de X_2 sachant X_1. Définissons maintenant

$$E_1 \;=\; \mathrm{Ent}_{r_1}\left(\frac{q_1}{r_1}\right) \quad \text{et} \quad E_2 \;=\; \int_0^1 \mathrm{Ent}_{r_2}\left(\frac{q_2\,(\bullet\,|x_1\,)}{r_2}\right)\, q_1\,(x_1)\;\, dx_1\,. \tag{4}$$

On montre que (tensorisation de l'entropie) :

$$\mathrm{Ent}_r\left(\frac{q}{r}\right) \;=\; E_1 \,+\, E_2\,. \tag{5}$$

Nous allons montrer que pour l'analogue de la forme Λ définie en II (voir ci-dessous), on a

$$\sup_\alpha \;\; \Lambda\,(h) \;\leq\; E_1 \,+\, E_2 \tag{6}$$

pour $h(x,\, y) \;=\; \displaystyle\sum_{i=1}^{2} \alpha_i(y)\, \mathbf{1}_{x_i \neq y_i}$, la borne supérieure étant prise sur l'ensemble des fonctions $\alpha =$ $(\alpha_1,\, \alpha_2)$ telles que $\displaystyle\int \sum_{i=1}^{2} \alpha_i^2(y)\; r(y)\;\; dy \leq 1$. On procéderait de manière similaire pour $d_2(r|\, q)$, en intervertissant q et r. Soient

$$h_1\,(x_1,\, x_2,\, y_1,\, y_2) \;=\; \alpha_1\,(y_1,\, y_2)\; \mathbf{1}_{x_1 \neq y_1} \quad \text{et} \quad h_2\,(x_1,\, x_2,\, y_1,\, y_2) \;=\; \alpha_2\,(y_1,\, y_2)\; \mathbf{1}_{x_2 \neq y_2}\,.$$

Nous allons montrer que si $\displaystyle\int \sum_{i=1}^{2} \alpha_i^2(y)\; r(y)\;\; dy \leq 1$, alors $\Lambda(h_1) \leq \sqrt{2\,E_1}$ et $\Lambda(h_2) \leq \sqrt{2\,E_2}$. En utilisant l'inégalité de Cauchy-Schwarz $|a_1 b_1 \,+\, a_2 b_2| \leq \sqrt{a_1^2 + a_2^2}\; \sqrt{b_1^2 + b_2^2}$, on en déduira (**??**), d'où la première moitié de l'inégalité cherchée via la définition (21).

Plutôt que de détailler les calculs, mieux vaut d'abord indiquer comment est construite la forme linéaire positive (i.e., la mesure de probabilité) Λ. Le but de cette construction est d'obtenir un couplage optimal entre les densités de probabilité q et r en enchaînant le couplage optimal $\Lambda_1 =$ $\Lambda_1(dx_1,\, dy_1)$ entre q_1 et r_1 (construit comme en II) et, pour chaque couple de valeurs possibles x_1 et y_1, le couplage conditionnel optimal $\Lambda_2(\bullet|x_1,\, y_1) = \Lambda_2(dx_2,\, dy_2|x_1,\, y_1)$ entre les entre les densités conditionnelles $q_2(\bullet|x_1)$ et $r_2(\bullet|y_1)$. Comme r est ici une densité produit, $r_2(\bullet|y_1)$ est indépendante de y_1.

Soient (par analogie avec II)

$$\phi_1(x_1) \;=\; \min\left(q_1(x_1),\, r_1(x_1)\right), \qquad \phi_2\,(x_2\,|x_1\,) \;=\; \min\left(q_2\,(x_2\,|x_1\,)\,,\, r_2(x_2)\right) \tag{7}$$

et

$$\begin{aligned} \psi_1(x_1,\, y_1) \;\;&=\;\; \frac{[q_1(x_1) \,-\, r_1(x_1)]_+\, [r_1(y_1) \,-\, q_1(y_1)]_+}{d_{\mathrm{VT}}(q_1,\, r_1)}\,, \\ \psi_2\,(x_2,\, y_2\,|x_1\,) \;\;&=\;\; \frac{[q_2\,(x_2\,|x_1\,) \,-\, r_2(x_2)]_+\, [r_2(y_2) \,-\, q_2\,(y_2\,|x_1\,)]_+}{d_{VT}\,(q_2\,(\cdot\,|x_1\,)\,,\, r_2)}\,. \end{aligned} \tag{8}$$

