

**MÉTHODES ENTROPIQUES
POUR LES CONVOLUTIONS DE BERNOULLI
[d'après Hochman, Shmerkin, Breuillard, Varjú]**

par **Sébastien GOUËZEL**

1. INTRODUCTION

L'exemple le plus classique d'ensemble autosimilaire dans \mathbb{R} , l'ensemble triadique de Cantor $C_{1/3}$, peut être décrit comme suit. Pour $\lambda = 1/3$, considérons l'ensemble $\Phi = \Phi_\lambda = \{x \mapsto \lambda x, x \mapsto 1 + \lambda(x - 1)\}$ des deux contractions affines de rapport λ et de points fixes respectifs 0 et 1. Alors l'unique compact non vide de \mathbb{R} invariant par les éléments de Φ est $C_{1/3}$. Cette description se prête à une reformulation dynamique : partons d'un compact non vide K_0 de \mathbb{R} , par exemple $K_0 = [0, 1]$, et définissons une suite par $K_{n+1} = \Phi(K_n) = \bigcup_{\varphi \in \Phi} \varphi(K_n)$. Alors cette suite converge (en topologie de Hausdorff) vers $C_{1/3}$, quel que soit le compact de départ K_0 , puisque l'application $K \mapsto \Phi(K)$ est contractante d'un facteur λ pour la distance de Hausdorff et a $C_{1/3}$ pour point fixe. Pour $K_0 = [0, 1]$, alors $K_1 = [0, 1/3] \cup [1/3, 2/3]$, et K_n correspond à la n -ième étape dans la description géométrique habituelle de $C_{1/3}$, où l'on retire le tiers central des intervalles conservés à l'étape précédente.

Cette construction fonctionne en fait pour tout $\lambda \in]0, 1[$. Lorsque $\lambda \in]0, 1/2[$, l'ensemble C_λ obtenu est encore un ensemble de Cantor, tout à fait analogue à $C_{1/3}$. En revanche, si $\lambda \in [1/2, 1[$, on a $C_\lambda = [0, 1]$. En effet, si l'on part de $K_0 = [0, 1]$, alors $K_1 = [0, \lambda] \cup [1 - \lambda, 1]$ coïncide avec K_0 (il y a même un recouvrement). Il en va donc de même de tous les K_n , et de leur limite.

La situation devient plus intéressante si l'on remplace les ensembles par des mesures de probabilité. Partons d'une mesure de probabilité arbitraire κ_0 à support compact, et définissons $\kappa_{n+1} = \frac{1}{2} \sum_{\varphi \in \Phi_\lambda} \varphi_*(\kappa_n)$. En notant K_0 le support de κ_0 , alors le support de κ_n est l'ensemble K_n défini ci-dessus. De plus, comme l'application $\kappa \mapsto \frac{1}{2} \sum_{\varphi \in \Phi_\lambda} \varphi_*(\kappa)$ est contractante de rapport λ pour une distance adaptée sur les mesures de probabilité à support compact (la distance de Wasserstein W_1), la suite κ_n converge vers une mesure limite $\bar{\mu}_\lambda$, de support égal à C_λ et indépendante du choix de κ_0 . C'est l'unique mesure de probabilité sur \mathbb{R} qui est *stationnaire* pour l'action de Φ_λ : elle est caractérisée par l'égalité $\bar{\mu}_\lambda = \frac{1}{2} \sum_{\varphi \in \Phi_\lambda} \varphi_*(\bar{\mu}_\lambda)$. Lorsque $\lambda < 1/2$, c'est la mesure uniforme classique sur l'ensemble de Cantor C_λ , donnant une mesure 2^{-n} à chacun des 2^n intervalles de n -ième génération dans la construction géométrique de C_λ (c'est aussi la mesure de Hausdorff de

dimension $\log 2/|\log \lambda|$ sur C_λ). Pour $\lambda = 1/2$, c'est exactement la mesure de Lebesgue sur $[0, 1]$ puisque, si l'on part de $\kappa_0 = \text{Leb}_{[0,1]}$, alors toutes les mesures κ_n coïncident avec $\text{Leb}_{[0,1]}$. En revanche, la situation est plus complexe pour $\lambda \in]1/2, 1[$: le recouvrement dans la décomposition $K_1 = [0, \lambda] \cup [1 - \lambda, 1]$ rend la mesure κ_1 inhomogène, donnant proportionnellement plus de poids à l'intervalle central $[1 - \lambda, \lambda]$ qu'aux deux bouts. Cette inhomogénéité se transfère ensuite, par itération, à toutes les échelles.

Le problème central auquel cet exposé est consacré est l'exploration de la régularité de la mesure $\bar{\mu}_\lambda$, et en particulier des paramètres $\lambda \in [1/2, 1[$ pour lesquels $\bar{\mu}_\lambda$ est absolument continue (par rapport à la mesure de Lebesgue) : les seuls paramètres pour lesquels on sait que $\bar{\mu}_\lambda$ n'est pas absolument continue sont les inverses des nombres de Pisot, et il est bien possible que ce soit les seuls, même si ce problème est encore ouvert. Les résultats récents que l'on décrira plus bas indiquent que les paramètres exceptionnels λ pour lesquels $\bar{\mu}_\lambda$ n'est pas absolument continue forment un petit ensemble, en un sens que l'on précisera.

En sus de leur description géométrique, les mesures $\bar{\mu}_\lambda$ admettent une description équivalente plus probabiliste, et souvent plus pratique pour les démonstrations, qui explique le nom de *convolutions de Bernoulli* qu'on leur donne. Soit $\xi_0, \xi_1, \xi_2, \dots$ une suite de variables aléatoires indépendantes identiquement distribuées de type Bernoulli, i.e., qui valent 1 ou -1 avec probabilité $1/2$ (on peut prendre par exemple les fonctions coordonnées sur l'espace $\Omega = \{-1, 1\}^{\mathbb{N}}$ muni de la mesure $\mathbb{P} = ((\delta_{-1} + \delta_1)/2)^{\otimes \mathbb{N}}$). Formons la somme $X_\infty = \sum_{n \geq 0} \xi_n \lambda^n$, c'est une série aléatoire convergente dont la loi μ_λ est une mesure de probabilité sur \mathbb{R} , donnée par

$$\mu_\lambda(A) = \mathbb{P}\left(\sum_{n \geq 0} \xi_n \lambda^n \in A\right).$$

On peut aussi écrire $\mu_\lambda = (X_\infty)_* \mathbb{P}$ si l'on souhaite éviter complètement le langage probabiliste. Comme la loi d'une somme de variables indépendantes est la convolution des lois de ces variables, la définition précédente permet aussi d'écrire μ_λ comme un produit de convolution infini de lois de Bernoulli :

$$(1.1) \quad \mu_\lambda = \bigstar_{n \geq 0} \frac{\delta_{-\lambda^n} + \delta_{\lambda^n}}{2}.$$

En isolant la variable ξ_0 (qui vaut -1 ou 1 avec probabilité $1/2$) et en notant que la loi de $\sum_{n \geq 1} \xi_n \lambda^n$ coïncide avec μ_λ à une contraction de facteur λ près, on obtient l'égalité

$$(1.2) \quad \mu_\lambda = \frac{1}{2} \left((x \mapsto -1 + \lambda x)_* \mu_\lambda + (x \mapsto 1 + \lambda x)_* \mu_\lambda \right).$$

C'est la même équation que celle satisfaite par $\bar{\mu}_\lambda$, à un changement affine de coordonnées près (envoyant $-1/(1-\lambda)$ sur 0 et $1/(1-\lambda)$ sur 1). Ce changement de coordonnées envoie donc μ_λ sur $\bar{\mu}_\lambda$ (il envoie aussi la loi de $X_N = \sum_{n < N} \xi_n \lambda^n$ sur la mesure κ_N issue par le procédé décrit plus haut de $\kappa_0 = \delta_{1/2}$), et les propriétés de régularité des mesures sont exactement les mêmes. Dans la suite, on travaillera uniquement avec μ_λ , mais le lecteur est invité à garder en tête le point de vue géométrique expliqué plus haut.

La question de la régularité de cette mesure peut sembler anecdotique. Il faut cependant y penser comme à un modèle extrêmement basique pour toutes les situations où de l'autosimilarité est présente (comme on en rencontre souvent en systèmes dynamiques, en théorie géométrique de la mesure, en probabilités) : les outils sophistiqués qui ont été développés pour étudier les convolutions de Bernoulli ont tous trouvé des applications dans des domaines beaucoup plus larges (au prix de complications techniques). C'est donc une sorte de terrain de jeux, où tester de nouvelles idées et de nouvelles techniques dans un contexte aussi simplifié que possible (mais de nombreuses questions fondamentales y restent encore ouvertes !)

Il est facile de voir que l'ensemble $\mathcal{E}_\perp \subseteq]1/2, 1[$ des paramètres exceptionnels pour lesquels μ_λ n'est pas absolument continue par rapport à la mesure de Lebesgue contient les inverses des nombres de Pisot (pour ces paramètres, on sait même que μ_λ est alors de dimension < 1 , voir le théorème 3.5), mais on ne connaît pas d'autre élément de \mathcal{E}_\perp . La question fondamentale sur les convolutions de Bernoulli est de comprendre cet ensemble. En ordonnant les énoncés par force, on peut se poser les questions suivantes :

1. Est-il de mesure de Lebesgue nulle ?
2. Est-il de dimension de Hausdorff nulle ?
3. Est-il dénombrable ?
4. Est-il uniquement composé de nombres algébriques ?
5. Est-il égal à l'ensemble des inverses des nombres de Pisot dans $]1/2, 1[$?

On peut aussi se poser les mêmes questions sur l'ensemble exceptionnel $\mathcal{E}_{<1}$ (plus petit *a priori*) des paramètres exceptionnels pour lesquels la dimension $\text{Dim}(\mu_\lambda)$ de la mesure μ_λ , définie dans le théorème 2.2, vérifie $\text{Dim}(\mu_\lambda) < 1$.

Ce texte est consacré aux résultats impressionnants démontrés récemment sur les convolutions de Bernoulli par Hochman, Shmerkin, Breuillard et Varjú [Hoc14, Shm14, Shm16, Var16, BV15, BV16] suite au travail fondateur de Hochman [Hoc14] : il a montré comment des méthodes de combinatoire additive pouvaient être utilisées pour démontrer un énoncé sur la croissance de l'entropie des mesures par convolution, et pourquoi un tel énoncé était pertinent dans l'étude des convolutions de Bernoulli. Suite à ces résultats, on sait que \mathcal{E}_\perp est de dimension de Hausdorff nulle (théorème 2.8), et très probablement que $\mathcal{E}_{<1}$ est uniquement composé de nombres algébriques (voir la conjecture 2.6, la proposition 2.7 et le théorème 4.15). Les questions 3, 4, 5 ci-dessus sont toutefois toujours ouvertes.

