PUBLICATIONS MATHÉMATIQUES ET INFORMATIQUES DE RENNES

Y. DERRIENNIC

Entropie, théorèmes limite et marches aléatoires

Publications de l'Institut de recherche mathématiques de Rennes, 1985, fascicule 1
« Séminaire de probabilités », , p. 9-53

<http://www.numdam.org/item?id=PSMIR_1985___1_9_0>

© Département de mathématiques et informatique, université de Rennes, 1985, tous droits réservés.

L'accès aux archives de la série « Publications mathématiques et informatiques de Rennes » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

ENTROPIE, THEOREMES LIMITE ET MARCHES ALEATOIRES

Y. Derriennic

Résumé

Le comportement asymptotique d'une marche aléatoire $S_n = X_1, \ldots X_n$ définie sur un groupe localement compact est traduit, pour une part importante, par l'ordre de grandeur en n de l'information mutuelle $I(S_1, S_n)$ des deux variables aléatoires S_1 et S_n. Dans le cas discret

$$I(S_1, S_n) = H(\mu^n) - H(\mu^{n-1}),$$

$H(\mu^n)$ désignant l'entropie absolue, au sens de Shannon, de la $n^{ième}$ convolée de la loi commune μ des accroissements indépendants X_n. Dans le cas absolument continu une formule analogue est valable. Sont examinés en détail les liens entre $I(S_1, S_n)$ d'une part, la loi des grands nombres, le théorème limite central, les fonctions harmoniques d'autre part. La structure du groupe sous-jacent joue un rôle important ; en particulier la croissance et l'unimodularité.

Mots-clés : Entropie, information, marche aléatoire, fonction harmonique, frontière, loi des grands nombres, théorème limite central, croissance non exponentielle.

MR codification 60 G

ENTROPIE, THEOREMES LIMITE ET MARCHES ALEATOIRES

Y. DERRIENNIC
UNIVERSITE DE BRETAGNE OCCIDENTALE
BREST, FRANCE

La notion d'entropie d'une distribution de probabilité a pour origine la formule $H = -\sum p \log p$ de Boltzmann. Comme on le sait, elle est fondamentale dans de nombreux sujets. Le but poursuivi ici est d'exposer le rôle de cette notion dans l'étude des sommes de variables aléatoires indépendantes équidistribuées à valeurs réelles, vectorielles ou encore à valeurs dans un groupe localement compact.

Une suite de v.a. indépendantes X_n, équidistribuées sur un groupe localement compact G, engendre la marche aléatoire $S_1 = X_1$, $S_2 = X_1X_2$, ... $S_n = X_1 .... X_n$ (ou $S_n = X_1 +..... + X_n$ si G est abélien). Le comportement asymptotique de cette marche est lié aux propriétés de la distribution des X_n et à celles du groupe G. Pour une part importante ce comportement est traduit par l'ordre de grandeur, quand n devient grand, de l'entropie $H(S_n)$ de la v.a. S_n. En effet, dans de nombreuses situations, il apparait que $H(S_n)$ a un développement asymptotique de la forme $H(S_n) \sim hn + c \log n$ dont la signification probabiliste est remarquable. Le premier terme hn, dans lequel figure la constante h, appelée dans la suite entropie asymptotique de la marche aléatoire, est lié à la loi des grands nombres ; sa nullité caractérise les marches aléatoires ne présentant qu'un seul comportement asymptotique distinguable, ou encore n'admettant pas d'autres fonctions harmoniques que les constantes. Le second terme $c \log n$ est lié au théorème limite central. Découvrir les liens entre la distribution des X_n, la structure de G et l'ordre de grandeur de l'entropie $H(S_n)$ est donc le sujet traité ici. Ce sujet n'est pas clos. La détermination de toutes les situations où $H(S_n)$ a le développement asymptotique donné ci-dessus, est loin d'être achevée.

Nous nous sommes efforcés de tenir compte de tous les résultats en notre connaissance (certains travaux très récents mis à part [27] [35]

[36]), mais bon nombre d'entre eux sont exposés sans démonstration complète, afin de ne pas répéter inutilement d'autres publications aisément accessibles. Le cas des marches aléatoires sur les groupes discrets a fait l'objet d'une publication récente très riche, comportant une bibliographie étoffée ([25]). Nous avons essayé de traiter le cas général ou au moins le cas où la loi de la marche aléatoire a une densité et les résultats de ce travail paraîssent ici pour la première fois sous forme détaillée (un résumé, contenant quelques erreurs et imprécisions, est paru dans [10]).

Voici, pour conclure cette introduction, le plan de l'article :

I

PRELIMINAIRES. ENTROPIE ET CONVOLUTION

Un groupe localement compact à base dénombrable G et une mesure de probabilité μ définie sur la tribu borélienne de G sont donnés. Une suite (X_n) de v.a. indépendantes de loi μ engendre alors la suite des produits, suivant la loi de groupe :

$$S_n = X_1 \ldots\ldots X_n,$$

qui constitue la marche aléatoire droite définie par μ, issue de e, l'élément neutre de G. Si G est abélien, en particulier $\mathbb{R}^d$ ou $\mathbb{Z}^d$, la loi de groupe est notée, comme d'habitude, additivement et

$$S_n = X_1 + \ldots\ldots + X_n.$$

La distribution de probabilité de la v.a. S_n est la $n^{\text{ième}}$ puissance de convolution μ^n de μ :

$$\int_G f(s)\,d\mu^n(s) = \int_{G^n} f(x_1 \ldots\ldots x_n)\; d\mu(x_1)\ldots\ldots d\mu(x_n).$$

La loi de probabilité de la marche aléatoire (S_n) est la loi pour laquelle les X_n sont indépendantes, de même distribution μ. Il est parfois utile de fixer les idées en considérant que les X_n sont les projections canoniques sur G de l'espace $\Omega = G^{\mathbb{N}}$, muni de sa tribu borélienne et de la mesure produit infini $\mu^{\otimes \mathbb{N}}$.

Une mesure de Haar à gauche sur G, notée m, est fixée. Si μ a une densité par rapport à m, elle est notée φ. Alors μ^n a une densité, notée φ_n, donnée par :

$$\varphi_n(y) = \int_G \varphi(x)\; \varphi_{n-1}\,(x^{-1}y)\,dm(x)$$

Dans le cas où G est un groupe discret dénombrable, m est la mesure de dénombrement ; on peut alors identifier μ et φ. L'entropie de S_n est définie par :

$$H(S_n) = H(\mu^n) = -\sum_{x \in G} \mu^n(x)\, \log\, \mu^n(x)\ ;$$

c'est l'entropie absolue de la distribution μ^n.

Dans le cas absolument continu, μ ayant la densité φ, l'entropie

différentielle de S_n est définie par :

$$\tilde{H}(S_n) = \tilde{H}(\varphi_n) = -\int_G \varphi_n(x) \log \varphi_n(x)\, dm(x).$$

La première définition peut être considérée comme un cas particulier de la seconde.

Pour comprendre le sens de ces quantités, il faut aussi considérer l'information mutuelle des v.a. S_1 et S_n :

$$I(S_1, S_n) = H(\lambda(S_1, S_n) ; \lambda(S_1) \otimes \lambda(S_n))$$

qui est l'entropie relative de la loi du couple $\lambda(S_1, S_n)$ par rapport à la loi produit des marginales $\lambda(S_1) \otimes \lambda(S_n)$. Dans un appendice sont résumées les propriétés de l'entropie H et de l'information I. On se propose d'indiquer, dans cette première partie, quelques propriétés liant l'entropie, l'information et la convolution.

Considérons d'abord le cas discret. Si la série $-\sum \mu(x) \log \mu(x)$ converge, c'est à dire si $H(\mu)$ est finie, il en est de même pour μ^n et $H(\mu^n) \leqslant n\, H(\mu)$. En effet μ^n est image de la mesure produit $\mu \otimes \ldots\ldots \otimes \mu$ (n facteurs) dont l'entropie est évidemment $nH(\mu)$ (voir l'appendice). Le même argument montre que $H(\mu^n)$ est une suite sous-additive, c'est à dire $H(\mu^{n+k}) \leqslant H(\mu^n) + H(\mu^k)$. Ceci a été prouvé, avec un argument différent, par Avez [1], puis par Kaimanovich-Vershik [25]. Le rôle de l'information mutuelle $I(S_1, S_n)$ apparaît dans la proposition suivante.

Proposition : *Dans le cas discret, si* $H(\mu) < \infty$, *alors*

$$I(S_1, S_n) = H(\mu^n) - H(\mu^{n-1}).$$

Démonstration : Par définition

$$I(S_1, S_n) = \sum_{x,y \in G} \left(\log \frac{\mu(x)\, \mu^{n-1}(x^{-1}y)}{\mu(x)\, \mu^n(y)} \right) \mu(x)\, \mu^{n-1}(x^{-1}y).$$

Après simplification, par la convergence des séries :

$$I(S_1, S_n) = \sum_{x,y} (\log \mu^{n-1}(x^{-1}y))\, \mu(x)\, \mu^{n-1}(x^{-1}y) - \sum_y \mu^n(y) \log \mu^n(y).$$

$$= H(\mu^n) - H(\mu^{n-1}).$$

Si l'entropie $H(\mu)$ est infinie, la quantité $I(S_1, S_n)$ peut être soit infinie, soit finie ; les deux cas peuvent se présenter. Cependant la loi conjointe $\lambda(S_1, S_n)$ est toujours absolument continue par rap-

port à la loi produit $\lambda(S_1) \otimes \lambda(S_n)$.

Le cas absolument continu est plus compliqué. Mais si l'on suppose que toutes les intégrales $\int_G \varphi_n(x) \, |\log \varphi_n(x)| \, dm(x)$ convergent, la proposition précédente est encore valide.

Proposition : *Dans le cas absolument continu, si pour tout* n $\varphi_n \log \varphi_n$ *est intégrable, alors*

$$I(S_1, S_n) = \tilde{H}(\varphi_n) - \tilde{H}(\varphi_{n-1}).$$

Démonstration : La densité de la loi conjointe $\lambda(S_1, S_n)$ par rapport à $m \otimes m$ est $\varphi(x) \, \varphi_{n-1}(x^{-1}y)$; celle de la loi produit $\lambda(S_1) \otimes \lambda(S_n)$ est $\varphi(x) \, \varphi_n(y)$. Comme $\varphi_n(y) = 0$ si et seulement si $\varphi(x) \, \varphi_{n-1}(x^{-1}y) = 0$ m p.p. en x, la fonction $\varphi(x) \, \varphi_{n-1}(x^{-1}y) = 0$ $m \otimes m$ p.p. sur l'ensemble où $\varphi(x) \, \varphi_n(y) = 0$. D'après le théorème de Gelfand-Yaglom-Perez rappelé en appendice, on trouve, après simplification :

$$I(S_1, S_n) = \int \log \left(\frac{\varphi_{n-1}(x^{-1}y)}{\varphi_n(y)} \right) \varphi(x) \, \varphi_{n-1}(x^{-1}y) \, dm(x) \, dm(y).$$

Cette formule est valide sans hypothèse sur φ ; l'intégrale est définie au sens large, prenant une valeur positive, finie ou infinie. Sous l'hypothèse faite sur φ on trouve, comme précédemment, par l'invariance à gauche de m :

$$I(S_1, S_n) = \int \varphi_{n-1}(y) \log \varphi_{n-1}(y) \, dm(y) - \int \varphi_n(y) \log \varphi_n(y) \, dm(y)$$
$$= \tilde{H}(\varphi_n) - \tilde{H}(\varphi_{n-1}).$$

Dans la suite des conditions de moments sur φ, suffisantes pour que les $\varphi_n \log \varphi_n$ soient intégrables, seront données. Observons déjà que cela est réalisé pour φ bornée et portée par un ensemble de mesure m finie. Les deux propositions qui viennent d'être démontrées unifient en quelque sorte, les notions différentes d'entropie absolue et d'entropie relative. Le cas discret apparaît comme un cas particulier du cas absolument continu. C'est un des avantages de la considération de l'information mutuelle $I(S_1, S_n)$.

Dans le cas général, μ ayant une partie singulière avec m, la quantité $I(S_1, S_n)$ conserve un sens ; elle prend une valeur positive

finie ou infinie (voir l'appendice). C'est alors cette quantité dont il faut étudier la croissance avec n. Il est important de remarquer que le cas $I(S_1, S_n) = +\infty$ se scinde en deux. On peut avoir $I(S_1, S_n)=+\infty$ en raison de la divergence des intégrales ou des séries définissant l'information mutuelle, comme dans le cas absolument continu ou discret. Un exemple est facile à donner en prenant une mesure μ sur les entiers telle que, à l'infini, $\mu(n) \sim \frac{C}{n(\log n)^\gamma}$, $1 < \gamma < 2$. Mais si μ n'est pas étalée, c'est à dire si aucune de ses puissances μ^k n'a de partie absolument continue non nulle, si de plus μ n'a pas de partie discrète, il est possible que, pour tout n, les lois $\lambda(S_1, S_n)$ et $\lambda(S_1) \otimes \lambda(S_n)$ soient étrangères. Un exemple de ce phénomène est donné par la mesure μ sur $\mathbb{R}$, construite par Fourt [15]. Cette mesure est diffuse et portée par un ensemble A tel que $A \cap (A + x)$ a au plus un élément si $x \neq 0$. L'ensemble $B = \{(x,y) \in \mathbb{R}^2 ; y - x \in A\}$ est de mesure 1 pour la loi conjointe car :

$$\lambda(S_1, S_2)(B) = \int_A d\mu(x)\, \mu(A) = 1,$$

mais de mesure 0 pour le produit des marginales car :

$$\lambda(S_1) \otimes \lambda(S_2)(B) = \int_A d\mu(x)\, \mu^2(A+x)$$

$$= \int_A d\mu(x) \int_A d\mu(u)\, \mu(A+x-u) = 0$$

Il n'est pas surprenant que ce cas se révèle plus difficile que l'autre.

II

ENTROPIE ET THEOREME LIMITE CENTRAL

Le maximum de l'entropie d'une distribution de probabilité à support fini est réalisé par la distribution uniforme. Cela est bien connu depuis longtemps. Le travail de Shannon contient une observation du même type concernant les lois gaussiennes : parmi les densités de variance donnée, sur $\mathbb{R}$ ou $\mathbb{R}^d$, le maximum de l'entropie différentielle est réalisé par la densité gaussienne ([34], p.88). Ces deux propriétés "variationnelles" ont d'intéressantes connexions avec les théorèmes de convergence en loi. C'est le thème de cette seconde partie.