Avec les notations introduites, la mesure de probabilité $\Lambda_1(dx_1,\, dy_1)$ est définie, avec un abus de notation explicité ci-dessous, par

$$\Lambda_1\,(dx_1,\, dy_1) \;=\; \phi_1(x_1)\,\mathbf{1}_{x_1=y_1}(dx_1) \;+\; \psi_1(x_1,\, y_1)\,dx_1dy_1\,,$$

où $\mathbf{1}_{x_1=y_1}(dx_1)$ désigne la mesure de probabilité concentrée sur la diagonale $x_1 = y_1$ et uniforme sur celle-ci. De même,

$$\Lambda_2\,(dx_2,\, dy_2\,|x_1,\, y_1\,) \;=\; \phi_2\,(x_2\,|x_1\,)\;\mathbf{1}_{x_2=y_2}(dx_2) \;+\; \psi_2\,(x_2,\, y_2\,|x_1\,)\;dx_2dy_2\,.$$

Pour enchaîner ces couplages, on forme le produit

$$\Lambda\,(dx_1,\, dy_1,\, dx_2,\, dy_2) \;=\; \Lambda_1\,(dx_1,\, dy_1) \otimes \Lambda_2\,(dx_2,\, dy_2\,|x_1,\, y_1\,)\;,$$

défini sur les pavés par

$$\Lambda\,(A_1 \times B_1 \times A_2 \times B_2) \;=\; \int_{A_1\times B_1} \Lambda_1\,(dx_1,\, dy_1)\,\Lambda_2\,(A_2 \times B_2\,|x_1,\, y_1\,)\;.$$

Notons E l'espace entier. La première marginale est donc définie par

$$\Lambda\,(A_1 \times E \times A_2 \times E) \;=\; \int_{A_1\times E} \Lambda_1\,(dx_1,\, E)\,\Lambda_2\,(A_2 \times E\,|x_1,\, y_1\,)\;,$$

donc par

$$\int_{A_1\times A_2} q_1\,(x_1)\,q_2\,(x_2\,|x_1\,)\;dx_1dx_2\,,$$

c'est donc bien la loi de densité $q(x_1,\, x_2)$, et on a le résultat analogue pour la seconde marginale.

Traduisons cela dans le formalisme de II. Pour toute fonction $h : [0,\, 1]^4 \to \mathbb{R}$ bornée mesurable, posons

$$\begin{aligned}
\Lambda(h) \;\;=\;\; & \int \phi_1\,(x_1)\,\phi_2\,(x_2\,|x_1\,)\;h\,(x_1,\, x_2,\, x_1,\, x_2)\;\;dx_1\,dx_2 \\
& + \int \phi_1\,(x_1)\,\psi_2\,(x_2,\, y_2\,|x_1\,)\;h\,(x_1,\, x_2,\, x_1,\, y_2)\;\;dx_1\,dx_2\,dy_2 \\
& + \int \phi_2\,(x_2\,|x_1\,)\,\psi_1\,(x_1,\, y_1)\;h\,(x_1,\, x_2,\, y_1,\, x_2)\;\;dx_1\,dx_2\,dy_1 \\
& + \int \psi_1\,(x_1,\, y_1)\,\psi_2\,(x_2,\, y_2\,|x_1\,)\;h\,(x_1,\, x_2,\, y_1,\, y_2)\;\;dx_1\,dx_2\,dy_1\,dy_2\,.
\end{aligned} \tag{9}$$

On vérifie que Λ est bien une forme linéaire positive telle que $\Lambda(\mathbf{1}) = 1$, et qu'elle admet les densités de probabilité marginales $\ell_1 = q$ et $\ell_2 = r$, c'est-à-dire que $\Lambda \in \mathcal{L}(q,\, r)$.

Formant $\Lambda_1(dx_1,\, dy_1) \otimes \Lambda_2(dx_2,\, dy_2|x_1,\, y_1)$, on obtient

$$\begin{aligned}
&\phi_1(x_1)\phi_2\,(x_2\,|x_1\,)\;\mathbf{1}_{x_1=y_1}(dx_1)\,\mathbf{1}_{x_2=y_2}(dx_2) \;+\; \phi_1(x_1)\psi_2\,(x_2,\, y_2\,|x_1\,)\;\mathbf{1}_{x_1=y_1}(dx_1)\,dx_2dy_2 \\
&+\; \psi_1(x_1,\, y_1)\phi_2\,(x_2\,|x_1\,)\;dx_1dy_1\,\mathbf{1}_{x_2=y_2}(dx_2) \;+\; \psi_1(x_1,\, y_1)\psi_2\,(x_2,\, y_2\,|x_1\,)\;dx_1dy_1dx_2dy_2\,.
\end{aligned}$$