Cet article est organisé comme suit. La partie 2 contient les énoncés des théorèmes sur les convolutions de Bernoulli dont on souhaite évoquer les preuves par la suite. Dans la partie 3, on démontre complètement les résultats (élémentaires) sur les mesures μ_λ lorsque λ est l'inverse d'un nombre de Pisot. C'est l'occasion d'introduire l'entropie, et de se familiariser avec les problèmes qui se poseront pour les paramètres généraux. La partie 4 étudie les paramètres généraux avec un point de vue entropique. On y énonce

le théorème de Hochman sur l’entropie des convolutions de Bernoulli (théorème 4.11) et on montre comment il implique plusieurs des résultats de la partie 2. Enfin, la partie 5 évoque la preuve du théorème de Hochman, en lien avec la croissance de l’entropie des mesures par convolution, et discute des extensions et raffinements de ces résultats.

2. RÉSULTATS CONNUS SUR LES CONVOLUTIONS DE BERNOULLI

Ce paragraphe est consacré à un exposé à la Prévert des propriétés connues des convolutions de Bernoulli pour les paramètres $\lambda \in [1/2, 1[$, le cas $\lambda < 1/2$ étant complètement explicite. L’ordre d’exposition est purement thématique, il ne respecte pas du tout la chronologie dans laquelle les énoncés ont été démontrés, ni leur difficulté. On reviendra plus loin sur la plupart de ces énoncés, une fois qu’on aura mis en place les outils permettant d’expliquer leurs preuves. Pour faciliter leur localisation, la numérotation d’un énoncé donnée ici correspond à l’endroit de l’article où il réapparaîtra.

Résultats pour tous les paramètres

PROPOSITION 2.1 (Jessen et Wintner [JW35]). — *Pour $\lambda \in [1/2, 1[$, la mesure μ_λ est soit absolument continue par rapport à la mesure de Lebesgue, soit totalement singulière par rapport à celle-ci.*

Démonstration. — Si μ_λ a une composante absolument continue ν non nulle, celle-ci vérifie également l’égalité de stationnarité (1.2). Comme l’application $\kappa \mapsto \frac{1}{2} \sum_{\varphi \in \Phi_\lambda} \varphi_*(\kappa_n)$ est contractante de rapport λ pour la distance de Wasserstein W_1 , cette équation (1.2) admet une unique solution parmi les mesures de probabilité. Ainsi, ν est un multiple (non nul) de μ_λ , ce qui montre que μ_λ elle-même est absolument continue. \square

THÉORÈME 2.2 (Ledrappier [Led92], Feng et Hu [FH09]). — *Soit $\lambda \in [1/2, 1[$. Il existe alors un réel $\alpha(\lambda) \in]0, 1]$ tel que, pour μ_λ -presque tout x ,*

$$\frac{\log \mu_\lambda([x - r, x + r])}{\log r} \xrightarrow{r \rightarrow 0} \alpha(\lambda).$$

On dit que μ_λ est dimensionnellement exacte, de dimension $\alpha(\lambda)$, et on note $\text{Dim}(\mu_\lambda) = \alpha(\lambda)$.

Cet énoncé est la conséquence d’un résultat beaucoup plus général en systèmes dynamiques, dû à Ledrappier et Young, qui affirme que toute mesure invariante par une application régulière sur une variété compacte est dimensionnellement exacte le long de sous-variétés adaptées à la dynamique. Ce théorème est démontré par des techniques entropiques, mais complètement différentes de celles dont il sera question dans cet article. Nous l’admettrons donc comme une boîte noire.

Pour clarifier le contenu de ce théorème, introduisons la notation suivante : on écrira

$$u_n \sim_{\log} v_n$$

lorsque $\log u_n / \log v_n \rightarrow 1$. Le théorème précédent affirme alors que $\mu_\lambda([x - r, x + r]) \sim_{\log} r^{\alpha(\lambda)}$ (les fluctuations étant en $r^{o(1)}$). Dans \mathbb{R}^d , la mesure de Lebesgue des boules se comporte exactement en $C(d)r^d$. Cela justifie qu'on considère $\alpha(\lambda)$ comme une dimension.

Remarque 2.3. — Si μ_λ est absolument continue par rapport à la mesure de Lebesgue, alors $\text{Dim}(\mu_\lambda) = 1$, mais la réciproque est a priori fautive. Il est donc plus faible (et plus facile) de démontrer une assertion $\text{Dim}(\mu_\lambda) = 1$ que de démontrer l'absolue continuité de μ_λ , même si c'est déjà un résultat fort.

L'intérêt et la difficulté du problème viennent du fait que $\text{Dim}(\mu_\lambda)$ ne dépend pas continûment de λ . Néanmoins, cette fonction est semi-continue inférieurement :

PROPOSITION 4.10. — *La fonction $\lambda \mapsto \text{Dim}(\mu_\lambda)$ est semi-continue inférieurement : si $\lambda_n \rightarrow \lambda$, alors $\liminf \text{Dim}(\mu_{\lambda_n}) \geq \text{Dim}(\mu_\lambda)$.*

Il se peut donc que $\text{Dim}(\mu_\lambda) = 1$ pour une grande majorité de paramètres (dense), avec au milieu de ceux-ci quelques paramètres où $\text{Dim}(\mu_\lambda) < 1$. C'est précisément le comportement auquel on s'attend.

On peut minorer la dimension de μ_λ pour *tous* les paramètres, en utilisant les résultats profonds qui suivent et une preuve numérique assistée par ordinateur :

PROPOSITION 6.1 (Hare et Sidorov [HS16]). — *Pour tout $\lambda \in [1/2, 1[$, on a $\text{Dim}(\mu_\lambda) \geq 0,82$.*

Résultats pour les paramètres algébriques

DÉFINITION 2.4. — *Un nombre de Pisot est un entier algébrique β de module > 1 dont tous les conjugués (distincts de β) sont de module < 1 .*

Pensons par exemple au nombre d'or $(1 + \sqrt{5})/2$ ou à la racine maximale de $x^3 - x - 1$ (ce dernier est le plus petit nombre de Pisot). De manière surprenante, les nombres de Pisot forment un sous-ensemble fermé de \mathbb{R} . Le nombre d'or est le plus petit point d'accumulation des nombres de Pisot [BDGGH⁺92]. Par conséquent, il y a une infinité de nombres de Pisot dans $]1, 2[$.

THÉORÈME 3.5 (Erdős [Erd39], Garsia [Gar63]). — *Soit $\lambda \in]1/2, 1[$ tel que λ^{-1} soit un nombre de Pisot. Alors μ_λ est singulière par rapport à la mesure de Lebesgue, et on a même $\text{Dim}(\mu_\lambda) < 1$.*

Conjecturalement, les inverses des nombres de Pisot sont les seuls paramètres pour lesquels μ_λ est singulière par rapport à la mesure de Lebesgue.

Un cas particulier d'un théorème de Hochman (énoncé plus loin comme le théorème 4.12) montre que, pour beaucoup de nombres algébriques, la dimension de la mesure μ_λ est maximale :

PROPOSITION 4.13 (Hochman [Hoc14]). — *Soit $\lambda \in [1/2, 1[$ algébrique sur \mathbb{Q} , mais qui n'est pas racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$. Alors $\text{Dim}(\mu_\lambda) = 1$.*

Ce théorème s'applique par exemple à tout λ rationnel. Avec le théorème 3.5, ceci implique qu'un inverse dans $]1/2, 1[$ d'un nombre de Pisot est racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$. On donnera une démonstration directe de ce fait plus loin, après le lemme 3.3.

Les convolutions de Bernoulli font apparaître les sommes $\sum_{n < N} \xi_n \lambda^n$ avec $\xi_n \in \{-1, 1\}$. Lorsque λ est algébrique, c'est le nombre de coïncidences entre ces sommes qui va dicter la dimension de la mesure μ_λ (voir le théorème 4.12 pour un énoncé précis). Sous les hypothèses de la proposition 4.13, ces sommes sont toutes distinctes, et la dimension est alors maximale. Lorsque λ est racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$, Breuillard et Varjú ont réussi à quantifier la fréquence de ces coïncidences en termes de la *mesure de Mahler* de λ . Lorsque λ est un nombre algébrique, de polynôme minimal $a(X - z_1) \cdots (X - z_d) \in \mathbb{Z}[X]$, on définit la mesure de Mahler M_λ comme

$$(2.1) \quad M_\lambda = |a| \prod_j \max(|z_j|, 1) \geq 1.$$

On rappelle la conjecture de Lehmer sur les mesures de Mahler :

CONJECTURE 2.5 (Lehmer). — *Il existe $c > 1$ tel que la mesure de Mahler de tout nombre algébrique est soit égale à 1, soit supérieure ou égale à c .*

Plus précisément, il est même conjecturé qu'on peut prendre $c = 1,1762\dots$, correspondant à λ racine du polynôme $X^{10} + X^9 - X^7 - X^6 - X^5 - X^4 - X^3 + X + 1$.

La mesure de Mahler des nombres algébriques peut être utilisée pour contrôler la dimension de la convolution de Bernoulli correspondante :

THÉORÈME 4.14 (Breuillard-Varjú [BV15]). — *Il existe c_0 avec la propriété suivante ($c_0 = 0,44$ fonctionne). Soit $\lambda \in [1/2, 1[$ un nombre algébrique sur \mathbb{Q} tel que*

$$\lambda > \min(2, M_\lambda)^{-c_0}.$$

Alors $\text{Dim}(\mu_\lambda) = 1$.

Si la conjecture de Lehmer est vraie, on en déduit que, pour tout λ algébrique assez proche de 1, alors $\text{Dim}(\mu_\lambda) = 1$ puisque la condition du théorème est satisfaite.

Des raffinements supplémentaires permettent même de garantir l'absolue continuité de la mesure, sous des conditions beaucoup plus restrictives.

THÉORÈME 5.6 (Varjú [Var16]). — *Pour tout $\varepsilon > 0$, il existe $c(\varepsilon) > 0$ avec la propriété suivante. Soit $\lambda \in [1/2, 1[$ un nombre algébrique sur \mathbb{Q} tel que*

$$\lambda > 1 - c(\varepsilon) \min(\log M_\lambda, (\log M_\lambda)^{-1-\varepsilon}).$$

Alors μ_λ est absolument continue.