La démonstration la plus courante de ces deux résultats repose sur l'inégalité

$$H(f;g) = \int f(x) \log \frac{f(x)}{g(x)} \, dx \geqslant 0$$

toujours satisfaite par l'entropie relative d'une densité f par rapport à une densité g dominant f, dans laquelle l'inégalité n'a lieu que si $f = g$ (c'est un corollaire direct de l'inégalité $\log x < x-1$). En posant $g \equiv 1$ on trouve la première propriété. Plus généralement, pour une densité φ portée par un ensemble de mesure finie E, on a

$$\tilde{H}(\varphi) = - \int_E \varphi(x) \log \varphi(x) \, dm(x) \leqslant \log m(E),$$

mais $\tilde{H}(\varphi)$ peut valoir $-\infty$. En posant $g(x) = \psi_\sigma(x)$ où

$$\psi_\sigma(x) = \frac{1}{\sigma \sqrt{2\Pi}} e^{-x^2/2\sigma^2}$$

notation qui sera utilisée dans la suite, on trouve la seconde propriété, car pour f centrée et de variance σ^2 on a :

$$- \int f(x) \log \psi_\sigma(x) \, dx = \frac{1}{2} \log (2\Pi e \, \sigma^2) = \tilde{H}(\psi_\sigma).$$

Le centrage n'est pas une restriction : l'entropie différentielle est invariante par translation. Cela prouve aussi que toute densité bornée ayant une variance finie, sur $\mathbb{R}$ ou $\mathbb{R}^d$, a une entropie différentielle finie.

Il n'est peut-être pas inutile de rappeler l'argument original de Shannon. Le maximum de $- \int f(x) \log f(x) \, dx$ sous les conditions $\int f(x) \, dx = 1$ et $\int x^2 f(x) \, dx = \sigma^2$, ne peut être atteint que si

$$\frac{d}{df} \int (-f(x) \log f(x) + \alpha f(x) x^2 + \beta f(x)) \, dx = 0.$$

En dérivant sous le signe somme on trouve

$$\int (-1 - \log f(x) + \alpha x^2 + \beta) \, dx = 0.$$

En fixant les multiplicateurs de Lagrange α et β de façon à vérifier les conditions on trouve que $f(x) = \psi_\sigma(x)$ vérifie cette égalité. On a reproché à cette preuve de manquer de rigueur ([30] note 2), mais il n'est pas difficile de la rendre précise, en calculant la dérivée par

rapport à un paramètre bien choisi. Par exemple, si l'on pose $F(t) = \tilde{H}(t\,\Psi_\sigma + (1-t)f)$ pour $0 \leqslant t \leqslant 1$, on vérifie facilement $F'(0) = 0$ et $F''(t) < 0$, pour f permettant les dérivations sous le signe somme. Cela montre que F a un maximum strict en $t = 0$ et donc que $\tilde{H}(\Psi_\sigma) > \tilde{H}(f)$; on passe alors à des densités f quelconques par approximation.

Pour faire apparaître la relation avec les théorèmes limite, considérons d'abord le cas où G est un groupe compact et où μ a une densité φ par rapport à la mesure de Haar m normalisée par $m(G) = 1$. Par le premier principe variationnel $\tilde{H}(\varphi) \leqslant 0$, avec égalité seulement si $\varphi = 1$ m p.p. Calculons $\tilde{H}(\varphi_{n+1})$. Par convexité

$$-\varphi_{n+1}(y) \log \varphi_{n+1}(y) \geqslant \int_G (-\varphi_n (x^{-1}y) \log \varphi_n (x^{-1}y))\, \varphi(x)\, dm(x) ;$$

en intégrant en y on trouve

$$\tilde{H}(\varphi_{n+1}) \geqslant \int_G \tilde{H}(\varphi_n)\, \varphi_n(x)\, dm(x) = \tilde{H}(\varphi_n).$$

La suite $\tilde{H}(\varphi_n)$, qui est donc croissante et négative, a une limite. Si φ est continue, la suite φ_n est équicontinue ; par un argument de compacité on montre alors que cette limite est 0, autrement dit :

$$\lim_n \tilde{H}(\varphi_n) = \lim_n H(\mu^n ; m) = 0$$

ce qui implique

$$\lim_n \int_G |\varphi_n(x) - 1|\, dm(x) = 0$$

(d'après [31] p.20). Cet argument a été donné par Csiszar pour démontrer le théorème de Ito-Kawada selon lequel la suite μ^n converge faiblement vers m quand μ est strictement apériodique ([7]). Il montre que la tendance vers l'équirépartition n'est que la tendance vers le maximum de l'entropie.

Le travail de Csiszar s'inspirait d'un article de Linnik, dans lequel le théorème limite central était démontré à partir de l'étude de $\tilde{H}(\varphi_n)$, φ étant une densité sur $\mathbb{R}$, ([30]). Il ressort de l'article de Linnik que la convergence en loi de $\frac{1}{\sqrt{n}} S_n$ vers une loi gaussienne ne fait qu'exprimer la tendance vers le maximum de l'entropie, étant donné les variances. Commentant ce travail dans son livre, Renyi a écrit "ainsi le théorème limite central apparaît analogue au second principe de la

thermodynamique". ([32] p 554).

Pour comprendre la relation entre la croissance de $\tilde{H}(\varphi_n)$, φ étant une densité sur $\mathbb{R}$, ou de $H(\mu^n)$, μ étant une probabilité sur $\mathbb{Z}$, et le théorème limite central, commençons par quelques observations simples.

Proposition : *Si φ est une densité sur $\mathbb{R}$, bornée, de variance σ^2, de fonction caractéristique intégrable, alors*

$$\lim_n \tilde{H}(\varphi_n) - \frac{1}{2}\log n\sigma^2 = \frac{1}{2}\log(2\pi e)$$

i.e.

$$\lim_n (\tilde{H}(\varphi_n) - \tilde{H}(\Psi_{\sigma\sqrt{n}})) = 0$$

(Si φ est centrée, ce qui ne change rien au problème le résultat s'écrit encore $\lim_n H(\varphi_n ; \Psi_{\sigma\sqrt{n}}) = 0$*).*

Démonstration : Supposons φ centrée. Soit $\widehat{\varphi_n}(x) = \sigma\sqrt{n}\,\varphi_n(x\sigma\sqrt{n})$ qui est la densité de $\frac{1}{\sigma\sqrt{n}} S_n$. Sous l'hypothèse faite sur la fonction caractéristique le théorème limite local est valide :

$$\lim_n \widehat{\varphi_n}(x) = \Psi_1(x) \quad \text{uniformément sur } \mathbb{R} \quad [14]$$

Pour n assez grand, $\widehat{\varphi_n}(x) < 1$ donc on a

$$\tilde{H}(\widehat{\varphi_n}) \geq \int_{-L}^{+L} -\widehat{\varphi_n}(x)\log\widehat{\varphi_n}(x)\,dx, \quad \text{pour tout } L.$$

Cette intégrale tend avec n, vers $\int_{-L}^{L} -\Psi_1(x)\log\Psi_1(x)\,dx$.

Comme $\tilde{H}(\Psi_1) \geq \tilde{H}(\widehat{\varphi_n})$ et que L est arbitraire, on obtient :

$$\lim_n \tilde{H}(\widehat{\varphi_n}) = \tilde{H}(\Psi_1).$$

C'est le résultat cherché car $\tilde{H}(\widehat{\varphi_n}) = \tilde{H}(\varphi_n) - \frac{1}{2}\log n\sigma^2$.

Proposition : *Si μ est la loi de Bernoulli sur les entiers :*

$\mu(1) = p$, $\mu(0) = 1-p$ $(0 < p < 1)$, *la suite* $H(\mu^{n+1}) - H(\mu^n)$ *est décroissante*, $H(\mu^{n+1}) - H(\mu^n) \sim \frac{1}{2}$ *et* $H(\mu^n) \sim \frac{1}{2}\log n$.

Démonstration : En utilisant la formule de Kolmogorov (voir l'appendice) on a :

$$H(S_{n+1}, X_{n+1}) = H(X_{n+1}) + H(S_{n+1} \mid X_{n+1}) = H(X_{n+1}) + H(S_n)$$

et d'autre part :

$$H(S_{n+1}, X_{n+1}) = H(S_{n+1}) + H(X_{n+1} \mid S_{n+1}).$$

Soit $g(x) = -x \log x - (1-x) \log (1-x)$, pour $0 < x < 1$. Sachant $S_{n+1} = k$, la loi de $(X_1, \ldots, X_{n+1})$ est équirépartie sur les suites de 0 ou 1 de longueur n+1, comportant k fois 1. La loi conditionnelle de X_{n+1} est donc

$$P(X_{n+1}=1 \mid S_{n+1} = k) = \frac{k}{n+1}.$$

Alors $$H(X_{n+1} \mid S_{n+1}) = \sum_{k=0}^{n+1} \binom{n+1}{k} p^k (1-p)^{n+1-k} \; g(\frac{k}{n+1}),$$

ce qui donne

$$H(\mu^{n+1}) - H(\mu^n) = H(S_{n+1}) - H(S_n) = g(p) - B_{n+1} g(p)$$

où $B_{n+1}\, g$ est le polynôme de Bernstein de g de degré n+1. Le résultat annoncé résulte alors directement du théorème de de Moivre-Laplace et de la concavité de g.

Ces propositions suggèrent que, μ étant une probabilité de variance finie sur $\mathbb{R}$ ou $\mathbb{Z}$, le théorème limite central implique l'ordre de grandeur $\frac{1}{2} \log n$ pour $\tilde{H}(\varphi_n)$ ou $H(\mu^n)$. (L'hypothèse minimale sur μ sous laquelle cette estimation est valide reste à préciser, en particulier dans le cas discret). Dans l'espace de dimension d l'estimation est $\frac{d}{2} \log n$. La démarche de Linnik consiste à contrôler de façon directe l'ordre de grandeur de $\tilde{H}(\varphi_n)$ afin d'en déduire le théorème limite central. Le point essentiel de son argumentation est l'inégalité suivante.

Inégalité de Linnik Soit φ une densité sur $\mathbb{R}$, centrée, de variance σ^2 et de la forme $\varphi = \Psi_{\delta^2} * \gamma$ où γ est une densité à support compact. Alors

$$(\tilde{H}(\varphi_{n+1}) - \frac{1}{2} \log (n+1) \sigma^2) - (\tilde{H}(\varphi_n) - \frac{1}{2} \log n \sigma^2)$$

$$= \frac{1}{2n} \left(n\sigma^2 \int_{-\infty}^{+\infty} (\frac{\varphi_n'}{\varphi_n})^2(x) \, \varphi_n(x) \, dx - 1 \right) + O\left(\frac{1}{n} \left(\tau + \int_{|x| > \tau\sigma\sqrt{n}} x^2 \varphi(x) dx\right)\right)$$

(le dernier terme est 0 par rapport à $n > 1$ et $\tau > 0$).

Le premier membre s'écrit encore

$$H(\varphi_{n+1} \; ; \; \Psi_{(n+1)\sigma^2}) - H(\varphi_n \; ; \; \Psi_{n\sigma^2}).$$

L'expression $\int_{-\infty}^{+\infty} (\varphi'_n / \varphi_n)^2(x)\ \varphi_n(x)\ dx$ est l'information de Fisher de la densité φ_n. D'après l'inégalité de Cramer-Rao-Fréchet, elle est supérieure ou égale à $1/n\sigma^2$ avec égalité seulement si φ_n est gaussienne. La démonstration, assez difficile, de Linnik exploite à la fois les propriétés de H et celles de l'information de Fisher. Un article récent de Brown permet de la simplifier un peu, en particulier grâce à l'inégalité

$$\text{Inf. Fisher } (\varphi * \varphi) < 2 \text{ Inf. Fisher}(\varphi)$$

dans laquelle l'égalité n'est réalisée que si φ est gaussienne ([6]).

Le raisonnement que l'on vient de résumer semble assez peu commode. Cependant il conduit à plusieurs questions intéressantes. La propriété d'entropie maximum à variance donnée, caractérise-t-elle les lois gaussiennes sur un groupe de Lie autre que IR^d ? Le théorème limite central prend-il alors le sens de tendance vers le maximum de l'entropie ? La considération de l'entropie n'élimine pas la difficulté liée à la définition de la notion de variance, mais elle libère l'énoncé du théorème limite du problème de la "normalisation" : en effet le théorème peut se formuler directement sous la forme $\lim_n H(\varphi_n ; \Psi_n) = 0$ (le même avantage est présenté par la considération de la distance en variation au lieu de l'entropie relative).

Pour une loi stable d'indice α, $0 < \alpha < 2$, sur IR on vérifie immédiatement que

$$\tilde{H}(\varphi_n) \sim \frac{1}{\alpha} \log n$$

On peut aussi se demander si la convergence vers une loi stable autre que gaussienne correspond à un principe variationnel ? La notion d'entropie différentielle ne conduit pas simplement à un tel principe. En effet, pour que φ appartienne au domaine d'attraction "standard" d'une densité stable d'indice α, la condition nécessaire et suffisante ne porte que sur le comportement à l'infini de φ; il est donc clair que la densité stable ne maximise pas l'entropie différentielle dans son domaine d'attraction "standard".

III

ENTROPIE ASYMPTOTIQUE D'UNE MARCHE ALEATOIRE

Alors que le travail de Linnik est resté peu exploité, l'étude des fonctions harmoniques sur les groupes de Lie, qui s'est développée depuis vingt-cinq ans, a fait usage de la notion d'entropie suivant un point de vue différent, suggéré par la théorie de Kolmogorov-Sinai-Ornstein. Déjà dans le travail de Fürstenberg, la notion d'*entropie asymptotique* d'une marche aléatoire est considérée, quoique l'expression ne soit pas employée ([16]). Dans l'étude des fonctions harmoniques sur les groupes discrets l'idée a été employée avec un certain succès par Avez ([1] [2] [3]) puis par Kaimanovich et Vershik ([24] [25]). On se propose dans cette partie de définir l'entropie asymptotique et de donner ses premières propriétés, pour une marche aléatoire sur un groupe localement compact séparable G définie par une probabilité μ *quelconque*.

