ANNALES DE L'I. H. P., SECTION B

LAURENT MICLO

Une majoration sous-exponentielle pour la convergence de l'entropie des chaînes de Markov à trou spectral

Annales de l'I. H. P., section B, tome 35, n° 3 (1999), p. 261-311
<http://www.numdam.org/item?id=AIHPB_1999__35_3_261_0>

Ann. Inst. Henri Poincaré,

Vol. 35, n° 3, 1999, p. 261-311 *Probabilités et Statistiques*

Une majoration sous-exponentielle pour la convergence de l'entropie des chaînes de Markov à trou spectral

par

Laurent MICLO

Laboratoire de Statistique et Probabilités, Université Paul Sabatier,
118, route de Narbonne, 31062 Toulouse cedex, France
E-mail: miclo@cict.fr

ABSTRACT. – Let P be a Markovian kernel with an invariant probability μ. Let P^* be the adjoint operator of P in $\mathbb{L}^2(\mu)$. We will see how a spectral gap for $I - PP^*$ (where I is the identity) allows one to obtain an upper bound on the speed of convergence toward equilibrium in the entropy sense, under various hypotheses on the degree of integrability of the density of the initial law with respect to μ. Then we will show how to adapt this approach to get strong ergodicity results for certain inhomogeneous Markov chains. © Elsevier, Paris

Key words and phrases: General Markov chains with a spectral gap, restricted entropy-energy inequalities, upper bounds on the entropy evolution, strong ergodicity in law for inhomogeneous Markov chains.

RÉSUMÉ. – Soit P un noyau markovien admettant une probabilité invariante μ. Notons P^* l'opérateur adjoint de P dans $\mathbb{L}^2(\mu)$. Nous allons voir comment l'existence d'un trou spectral pour $I - PP^*$ (I étant l'identité) permet d'obtenir des majorations de la vitesse de convergence vers l'équilibre au sens de l'entropie, sous diverses hypothèses d'intégrabilité de la densité de la loi initiale par rapport à μ, puis on en déduira des critères d'ergodicité forte en loi pour des chaînes de Markov inhomogènes. © Elsevier, Paris

American Mathematical Society 1991 *Subject Classifications:* Primary 60 J 27 and 60 J 10; secondary 47 A 35, 47 A 50 and 47 A 75.

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques - 0246-0203
Vol. 35/99/03/© Elsevier, Paris

1. INTRODUCTION

Souvent pour étudier quantitativement l'évolution de l'entropie des processus de Markov à temps continu, on se ramène à considérer certaines inégalités de Sobolev-logarithmiques, qui permettent classiquement d'obtenir une convergence exponentiellement rapide. Si le temps est discret, on a vu dans [11] comment il était possible d'adapter ces preuves, cependant les inégalités de Sobolev-logarithmiques qui y apparaissent naturellement ne sont satisfaites que si l'espace des états est fini (et si la chaîne est irréductible apériodique). Le but de cet article est d'étendre quand même ce type de résultat à des chaînes de Markov plus générales, en obtenant une estimée de la décroissance vers 0 dans le temps de l'entropie par rapport à une probabilité invariante μ, sous l'hypothèse d'existence d'un trou spectral dans $\mathbb{L}^2(\mu)$ pour le symétrisé multiplicatif du noyau markovien, et si la mesure initiale (ou l'une des lois de la chaîne à un instant donné) admet une densité par rapport à μ qui est un peu plus intégrable que simplement dans $(\mathbb{L}\ln(\mathbb{L}))(\mu)$. La majoration de la vitesse de convergence qui en résulte n'est plus alors nécessairement exponentielle en le temps et dépend notamment du degré d'intégrabilité de la densité initiale (de plus la borne que l'on obtient pour l'entropie à un instant $n \in \mathbb{N}^*$ ne sera pas linéaire en l'entropie initiale, ce qui ne permet pas de se ramener à une convergence exponentielle).

Précisons maintenant notre cadre : soit $(S, \mathcal{S})$ un espace mesurable muni d'un noyau strictement markovien $P : S \times \mathcal{S} \to [0,1]$ satisfaisant les conditions de mesurabilité usuelles. Comme d'habitude également, on fait agir P sur les fonctions mesurables bornées f définies sur S (dont l'ensemble sera noté $\mathcal{B}(S)$) par

$$\forall\, x \in S, \qquad Pf(x) = \int f(y)\, P(x, dy)$$

et sur les probabilités m définies sur S (dont l'ensemble sera noté $\mathcal{P}(S)$) par

$$\forall\, A \in \mathcal{S}, \qquad mP(A) = \int P(y, A)\, m(dy)$$

Notre première hypothèse est que P admet au moins une probabilité μ invariante, i.e. qui vérifie $\mu P = \mu$. On n'a pas obligatoirement unicité d'une telle mesure μ, mais on suppose que l'on en a choisit une, fixée pour toute la suite.

Le fait que μ est invariante implique que si $A \in \mathcal{S}$ est négligeable pour μ, alors il en est de même pour l'application $P\mathbf{I}_A$, ce qui permet de prolonger de manière unique P comme un opérateur sur $\mathbb{L}^2(\mu)$ (et plus généralement sur tous les $\mathbb{L}^p(\mu)$, pour $1 \leq p \leq \infty$). Celui-ci est alors positif ($f \in \mathbb{L}^2(\mu)$ et $f \geq 0$ implique que $Pf \geq 0$) et par une application de l'inégalité de Schwarz il apparaît qu'il est de norme égale à 1 ($\sup_{f \in \mathbb{L}^2(\mu),\, \|f\|_{\mathbb{L}^2(\mu)}=1} \|Pf\|_{\mathbb{L}^2(\mu)} = 1$).

On notera P^* l'adjoint de P dans $\mathbb{L}^2(\mu)$ (plus généralement, par dualité on peut définir P^* sur $\mathbb{L}^p(\mu)$, pour $1 < p \leq \infty$, on rappelera dans la section suivante comment construire directement P^* sur $\mathbb{L}^1(\mu)$), et remarquons que P^* est aussi positif et contractant (ce qui permet de le considérer comme un opérateur de Markov généralisé), qui de plus admet également μ comme probabilité invariante, au sens où pour tout $f \in \mathbb{L}^2(\mu)$, $\mu(P^*f) = \mu(fP1) = \mu(f)$.

Soit I l'opérateur identité, notre seconde et principale hypothèse est que $I - PP^*$ admet un trou spectral :

$$\lambda(I - PP^*) \stackrel{\text{déf.}}{=} \inf_{f \in \mathbb{L}^2(\mu) \setminus \mathrm{Vect}(\mathbf{1})} \frac{\mathcal{E}_{I-PP^*}(f,f)}{\mathcal{E}_{I-E_\mu}(f,f)} > 0 \tag{1}$$

où pour tout opérateur continu $Q : \mathbb{L}^2(\mu) \to \mathbb{L}^2(\mu)$, on a posé

$$\forall\, f \in \mathbb{L}^2(\mu), \qquad \mathcal{E}_Q(f,f) = \int fQf\, d\mu$$

où E_μ est l'opérateur correspondant à l'espérance par rapport à μ,

$$\forall\, f \in \mathbb{L}^2(\mu), \qquad E_\mu f = \left(\int f\, d\mu\right)\mathbf{1}$$

et où $\mathbf{1}$ est la fonction prenant toujours la valeur 1 (ainsi $\mathcal{E}_{I-E_\mu}(f,f) = \mu(f^2) - (\mu(f))^2$ n'est autre que la variance de f par rapport à μ).

Notons plus précisément que ci-dessus, $\mathcal{E}_{I-PP^*}$ et $\mathcal{E}_{I-E_\mu}$ sont des formes de Dirichlet de domaine $\mathbb{L}^2(\mu)$ (et de générateurs respectifs $I - PP^*$ et $I - E_\mu$).

Remarquons que l'on a toujours $\lambda(I - PP^*) = \lambda(I - P^*P)$, bien que contrairement à la situation où S est fini, les spectres de PP^* et de P^*P puissent être différents (considérer l'exemple classique où $S = \{0,1\}^{\mathbb{N}}$ muni de sa tribu produit habituelle, où pour tout $x = (x_i)_{i \in \mathbb{N}} \in S$, $P(x,\cdot) = \delta_{(x_{i+1})_{i\in\mathbb{N}}}(\cdot)$ et où $\mu = ((\delta_0 + \delta_1)/2)^{\otimes \mathbb{N}}$, auquel cas P^* admet une version régulière donnée par $P^*(x,\cdot) = (\delta_{(0,x_0,x_1,\cdots)}(\cdot) + \delta_{(1,x_0,x_1,\cdots)}(\cdot))/2$ pour tout $x = (x_i)_{i\in\mathbb{N}} \in S$, et il en resort que $P^*P = I$ alors que PP^* est la projection orthogonale sur $\mathbb{L}^2(S, \widetilde{\mathcal{S}}, \mu)$, où $\widetilde{\mathcal{S}}$ est tribu engendrée par les coordonnées sur S indexées par $\mathbb{N}^*$). En effet, soit l'espace de Hilbert $V = \{f \in \mathbb{L}^2(\mu) \,/\, \mu(f) = 0\}$ muni du produit scalaire hérité de $\mathbb{L}^2(\mu)$. Alors V est stable par P et P^*. Notons Q la restriction de P à V, il est clair que l'adjoint Q^* de Q est la restriction de P^* à V. Or $1 - \lambda(I - PP^*)$ n'est autre que la norme d'opérateur de Q, qui est aussi la norme de Q^*, laquelle vaut $1 - \lambda(I - P^*P)$.

Plus généralement, on dit qu'une chaîne de Markov admet un trou spectral, si ses probabilités de transitions sont données par un noyau P qui satisfait pour un certain $n \in \mathbb{N}^*$,

$$\inf_{f \in \mathbb{L}^2(\mu) \setminus \mathrm{Vect}(\mathbf{1})} \frac{\mathcal{E}_{I - P^n P^{n*}}(f, f)}{\mathcal{E}_{I - E_\mu}(f, f)} > 0$$

où μ est toujours une probabilité invariante pour P, et si elle admet une loi initiale absolument continue par rapport à μ. Comme d'habitude, on se ramène alors au cas précédent en remplaçant P par P^n.

Notre dernière hypothèse va porter sur la loi initiale de la chaîne de Markov que l'on considère et pour la décrire, introduisons la notion suivante : on dit qu'une fonction convexe $\varphi : \mathbb{R}_+ \to \mathbb{R}$ est un « bon majorant » (sous-entendu de l'application $\mathbb{R}_+ \ni x \mapsto x\ln(x)$), s'il existe un $x_0 > 1$ tel que φ soit de classe C^4 et strictement positive sur $]x_0, +\infty[$, et vérifie pour tout $x > x_0$,

$$\frac{d}{dx}\frac{x\ln(x)}{\varphi(x)} < 0 \tag{2}$$

$$\frac{d}{dx}\left(x\frac{d^3}{dx^3}(\varphi(x^2))\right) \geq 0 \tag{3}$$

$$\lim_{x \to +\infty} \frac{x\ln(x)}{\varphi(x)} = 0 \tag{4}$$

Les conditions (2) et (4) assurent que φ croît de manière régulière plus vite en $+\infty$ que l'application $\mathbb{R}_+ \ni x \mapsto x\ln(x)$. L'hypothèse (3) peut paraître étrange, elle nous servira à imposer que pour tous K_1, K_2, K_3, $K_4 \in \mathbb{R}$, l'équation

$$K_1 + K_2 x + K_3 x\ln(x) + K_4\varphi'(x^2)x = 0$$

admet au plus 5 zéros, ce qui permettra de ramener certains problèmes variationnels sur des espaces finis à des considérations sur un ensemble ne contenant que 5 points. Les estimées que l'on obtient ne dépendent plus alors du cardinal de l'ensemble initialement considéré, ce qui en passant à la limite, fournit des évaluations sur des espaces mesurables quelconques.

Cependant la condition (3) n'est certainement pas optimale pour ceci, on l'étendra d'ailleurs un peu dans la section 3, mais dans la pratique les hypothèses ci-dessus sont en général suffisantes, car elles permettent de voir par exemple que les fonctions

$$x \mapsto x^{1+\eta}, \quad x \mapsto x|\ln(x)|^{1+\eta}, \quad x \mapsto x\ln(x)\ln(|\ln(x)|)$$

sont des bons majorants, pour tout $\eta > 0$.

Rappelons par ailleurs que l'entropie d'une probabilité $m \in \mathcal{P}(S)$ par rapport à μ est définie par

$$\mathrm{Ent}(m|\mu) = \begin{cases} \int \ln(dm/d\mu)\, dm & , \text{si } m \ll \mu \\ +\infty & , \text{sinon} \end{cases}$$

Soit $m \in \mathcal{P}(S)$, qui sera la loi initiale de la chaîne, une probabilité absolument continue par rapport à μ pour laquelle il existe un bon majorant φ tel que

$$K \stackrel{\text{déf.}}{=} \mu(\varphi(dm/d\mu)) < +\infty$$

ce qui revient à demander un peu plus que la simple intégrabilité sous μ de $\ln(dm/d\mu)dm/d\mu$ nécessaire pour que $\mathrm{Ent}(m|\mu)$ soit finie.

Notre but est de construire de manière relativement explicite une fonction $v : \mathbb{N} \to \mathbb{R}_+$, qui ne dépende que de $\mathrm{Ent}(m|\mu)$, $\lambda(I - PP^*)$, φ et K, telle que

$$\lim_{n\to\infty} v(n) = 0$$

et surtout

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(mP^n|\mu) \leq v(n)$$

ce qui permet d'estimer en un certain sens la vitesse de convergence vers l'équilibre d'une chaîne de Markov de loi initiale m et dont les probabilités de transitions sont données par P.

Ainsi par exemple si l'entropie initiale $\mathrm{Ent}(m|\mu)$ vaut 1 et si les bons majorants sont ceux présentés ci-dessus, on obtient respectivement les fonctions v définies pour tous $n \in \mathbb{N}$ par

$$\begin{aligned} v(n) &= \exp(1 - \sqrt{1 + A\lambda(I - PP^*)n}) \\ v(n) &= (1 + A\lambda(I - PP^*)n)^{-\eta} \\ v(n) &= 8K/\ln[\exp(8K) + \ln(1 + A\lambda(I - PP^*)n)] \end{aligned}$$

où $A > 0$ est une constante « universelle », i.e. qui ne dépend du problème considéré que par l'intermédiaire de φ et de K, mais on verra aussi ultérieurement comment préciser la dépendance de A en K.

Le plan de l'article est le suivant : dans la section suivante, qui est indépendante du reste, on verra que si l'on ne s'intéresse pas à la rapidité de convergence, pour obtenir seulement la convergence de l'entropie vers 0, il suffit de supposer que l'entropie initiale est finie et que la tribu de queue est triviale pour la loi de la chaîne de Markov stationnaire dont la loi initiale est μ et dont les probabilités de transitions sont données par P, ce qui affaiblit beaucoup les hypothèses précédentes. Dans la section 3, on s'intéressera à des inégalités énergie-entropie restreintes à certains sous-ensembles particuliers de $\mathbb{L}^2(\mu)$, dans le cas où S est fini et où $P = E_\mu$. Cependant, comme on l'a mentionné précédemment, cette étude n'est pas effectuée en vue d'applications pour des ensembles finis, mais pour obtenir des estimées indépendantes du cardinal de l'ensemble, de manière à pouvoir les étendre à des ensembles mesurables quelconques. On en déduira ensuite à l'aide du trou spectral des inégalités énergie-entropie restreintes plus générales dans la section 4, sans limitations sur la forme de P, que l'on « intégrera » dans la cinquième section pour obtenir les estimées escomptées (pour essayer de justifier l'aspect technique des sections 3 et 4, on pourra commencer par lire cette partie, en admettant les résultats précédents). Enfin dans une dernière section, on verra comment ceci peut s'appliquer à l'étude de l'ergodicité forte en loi de certaines chaînes de Markov inhomogènes.

Cependant nous pensons que les majorations présentées des taux de décroissance ne sont pas asymptotiquement du même ordre que les véritables vitesses de convergence. Ainsi par exemple si l'on sait que la densité initiale $dm/d\mu$ appartient à $\mathbb{L}^2(\mu)$, alors en utilisant d'une part l'estimée

$$\forall\, n \in \mathbb{N}, \qquad \left\| \frac{dmP^n}{d\mu} - 1 \right\|^2_{\mathbb{L}^2(\mu)} \leq (1 - \lambda(I - PP^*))^n \left\| \frac{dm}{d\mu} - 1 \right\|^2_{\mathbb{L}^2(\mu)}$$

qui découle immédiatement de la définition du trou spectral et de l'identité $dmP/d\mu = P^*(dm/d\mu)$, et d'autre part la majoration

$$\forall\, m \in \mathcal{P}(S),\, m \ll \mu, \qquad \mathrm{Ent}(m|\mu) \leq \ln\left(1 + \left\| \frac{dm}{d\mu} - 1 \right\|^2_{\mathbb{L}^2(\mu)}\right)$$

qui provient de la concavité du logarithme, comme l'a remarqué Francis Su dans sa thèse de Ph.D. à l'université d'Harvard, on fait apparaître que l'entropie finit par décroître exponentiellement vite en le temps.

Mais l'hypothèse $dm/d\mu \in \mathbb{L}^2(\mu)$ est typiquement de celles que l'on veut éviter, et même si elle est satisfaite, nos estimées peuvent être meilleures sur un certain intervalle de temps que celles déduites de ce qui précède. En effet, il se peut que $\ln(1 + \left\|\frac{dm}{d\mu} - 1\right\|^2_{\mathbb{L}^2(\mu)})$ soit très grand par rapport à $\operatorname{Ent}(m|\mu)$, et en étudiant directement l'évolution de l'entropie, on évite ce saut initial dans l'estimation. Il semblerait ainsi, du moins si la densité initiale appartient à un $\mathbb{L}^p(\mu)$, avec $p \geq 2$ (les cas où $1 < p < 2$ restent par contre problématiques, si l'on n'a pas recours à l'entropie), que les calculs qui suivent seront surtout intéressants en temps moyen, pour estimer des temps de relaxation à l'équilibre au sens de l'entropie, en fonction de l'intégrabilité de la densité initiale, et non pas pour évaluer des comportements asymptotiques en temps grand. Mais la motivation principale provient des situations inhomogènes, où l'entropie est plus maniable que les normes $\mathbb{L}^2$, dès que l'on s'affranchit de certaines hypothèses de bornitude, comme on le verra dans la section 6.