Toutes les constantes dans l'article de Varjú peuvent être explicitées. Cela lui permet d'obtenir l'énoncé suivant : si a et b sont des entiers positifs premiers entre eux, alors $\mu_{1-a/b}$ est absolument continue dès que

$$a < 10^{-37} \frac{b}{\log(b+1)(\log \log(b+2))^3}.$$

Résultats pour les paramètres génériques

Considérons l'ensemble des polynômes de degré au plus n à coefficients dans $\{-1, 0, 1\}$ (de cardinal 3^{n+1}), et notons \mathcal{Z}_n l'ensemble des racines (complexes) de ces polynômes. C'est un ensemble de cardinal majoré par $n3^{n+1}$. On définit l'ensemble de Hochman $\mathcal{H} \subseteq [1/2, 1[$ comme l'ensemble des paramètres approchés superexponentiellement par des éléments de \mathcal{Z}_n . Formellement, pour $A > 0$ et $N_0 > 0$, posons

$$\mathcal{H}_{A, N_0} = \{\lambda \in [1/2, 1[: \forall n \geq N_0, d(\lambda, \mathcal{Z}_n) \leq e^{-An}\},$$

où d désigne la distance euclidienne dans \mathbb{C} . Alors \mathcal{H} est défini par

$$(2.2) \quad \mathcal{H} = \bigcap_{A>0} \bigcup_{N_0} \mathcal{H}_{A, N_0}.$$

Cet ensemble est extrêmement mince. En effet, si l'on fixe $A > \log 3$ et N_0 , alors pour tout $n \geq N_0$ l'ensemble \mathcal{H}_{A, N_0} est contenu dans $n3^{n+1}$ intervalles de taille $2e^{-An}$. Comme $n3^{n+1} \cdot 2e^{-An} \rightarrow 0$, l'ensemble \mathcal{H}_{A, N_0} est donc de mesure nulle, mais on a plus. On dit qu'un ensemble est de dimension de boîte (ou de Minkowski) au plus d si, pour tout ε , pour tout r assez petit, on peut le recouvrir par $r^{-d-\varepsilon}$ intervalles de longueur r . Alors \mathcal{H}_{A, N_0} est de dimension de Minkowski au plus $(\log 3)/A$. La dimension de Minkowski n'est pas stable par réunion dénombrable croissante, mais c'est le cas d'une notion de dimension un tout petit peu plus faible appelée dimension de packing (voir par exemple [Fal90] pour ces notions de dimensions). Ainsi, $\bigcup_{N_0} \mathcal{H}_{A, N_0}$ a dimension de packing au plus $(\log 3)/A$. Finalement, \mathcal{H} a dimension de packing nulle. Ceci implique par exemple que \mathcal{H} a dimension de Hausdorff nulle, mais c'est plus fort.

Si l'on tire uniformément $n3^{n+1}$ points indépendants dans le disque de rayon 2 (qui contient tous les points de \mathcal{Z}_n), alors avec très grande probabilité la distance minimale entre deux points est minorée par 10^{-n} . On s'attend naïvement à ce que le comportement de l'ensemble non aléatoire \mathcal{Z}_n soit analogue. En particulier, la conjecture suivante (qu'Hochman pose comme question dans [Hoc14]) est très naturelle :

CONJECTURE 2.6. — *Il existe $M > 0$ tel que la distance entre deux points distincts de \mathcal{Z}_n est minorée par M^{-n} , pour n assez grand.*

PROPOSITION 2.7. — *Si la conjecture 2.6 est vraie, alors \mathcal{H} est inclus dans $\bigcup_n \mathcal{Z}_n$, il est donc dénombrable et composé de nombres algébriques.*

Démonstration. — Si la conjecture est vraie, on peut fixer $M = e^A > 0$ tel que la distance entre deux éléments distincts de \mathcal{Z}_n soit au moins M^{-n} pour tout $n \geq n_0$. Soit $x \in \mathcal{H}_{2A, N_0}$. Pour tout $n \geq N_0$, il existe alors $z_n \in \mathcal{Z}_n$ tel que $|x - z_n| \leq e^{-2An}$. Pour $n > N_0$, les points z_n et z_{n-1} sont deux éléments de \mathcal{Z}_n , à distance au plus $e^{-2An} + e^{-2A(n-1)} = (1 + e^{-2A})e^{-2An}$, qui est $< e^{-An}$ si n est assez grand. Ceci implique que $z_n = z_{n-1}$ au vu de la séparation minimale entre les éléments distincts de \mathcal{Z}_n . Ainsi, la suite z_n stationne, en un point z appartenant à $\bigcup_n \mathcal{Z}_n$. Comme elle converge vers x , on obtient $x \in \bigcup_n \mathcal{Z}_n$. \square

La meilleure borne démontrée sur la séparation des éléments de \mathcal{Z}_n est un petit peu plus faible qu’une borne exponentielle : deux points distincts de \mathcal{Z}_n sont à distance au moins $2e^{-4n \log n}$ d’après [Mah64] (voir [BV16, Theorem 21]).

Le théorème suivant, dû à Hochman, montre que les paramètres pour lesquels la dimension de la convolution de Bernoulli μ_λ n’est pas maximale appartiennent nécessairement à l’ensemble de Hochman \mathcal{H} . Ainsi, ces paramètres sont extrêmement rares, et très probablement tous algébriques d’après la proposition 2.7 – cela justifie de s’intéresser tout particulièrement aux paramètres algébriques comme nous l’avons fait plus haut.

THÉORÈME 4.15 (Hochman [Hoc14]). — *Soit λ tel que $\text{Dim}(\mu_\lambda) < 1$. Alors $\lambda \in \mathcal{H}$.*

En particulier, les paramètres pour lesquels $\text{Dim}(\mu_\lambda) < 1$ ont dimension de packing nulle et donc aussi dimension de Hausdorff nulle.

Cet énoncé peut être reformulé ainsi : si $\text{Dim}(\mu_\lambda) < 1$, alors pour tout $A > 0$ il existe pour tout n assez grand $\lambda_n \in \mathcal{Z}_n$ avec $|\lambda - \lambda_n| \leq e^{-An}$. Ainsi, λ est extrêmement bien approché par des éléments de \mathcal{Z}_n (qui sont en particulier des entiers algébriques).

Lorsque $\text{Dim}(\mu_\lambda) < 1$, le théorème 4.15 donne pour tout n assez grand une bonne approximation λ_n de λ dans \mathcal{Z}_n . Cependant, il ne garantit pas que λ_n soit réel, et même si λ_n est réel il ne dit rien sur les propriétés de la mesure de Bernoulli μ_{λ_n} . Avec des techniques plus fines, Breuillard et Varjú montrent que, pour une infinité de valeurs de n , on peut garantir que λ_n soit réel et que μ_{λ_n} ait elle-même une dimension < 1 . De plus, ils obtiennent une meilleure vitesse de convergence de λ_n vers λ :

THÉORÈME 5.7 (Breuillard-Varjú [BV16]). — *Soit $\lambda \in [1/2, 1[$ tel que $\text{Dim}(\mu_\lambda) < 1$. Il existe alors une suite $n_i \rightarrow \infty$ et une suite d’approximations $\lambda_{n_i} \in \mathcal{Z}_{n_i} \cap [1/2, 1[$ telles que $|\lambda - \lambda_{n_i}| \leq e^{-n_i^{\log \log \log n_i}}$ et $\text{Dim}(\mu_{\lambda_{n_i}}) \rightarrow \text{Dim}(\mu_\lambda)$.*

En particulier, on déduit de cet énoncé que, pour tout intervalle ouvert $I \subseteq [1/2, 1[$,

$$(2.3) \quad \{\lambda \in I : \text{Dim}(\mu_\lambda) < 1\} \subseteq \overline{\{\lambda \in \overline{\mathbb{Q}} \cap I : \text{Dim}(\mu_\lambda) < 1\}}.$$

Ce résultat permet de déduire, à partir des résultats démontrés pour les paramètres algébriques, des résultats qui sont valables pour tous les paramètres. Par exemple, si

la conjecture de Lehmer est vraie, on obtient directement que tous les paramètres λ assez proches de 1 (algébriques ou non) vérifient $\text{Dim}(\mu_\lambda) = 1$, puisque le théorème 4.14 donne cette assertion pour les nombres algébriques et que l'inclusion (2.3) permet de l'étendre à tous les paramètres.

De même, si on savait que les seuls paramètres algébriques avec $\text{Dim}(\mu_\lambda) < 1$ étaient les inverses des nombres de Pisot, alors on en déduirait immédiatement que les seuls paramètres (algébriques ou non) avec $\text{Dim}(\mu_\lambda) < 1$ seraient les inverses des nombres de Pisot, puisque l'ensemble des nombres de Pisot est fermé.

Le théorème 5.7 implique aussi des résultats inconditionnels qui reposent sur des résultats quantitatifs de transcendance : si on sait qu'un paramètre λ n'est pas approché suffisamment rapidement par des nombres algébriques, on en déduit que $\text{Dim}(\mu_\lambda) = 1$. C'est par exemple vrai pour $\lambda = \log 2$, $e^{-1/2}$ ou $\pi/4$ [Var16].

Les méthodes de preuve de Hochman ne permettent pas directement de montrer l'absolue continuité des mesures pour les paramètres hors de \mathcal{H} . Cependant, en combinant ces arguments avec des techniques d'analyse de Fourier (moins explicites, et qui en particulier ne permettent pas de décrire l'ensemble de paramètres exceptionnels), Shmerkin a pu aller plus loin :

THÉORÈME 2.8 (Shmerkin [Shm14, Shm16]). — *Il existe un ensemble (non explicite) $\mathcal{S} \subseteq [1/2, 1[$ de dimension de Hausdorff nulle tel que, pour $\lambda \notin \mathcal{S}$, la mesure μ_λ est absolument continue par rapport à la mesure de Lebesgue, et de plus sa densité est dans L^p pour tout $p \in [1, \infty[$.*

Ainsi, même pour le problème de l'absolue continuité, l'ensemble des paramètres exceptionnels est extrêmement petit (toutefois dans un sens moins fort que plus haut, car on parle ici de dimension de Hausdorff et non de dimension de packing). Il y a en fait deux énoncés dans ce théorème : d'une part l'absolue continuité pour la plupart des paramètres [Shm14], et d'autre part le fait que la densité soit dans L^p pour tout p [Shm16]. Le second point est nettement plus délicat que le premier, et fait appel à des généralisations profondes des résultats de Hochman dont nous ne parlerons pas ici (disons seulement qu'il s'agit de remplacer l'entropie par d'autres quantités analogues mais reliées à la norme L^p dans tous les arguments).

Esquissons rapidement la preuve de l'absolue continuité dans le théorème de Shmerkin, car les arguments n'ont rien à voir avec les techniques entropiques qu'on développera dans la suite.

Esquisse de démonstration. — Le point clé de la preuve est la remarque (élémentaire) suivante : si une mesure μ s'écrit comme $\mu^{(1)} * \mu^{(2)}$ où $\mu^{(1)}$ a des coefficients de Fourier qui décroissent polynomialement à l'infini et $\text{Dim}(\mu^{(2)}) = 1$, alors μ est absolument continue [Shm14, Lemma 2.1].