La loi de probabilité de la marche aléatoire issue de e, $S_n = X_1 \ldots X_n$ où les X_n sont des v.a. indépendantes de même distribution μ, est l'image de la mesure produit infini $\mu^{\otimes \mathbb{N}}$ par l'application :

$$(x_1, \ldots, x_n \ldots) \mapsto (s_1 = x_1, \ldots, s_n = (x_1 \ldots x_n), \ldots) ;$$

on la note P_e. Une v.a. Z fonction de la suite $(S_1, \ldots, S_n \ldots)$ est dite "asymptotique" si elle ne dépend que des coordonnées S_n pour $n \geqslant k$ et ceci pour tout k. Les événements asymptotiques pour (S_n) forment la tribu "asymptotique" ; on note $\mathcal{A} = \bigcap_k \sigma(S_n ; n \geqslant k)$ (ne pas confondre $\mathcal{A}$ avec la tribu asymptotique de la suite indépendante (X_n)). Une v.a. Y fonction de la suite $(S_1, \ldots S_n \ldots)$ est dite "invariante" si elle vérifie :

$$Y(S_1, \ldots, S_n, \ldots) = Y(S_2, \ldots, S_{n+1} \ldots)$$

autrement dit si elle est invariante sous l'opérateur décalage

$$\vartheta(S_1, \ldots, S_n, \ldots) = (S_2, \ldots, S_{n+1}, \ldots).$$

Les événements invariants forment la tribu "invariante" notée $\mathcal{I}$. L'inclusion $\mathcal{I} \subset \mathcal{A}$ est évidente (pour plus de détails voir l'appendice 2).

Définition : L'entropie asymptotique de la marche aléatoire définie par μ *, quelconque, est l'information mutuelle des tribus* $\sigma(S_1)$ *et* $\mathcal{A}$ *pour* P_e *. On note*

$$h(\mu) = I(S_1, \mathcal{A}).$$

Si le comportement asymptotique de la marche, c'est à dire la tribu $\mathcal{A}$ mod. P_e, est connu, on peut, en principe, en déduire $h(\mu)$. Réciproquement, s'il est possible de connaître, a priori, $h(\mu)$ on espère en déduire des informations sur la tribu $\mathcal{A}$. Le principal moyen pour atteindre $h(\mu)$ est l'approximation par $I(S_1, S_n)$.

Proposition : $I(S_1, S_n) = I(S_1, \sigma(S_k ; k \geqslant n)$

La suite $I(S_1, S_n)$ *décroît, au sens large. Elle peut être constante égale à* $+\infty$ *S'il existe* n *tel que* $I(S_1; S_n) < \infty$ *alors* $h(\mu) = \lim_n I(S_1, S_n)$.

Démonstration : La propriété de Markov signifie que, sachant S_n, S_1 et $(S_{n+1} \ldots\ldots)$ sont indépendantes. En vertu de la formule de Kolmogorov, cela entraîne la première égalité (voir [31] partie 3.4). La décroissance de $I(S_1, S_n)$ et l'égalité $h(\mu) = \lim_n I(S_1, S_n)$ dans le cas fini, résultent alors de la propriété 5 de l'information mutuelle, rappelée en appendice.

Les calculs effectués dans la partie I donnent l'énoncé suivant :

Proposition : Dans le cas discret, si $H(\mu) < \infty$ *, on a*

$$h(\mu) = \lim_n (H(\mu^n) - H(\mu^{n-1})) = \lim_n \frac{1}{n} H(\mu^n).$$

Dans le cas absolument continu, si $\tilde{H}(\varphi_n)$ *est finie pour tout* n*, on a*

$$h(\mu) = \lim_n (\tilde{H}(\varphi_n) - \tilde{H}(\varphi_{n-1})) = \lim_n \frac{1}{n} \tilde{H}(\varphi_n).$$

Si les quantités $I(S_1, S_n)$ sont toutes infinies, $h(\mu)$ est plus difficile à atteindre.

Proposition : En désignant par $\mathcal{C}$ *une partition finie mesurable, arbitraire, de* G *et* $S_1^{-1}(\mathcal{C})$ *la sous-tribu de* $\sigma(S_1)$*engendrée par* $\mathcal{C}$ *, on a, pour* μ *quelconque :*

$$h(\mu) = \sup_{\mathcal{C}} I(S_1^{-1}(\mathcal{C}), \mathcal{A}) = \sup_{\mathcal{C}} \lim_n I(S_1^{-1}(\mathcal{C}), S_n).$$

Démonstration : La première égalité répète la définition de l'information mutuelle $I(S_1, \mathcal{A})$. Pour la seconde, on observe que

$I(S_1^{-1}(\mathcal{C}), \mathcal{A}) = \lim_n I(S_1^{-1}(\mathcal{C}), S_n)$; ceci résulte, comme précédemment, de la propriété de Markov qui implique la monotonie de la suite qui est dans ce cas finie.

Dans le travail d'Avez la quantité $h(\mu)$ est définie, dans le cas discret, par l'égalité

$$h(\mu) = \lim_n \frac{1}{n} H(\mu^n) ;$$

dans le cas absolument continu, la convergence de $\frac{1}{n} \tilde{H}(\varphi_n)$ n'est pas prouvée et seule $\limsup_n \frac{1}{n} \tilde{H}(\varphi_n)$ est considérée. Avez appelle $h(\mu)$ l'entropie de la marche aléatoire. La définition, plus générale, donnée ici est suggérée par le travail de Kaimanovich et Vershik, bien que seul le cas discret y soit envisagé. Il semble préférable de nommer $h(\mu)$ "entropie asymptotique" de la marche aléatoire. En effet l'entropie absolue de la loi P_e est infinie, sauf dans le cas déterministe ; d'autre part le gain d'entropie par unité de temps est celui du schéma de Bernoulli de base μ et vaut donc $H(\mu)$. La quantité $h(\mu)$ ne représente que la part d'entropie de la marche restant dans le comportement asymptotique $\mathcal{A}$, quand on connait la position S_1. Si les quantités considérées sont finies on a :

$$h(\mu) = H_{P_e}(\mathcal{A}) - H_{P_e}(\mathcal{A}/S_1).$$

La définition de $h(\mu)$ fait intervenir $\mathcal{A}$, la tribu asymptotique. On peut se demander quel rôle joue exactement la tribu invariante $\mathcal{I}$.

Théorème *: Les tribus asymptotique $\mathcal{A}$ et invariante $\mathcal{I}$ de la marche aléatoire coïncident modulo la loi de la marche aléatoire P_e. L'entropie asymptotique vérifie* $h(\mu) = I(S_1, \mathcal{A}) = I(S_1, \mathcal{I})$.

La démonstration qui sort un peu du sujet traité, est repoussée en appendice. Rappelons seulement ici que, dans le cas où μ^n et μ^{n+1} ne sont pas étrangères pour au moins un n, l'égalité $\mathcal{A} = \mathcal{I}$ mod P_e résulte de la loi "zéro ou deux" et est bien connue ([8]). La motivation d'Avez était l'étude des fonctions μ-harmoniques en relation avec la structure du groupe.

Définition *: Une fonction réelle, mesurable g définie sur G est dite μ-harmonique à droite si, pour tout* $x \in G$

$$g(x) = \int_G g(xy)\, d\mu(y).$$

(dans la suite on omet la précision "à droite").

Le lien entre fonctions μ-harmoniques et entropie asymptotique apparaît tout d'abord dans l'énoncé général suivant

Théorème *: La tribu invariante* $\mathcal{I}$ *(ou la tribu asymptotique* $\mathcal{A}$*) est triviale* mod P_e *si et seulement si* $h(\mu) = 0$.

Si μ *est adaptée (i.e. si* G *est le plus petit sous-groupe fermé contenant le support de* μ, $\mathrm{Supp}(\mu)$*), si la tribu* $\mathcal{I}$ *est triviale* mod P_e *, les fonctions* μ*-harmoniques bornées, continues, sont constantes. Réciproquement, si* μ *est étalée, si les fonctions* μ*-harmoniques, bornées, continues sont constantes alors* $\mathcal{I}$ *est triviale* mod P_e.

La démonstration du premier point repose sur le lemme suivant.

Lemme *: Si* $\mathcal{D}$ *est une sous-tribu de* $\mathcal{A}$ *, stable sous l'opérateur décalage* ϑ *,* $\vartheta^{-1}\mathcal{D} = \mathcal{D}$ *, alors pour tout* k, $I(S_k, \mathcal{D}) = kI(S_1, \mathcal{D})$.

Démonstration : D'après la propriété de Markov et la formule de Kolmogorov :

$$I(S_k, \mathcal{D}) = I((S_1, \ldots, S_k), \mathcal{D}) = I((S_1 \ldots S_{k-1}), \mathcal{D}) + EI(S_k, \mathcal{D}/S_1 \ldots S_{k-1}).$$

Par stationnarité $EI(S_k, \mathcal{D}/S_1 \ldots S_{k-1}) = I(S_1, \mathcal{D})$, donc $I(S_k, \mathcal{D}) = kI(S_1, \mathcal{D})$.

Démonstration du théorème: Si $h(\mu) = I(S_1, \mathcal{I}) = I(S_1, \mathcal{A}) = 0$ alors d'après le lemme $I(S_k, \mathcal{A}) = I(S_k, \mathcal{I}) = 0$ pour tout k.

D'après la propriété 2 de l'information mutuelle (voir l'appendice), cela signifie que, pour tout k, les tribus $\sigma(S_1, \ldots, S_k)$ et $\mathcal{A}$ (ou $\mathcal{I}$) sont indépendantes pour P_e. La tribu $\mathcal{A}$ (ou $\mathcal{I}$) étant une sous-tribu de $\sigma(S_1, \ldots, S_k \ldots)$ cela prouve que $\mathcal{A}$ (ou $\mathcal{I}$) est triviale mod. P_e. La réciproque est évidente.

La démonstration du second point repose sur la correspondance bien connue entre v.a. invariantes et fonctions harmoniques d'une chaîne de Markov. Soit g une fonction μ-harmonique bornée continue. La suite $g(S_n)$ forme une martingale pour P_e, qui converge vers une v.a. invariante Z. Si $\mathcal{I}$ est triviale mod P_e, alors

$$Z = E_e(Z) = g(e) \quad P_e \text{ p.s.}$$

et $g(S_1) = E_e(Z/S_1) = g(e)$ P_e p.s.

Ceci donne $g(y) = g(e)$ μp.p. en y . La fonction g étant continue

on obtient $g(y) = g(e)$ pour tout $y \in \mathrm{Supp}(\mu)$. On trouve le même résultat pour la translatée à gauche $g_a(x) = g(ax)$ qui est aussi μ-harmonique. Donc le sous-groupe fermé des périodes à droite de g contient $\mathrm{Supp}(\mu)$. L'hypothèse d'adaptation implique que g est constante.

Réciproquement si μ est étalée, toutes les fonctions μ-harmoniques bornées sont continues donc ici constantes. Il en résulte que $\mathcal{J}$ est triviale mod P_e, et même mod. la loi de la marche aléatoire de distribution initiale arbitraire sur G ([8] p. 119).

La constance des fonctions μ-harmoniques sous l'hypothèse $h(\mu) = 0$, a été démontrée, dans le cas discret ou absolument continu avec $\tilde{H}(\varphi_n)$ fini par Avez, avec un argument direct, élégant ([2], [4]). La réciproque a été prouvée, dans le cas discret dans [9] et [24]. La démonstration générale donnée ici s'inspire de celle de [25].

Dans la réciproque, l'hypothèse d'étalement est indispensable. Sinon il est possible qu'existent des fonctions μ-harmoniques constantes m.p.p. mais non constantes relativement à la mesure $\sum_0^\infty \mu^n/2^{n+1}$. Alors la tribu $\mathcal{J}$ est triviale pour la loi de la marche aléatoire si la distribution initiale est absolument continue, mais n'est pas triviale pour P_e et on a $h(\mu) > 0$. Un exemple est le suivant : dans $SO(3, \mathbb{R})$ il existe un sous-groupe à deux générateurs a et b qui est libre ([18] p. 9) ; la moyenne des mesures de Dirac $\mu = \frac{1}{2}(\delta_a + \delta_b)$ est adaptée au sous-groupe *fermé* G engendré par a et b, qui est compact ; les fonctions μ-harmoniques continues sur G sont donc constantes, mais sur le sous-groupe libre engendré par a et b, il existe des fonctions μ-harmoniques non constantes, bornées, donc $h(\mu) > 0$. Sur un groupe abélien ce phénomène est impossible en vertu de la loi "zéro ou un" de Hewitt et Savage.

Examinons, pour conclure cette partie, le problème de la stabilité de l'entropie asymptotique $h(\mu)$: une suite de probabilités μ_j est donnée sur G, qui converge vers μ étroitement. Existe-t-il une relation, valide en général, entre $h(\mu)$ et $h(\mu_j)$? La réponse est négative et c'est une des difficultés du sujet. Si $\lim_j \mu_j = \mu$ alors $\lim_j \mu_j^n = \mu^n$ et $\lim_j \lambda_j(S_1, S_n) = \lambda(S_1, S_n)$ (loi conjointe de S_1 et S_n). Par un théorème de Dobrushin ([11], [31] p. 9) on a

$$I(S_1, S_n) = \sup_{\mathcal{E}} \sum_{E,F} \log\left(\frac{\int_E d\mu(x)\, \mu^{n-1}(xF)}{\mu(E)\, \mu^n(F)}\right) \int_E d\mu(x)\, \mu^{n-1}(xF),$$

où $\mathcal{E}$ désigne la collection des partitions finies de $G \times G$ formées d'ensembles $E \times F$ boréliens, de frontière négligeable pour $\lambda(S_1) \otimes \lambda(S_n) = \mu \otimes \mu^n$ et pour $\lambda(S_1, S_n)$. La même égalité a lieu pour μ_j ; en passant à la limite dans la somme finie on trouve :

$$\liminf_j \; I_j(S_1, S_n) \geqslant I(S_1, S_n) \qquad ([31] \text{ p13}).$$

Quand on passe à la limite en n, même si toutes les quantités considérées sont finies, cette inégalité n'a pas de raison d'être préservée. Kaimanovich et Vershik ont donné à ce sujet deux exemples remarquables. Sur $\mathfrak{S}_\infty$ le groupe des permutations d'ordre fini d'un ensemble infini dénombrable, il existe une probabilité symétrique μ pour laquelle $H(\mu) < \infty$ et $H(\mu) > 0$ alors que, tout élément de $\mathfrak{S}_\infty$ étant d'ordre fini, toute probabilité à support fini, a une entropie asymptotique nulle ([25] p. 484). D'autre part sur le groupe des configurations finies de $\mathbb{Z}^3$, $F_0(\mathbb{Z}^3, \mathbb{Z}/2\mathbb{Z})$ décrit dans la partie IV, il existe une probabilité μ telle que $h(\mu) = 0$, alors que pour toute probabilité à support fini, adaptée, l'entropie asymptotique est non nulle. Il n'y a donc aucun moyen général d'approximer, ni de majorer ou minorer, l'entropie asymptotique d'une probabilité μ par l'entropie asymptotique de "restrictions" finies de μ convergeant vers μ.