D'autre part, précisons que l'on pourrait remplacer l'inégalité de trou spectral par d'autres inégalités, par exemple celles introduites récemment par Mathieu [8], qui reviennent à supposer que pour un $0 < p \leq 2$, on a

$$\inf_{f \in \mathbb{L}^2(\mu) \setminus \operatorname{Vect}(\mathbf{1})} \frac{\mathcal{E}_{I-PP^*}(f,f)}{\mu((f-\mu(f))^p)^{2/p}} > 0$$

car de nombreux calculs présentés ci-dessous peuvent s'étendre à cette situation.

Nous avons préféré travailler avec le trou spectral, car il existe de nombreux moyens pour l'estimer (voir la fin de la section 4).

2. CONVERGENCE QUALITATIVE DE L'ENTROPIE VERS 0

Pour pouvoir par la suite nous intéresser uniquement à des résultats quantitatifs, nous allons prouver ici, sous des hypothèses plus restreintes que celles présentées dans l'introduction, que l'on a souvent une convergence de l'entropie vers 0. La démonstration que nous en donnons ne permet pas d'en estimer la vitesse, car elle basée partiellement sur le théorème de convergence des martingales inverses positives. Par contre, elle s'adapte immédiatement pour fournir d'autres résultats d'ergodicité du même type dont nous discuterons.

Soit à nouveau P un noyau de Markov admettant une probabilité invariante μ, on dit que le couple (P, μ) satisfait la loi du 0-1, si en notant $X = (X_n)_{n\in\mathbb{N}}$ la chaîne de Markov canonique (i.e. définie comme le processus des coordonnées sur l'espace $(S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}})$) stationnaire de probabilités de transitions données par P et de loi initiale μ, on a pour tout événement A appartenant à la tribu de queue $\mathcal{T} = \cap_{n\in\mathbb{N}}\sigma(X_p\,;\, p \geq n)$,

$$\mathbb{P}_\mu(A) = 0 \text{ ou } 1$$

Il s'agit de l'une des conditions classiques d'ergodicité en théorie générale des chaînes de Markov (voir par exemple la section 2 du chapitre 6 du livre de Revuz [13]) et sera notre unique hypothèse dans cette section (avec l'existence de la probabilité initiale μ).

Par ailleurs, rappelons comment il est possible de définir P^* directement sur $\mathbb{L}^1(\mu)$ (cf. par exemple le début du chapitre VII de [5] ou la définition 1.2 p. 107 de [13]). Si $f \in \mathbb{L}^1(\mu)$, considérons la mesure signée bornée donnée par $m = f\mu$. Du fait de l'invariance de μ par P, la mesure signée bornée mP est encore absolument continue par rapport à μ, et on désigne alors par $P^*f \in \mathbb{L}^1(\mu)$ sa densité. On vérifie facilement que P^* admet aussi l'interprétation suivante : si $f \in \mathbb{L}^1(\mu)$ (ou si f est $\mathcal{S}$-mesurable et positive), on a pour tout $n \geq 0$, $\mathbb{P}_\mu$-p.s.,

$$\mathbb{E}_\mu[f(X_n)|\mathcal{F}^{n+1}] = P^*f(X_{n+1})$$

où $\mathcal{F}^{n+1} = \sigma(X_p\,;\, p \geq n+1)$.

On en déduit la

Proposition 1. – *Supposons que (P, μ) vérifie la loi du 0-1. Soit m une probabilité initiale sur S, alors soit*

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(mP^n|\mu) = +\infty$$

soit

$$\lim_{n\to\infty} \mathrm{Ent}(mP^n|\mu) = 0$$

Preuve. – Clairement il suffit de montrer que si $m \in \mathcal{P}(S)$ est telle que $\mathrm{Ent}(m|\mu) < +\infty$, alors

$$\lim_{n\to\infty} \mathrm{Ent}(mP^n|\mu) = 0$$

Pour ceci, notons pour $n \in \mathbb{N}$, f_n la densité de mP^n par rapport à μ. D'après ce qui précède, ces fonctions se déduisent les unes des autres par la relation de récurrence

$$f_{n+1} = P^*f_n$$

ainsi en posant $Z_n = f_n(X_n)$, le processus $(Z_n)_{n\in\mathbb{N}}$ est une martingale inverse par rapport à la famille de tribus $(\mathcal{F}^n)_{n\in\mathbb{N}}$. Comme de plus les variables aléatoires Z_n sont positives, il est bien connu (voir par exemple le paragraphe V-3 de [12]) que $\mathbb{P}_\mu$-p.s.,

$$\lim_{n\to\infty} Z_n = \mathbb{E}_\mu[Z_0|\mathcal{T}]$$

Mais vu l'hypothèse faite sur la tribu de queue, le terme de droite est $\mathbb{P}_\mu$-p.s. constant, et son espérance vaut $\mu(f_0) = 1$, on a donc en fait

$$\lim_{n\to\infty} Z_n = 1$$

Cependant, $\mu(f_0 \ln(f_0)) = \mathrm{Ent}(m|\mu) < +\infty$ et le théorème de La Vallée-Poussin (cf. par exemple le théorème 22 p. 38 de [9]) implique qu'il existe une fonction croissante et convexe $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ telle que pour x grand $\psi(x) \gg x$ et telle que

$$\mu(\psi(f_0 \ln(f_0))) < +\infty$$

Posons $G(x) = \psi(x\ln(x))$ pour $x \geq 0$. Du fait que ψ et $\mathbb{R}_+ \ni x \mapsto x\ln(x)$ sont convexes et que ψ est croissante, il découle que G est convexe. En appliquant alors l'inégalité de Jensen à des espérances conditionnelles, on voit que $\mathbb{P}_\mu$-p.s.,

$$\begin{aligned} G(f_{n+1}(X_{n+1})) &= G(\mathbb{E}_\mu[f_n(X_n)|\mathcal{F}^{n+1}]) \leq \mathbb{E}_\mu[G(f_n(X_n))|\mathcal{F}^{n+1}] \\ &= P^*(G(f_n))(X_{n+1}) \end{aligned}$$

c'est-à-dire que μ-p.s., $G(f_{n+1}) \leq P^*(G \circ f_n)$, d'où $\mu(G(f_{n+1})) \leq \mu(P^*(G \circ f_n)) = \mu(G(f_n))$, et en conséquence pour tout $n \in \mathbb{N}$,

$$\mu(G(f_n)) \leq \mu(G(f_0)) < +\infty$$

ce qui s'écrit aussi sous la forme

$$\mathbb{E}_\mu[\psi(Z_n \ln(Z_n))] \leq \mu(G(f_0))$$

Ainsi une application de la réciproque du théorème de La Vallée-Poussin permet d'obtenir que $(Z_n \ln(Z_n))_{n\in\mathbb{N}}$ est uniformément intégrable, puis que

$$\lim_{n\to\infty} \mathbb{E}_\mu[Z_n \ln(Z_n)] = \mathbb{E}_\mu[1\ln(1)] = 0$$

ce qui n'est autre que le résultat annoncé. □

En fait on ne s'est pas vraiment servi ci-dessus de la forme particulière de la fonction $\mathbb{R}_+ \ni x \mapsto x\ln(x)$ (celle-ci permettrait d'avoir des renseignements supplémentaires du type de ceux donnés dans la proposition IV-2-10 de [12], qui s'adapte immédiatement aux martingales inverses), ainsi plus généralement, soit $\varphi : \mathbb{R}_+ \to \mathbb{R}$ une application convexe minorée telle que $\varphi(1) = 0$. Notons pour toute probabilité $m \in \mathcal{P}(S)$,

$$E_\varphi(m) = \begin{cases} \displaystyle\int \varphi\left(\frac{dm}{d\mu}\right) d\mu & \text{, si } m \ll \mu \\ +\infty & \text{, sinon} \end{cases}$$

et remarquons que cette quantité est toujours positive d'après l'inégalité de Jensen.

On montre alors de la même manière que pour toute probabilité initiale m, soit

$$\forall\, n \in \mathbb{N}, \qquad E_\varphi(mP^n|\mu) = +\infty$$

soit

$$\lim_{n\to\infty} E_\varphi(mP^n|\mu) = 0$$

Supposons maintenant de plus que φ admette un développement limité d'ordre 2 au voisinage de 1 de la forme $\varphi(1+x) = \varphi'(1)x + ax^2 + \mathrm{o}(x^2)$ avec $a > 0$ (en fait une minoration suffirait pour ce qui suit, notamment une telle inégalité existe toujours si φ n'est pas dérivable en 1, en remplaçant $\varphi'(1)$ par un nombre strictement compris entre la dérivée à gauche et à droite en 1). En reprenant la preuve de la majoration $\|m-\mu\|_{\mathrm{vt}} \le \sqrt{2\mathrm{Ent}(m|\mu)}$ donnée p. 198 de [16], on montre qu'il existe une constante $K > 0$ telle que

$$\forall\, m \in \mathcal{P}(S), \qquad \|m-\mu\|_{\mathrm{vt}} \le K\sqrt{E_\varphi(m)}$$

En effet, en considérant séparément ce qui se passe au voisinage de 1, de 0 et de $+\infty$, on voit qu'il existe deux constantes $b, c > 0$ telles que

$$\forall\, x \ge 0, \qquad (x-1)^2 \le (b+cx)(\varphi(x) - \varphi'(1)(x-1))$$

Il en découle que si $m \ll \mu$ et si $f = dm/d\mu$,

$$\begin{aligned} \|m-\mu\|_{\mathrm{vt}}^2 = \|f-1\|_{\mathbb{L}^1(\mu)}^2 &\le \left\|(b+cf)^{\frac12}(\varphi(f) - \varphi'(1)(f-1))^{\frac12}\right\|_{\mathbb{L}^2(\mu)}^2 \\ &\le \|(b+cf)\|_{\mathbb{L}^1(\mu)}\|(\varphi(f)-\varphi'(1)(f-1))\|_{\mathbb{L}^1(\mu)} \\ &= (b+c)E_\varphi(m) \end{aligned}$$

car on a $f \geq 0$ et $\varphi(f) \geq \varphi'(1)(f-1)$ par convexité.

A partir de ces deux observations on peut retrouver le résultat suivant, qui découle aussi du corollaire 2.5 de [13] :

COROLLAIRE 2. – *Plaçons nous à nouveau sous l'hypothèse de la proposition précédente et soit m une probabilité initiale quelconque. Alors soit il existe un $n \in \mathbb{N}$ tel que $mP^n \ll \mu$, auquel cas on a*

$$\lim_{n\to\infty} \|mP^n - \mu\|_{\mathrm{vt}} = 0$$

soit pour tout $n \in \mathbb{N}$, mP^n n'est pas absolument continu par rapport à μ.

Preuve. – Comme d'habitude il suffit de considérer le cas où la probabilité initiale m est telle que $m \ll \mu$. Notons f_0 sa densité. Une nouvelle application du théorème de La Vallée-Poussin montre qu'il existe une fonction φ croissante, convexe et satisfaisant $\lim_{x\to+\infty} \varphi(x)/x = +\infty$ telle que $\mu(\varphi(f_0)) < +\infty$. On peut supposer de plus que φ est de classe C^2 (par convolution avec une fonction régularisante de support inclus dans $\mathbb{R}_-$) et, quitte à lui ajouter une constante et à augmenter un peu φ'' sur un compact (ce qui en l'infini n'a pour conséquence que de rajouter à φ une fonction linéaire), que $\varphi(1) = 0$ et $\varphi''(1) > 0$.

Ainsi d'après la discussion qui précède, on a

$$\lim_{n\to\infty} E_\varphi(mP^n) = 0$$

puis en fin de compte

$$\lim_{n\to\infty} \|mP^n - \mu\|_{\mathrm{vt}} = 0 \qquad \square$$

Notons que la preuve du corollaire 2.5 de [13] fait également intervenir le théorème de convergence des martingales inverses positives, mais d'une manière un peu différente, ce qui fait qu'il est plus général. En revanche, l'approche précédente permet d'être un peu plus précis, puisqu'elle montre que la densité de $mP^n - \mu$ par rapport à μ converge en moyenne vers 0 dans l'espace d'Orlicz $\mathbb{L}^\Phi(\mu)$, si $m \ll \mu$ et si Φ est une fonction de Young telle que $\int \Phi(dm/d\mu)\, d\mu < +\infty$ (pour une présentation des espaces d'Orlicz, voir le livre de Krasnosel'skii et Rutickii [7] ou l'appendice de Neveu dans [12], avec plus de prudence toutefois, car son affirmation p. 199 « puisque les v.a. étagées sont denses dans $\mathbb{L}^\Phi$ » est en général fausse, ce qui nous empêche aussi d'appliquer sa proposition IX-3-4).

En effet, ceci signifie simplement, en reprenant les notations de la preuve de la proposition 1, que

$$\lim_{n\to\infty} \int \Phi(|f_n - 1|)\, d\mu = 0$$

ce qui s'obtient en considérant la fonction convexe φ définie sur $\mathbb{R}_+$ par $\forall\ x \geq 0,\ \varphi(x) = \Phi(|x-1|)$, qui vérifie bien $E_\varphi(m) < +\infty$ car $\varphi(\cdot) \leq \Phi(\cdot) \vee \max_{0\leq t\leq 1} \Phi(t)$ sur $\mathbb{R}_+$.

Notamment si Φ satisfait la condition $(\triangle_2)$ de [7] (c'est-à-dire s'il existe $u_0 > 0$ et $k > 1$ tels que pour tout $u \geq u_0$, $\Phi(2u) \leq k\Phi(u)$), alors $dmP^n/d\mu$ converge vers 1 en norme dans $\mathbb{L}^\Phi(\mu)$, ce qui peut s'écrire sous la forme : pour tout $M > 0$ fixé,

$$\lim_{n\to\infty} \sup_{g\in\mathbb{L}^\Psi(\mu)\,;\,\mu(\Psi(|g|))\leq M} \left| \int g\, d(mP^n - \mu) \right| = 0$$

où Ψ est la fonction de Young conjuguée de Φ (cf. le théorème 9.4 p. 76, le théorème 9.3 p. 74 et l'inégalité (9.12) p. 72 de [7], ainsi que la remarque à la fin de la page 77 pour la réciproque, grâce à $(\triangle_2)$). Notons que cette condition $(\triangle_2)$ permet aussi d'utiliser les preuves de l'appendice de Neveu [12] et donc d'obtenir directement la convergence précédente (voir le paragraphe 9 de [7]).

Remarquons par ailleurs que l'on a une équivalence dans le corollaire 2, car Totoki (cf. la preuve de la proposition 3 de [17]) a obtenu très simplement un résultat qui dans notre contexte se réécrit de la manière suivante :

$$(P,\mu) \text{ satisfait la loi du 0-1} \iff \forall\ h \in \mathcal{B}(S),\ \lim_{n\to\infty} \mu(|P^{n*}h - \mu(h)|) = 0$$

On peut clairement remplacer dans cette équivalence $\mathcal{B}(S)$ par

$$\left\{ f \in \mathcal{B}(S)\,/\, f \geq 0 \text{ et } \int f\, d\mu = 1 \right\}$$

ce qui montre que la loi du 0-1 pour (P,μ) est notamment impliquée par l'hypothèse

$$\forall\ m \in \mathcal{P}(S),\ m \ll \mu, \qquad \lim_{n\to\infty} \|mP^n - \mu\|_{\mathrm{vt}} = 0$$

et est donc équivalente à cette propriété.

Il apparaît également que la proposition 1 donne une condition nécessaire et suffisante, c'est-à-dire que l'on a aussi réciproquement ; si pour toute probabilité initiale sur S, soit $\forall\ n \in \mathbb{N}$, $\mathrm{Ent}(mP^n|\mu) = +\infty$, soit $\lim_{n\to\infty} \mathrm{Ent}(mP^n|\mu) = 0$, alors (P,μ) vérifie la loi du 0-1.

De la même manière on constate que cette loi du 0-1 est équivalente à

$$\forall\ f \in \mathbb{L}^2(\mu), \qquad \lim_{n\to\infty} \|P^{n*}f - \mu(f)\|_{\mathbb{L}^2(\mu)} = 0$$

ce qui permet de se rendre compte en quel sens on va renforcer cette hypothèse, car l'existence d'un $n \in \mathbb{N}^*$ tel que $\lambda(I - P^n P^{n*}) > 0$ revient exactement à dire que la convergence précédente est uniforme en les $f \in \mathbb{L}^2(\mu)$ tels que $\|f\|_{\mathbb{L}^2(\mu)} \leq 1$, auquel cas la convergence à lieu exponentiellement rapidement.

3. CAS DES ENSEMBLES FINIS

On va considérer dans cette section, tout d'abord la situation où $S = \{1, \cdots, N\}$, avec $N \in \mathbb{N}$, $N \geq 2$ fixé, puis grâce à une astuce basée sur un argument apparaissant dans la preuve du théorème 5.1 de [3], on ramènera au cas précédent certains problèmes sur $S = \{1, \cdots, n\}$, avec $n \geq N$. Notre unique but est de montrer le corollaire 5 ci-dessous, dont la preuve est technique mais très élémentaire.

Soit $\varphi : \mathbb{R}_+ \to \mathbb{R}$ une fonction continue pour laquelle il existe $x_0 > 1$ tel que l'application $x \mapsto x\ln(x)/\varphi(x)$ soit strictement décroissante sur $[x_0, +\infty[$ et dont la limite en $+\infty$ soit 0. Il en est alors de même pour la fonction $x \mapsto x/\varphi(x)$, et ainsi si $K > 0$ est fixé, il existe une unique bijection ψ entre $]0, 2K(N-1)x_0\ln(x_0)/\varphi(x_0)]$ et $]0, x_0\varphi(x_0)]$ telle que

$$\forall\, x \in [x_0, +\infty[, \qquad \psi(2K(N-1)x\ln(x)/\varphi(x)) = x/\varphi(x)$$

On prolonge ensuite ψ sur $\mathbb{R}_+$ en posant $\psi(0) = 0$ et par exemple $\psi(x) = x_0/\varphi(x_0)$ pour $x \geq 2K(N-1)x_0\ln(x_0)/\varphi(x_0)$, remarquons bien que ψ ne dépend que du triplet $(K(N-1), \varphi, x_0)$.