Fixons k un entier assez grand. On prend pour $\mu_\lambda^{(1)}$ la loi de $\sum_{n \equiv 0 [k]} \xi_n \lambda^n$, et pour $\mu_\lambda^{(2)}$ la loi de $\sum_{n \not\equiv 0 [k]} \xi_n \lambda^n$, de telle sorte que $\mu_\lambda = \mu_\lambda^{(1)} * \mu_\lambda^{(2)}$. La mesure $\mu_\lambda^{(1)}$ est en fait égale à μ_{λ^k} . Un théorème de Erdős et Kahane assure que les coefficients de Fourier

de cette mesure décroissent polynomialement à l'infini, en dehors d'un ensemble de paramètres de dimension de Hausdorff nulle [Shm14, Proposition 2.3]. La mesure $\mu_\lambda^{(2)}$, quant à elle, n'est pas une convolution de Bernoulli, mais elle y ressemble suffisamment pour que des variantes des résultats qu'on a énoncés plus haut s'y appliquent. En particulier, en dehors d'un ensemble de paramètres de dimension de packing nulle, on a $\text{Dim}(\mu_\lambda^{(2)}) = 1$ lorsque $\lambda > 2^{-1+1/k}$. On obtient ainsi que, dans l'intervalle $[2^{-1+1/k}, 1[$, les paramètres exceptionnels pour lesquels μ_λ n'est pas absolument continue ont dimension de Hausdorff nulle. La même conclusion dans $[1/2, 1[$ s'obtient en faisant tendre k vers l'infini. \square

Ce théorème est le dernier en date d'une longue série de raffinements successifs garantissant l'absolue continuité de la mesure μ_λ pour des ensembles de paramètres λ de plus en plus grands. Historiquement, on peut mentionner les résultats intermédiaires suivants :

1. Méthodes de Fourier [Erd40, Kah71]. Il existe $a < 1$ tel que μ_λ est absolument continue pour presque tout $\lambda \in [a, 1[$. Plus généralement, pour tout k , il existe $a_k < 1$ tel que μ_λ est absolument continue avec densité C^k pour presque tout $\lambda \in [a_k, 1[$. Mieux, la dimension de Hausdorff des paramètres $\lambda \in [a, 1[$ pour lesquels μ_λ n'est pas absolument continue avec une densité C^k tend vers 0 lorsque a tend vers 1.
2. Méthodes de transversalité [Sol95, PS00]. La mesure μ_λ est absolument continue pour presque tout $\lambda \in [1/2, 1[$. De plus, pour tout $a > 1/2$, la dimension de Hausdorff des paramètres exceptionnels dans $[a, 1[$ est < 1 .

Pour tous ces résultats, antérieurs à l'introduction par Hochman de méthodes liées à la combinatoire additive et dont nous ne parlerons pas ici, le survol de Peres, Schlag et Solomyak [PSS00] est une référence incontournable. Signalons juste que toutes ces approches ne donnent jamais des ensembles explicites de paramètres. En ce sens, les arguments de Hochman sont un progrès décisif.

3. ENTROPIE ET NOMBRES DE PISOT

Dans ce paragraphe, on démontre de manière détaillée le fait que la convolution de Bernoulli de paramètre λ inverse d'un nombre de Pisot est de dimension < 1 . La preuve est classique et élémentaire, c'est surtout l'occasion de nous familiariser avec l'entropie dans un cas simple, et de voir quelles sont les difficultés qu'il faudra surmonter pour des paramètres λ généraux.

Si X est une variable aléatoire prenant un nombre fini de valeurs, on définit son entropie $H(X)$ par la formule

$$(3.1) \quad H(X) = - \sum_x \mathbb{P}(X = x) \cdot \log \mathbb{P}(X = x).$$

On interprète habituellement cette formule en disant que l'entropie mesure l'information contenue par une variable aléatoire. L'entropie d'une variable pouvant prendre N valeurs est majorée par $\log N$, et vaut $\log N$ si et seulement si la variable est uniforme (cela résulte de la convexité de $x \mapsto -\log x$). Dès que X prend N valeurs de manière à peu près équiprobable, l'entropie $H(X)$ mesure donc (le logarithme de) la taille N de cet ensemble de valeurs. De manière plus générale, si X est à peu près uniforme sur un ensemble de grande mesure (on appellera les valeurs qu'elle y prend les valeurs typiques de X) et le reste des valeurs est négligeable, on s'attend à ce que l'entropie mesure le (logarithme du) nombre de valeurs typiques de X . C'est faux en général (si X n'a pas de bonnes propriétés), mais c'est souvent vrai quand X est engendrée par un processus « naturel ». C'est l'intuition sur laquelle je vais insister dans ce texte : l'entropie d'une variable discrète X mesure le logarithme du nombre de valeurs typiques prises par X .

Fixons $\lambda \in [1/2, 1[$. La variable aléatoire $X_N = \sum_{n < N} \xi_n \lambda^n$ est une variable aléatoire discrète, qui prend au plus 2^N valeurs. On peut donc considérer son entropie $H(X_N)$. Des résultats de théorie ergodique sous-additive permettent de démontrer aisément le résultat suivant, qui justifie l'interprétation précédente de l'entropie dans ce contexte.

PROPOSITION 3.1. — *La suite des entropies $H(X_N)$ est sous-additive : pour tous les entiers M et N , on a $H(X_{N+M}) \leq H(X_N) + H(X_M)$. Ainsi, $H(X_N)/N$ converge vers une limite*

$$(3.2) \quad h(\lambda) = \inf_{N > 0} H(X_N)/N,$$

appelée l'entropie de λ .

Cette quantité mesure la taille des atomes typiques de X_N : pour presque tout ω ,

$$(3.3) \quad \mathbb{P}\{\omega' : X_N(\omega') = X_N(\omega)\} \sim_{\log} e^{-Nh(\lambda)}.$$

Enfin, le cardinal des valeurs typiques de X_N croît comme $e^{Nh(\lambda)}$, dans le sens suivant. Fixons un seuil $s \in]0, 1[$, par exemple $s = 1/2$. Alors

$$(3.4) \quad \min\{\text{Card } A : \mathbb{P}(X_N \in A) \geq s\} \sim_{\log} e^{Nh(\lambda)}.$$

Démonstration. — On travaille sur l'espace $\Omega = \{-1, 1\}^{\mathbb{N}}$ muni de la mesure \mathbb{P} , produit des mesures de Bernoulli, et on note ξ_n la fonction n -ième coordonnée. Ces fonctions forment une suite de variables i.i.d. de Bernoulli. Notons

$$(3.5) \quad Q_N(\omega) = \left\{ \omega' : \sum_{n < N} \xi_n(\omega') \lambda^n = \sum_{n < N} \xi_n(\omega) \lambda^n \right\},$$

et $f_N(\omega) = \mathbb{P}(Q_N(\omega))$. Les ensembles $Q_N(\omega)$ forment une partition \mathcal{Q}_N de l'espace de probabilité Ω suivant la valeur de X_N , et l'entropie de X_N est donnée par l'entropie de cette partition, soit

$$(3.6) \quad H(X_N) = - \sum_{Q \in \mathcal{Q}_N} \mathbb{P}(Q) \log \mathbb{P}(Q) = - \int \log f_N \, d\mathbb{P}.$$

Notons σ le décalage à gauche sur Ω , envoyant $(x_n)_{n \geq 0}$ sur $(x_{n+1})_{n \geq 0}$. Alors on a

$$(3.7) \quad Q_N(\omega) \cap \sigma^{-N}Q_M(\sigma^N\omega) \subseteq Q_{N+M}(\omega).$$

En effet, si un point ω' vérifie les deux égalités

$$\sum_{n < N} \xi_n(\omega')\lambda^n = \sum_{n < N} \xi_n(\omega)\lambda^n \quad \text{et} \quad \sum_{m < M} \xi_{m+N}(\omega')\lambda^i = \sum_{m < M} \xi_{m+N}(\omega)\lambda^i,$$

alors il vérifie évidemment l'égalité $\sum_{i < N+M} \xi_i(\omega')\lambda^i = \sum_{i < N+M} \xi_i(\omega)\lambda^i$.

En considérant la mesure des ensembles dans (3.7) et en utilisant l'indépendance, on obtient l'inégalité $f_N(\omega)f_M(\sigma^N\omega) \leq f_{N+M}(\omega)$. Ainsi,

$$(3.8) \quad -\log f_{N+M}(\omega) \leq -\log f_N(\omega) - \log f_M(\sigma^N\omega).$$

C'est une équation de sous-additivité. En l'intégrant et en utilisant la formule (3.6) pour $H(X_N)$ ainsi que le fait que σ préserve \mathbb{P} , on obtient l'inégalité $H(X_{N+M}) \leq H(X_N) + H(X_M)$ annoncée. C'est alors un résultat général de sous-additivité, appelé lemme de Fekete, que $H(X_N)/N$ converge vers $\text{Inf}_{N > 0} H(X_N)/N$. On note $h = h(\lambda)$ cette limite.

On peut même dire plus : grâce à (3.8), un résultat fondamental de théorie ergodique sous-additive, le théorème de Kingman, assure que $-\log f_N(\omega)/N$ converge presque sûrement vers h . Ainsi, presque sûrement, $\mathbb{P}(Q_N(\omega)) \sim_{\log} e^{-Nh}$, ce qui démontre (3.3).

Montrons enfin (3.4). Si on fixe $\varepsilon > 0$, il s'agit de voir qu'on peut trouver un ensemble de cardinal $e^{N(h+\varepsilon)}$ auquel X_N appartient avec probabilité au moins s , et que c'est impossible avec un ensemble de cardinal $e^{N(h-\varepsilon)}$. Les deux arguments étant semblables, on va faire le premier.