IV

CROISSANCE, LOI DES GRANDS NOMBRES ET ENTROPIE ASYMPTOTIQUE NULLE

Comme on vient de le voir, la nullité de $h(\mu)$ équivaut à la constance des fonctions μ-harmoniques continues bornées, au moins dans le cas étalé. Alors la marche aléatoire ne présente qu' ***un*** comportement limite distinguable avec probabilité positive. La partie IV est consacrée aux problèmes suivants :

- caractériser les groupes G sur lesquels il existe μ adaptée telle que $h(\mu) = 0$;
- caractériser les groupes G tels que pour toute probabilité μ, adaptée, $h(\mu) = 0$;
- dans le cas intermédiaire, décrire les probabilités adaptées telles que $h(\mu) = 0$.

Pour que les problèmes soient bien posés on suppose toujours, dans cette partie, que μ est adaptée (i.e. G est le plus petit sous-

groupe fermé portant μ).

Le premier problème est résolu par le théorème suivant.

Théorème : *Une condition nécessaire et suffisante pour qu'il existe sur* G *une probabilité* μ , *adaptée, telle que* $h(\mu) = 0$ *est que* G *soit moyennable.*

Il est connu depuis longtemps que l'existence sur G d'une probabilité μ pour laquelle les fonctions μ-harmoniques continues, bornées, sont constantes implique la moyennabilité. ([5] p. 43). Cela prouve la nécessité. La suffisance a été prouvée par Rosenblatt et indépendamment par Kaimanovich et Vershik ([33], [24], [25]) : sur tout groupe moyennable on peut construire une probabilité μ symétrique, absolument continue, pour laquelle les μ-harmoniques bornées sont constantes et donc $h(\mu) = 0$.

Le second problème n'a pas encore de solution complète. D'après la loi "zéro ou un" de Hewitt et Savage, qui implique le théorème de Choquet-Deny, si G est abélien $h(\mu) = 0$ pour tout μ . Il est aussi connu que, si G est nilpotent de degré 2, les fonctions μ-harmoniques bornées continues sont constantes, quelle que soit μ ; en degré supérieur le résultat reste vrai si μ est étalée , ou a un "moment" ([19]). Pour décrire complètement la classe des groupes G pour lesquels $h(\mu) = 0$, quelle que soit μ adaptée, il faudrait au moins savoir si cette classe contient les groupes à croissance non exponentielle discrets ou non. C'est dans l'étude de ce problème que se présente le résultat le plus remarquable d'Avez.

L'énoncé demande quelques rappels sur la croissance.

On suppose que G est à génération compacte, non compact. A un voisinage V de e, symétrique, compact, qui engendre G, on associe la *"longueur"* :

$$L_V(x) = \inf \{n \in \mathbb{N} \; ; \; x \in V^n \};$$

V^n est l'ensemble des produits $x_1 \ldots . x_n$ où $x_i \in V$. On a

$$L_V(xy) \leqslant L_V(x) + L_V(y).$$

En posant $C_V(n) = m(V^n) = m \{L_V \leqslant n\}$ on a

$$\log C_V(n+k) \leqslant \log C_V(n) + \log C_V(k)$$

et donc, par sous-additivité :

$$\lim_n \frac{1}{n} \log C_V(n) = \inf_n \frac{1}{n} \log C_V(n)$$

Le groupe G est dit à croissance *non exponentielle* si $\lim_n \frac{1}{n} \log C_V(n) = 0$ Il est dit à croissance *polynomiale* si $C_V(n) = O(n^d)$ pour un $d > 0$. Ces deux propriétés ne dépendent pas de l'ensemble V choisi ; elles ne dépendent que de G ([19]).

Théorème *: (Avez [4]) Si G est à croissance non exponentielle, si μ a une densité bornée à support compact, alors $h(\mu) = 0$ (i.e. les fonctions μ-harmoniques, bornées sont constantes).*

En raffinant les arguments on peut prouver un peu plus. Commençons par un lemme dans lequel les notations sont celles introduites ci-dessus

Lemme *: Si μ a une densité bornée φ et si*

$$\sum_{k=1}^{\infty} \mu\{L_V = k\} \log C_V(k) = E[\log C_V(L_V(X_1))] < \infty$$

alors $\tilde{H}(\varphi_n) < \infty$ pour tout n et

$$\tilde{H}(\varphi_n) \leqslant E[\log C_V(L_V(S_n))] - \sum_{k=1}^{\infty} \mu^n\{L_V = k\} \log \mu^n\{L_V = k\} .$$

Démonstration : Sur une partie B de mesure $m(B) < \infty$ le maximum de l'entropie différentielle est réalisé par la densité uniforme (voir partie II). Donc

$$-\int_{\{L_V=k\}} \varphi_n(x) \log \varphi_n(x)\, dm(x) \leqslant \mu^n\{L_V=k\} \log \frac{m\{L_V=k\}}{\mu^n\{L_V=k\}}$$

En sommant sur k on obtient :

$$\tilde{H}(\varphi_n) \leqslant E[\log C_V(L_V(S_n))] - \sum_{k=1}^{\infty} \mu^n\{L_V=k\} \log \mu^n\{L_V=k\} .$$

Comme $\log C_V(L_V(S_n)) \leqslant \sum_1^n \log C_V(L_V(X_i))$, l'hypothèse donne $E[\log C_V(L_V(S_n))] < \infty$. D'autre part

$$\sum_{k=1}^{\infty} \mu^n\{L_V=k\} \log k = E[\log L_V(S_n)] \leqslant E[\log C_V(L_V(S_n))] < \infty \quad \text{car}$$

$k = O(C_V(k))$. Cela implique $-\sum_{k=1}^{\infty} \mu^n\{L_V=k\} \log \mu^n\{L_V=k\} < \infty$

(voir partie VII)

Théorème : *Si* G *est un groupe à croissance non exponentielle, si* μ *a une densité* φ *bornée et vérifie :* $E[\log C_V(L_V(X_1))] < \infty$ *et* $E(L_V(X_1)) < \infty$ *alors* $\lim_n \frac{1}{n} \tilde{H}(\varphi_n) = h(\mu) = 0$ *(les notations sont celles introduites ci-dessus les deux hypothèses de "moments" finis ne dépendent pas de l'ensemble* V *choisi).*

Plus précisément si $E[\log C_V(L_V(X_1))] < \infty$ *est vérifié pour un choix de* V *tel que* $\lim_n \frac{C_V(n)}{C_V(n-1)} = 1$ *alors* $\lim_n \frac{1}{n} \tilde{H}(\varphi_n) = h(\mu) = 0$ *; en particulier si,* G *est à croissance polynômiale, l'hypothèse* $E[\log (L_V(X_1))] < \infty$ *suffit pour que* $\lim_n \frac{1}{n} \tilde{H}(\varphi_n) = h(\mu) = 0$.

Démonstration : Soit $\mathcal{V}$ la partition de G en "couronnes" $\{L_V = k\}$, $k = 1, 2, \ldots$. Alors

$$- \sum_{k=1}^{\infty} \mu^n \{L_V = k\} \log \mu^n \{L_V = k\} = H(S_n^{-1}(\mathcal{V})).$$

Posons $T_n = \sum_{i=1}^{n} L_V(X_i)$. On a $L_V(S_n) \leqslant T_n$. Les propriétés de H, rappelées en appendice, permettent d'écrire :

$H(S_n^{-1}(\mathcal{V})) \leqslant H(S_n^{-1}(\mathcal{V}), T_n) = H(T_n) + EH(S_n^{-1}(\mathcal{V})/T_n)$. Comme T_n est une marche aléatoire sur $\mathbb{Z}$, avec $H(T_1) < \infty$ car $\sum_{k=1}^{\infty} \mu\{T_1 = k\} \log k < \infty$, on a $\lim_n \frac{1}{n} H(T_n) = 0$. (Voir partie VII). Sachant $T_n = k$, S_n ne peut se trouver que dans, au plus, k atomes de $\mathcal{V}$; donc $H(S_n^{-1}(\mathcal{V})/T_n = k) \leqslant \log k$. Cela donne $EH(S_n^{-1}(\mathcal{V})/T_n) \leqslant E(\log T_n)$.

En vertu du lemme, on obtient :

$$\tilde{H}(\varphi_n) \leqslant E[\log C_V(L_V(S_n))] + E(\log T_n) + o(n).$$
$$\leqslant 2E[\log C_V(T_n)] + o(n).$$

La suite sous-additive $\log C_V(T_n)$ est intégrable et positive, donc $\frac{1}{n} \log C_V(T_n)$ converge p.s. et en moyenne d'après le théorème ergodique sous-additif. D'après l'hypothèse $E(L_V(X_1)) = E(T_1) < \infty$, $T_n = O(n)$ p.s. Le groupe G étant à croissance non exponentielle, on obtient $\lim_n \frac{1}{n} \log C_V(T_n) = 0$ p.s. et donc $\lim_n \frac{1}{n} E(\log C_V(T_n)) = 0$.

La première partie du théorème est ainsi démontrée.

Supposons que $\lim \frac{C_V(n)}{C_V(n-1)} = 1$; c'est évidemment le cas si G est

à croissance polynomiale.

Alors $\log C_V(T_n) - \log C_V(T_n - T_1) = \log \frac{C_V(T_n)}{C_V(T_n - T_1)} \xrightarrow[n\to\infty]{} 0$ p.s. D'après la méthode de [9], cela implique $\lim_n \frac{1}{n} E(\log C_V(T_n)) = 0$, sans l'hypothèse $E(T_1) < \infty$. Si G est à croissance polynomiale $C_V(n) \simeq n^d$, donc $E[\log L_V(X_1)] < \infty$ implique $E[\log C_V(L_V(X_1)] < \infty$ et le théorème est complètement démontré.

L'énoncé précédent conduit à la question naturelle, mais difficile, suivante :

"Sur un groupe G *à croissance non exponentielle, a-t-on* $h(\mu)=0$, *quelle que soit* μ *(absolument continue) ?"*. i.e. les hypothèses de "moments" finis posées dans le théorème sont-elles indispensables ? Cette question se pose d'autant plus qu'on sait que pour G discret à croissance polynomiale la réponse est positive ; dans ce cas, l'hypothèse de moment logarithmique fini pour μ, posée dans le théorème, sous laquelle on a vu que $\lim_n \frac{1}{n} \tilde{H}(\varphi_n) = h(\mu) = 0$, n'est pas indispensable. En effet, Gromov a démontré que tout groupe discret à croissance polynomiale est une extension finie d'un groupe nilpotent [19] ; sur un groupe nilpotent discret les fonctions μ-harmoniques bornées sont constantes, comme on l'a déjà mentionné, et cette propriété est conservée par extension finie. Cela ne suffit pas à résoudre complètement la question, car Grigorchuk a démontré qu'il existe des groupes discrets à croissance non exponentielle qui ne sont pas à croissance polynomiale [20].

Concernant le problème de la description des probabilités μ, adaptées sur G moyennable, telles que $h(\mu) = 0$, le résultat général le plus frappant a été obtenu par Guivarc'h, dans le cadre de son étude de la loi des grands nombres sur les groupes de Lie [22]. Guivarc'h a introduit la notion de *croissance* d'une probabilité μ dans G, qui est la suivante dans le cas où μ a une densité continue.

Définition *: On appelle croissance dans* G *d'une probabilité* μ *de densité continue* φ *sur* G *, la borne inférieure* $\mathcal{C}(\mu)$ *des réels* $\alpha>0$ *tels qu'il existe une suite croissante exhaustive de boréliens* A_n *dans* G *vérifiant* $\lim_n \mu^n(A_n) = 1$ et $\limsup_n \frac{1}{n} \log m(A_n) \leqslant \alpha$.

Cette notion apparaît comme plus primitive que la notion d'entropie asymptotique elle-même.

Proposition *: Si* μ *a une densité continue* φ *à support compact* V *, on a* $h(\mu) \leqslant \mathcal{C}(\mu)$.

Démonstration : On peut alors prendre les A_n de la définition précédente vérifiant $A_n \subset V^n$. On a

$$\tilde{H}(\varphi_n) = -\int_{A_n} \varphi_n(x) \log \varphi_n(x)\, dm(x) - \int_{V^n \setminus A^n} \varphi_n(x) \log \varphi_n(x)\, dm(x).$$

Le maximum de l'entropie différentielle étant réalisé par la densité uniforme, on obtient :

$$\tilde{H}(\varphi_n) \leqslant \mu^n(A_n) \log \frac{m(A_n)}{\mu^n(A_n)} + \mu^n(V^n \setminus A_n) \log \frac{m(V^n \setminus A_n)}{\mu^n(V^n \setminus A_n)}, \text{ d'où on tire}$$

$$\tilde{H}(\varphi_n) \leqslant \log m(A_n) + (1-\mu^n(A_n)) \log m(V^n) - \mu^n(A_n) \log \mu^n(A_n) - \mu^n(V^n \setminus A_n) \log \mu^n(V^n \setminus A_n).$$

Comme $\frac{1}{n} \log m(V^n)$ est borné, on obtient

$$\lim_n \frac{1}{n} \tilde{H}(\varphi_n) = h(\mu) \leqslant \operatorname{limsup}_n \frac{1}{n} \log m(A_n),$$ ce qui prouve la proposition.

Dans le cas où G est un groupe de Lie connexe moyennable, la version de la loi forte des grands nombres de [22], s'énonce ainsi : $\lim_n \frac{1}{n} L_V(S_n) = 0$ p.s. si μ est *centrée*, c'est à dire si $\int_G \gamma(x)\, d\mu(x) = 0$ pour tout homomorphisme de groupe γ, de G dans $\mathbb{R}$; cela implique $\mathcal{C}(\mu) = 0$. On arrive ainsi au théorème suivant ([22] p. 75)).