Notons $\mathcal{P}_N^*$ l'ensemble des probabilités sur $\{1, \cdots, N\}$ qui en chargent tous les points, et $\mathcal{B}_N^+$ l'ensemble des fonctions positives définies sur $\{1, \cdots, N\}$. Pour $\mu \in \mathcal{P}_N^*$ et $f \in \mathcal{B}_N^+$, posons

$$\begin{aligned}
\mathcal{E}_{I-E_\mu}^{(\mu)}(f,f) &= \sum_{1\leq x,y\leq N} (f(y) - f(x))^2 \mu(x)\mu(y) \\
\mathcal{L}^{(\mu)}(f) &= \sum_{1\leq x\leq N} f^2(x)\ln(f^2(x))\,\mu(x)
\end{aligned}$$

On note enfin pour $K > 0$ fixé,

$$\alpha_{\varphi,K}(\mu) = \inf_{f\in D(\mu,\varphi,K)} \frac{\mathcal{E}_{I-E_\mu}^{(\mu)}(f,f)}{\psi(\mathcal{L}^{(\mu)}(f))}$$

où $D(\mu, \varphi, K) = \{f \in \mathcal{B}_N^+ \setminus \{\mathbb{1}\} \,/\, \mu(f^2) = 1 \text{ et } \mu(\varphi(f^2)) \leq K\}$. Alors, PROPOSITION 3. – *On a*

$$\inf_{\mu\in\mathcal{P}_N^*} \alpha_{\varphi,K}(\mu) > 0$$

Par la suite on notera aussi $C(N,K,\varphi) > 0$ *la constante définie par l'infimum ci-dessus.*

Preuve. – La démonstration va s'effectuer par l'absurde : supposons qu'il existe deux suites $(\mu_n)_{n\in\mathbb{N}}$ et $(f_n)_{n\in\mathbb{N}}$ d'éléments respectifs de $\mathcal{P}_N^*$ et de $\mathcal{B}_N^+$ satisfaisant

$$\sum_{1\leq i\leq N} f_n(i)\mu_n(i) = 1 \qquad \text{et} \qquad \sum_{1\leq i\leq N} \varphi(f_n(i))\mu_n(i) \leq K$$

telles qu'en posant

$$\mathcal{E}_n = 1 - \sum_{1\leq i\leq N} \sqrt{f_n(i)}\mu_n(i)$$

$$\text{et} \qquad \mathcal{L}_n = \sum_{1\leq i\leq N} f_n(i)\ln(f_n(i))\mu_n(i) = \mathcal{L}^{(\mu_n)}(\sqrt{f_n})$$

on ait

$$\lim_{n\to\infty} \frac{\mathcal{E}_n}{\psi(\mathcal{L}_n)} = 0$$

(on aura remarqué que $\mathcal{E}_{I-E_{\mu_n}}^{(\mu_n)}(\sqrt{f_n},\sqrt{f_n}) = 1-(1-\mathcal{E}_n)^2 \geq \mathcal{E}_n$).

Par symétrie du problème, on peut supposer que pour tout $n\in\mathbb{N}$,

$$0 < \mu_n(1) \leq \mu_n(2) \leq \cdots \leq \mu_n(N)$$

De plus par compacité, quitte à extraire une sous-suite, on fera également l'hypothèse que les μ_n convergent vers une certaine probabilité μ sur $\{1,\cdots,N\}$:

$$\forall\, 1\leq i\leq N, \qquad \lim_{n\to\infty} \mu_n(i) = \mu(i)$$

Dans un premier temps on va s'intéresser au cas où $\mu\notin\mathcal{P}_N^*$: soit alors $i_0 = \sup\{1\leq i\leq N\,/\,\mu(i)=0\}$, de sorte que $\forall\, i\leq i_0,\ \mu(i)=0$ et $\forall\, i>i_0,\ \mu(i)>0$. On conviendra désormais de noter $\sum_*$ la somme sur les $1\leq i\leq i_0$, $\sum^*$ la somme sur les $i_0<i\leq N$ et $\sum = \sum_* + \sum^*$.

Par ailleurs, les $f_n(i)$ étant positifs et $\sum f_n(i)\mu_n(i) = 1$, on peut aussi supposer que pour $i > i_0$, $f_n(i)$ converge pour n grand, disons vers

$f(i) \geq 0$. Commençons par vérifier que l'on a nécessairement $f(i) = 1$ pour $i > i_0$.

Pour ceci montrons que pour tout $1 \leq i \leq i_0$,

$$\lim_{n \to \infty} f_n(i) \ln(f_n(i)) \mu_n(i) = 0 \tag{5}$$

Puisque $\lim_{n\to\infty} \mu_n(i) = 0$, si ceci n'était pas le cas, il existerait une sous-suite $(n_k)_{k \in \mathbb{N}}$ telle que $\lim_{k\to\infty} f_{n_k}(i) = +\infty$ et $\liminf_{k\to\infty} f_{n_k}(i) \ln(f_{n_k}(i)) \mu_{n_k}(i) > 0$. Cependant pour k grand on a $\varphi(f_{n_k}(i)) \gg f_{n_k}(i) \ln(f_{n_k}(i))$ d'où $\liminf_{k\to\infty} \varphi(f_{n_k}(i)) \mu_{n_k}(i) = +\infty$ ce qui est en contradiction avec $\sum \varphi(f_n(i)) \mu_n(i) \leq K$, car φ est minorée sur $\mathbb{R}_+$.

De la même manière on voit que pour tout $1 \leq i \leq i_0$,

$$\lim_{n \to \infty} f_n(i) \mu_n(i) = 0$$
$$\lim_{n \to \infty} \sqrt{f_n(i)} \mu_n(i) = 0$$

Ainsi on obtient que

$$\lim_{n \to \infty} \mathcal{E}_n = 1 - \sum\nolimits^* \sqrt{f(i)} \mu(i)$$
$$\lim_{n \to \infty} \mathcal{L}_n = \sum\nolimits^* f(i) \ln(f(i)) \mu(i)$$
$$\sum\nolimits^* f(i) \mu(i) = 1$$

Or le fait que $\lim_{n\to\infty} \mathcal{E}_n / \psi(\mathcal{L}_n) = 0$ implique (car ψ est bornée) que $\lim_{n\to\infty} \mathcal{E}_n = 0$ d'où $\sum^* \sqrt{f(i)} \mu(i) = 1 = \sum^* f(i) \mu(i)$, ce qui constitue un cas d'égalité dans l'inégalité de Schwartz et on en déduit immédiatement que pour tout $i > i_0$, $f(i) = 1$.

Posons pour $i > i_0$,

$$\alpha_n(i) = \mu_n(i) - \mu(i) \qquad \text{et} \qquad \beta_n(i) = f_n(i) - 1$$

de sorte que $\lim_{n\to\infty} \alpha_n(i) = \lim_{n\to\infty} \beta_n(i) = 0$ et on a évidemment

$$\sum\nolimits^* \alpha_n(i) = -\sum\nolimits_* \mu_n(i)$$
$$\sum\nolimits^* \beta_n(i) \mu(i) + \alpha_n(i) + \beta_n(i) \alpha_n(i) = -\sum\nolimits_* f_n(i) \mu_n(i)$$

Pour minorer $\mathcal{E}_n$, commençons par trouver $\gamma_1 > 0$ assez petit tel que pour tout $n \in \mathbb{N}$ et tout $i > i_0$,

$$\sqrt{1 + \beta_n(i)} \leq 1 + \frac{\beta_n(i)}{2} - \gamma_1 \beta_n^2(i)$$

ce qui est possible vu le comportement asymptotique de $\beta_n(i)$ et celui de $x \mapsto \sqrt{1+x}$ au voisinage de 0.

On écrit alors que

$$\begin{aligned}
&\sum\nolimits^* \sqrt{f_n(i)}\mu_n(i)\\
&= \sum\nolimits^* \sqrt{1+\beta_n(i)}(\mu(i)+\alpha_n(i))\\
&\leq \sum\nolimits^* (1+\beta_n(i)/2)(\mu(i)+\alpha_n(i)) - \gamma_1 \sum\nolimits^* \beta_n^2(i)\mu_n(i)\\
&= \sum\nolimits^* \mu(i) + \frac{1}{2}\sum\nolimits^* (\beta_n(i)\mu(i)+\alpha_n(i)+\beta_n(i)\alpha_n(i))\\
&\quad + \frac{1}{2}\sum\nolimits^* \alpha_n(i) - \gamma_1 \sum\nolimits^* \beta_n^2(i)\mu_n(i)\\
&= 1 - \frac{1}{2}\sum\nolimits_* (1+f_n(i))\mu_n(i) - \gamma_1 \sum\nolimits^* \beta_n^2(i)\mu_n(i)
\end{aligned}$$

ce qui permet d'obtenir, avec $A_n = \sum^* \beta_n^2(i)\mu_n(i) \geq 0$ et $B_n = [\sum_* \mu_n(i)]/2 \geq 0$,

$$\begin{aligned}
\mathcal{E}_n &\geq \frac{1}{2}\sum\nolimits_* (1+f_n(i))\mu_n(i) + \sum\nolimits_* \sqrt{f_n(i)}\mu_n(i) + \gamma_1 A_n\\
&\geq \left(\frac{1}{2}\max_{1\leq i\leq i_0} f_n(i)\mu_n(i)\right) \vee (\gamma_1 A_n + B_n)
\end{aligned}$$

Trouvons ensuite une autre constante $\gamma_2 > 0$ assez grande telle que pour tout $n \in \mathbb{N}$ et tout $i > i_0$,

$$f_n(i)\ln(f_n(i)) = (1+\beta_n(i))\ln(1+\beta_n(i)) \leq \beta_n(i) + \gamma_2\beta_n^2(i)$$

inégalité que l'on utilise pour obtenir une majoration de $\mathcal{L}_n$:

$$\begin{aligned}
\mathcal{L}_n &\leq \sum\nolimits_* f_n(i)\ln(f_n(i))\mu_n(i) + \sum\nolimits^* \beta_n(i)(\mu(i)+\alpha_n(i)) + \gamma_2 A_n\\
&= \sum\nolimits_* f_n(i)\ln(f_n(i))\mu_n(i) - \sum\nolimits_* f_n(i)\mu_n(i) + \sum\nolimits_* \mu_n(i) + \gamma_2 A_n\\
&\leq \sum\nolimits_* (1+f_n(i)\ln(f_n(i)))\mu_n(i) + \gamma_2 A_n\\
&\leq \left(2i_0 \max_{1\leq i\leq i_0} f_n(i)\ln(f_n(i))\mu_n(i)\right) \vee (2\gamma_2 A_n + 4B_n)
\end{aligned}$$

Soit $1 \leq i_1 \leq i_0$ tel que l'ensemble N_1 des $n \in \mathbb{N}$ tels que le maximum ci-dessus sur $1 \leq i \leq i_0$ soit atteint en i_1 soit infini. Alors pour $n \in N_1$, on a

$$\frac{\mathcal{E}_n}{\psi(\mathcal{L}_n)} \geq \frac{[f_n(i_1)\mu_n(i_1)/2] \vee (\gamma_1 A_n + B_n)}{\psi([2(N-1)f_n(i_1)\ln(f_n(i_1))\mu_n(i_1)] \vee (2\gamma_2 A_n + 4B_n))}$$

car on aura remarqué que ψ est croissante.

• Considérons d'abord le cas où il existe une infinité de $n \in N_1$ tels que $2\gamma_2 A_n + 4B_n > 2N f_n(i_1)\ln(f_n(i_1))\mu_n(i_1)$ et notons N_2 l'ensemble de ces n. On a donc pour $n \in N_2$,

$$\frac{\mathcal{E}_n}{\psi(\mathcal{L}_n)} \geq \frac{\gamma_1 A_n + B_n}{\psi(2\gamma_2 A_n + 4B_n)}$$

ainsi en utilisant que $\lim_{n\to\infty} A_n = 0 = \lim_{n\to\infty} B_n$, la positivité des A_n, B_n et le fait que pour x petit on a $\psi(x) \ll x$, on obtient une contradiction car le terme de droite diverge pour n grand vers $+\infty$.

• Intéressons-nous maintenant au cas où N_2 est fini, et où il existe une infinité de $n \in N_1 \setminus N_2$ tels que $x_0 \leq f_n(i_1)$ (notons N_3 l'ensemble de ces n). Alors en vertu de (5), pour tout $n \in N_3$ assez grand, il existe un unique $y_n \geq x_0$ tel que

$$\frac{K y_n \ln(y_n)}{\varphi(y_n)} = f_n(i_1)\ln(f_n(i_1))\mu_n(i_1)$$

et on a évidemment $\lim_{n\in N_3,\, n\to\infty} y_n = +\infty$ ainsi que $\psi(2(N-1) f_n(i_1)\ln(f_n(i_1))\mu_n(i_1)) = y_n/\varphi(y_n)$.

Vérifions que $K y_n/\varphi(y_n) \leq f_n(i_1)\mu_n(i_1)$. Ceci revient à voir que

$$\frac{f_n(i_1)\ln(f_n(i_1))\mu_n(i_1)}{\ln(y_n)} \leq f_n(i_1)\mu_n(i_1)$$

i.e.

$$\frac{\ln(f_n(i_1))}{\ln(y_n)} \leq 1$$

puis $y_n \geq f_n(i_1)$ (car $\ln(y_n) > 0$). Par stricte décroissance de $[x_0, +\infty[\ni x \mapsto x\ln(x)/\varphi(x)$, il suffit de constater que

$$\frac{K y_n \ln(y_n)}{\varphi(y_n)} \leq \frac{K f_n(i_1)\ln(f_n(i_1))}{\varphi(f_n(i_1))}$$

ce qui découle de la définition de y_n et de la majoration $\varphi(f_n(i_1))\mu_n(i_1) \leq K$. Ainsi en fin de compte on a bien pour $n \in N_3$, n assez grand

$$\begin{aligned}\frac{\mathcal{E}_n}{\psi(\mathcal{L}_n)} &\geq \frac{f_n(i_1)\mu_n(i_1)}{2\psi(2(N-1)f_n(i_1)\ln(f_n(i_1))\mu_n(i_1))}\\ &\geq \frac{f_n(i_1)\mu_n(i_1)\varphi(y_n)}{2y_n} \geq K/2\end{aligned}$$

ce qui est la contradiction recherchée.

• Enfin, toujours si N_2 est fini, considérons le cas où pour une infinité de $n \in N_1 \setminus N_2$, on a $x_0 \geq f_n(i_1)$ (notons N_4 l'ensemble de ces n). En utilisant à nouveau le fait que pour x petit on a $\psi(x) \ll x$, on obtient rapidement une contradiction car

$$\lim_{n \in N_4,\, n \to \infty} \frac{\mu_n(i_1)}{2\psi(2(N-1)x_0 \ln(x_0)\mu_n(i_1))} = +\infty$$

Pour terminer la démonstration, il reste à traiter le cas où $\mu \in \mathcal{P}_N^*$, mais celui-ci est encore plus facile, car les considérations précédentes sont valables et se simplifient, du fait que toutes les sommes $\sum_*$ sont nulles, et n'intervient finalement que A_n. □

Soit μ une probabilité quelconque sur $\{1, \cdots, N\}$ (on notera leur ensemble $\mathcal{P}_N$), posons désormais

$$D(\mu, \varphi, K) = \{f \in \mathcal{B}_N^+ \,/\, \mu(f^2) = 1 \text{ et } \mu(\varphi(f^2)) \leq K\}$$

puis

$$\alpha_{\varphi,K}(\mu) = \inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\psi(\mathcal{L}^{(\mu)}(f))}$$

en convenant que $\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)/\psi(\mathcal{L}^{(\mu)}(f)) = +\infty$ si $f \in D(\mu, \varphi, K)$ est telle que $\mathcal{L}^{(\mu)}(f) = 0$, i.e. si f est égal à 1 sur le support de μ.

On vérifie alors sans difficulté que l'on a aussi

$$C(N, K, \varphi) = \inf_{\mu \in \mathcal{P}_N} \alpha_{\varphi,K}(\mu)$$

Remarques. – a) La dépendance de ψ en N et K peut apparaître artificielle, d'ailleurs soit $\widetilde{\psi}$ la bijection entre $[0, x_0 \ln(x_0)/\varphi(x_0)]$ et $[0, x_0\varphi(x_0)]$ que l'on obtient en imposant que

$$\forall\, x \in [x_0, +\infty], \qquad \widetilde{\psi}(x \ln(x)/\varphi(x)) = x/\varphi(x)$$

On prolonge ensuite $\widetilde{\psi}$ sur $\mathbb{R}_+$ en la rendant par exemple constante sur $[x_0 \ln(x_0)/\varphi(x_0), +\infty[$. Supposons que pour tout $0 < A < 1$,

$$(6) \qquad \liminf_{x \to 0_+} \frac{\widetilde{\psi}(Ax)}{\widetilde{\psi}(x)} > 0$$

alors on aurait pu considérer pour $\mu \in \mathcal{P}_N$,

$$\widetilde{\alpha}_{\varphi,K}(\mu) = \inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widetilde{\psi}(\mathcal{L}^{(\mu)}(f))}$$

car on aboutit également à la conclusion que

$$\widetilde{C}(N,K,\varphi) \stackrel{\text{déf.}}{=} \inf_{\mu\in\mathcal{P}_N} \widetilde{\alpha}_{\varphi,K}(\mu) > 0$$

et de plus les constantes $\widetilde{C}(N,K,\varphi)$ sont clairement décroissantes en N et en K.