La suite de fonctions $-\log f_N/N$ converge presque sûrement vers h . Elle converge donc uniformément vers h sur un ensemble K de mesure arbitrairement proche de 1, en particulier avec $\mathbb{P}(K) \geq s$. Si N est assez grand, on a $f_N(\omega) \geq e^{-N(h+\varepsilon)}$ pour tout $\omega \in K$, i.e., tout élément de \mathcal{Q}_n qui rencontre K a mesure au moins $e^{-N(h+\varepsilon)}$. Soit A l'ensemble des valeurs prises par X_N sur K . Alors $\mathbb{P}(X_N \in A) \geq \mathbb{P}(K) \geq s$. De plus,

$$\text{Card } A = \sum_{Q \in \mathcal{Q}_n, Q \cap K \neq \emptyset} 1 \leq \sum_{Q \in \mathcal{Q}_n, Q \cap K \neq \emptyset} \mathbb{P}(Q)/e^{-N(h+\varepsilon)} \leq e^{N(h+\varepsilon)} \sum_{Q \in \mathcal{Q}_n} \mathbb{P}(Q) = e^{N(h+\varepsilon)}.$$

Ainsi, le terme de gauche de (3.4) est borné par $e^{N(h+\varepsilon)}$. La borne inférieure en $e^{N(h-\varepsilon)}$ est analogue. \square

COROLLAIRE 3.2. — *La dimension $\text{Dim}(\mu_\lambda)$ vérifie toujours l'inégalité*

$$(3.9) \quad \text{Dim}(\mu_\lambda) \leq \min\left(1, \frac{h(\lambda)}{|\log \lambda|}\right).$$

Démonstration. — L'inégalité $\text{Dim}(\mu_\lambda) \leq 1$ est évidente puisqu'on est en dimension 1. Pour l'autre inégalité, fixons un aléa ω typique. Alors $x = X_\infty(\omega) = \sum \xi_n(\omega)\lambda^n$ est typique pour $\mu_\lambda = (X_\infty)_*\mathbb{P}$. En particulier, le théorème 2.2 assure que $\mu_\lambda([x - C\lambda^N, x + C\lambda^N]) \sim_{\log} \lambda^{N \text{Dim}(\mu_\lambda)}$, pour tout $C > 0$. Si ω' vérifie $X_N(\omega') = X_N(\omega)$, alors $X_\infty(\omega')$ et

$X_\infty(\omega)$ diffèrent au plus de $\sum_{n \geq N} \lambda^n = C_0 \lambda^N$ pour $C_0 = 1/(1 - \lambda)$. Ainsi, pour tout $\omega' \in Q_N(\omega)$ défini en (3.5), on a $X_\infty(\omega') \in [x - C_0 \lambda^N, x + C_0 \lambda^N]$. On obtient donc

$$(3.10) \quad e^{-N \operatorname{Dim}(\mu_\lambda) |\log \lambda|} = \lambda^{N \operatorname{Dim}(\mu_\lambda)} \sim_{\log} \mu_\lambda([x - C_0 \lambda^N, x + C_0 \lambda^N]) \\ \geq \mathbb{P}(Q_N(\omega)) \sim_{\log} e^{-Nh(\lambda)},$$

le dernier équivalent découlant de (3.3). Ainsi, $\operatorname{Dim}(\mu_\lambda) \cdot |\log \lambda| \leq h(\lambda)$ comme annoncé. \square

On voit dans la preuve que l'inégalité (3.9) risque d'être stricte si beaucoup de valeurs de X_N sont regroupées dans un intervalle très court, car l'inégalité dans (3.10) est alors très grossière. Conjecturalement, cependant, on a toujours égalité dans (3.9) : ce phénomène de regroupement excessif ne pourrait se produire que lorsque X_N prend de très nombreuses valeurs, i.e., $h(\lambda)/|\log \lambda| \geq 1$. Si cette conjecture était vraie, on aurait par exemple $\operatorname{Dim}(\mu_\lambda) = 1$ dès que λ est transcendant (ou plus généralement dès que λ n'est pas racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$) puisqu'alors X_N est distribué uniformément sur 2^N points, et $H(X_N) = N \log 2$. On discutera plus loin les méthodes sophistiquées de Hochman permettant de montrer cette conjecture pour une très grande majorité de paramètres λ ; elle est par exemple vraie pour tout λ algébrique, voir le théorème 4.12.

Pour l'instant, nous allons nous restreindre à un cas où l'égalité est élémentaire, celui des inverses des nombres de Pisot (voir la définition 2.4). Le lemme suivant assure que, dans ce cas, les valeurs de X_N sont toutes bien séparées.

LEMME 3.3. — *Soit $\beta > 1$ un nombre de Pisot. Considérons deux sommes $\sum_{n < N} \varepsilon_n \beta^n$ et $\sum_{n < N} \varepsilon'_n \beta^n$ avec $\varepsilon_n, \varepsilon'_n \in \{-1, 1\}$. Si ces sommes sont distinctes, alors elles sont séparées d'au moins $c = c(\beta)$ indépendant de $\varepsilon, \varepsilon'$ et N .*

On en déduit, pour $\lambda = \beta^{-1}$, que deux sommes distinctes $\sum_{n < N} \varepsilon_n \lambda^n$ et $\sum_{n < N} \varepsilon'_n \lambda^n$, sont séparées de $c \lambda^N$.

Si $\beta \in [1, 2[$, alors on a 2^N sommes $\sum_{n < N} \varepsilon_n \beta^n$, uniformément séparées si elles sont distinctes, dans un intervalle de taille $2\beta^N/(\beta - 1) = o(2^N)$. Par conséquent, certaines de ces sommes doivent coïncider : β est nécessairement racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$, ce qui n'est pas évident directement au vu de la définition d'un nombre de Pisot.

Démonstration. — Posons $\eta_n = \varepsilon_n - \varepsilon'_n$. Il s'agit de voir que, si $\sum \eta_n \beta^n$ est de module au plus c où c est assez petit, alors $\sum \eta_n \beta^n = 0$. Notons $(\beta_k)_{1 \leq k \leq K}$ les conjugués de Galois de β distincts de β . Comme ils sont de module < 1 , les sommes $\sum \eta_n \beta_k^n$ sont toutes uniformément bornées par une constante C , indépendamment de N et η . Ainsi,

$$\left| \left(\sum \eta_n \beta^n \right) \cdot \prod_k \left(\sum \eta_n \beta_k^n \right) \right| \leq c \cdot C^k,$$

qui est < 1 si c est assez petit. Le nombre à gauche étant un entier, il est donc nul, ce qui donne $\sum \eta_n \beta^n = 0$, ou $\sum \eta_n \beta_k^n = 0$ pour un certain k . Dans le deuxième cas, comme

β_k est conjugué à β , on obtient encore $\sum \eta_n \beta^n = 0$ en appliquant l'automorphisme qui envoie β_k sur β . \square

On réutilisera plus loin le même type d'arguments dans le cas de nombres algébriques généraux, durant la preuve du théorème 4.12. Au lieu d'une minoration uniforme comme dans le lemme 3.3, on trouvera une minoration exponentiellement petite en terme de N , ce qui nous suffira alors.

COROLLAIRE 3.4. — *Si $\lambda \in [1/2, 1[$ est l'inverse d'un nombre de Pisot, alors $\text{Dim}(\mu_\lambda) = h(\lambda)/|\log \lambda|$.*

Démonstration. — Le corollaire 3.2 donne déjà une inégalité, il s'agit donc de montrer que $\text{Dim}(\mu_\lambda) \cdot |\log \lambda| \geq h - \varepsilon$ pour $\varepsilon > 0$ fixé. On reprend la preuve de ce corollaire, mais en renversant les inégalités. Le lemme 3.3 assure que le nombre de valeurs possibles prises par X_N dans un intervalle de longueur $C\lambda^N$ est borné par un nombre $D = D(C)$ indépendant de N , ce qui va donner une borne supérieure pour $\mu_\lambda([x - \lambda^N, x + \lambda^N])$ pour x typique et permettra de conclure.

Les atomes de \mathcal{Q}_N de masse $> e^{-Nh+N\varepsilon}$ ont une probabilité δ_N qui tend vers 0 d'après la proposition 3.1. On en déduit que la probabilité qu'il existe ω' avec $|X_N(\omega') - X_N(\omega)| \leq C\lambda^N$ et $\mathbb{P}(Q_N(\omega')) > e^{-Nh+N\varepsilon}$ est alors bornée par $D(C)\delta_N$, qui tend aussi vers 0. En effet, en considérant une fonction φ qui associe à un atome $Q \in \mathcal{Q}_N$ un atome Q' avec $|X_N(Q') - X_N(Q)| \leq C\lambda^N$ de masse maximale, cette probabilité vaut

$$\begin{aligned} \sum_{Q, \mathbb{P}(\varphi(Q)) > e^{-Nh+N\varepsilon}} \mathbb{P}(Q) &\leq \sum_{Q, \mathbb{P}(\varphi(Q)) > e^{-Nh+N\varepsilon}} \mathbb{P}(\varphi(Q)) \\ &\leq \sum_{Q', \mathbb{P}(Q') > e^{-Nh+N\varepsilon}} \mathbb{P}(Q') \cdot \text{Card}\{Q : Q' = \varphi(Q)\} \\ &\leq D(C) \sum_{Q', \mathbb{P}(Q') > e^{-Nh+N\varepsilon}} \mathbb{P}(Q') = D(C)\delta_N. \end{aligned}$$

Par conséquent, avec grande probabilité, toutes les valeurs de X_N dans $[X_N(\omega) - C\lambda^N, X_N(\omega) + C\lambda^N]$ sont prises avec probabilité au plus $e^{-Nh+N\varepsilon}$. Ceci donne, avec grande probabilité,

$$\mathbb{P}\{\omega' : |X_N(\omega') - X_N(\omega)| \leq C\lambda^N\} \leq D(C)e^{-Nh+N\varepsilon}.$$

Finalement, $X_\infty(\omega')$ et $X_\infty(\omega)$ ne peuvent être λ^N -proches que si $X_N(\omega')$ et $X_N(\omega)$ étaient distants de au plus $C_0\lambda^N$ pour $C_0 = 1 + 2/(1 - \lambda)$, puisque $|X_N - X_\infty| \leq \lambda^N/(1 - \lambda)$. Ainsi, avec grande probabilité, $\mu_\lambda([x - \lambda^N, x + \lambda^N]) \leq D(C_0)e^{-Nh+N\varepsilon}$. De plus, pour presque tout x

$$\mu_\lambda([x - \lambda^N, x + \lambda^N]) \sim_{\log} \lambda^{N \text{Dim}(\mu_\lambda)} = e^{-N \text{Dim}(\mu_\lambda) |\log \lambda|}.$$

Cela donne l'inégalité recherchée $\text{Dim}(\mu_\lambda) \cdot |\log \lambda| \geq h - \varepsilon$. \square

THÉORÈME 3.5 (Erdős [Erd39], Garsia [Gar63]). — Soit $\lambda \in]1/2, 1[$ tel que λ^{-1} soit un nombre de Pisot. Alors μ_λ est singulière par rapport à la mesure de Lebesgue, et on a même $\text{Dim}(\mu_\lambda) < 1$.

Démonstration. — On commence par montrer que la mesure μ_λ est singulière, en calculant sa transformée de Fourier. En partant de l'expression (1.1) de μ_λ comme convolution infinie, et en utilisant le fait que la transformée de Fourier d'une convolution est le produit des transformées de Fourier, on obtient l'identité

$$\int e^{itx} d\mu_\lambda(x) = \prod_n \cos(\lambda^n t).$$

Si μ_λ était absolument continue, alors sa transformée de Fourier tendrait vers 0 en l'infini d'après le lemme de Riemann-Lebesgue. On va voir que ce n'est pas le cas en estimant la transformée de Fourier pour $t = 2\pi\beta^N$ où $\beta = \lambda^{-1}$ est Pisot. On décompose le produit ci-dessus comme

$$\prod_{n \leq N} \cos(2\pi\lambda^n \beta^N) \cdot \prod_{n > N} \cos(2\pi\lambda^n \beta^N) = \prod_{n \leq N} \cos(2\pi\beta^n) \cdot \prod_{n > 0} \cos(2\pi\lambda^n).$$

Le produit $\prod_{n > 0} \cos(2\pi\lambda^n)$ est convergent et non nul car $\cos(2\pi\lambda^n)$ tend exponentiellement vite vers 1. Notons $(\beta_k)_{k \leq K}$ les conjugués de β , ils sont de module < 1 . Alors $\beta^n + \sum_k \beta_k^n$ est entier, donc $\cos(2\pi\beta^n) = \cos(2\pi \sum_k \beta_k^n)$. Ainsi,

$$\prod_{n \leq N} \cos(2\pi\beta^n) \xrightarrow{N \rightarrow \infty} \prod_n \cos\left(2\pi \sum_{k \leq K} \beta_k^n\right),$$

qui est également un produit convergent non nul. Cela conclut la preuve de la singularité de μ_λ . D'après la proposition 2.1, la mesure μ_λ est même purement singulière.