Théorème : *Si G est un groupe de Lie connexe moyennable, si μ est une probabilité centrée, de densité φ continue à support compact, alors*

$$\lim_n \frac{1}{n} \tilde{H}(\varphi_n) = h(\mu) = 0$$ *(i.e. les fonctions μ-harmoniques bornées sont constantes).*

Pour conclure cette partie, décrivons la classe des groupes des *"configurations finies"* $F_0(\mathbb{Z}^k, \mathbb{Z}/2\mathbb{Z})$, introduite par Kaimanovich et Vershik pour résoudre la question suivante d'Avez [3] :

"Sur un groupe, discret, à croissance exponentielle, peut-il exister une probabilité à support fini μ, adaptée, telle que $h(\mu) = 0$?".

Le groupe $F_0(\mathbb{Z}^k, \mathbb{Z}/2\mathbb{Z})$ est le produit semi-direct de $\mathbb{Z}^k$, avec le groupe des applications à support fini dans $\mathbb{Z}^k$, à valeurs dans le groupe des entiers modulo 2, l'action de $\mathbb{Z}^k$ sur la seconde composante étant définie par les tranlations. Un élément de ce groupe a la forme (i, χ_A) où $i \in \mathbb{Z}^k$ et A est une partie finie de $\mathbb{Z}^k$ (χ désignant l'indicatrice); la loi de groupe s'écrit

$$(i, \chi_A)\,(j, \chi_B) = (i+j, \chi_A + \chi_{B+i} \bmod 2).$$

Ces groupes ($k \geqslant 1$) sont résolubles de degré 2 et à croissance exponentielle (donc ils ne sont pas nilpotents). Kaimanovich et Vershik ont montré que la mesure μ symétrique, élémentaire sur $F_o(\mathbb{Z}, \mathbb{Z}/2\mathbb{Z})$, définie par $\mu(1,0) = \mu(-1, 0) = 1/4$ et $\mu(0, \chi_{\{0\}}) = 1/2$, a une croissance $\mathcal{C}(\mu)$ nulle dans ce groupe [25]. D'après ce qu'on a vu ci-dessus, cela donne $h(\mu) = 0$, ce qui répond positivement à la question posée. De façon plus précise, ces auteurs démontrent le résultat suivant :

> ***Proposition*** *: Soit μ à support fini sur $F_o(\mathbb{Z}^k, \mathbb{Z}/2\mathbb{Z})$. Une condition nécessaire et suffisante pour que les fonctions μ-harmoniques bornées soient constantes (i.e. $h(\mu) = 0$) est que la première composante de S_n, qui forme une marche aléatoire sur $\mathbb{Z}^k$, soit récurrente. Si de plus μ est symétrique, cela est réalisé seulement pour* k=1 et k=2 ([25], p. 482).

L'idée de la démonstration est que, si la première composante de S_n est transitoire sur $\mathbb{Z}^k$, alors la seconde composante de S_n a une "limite" p.s. non constante, qui engendre des fonctions μ-harmoniques bornées non constantes.

Chacun de ces groupes est moyennable, donc porte une mesure μ adaptée telle que $h(\mu) = 0$. Kaimanovich et Vershik affirment aussi que, pour $k \geqslant 3$, cette mesure μ vérifie nécessairement $H(\mu) = \infty$. (voir fin de la partie III). ([25] p. 481).

V

FRONTIERE, ENTROPIE ASYMPTOTIQUE ET FORMULE DE FÜRSTENBERG

La notion d'entropie asymptotique $h(\mu)$ étudiée dans les deux parties précédentes est apparue pour la première fois sous une forme quelque peu différente, dans le travail de Fürstenberg sur les produits de matrices unimodulaires indépendantes ([16]). Considérons $G = Sl(2,\mathbb{R})$ et μ une probabilité adaptée sur G, suffisamment régulière. S'il existe sur l'espace projectif $\mathbb{P}_1$ une mesure Π, équivalente à la mesur de Lebesgue, telle que $\mu * \Pi = \Pi$, la convolution étant déduite de l'action canonique de $Sl(2, \mathbb{R})$ sur $\mathbb{P}_1$, Fürstenberg démontre que

$$\lim_n \frac{1}{n} \log \|S_n\| = -\frac{1}{2} \iint \log \frac{dx^{-1}\Pi}{d\Pi}(y)\, d\mu(x)\, d\Pi(y) \quad \text{p.s.}$$

($\|\ .\ \|$ désigne la norme euclidienne d'une matrice de $Sl(2, \mathbb{R})$). Il apparaît, a posteriori, que cette quantité n'est autre que l'entropie asymptotique de la marche aléatoire S_n. C'est ce qu'on se propose d'expliquer, en général, dans cette partie.

On reprend les notations des parties précédentes. On considère de plus l'action mesurable de G sur l'espace Ω des trajectoires $(s_1,\ldots, s_n\ldots)$ de la marche aléatoire, définie par :

$$(x, (s_1,\ldots, s_n,\ldots)) \mapsto (xs_1,\ldots, xs_n,\ldots).$$

Cette action commute avec l'opérateur de décalage ϑ. Elle engendre un produit de convolution $\rho * \nu$, pour une mesure ρ sur G et ν sur Ω. La convolution $\delta_x * P_e$, $x \in G$, représente la loi de la marche aléatoire issue de x, $(x\, S_n)_n$. On a donc $\vartheta P_e = \mu * P_e$.

La loi conjointe de S_1 et S_{n+1} s'écrit :

$$P_e(S_1 \in B_1,\ S_{n+1} \in B_{n+1}) = \int_{B_1} d\mu(x)\ (\delta_x * \mu^n)\ (B_{n+1})$$

$$= \int_{B_1} d\mu(x)\ (\delta_x * P_e)\ (S_n \in B_{n+1})$$

La "loi conjointe" de S_1 et $\mathcal{J}$ s'écrit :

$$P_e\ ((S_1 \in B_1) \cap \mathcal{J}) = \int_{B_1} d\mu(x)\ (\delta_x * P_e)\ (\mathcal{J})$$

où $J \in \mathcal{J}$. Le calcul de l'information mutuelle $I(S_1, S_n)$ ou $I(S_1, \mathcal{J})$ obéit à un principe général exprimé par la proposition suivante :

Proposition : *Soit* $(E, \mathcal{E}, m)$ *un espace probabilisé. Soit* $T(x,B)$ *une probabilité de transition, définie sur* E *, à valeurs dans* $(F, \mathcal{F})$ *, la tribu* $\mathcal{F}$ *étant à base dénombrable. Soit* $m' = Tm$ *la probabilité image sur* $(F, \mathcal{F})$:

$m'(B) = \int_E T(x,B)\, dm(x)$. *Soit* ρ *la mesure définie sur l'espace produit* $E \times F$, $\mathcal{E} \otimes \mathcal{F}$, *par* $\rho(A \times B) = \int_A T(x, B)\, dm(x)$.

Alors ρ *est absolument continue par rapport à* $m \otimes m'$ *si et seulement si, pour* m *presque tout* x, $T(x, .)$ *est absolument continue par rapport à* m' *et alors*

$$\frac{d\rho}{dm \otimes dm'}(x,y) = \frac{dT(x,.)}{dm'}(y) \quad m \otimes m' \text{ p.p.}$$

L'entropie relative $H(\rho ; m \otimes m')$ *est finie si et seulement si l'intégrale*

$$\iint \left(\log \frac{dT(x,.)}{dm'}(y)\right) dT(x,.)(y)\, dm(x)$$

existe et alors elles sont égales.

Démonstration : Elle repose sur les arguments ordinaires de la différentiation, aussi on ne fait que l'esquisser. Soit $\mathcal{F}_k$ une suite croissante de sous-tribus finies de $\mathcal{F}$, formées d'ensembles de m'-mesure > 0 et telles que $\mathcal{F}$ soit engendrée par $\bigcup_k \mathcal{F}_k$ mod m. La fonction $\frac{T(x, F_k(y))}{m'(F_k(y))}$, où $F_k(y)$ désigne l'atome de $\mathcal{F}_k$ contenant y, est une version de la densité de ρ par rapport à $m \otimes m'$ en restriction à $\mathcal{E} \otimes \mathcal{F}_k$; ρ est absolument continue par rapport à $m \otimes m'$ si et seulement si cette suite de fonctions, qui forme une martingale, converge dans $L^1(m \otimes m')$; $T(x,.)$ est absolument continue par rapport à m', si et seulement si cette suite converge, à x fixé, dans $L^1(m')$. Cela démontre la première partie de la proposition. L'assertion concernant l'entropie relative résulte de la première partie, en vertu du théorème de Gelfand-Yaglom - Perez rappelé dans l'appendice 1.

Il faut souligner que cette proposition peut être en défaut si la tribu $\mathcal{F}$ n'est pas à base dénombrable et ceci complique certains des énoncés qui vont suivre. En prenant $(E, \mathcal{E}) = (G, \mathcal{B})$, $(F, \mathcal{F}) = (\Omega, \sigma(S_n, \ldots))$, $T(x, B) = \delta_x * P_e(B)$ on trouve tout d'abord la formule suivante pour $I(S_1, S_{n+1})$:

$$I(S_1, S_{n+1}) = \iint (\log \frac{d\delta_x * \mu^n}{d\mu^{n+1}}(y))\, d(\delta_x * \mu^n)(y)\, d\mu(x)$$

$$= \iint (\log \frac{d\delta_x * P_e}{d\vartheta P_e}\Big|_{\sigma(S_n \ldots)}(\omega))\, d\delta_x * P_e(\omega)\, d\mu(x).$$

Dans la seconde la densité de $\delta_x * P_e$ par rapport à ϑP_e est prise en restriction à la tribu $\sigma(S_n, \ldots)$. Le groupe G étant supposé à base dénombrable, l'hypothèse de la proposition est bien remplie.

Considérons maintenant $(F, \mathcal{F}) = (\Omega, \mathcal{J})$. Sur $\mathcal{J}$ on a $P_e = \vartheta P_e = \int_G \delta_x * P_e \, d\mu(x)$. Mais il n'est plus possible, même dans les cas usuels, de considérer $\mathcal{J}$ à base dénombrable : par exemple, pour une rotation ergodique sur le cercle la tribu des invariants n'est pas à base dénombrable. Cependant l'espace mesuré $(\Omega, \mathcal{J}, P_e)$ est "séparable" en tant que sous-espace de $(\Omega, \mathcal{B}^\infty, P_e)$. Il existe donc au moins une sous-tribu $\mathcal{J}'$ de $\mathcal{J}$ qui soit à base dénombrable et qui soit égale à $\mathcal{J}$ mod. P_e. On a évidemment $h(\mu) = I(S_1, \mathcal{J}) = I(S_1, \mathcal{J}')$. La proposition précédente peut alors être appliquée à $\mathcal{J}'$.

Théorème : *Soit $\mathcal{J}'$ une sous tribu à base dénombrable de la tribu $\mathcal{J}$ telle que $\mathcal{J}' = \mathcal{J}$ mod. P_e. L'entropie asymptotique $h(\mu)$ de la marche aléatoire s'écrit :*

$$h(\mu) = \iint (\log \frac{d\delta_x * P_e}{dP_e}\Big|_{\mathcal{J}'}(\omega))\, d\delta_x * P_e(\omega)\, d\mu(x) \; ; \; \textit{si } \mathcal{J}' \textit{ est stable sous } G$$

$$h(\mu) = -\int_G d\mu(x) \int_\Omega (\log \frac{d\delta_{x^{-1}} * P_e}{dP_e}\Big|_{\mathcal{J}'}(\omega))\, dP_e(\omega).$$

Cet énoncé signifie que $h(\mu)$ est finie si et seulement si l'intégrale considérée existe et alors ces deux quantités sont égales. On passe de la 1ère intégrale à la 2ème par le changement de variables $\omega' = x\omega$ (action de G sur Ω).

Quand $h(\mu) > 0$, l'espace des fonctions μ-harmoniques bornées est non trivial. Sa description complète est un problème difficile. Pour ce faire on introduit la notion de μ-frontière.

Soit $(M, \mathcal{M})$ un G-espace mesurable, la tribu $\mathcal{M}$ étant à base dénombrable. Soit ν une probabilité μ-invariante sur M :

$$\mu * \nu = \int_G \delta_x * \nu \, d\mu = \nu ,$$

la convolution étant déduite de l'action de G sur M.

Pour toute fonction f réelle, $\mathcal{B}$-mesurable, bornée sur M on définit une fonction μ-harmonique bornée sur G, $g = Rf$ par la formule :

$$g(x) = \int_M f(xz)\, d\nu(z) = \int_M f(z)\, d\, \delta_x * \nu(z).$$

L'opérateur R est linéaire, contractant pour la norme sup.

Définition : *L'espace mesuré* $(M, \mathcal{B}, \nu)$ *où* $\mathcal{B}$ *est à base dénombrable et où* ν *est* μ*-invariante, est une* μ*-frontière si l'opérateur* R *défini ci-dessus possède les deux propriétés suivantes :*

1) R *est injectif i.e.* $Rf = 0$ *sur* G *si et seulement si* $f = 0\ \delta_x * \nu$ p.p. *pour tout* $x \in G$.

2) R *est multiplicatif pour le produit propre* $\square$ *des fonctions* μ*-harmoniques bornées :* $R(ff') = Rf \square Rf'$.

Le produit propre de deux fonctions μ-harmoniques bornées g et g' est défini par :

$$g''(x) = g \square g'(x) = \lim_n \int_G g(xy)\, g'(xy)\, d\mu^n(y)$$

$$= \int_\Omega \lim_n g(x\, S_n)\, g'(x\, S_n)\, dP_e.$$

Il introduit une structure d'algèbre sur l'espace des fonctions μ-harmoniques bornées, isomorphe à celle de l'algèbre des v.a. invariantes, considérées modulo $\delta_x * P_e$, pour tout $x \in G$.

Définition : *La* μ*-frontière* $(M, \mathcal{B}, \nu)$ *est appelée frontière de Poisson, associée à* (G, μ) *, si de plus l'opérateur* R *est surjectif.*

A l'origine les notions de μ-frontière et de frontière de Poisson, ont été introduites par Fürstenberg de façon un peu différente ([17]). Fürstenberg définit une μ-frontière comme un G-espace compact M, métrisable, portant une probabilité ν, μ-invariante, de telle sorte que $\delta_{S_n} * \nu$ converge p.s. vers une mesure de Dirac sur M. On vérifie aussitôt qu'alors les deux conditions de la définition donnée ci-dessus sont vérifiées.