Cependant (6) n'est pas toujours satisfait, comme on s'en rend compte en considérant l'exemple donné par $\varphi(x) = x\ln(x)\ln(|\ln(x)|)$, pour lequel il faut prendre

$$\forall\, x \geq 0, \qquad \psi(x) = \exp\left(-\exp\left(\frac{8K}{x}\right)\right)\exp\left(-\frac{8K}{x}\right)$$

b) Plus généralement, seul importe le comportement de $\psi(x)$ (ou de $\widetilde{\psi}(x)$ si (6) est satisfait) pour $x > 0$ petit, ainsi si pour une certaine fonction croissante $\widehat{\psi} : I\!R_+ \to I\!R_+$ on a $\widehat{\psi} = \mathcal{O}(\psi)$ (la constante de majoration pouvant dépendre de N et K, remarquons notamment que $\widehat{\psi}$ sera aussi bornée), alors on a encore

$$\widehat{C}(N,K,\varphi) \stackrel{\text{déf.}}{=} \inf_{\mu\in\mathcal{P}_N} \widehat{\alpha}_{\varphi,K}(\mu) > 0$$

en ayant évidemment posé

$$\widehat{\alpha}_{\varphi,K}(\mu) = \inf_{f\in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))}$$

Dans la pratique on choisira plutôt de telles fonctions $\widehat{\psi}$ admettant une forme simple. Ainsi par exemple, si pour tout x assez grand $\varphi(x) = x(\ln(x))^{1+\eta}$ (resp. $\varphi(x) = x^{1+\eta}$), avec $\eta > 0$, alors on a pour x petit $\widetilde{\psi}(x) = K^{-1/\eta}x^{(1+\eta)/\eta}$ (resp. $\widetilde{\psi}(x) \sim -x/(\eta\ln(x))$) et on préférera donc prendre $\widehat{\psi}(x) = 1 \wedge x^{(1+\eta)/\eta}$ (resp. $\widehat{\psi}(x) = (-x/\ln(e^{-1}x))\mathbf{1}_{x\leq 1} + \mathbf{1}_{x>1}$).

c) Soit $\widehat{\psi}$ une fonction comme ci-dessus. Montrons qu'il existe une application croissante de classe C^1 sur $I\!R_+^*$, $\check{\psi}$, telle que pour tout $x > 0$,

$$\widehat{\psi}(x) \leq \check{\psi}(x) \leq 2\widehat{\psi}(x)$$

Pour ceci, il suffit de modifier légèrement la procédure usuelle de régularisation par convolution : en effet, on construit d'abord une application croissante de classe C^1, $\epsilon : I\!R_+^* \to I\!R_+^*$, telle que pour tout $x > 0$,

$$\widehat{\psi}(x + 2\epsilon(x)) \leq 2\widehat{\psi}(x)$$

(ce qui peut se faire en considérant dans un premier temps une telle application qui soit croissante et constante par morceaux dans tout intervale compact de $\mathbb{R}_+^*$, que l'on convole ensuite avec une fonction régularisante comme ci-dessous, mais dont le support est égal à $[-2,0]$). Soit $\rho : \mathbb{R} \to [0,1]$ de classe C^1, vérifiant $\int \rho(x)\,dx = 1$ et de support inclus dans $[0,2]$. Il suffit alors de considérer pour $x > 0$,

$$\check{\psi}(x) = \int_{[0,2]} \widehat{\psi}(x + \epsilon(x)t)\rho(t)\,dt$$

Notons que de la même manière on construirait une telle fonction $\check{\psi}$ de classe C^∞ sur $\mathbb{R}_+^*$. Ainsi il existe toujours une fonction $\widehat{\psi}$ qui satisfait la condition de la remarque précédente et qui soit de classe C^1.

d) Parfois il est commode également de pouvoir perturber légèrement φ (voir par exemple la fin de cette section). Supposons que l'on dispose d'une fonction $\varphi_1 : \mathbb{R}_+ \to \mathbb{R}$ qui soit minorée et telle qu'il existe une constante $a > 0$ telle que pour tout x assez grand, on ait

$$\varphi_1(x) \geq a\varphi(x)$$

Alors pour toute application $\widehat{\psi}$ associée à φ comme dans la remarque (b) ci-dessus, on a

$$\inf_{\mu \in \mathcal{P}_N} \inf_{f \in D(\mu,\varphi_1,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))} > 0$$

car ceci découle trivialement de l'inclusion $D(\mu,\varphi_1,K) \subset D(\mu,\varphi,a^{-1}(K+b))$, où $b \geq 0$ est tel que pour tout $x \in \mathbb{R}_+$, $\varphi_1(x) + b \geq a\varphi(x)$.

Supposons maintenant de plus que φ soit au moins de classe C^4 sur $\mathbb{R}_+^*$ et que

$$\frac{d}{dx}\left(x\frac{d^3}{dx^3}(\varphi(x^2))\right)$$

ne s'annule qu'un nombre fini de fois sur $\mathbb{R}_+^*$, disons $\widetilde{N} \in \mathbb{N}$. Le résultat qui va nous permettre de sortir du cas d'ensembles finis s'énonce alors

PROPOSITION 4. – *Soit $N = \widetilde{N} + 4$, $n \geq N$ et μ une probabilité sur $\{1, \cdots, n\}$. On a alors pour toute fonction $\widehat{\psi}$ comme dans la discussion qui précède et de classe C^1 sur $\mathbb{R}_+^*$,*

$$\inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))} \geq \widehat{C}(N,K,\varphi)$$

Preuve. – Commençons par montrer que l'infimum dans la proposition ci-dessus est atteint. En effet, s'il en était autrement, il existerait une suite $(f_n)_{n\in\mathbb{N}}$ d'éléments de $D(\mu,\varphi,K)$ telle que

$$\lim_{n\to\infty} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f_n,f_n)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f_n))} = \inf_{f\in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))}$$

et

$$\lim_{n\to\infty} f_n = 1$$

Cependant en effectuant un développement limité on s'aperçoit que

$$\mathcal{L}^{(\mu)}(f_n) = 2\mathcal{E}^{(\mu)}_{I-E_\mu}(f_n,f_n) + o(\mathcal{E}^{(\mu)}_{I-E_\mu}(f_n,f_n))$$

et il existe donc un $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$,

$$\widehat{\psi}(\mathcal{L}^{(\mu)}(f_n)) \leq \widehat{\psi}(3\mathcal{E}^{(\mu)}_{I-E_\mu}(f_n,f_n))$$

On obtient alors une contradiction comme dans la preuve de la proposition 3, du fait que $\widehat{\psi}(x) \ll x$ pour $x > 0$ petit.

Soit donc $f_0 \in D(\mu,\varphi,K)$ tel que

$$\frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f_0,f_0)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f_0))} = \inf_{f\in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))}$$

et notons $K_0 = \mu(\varphi(f_0^2))$ et $I = \{1 \leq i \leq n \,/\, f_0(i) > 0\}$. Soit également $\mathcal{B}(I)$ l'ensemble des fonctions définies sur I, que l'on identifie avec les fonctions définies sur $\{1,\cdots,n\}$ qui s'annulent en dehors de I. L'application F définie par

$$\mathcal{B}(I) \ni f \to \frac{1-(\mu(f))^2}{\widehat{\psi}(\int f^2 \ln(f^2)\, d\mu)}$$

est différentiable au voisinage de f_0 et sa dérivée dans la direction $h \in \mathcal{B}(I)$ est donnée par

$$\frac{-2\mu(f_0)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f_0))}\mu(h) - \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f_0,f_0)}{\widehat{\psi}^2(\mathcal{L}^{(\mu)}(f_0))}\widehat{\psi}'(\mathcal{L}^{(\mu)}(f_0))\mu((4f_0\ln(f_0)+2f_0)h)$$

Ainsi en considérant f_0 comme un minimum de F sur la surface $\{f \in \mathcal{B}(I) \,/\, \mu(f^2) = 1 \text{ et } \mu(\varphi(f^2)) = K_0\}$, on voit qu'il existe des constantes $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ (les multiplicateurs de Lagrange) telles que pour tout $p \in I$,

$$\lambda_1 \left[\frac{-2\mu(f_0)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f_0))} - \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f_0, f_0)\widehat{\psi}'(\mathcal{L}^{(\mu)}(f_0))}{\widehat{\psi}^2(\mathcal{L}^{(\mu)}(f_0))}(4f_0(p)\ln(f_0(p)) + 2f_0(p)) \right]$$
$$+\lambda_2 2f_0(p) + \lambda_3 2\varphi'(f_0^2(p))f_0(p) = 0$$

Les valeurs de f_0 (outre 0) sont donc aussi des solutions d'une équation en x du type

$$K_1 + K_2 x + K_3 x \ln(x) + K_4 \varphi'(x^2)x = 0$$

où K_1, K_2, K_3, K_4 sont certains réels. Notons $G(x)$ le membre de gauche ci-dessus, on a

$$\frac{d^2}{dx^2}G(x) = \frac{K_3}{x} + K_4 \frac{d^3}{dx^3}(\varphi(x^2))$$

et cette fonction admet autant de zéros sur $\mathbb{R}_+^*$ que $K_3 + K_4 x \frac{d^3}{dx^3}\varphi(x^2)$. Il en résulte, vu l'hypothèse supplémentaire faite sur φ, que f_0 prend au plus N valeurs, disons $v_1, \cdots, v_N$.

Soient pour $1 \leq i \leq N$, $A_i = \{1 \leq p \leq n \,/\, f_0(p) = v_i\}$, et $\widetilde{\mu}$ la probabilité sur $\{1, \cdots, N\}$ définie par

$$\forall\, 1 \leq i \leq N, \qquad \widetilde{\mu}(i) = \mu(A_i)$$

il apparaît que

$$\frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f_0, f_0)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f_0))} \geq \widehat{\alpha}_{\varphi,K}(\widetilde{\mu}) \geq \widehat{C}(N, K, \varphi)$$

d'où le résultat annoncé. □

Ainsi sous les hypothèses de la proposition précédente, on a pour tout $n \geq N$,

$$\inf_{\mu \in \mathcal{P}_n} \inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f, f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))} = \widehat{C}(N, K, \varphi)$$

D'autre part, il n'est pas difficile d'assouplir la condition donnée sur φ avant la proposition 4.

Commençons par remarquer que s'il existe $a \in \mathbb{R}$ tel que pour tout x assez grand on ait

$$x\frac{d^3}{dx^3}(\varphi(x^2)) \leq a$$

alors en intégrant trois fois, il apparaît immédiatement que

$$\limsup_{x\to+\infty} \frac{\varphi(x^2)}{x^2\ln(x^2)} < +\infty$$

Ainsi sous nos hypothèses on a nécessairement

$$\limsup_{x\to+\infty} x\frac{d^3}{dx^3}(\varphi(x^2)) = +\infty$$

et $\frac{d}{dx}(x\frac{d^3}{dx^3}\varphi(x^2))$ finit donc par être strictement positif (ce qui à son tour implique qu'en fait ci-dessus $\lim_{x\to+\infty} x\frac{d^3}{dx^3}(\varphi(x^2)) = +\infty$, cette divergence permettant de retrouver que $\lim_{x\to+\infty} x^2\ln(x^2)/\varphi(x^2) = 0$).

Réciproquement, supposons seulement qu'il existe $x_1 \geq 0$ tel que φ soit de classe C^4 sur $]x_1, +\infty[$ et satisfasse pour $x > x_1$,

$$\frac{d}{dx}\left(x\frac{d^3}{dx^3}(\varphi(x^2))\right) \geq 0$$

alors la proposition 4 est encore vérifiée (en faisant toujours l'hypothèse que φ a le comportement asymptotique décrit avant la proposition 3) et on peut même y prendre $N = 5$.

En effet, il est possible de construire une fonction $\varphi_1 : \mathbb{R}_+ \to \mathbb{R}_+^*$ de classe C^4 sur $\mathbb{R}_+$ telle que φ_1 majore φ, vérifie $\varphi_1(x) = \varphi(x)$ pour tout $x \geq x_1 + 1$ et ne satisfait pas $\frac{d}{dx}(x\frac{d^3}{dx^3}(\varphi_1(x^2))) \equiv 0$ sur $]0, x_1 + 1[$. Par ailleurs, en considérant des applications de la forme $\mathbb{R}_+ \ni x \mapsto a/(b + \sqrt{x})$ avec $a, b > 0$ bien choisis, on montre sans difficulté que l'on peut trouver une fonction bornée $h : \mathbb{R}_+ \to \mathbb{R}_+$ telle qu'en posant $\varphi_2 = \varphi_1 + h$, $\frac{d}{dx}(x\frac{d^3}{dx^3}(\varphi_2(x^2)))$ ne s'annule qu'une fois sur $\mathbb{R}_+^*$. Si on a bien $\lim_{x\to+\infty} x\ln(x)/\varphi_2(x) = 0$, par contre il n'est pas clair a priori que $x \mapsto x\ln(x)/\varphi_2(x)$ finit par décroître, néanmoins ceci n'est pas grave d'après la remarque (d) qui précède et la preuve de la proposition 4, car φ_2 majore φ. Ainsi, il existe une constante $\widehat{C}(K, \varphi) > 0$ ne dépendant pas de μ ni de n, telle que

$$\inf_{f\in D(\mu,\varphi_2,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))} \geq \widehat{C}(K, \varphi)$$

Soit $K_1 = \sup_{x \geq 0} \varphi_2(x) - \varphi(x) < +\infty$, il reste à utiliser que si $\mu \in \mathcal{P}_n$ et $K > 0$, alors $D(\mu, \varphi, K) \subset D(\mu, \varphi_2, K + K_1)$, pour se persuader qu'il existe une autre constante $\widehat{C}(K, \varphi) > 0$, telle que

$$\inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\widehat{\psi}(\mathcal{L}^{(\mu)}(f))} \geq \widehat{C}(K, \varphi)$$

En résumé, en utilisant la remarque (c), on a donc montré le résultat suivant, pour lequel la convexité de φ est en fait inutile,

COROLLAIRE 5. – *Soit φ un bon majorant (de l'application $\mathbb{R}_+ \ni x \mapsto x\ln(x)$) et $K > 0$, il existe des fonctions croissantes et bornées ψ telles que pour x grand*

$$\psi(8Kx\ln(x)/\varphi(x)) = \mathcal{O}(x/\varphi(x))$$

et si une telle application est fixée, alors

$$\inf_{n \in \mathbb{N}^*} \inf_{\mu \in \mathcal{P}_n} \inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\psi(\mathcal{L}^{(\mu)}(f))} > 0$$

Notons que l'introduction de ψ n'est pas innocente, car on a vérifié dans [11] que

$$\inf_{\mu \in \mathcal{P}_3} \inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)}{\mathcal{L}^{(\mu)}(f)} = 0$$

4. INÉGALITÉS ÉNERGIE-ENTROPIE RESTREINTES

Notre but ici est d'étendre les résultats de la section précédente, ce qui est relativement immédiat vu nos hypothèses.

On suppose que φ est un bon majorant fixé et que ψ est l'une des fonctions lui est associée comme dans le corollaire 5. Soit μ une probabilité sur un espace mesurable quelconque $(S, \mathcal{S})$. Comme d'habitude, on note pour $K > 0$ fixé (celui qui intervient aussi dans la définition de ψ),

$$D(\mu, \varphi, K) = \{f \in \mathbb{L}^2(\mu) \,/\, \mu(f^2) = 1 \text{ et } \mu(\varphi(f^2)) \leq K\}$$

PROPOSITION 6. – *Il existe une constante universelle (i.e. ne dépendant pas de l'espace de probabilité $(S, \mathcal{S}, \mu)$) $A > 0$ telle que pour tout $f \in D(\mu, \varphi, K)$,*

$$\mathcal{E}_{I-E_\mu}(f,f) \geq A\psi(\mathcal{L}(f))$$

où $\mathcal{L}(f) = \int f^2 \ln(f^2)\, d\mu$.

Preuve. – On va montrer que l'on peut prendre pour A la dernière expression apparaissant dans le corollaire 5. Du fait que pour tout $f \in \mathbb{L}^2(\mu)$, $\mathcal{E}_{I-E_\mu}(f,f) \geq \mathcal{E}_{I-E_\mu}(|f|,|f|)$ et $\mathcal{L}(f) = \mathcal{L}(|f|)$, il suffit de considérer des fonctions positives. Soit donc $f \in D(\mu,\varphi,K)$, $f \geq 0$. Pour $m \in \mathbb{N}$ et $0 \leq k \leq m2^m$, posons

$$A_{m,k} = \begin{cases} \{k2^{-m} \leq f < (k+1)2^{-m}\} & \text{, si } k < m2^m \\ \{f \geq m\} & \text{, sinon} \end{cases}$$

$$a_{m,k} = \sqrt{\int_{A_{k,m}} \frac{f^2}{\mu(A_{k,m})}\, d\mu}$$

en convenant que cette dernière quantité est nulle si $\mu(A_{m,k}) = 0$, puis notons

$$f_m = \sum_{k=0}^{m2^m} a_{m,k} \mathbf{1}_{A_{m,k}}$$

de sorte que

$$\lim_{m\to\infty} \mu(f_m) = \mu(f)$$
$$\lim_{m\to\infty} \mu(f_m^2 \ln(f_m^2)) = \mu(f^2 \ln(f^2))$$

De plus pour tout $m \geq 0$, on a $\mu(f_m^2) = \mu(f^2) = 1$, et par l'inégalité de Jensen il apparaît que

$$\begin{aligned} \mu(\varphi(f_m^2)) &= \sum_{k=0}^{m2^m} \varphi(a_{m,k}^2)\mu(A_{m,k}) \\ &= \sum_{k=0}^{m2^m} \varphi\left(\int_{A_{k,m}} \frac{f^2}{\mu(A_{k,m})}\, d\mu\right)\mu(A_{m,k}) \\ &\leq \sum_{k=0}^{m2^m} \int_{A_{k,m}} \varphi(f^2)\, d\mu \\ &\leq K \end{aligned}$$

On peut alors appliquer le corollaire 5 pour se rendre compte que

$$\mathcal{E}_{I-E_\mu}(f_m, f_m) \geq A\psi(\mathcal{L}(f_m))$$

puis en passant à la limite pour m grand, on obtient la même inégalité pour f. □

Nous pouvons maintenant à nouveau considérer un noyau de Markov P qui laisse μ invariante.