On va en déduire le fait suivant : pour tout $M > 0$,

$$(3.11) \quad \mathbb{P}\{\omega : \mathbb{P}(Q_N(\omega)) < M\lambda^N\} \rightarrow 0.$$

Autrement dit, la mesure $(X_N)_*\mathbb{P}$, qui est concentrée sur au plus $D_0\lambda^{-N}$ points d'après le lemme 3.3, est en fait loin d'être la mesure uniforme sur cet ensemble, elle est nettement plus concentrée. Pour montrer (3.11), remplaçons $(X_N)_*\mathbb{P}$ par une mesure à densité ν_N , dont la densité autour d'une valeur x prise par X_N est donnée par $1_{[x-c\lambda^N, x+c\lambda^N]}\mathbb{P}(X_N = x)/(2c\lambda^N)$, où c est fixé assez petit pour que ces intervalles soient deux à deux disjoints. Comme ν_N est proche de $(X_N)_*\mathbb{P}$, elle converge également vers la mesure purement singulière μ_λ . En particulier, la probabilité que sa densité soit bornée par une constante donnée tend vers 0. Cela donne (3.11).

Montrons maintenant que, pour N assez grand, on a $H(X_N) < N|\log \lambda|$. Comme $\text{Dim}(\mu_\lambda) = (\text{Inf } H(X_N)/N)/|\log \lambda|$ d'après le corollaire 3.4 et la définition de $h(\lambda)$, cela prouvera que $\text{Dim}(\mu_\lambda) < 1$, comme annoncé. On décompose \mathcal{Q}_N comme union disjointe de deux ensembles, \mathcal{P}_N et \mathcal{S}_N , correspondant respectivement aux atomes principaux, c'est-à-dire ceux pour lesquels $\mathbb{P}(Q_N(\omega)) \geq M\lambda^N$, et aux atomes secondaires pour lesquels $\mathbb{P}(Q_N(\omega)) < M\lambda^N$ (où M est fixé assez grand, $M = 2$ suffira). On notera P_N et

S_N les unions des atomes principaux et secondaires respectivement. Il y a peu d'atomes principaux, ce qui va permettre de montrer que leur contribution à l'entropie est assez petite. Et la probabilité des atomes secondaires est petite d'après (3.11), donc ils auront aussi une petite contribution. Le calcul se fait comme suit. Estimons tout d'abord la contribution des atomes principaux. Puisqu'ils vérifient tous $\mathbb{P}(Q) \geq M\lambda^N$, on a

$$\begin{aligned} - \sum_{Q \in \mathcal{P}_N} \mathbb{P}(Q) \log \mathbb{P}(Q) &\leq - \sum_{Q \in \mathcal{P}_N} \mathbb{P}(Q) \log(M\lambda^N) = \mathbb{P}(P_N)(N|\log \lambda| - \log M) \\ &\leq \mathbb{P}(P_N)N|\log \lambda| - \log M/2, \end{aligned}$$

la dernière inégalité étant valide si N est assez grand car $\mathbb{P}(P_N) \rightarrow 1$ d'après (3.11). Pour les atomes secondaires, on fait apparaître la mesure de probabilité $\mathbb{P}/\mathbb{P}(S_N)$ sur \mathcal{S}_N . Comme il y a au plus $D_0\lambda^{-N}$ atomes, l'entropie de cette mesure est bornée par $\log(D_0\lambda^{-N})$. Comme $\mathbb{P}(S_N) \rightarrow 0$, on obtient

$$\begin{aligned} (3.12) \quad - \sum_{Q \in \mathcal{S}_N} \mathbb{P}(Q) \log \mathbb{P}(Q) &= \mathbb{P}(S_N) \left(-\log \mathbb{P}(S_N) - \sum_{Q \in \mathcal{S}_N} \frac{\mathbb{P}(Q)}{\mathbb{P}(S_N)} \log \left(\frac{\mathbb{P}(Q)}{\mathbb{P}(S_N)} \right) \right) \\ &\leq -\mathbb{P}(S_N) \log \mathbb{P}(S_N) + \mathbb{P}(S_N) \cdot \log(D_0\lambda^{-N}) = o(1) + \mathbb{P}(S_N)N|\log \lambda|. \end{aligned}$$

En sommant les deux inégalités, et comme $\mathbb{P}(P_N) + \mathbb{P}(S_N) = 1$, on trouve

$$(3.13) \quad H(X_N) \leq N|\log \lambda| + o(1) - \log M/2.$$

Cela donne bien $H(X_N) < N|\log \lambda|$ si N est assez grand. \square

Remarque 3.6. — La preuve de la dimension < 1 est un peu subtile, car (3.13) ne donne pas une asymptotique de la forme $H(X_N) \leq (1 - \varepsilon)N|\log \lambda|$ mais une inégalité plus faible, sans décroissance en N (qui implique toutefois le résultat recherché par sous-additivité). En particulier, cette preuve n'est pas quantitative, elle transforme une assertion de singularité en une assertion plus forte mais sans borne explicite. Je ne connais pas de preuve directe de l'assertion sur la dimension, qui ne passe pas d'abord par le calcul de la transformée de Fourier pour obtenir la singularité de μ_λ .

4. ENTROPIE À PETITE ÉCHELLE POUR LES CONVOLUTIONS DE BERNOULLI

La discussion précédente montre que, en général, $h(\lambda)$ risque de ne pas capturer la dimension lorsqu'il y a des sommes X_N qui sont très proches, mais distinctes. Une manière plus sûre de calculer la dimension devrait être de regrouper toutes les sommes assez proches par paquets, et de calculer l'entropie ainsi produite. C'est ce qu'on va faire maintenant.

4.1. Compléments sur l'entropie

On doit tout d'abord introduire un peu plus de matériel autour de la notion d'entropie. Si μ est une mesure de probabilité et \mathcal{Q} est une partition finie, on notera $H(\mu, \mathcal{Q})$ (ou simplement $H(\mathcal{Q})$ si μ est implicite) pour l'entropie de la partition, donnée par

$$H(\mu, \mathcal{Q}) = - \sum_{Q \in \mathcal{Q}} \mu(Q) \log \mu(Q).$$

C'est l'information moyenne donnée par la connaissance de l'élément de \mathcal{Q} dans laquelle on se trouve (où l'information est $I(x) = -\log \mu(Q(x))$, avec $Q(x)$ l'atome de \mathcal{Q} contenant x). Ainsi, l'entropie d'une variable aléatoire discrète X sur un espace (Ω, \mathbb{P}) , définie en (3.1), est l'entropie de la partition formée des atomes où X est constante, de la forme $X^{-1}(\{x\})$.

L'entropie des partitions admet aussi une version relative : si \mathcal{R} est une partition plus fine que \mathcal{Q} (i.e., chaque atome de \mathcal{Q} est union d'atomes de \mathcal{R}), on pose

$$H(\mu, \mathcal{R} \mid \mathcal{Q}) = \sum_{Q \in \mathcal{Q}} \mu(Q) \left(- \sum_{R \in \mathcal{R}, R \subseteq Q} \frac{\mu(R)}{\mu(Q)} \log \left(\frac{\mu(R)}{\mu(Q)} \right) \right).$$

Dans cette définition, le terme entre parenthèses à droite est l'entropie de la partition \mathcal{R} pour la mesure de probabilité $\mu|_Q/\mu(Q)$, il est donc positif et majoré par $\log \text{Card}\{R \in \mathcal{R} : R \subseteq Q\}$. En développant $\log(\mu(R)/\mu(Q)) = \log \mu(R) - \log \mu(Q)$ et en regroupant les termes, on obtient l'égalité

$$H(\mu, \mathcal{R} \mid \mathcal{Q}) = H(\mu, \mathcal{R}) - H(\mu, \mathcal{Q}).$$

Ceci fournit donc le lemme suivant :

LEMME 4.1. — *Si \mathcal{R} est une partition finie sur un espace probabilisé (Ω, μ) , plus fine qu'une partition \mathcal{Q} , alors*

$$0 \leq H(\mu, \mathcal{R}) - H(\mu, \mathcal{Q}) \leq \max_{Q \in \mathcal{Q}} \log \text{Card}\{R \in \mathcal{R} : R \subseteq Q\}.$$

Notons que le calcul (3.12) reposait déjà implicitement sur cette notion d'entropie relative.

Plus généralement, si \mathcal{Q} et \mathcal{R} sont deux partitions finies mais non comparables, on peut considérer la partition $\mathcal{Q} \vee \mathcal{R}$ dont les atomes sont formés des intersections des atomes de \mathcal{Q} et de \mathcal{R} (qui est plus fine que \mathcal{Q} et \mathcal{R}) et lui appliquer le résultat précédent. On obtient donc :

$$(4.1) \quad |H(\mu, \mathcal{Q}) - H(\mu, \mathcal{R})| \leq \max \left(\max_{Q \in \mathcal{Q}} \log \text{Card}\{R \in \mathcal{R} : Q \cap R \neq \emptyset\}, \right. \\ \left. \max_{R \in \mathcal{R}} \log \text{Card}\{Q \in \mathcal{Q} : Q \cap R \neq \emptyset\} \right).$$

Ainsi, deux partitions qui se ressemblent ont des entropies comparables.

Donnons quelques résultats faciles sur l'entropie dont nous aurons besoin plus loin.

LEMME 4.2. — *L'entropie d'une variable aléatoire X sur un espace probabilisé, à valeurs dans un ensemble fini, vérifie les propriétés suivantes :*

1. *pour toute fonction f , la variable aléatoire discrète $f(X)$ vérifie $H(f(X)) \leq H(X)$.*
2. *si Y est définie sur le même espace que X , alors la variable aléatoire discrète (X, Y) vérifie $H(X, Y) \leq H(X) + H(Y)$, avec égalité si et seulement si X et Y sont indépendantes.*
3. *Si Y et X sont toutes deux à valeurs réelles (ou dans un groupe commutatif), alors*

$$|H(X) - H(Y)| \leq H(X - Y).$$

Démonstration. — Pour la première assertion, les ensembles où $f(X)$ est constante sont union d'ensembles où X est constante. La partition correspondante a donc une entropie plus petite, d'après le lemme 4.1.