La relation avec l'entropie asymptotique apparaît dans l'énoncé suivant :

Théorème : *Si* $(M, \mathcal{M}, \nu)$ *est une* μ*-frontière, on a l'inégalité :*

$$h(\mu) \geq -\int_G d\mu(x) \int_M (\log \frac{d\delta_{x^{-1}} * \nu}{d\nu}(z))\, d\nu(z).$$

C'est une égalité si et seulement si $(M, \mathcal{M}, \nu)$ *est frontière de Poisson, dans le cas où* $h(\mu) < \infty$.

Démonstration : D'après la proposition donnée au début de cette partie, l'intégrale apparaissant ici est l'entropie relative de la mesure, sur $G \times M$, $\int_B d\mu(x)\, \delta_x * \nu(N)$, par rapport à la mesure $\mu \otimes \nu\ (B \times N)$. D'après la définition de μ-frontière, la tribu $\mathcal{M}$ munie de la famille des probabilités $\delta_x * \nu$, $x \in G$, est représentée par une sous-tribu $\mathcal{Y}$ de $\mathcal{I}$, stable sous l'action de G, à base dénombrable, munie de la famille $\delta_x * P_e$, $x \in G$. En effet à $N \in \mathcal{M}$ correspond la fonction μ-harmonique $R\chi_N$; à cette fonction correspond une classe mod. $\delta_x * P_e$, pour tout x, de v.a. invariante. Comme R est multiplicatif cette v.a. invariante est l'indicatrice d'un événement invariant. L'intégrale considérée ci-dessus n'est autre que l'information mutuelle $I(S_1, \mathcal{Y})$. L'inégalité avec $h(\mu)$ est alors évidente.

Si c'est une égalité $h(\mu) = I(S_1, \mathcal{I}) = I(S_1, \mathcal{Y})$; par le lemme de la partie III on trouve aussi $I((S_1 \ldots S_k), \mathcal{Y}) = I(S_k, \mathcal{Y}) = kI(S_1, \mathcal{Y})$ $= I((S_1, \ldots S_k), \mathcal{I})$. Ceci étant vrai pour tout k, les tribus $\mathcal{I}$ et $\mathcal{Y}$ sont égales mod P_e. De même on montre leur égalité mod $\delta_x * P_e$, pour tout $x \in G$. Alors toutes les fonctions μ-harmoniques bornées sont représentées sur Ω par une v.a. $\mathcal{Y}$ mesurable, mod $\delta_x * P_e$ pour tout x. Cela prouve la surjectivité de l'opérateur R.

Ce théorème donne un critère qui permet de reconnaitre la frontière de Poisson parmi les μ-frontières éventuelles ; autrement dit, on peut ainsi reconnaître si la μ-frontière considérée permet la représentation de toutes les fonctions μ-harmoniques bornées.

Dans le cas discret Kaimanovich et Vershik ont développé le point précédent de façon plus simple et introduit la notion de transformée de Radon-Nikodym [25]. Ils ont montré, en particulier, que l'inégalité du théorème précédent est valide pour tout G-espace muni d'une probabilité μ-invariante, même si ce n'est pas une μ-frontière, mais sous l'hypothèse $H(\mu) < \infty$. Il serait intéressant de savoir si cela reste vrai en général.

Pour illustrer le principe qu'on vient de décrire, considérons l'exemple de la marche aléatoire élémentaire sur le groupe libre F_2 à deux générateurs a et b: La probabilité μ est alors $\mu = \frac{1}{4}(\delta_a + \delta_b + \delta_{a^{-1}} + \delta_{b^{-1}})$. Calculons, tout d'abord, de façon directe l'entropie asymptotique $h(\mu)$. En notant $L(x)$ la longueur de $x \in F_2$, c'est à dire le nombre de symboles de son écriture réduite, on a :

$$H(S_n) = H(S_n, L(S_n)) = H(L(S_n)) + EH(S_n/L(S_n)).$$

Sachant $L(S_n) = k$, S_n est équirépartie sur les $4(3^{k-1})$ éléments de F_2, de longueur k. Donc :

$$EH(S_n/L(S_n)) = \sum_{k \geq 1} \log\,(4(3^{k-1}))\ P(L(S_n) = k)$$

$$= (\log 4) + (\log 3)\, E(L(S_n) - 1).$$

Pour n grand, $L(S_n)$ se comporte comme la marche aléatoire sur $\mathbb{Z}$ définie par $\frac{3}{4}\delta_1 + \frac{1}{4}\delta_{-1}$. Cela donne

$$\lim_n \frac{1}{n} H(L(S_n)) = 0,$$

$$\lim_n \frac{1}{n} E(L(S_n)) = \frac{1}{2}$$

d'où $$\lim_n \frac{1}{n} H(S_n) = h(\mu) = \frac{1}{2} \log 3.$$

Considérons maintenant la μ-frontière formée par l'espace M des mots réduits infinis écrits avec les quatre symboles a, b, a^{-1}, b^{-1}, et ν la probabilité sur M obtenue comme la loi de la chaîne de Markov à quatre états, de transition :

	a	b	a^{-1}	b^{-1}
a	1/3	1/3	0	1/3
b	1/3	1/3	1/3	0
a^{-1}	0	1/3	1/3	1/3
b^{-1}	1/3	0	1/3	1/3

et de distribution initiale uniforme ($\frac{1}{4}$ sur chaque état). Alors

$$\frac{d\delta_a * \nu}{d\nu}(m) = 3 \quad \text{pour} \quad m \in \{m_o = a \text{ ou } b \text{ ou } b^{-1}\}$$

$$= \frac{1}{3} \quad \text{pour} \quad m \in \{m_o = a^{-1}\},$$

d'où $$\int \log \frac{d\delta_a * \nu}{d\nu}(m)\ d\nu(m) = \frac{3}{4} \log 3 - \frac{1}{4} \log 3 = \frac{1}{2} \log 3$$

On obtient finalement :

$$\int_G d\mu(x) \int_M \log \frac{d\delta_{x^{-1}} * \nu}{d\nu}(m)\ d\nu(m) = \frac{1}{2} \log 3.$$

Le théorème précédent nous permet de conclure que (M, ν) est l'espace de Poisson de (F_2, μ). Autrement dit l'ensemble des fonctions μ-harmoniques bornées est exhaustivement représenté par l'espace des fonctions mesurables bornées sur M, par l'intermédiaire de la formule :

$$g(x) = \int_M f(x\ m)\ d\nu(m).$$

Ce résultat a été prouvé pour la première fois, par un argument direct différent, par Dynkin et Maliutov ([12]). Le calcul direct de l'entropie effectué ci-dessus, montre aussi que le développement au "2ème ordre" de $H(S_n)$ donne un terme de l'ordre de $\frac{1}{2} \log n$, comme sur $\mathbb{Z}$ (voir partie II).

Dans [28] et [29], Ledrappier a utilisé le même principe "entropique" pour obtenir la frontière de Poisson de certains groupes discrets de matrices.

VI

ENTROPIE ET SYMETRIE

On se propose, dans cette partie, de comparer les fonctions harmoniques et les entropies des marches aléatoires droite $S_n = X_1 \ldots X_n$ et gauche $S'_n = X_n \ldots X_1$ définies par une probabilité μ sur G. Par l'automorphisme $x \mapsto x^{-1}$ la marche aléatoire droite $S_n = X_1 \ldots X_n$ est changée en la marche aléatoire gauche $S_n^{-1} = X_n^{-1} \ldots X_1^{-1}$ définie par la probabilité $\check{\mu}$, image de μ par l'automorphisme inverse. L'étude de la marche aléatoire gauche définie par μ se ramène ainsi à celle de la marche aléatoire droite définie par $\check{\mu}$. On se propose donc de comparer les fonctions μ-harmoniques et $\check{\mu}$-harmoniques, ainsi que les entropies asymptotiques $h(\mu)$ et $h(\check{\mu})$ en ne considérant, comme précédemment, que des marches aléatoires droites.

La fonction module du groupe G est notée Δ :

$$\int_G f(x)\ dm(x) = \int_G f(x^{-1})\ \Delta(x)^{-1}\ dm(x)$$

(m désigne comme précédemment la mesure de Haar gauche). Si μ^n a une densité φ_n par rapport à m, alors $(\check{\mu})^n$ a une densité, notée Ψ_n donnée par $\Psi_n(x) = \varphi_n(x^{-1})\,\Delta(x)^{-1}$.

En effet, $X_1....X_n$ et $X_n....X_1$ ayant même loi μ^n on trouve :

$$\int_G f(x)\, d(\check{\mu})^n(x) = E\,[f(X_1^{-1}...X_n^{-1})] = E\,[f(X_1....X_n)^{-1}]$$

$$= \int_G f(x^{-1})\, d\mu^n(x) = \int_G f(x^{-1})\varphi_n(x)\, dm(x) = \int_G f(x)\varphi_n(x^{-1})\Delta(x)^{-1} dm(x)$$

Il est alors facile de comparer les entropies différentielles $\tilde{H}(\varphi_n)$ et $\tilde{H}(\Psi_n)$.

Proposition *: Si l'intégrale* $\int_G \log\Delta(x)\, d\mu(x) = E(\log\Delta(X_1))$ *converge, l'entropie différentielle* $\tilde{H}(\mu^n) = \tilde{H}(\varphi_n)$ *est finie si et seulement si l'entropie différentielle* $\tilde{H}(\check{\mu}^n) = \tilde{H}(\Psi_n)$ *l'est. Alors* $\tilde{H}(\Psi_n) = \tilde{H}(\varphi_n) - n\,E(\log\Delta(X_1))$.

Démonstration : C'est un calcul direct :

$$\tilde{H}(\Psi_n) = -\int \log\,(\varphi_n(x^{-1})\,\Delta(x)^{-1})\varphi_n(x^{-1})\,\Delta(x)^{-1}\, dm(x)$$

$$= -\int \log\,(\varphi_n(x)\,\Delta(x))\,\varphi_n(x)\, dm(x)$$

$$= \tilde{H}(\varphi_n) - \int (\log\Delta(x))\,\varphi_n(x)\, dm(x)\ ;$$

Δ étant multiplicatif, $\int (\log\Delta(x))\,\varphi_n(x)\, dm(x) = E(\log\Delta(S_n))$ $= n\,E(\log\Delta(X_1))$.

Théorème *: Dans le cas absolument continu, les entropies différentielles* $\tilde{H}(\mu^n) = \tilde{H}(\varphi_n)$ *étant finies, si l'intégrale* $E(\log\Delta(X_1)) = \int_G \log\Delta(x)\, d\mu(x)$ *converge, l'entropie asymptotique vérifie l'inégalité :*

$$h(\mu) \geqslant E(\log\Delta(X_1)).$$

De plus

$$h(\check{\mu}) = h(\mu) - E(\log\Delta(X_1)).$$

Démonstration : D'après la partie III, $h(\mu) = \lim_n \frac{1}{n}\,\tilde{H}(\varphi_n)$, et $h(\check{\mu}) = \lim_n \frac{1}{n}\,\tilde{H}(\Psi_n)$. La proposition donne alors

$$h(\check{\mu}) = h(\mu) - E(\log\Delta(X_1)).$$

La quantité $h(\check{\mu})$ est positive, donc $h(\mu) \geqslant E(\log\Delta(X_1))$.

Ce théorème a plusieurs corollaires remarquables.

Corollaire: Si G est unimodulaire (i.e. $\Delta \equiv 1$), si μ est absolument continue avec les entropies différentielles $\tilde{H}(\varphi_n)$ finies, alors $h(\check{\mu}) = h(\mu)$. En particulier, si toutes les fonctions μ-harmoniques bornées sont constantes il en est de même pour $\check{\mu}$.

La seconde assertion du corollaire résulte de la première en vertu de la partie II. On peut se demander si l'égalité $h(\check{\mu}) = h(\mu)$ est vraie, pour G unimodulaire, sans restriction sur μ. D'après Kaimanovich et Vershik la réponse à cette question est négative, même sur un groupe discret. Dans le cas discret, le corollaire précédent dit que $h(\mu)=h(\check{\mu})$ dès que $H(\mu) < \infty$. Mais si $H(\mu) = \infty$, des fonctions bornées non constantes harmoniques pour μ peuvent exister alors qu'il n'en existe pas pour $\check{\mu}$ ([25] p. 483). Cela montre, de nouveau, qu'il n'y a pas en général de principe d'approximation de $h(\mu)$ par des mesures à support fini. (voir la fin de la partie III).

Corollaire : Si G n'est pas unimodulaire, si μ est absolument continue avec les entropies différentielles $\tilde{H}(\varphi_n)$ finies, si $\int_G \log\Delta(x)\, d\mu(x) > 0$, alors il existe des fonctions μ-harmoniques bornées non constantes.

Considérons l'exemple du groupe affine de la droite réelle : $G = \left\{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix};\ a \neq 0,\ a, b, \in \mathbb{R}\right\}$. La mesure de Haar gauche est donnée par $dm(a,b) = \frac{1}{a^2}\, dadb$; la fonction module est $\Delta \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = 1/a$. Le corollaire précédent nous dit que pour μ à densité bornée à support compact (ou telle que les entropies différentielles $\tilde{H}(\varphi_n)$ soient finies), il existe des fonctions μ-harmoniques bornées non constantes dès que $\int_G \log a(x)\, d\mu(x) < 0$ ($a(x)$ désigne le 1er coefficient de la matrice $x \in G$). On retrouve ainsi, pour une part, un énoncé d'Azencott ([5] ; voir aussi [13]). On sait aussi, dans ce cas, que si $\int_G \log a(x) d\mu(x) > 0$ les fonctions μ-harmoniques bornées sont nécessairement constantes, donc $h(\mu) = 0$. Le théorème précédent donne alors l'énoncé suivant.

Corollaire : Sur le groupe affine de la droite réelle $G = \left\{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} ;\ a \neq 0,\ a,b \in \mathbb{R}\right\}$, si μ a une densité bornée à support

compact, alors

$$h(\mu) = -\int \log a(x)\, d\mu(x) = -E(\log a(X_1))$$

si cette quantité est positive, et 0 sinon.

Sous ces hypothèses, la méthode de la partie V s'applique alors pour décrire l'espace de Poisson de la marche aléatoire définie par μ.

VII

QUELQUES COMPLEMENTS

A) L'entropie, au sens de Kolmogorov-Sinai, d'un processus stationnaire peut être obtenue à partir de la connaissance d'une trajectoire typique, par un passage à la limite. C'est le contenu du théorème de Shannon-Mc Millan-Breiman. Dans [1], Avez a demandé si un énoncé analogue était valide pour l'entropie asymptotique d'une marche aléatoire. Dans le cas discret, avec $H(\mu) < \infty$, la réponse est positive.