PROPOSITION 7. – *Supposons que* $I - PP^*$ *admette dans* $\mathbb{L}^2(\mu)$ *un trou spectral* $\lambda \overset{\text{déf.}}{=} \lambda(I - PP^*)$, *alors avec les notations précédentes,*

$$\inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}_{I-PP^*}(f,f)}{\psi(\mathcal{L}(f))} \geq A\lambda$$

Preuve. – En effet, à l'instar de la démonstration du corollaire 5.4 de Diaconis et Saloff-Coste [3], il suffit d'écrire que pour tout $f \in D(\mu, \varphi, K)$, $f^2 \neq \mathbf{I}$,

$$\frac{\mathcal{E}_{I-PP^*}(f,f)}{\psi(\mathcal{L}(f))} = \frac{\mathcal{E}_{I-PP^*}(f,f)}{\mathcal{E}_{I-E_\mu}(f,f)} \frac{\mathcal{E}_{I-E_\mu}(f,f)}{\psi(\mathcal{L}(f))} \qquad \square$$

Remarque. – On pourrait évidemment remplacer λ par l'une des constantes suivantes

$$\inf_{f \in D(\mu,\varphi,K)} \frac{\mathcal{E}_{I-PP^*}(f,f)}{\mathcal{E}_{I-E_\mu}(f,f)}$$

$$\inf_{f \in \mathbb{L}^\Phi(\mu)} \frac{\mathcal{E}_{I-PP^*}(f,f)}{\mathcal{E}_{I-E_\mu}(f,f)}$$

où Φ est une fonction de Young dominant la fonction convexe $\mathbb{R} \ni x \mapsto \varphi(x^2)$, malheureusement nous ne savons pas évaluer celles-ci, sauf à estimer le trou spectral de $I - PP^*$!

Pour des exemples d'évaluation de trou spectral par la méthode de Poincaré dans le cas où S n'est pas fini (dénombrable ou non), on renvoit à un article récent de Rosenthal [14].

Mais il existe des situations où le trou spectral de $I - PP^*$ se calcule naturellement et ci-dessous se trouvent trois exemples pour lesquels on pourrait en fait expliciter tout le spectre :

a) Supposons que $S = (\mathbb{R}/\mathbb{Z})^n$, avec $n \in \mathbb{N}^*$, que l'on munit de sa tribu borélienne $\mathcal{S}$. Soit μ la mesure de Lebesgue et $p \in \mathbb{L}^2(\mu)$, satisfaisant $p \geq 0$ et $\mu(p) = 1$. On définit un noyau markovien P en posant (une version de p ayant été choisie)

$$\forall\, x \in S, \forall\, A \in \mathcal{S}, \qquad P(x, A) = \int p(y - x)\mathbf{I}_A(y)\, \mu(dy)$$

Il est clair que P admet μ pour probabilité invariante et qu'une version de P^* est donnée par

$$\forall\, x \in S, \forall\, A \in \mathcal{S}, \qquad P^*(x, A) = \int p(x - y)\mathbf{I}_A(y)\, \mu(dy)$$

(notamment si p est pair alors μ est réversible pour P).

On peut calculer le trou spectral par le biais de l'analyse de Fourier : pour $k \in \mathbb{N}^n$, posons

$$e_k : S \ni x \mapsto \exp(2\pi i \langle k, x\rangle) \in \mathbb{C}$$

Cette base $(e_k)_{k\in\mathbb{N}^n}$ de $\mathbb{L}^2(S,\mu,\mathbb{C})$ diagonalise P, car on vérifie que pour tout $k \in \mathbb{N}^n$, $Pe_k = p_{-k}e_k$, où $p_{-k} = \mu(pe_k)$ est un des coefficients de Fourier de p. Il apparaît ainsi que les valeurs propres de PP^* sont les $|p_k|^2$, pour $k \in \mathbb{N}^n$, de sorte que le trou spectral de $I - PP^*$ est $1 - \sup_{k\in\mathbb{N}^n\setminus\{0\}} |p_k|^2 = 1 - \max_{k\in\mathbb{N}^n\setminus\{0\}} |p_k|^2$ (car $\lim_{|k|\to+\infty} |p_k| = 0$), quantité qui est toujours non nulle.

Plus généralement, supposons $\mathcal{S}$ séparable et soit P un noyau markovien admettant une densité par rapport à une probabilité invariante μ, c'est-à-dire qu'il existe une fonction $S^2 \ni (x,y) \mapsto p(x,y) \in \mathbb{R}_+$ telle que l'on ait la représentation

$$\forall\, x \in S,\ \forall\, A \in \mathcal{S}, \qquad P(x,A) = \int p(x,y)\mathbf{1}_A(y)\,\mu(dy)$$

Pour $x,y \in S$, notons $q(x,y) = \int p(x,z)p(y,z)\,\mu(dz)$ et faisons l'hypothèse que

$$\int q^2(x,y)\,\mu(dx)\mu(dy) < +\infty \tag{7}$$

et que les seules fonctions harmoniques pour PP^* dans $\mathbb{L}^2(\mu)$ sont les constantes. Alors $I - PP^*$ admet un trou spectral strictement positif.

En effet, remarquons que le noyau PP^* admet aussi une version à densité, et que cette dernière n'est autre que q. La condition (7) implique alors que PP^* est un opérateur compact dans $\mathbb{L}^2(\mu)$ (voir par exemple l'exercice 52 p. 518 de [4]), et comme il est également symétrique, il est diagonalisable par séparabilité de $\mathbb{L}^2(\mu)$. La compacité entraîne qu'il ne peut avoir une suite de valeurs propres différentes de 1 convergeant vers 1, et puisque l'on a fait aussi l'hypothèse que l'espace propre associé à 1 est la droite des fonctions constantes, on en déduit le résultat annoncé.

Notons que la condition

$$\int p^2(x,y)\,\mu(dx)\mu(dy) < +\infty$$

implique (7) par une application de l'inégalité de Cauchy-Schwarz, ou plus directement que P est compact, d'où la compacité requise de PP^*.

b) Voici un exemple bien connu où le critère ci-dessus ne s'applique jamais : pour $t > 0$ et $\beta > 0$, notons $R_t^{(\beta)}$ le noyau markovien défini sur $\mathbb{R}^d$, $d \geq 1$, en posant pour tout $x \in \mathbb{R}^d$,

$$R_t^{(\beta)}(x, dy) = \left(\frac{\beta}{\pi(1 - \exp(-2\beta t))}\right)^{\frac{d}{2}} \exp\left(-\beta\frac{|y - \exp(-\beta t)x|^2}{1 - \exp(-2\beta t)}\right) dy$$

où $| \cdot |$ désigne la norme euclidienne usuelle.

Il apparaît que $(R_t^{(\beta)})_{t\geq 0}$ est le semi-groupe d'Ornstein-Uhlenbeck à la température β^{-1}, dont le générateur est $L_\beta \cdot \stackrel{\text{déf.}}{=} \triangle/2 \cdot -\beta\langle x, \nabla\cdot\rangle$ (du moins sa restriction sur les fonctions de classe C^∞ à support compact). Or le spectre de ce dernier est bien connu, il s'agit de $-\beta\mathbb{N}$ (les vecteurs propres associés sont certains polynômes d'Hermite, à une homothétie près de rapport $(2\beta)^{-\frac{1}{2}}$ de l'espace, cf. par exemple [1]), notamment L_β admet β pour trou spectral.

Par ailleurs, rappelons que $R_t^{(\beta)}$ est réversible par rapport à la mesure gaussienne ν_β de moyenne nulle et de matrice de covariance $(2\beta)^{-1}\mathrm{Id}$, avec Id la matrice identité $d \times d$.

On en déduit que

$$\lambda(I - R_t^{(\beta)}R_t^{(\beta)*}) = \lambda(I - R_{2t}^{(\beta)}) = 1 - \exp(-2\beta t)$$

c) Nous allons nous intéresser ici à une chaîne de Markov considérée par Föllmer [6] (qui lui-même l'avait emprunté à Wasserstein [19]).

On prend $S = \{0, 1\}^{\mathbb{Z}}$ muni de sa tribu produit usuelle $\mathcal{S}$. Soit $x = (x_i)_{i\in\mathbb{Z}} \in S$, pour $i \in \mathbb{Z}$ on définit une probabilité $P_i(x, \cdot)$ sur $\{0, 1\}$ par

$$(8) \qquad P_i(x, \{1\}) = \begin{cases} p_i & , \text{si } x_{i+1} = 1 \\ 1 - p_i & , \text{si } x_{i+1} = 0 \end{cases}$$

où les p_i sont des nombres appartenant à $[1/2, 1[$. Soit ensuite la probabilité produit sur $\mathcal{S}$ donnée par

$$P(x, \cdot) = \bigotimes_{i\in\mathbb{Z}} P_i(x, \cdot)$$

L'application P ainsi construite sur $S \times \mathcal{S}$ est bien un noyau de probabilités de transitions.

Pour $r \in [0,1]$, soit ν_r la probabilité produit définie sur $\mathcal{S}$ par

$$\forall\, i \in \mathbb{Z}, \qquad \nu_r(x_i = 1) = r_i \stackrel{\text{déf.}}{=} \frac{1}{2} + \left(r - \frac{1}{2}\right) \prod_{j \geq i} (2p_j - 1)$$

on constate sans difficulté que chacune de ces probabilités est invariante pour P. Fixons donc un tel r et prenons $\mu = \nu_r$.

On vérifie par un argument usuel que P^* admet une version régulière sous forme d'un noyau de probabilités de transitions tel que pour tout $x \in S$, $P^*(x, \cdot)$ est également un produit $\otimes_{i\in\mathbb{Z}} P_i^*(x, \cdot)$, où la probabilité sur le $i^{\text{ième}}$ facteur $\{0,1\}$ est donnée par

$$P_i^*(x,\{1\}) = \begin{cases} \dfrac{r_i}{r_{i-1}} p_{i-1} & , \text{si } x_{i-1} = 1 \\[2ex] \dfrac{r_i}{1 - r_{i-1}}(1 - p_{i-1}) & , \text{si } x_{i-1} = 0 \end{cases}$$

(on aura remarqué que $0 < r_j < 1$).

On calcule alors que la composition $Q(x,\cdot) = PP^*(x,\cdot)$ est également une probabilité produit $Q(x,\cdot) = \otimes_{i\in\mathbb{Z}} Q_i(x, \cdot)$ dont le $i^{\text{ième}}$ facteur est défini par

$$Q_i(x,\{1\}) = \begin{cases} q_i(1) & , \text{si } x_i = 1 \\ q_i(0) & , \text{si } x_i = 0 \end{cases}$$

avec pour tout $i \in \mathbb{Z}$,

$$\begin{aligned} q_i(0) &= \frac{r_i}{r_{i-1}(1 - r_{i-1})} p_{i-1}(1 - p_{i-1}) \\ q_i(1) &= \frac{r_i}{r_{i-1}(1 - r_{i-1})} ((1 - r_{i-1})p_{i-1}^2 + r_{i-1}(1 - p_{i-1})^2) \end{aligned}$$

et on vérifie immédiatement que $0 < q_i(0), q_i(1) < 1$.

Remarquons que le noyau Q fait évoluer les différentes coordonnées de S indépendamment, et vu les structures produits de Q et de μ, chaque Q_i (considéré comme un noyau de probabilités de transitions sur $\{0,1\}$) admet nécessairement la restriction μ_i de μ au $i^{\text{ième}}$ facteur de S comme probabilité réversible, pour $i \in \mathbb{Z}$. La matrice Q_i est donc diagonalisable et ses valeurs propres sont 1 et $\lambda_i = -q_i(0) + q_i(1) < 1$ (formule de la trace). D'autre part on se convainc facilement que

$$\sup_{f \in \mathbb{L}^2(\mu),\, \mu(f)=0,\, f \not\equiv 0} \frac{\mu(fQf)}{\mu(f^2)} = \sup_{i\in\mathbb{Z}} \lambda_i$$

de sorte que

$$\lambda(I - PP^*) = 1 - \sup_{i\in\mathbb{Z}} \lambda_i$$

Montrons que si $r \neq 0$ et $r \neq 1$,

$$\sup_{i\in\mathbb{Z}} \lambda_i = \sup_{i\in\mathbb{Z}}(2p_i - 1)^2 = (2\sup_{i\in\mathbb{Z}} p_i - 1)^2 \tag{9}$$

à partir de l'expression

$$\lambda_i = q_i(1) - q_i(0) = (2p_{i-1} - 1)^2 \frac{r_i(1 - r_i)}{r_{i-1}(1 - r_{i-1})} \tag{10}$$

et pour cela considérons deux cas.

• Si $\prod_{l\geq 0}(2p_l - 1) > 0$, alors nécessairement $\lim_{i\to+\infty} p_i = 1$ et $\lim_{i\to+\infty} r_i = r$, ainsi si $0 < r < 1$, on a immédiatement

$$\lim_{i\to+\infty} \lambda_i = 1$$

d'où la validité de (9), puisque ses trois termes sont nuls.

• Si $\prod_{l\geq 0}(2p_l - 1) = 0$, alors pour tous $j \in \mathbb{Z}$, on a $\prod_{l\geq j}(2p_l - 1) = 0$, i.e. $r_j = 1/2$. L'égalité annoncé (9) est alors claire, puisque $\lambda_i = (2p_{i-1} - 1)^2$.

Notons que dans cet exemple l'hypothèse $\sup_{i\in\mathbb{Z}} p_i < 1$ est équivalente à la condition de Dobrushin (voir [6], équation (1.9)), et dans le cas où $r \in]0,1[$, on a donc montré que celle-ci est équivalente à l'existence d'un trou spectral strictement positif. Mais nous allons maintenant voir que dans les cas où $r = 0$ ou $r = 1$, on peut avoir un trou spectral sans que la condition de Dobrushin soit satisfaite.

Supposons par exemple que $r = 0$. On remarque tout d'abord que si $\prod_{l\geq 0}(2p_l - 1) = 0$, alors l'égalité (9) est encore valable. Faisons donc l'hypothèse que $\prod_{l\geq 0}(2p_l - 1) > 0$. D'autre part puisque $\lim_{i\to-\infty} r_i$ existe toujours par monotonicité et qu'elle appartient à $]0,1[$, on voit que $\sup_{i\leq 0} \lambda_i < 1$ équivaut à $\limsup_{i\to-\infty} p_i < 1$. Supposons donc de plus ceci réalisé, et écrivons pour tout $l \geq 0$, $2p_l - 1 = 1 - \epsilon_l$, de sorte que $\epsilon_l > 0$ et $\sum_{l\geq 0} \epsilon_l < +\infty$. On calcule que pour i grand, on a l'équivalence

$$\lambda_i \sim \frac{\sum_{j\geq i} \epsilon_j}{\sum_{j\geq i-1} \epsilon_j}$$

ainsi dans cette situation on a l'existence d'un trou spectral si et seulement si

$$\liminf_{i\to+\infty} \frac{\epsilon_i}{\sum_{j\geq i}\epsilon_j} > 0$$

De même si $r = 1$ et si $\prod_{l\geq 0}(2p_l - 1) > 0$, l'existence d'un trou spectral est équivalente à

$$\begin{gathered} \limsup_{i\to-\infty} p_i < 1 \\ \liminf_{i\to+\infty} \frac{p_i - 1}{\sum_{j\geq i}(p_j - 1)} > 0 \end{gathered}$$

Ces conditions sont par exemple satisfaites si on prend $p_i = 3/4$ pour $i \leq 0$ et $p_i = 1 - \exp(-i)$ pour $i \geq 1$.

Enfin notons que sachant que (1) est vérifié par une certaine chaîne de Markov, on peut en déduire que des chaînes relativement proches admettent également un trou spectral, par des procédés de comparaisons (cf. [15]). Nous allons illustrer ceci sur un exemple qui nous sera utile ultérieurement.

Reconsidérons l'exemple (b) ci-dessus, où l'on prend $t = \beta^{-1}$. Soit $V : \mathbb{R}^d \to \mathbb{R}_+$ une fonction mesurable bornée, et posons pour tout $x \in \mathbb{R}^d$,

$$\widetilde{P}_\beta(x, dy) = \exp(-\beta(V(y) - V(x))_+) R^{(\beta)}_{\beta^{-1}}(x, dy) + A_\beta(x)\delta_x(dy)$$

avec $A_\beta(x) = 1 - \int R^{(\beta)}_{\beta^{-1}}(x, dy) \exp(-\beta(V(y) - V(x))_+)$.

Il s'agit d'une perturbation de type Metropolis du noyau $R^{(\beta)}_{\beta^{-1}}$, ce qui permet de se rendre compte que $\widetilde{P}_\beta$ est réversible par rapport à la probabilité $\widetilde{\mu}_\beta$ définie par

$$\widetilde{\mu}_\beta(dx) = Z_\beta^{-1} \exp(-\beta U(x))\, dx$$

avec $U(x) = |x|^2 + V(x)$ et $Z_\beta = \int \exp(-\beta U(x))\, dx$, on aura constaté que U est une fonction positive, ce qui assure que l'application $\mathbb{R}_+^* \ni \beta \mapsto Z_\beta$ est décroissante.