On démontre que l'on peut encore définir $\Lambda(h)$ pour les fonctions

$$h_1\,(x_1,\, x_2,\, y_1,\, y_2) \;=\; \alpha_1\,(y_1,\, y_2)\;\mathbf{1}_{x_1\neq y_1} \quad \text{et} \quad h_2\,(x_1,\, x_2,\, y_1,\, y_2) \;=\; \alpha_2\,(y_1,\, y_2)\;\mathbf{1}_{x_2\neq y_2}\,,$$

ainsi que pour

$$h_3\,(x_1,\, x_2,\, y_1,\, y_2) \;=\; \beta_1\,(x_1,\, x_2)\;\mathbf{1}_{x_1\neq y_1} \quad \text{et} \quad h_4\,(x_1,\, x_2,\, y_1,\, y_2) \;=\; \beta_2\,(x_1,\, x_2)\;\mathbf{1}_{x_2\neq y_2}\,,$$

pourvu que les intégrales

$$\int \alpha_i^2\,(y_1,\, y_2)\; r\,(y_1,\, y_2)\;\; dy_1\, dy_2 \quad \text{et} \quad \int \beta_i^2\,(x_1,\, x_2)\; q\,(x_1,\, x_2)\;\; dx_1\, dx_2$$

soient finies.

Nous allons considérer seulement le cas de h_1 et h_2, c'est analogue pour h_3 et h_4.

— Pour h_1, on peut mener les calculs directement : à cause du facteur $\mathbf{1}_{x_1 \neq y_1}$, il ne reste que (notations abrégées, avec $x_2 = y_2$ dans la première intégrale) :

$$\Lambda(h_1) \;=\; \int \phi_2\, \psi_1\, \alpha_1\,(y_1,\, y_2)\;\, dx_1 dy_1 dy_2 \;+\; \int \psi_1\, \psi_2\, \alpha_1\,(y_1,\, y_2)\;\, dx_1 dx_2 dy_1 dy_2\,.$$

Or, $\phi_2 + \displaystyle\int \psi_2\, dx_2 = r_2$. Comme de plus $\displaystyle\int \psi_1\, dx_1 = [1 - q_1/r_1]_+ r_1$, on obtient le rśultat en utilisant l'inégalité de Cauchy-Schwarz et I.

— Pour h_2, il semble impossible de ne pas utiliser la mesure $\Lambda_1(dx_1,\, dy_1)$. En intégrant par rapport à x_2 dans l'expression de $\Lambda(h_2)$, on obtient

$$\Lambda(h_2) \;=\; \int \alpha_2\,(y_1,\, y_2)\; \left[1 \;-\; \frac{q_2}{r_2}\right]_+ \,(y_2|\, x_1)\;\, r_2(y_2)\, dy_2\, \Lambda_1\,(dx_1,\, dy_1)\;\,.$$

On applique l'inégalité de Cauchy-Schwarz aux intégrales par rapport à $r_2(y_2)\, dy_2\;\, \Lambda_1(dx_1,\, dy_1)$. Dans un des facteurs, on peut intégrer $\Lambda_1(dx_1,\, dy_1)$ en x_1, ce qui donne la marginale $r_1(y_1)\, dy_1$ et conduit à $\sqrt{\displaystyle\int \alpha_i^2 r\, dy_1 dy_2}$. Dans l'autre, on intègre $\Lambda_1(dx_1,\, dy_1)$ en y_1, ce qui donne la marginale $q_1(x_1)\, dx_1$ et, par I, conduit à $\sqrt{2\, E_2}$.

— C'est ce schéma qui se généralise au cas $n \geq 2$, voir ci-dessous.

— Pour conclure, on utilise l'inégalité de Cauchy-Schwarz $|a_1 b_1\, +\, a_2 b_2| \,\leq\, \sqrt{a_1^2 + a_2^2}\; \sqrt{b_1^2 + b_2^2}$, où on prend $a_i \,=\, \sqrt{2\, E_i}$ et $b_i \,=\, \sqrt{\displaystyle\int \alpha_i^2 r\, dy_1 dy_2}$, avec $\displaystyle\int \sum_{i=1}^{2} \alpha_i^2 r\, dy_1 dy_2 \leq 1$.