Pour le deuxième point, on calcule comme suit.

$$\begin{aligned} H(X, Y) &= - \sum_{x,y} \mathbb{P}(X = x, Y = y) \log \mathbb{P}(X = x, Y = y) \\ &= - \sum_{x,y} \mathbb{P}(X = x, Y = y) \log \left(\frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)} \right) \\ &\quad - \sum_{x,y} \mathbb{P}(X = x, Y = y) (\log \mathbb{P}(X = x) + \log \mathbb{P}(Y = y)). \end{aligned}$$

La dernière somme donne $H(X) + H(Y)$ en séparant les deux termes, il s'agit donc de voir que le premier terme est négatif. On le réécrit comme

$$\begin{aligned} \sum_{x,y} \mathbb{P}(X = x, Y = y) \log \left(\frac{\mathbb{P}(X = x)\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \right) \\ \leq \log \left(\sum_{x,y} \mathbb{P}(X = x, Y = y) \frac{\mathbb{P}(X = x)\mathbb{P}(Y = y)}{\mathbb{P}(X = x, Y = y)} \right) = \log 1 = 0, \end{aligned}$$

l'inégalité découlant de la concavité du logarithme et de l'inégalité de Jensen.

Pour le troisième point, $X = (X - Y) + Y$ est une fonction de $X - Y$ et Y . Ainsi, $H(X) \leq H(X - Y, Y)$ par le premier point. Cette entropie est bornée par $H(X - Y) + H(Y)$ d'après le deuxième point. Ainsi, $H(X) - H(Y) \leq H(X - Y)$. On contrôle de même $H(Y) - H(X)$, pour obtenir l'inégalité recherchée. \square

4.2. Entropie pour les convolutions de Bernoulli

On note \mathcal{D}_n la partition dyadique de \mathbb{R} en intervalles de taille 2^{-n} , de la forme $[k2^{-n}, (k+1)2^{-n}[$ pour $k \in \mathbb{Z}$. On note $D_n(x)$ l'intervalle de \mathcal{D}_n qui contient x . On rappelle qu'on a défini la notion de mesure dimensionnellement exacte dans le théorème 2.2.

PROPOSITION 4.3. — Soit μ une mesure réelle dimensionnellement exacte de dimension $\alpha > 0$. Alors, pour presque tout x ,

$$\mu(D_n(x)) \sim_{\log} 2^{-\alpha n}.$$

Démonstration. — La fonction $r \mapsto \log \mu([x-r, x+r]) / \log r$ converge pour μ -presque tout x vers α , par définition d'une mesure dimensionnellement exacte. Elle converge uniformément sur un ensemble K de mesure arbitrairement grande.

On va commencer par voir que presque tout point est éloigné des bords des éléments de \mathcal{D}_n . Fixons $\delta > 0$. Montrons qu'il existe alors $\rho > 0$ tel que, pour n assez grand, pour tout $x \in \partial\mathcal{D}_n$,

$$(4.2) \quad \mu(K \cap [x - 2^{-(1+\delta)n}, x + 2^{(1-\delta)n}]) \leq e^{-\rho n} \mu([x - 2^{-n-2}, x + 2^{-n-2}]).$$

Si l'ensemble à gauche est vide, il n'y a rien à faire. Sinon, prenons $y \in K \cap [x - 2^{-(1+\delta)n}, x + 2^{(1-\delta)n}]$. Alors $K \cap [x - 2^{-(1+\delta)n}, x + 2^{(1-\delta)n}] \subseteq [y - 2^{-(1+\delta)n+1}, y + 2^{-(1+\delta)n+1}]$. Ce dernier ensemble a mesure au plus $2^{-\alpha(1+\delta)n+\varepsilon n}$ si n est assez grand, par définition de K (où $\varepsilon > 0$ est fixé, assez petit). De plus, $[x - 2^{-n-2}, x + 2^{-n-2}]$ contient $[y - 2^{-n-3}, y + 2^{-n-3}]$, qui a mesure au moins $2^{-\alpha n - \varepsilon n}$ par définition de K . Si ε est assez petit, on a $-(1+\delta)\alpha + \varepsilon < -\alpha - \varepsilon$, ce qui donne alors (4.2).

En sommant l'inégalité (4.2) sur les $x \in \partial\mathcal{D}_n$, on en déduit que $\{y \in K : d(y, \partial\mathcal{D}_n) \leq 2^{-(1+\delta)n}\}$ a mesure au plus $e^{-\rho n}$ (car les intervalles apparaissant à droite de (4.2) sont deux à deux disjoints), qui est sommable. Ainsi, par le lemme de Borel-Cantelli, presque tout $x \in K$ vérifie $d(x, \partial\mathcal{D}_n) \geq 2^{-(1+\delta)n}$ pour n assez grand.

Alors $D_n(x)$ contient $[x - 2^{-(1+\delta)n}, x + 2^{-(1+\delta)n}]$ et est contenu dans $[x - 2^{1-n}, x + 2^{1-n}]$. Le premier intervalle a une mesure au moins $2^{-\alpha(1+\delta)n - \varepsilon n}$, le second une mesure au plus $2^{-\alpha n + \varepsilon n}$. Les deux exposants étant arbitrairement proches de α , on en déduit que $\mu(D_n(x)) \sim_{\log} 2^{-\alpha n}$ pour presque tout $x \in K$. Comme la mesure de K est arbitrairement proche de 1, cela conclut la preuve. \square

On déduit de cette proposition que le nombre d'atomes de \mathcal{D}_n nécessaires pour couvrir une proportion significative de μ croît comme $2^{\alpha n}$, exactement comme dans la preuve de la proposition 3.1. Cela pointe vers une interprétation entropique de la dimension de la mesure μ .

Si μ est une mesure à support compact sur \mathbb{R} et \mathcal{D}_n est la partition dyadique, alors $H(\mu, \mathcal{D}_n)$ peut aussi être vu comme l'entropie de la variable discrète $[2^n X]$, où X est distribuée suivant μ . Cette entropie mesure la complexité de μ à l'échelle 2^{-n} . Il découle de la proposition 4.3 que cette entropie calcule la dimension de la mesure μ :

PROPOSITION 4.4. — Soit μ une mesure réelle dimensionnellement exacte à support compact. Alors $H(\mu, \mathcal{D}_n) / \log(2^n) \rightarrow \text{Dim}(\mu)$ lorsque $n \rightarrow \infty$.

Démonstration. — Cette proposition se déduit de la proposition 4.3 en utilisant les mêmes techniques que dans la partie 3. On sépare les atomes de \mathcal{D}_n dont la mesure est proche de $2^{-n \text{Dim}(\mu)}$ (ils ont une probabilité proche de 1 d'après la proposition 4.3), et les autres atomes (de probabilité presque nulle). Les premiers donnent une contribution

à l'entropie de l'ordre de n $\text{Dim}(\mu) \log 2$, les seconds ont une contribution négligeable (comme dans (3.12)). \square

Les travaux de Hochman [Hoc14] sont formulés en termes de l'entropie dyadique. Cependant, il est apparu plus tard chez Breuillard et Varjú qu'il était techniquement plus efficace de rajouter une moyennisation supplémentaire : quand on utilise la partition dyadique \mathcal{D}_n , on fixe arbitrairement et inutilement un point de référence 0. Étant données une mesure réelle μ et une échelle $r > 0$ (le cas dyadique correspond à $r = 2^{-n}$), on notera dans la suite

$$(4.3) \quad H(\mu, r) = \int_{t=0}^1 H(\mu, \{[tr + kr, tr + (k+1)r[, k \in \mathbb{Z}\}) dt,$$

l'entropie de μ par rapport à une partition de \mathbb{R} en intervalles de taille r , dont le point base tr est choisi aléatoirement. Notons que le choix de différents points base donne lieu à des partitions qui se ressemblent beaucoup (chaque intervalle de la nouvelle partition rencontre au plus deux intervalles de l'ancienne, et inversement), si bien que les entropies de ces deux partitions diffèrent au plus de $\log 2$ d'après (4.1). Ainsi,

$$(4.4) \quad |H(\mu, r) - H(\mu, \{[kr, (k+1)r[, k \in \mathbb{Z}\})| \leq \log 2.$$

En particulier, les comportements asymptotiques sont les mêmes. En revanche, quand on voudra faire des estimées fines, la quantité moyennée se comportera mieux (avec des inégalités valables en permanence, sans terme d'erreur). C'est ce langage qu'on utilisera dans la suite.

La proposition 4.4 se reformule ainsi :

PROPOSITION 4.5. — *Soit μ une mesure réelle dimensionnellement exacte à support compact. Alors $H(\mu, r)/|\log r| \rightarrow \text{Dim}(\mu)$ lorsque $r \rightarrow 0$.*

Démonstration. — Si r_n tend exponentiellement vite vers 0, alors on peut faire fonctionner les arguments à base de Borel-Cantelli dans la preuve de la proposition 4.3, et démontrer ainsi que $H(\mu, \{[kr_n, (k+1)r_n[, k \in \mathbb{Z}\})/|\log r_n|$ tend vers $\text{Dim}(\mu)$. La convergence de $H(\mu, r_n)/|\log r_n|$ vers $\text{Dim}(\mu)$ en découle d'après (4.4). Enfin, si $H(\mu, r)/|\log r|$ ne convergerait pas vers $\text{Dim}(\mu)$, on pourrait trouver une sous-suite r_n tendant vers 0 arbitrairement vite le long de laquelle cette convergence n'aurait pas lieu, ce qui donnerait une contradiction. \square

Une convolution de Bernoulli μ_λ ressemble à l'échelle λ^N à la loi de $X_N = \sum_{n < N} \xi_n \lambda^n$. Introduisons une notation utile pour celle-ci : pour I une partie de $[0, 1]$, on notera μ_λ^I la loi de $\sum \xi_n \lambda^n$, où la somme est restreinte aux n tels que $\lambda^n \in I$. Ainsi, la loi de X_N est par définition $\mu_\lambda^{] \lambda^N, 1]}$. Moralement, μ_λ^I capture le comportement de μ_λ qui se manifeste aux échelles dans I . Pour $I =] \lambda^N, 1]$ (ou plus généralement $]r, 1]$ avec $r \in [\lambda^N, \lambda^{N-1}[$), on retrouve la loi de X_N . Si I et J sont deux ensembles disjoints, on a

$$(4.5) \quad \mu_\lambda^{I \cup J} = \mu_\lambda^I * \mu_\lambda^J.$$

Cela traduit de manière efficace que μ_λ est une mesure obtenue comme convolution, en (1.1).

La proposition qui suit est naturelle : comme l'entropie à l'échelle r ne dépend que du comportement de la mesure à l'échelle r , toute contribution à une échelle plus petite n'est pas pertinente pour la calculer, et on peut donc remplacer μ par $\mu^{[r,1]}$ dans l'énoncé de la proposition 4.5.