Cependant, pour toute fonction $f \in \mathbb{L}^2(\widetilde{\mu}_\beta)$, on a

$$\begin{aligned}
&\mathcal{E}_{I-\widetilde{P}_\beta \widetilde{P}_\beta^*}(f,f) \\
&\quad = \mathcal{E}_{I-\widetilde{P}_\beta^2}(f,f) \\
&\quad = \frac{1}{2}\int \widetilde{\mu}_\beta(dx)\widetilde{P}_\beta(x,dz)\widetilde{P}_\beta(z,dy)(f(y)-f(x))^2 \\
&\quad = \frac{1}{2Z_\beta}\left(\frac{\pi}{\beta}\right)^{\frac{d}{2}}\left[\int \nu_\beta(dx)R^{(\beta)}_{\beta^{-1}}(x,dz)R^{(\beta)}_{\beta^{-1}}(z,dy)\right. \\
&\qquad \exp(-\beta[V(x)+(V(z)-V(x))_+ + (V(y)-V(z))_+])(f(y)-f(x))^2 \\
&\qquad \left. + \int \nu_\beta(dx)[A_\beta(x)+A_\beta(y)]R^{(\beta)}_{\beta^{-1}}(x,dy)(f(y)-f(x))^2\right] \\
&\quad \geq \exp(-2M\beta)\frac{1}{2}\int \nu_\beta(dx)R^{(\beta)}_{\beta^{-1}}(x,dz)R^{(\beta)}_{\beta^{-1}}(z,dy)(f(y)-f(x))^2
\end{aligned}$$

avec $M = \sup_{x,y\in\mathbb{R}^d}|V(y)-V(x)|$, et il en découle que

$$\lambda(I-\widetilde{P}_\beta\widetilde{P}_\beta^*) \geq \exp(-3M\beta)\lambda(I-R^{(\beta)2}_{\beta^{-1}}) = (1-e^{-2})\exp(-3M\beta)$$

car

$$\begin{aligned}
\mathcal{E}_{I-E_{\widetilde{\mu}_\beta}}(f,f) &= \int \widetilde{\mu}_\beta(dx)(f(x)-\widetilde{\mu}_\beta(f))^2 \\
&\leq \int \widetilde{\mu}_\beta(dx)(f(x)-\nu_\beta(f))^2 \\
&\leq \exp(M\beta)\int \nu_\beta(dx)(f(x)-\nu_\beta(f))^2
\end{aligned}$$

5. CONVERGENCE QUANTITATIVE DE L'ENTROPIE

A partir des inégalités de la section précédente, l'étude de l'évolution temporelle de l'entropie est aussi relativement directe, si on reprend une estimée générale donnée dans [11].

En effet, on y a vu que

$$\mathrm{Ent}(mP|\mu) \leq \mathrm{Ent}(m|\mu) - \mathcal{E}_{I-PP^*}(\sqrt{dm/d\mu}, \sqrt{dm/d\mu})$$

du moins si m est une probabilité telle que $\mathrm{Ent}(m|\mu) < +\infty$ (la preuve n'y est présentée que dans le cas où S est fini, mais elle s'étend immédiatement au cas général).

Mais supposons de plus qu'il existe un bon majorant φ, désormais fixé, tel la densité initiale $f = dm/d\mu$ satisfasse

$$K \stackrel{\text{déf.}}{=} 1 \vee \mu(\varphi(f)) < +\infty \tag{11}$$

Une application de l'inégalité de Jensen fait alors apparaître que pour tout $n \in \mathbb{N}$,

$$\mu(\varphi(f_n)) \leq K$$

où $f_n = P^{*n}f$ désigne la densité de mP^n par rapport à μ, c'est-à-dire que chacune des fonctions $\sqrt{f_n}$ appartient à $D(\mu, \varphi, K)$. Fixons une application ψ vérifiant les conditions du corollaire 5 avec cette constante $K \geq 1$.

On peut alors appliquer les estimées précédentes pour obtenir

LEMME 8. – *Supposons que (11) soit satisfait, on a alors pour tout $n \in \mathbb{N}$,*

$$\operatorname{Ent}(mP^{n+1}|\mu) \leq \operatorname{Ent}(mP^n|\mu) - A\lambda\psi(\operatorname{Ent}(mP^n|\mu))$$

où λ est le trou spectral de $I - PP^*$ dans $\mathbb{L}^2(\mu)$ et A une constante qui ne dépend que de φ, K et ψ.

Pour pouvoir exploiter ces inégalités aux différences, on va les comparer à l'équation différentielle qui leur correspond :

Soit $v : \mathbb{R}_+ \to \mathbb{R}_+$ la solution de

$$\begin{cases} v'_t &= -A\lambda\psi(v_t) \\ v_0 &= \operatorname{Ent}(m|\mu) \end{cases}$$

PROPOSITION 9. – *On a pour tout $n \in \mathbb{N}$,*

$$\operatorname{Ent}(mP^n|\mu) \leq v_n$$

Preuve. – Ceci découle aisément de la positivité et de la croissance de l'application ψ. □

Il suffit donc désormais d'évaluer v_t pour t grand, et pour cela, fixons un $x_0 > 0$ (il est généralement commode de choisir $x_0 = \operatorname{Ent}(m|\mu)$) et soit

$$\begin{aligned} F : \mathbb{R}_+^* &\to \mathbb{R} \\ x &\mapsto \int_{x_0}^{x} \frac{1}{\psi(t)}\, dt \end{aligned}$$

Puisque pour t petit, $0 < \psi(t) \ll t$, on a le comportement suivant pour F : $\lim_{x\to 0_+} F(t) = -\infty$, $\lim_{x\to +\infty} F(t) = +\infty$ et F est strictement croissante sur $\mathbb{R}_+^*$, c'est-à-dire que F effectue une bijection croissante entre $\mathbb{R}_+^*$ et $\mathbb{R}$. Mais l'intérêt de cette fonction provient surtout du fait que pour tout $t \geq 0$, $F(v_t) = F(v_0) - A\lambda t$, d'où

$$v_t = F^{-1}(F(\mathrm{Ent}(m|\mu)) - A\lambda t)$$

Notamment, on a bien que $\lim_{t\to+\infty} v_t = 0$.

Pour terminer cette courte section, illustrons ces calculs en considérant les situations présentées dans l'introduction. Ci-dessous λ désignera toujours le trou spectral de $I - PP^*$.

• Si la densité initiale $dm/d\mu$ appartient à l'un des $\mathbb{L}^{1+\eta}(\mu)$, avec $\eta > 0$, on a vu que l'on pouvait prendre

$$\psi(t) = -\frac{t}{\ln(e^{-1}t)}\mathbf{1}_{t\leq 1} + \mathbf{1}_{t>1}$$

ce qui nous fournit tout naturellement les fonctions

$$\forall\, x > 0, \qquad F(x) = \begin{cases} (1-\ln^2(e^{-1}x))/2 & ,\ \text{si } x \leq 1 \\ x-1 & ,\ \text{si } x \geq 1 \end{cases}$$

$$\forall\, x \in \mathbb{R}, \qquad F^{-1}(x) = \begin{cases} \exp(1-\sqrt{1-2x}) & ,\ \text{si } x \leq 0 \\ x+1 & ,\ \text{si } x \geq 0 \end{cases}$$

Supposons que $\mathrm{Ent}(m|\mu) \leq 1$ (en effet, sinon les estimées précédentes permettent de majorer $\mathrm{Ent}(mP^n|\mu)$ par un terme décroissant linéairement en n, du moins tant que ce terme reste supérieur à 1, mais ceci est certainement trop grossier pour rendre compte de la véritable évolution de l'entropie sur l'intervalle de temps ainsi déterminé. Pour être un peu plus précis il aurait fallu s'intéresser plus sérieusement dans la section 3 au comportement global que l'on peut permettre à ψ, i.e. pas seulement au voisinage de 0, voir aussi à ce sujet l'appendice 1, mais remarquons que le comportement de ψ en l'infini n'est pas important, car $\mathcal{L}(\cdot)$ est bornée sur $D(\mu, \varphi, K)$. Il serait aussi possible de changer la définition de ψ en prenant par exemple $\psi(t) = -\frac{t}{\ln(e^{-1}a^{-1}t)}\mathbf{1}_{t\leq a} + a\mathbf{1}_{t>1}$, avec $a \geq 1 \vee \mathrm{Ent}(m|\mu)$, cependant la constante A dépendra alors de l'entropie initiale par l'intermédiaire de son majorant a. On obtient alors

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(mP^n|\mu) \leq \exp(1-\sqrt{\ln^2(e^{-1}\mathrm{Ent}(m|\mu)) + A\lambda n})$$

où $A > 0$ est une constante dépendant de $\int (dm/d\mu)^{1+\eta}\, d\mu$ et de η.

• Au bon majorant $\mathbb{R}_+ \ni x \mapsto x|\ln(x)|^{1+\eta}$, où $\eta > 0$, on peut associer les applications

$$\begin{aligned}
&\forall\, x \geq 0, && \psi(x) = 1 \wedge x^{(1+\eta)/\eta} \\
&\forall\, x > 0, && F(x) = \begin{cases} \eta(1 - x^{-1/\eta}) & , \text{si } x \leq 1 \\ x - 1 & , \text{si } x \geq 1 \end{cases} \\
&\forall\, x \in \mathbb{R}, && F^{-1}(x) = \begin{cases} (1 - x/\eta)^{-\eta} & , \text{si } x \leq 0 \\ x + 1 & , \text{si } x \geq 0 \end{cases}
\end{aligned}$$

ce qui amène les inégalités

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(mP^n|\mu) \leq (1 + A\lambda n)^{-\eta}$$

si $\mathrm{Ent}(m|\mu) = 1$, où $A > 0$ est une constante dépendant de $\int |\ln(dm/d\mu)|^{1+\eta}\, dm$ et de η.

• On a aussi déjà vu qu'au bon majorant $\mathbb{R}_+ \ni x \mapsto x\ln(x)\ln(|\ln(x)|)$, on pouvait associer l'application

$$\forall\, x \geq 0, \qquad \psi(x) = \exp\left(-\exp\left(\frac{8K}{x}\right)\right)\exp\left(-\frac{8K}{x}\right)$$

où $K = 1 \vee \int \ln(dm/d\mu)\ln[|\ln(dm/d\mu)|]\, dm$, cependant on n'obtient pas une formule « fermée » pour F. Supposons comme précédemment que $\mathrm{Ent}(m|\mu) = 1$ et prenons donc $x_0 = 1$. Même s'il apparaît facilement que pour x petit

$$-F(x) \sim \frac{x^2}{8K}\exp\left(\exp\left(\frac{8K}{x}\right)\right)$$

on préfère utiliser l'inégalité

$$\forall\, 0 < x \leq 1, \qquad F(x) \geq -\frac{1}{8K}(\exp(\exp(8K/x)) - \exp(\exp(8K)))$$

qui montre qu'il existe une constante $A > 0$ ne dépendant que de K telle que

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(mP^n|\mu) \leq \frac{8K}{\ln(\ln(\exp(\exp(8K)) + A\lambda n))}$$

Un des buts de la section suivante est d'expliciter un peu plus la dépendance en K de A.

6. APPLICATION À DES CHAÎNES DE MARKOV INHOMOGÈNES

Une des motivations des considérations précédentes provient de problèmes d'ergodicité forte (en loi) pour certaines chaînes de Markov qui sont inhomogènes par l'intermédiaire d'une température évanescente avec le temps (algorithmes de recuit simulé en temps discret sur des espaces généraux). Pour obtenir de tels résultats de convergence, on procède classiquement en deux étapes : tout d'abord on montre que la loi de la position du processus à un temps donné s'approche asymptotiquement de la probabilité invariante instantanée associée, puis on s'intéresse au comportement de ces dernières. Pour mesurer l'écartement entre des lois, il est souvent commode de considérer leur entropie relative et les calculs qui précèdent sont alors utiles pour étudier l'évolution de ces quantités. L'intérêt de l'entropie par rapport à des normes $\mathbb{L}^p$, avec $p > 1$, est qu'elle se comporte plus agréablement quand les lois invariantes instantanées varient avec le temps. En effet, en reprenant les calculs présentés dans [2] pour lesquels la finitude de l'espace des phases n'est pas nécessaire, on se rend compte qu'une des hypothèses minimales pour une bonne étude de l'évolution des normes $\mathbb{L}^2$, est que les lois invariantes instantanées, $(\mu_n)_{n\geq 0}$, doivent être équivalentes entre elles et que les densités $d\mu_n/d\mu_{n+1}$ appartiennent respectivement aux espaces $\mathbb{L}^\infty(\mu_n)$ (qui sont alors tous les mêmes), pour $n \in \mathbb{N}$. Mais en pratique ce dernier point est trop exigeant, et on en donnera un exemple ci-dessous.

Cependant pour les chaînes inhomogènes, la constante de majoration K aura tendance à croître avec le temps (il faudra d'ailleurs s'arranger pour qu'elle ne diverge pas trop rapidement, en jouant avec le contrôle que l'on peut avoir sur l'évolution de la température). Notre première tâche consiste donc à obtenir un résultat de dépendance assez explicite de A et ψ en fonction de K dans la proposition 6.

Plus précisément, soit φ un bon majorant fixé et $K_0 > \varphi(1)$. Soit également une fonction $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ telle que pour tout espace probabilisé $(S, \mathcal{S}, \mu)$ et pour tout $f \in D(\mu, \varphi, K_0)$, on ait

$$\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f) \geq 2\psi(\mathcal{L}^{(\mu)}(f)) \tag{12}$$

(on inclut la constante $A > 0$ qui intervient dans la proposition 6 dans la définition de ψ).

Alors, on a

PROPOSITION 10. – *Soit $K \geq K_0$, on a pour tout $f \in D(\mu, \varphi, K)$,*

$$\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f) \geq \frac{K-\varphi(1)}{K_0-\varphi(1)}\psi\left(\frac{K_0-\varphi(1)}{K-\varphi(1)}\mathcal{L}^{(\mu)}(f)\right)$$

Preuve. – On va se servir du fait que A et ψ ne dépendent pas de l'espace probabilisé considéré : soit $\triangle$ un point n'appartenant pas à S et intéressons-nous à $(\bar{S}, \bar{\mathcal{S}}, \bar{\mu})$, où on a posé

$$\begin{aligned}
\bar{S} &= S \sqcup \{\triangle\}\\
\bar{\mathcal{S}} &= \sigma(\mathcal{S} \sqcup \{\{\triangle\}\})\\
\bar{\mu} &= (1-\alpha)\delta_\triangle + \alpha \mathbf{1}_S \mu\\
\alpha &= \frac{K_0-\varphi(1)}{K-\varphi(1)} \in]0,1]
\end{aligned}$$

Soit ensuite la fonction $\bar{f}$ définie par

$$\forall\ x \in \bar{S}, \qquad \bar{f}(x) = \begin{cases} f(x) & \text{, si } x \in S\\ 1 & \text{, si } x = \triangle \end{cases}$$

(où rappelons que $f \in D(\mu, \varphi, K)$).

On a

$$\bar{\mu}(\bar{f}^2) = (1-\alpha) + \alpha\mu(f^2) \ = \ 1$$

et

$$\begin{aligned}
\bar{\mu}(\varphi(\bar{f}^2)) &= (1-\alpha)\varphi(1) + \alpha\mu(\varphi(f^2))\\
&\leq (1-\alpha)\varphi(1) + \alpha K \ = \ K_0
\end{aligned}$$

de sorte que $\bar{f} \in D(\bar{\mu}, \varphi, K_0)$.

Ainsi on a bien d'après (12),

$$\mathcal{E}^{(\bar{\mu})}_{I-E_{\bar{\mu}}}(\bar{f},\bar{f}) \geq \psi(\mathcal{L}^{(\bar{\mu})}(\bar{f}))$$

et il reste donc à comparer, d'une part $\mathcal{L}^{(\bar{\mu})}(\bar{f})$ et $\mathcal{L}^{(\mu)}(f)$, et d'autre part les variances.

Mais remarquons que

$$\begin{aligned}
\mathcal{L}^{(\bar{\mu})}(\bar{f}) &= \alpha\mathcal{L}^{(\mu)}(f)\\
\mathcal{E}^{(\bar{\mu})}_{I-E_{\bar{\mu}}}(\bar{f},\bar{f}) &= \alpha(1-\alpha)(1-\mu(f))^2 + \alpha\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)\\
&\leq 2\alpha\mathcal{E}^{(\mu)}_{I-E_\mu}(f,f)
\end{aligned}$$

d'où en fin de compte le résultat annoncé. □

Pour voir comment on peut mettre en pratique ces estimées, on va d'abord effectuer un calcul relativement générique, dont l'on précisera ensuite les hypothèses sur un exemple plus spécifique au recuit simulé.

Soit $(P_n)_{n\geq 0}$ une suite de noyaux markoviens sur $(S, \mathcal{S})$ et supposons-nous donné une suite $(\mu_n)_{n\geq 0}$ de probabilités invariantes respectives. On fait l'hypothèse que toutes ces mesures μ_n, $n \in \mathbb{N}$, sont équivalentes entre elles et on note pour tout $n \in \mathbb{N}$, λ_n le trou spectral de $I - P_n P_n^*$ dans $\mathbb{L}^2(\mu_n)$.

Soit m_0 une probabilité absolument continue par rapport à μ_0, on désignera alors par m_n la loi à l'instant $n \geq 0$ d'une chaîne de Markov inhomogène de loi initiale m_0 et dont la probabilité de transition est P_m pour tout temps $m \geq 0$. Remarquons tout d'abord que si pour un $n \in \mathbb{N}$, $m_n \ll \mu_0$, alors $m_n \ll \mu_n$ ce qui implique que $m_{n+1} = m_n P_n \ll \mu_n P_n = \mu_n$, d'où $m_{n+1} \ll \mu_0$. Ainsi par récurrence on obtient que pour tout $n \in \mathbb{N}$, $m_n \ll \mu_n$.

Soit φ un bon majorant, on fait l'hypothèse que

$$K_0 \stackrel{\text{déf.}}{=} \varphi(1) + \mu_0\left(\varphi\left(\frac{dm_0}{d\mu_0}\right)\right) < +\infty$$

et considérons ψ une fonction positive et croissante telle que (12) soit satisfait pour toute probabilité μ et toute fonction $f \in D(\mu, \varphi, K_0)$. On note $\widetilde{\psi}$ l'application définie par

$$\forall\, x \geq 0, \qquad \widetilde{\psi}(x) = x - \sup_{0\leq y\leq x} y - \psi(y)$$

de manière à ce que $\mathbb{R}_+ \ni x \mapsto x - \widetilde{\psi}(x)$ soit croissante et que $\widetilde{\psi} \leq \psi$, ainsi notamment (12) est aussi satisfait avec cette fonction $\widetilde{\psi}$.