Théorème : *Sur un groupe* G *discret, si* $H(\mu) < \infty$, *alors* $\lim_n -\frac{1}{n} \log \mu^n(S_n) = h(\mu)$ P_e p.s. *et dans* $L^1(P_e)$.

Ce théorème a été démontré comme un corollaire du théorème ergodique sous-additif dans [9]. Une démonstration plus directe, plus proche de la démonstration habituelle du théorème de Shannon-Mc Millan-Breiman, a été donnée indépendamment dans [25]. On peut se demander quel énoncé du même type est valide dans le cas non discret.

B) La démonstration du lemme de la partie IV utilise la proposition suivante

Proposition : *Soit* $(\rho_n)_{n \in \mathbb{N}}$ *une distribution de probabilité discrète. L'entropie* $H(\rho) = -\sum_{n=1}^{\infty} \rho_n \log \rho_n$ *est finie si et seulement si la série* $\sum_{n=1}^{\infty} \rho'_n \log n$ *converge, où* (ρ'_n) *désigne la réarrangée décroissante de la suite* ρ_n . *La convergence de la série* $\sum_{n=1}^{\infty} \rho_n \log n$ *est donc une condition suffisante, impossible à améliorer, pour que l'entropie* $H(\rho)$ *soit finie.*

Démonstration : L'entropie est invariante par réarrangement: $H(\rho) = H(\rho')$.

Supposons $\sum_{n=1}^{\infty} \rho_n \log n < \infty$. Soit $A = \{n ; -\log \rho_n < 2 \log n\}$. Alors $-\sum_{n\in A} \rho_n \log \rho_n$ converge. Pour $n \notin A$, $\rho_n < \frac{1}{n^2}$; la convergence de la série $\sum_{n=1}^{\infty} \frac{1}{n^2} \log n^2$ implique donc celle de $-\sum_{n\notin A} \rho_n \log \rho_n$.

Réciproquement, supposons (ρ_n) décroissante et $-\sum_{n=1}^{\infty} \rho_n \log \rho_n < \infty$. Soit $B = \{n; \rho_n > \frac{1}{n}\}$. Si B était infini on construirait la suite croissante t_k : $t_1 = 1$ et $t_{k+1} = \inf \{n > t_k ; \rho_n > \frac{1}{n}\}$. On aurait $\rho_n > \frac{1}{t_{k+1}}$ pour $t_{k+1} \geqslant n > t_k$ en raison de la décroissance de ρ_n. La convergence de $-\sum_{n=1}^{\infty} \rho_n \log \rho_n$ donnerait alors celle de $\sum_{k=1}^{\infty} \frac{t_{k+1} - t_k}{t_{k+1}} \log t_{k+1}$. Or la série $\sum_{k=1}^{\infty} \frac{t_{k+1} - t_k}{t_{k+1}}$ diverge nécessairement. Donc l'ensemble B est fini. Comme $\sum_{n\in B^c} \rho_n \log n \leqslant \sum_{n\in B^c} \rho_n \log \rho_n$, l'hypothèse $H(\rho) < \infty$ implique $\sum_{n=1}^{\infty} \rho_n \log n < \infty$.

Un énoncé analogue est valide pour une densité bornée φ sur $\mathbb{R}$ ou $\mathbb{R}^d$; L'entropie différentielle $\tilde{H}(\varphi) = -\int \varphi(x) \log \varphi(x)\, dx$ est finie si et seulement si le moment logarithmique $\int_1^{+\infty} \bar{\varphi}(x) \log x\, dx$ est fini pour la réarrangée décroissante $\bar{\varphi}$ de φ. D'après le lemme de la partie IV, quand le groupe G est à croissance polynomiale, le moment logarithmique fini pour φ, densité bornée, est une condition suffisante pour que toutes les $\tilde{H}(\varphi_n)$ soient finies. La proposition précédente montre que cette condition de moment ne peut pas être affaiblie.

C) D'après le théorème de la partie IV, si μ est une probabilité sur $\mathbb{Z}$ ayant un moment logarithmique fini, $\lim_n \frac{1}{n} H(\mu^n) = h(\mu) = 0$. L'un des arguments de la démonstration est le suivant :

Proposition : *Si μ est une probabilité sur $\mathbb{Z}$ telle que $\sum \mu(n) \log |n| < \infty$, alors $\lim_n \frac{1}{n} \log S_n = 0$ p.s. et en moyenne.*

Pour μ quelconque sur $\mathbb{Z}$, il résulte du théorème de Choquet-Deny que $h(\mu) = 0$. Il serait intéressant de produire une démonstration directe de ce fait. Cela fournirait probablement de nouveaux arguments conduisant à des versions du théorème de Choquet-Deny pour d'autres situations.

Si μ a un moment d'ordre 2, l'ordre de grandeur attendu pour $H(\mu^n)$ est $\frac{1}{2} \log n$; si μ appartient au domaine d'attraction d'une loi stable d'indice α, c'est $\frac{1}{\alpha} \log n$ (voir la partie II). En considérant des exemples où $\mu(n) \sim \frac{1}{n(\log n)^a}$, $a > 2$, on peut se convaincre que $H(\mu^n)$ peut être de l'ordre de $n^{1-\varepsilon}$, quel que soit $\varepsilon > 0$. L'énoncé $\lim_n \frac{1}{n} H(\mu^n) = 0$, pour $H(\mu) < \infty$, semble donc le meilleur possible, même sur $\mathbb{Z}$.

D) On peut aussi se demander si la croissance $\mathcal{C}(\mu)$ dans $\mathbb{Z}$ ou $\mathbb{R}$ d'une probabilité μ, de densité continue φ, est une quantité accessible. Plus précisément, en suivant la définition de $\mathcal{C}(\mu)$ donnée dans la partie IV, peut-on préciser l'ordre de grandeur de la suite des mesures de Lebesgue $m(A_n)$, pour une suite croissante de boréliens A_n telle que $\int_{A_n} \varphi_n(x)\, dx > 1 - \varepsilon$? Il apparaît qu'on peut avoir $\mathcal{C}(\mu) = \infty$ et même que la suite $m(A_n)$ peut croître à une vitesse arbitrairement grande. La démonstration de ce fait repose sur un théorème de Polya : "Toute fonction réelle f, paire, telle que $f(0) = 1$, convexe sur $[0, +\infty[$, est la fonction caractéristique d'une mesure de probabilité sur $\mathbb{R}$" ([14] p. 509). Indiquons rapidement cette démonstration. Pour une densité continue φ, paire, décroissante sur $[0, +\infty[$, la croissance de $m(A_n)$ est celle de la suite

$$\inf \{a > 0 ; \int_{-a}^{+a} \varphi_n(x)\, dx > 1 - \varepsilon\}$$

ou encore celle de

$$\inf \{a > 0 ; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2}\, \varphi_n(ua)\, adu > 1-\varepsilon\}.$$

Par l'identité de Parseval et après troncation de l'intégrale on est ramené à considérer

$$\inf \{a > 0 ; \frac{2}{\sqrt{2\pi}} \int_0^2 e^{-u^2/2}\, (\hat{\varphi})^n(\tfrac{u}{a})\, du > 1-\varepsilon\}.$$

où $\hat{\varphi}$ désigne la fonction caractéristique de φ. En prenant $\hat{\varphi}$ comme

dans le théorème de Polya on a

$$1 + (\hat{\varphi})^n \left(\frac{2}{a}\right) > \int_0^2 (\hat{\varphi})^n \left(\frac{u}{a}\right) du$$

et l'ordre de grandeur cherché est minoré par celui de

$\inf \{a > 0 ; (\hat{\varphi})^n \left(\frac{2}{a}\right) > \delta\}$ pour un δ, $0 < \delta < 1$. La pente de $\hat{\varphi}$ au voisinage de 0 pouvant être arbitraire, cette quantité peut croître avec n aussi vite que l'on veut.

APPENDICE 1

QUELQUES GENERALITES SUR LES NOTIONS D'ENTROPIE ET D'INFORMATION

Cet appendice ne contient que des rappels, sans démonstration, des résultats, plus ou moins classiques, qui ont été utilisés dans l'exposé précédent. La référence principale est le livre de Pinsker [31].

L'entropie d'une distribution de probabilité discrète (p_k) est définie par la formule de Boltzmann-Shannon :

$$H(p) = -\sum p_k \log p_k.$$

Grâce à la convexité de la fonction $-t \log t$ il est facile de voir que ce nombre est toujours positif, éventuellement $+\infty$, et n'est nul que si (p_k) est une distribution de Dirac. Si X est une v.a. de distribution (p_k) on dit que $H(p) = H(X)$ est l'entropie de X. Une permutation des p_k ne change pas $H(p)$ donc $H(X)$ ne dépend que de la distribution réarrangée de X.

Pour définir l'entropie d'une loi de probabilité ayant une densité φ sur $\mathbb{R}$ ou sur G, on introduit suivant Shannon ([34]) la quantité

$$\tilde{H}(\varphi) = -\int \varphi(x) \log \varphi(x)\, dx.$$

Malgré la similarité avec la formule précédente, cette quantité n'est pas en général positive ; elle peut prendre toute valeur de $-\infty$ à $+\infty$ inclus ou même ne pas être définie. De plus elle est modifiée par tout changement de variable de déterminant jacobien différent de 1. Pour ces raisons quand l'intégrale $-\int \varphi(x) \log \varphi(x)\, dx$ a un sens, on l'appelle *entropie différentielle* de la densité φ, et on la note $\tilde{H}(\varphi)$. Si φ est la loi d'une v.a. X, on dit que $\tilde{H}(\varphi) = \tilde{H}(X)$ est l'entropie différentielle de X. Cette quantité ne dépend que de la densité réarrangée de X.

Pour faire apparaître le lien entre l'entropie absolue et l'entropie différentielle et atteindre une généralité suffisante, il est nécessaire d'introduire la notion d'*entropie relative* suivante. Etant données deux probabilités P et Q sur un espace mesurable $(\Omega, \mathcal{B})$, l'entropie de

Q relative à P est définie par

$$H(Q\ ;\ P) = \sup_{\mathfrak{F}} \sum_{F\in\mathfrak{F}} (\log \frac{Q(F)}{P(F)})\ Q(F)$$

où la somme est prise sur l'ensemble des parties, de P-mesure positive, F d'une partition finie mesurable $\mathfrak{F}$ et où le sup. est pris sur l'ensemble de toutes ces partitions. C'est une quantité positive, éventuellement infinie, nulle seulement si P = Q. Le théorème suivant dû à Gelfand, Yaglom et Perez ([11] ; voir aussi [31]) est alors fondamental.

Théorème : *Si* $H(Q\ ;\ P) < \infty$ *alors* Q *est absolument continue par rapport à* P.

Si Q *est absolument continue par rapport à* P *alors*

$$H(Q\ ;\ P) = \int (\log \frac{dQ}{dP})\ dQ$$

(intégrale éventuellement infinie).

L'intégrale apparaissant ici est aussi appelée "information de Kullback". L'entropie relative $H(Q\ ;\ P)$ est donc infinie dans deux cas si Q n'est pas absolument continue par rapport à P, ou si l'intégrale $\int (\log \frac{dQ}{dP})\ dQ$ est infinie.

La notion d'*information mutuelle* de deux v.a. X et Y est déduite de celle d'entropie relative par la formule :

$$I(X,Y) = H(\lambda(X,Y)\ ;\ \lambda(X) \otimes \lambda(Y))$$

où $\lambda(X,Y)$ est la loi conjointe et $\lambda(X) \otimes \lambda(Y)$ la loi produit des marginales du couple (X,Y). Cette définition ne dépend que des tribus $\sigma(X)$ et $\sigma(Y)$ engendrées par les v.a. X et Y. Dans le cas où X et Y ne prennent qu'un nombre fini de valeurs on a :

$$I(X,Y) = H(X,Y) - H(X) - H(Y) = H(X) - H(X/Y)$$

formule d'usage constant dans la théorie de Kolmogorov-Sinai-Ornstein. Grâce à la concavité de la fonction log on démontre alors les propriétés suivantes ([31] chap II).

Propriétés de l'information mutuelle

1) $I(X,Y) \geqslant 0$
2) $I(X,Y) = 0$ si et seulement si X et Y sont indépendantes.
3) $I(X, f(Y)) \leqslant I(X,Y)$, f étant une fonction mesurable quelconque.

4) $\lim_n \uparrow I(X,(Y_1....Y_n)) = I(X, (Y_1....Y_n....))$.

5) s'il existe n tel que $I(X, (Y_n, Y_{n+1}....))$ soit fini,
$\lim_n \downarrow I(X, (Y_n, Y_{n+1},....)) = I(X, \mathcal{C})$ où $\mathcal{C} = \bigcap_n \sigma (Y_n, Y_{n+1}...))$.

Pour retrouver dans ce cadre la première formule de Shannon on observe que $I(X,X) < \infty$ si et seulement si X est discrète de distribution (p_k) vérifiant $-\sum p_k \log p_k < \infty$; alors $I(X,X) = -\sum p_k \log p_k$. On pose donc, en général :

$$H(X) = I(X,X).$$

On a l'inégalité $I(X,Y) \leqslant H(X)$.

Pour retrouver la seconde formule de Shannon, on considère deux densités de probabilité φ et Ψ , sur $\mathbb{R}$ ou G, avec $\varphi(x)/\Psi(x) < \infty$ p.p. Alors

$$H(\varphi;\Psi) = \int (\log \frac{\varphi(x)}{\Psi(x)}\ \varphi(x)\ dx = \int (\log \varphi(x))\ \varphi(x)\ dx - \int (\log \Psi(x))\ \varphi(x) dx.$$

La notion d'entropie ou information *conditionnelle* définie et étudiée par Dobrushin ([11] [31]) est essentielle. Etant données trois v.a. X, Y, Z, les deux premières prenant leurs valeurs dans un espace "séparable" afin que les versions régulières des probabilités conditionnelles $\lambda(X,Y/Z=z)$, $\lambda(X/Z=z)$, $\lambda(Y/Z=z)$ existent, on définit l'information mutuelle de X et Y sachant Z par :

$$EI(X,Y/Z) = \int I(X,Y/Z=z) d\ \lambda_Z(z).$$

($I(X,Y/Z=z)$ étant l'entropie de $\lambda(X,Y/Z=z)$ relative à $\lambda(X/Z=z) \otimes \lambda(Y/Z=z)$). On a alors l'importante formule

$$I((X,Z),Y) = I(Y,Z) + EI\ (X,Y/Z),$$

dite formule de Kolmogorov.