Cas général . — On définit

$$x^{(j)} = (x_1, \ldots, x_j) \quad \text{et} \quad y^{(j)} = (y_1, \ldots, y_j)\;\,,$$

les densités marginales partielles $q^{(j)}(x^{(j)})$ et $r^{(j)}(y^{(j)})$ (qui est le produit $r_1(y_1) \ldots r_j(y_j)$), obtenues par intégrations partielles de $q(x)$ et de $r(y)$, les densités de probabilité conditionnelles $q_j(x_j|\, x^{(j-1)})$ $(j \geq 2)$ et $r_j(y_j|\, y^{(j-1)})$ (qui se réduit à $r_j(y_j)$), les entropies partielles

$$E_j \;=\; \int \mathrm{Ent}_{r_j}\left(\frac{q_j\left(\bullet\,|x^{(j-1)}\right)}{r_j}\right)\, q^{(j-1)}\left(x^{(j-1)}\right)\;\, dx^{(j-1)}\,,$$

les couplages conditionnels optimaux $\Lambda_j(dx_j,\, dy_j|x^{(j-1)},\, y^{(j-1)})$ $(j \geq 2)$, et les couplages enchaînés partiels optimaux

$$\Lambda^{(j)}\left(dx^{(j)},\, dy^{(j)}\right) \;=\; \Lambda_1\,(dx_1,\, dy_1) \otimes \Lambda_2\,(dx_2,\, dy_2\,|x_1,\, y_1\,) \otimes \ldots \otimes \Lambda_j\left(dx_j,\, dy_j\,\middle|\,x^{(j-1)},\, y^{(j-1)}\right)\,.$$

Les marginales de $\Lambda_j(dx_j,\, dy_j | x^{(j-1)},\, y^{(j-1)})$ sont $q_j(\bullet | \, x^{(j-1)})$ et r_j :

$$\int_{y_j} \Lambda_j \left(dx_j,\, dy_j \left| x^{(j-1)},\, y^{(j-1)} \right.\right) \;=\; q_j \left(x_j \left| x^{(j-1)} \right.\right) \, dx_j$$

et

$$\int_{x_j} \Lambda_j \left(dx_j,\, dy_j \left| x^{(j-1)},\, y^{(j-1)} \right.\right) \;=\; r_j \left(y_j \right) \, dy_j \, .$$

D'autre part,

$$\int_{y_j} \mathbf{1}_{x_j \neq y_j} \, \Lambda_j \left(dx_j,\, dy_j \left| x^{(j-1)},\, y^{(j-1)} \right.\right) \;=\; \left[q_j \left(x_j \left| x^{(j-1)} \right.\right) \;-\; r_j \left(x_j \right) \right]_+ \, dx_j$$

et

$$\int_{x_j} \mathbf{1}_{x_j \neq y_j} \, \Lambda_j \left(dx_j,\, dy_j \left| x^{(j-1)},\, y^{(j-1)} \right.\right) \;=\; \left[r_j \left(y_j \right) \;-\; q_j \left(y_j \left| x^{(j-1)} \right.\right) \right]_+ \, dy_j \, .$$

Il faut établir des analogues pour des fonctions $\alpha_j(y)\ \mathbf{1}_{x_j \neq y_j}$ ou $\beta_j(x)\ \mathbf{1}_{x_j \neq y_j}$. Dans le cas des fonctions $\alpha_j(y)\ \mathbf{1}_{x_j \neq y_j}$, par exemple, on procède de manière similaire au cas h_2 ci-dessus : on intègre par rapport à $dx_{j+1}\ \ldots\ dx_n$ dans l'expression de

$$\Lambda^{(n)} \left(\alpha_j(y) \mathbf{1}_{x_j \neq y_j} \right) \, ,$$

puis on intègre $\mathbf{1}_{x_j \neq y_j}\ \ \Lambda_j(dx_j,\, dy_j | x^{(j-1)},\, y^{(j-1)})$ par rapport à dx_j, ce qui donne $[r_j(y_j) - q_j(y_j | x^{(j-1)})]_+ \ dy_j$. On applique l'inégalité de Cauchy-Schwarz aux intégrales par rapport à la mesure

$$\prod_{k=j}^{n} r_k(y_k) \, dy_k \, \Lambda^{(j-1)} \left(dx^{(j-1)},\, dy^{(j-1)} \right) \, .$$

Dans un des facteurs, on intègre $\Lambda^{(j-1)}$ en $x^{(j-1)}$, ce qui conduit à $\sqrt{\int \alpha_j^2 r \, dy}$. Dans l'autre, on intègre $\Lambda^{(j-1)}$ en $y^{(j-1)}$, ce qui conduit à $\sqrt{2\, E_j}$. On conclut alors comme ci-dessus.

Conclusion . — L'intérêt de cette méthode par enchaînement de couplages optimaux est de permettre des extensions à des situations où les variables aléatoires X_i ne sont plus indépendantes, mais seulement asymptotiquement indépendantes en un certain sens (par exemple, une classe de chaînes de Markov), ou encore dépendent les unes des autres par des mécanismes d'interaction locale (systèmes de particules).