Enfin supposons que l'on sache calculer des suites $(\gamma_n)_{n\geq 0}$ et $(K_n)_{n\geq 0}$ qui majorent respectivement les quantités

$$\forall\, n \in \mathbb{N}, \qquad \int \ln\left(\frac{\mu_n}{\mu_{n+1}}\right) dm_{n+1} \leq \gamma_n$$

$$\forall\, n \in \mathbb{N}, \qquad \int \varphi\left(\frac{m_n}{\mu_n}\right) d\mu_n \leq K_n$$

et posons pour tout $n \in \mathbb{N}$,

$$\Gamma_n = \frac{K_n - \varphi(1)}{K_0 - \varphi(1)}$$

PROPOSITION 11. – *Faisons l'hypothèse supplémentaire que la suite* $(J_n)_{n\geq 0}$ *définie par la récurrence*

$$\begin{aligned} & J_0 = \mathrm{Ent}(m_0|\mu_0) \\ \forall\, n \in I\!N, \qquad & J_{n+1} = J_n - \lambda_n \widetilde{\psi}(J_n) + \gamma_n \Gamma_n^{-1} \end{aligned}$$

satisfasse

$$\lim_{n\to\infty} \Gamma_n J_n = 0$$

Alors on a

$$\lim_{n\to\infty} \mathrm{Ent}(m_n|\mu_n) = 0$$

Preuve. – Les hypothèses précédentes nous fournissent effectivement tous les éléments pour l'étude traditionnelle de l'évolution de l'entropie. Pour $n \in I\!N$, notons $I_n = \mathrm{Ent}(m_n|\mu_n)$. On a, d'après la proposition 10,

$$\begin{aligned} I_{n+1} &= \mathrm{Ent}(m_{n+1}|\mu_n) + \int \ln\left(\frac{\mu_n}{\mu_{n+1}}\right) dm_{n+1} \\ &\leq \mathrm{Ent}(m_n|\mu_n) - \lambda_n \Gamma_n \psi\big(\Gamma_n^{-1}\mathrm{Ent}(m_n|\mu_n)\big) + \gamma_n \\ &\leq I_n - \lambda_n \Gamma_n \widetilde{\psi}(\Gamma_n^{-1} I_n) + \gamma_n \end{aligned}$$

ce qui nous amène à poser pour tout $n \in I\!N$, $\widetilde{J}_n = I_n/\Gamma_n$, car on a alors

$$\begin{aligned} \widetilde{J}_{n+1} &\leq \Gamma_n^{-1} I_{n+1} \\ &\leq \widetilde{J}_n - \lambda_n \widetilde{\psi}(\widetilde{J}_n) + \gamma_n \Gamma_n^{-1} \end{aligned}$$

et vu la croissance de l'application $I\!R_+ \ni x \mapsto x - \lambda\widetilde{\psi}(x)$ pour tout $0 \leq \lambda \leq 1$, on en déduit que pour tout $n \in I\!N$, $\widetilde{J}_n \leq J_n$, d'où le résultat annoncé. □

La proposition précédente donne donc un critère assurant le rapprochement asymptotique de m_n et de μ_n, mais en suivant la même démarche, on peut en imaginer de nombreux autres, en voici une variante qui se sert de la croissance de ψ plutôt que de $\widetilde{\psi}$.

PROPOSITION 12. – *On reprend le contexte décrit avant la proposition 11, mais supposons que* $\lim_{n\to\infty} \gamma_n = 0$ *et que pour tout* $b > 0$,

$$\sum_{n=0}^{\infty} \gamma_n - \lambda_n \Gamma_n \psi\big(\Gamma_n^{-1} b\big) = -\infty$$

On a alors la convergence

$$\lim_{n\to\infty} \mathrm{Ent}(m_n|\mu_n) = 0$$

Preuve. – Soit $b > 0$ fixé, et notons pour $n \in \mathbb{N}$, $c(n,b) = \gamma_n - \lambda_n\Gamma_n\psi(\Gamma_n^{-1}b)$. Par croissance de ψ et d'après la preuve de la proposition 11, tant que $I_n \geq b$, on a $I_{n+1} \leq I_n + c(n,b)$, et par hypothèse il existe donc un $p > 0$ tel que $I_{n+p} < b$. Maintenant si $I_n < b$ et si $p = \inf\{m > 0 \,/\, I_{n+m} \geq b\}$, alors $I_{n+p} \leq b + \gamma_{n+p-1}$. Il apparaît ainsi que

$$\limsup_{n\to\infty} I_n \leq \limsup_{n\to\infty} \left(b + \sup_{p\geq n} \gamma_p \right) = b$$

et on en déduit le résultat annoncé car b peut être choisi arbitrairement petit. □

Exemples. – a) Voici une application certainement trop directe de ce qui précède : notons que l'on peut toujours considérer

$$\gamma_n = \operatorname*{ess\,sup}_{\mu_0} \frac{d\mu_n}{d\mu_{n+1}} - 1 \in \overline{\mathbb{R}}_+$$

De plus, si l'on suppose que le bon majorant φ vérifie $\varphi(1) \geq 0$ et qu'il existe une certaine constante $a > 0$ telle que l'on soit assuré de

$$(13) \qquad \forall\, f \geq 0, \forall\, x > 0, \qquad \varphi(xf)/x \leq \varphi(f)\exp(a(x-1)_+)$$

il apparaît que l'on peut prendre pour tout $n \in \mathbb{N}$,

$$(14) \qquad K_n = \exp\left(a\sum_{k=0}^{n-1}\gamma_k\right)\mu_0\left(\varphi\left(\frac{dm_0}{d\mu_0}\right)\right) + \varphi(1)$$

et donc

$$\Gamma_n = \exp\left(a\sum_{k=0}^{n-1}\gamma_k\right)$$

En effet, puisque φ est convexe et que μ_n est invariante pour P_n, on calcule que

$$\begin{aligned}
\int \varphi\left(\frac{dm_{n+1}}{d\mu_{n+1}}\right) d\mu_{n+1} &= \int \varphi\left(\frac{dm_{n+1}}{d\mu_n}\frac{d\mu_n}{d\mu_{n+1}}\right)\frac{d\mu_{n+1}}{d\mu_n}\, d\mu_n \\
&\leq \exp(a\gamma_n)\int \varphi\left(\frac{dm_{n+1}}{d\mu_n}\right) d\mu_n \\
&\leq \exp(a\gamma_n)\int \varphi\left(\frac{dm_n}{d\mu_n}\right) d\mu_n
\end{aligned}$$

Les propositions 11 et 12 permettent alors d'obtenir des hypothèses sur les suites $(\lambda_n)_{n\geq 0}$ et $(\gamma_n)_{n\geq 0}$ qui assurent la convergence

$$\lim_{n\to\infty} \mathrm{Ent}(m_n|\mu_n) = 0$$

par exemple, il est facile de voir que les conditions

$$\begin{cases} \displaystyle\sum_{n\geq 0} \lambda_n \left(1 + \sum_{k=0}^{n-1} \gamma_k\right)^{-1} & = + \infty \\ \displaystyle\lim_{n\to\infty} \gamma_n \left(\sum_{k=0}^{n-1} \gamma_k\right) \lambda_n^{-1} & = 0 \end{cases}$$

sont suffisantes pour ceci, s'il existe un $r > 0$ tel que $\int \left|\frac{dm_0}{d\mu_0}\right|^{1+r} d\mu_0 < +\infty$.

Cependant, une des hypothèses que l'on est amené à faire, $\gamma_n < +\infty$ pour tout $n \in \mathbb{N}$, est parfois trop forte, car à partir de celle-ci on peut appliquer directement des arguments $\mathbb{L}^2$, pour obtenir, du moins si l'on suppose que $\int \left|\frac{dm_0}{d\mu_0}\right|^2 d\mu_0 < +\infty$, des critères un peu meilleurs de convergence vers 0 pour l'expression

$$\int \left|\frac{dm_n}{d\mu_n} - 1\right|^2 d\mu_n$$

Mais à nouveau, contrairement à une approche par l'entropie, ceci ne permet pas d'aborder les situations où l'on suppose seulement que $dm_0/d\mu_0 \in \mathbb{L}^{1+r}(\mu_0)$, avec $0 < r < 1$.

Cependant, les estimations (14) sont trop grossières et trop générales, dès que φ croît moins vite que toute fonction puissance d'exposant strictement plus grand que 1. On peut alors souvent les améliorer en étudiant plus précisément la chaîne de Markov particulière dont l'on dispose, comme va l'illustrer l'exemple suivant.

b) Pour présenter une situation difficilement traitable uniquement par des arguments $\mathbb{L}^2$, même si la densité initiale appartient à $\mathbb{L}^2(\mu_0)$, nous allons maintenant reprendre l'exemple introduit à la fin de la section 4. Donnons-nous de plus une évolution croissante de l'inverse de la température, $\mathbb{N} \ni n \mapsto \beta_n \in \mathbb{R}_+^*$, telle que $\lim_{n\to\infty} \beta_n = +\infty$, et posons pour tout $n \in \mathbb{N}$,

$$\begin{aligned} P_n &= \widetilde{P}_{\beta_n} \\ \mu_n &= \widetilde{\mu}_{\beta_n} \end{aligned}$$

ce qui nous replace dans le contexte précédent.

Une première étape pour obtenir de bonnes suites $(\gamma_n)_{n\in\mathbb{N}}$ et $(\Gamma_n)_{n\in\mathbb{N}}$ consiste à montrer le résultat technique suivant

LEMME 13. – *Soit $r > 0$ et supposons que $\int U^{1+r}(x)\,m_0(dx) < +\infty$. Il existe alors une constante $A > 0$ telle que pour tout $n \in \mathbb{N}$,*

$$\int U^{1+r}(x)\,m_n(dx) \leq A\exp(M\beta_{n-1})$$

où rappelons que $M = \sup_{x,y\in\mathbb{R}^d} |V(y) - V(x)|$.

Preuve. – Du fait que V est borné, il suffit de voir qu'il existe une constante $A_1 > 0$ telle que pour tout $n \in \mathbb{N}$,

$$\int |x|^{2+2r}\,m_n(dx) \leq A_1\exp(M\beta_{n-1})$$

Pour ceci, remarquons que pour tout $x \in \mathbb{R}^d$, on a

$$\begin{aligned}
&P_n(x, |\cdot - e^{-1}x|^{2+2r})\\
&\quad= \int \exp(-\beta_n(V(y) - V(x))_+)R^{(\beta_n)}_{\beta_n^{-1}}(x, dy)|y - e^{-1}x|^{2+2r}\\
&\qquad+ A_{\beta_n}(x)(1 - e^{-1})^{2+2r}|x|^{2+2r}\\
&\quad\leq \int R^{(\beta_n)}_{\beta_n^{-1}}(x, dy)|y - e^{-1}x|^{2+2r} + (1 - \exp(-\beta_n M))(1 - e^{-1})^{2+2r}|x|^{2+2r}\\
&\quad\leq \frac{A_2}{\beta_n^{1+r}} + (1 - \exp(-\beta_n M))(1 - e^{-1})^{2+2r}|x|^{2+2r}
\end{aligned}$$

où $A_2 = (1 - e^{-1})^{1+r}\int_{\mathbb{R}^d} |y|^{2+2r}\exp(-|y|^2)\,dy/\pi^{\frac{d}{2}}$.

Or par convexité, on a

$$\begin{aligned}
P_n(x, |\cdot|^{2+2r}) &= P_n(x, |(1 - e^{-1})(\cdot - e^{-1}x)/(1 - e^{-1}) + e^{-1}x|^{2+2r})\\
&\leq (1 - e^{-1})^{1-2-2r}P_n(x, |\cdot - e^{-1}x|^{2+2r}) + e^{-1}|x|^{2+2r}
\end{aligned}$$

et en intégrant ceci par rapport à m_n, on fait apparaître que

$$\begin{aligned}
&\int |x|^{2+2r}(x)\,m_{n+1}(dx)\\
&\qquad\leq (1 - (1 - e^{-1})\exp(-\beta_n M))\int |x|^{2+2r}(x)\,m_n(dx) + \frac{A_3}{\beta_n^{1+r}}
\end{aligned}$$

avec $A_3 = (1 - e^{-1})^{1-2-2r}A_2$.

Pour $n \in I\!N$, on est donc amené à poser $a_n = \exp(-\beta_{n-1}M) \int |x|^{2+2r}(x)\, m_n(dx)$, car l'inégalité ci-dessus montre que

$$a_{n+1} \leq a_n + \exp(-M\beta_n)\left(\frac{A_3}{\beta_n^{1+r}} - (1-e^{-1})a_n\right)$$

Ainsi, si $a_n \geq A_3\beta_n^{-1-r}/(1-e^{-1})$, alors on a $a_{n+1} \leq a_n$, et d'autre part, $a_n < A_3\beta_n^{-1-r}/(1-e^{-1})$ implique que $a_{n+1} \leq A_3\beta_n^{-1-r}[(1-e^{-1})^{-1} + \exp(-M\beta_n)]$. Il en découle facilement que si la suite $(a_n)_{n\geq 0}$ ne finit pas par être décroissante, alors elle converge vers 0, et dans tous les cas on a donc

$$\limsup_{n\to\infty} a_n < +\infty$$

ce qui nous assure de la validité du lemme. □

On en déduit le résultat suivant :

LEMME 14. – *Soit $r > 0$ et supposons que $m_0 \ll \mu_0$ et que $\int \ln_+^{1+r}\left(\frac{dm_0}{d\mu_0}\right) dm_0 < +\infty$. Il existe alors une constante $B > 0$ telle que pour tout $n \in I\!N$,*

$$\int \ln_+^{1+r}\left(\frac{dm_n}{d\mu_n}\right) dm_n \leq B\exp(M\beta_n)$$

Preuve. – Commençons par vérifier que l'hypothèse du lemme 13 est satisfaite. Soit $K > 0$, et notons $\widetilde{U} = K \wedge U$. En utilisant, avec $f = \beta_0\widetilde{U}/2$ et $m = \widetilde{U}^r m_0/m_0(\widetilde{U}^r)$, la majoration générale

$$\int f\, dm \leq \mathrm{Ent}(m|\mu_0) + \ln\left(\int \exp(f)\, d\mu_0\right)$$

valable pour toute application mesurable et positive f et toute probabilité m, on fait apparaître que

$$\begin{aligned}
&\frac{\beta_0}{2}\int \widetilde{U}^{1+r}\, dm_0 \\
&\quad\leq \int \ln(\widetilde{U}^r)\widetilde{U}^r\, dm_0 + \int \ln_+\left(\frac{dm_0}{d\mu_0}\right)\widetilde{U}^r\, dm_0 - m_0(\widetilde{U}^r)\ln(m_0(\widetilde{U}^r)) \\
&\qquad + m_0(\widetilde{U}^r)\ln\left(\int \exp(\beta_0\widetilde{U}/2)\, d\mu_0\right) \\
&\quad\leq \int \ln(\widetilde{U}^r)\widetilde{U}^r\, dm_0 + \left(\int \ln_+^{1+r}\left(\frac{dm_0}{d\mu_0}\right) dm_0\right)^{1/(1+r)} \\
&\qquad \left(\int \widetilde{U}^{1+r}\, dm_0\right)^{r/(1+r)} \\
&\qquad + e^{-1} + m_0(U^r)\ln\left(\int \exp(\beta_0\widetilde{U}/2)\, d\mu_0\right)
\end{aligned}$$

Il reste alors à faire tendre K vers $+\infty$, en se servant du fait que $\int \exp(\beta_0 U/2)\, d\mu_0 < +\infty$ et $\int \ln_+^{1+r}(dm_0/d\mu_0)\, dm_0 < +\infty$, pour se rendre compte que l'on ne peut pas avoir

$$\int U^{1+r}(x)\, m_0(dx) = +\infty$$

sinon le membre de gauche dans les inégalités précédentes exploserait plus vite que le membre de droite.

Ainsi on sait déjà qu'il existe une constante $A > 0$ telle que pour tout $n \in I\!N$,

$$\int U^{1+r}\, dm_n \leq A \exp(M\beta_{n-1})$$

On en déduit que

$$\begin{aligned}
&\left(\int \ln_+^{1+r}\left(\frac{d\mu_n}{d\mu_{n+1}}\right) dm_{n+1}\right)^{1/(1+r)} \\
&= \left(\int \left((\beta_{n+1} - \beta_n)U + \ln \frac{Z_{\beta_{n+1}}}{Z_{\beta_n}}\right)_+^{1+r} dm_{n+1}\right)^{1/(1+r)} \\
&\leq \left(\int ((\beta_{n+1} - \beta_n)U)^{1+r}\, dm_{n+1}\right)^{1/(1+r)} \\
&= (\beta_{n+1} - \beta_n)\left(\int U^{1+r}\, dm_{n+1}\right)^{1/(1+r)} \\
&\leq A^{\frac{1}{1+r}} (\beta_{n+1} - \beta_n) \exp(M\beta_n/(1+r))
\end{aligned}$$

En utilisant cette estimée et le fait que l'application $I\!R_+ \ni x \mapsto x \ln_+^{1+r}(x)$ est convexe, on obtient

$$\begin{aligned}
&\left(\int \ln_+^{1+r}\left(\frac{dm_{n+1}}{d\mu_{n+1}}\right) dm_{n+1}\right)^{1/(1+r)} \\
&\leq \left(\int \left(\ln_+ \frac{dm_{n+1}}{d\mu_n} + \ln_+ \frac{d\mu_n}{d\mu_{n+1}}\right)^{1+r} m_{n+1}(dx)\right)^{1/(1+r)} \\
&\leq \left(\int \left(\ln_+ \frac{dm_{n+1}}{d\mu_n}\right)^{1+r} m_{n+1}(dx)\right)^{1/(1+r)} \\
&\quad + \left(\int \left(\ln_+ \frac{d\mu_n}{d\mu_{n+1}}\right)^{1+r} m_{n+1}(dx)\right)^{1/(1+r)}
\end{aligned}$$

$$\leq \left(\int \left(\ln_+ \frac{dm_n}{d\mu_n}\right)^{1+r} m_n(dx)\right)^{1/(1+r)}$$
$$+ A^{\frac{1}{1+r}} \exp(M\beta_n/(1+r))(\beta_{n+1} - \beta_n)$$
$$\leq \left(\int \left(\ln_+ \frac{dm_0}{d\mu_0}\right)^{1+r} m_0(dx)\right)^{1/(1+r)}$$
$$+ A^{\frac{1}{1+r}} \sum_{i=0}^{n} \exp(M\beta_{i+1}/(1+r)) - \exp(M\beta_i/(1+r))$$
$$\leq \left(\int \left(\ln_+ \frac{dm_0}{d\mu_0}\right)^{1+r} m_0(dx)\right)^{1/(1+r)} + A^{\frac{1}{1+r}} \exp(M\beta_{n+1}/(1+r))$$
$$\leq B_1 \exp(M\beta_{n+1}/(1+r))$$

pour une certaine constante $B_1 > 0$, indépendante de n, et le résultat annoncé en découle immédiatement, avec $B = B_1^{1+r}$. □

Soit $r > 0$ fixé, et considérons le bon majorant φ défini par

$$\forall\, x \geq 0, \qquad \varphi(x) = x \ln_+^{1+r}(x)$$

D'après ce qui précède, si $\int \varphi(dm_0/d\mu_0)\, d\mu_0 < +\infty$, il existe une constante $C > 0$ telle que l'on puisse prendre pour tout $n \in \mathbb{N}$,

$$\gamma_n = C \exp(\beta_n M/(1+r))(\beta_{n+1} - \beta_n)$$
$$\Gamma_n = C \exp(\beta_n M)$$

(pour la première égalité, on a utilisé la majoration $\int \ln(d\mu_n/d\mu_{n+1})$ $dm_{n+1} \leq (\int \ln_+^{1+r}(d\mu_n/d\mu_{n+1})\ dm_{n+1})^{1/(1+r)}$ et une inégalité de la preuve du lemme 14).