APPENDICE 2
SUR LES TRIBUS ASYMPOTIQUE ET INVARIANTE

On donne ici quelques indications supplémentaires sur les tribus asymptotique et invariante d'une marche aléatoire et la démonstration du théorème de la partie III.

Considérons $\Omega = G^{\mathbb{N}}$ muni de la tribu borélienne produit $\mathcal{B}^{\infty}$. Notons S_n les coordonnées canoniques. Munissons $(\Omega, \mathcal{B}^{\infty})$ de la probabilité P_e pour laquelle $(S_n)_n$ est une réalisation de la marche aléatoire, définie par μ, issue de e ; P_e est l'image de la probabilité produit $\mu^{\otimes \mathbb{N}}$ par l'application :

$$(x_1, \ldots, x_n \ldots) \longmapsto (s_1 = x_1,\ s_2 = x_1 x_2, \ldots,\ s_n = (x_1 \ldots x_n), \ldots).$$

C'est aussi la loi de la chaîne de Markov de transition

$$P(S_{n+1} \in B \ / \ S_n = x) = \mu(x^{-1}B)$$

et de distribution initiale δ_e. L'application décalage ϑ sur Ω est définie par

$$\vartheta(s_1, s_2, \ldots, s_n \ldots) = (s_2, s_3, \ldots, s_{n+1} \ldots).$$

C'est une application $\mathcal{B}^{\infty}$-mesurable, surjective sur Ω.

Le groupe G agit de façon mesurable sur Ω par

$$(x, (s_1, \ldots, s_n)) \longmapsto (xs_1, xs_2, \ldots, xs_n, \ldots).$$

Cette action permet de définir la "convolution" d'une mesure ρ sur G avec une mesure ν sur Ω:

$$\int_{\Omega} f(s_1, \ldots, s_n \ldots)\ d\ \rho * \nu\ (s_1, \ldots, s_n)$$
$$= \int_G d\rho(x) \int_{\Omega} f(xs_1, \ldots, xs_n, \ldots)\ d\nu(s_1, \ldots, s_n, \ldots).$$

On a alors la relation :

$$\vartheta P_e = \mu * P_e,$$

La tribu asymptotique de la marche aléatoire, notée $\mathcal{A}$, est définie par :

$$\mathcal{A} = \bigcap_{n \geq 1} \vartheta^{-n} \mathcal{B}^{\infty}.$$

La tribu invariante, notée $\mathcal{I}$, est définie par

$$\mathcal{I} = \{B \in \mathcal{B}^{\infty};\ \vartheta^{-1} B = B\}.$$

L'inclusion $\mathcal{I} \subset \mathcal{A}$ est évidente.

Proposition : *L'action de ϑ sur $\mathcal{A}$ est celle d'un automorphisme de tribu.*

Démonstration : La surjectivité de ϑ est essentielle. Par ϑ^{-1} on définit un endomorphisme surjectif de la tribu $\vartheta^{-n} \mathcal{B}^{\infty}$ sur $\vartheta^{-n-1} \mathcal{B}^{\infty}$. En passant à la limite on voit que ϑ^{-1} est un endomorphisme surjectif de $\mathcal{A}$. Comme ϑ est surjective, $\vartheta \vartheta^{-1} F = F$ pour tout $F \subset \Omega$, donc ϑ est inverse à gauche de l'endomorphisme surjectif ϑ^{-1} de $\mathcal{A}$; ϑ est aussi inverse à droite et cela prouve la proposition.

Théorème : *Quelle que soit μ, les tribus $\mathcal{A}$ et $\mathcal{I}$ sont égales* mod P_e *; c'est à dire que, pour tout $A \in \mathcal{A}$ il existe $J \in \mathcal{I}$ vérifiant*

$$A = J \quad P_e \text{ p.s.}$$

Démonstration : La démonstration se scinde en deux, suivant qu'il existe deux entiers n et n' tels que μ^n et $\mu^{n'}$ ne soient pas étrangères ou non.

Considérons le 1^{er} cas. Soit alors k le plus petit des entiers $j \geq 1$ tels qu'il existe $n \geq 0$ avec $\|\mu^{n+j} - \mu^n\| < 2$ ($\| \ \|$ note la variation totale des mesures ; μ^{n+j} et μ^n ne sont pas étrangères si et seulement si $\|\mu^{n+j} - \mu^n\| < 2$). Alors, d'après la loi "zéro ou deux" ([8] p. 120) $\lim_i \|\mu^{(i+1)k} - \mu^{ik}\| = 0$ et $\mathcal{A} = \mathcal{I}_k$ mod P_e où $\mathcal{I}_k = \{B \in \mathcal{B}^{\infty};\ \vartheta^{-k} B = B\}$ est la tribu invariante de ϑ^k. Si $k = 1$ c'est le résultat annoncé. Sinon on doit démontrer que $\mathcal{I}_k = \mathcal{I}$ mod P_e. L'inclusion $\mathcal{I} \subset \mathcal{I}_k$ est évidente. Pour alléger prenons $k=2$, le cas général étant similaire. Alors μ^{2i} et μ^{2j+1} sont étrangères pour tout i et j. Comme $\vartheta^{2i} P_e = \mu^{2i} * P_e$ est portée par tout ensemble de la forme $[S_1 \in C]$ où C est un borélien de G portant μ^{2i+1}, les mesures $\vartheta^{2i} P_e$ et $\vartheta^{2j+1} P_e$ sont étrangères pour tout i et j. Posons

$\xi = \sum_{i=0}^{\infty} \vartheta^{2i} P_e/2^{i+1}$ et $\rho = \vartheta\xi$; ξ et ρ sont étrangères sur Ω. Soient S et T deux ensembles de $\mathcal{B}^\infty$, disjoints, tels que $\xi(S) = \rho(T) = 1$. Les ensembles $S_1 = S \cap \bar{\vartheta}^1 T$ et $S_2 = \bar{\vartheta}^1 S_1$ sont disjoints ; le premier porte ξ et le second ρ. Soit $F \in \mathcal{J}_2$. Posons $\bar{F} = (F \cap S_1) \cup (\vartheta^1 F \cap S_2)$; alors $\bar{\vartheta}^1 \bar{F} = (\bar{\vartheta}^1 F \cap S_2) \cup (F \cap \bar{\vartheta}^1 S_2)$.

Comme $\vartheta^2 \xi << \xi$ on a $\xi(\bar{\vartheta}^2 S_1) = \xi(\bar{\vartheta}^1 S_2) = 1$ et $F \cap S_1 = F \cap \bar{\vartheta}^1 S_2$ mod ξ d'où $\bar{F} = \bar{\vartheta}^1 F$ mod $(\frac{\xi+\rho}{2})$. Comme $\vartheta(\frac{\xi+\rho}{2}) << \frac{\xi+\rho}{2}$, il existe $F' \in \mathcal{J}$ tel que $\bar{F} = F'$ mod $(\frac{\xi+\rho}{2})$ ([8] p. 114). On arrive ainsi à $F = \bar{F} = F'$ mod ξ donc aussi mod P_e.

Considérons le 2^d cas. Pour tout i et $j, i \neq j, \|\mu^i - \mu^j\| = 2$ donc $\vartheta^i P_e$ et $\vartheta^j P_e$ sont étrangères. Tout d'abord il est facile de construire une suite d'ensembles mesurables S_n, disjoints deux à deux, tels que $\vartheta^n P_e(S_n) = 1$. Posons $T_n^k = S_n \cap \bar{\vartheta}^1 S_{n+1} \cap \ldots \cap \bar{\vartheta}^k S_{n+k}$. Pout tout n et k, $\vartheta^n P_e(T_n^k) = 1$, donc $\vartheta^n P_e(T_n) = 1$ pour $T_n = \lim_k T_k^n$. On a aussi $T_n = S_n \cap \bar{\vartheta}^1 T_{n+1}$. Soit $F \in \mathcal{A}$. Posons $F_n = \vartheta^n F \cap T_n$ et $\bar{F} = \bigcup_{n \geq 0} F_n$; ces ensembles sont $\mathcal{B}^\infty$-mesurables, d'après la proposition. Par surjectivité de ϑ on a $\vartheta T_n \subset T_{n+1}$, donc $\vartheta \bar{F} \subset \bar{F}$ et $\bar{F} \subset \bar{\vartheta}^1 \vartheta \bar{F} \subset \bar{\vartheta}^1 \bar{F}$. La suite $\bar{\vartheta}^n \bar{F}$ est donc croissante avec pour limite $F' \in \mathcal{J}$. Comme ϑ est un automorphisme de $\mathcal{A}$ on a aussi, $F_n^c \cap T_n = \vartheta^n(F^c) \cap T_n$ d'où $\bar{\vartheta}^n \bar{F} \cap T_o = F \cap T_o$ et à la limite $F' \cap T_o = F \cap T_o$. Ceci prouve le résultat cherché : $F = F'$ P_e p.s.

BIBLIOGRAPHIE

[1] **AVEZ A.** (1972) - Entropie des groupes de type fini. C.R. Acad. Sc. Paris 275 A 1363-1366

[2] **AVEZ A.** (1974) - Théorème de Choquet-Deny pour les groupes à croissance non exponentielle. C.R. Acad. Sc. Paris 279 A, 25-28

[3] **AVEZ A.** (1976) - Croissance des groupes de type fini et fonctions harmoniques. L.N. in Math. Springer n° 532, 35-49.

[4] **AVEZ A.** (1976) - Harmonic functions on groups. Diff. Geom. and Relativity 27-32, Reidel (Holland)

[5] **AZENCOTT R.** (1970) - Espaces de Poisson des groupes localement compacts. L.N. in Math. Springer n°148

[6] **BROWN L.** (1982) - A proof of the central limit theorem motivated by the Carmer-Rao inequality. Statistics and Proba : essays in honor of C.R. Rao. North. Holland 141-148.

[7] **CSISZAR I.** (1964) - A note on limiting distributions on topological groups. Publ. Math. Inst. Hung. Acad. Sc. Vol 9, 595-598.

[8] **DERRIENNIC Y.** (1976) - Lois "zéro ou deux" pour les processus de Markov. Ann. Inst. H. Poincaré, Sect. B, 12, 111-129.

[9] **DERRIENNIC Y.** (1980) - Quelques applications du théorème ergodique sous-additif. Astérisque 74, 183-201

[10] **DERRIENNIC Y.** (1985) - Entropie et frontière d'une marche aléatoire : le cas général. Probabilités sur les structures géométriques. Publications de l'Université Paul Sabatier, Toulouse.

[11] **DOBRUSHIN R.L.** (1959) - General formulation of Shannon's basic theorems of the theory of information. Usp. Math. Nauk, 14, n5 6, 3-104. Traduction allemande : VEB Deutscher Verlag der Wissenschaften Berlin 1963.

[12] **DYNKIN E.B. et MALJUTOV M.B.** (1961) - Random walks on groups with a finite number of generators. Soviet Math. Dokl. 2, 399-402.

[13] **ELIE L.** (1978) - Fonctions harmoniques positives sur le groupe affine. L.N. in Math. Springer n° 706 96-110.

[14] **FELLER W.** (1966) - An introduction to probability theory Vol II (Wiley)

[15] **FOURT G.** (1972) - Existence de mesures à puissances singulières à toutes leurs translatées. C.R. Acad. Sc. Paris, 274 A, 648-650

[16] **FÜRSTENBERG H.** (1963) - Non communicating random products. T.A.M.S. 108, 377-428.

[17] **FÜRSTENBERG H.** (1971) - Random walks and discrete subgroups of Lie groups. Adv. Proba. Vol 1, 3-63 (Dekker).

[18] **GREENLEAF F.P.** (1969) - Invariant means on topological groups. Van Nostrand.

[19] **GRIGORCHUK R.** (1983) - On Milnor's problem of group growth. Soviet Math. Dokl. 28, n° 1, 23-26.

[20] **GROMOV M.** (1981) - Groups of polynomial growth and expanding maps. Publ. Math. IHES, 53, 53-78.

[21] **GUIVARC'H Y.** (1973) - Croissance polynomiale et période des fonctions harmoniques. Bull. SMF, 101, 333-379.

[22] **GUIVARC'H Y.** (1980) - Sur la loi des grands nombres. Astérisque, 74, 47-98.

[23] **GUIVARC'H Y.** (1980) - Quelques propriétés asymptotiques des produits de matrices aléatoires. L.N. in Math, Springer, n° 774, 177-250.

[24] **KAIMANOVICH V.A. et VERSHIK A.M.** (1979) - Random walks on groups : boundary, entropy, uniform distribution. Soviet Math. Dokl. 20, 1170-1173.

[25] **KAIMANOVICH V.A. et VERSHIK A.M.** (1983) - Random walks on discrete groups : boundary and entropy. The annals of Proba. Vol 11, n° 3, 457-490.

[26] **KAIMANOVICH V.A.** (1982) - The differential entropy of the boundary of a random walk on a group. Russian Math. Surveys

[27] KAIMANOVICH V.A. (1985) - An entropy criterion for maximality of the boundary of random walks on discrete groups. Soviet Math. Dokl. Vol 31, n° 1, 193-197.

[28] LEDRAPPIER F. (1983) - Une relation entre entropie, dimension et exposant pour certaines marches aléatoires. C.R. Acad. Sc. Paris 296 A, 369-372.

[29] LEDRAPPIER F. (1985) - Poisson boundaries of discrete groups of matrices. Israel J. of Math. Vol 50, n° 4, 319-336.

[30] LINNIK Yu. V. (1959) - An information-theoretic proof of the central limit theorem with the Lindeberg condition. Theory of Probability and its applications. Vol IV, n° 3, 288-299.

[31] PINSKER M.S. (1964) - Information and information stability of random variables and processes. Holden-Day.

[32] RENYI A. (1967) - Calcul des probabilités, DUNOD.

[33] ROSENBLATT J. (1981) - Ergodic and mixing random walks on locally compact groups. Math. Ann. 257, 31-42.

[34] SHANNON C.E. (1949) - The mathematical theory of communication. The University of Illinois Press.

[35] VAROPOULOS N.T. (1986) - Théorie du potentiel sur des groupes et des variétés. C.R. Acad. Sc. Paris t. 302, série I, n° 6, 203-205. (et les références du même auteur qui y sont indiquées).

[36] BARRON A.R. (1986) - Entropy and the central limit theorem. The Annals of Probability Vol 14, n° 1, 336-342.

Yves DERRIENNIC

Faculté des Sciences

29287 BREST CEDEX

FRANCE