Par ailleurs, quitte à prendre C encore plus grand, on a vu qu'une application ψ qui convenait était donnée par

$$\forall\, x > 0, \qquad \psi(x) = C(1 \wedge x^{\frac{1+r}{r}})$$

ainsi le critère de la proposition 12 montre que l'on sera assuré de

$$\lim_{n\to\infty} \operatorname{Ent}(m_n|\mu_n) = 0 \tag{15}$$

dès que pour tout $b > 0$,

$$\sum_{n=0}^{\infty} \exp(\beta_n M/(1+r))(\beta_{n+1} - \beta_n) - (1-e^{-2})C^{\frac{r-1}{r}} b^{\frac{1+r}{r}}$$
$$\exp(-(3+1/r)M\beta_n) = -\infty$$

Supposons plus précisément que l'évolution de l'inverse de la température soit donnée par

$$\forall\, n \in \mathbb{N}, \qquad \beta_n = k^{-1}\ln(e+n)$$

où $k > 0$ est une constante, la condition précédente revient alors à imposer que $k > (3+1/r)M$.

Dans le cas où $r = 2$, on aurait pu aboutir à la même conclusion en adaptant plutôt la méthode présentée dans [10], basée sur les inégalités

$$\frac{1}{\delta}\int (f-\mu(f))^2\,d\mu + 4\delta\int f^2(\ln(f+e)+1)^2\,d\mu \geq \int f^2\ln(f)\,d\mu$$

valables pour toute fonction $f \in \mathbb{L}^2(\mu)$, toute probabilité μ et tout $0 < \delta \leq \delta_0$, où $\delta_0 > 0$ est une certaine constante.

Cependant, en optimisant en δ, on retrouve les inégalités énergie-entropie restreintes utilisées ci-dessus si $r = 2$ (voir aussi l'appendice), et celles-ci peuvent donc également être vues comme une généralisation de l'approche donnée dans [10].

Remarque. – Supposons que l'on sache montrer qu'il existe une constante $c \geq 0$ telle que

$$\lim_{\beta\to\infty}\beta^{-1}\ln(\lambda(I-\widetilde{P}_\beta^2)) = -c$$

et que l'on fasse l'hypothèse que pour tout $R > 0$,

$$\int U^{1+R}(x)\,m_0(dx) < +\infty \tag{16}$$

on peut alors en déduire que (15) sera réalisé, dès que $k > c$, si l'on suppose que

$$\int \ln_+^{1+r}\left(\frac{dm_0}{d\mu_0}\right)dm_0 < +\infty \tag{17}$$

En effet, la relation (16) et la preuve du lemme 13 montrent que pour tout $\epsilon > 0$, il existe une constante $A_\epsilon > 0$ telle que pour tout $n \in \mathbb{N}$,

$$\int U^{1+r}(x)\,m_n(dx) \leq A_\epsilon\exp(\epsilon\beta_{n-1})$$

ce qui permet de voir que sous l'hypothèse supplémentaire (17), il existe une constante $B_\epsilon > 0$ telle que pour tout $n \in \mathbb{N}$,

$$\int \ln_+^{1+r}\left(\frac{dm_n}{d\mu_n}\right)dm_n \leq B_\epsilon\exp(\epsilon\beta_n)$$

et on conclut comme précédemment.

A l'aide des propositions 11 et 12, on obtient en pratique des critères qui assurent que pour tout $l \in \mathbb{N}$ et toute probabilité m absolument continue par rapport à μ_l et telle que $dm/d\mu_l \in \mathbb{L}^\infty(\mu_l)$, on soit assuré de

$$\lim_{n\to\infty} \mathrm{Ent}(m(P_l \cdots P_{l+n-1})|\mu_{l+n}) = 0 \tag{18}$$

Notamment, les conditions présentées dans l'exemple (b) ci-dessus impliqueront cette convergence.

Si l'on est un peu moins exigeant sur la qualité du rapprochement entre m_n et μ_n, il est alors facile d'étendre ce résultat :

COROLLAIRE 15. – *Supposons que (18) soit satisfait, alors pour toute probabilité initiale $m_0 \ll \mu_0$, on a en variation totale,*

$$\lim_{n\to\infty} \|m_n - \mu_n\|_{\mathrm{vt}} = 0$$

De plus, appelons $X = (X_n)_{n\geq 0}$ le processus des coordonnées canoniques sur $(\Omega, \mathcal{F}) = (S^{\mathbb{N}}, \mathcal{S}^{\otimes \mathbb{N}})$, et à toute probabilité m_0 sur S, associons $\mathbb{P}_{m_0}$ la loi sur $(\Omega, \mathcal{F})$ qui fait de X une chaîne de Markov inhomogène de transitions données par $(P_n)_{n\geq 0}$ et de loi initiale m_0. Soit A un événement de la tribu de queue de X, alors soit μ_0-p.s. en x, $\mathbb{P}_{\delta_x}(A) = 1$, soit μ_0-p.s. en x, $\mathbb{P}_{\delta_x}(A) = 0$.

Preuve. – Soit une loi initiale $m_0 \ll \mu_0$ et $0 < \epsilon < 1$. Il est possible de décomposer m_0 en $(1 - n_0(S))f_0\mu_0 + n_0$, où n_0 est une mesure positive de masse totale inférieure à ϵ et où f_0 est une densité (par rapport à μ_0) bornée.

On en déduit une décomposition de m_n en

$$\begin{aligned} m_n &= m_0(P_0 \cdots P_{n-1}) \\ &= (1 - n_0(S))(f_0\mu_0)(P_0 \cdots P_{n-1}) + (n_0)(P_0 \cdots P_{n-1}) \end{aligned}$$

ce qui permet de voir que

$$\begin{aligned} \|m_n - \mu_n\|_{\mathrm{vt}} &\leq (1 - n_0(S))\|(f_0\mu_0)(P_0 \cdots P_{n-1}) - \mu_n\|_{\mathrm{vt}} \\ &\quad + \|(n_0)(P_0 \cdots P_{n-1}) - n_0(S)\mu_n\|_{\mathrm{vt}} \\ &\leq \|(f_0\mu_0)(P_0 \cdots P_{n-1}) - \mu_n\|_{\mathrm{vt}} + 2\epsilon \end{aligned}$$

Cependant d'après l'hypothèse (18) appliquée avec $l = 0$ et $m = f_0\mu_0$,

$$\lim_{n\to\infty} \|(f_0\mu_0)(P_0 \cdots P_{n-1}) - \mu_n\|_{\mathrm{vt}} = 0$$

de sorte que

$$\limsup_{n\to\infty} \|m_n - \mu_n\|_{\mathrm{vt}} \leq 2\epsilon$$

et $\epsilon > 0$ pouvant être choisi arbitrairement petit, le premier résultat annoncé en découle.

De la même manière, il apparaît que pour tout $l \in \mathbb{N}$ et toute probabilité $m \ll \mu_0$,

$$\lim_{n\to\infty} \|m(P_l \cdots P_{l+n-1}) - \mu_{l+n}\|_{\mathrm{vt}} = 0$$

En adaptant la preuve du théorème 2 de [18], on voit que ceci admet pour conséquence que pour tout $m_0 \ll \mu_0$, la tribu de queue de X est $\mathbb{P}_{m_0}$-triviale.

Soit A un événement de la tribu de queue, et supposons par exemple que $\mathbb{P}_{m_0}(A) = 1$. S'il existait une distribution $m_0' \ll \mu_0$ telle que $\mathbb{P}_{m_0'}(A) = 0$, on aurait en posant $m_0'' = (m_0 + m_0')/2$, $\mathbb{P}_{m_0''}(A) = 1/2$, ce qui est absurde. On montre ainsi que $\mathbb{P}_m(A)$ ne dépend pas de la probabilité $m \ll \mu_0$ choisie. Reste à écrire que pour une telle probabilité on a

$$\mathbb{P}_m(\cdot) = \int_S \frac{dm}{d\mu_0}(x) \mathbb{P}_{\delta_x}(\cdot)\, \mu_0(dx)$$

pour se convaincre de la seconde affirmation du corollaire. □

Il est clair que ces calculs pourraient conduire à des estimations quantitatives sur la vitesse de convergence en variation totale. Sous de bonnes conditions, il est possible d'étudier directement l'évolution de la variation totale, mais ceci fait a priori plutôt intervenir les constantes ergodiques de Dobrushin (mais notons qu'elles sont aussi inutilisables dans l'exemple (b) ci-dessus), ce qui ne correspond pas à nos hypothèses.

Enfin précisons que les calculs de cette section peuvent s'adapter en temps continu, pour des processus de Markov inhomogènes de générateurs $(I - P_t)_{t\geq 0}$, où $(P_t)_{t\geq 0}$ est une famille mesurable de noyaux de probabilités de transition admettant une famille mesurable $(\mu_t)_{t\geq 0}$ de probabilités invariantes.

APPENDICE : UNE DÉMARCHE ALTERNATIVE

Dans le cas où φ est la fonction

$$\mathbb{R}_+ \ni x \mapsto x \ln_+^{1+\eta}(x) \tag{19}$$

avec $\eta > 0$ fixé, la minoration présentée dans la proposition 6 peut être vue comme une simple conséquence d'une inégalité de Hölder usuelle. Il est peut-être possible de généraliser cette approche à des fonctions convexes φ plus générales (satisfaisant toujours $\varphi(x) \gg x\ln(x)$, pour x grand), toutefois le fait que les normes d'Orlicz autres que celles des espaces $\mathbb{L}^p$, $0 \leq p \leq \infty$, ne sont pas très explicites nous a fait préférer les preuves plus élémentaires de la section 3, qui permettent de se figurer quelle doit être la forme de ψ rien que par une étude sur un ensemble à nombre de points réduit (typiquement 5).

Soient $0 < r < 2$, $0 < \alpha \leq 1$ et considérons la fonction

$$\begin{aligned} F \,:\, &[0,e] \to \mathbb{R} \\ &x \mapsto a(x-\alpha) + 5(x-\alpha)^2 - x^2\ln(x^2) \end{aligned}$$

avec $a = 2\alpha(1 + 2\ln(\alpha))$.

On calcule que $F''(x) = 4(1 - \ln(x)) \geq 0$ pour $x \in [0,e]$, et par choix de a, on a $F'(\alpha) = 0$, ainsi puisque $F(\alpha) \geq 0$, on a $F \geq 0$ sur son domaine de définition. Notons aussi que $a \geq -4e^{-3/2}$, ce qui montre que pour $x \geq e$, $a(x-\alpha) + 5(x-\alpha)^2 \geq 0$.

Par ailleurs, pour $x \geq e$, on a $x \leq e(e-1)^{-1}(x-\alpha)$, de sorte que

$$\begin{aligned} x^2\ln(x^2) &\leq \left(\frac{e}{e-1}\right)^r (x-\alpha)^r x^{2-r}\ln(x^2) \\ &\leq 5(x-\alpha)^r x^{2-r}\ln(x^2) \end{aligned}$$

et donc pour tout $x \geq 0$, on a

$$x^2\ln(x^2) \leq a(x-\alpha) + 5(x-\alpha)^2 + 5|x-\alpha|^r x^{2-r}\ln_+(x^2)$$

Soit maintenant une probabilité μ et une fonction positive $f \in \mathbb{L}^2(\mu)$ telle que $\mu(f^2) = 1$. On pose $\alpha = \mu(f)$ puis on intègre l'inégalité précédente avec f à la place de x, pour obtenir

$$\mu(f^2\ln(f^2)) \leq 5\mu((f-\mu(f))^2) + 5\mu(|f-\mu(f)|^r f^{2-r}\ln_+(f^2))$$

Appliquons l'inégalité de Hölder à ce dernier terme pour le majorer par

$$5[\mu((f-\mu(f))^2)]^{r/2}[\mu(f^2\ln_+^{2/(2-r)}(f^2)]^{(2-r)/2}$$

ce qui nous permet d'obtenir finalement

$$\mu(f^2\ln(f^2)) \leq 5[\mu((f-\mu(f))^2)]^{r/2}(1 + [\mu(f^2\ln_+^{2/(2-r)}(f^2))]^{(2-r)/2})$$

Ainsi si avec la fonction φ donnée en (19), où $\eta = r/(2-r)$, on a $\mu(\varphi(f^2)) \leq K$, alors

$$\frac{\mathcal{E}_{I-E_\mu}(f,f)}{\psi(\mathcal{L}(f))} \geq 5^{-2/r}[K^{(2-r)/2}+1]^{-2/r}$$

avec $\psi : \mathbb{R}_+ \ni x \mapsto x^{2/r}$, résultat qui dans ce cas est évidemment plus précis que la proposition 6, mais remarquons que la proposition 10 permet de rattraper le bon comportement en K, quand celui-ci devient grand.

Notons que ces calculs se généralisent au cas où $r = 2$, car il apparaît alors que si $\mu(f^2) = 1$ et si $\operatorname{ess\,sup}_\mu f^2 \leq K$ (nécessairement $K \geq 1$), on est assuré de

$$\frac{\mathcal{E}_{I-E_\mu}(f,f)}{\mathcal{L}(f)} \geq \frac{1}{5(1+\ln(K))}$$

Mais dans le cadre inhomogène de la section précédente, travailler avec la norme $\|\cdot\|_{\mathbb{L}^\infty(\mu_0)}$ (qui par équivalence des probabilités μ_n, pour $n \geq 0$, est aussi la norme $\|\cdot\|_{\mathbb{L}^\infty(\mu_n)}$) n'est intéressant que si l'on dispose de bonnes estimées sur $\operatorname{ess\,sup}_{\mu_0} d\mu_n/d\mu_{n+1}$, ce qui ne correspond pas en général aux situations pertinentes pour l'utilisation de l'entropie.

BIBLIOGRAPHIE

[1] D. Bakry, L'hypercontractivité et son utilisation en théorie des semigroupes. In P. Bernard, editor, *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXII-1992*, Lecture Notes in Mathematics 1581. Springer-Verlag, 1994.

[2] J.-D. Deuschel and C. Mazza, L^2 convergence of time nonhomogeneous Markov processes: I. spectral estimates. *The Annals of Applied Probability*, **4**(4) 1994, p. 1012–1056.

[3] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains. Préprint, Octobre 1995.

[4] N. Dunford and J. Schwartz, Linear operators, part I : general theory. Interscience publishers, 1964.

[5] S. R. Foguel, The ergodic theory of Markov processes. Van Nostrand, 1969.

[6] H. Föllmer, Tail structure of Markov chains on infinite product spaces. *Zeitschrift für Wahrscheinlichkeitstheorie verwandte Gebiete*, **50**, 1979, p. 273–285.

[7] M. A. Kranosel'skii and Y. B. Ruticku, Convex functions and Orlicz spaces. P. Noordhoff, 1961.

[8] P. Mathieu, Hitting times and spectral gap inequalities. *Annales de l'Institut Henri Poincaré*, **33**(4), 1997, p. 437–465.

[9] P.A. Meyer, Probabilités et potentiel. Hermann, 1966.

[10] L. Miclo, Recuit simulé sur $\mathbb{R}^n$. Etude de l'évolution de l'énergie libre. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, **28**(2), 1992, p. 235–266.

[11] L. Miclo, Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In J. Azéma, M. Emery, and M. Yor, editors, *Séminaire de Probabilités XXXI*, Lecture Notes in Mathematics 1655, p. 136–167. Springer-Verlag, Berlin, 1997.

[12] J. NEVEU, Martingales à temps discret. Masson et cie, 1972.
[13] D. REVUZ, Markov chains, North-Holland, 1975.
[14] J. S. ROSENTHAL, Markov chain convergence: from finite to infinite. *Stochastic Processes and their Applications*, **62**, 1996, p. 55–72.
[15] L. SALOFF-COSTE, Lectures on finite Markov chains. In P. Bernard, editor, *Lectures on Probability Theory and Statistics. Ecole d'Eté de Probabilités de Saint-Flour XXVI-1996*, Lecture Notes in Mathematics 1665. Springer-Verlag, Berlin, 1997.
[16] D. W. STROOCK, Logarithmic Sobolev inequalities for Gibbs states. In G. Dell'Antonio and U. Mosco, editors, *Dirichlet Forms*, Lecture Notes in Mathematics 1563, p. 194–228. Springer-Verlag, 1993.
[17] H. TOTOKI, On a class of special flows. *Zeitschrift für Wahrscheinlichkeitstheorie verwandte Gebiete*, **15**, 1970, p. 157–167.
[18] T. UENO, Some limit theorems for temporally discrete Markov processes. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, **7**, 1957, p. 449–462.
[19] L. N. WASSERSTEIN, Markov Processes over denumerable products of spaces describing large systems of automata. *Problemy Peredači Informacii*, **5**(3), 1969, p. 64–72.

(Manuscrit reçu le 10 février 1997;
version révisée reçue le 20 janvier 1998.)