PROPOSITION 4.6. — *Soit μ_λ une convolution de Bernoulli de paramètre $\lambda \in [1/2, 1[$. Alors $H(\mu_\lambda^{[r,1]}, r)/|\log r| \rightarrow \text{Dim}(\mu_\lambda)$ lorsque $r \rightarrow 0$.*

Démonstration. — La mesure $\mu_\lambda^{[r,1]}$ est la loi de la variable aléatoire X_N pour $N = N(r)$, tandis que μ_λ est la loi de X_∞ . De plus, X_N et X_∞ diffèrent au plus de Cr , où C est une constante. Les entropies de $\mu_\lambda^{[r,1]}$ et μ_λ pour la partition $\{[kr, (k+1)r]\}$ sont les entropies des variables aléatoires discrètes $[r^{-1}X_N]$ et $[r^{-1}X_\infty]$. La différence de ces variables est entière et uniformément bornée, donc d'entropie uniformément bornée. Le point 3 du lemme 4.2 assure donc que les entropies de ces variables diffèrent au plus d'une constante. Avec (4.4), on obtient $H(\mu_\lambda^{[r,1]}, r) = H(\mu_\lambda, r) + O(1)$. Ainsi, le résultat recherché découle du comportement asymptotique de $H(\mu_\lambda, r)$ donné dans la proposition 4.5. \square

L'intérêt de la formulation moyennée de l'entropie est qu'elle vérifie de nombreuses inégalités sans termes d'erreur, qu'on va donner maintenant. Pour les énoncer, on notera aussi $H(\mu, r | s) = H(\mu, r) - H(\mu, s)$ pour le gain d'entropie quand on passe de l'échelle s à l'échelle $r < s$. Lorsque $s = Nr$ est un multiple entier de r , il y a une description géométrique simple de ce gain d'entropie : on commence par regarder une partition de \mathbb{R} en intervalles de taille Nr (avec point base aléatoire), puis on la raffine en coupant chaque intervalle en N sous-intervalles (ce qui donne une partition de \mathbb{R} en intervalles de taille r , avec point base aléatoire). Ainsi, $H(\mu, r | Nr)$ s'écrit comme une moyenne d'entropies relatives. Avec le lemme 4.1, on obtient

$$0 \leq H(\mu, r | Nr) \leq \log N.$$

On peut aussi obtenir des estimées similaires lorsque s n'est pas un multiple entier de r , mais c'est plus délicat. Nous aurons besoin en particulier des lemmes suivants, que nous donnons sans preuve (les preuves sont élémentaires mais nécessitent un peu de travail).

LEMME 4.7 ([Var16, Lemma 8]). — *Soit μ une mesure à support compact sur \mathbb{R} . Alors, pour tous $r \leq s$,*

$$0 \leq H(\mu, r | s) \leq 2 \log(s/r).$$

Convolver μ par une mesure ν rend la mesure plus régulière, et devrait donc faire augmenter son entropie (puisque la mesure uniforme est d'entropie maximale). C'est ce qu'affirme le lemme suivant.

LEMME 4.8 ([Var16, Lemma 6]). — Soient μ et ν deux mesures à support compact sur \mathbb{R} . Alors, pour toute échelle r et pour tout entier N ,

$$H(\mu, r \mid Nr) \leq H(\mu * \nu, r \mid Nr).$$

Les deux lemmes précédents sont faux pour la version de l'entropie avec point base non aléatoire, ce qui illustre bien l'intérêt de la version moyennisée (4.3).

On déduit de ces deux lemmes la version suivante où le rapport des échelles n'a pas à être entier.

LEMME 4.9. — Soient μ et ν deux mesures à support compact sur \mathbb{R} . Alors, pour toutes échelles $r \leq s$,

$$H(\mu, r \mid s) \leq H(\mu * \nu, r \mid s) + 2 \log 2.$$

Démonstration. — Choisissons un entier N tel que $Nr \in [s, 2s]$. Alors

$$H(\mu, r \mid s) \leq H(\mu, r \mid Nr) \leq H(\mu * \nu, r \mid Nr) \leq H(\mu * \nu, r \mid s) + 2 \log 2,$$

où l'inégalité centrale découle du lemme 4.8, et les deux inégalités extrêmes proviennent du lemme 4.7. \square

En particulier, comme $\mu_\lambda = \mu_\lambda^{[0,1]}$ s'obtient en convolant μ_λ^I avec $\mu_\lambda^{[0,1]\setminus I}$ d'après (4.5), on obtient

$$(4.6) \quad H(\mu_\lambda^I, r \mid s) \leq H(\mu_\lambda, r \mid s) + 2 \log 2,$$

où le terme d'erreur $2 \log 2$ peut être enlevé si s est un multiple entier de r .

L'auto-similarité de la mesure se traduit aussi par la formule suivante :

$$(4.7) \quad H(\mu_\lambda^{\lambda^k I}, \lambda^k r \mid \lambda^k s) = H(\mu_\lambda^I, r \mid s),$$

puisque la dilatation de facteur λ^k envoie la mesure μ_λ^I sur la mesure $\mu_\lambda^{\lambda^k I}$ par définition.

On a maintenant tous les outils pour démontrer la semi-continuité inférieure de $\text{Dim}(\mu_\lambda)$ en fonction de λ . C'est pour nous surtout un exercice pour montrer comment utiliser les lemmes précédents et tirer parti de l'auto-similarité de la mesure.

PROPOSITION 4.10. — La fonction $\lambda \mapsto \text{Dim}(\mu_\lambda)$ est semi-continue inférieurement : si $\lambda_n \rightarrow \lambda$, alors $\liminf \text{Dim}(\mu_{\lambda_n}) \geq \text{Dim}(\mu_\lambda)$.

Démonstration. — On suit la preuve de [BV16, Lemma 16]. On vérifie (en utilisant le lemme 4.7) pour $k \in \mathbb{N}$ que la fonction

$$(4.8) \quad f_k(\lambda) = \frac{H(\mu_\lambda, \lambda^{2^k} \mid 1) - 2 \log 2}{2^k |\log \lambda|}$$

dépend continûment de λ . Lorsque k tend vers l'infini, $f_k(\lambda)$ converge vers $\text{Dim}(\mu_\lambda)$ d'après la proposition 4.5. Comme une limite croissante de fonctions continues est semi-continue inférieurement, il suffit donc de vérifier que $f_{k+1}(\lambda) \geq f_k(\lambda)$ pour conclure.

On écrit

$$(4.9) \quad H(\mu_\lambda, \lambda^{2^{k+1}} \mid 1) = H(\mu_\lambda, \lambda^{2^{k+1}} \mid \lambda^{2^k}) + H(\mu_\lambda, \lambda^{2^k} \mid 1).$$

On veut comparer le premier terme à $H(\mu_\lambda, \lambda^{2^k} \mid 1)$. En appliquant (4.6) avec $I = [0, \lambda^{2^k}]$, on obtient

$$H(\mu_\lambda^{[0, \lambda^{2^k}]}, \lambda^{2^{k+1}} \mid \lambda^{2^k}) \leq H(\mu_\lambda, \lambda^{2^{k+1}} \mid \lambda^{2^k}) + 2 \log 2.$$

Dans le terme de gauche de cette équation, si l'on dilate tout d'un facteur λ^{-2^k} comme en (4.7), on retrouve l'entropie de la mesure initiale μ_λ vue entre les échelles λ^{2^k} et 1. Ainsi,

$$H(\mu_\lambda, \lambda^{2^k} \mid 1) \leq H(\mu_\lambda, \lambda^{2^{k+1}} \mid \lambda^{2^k}) + 2 \log 2.$$

Par conséquent, en utilisant la décomposition (4.9), on obtient l'inégalité $2H(\mu_\lambda, \lambda^{2^k} \mid 1) \leq H(\mu_\lambda, \lambda^{2^{k+1}} \mid 1) + 2 \log 2$. Cela donne directement l'inégalité $f_{k+1}(\lambda) \geq f_k(\lambda)$, comme on le voulait. \square

4.3. Le théorème de Hochman, et applications

On a pour l'instant deux notions d'entropie : $H(X_N)$ capture toute l'entropie de la variable X_N , tandis que $H(X_N, \lambda^N)$ regroupe les valeurs de X_N qui sont distantes de moins de λ^N , et contient donc moins d'information. La première est assez facile à calculer car c'est une quantité purement algébrique, tandis que la seconde donne la dimension de la mesure d'après la proposition 4.6 mais elle est nettement plus difficile à estimer car elle requiert de comprendre quelles sommes coïncident *presque*, ce qui nous fait sortir du monde algébrique. Les deux quantités diffèrent typiquement, à moins que les valeurs de X_N ne soient séparées de $c\lambda^N$ comme dans le cas exceptionnel des nombres de Pisot. On peut dire que $H(X_N, \lambda^N)$ ne voit pas encore toute l'entropie de X_N . Si on regarde $H(X_N, s)$ avec $s \leq \lambda^N$, on commence à voir plus d'entropie, et lorsque $s \rightarrow \infty$ on finit par épuiser toute l'entropie de X_N .

Le théorème principal de Hochman, qui a lancé tous les progrès récents sur les convolutions de Bernoulli, est le renforcement suivant de la proposition 4.6. Il affirme qu'au lieu de regarder X_N à l'échelle λ^N dans la discussion précédente on peut la regarder à n'importe quelle échelle exponentiellement petite $s = e^{-AN}$ (ou $s = \lambda^{AN}$) sans perte, dès lors que $\text{Dim}(\mu_\lambda) < 1$.

THÉORÈME 4.11. — *Soit μ_λ une convolution de Bernoulli de paramètre $\lambda \in [1/2, 1[$. Supposons $\text{Dim}(\mu_\lambda) < 1$. Soit $A \geq 1$. Alors $H(\mu_\lambda^{[r, 1]}, r^A)/|\log r| \rightarrow \text{Dim}(\mu_\lambda)$ lorsque $r \rightarrow 0$. De manière équivalente, $H(X_N, \lambda^{AN})/(N|\log \lambda|) \rightarrow \text{Dim}(\mu_\lambda)$.*

Ce théorème a l'air d'un énoncé purement technique, mais il est en fait fondamental. Par rapport à la proposition 4.6, il affirme que la mesure $\mu_\lambda^{[r, 1]}$ n'est pas plus compliquée si on la regarde à l'échelle r^A qu'à l'échelle r dès que $\text{Dim}(\mu_\lambda) < 1$ (ce qui implique qu'il y a beaucoup de trous dans la distribution de μ_λ). Autrement dit, les valeurs typiques de $\sum_{n=0}^{N-1} \xi_n \lambda^n$ sont soit séparées de $c\lambda^N$, soit à distance au plus λ^{AN} , et ce pour tout $A > 0$. C'est un phénomène qu'on a déjà vu avec les inverses des nombres de Pisot, où la coïncidence est même exacte d'après le lemme 3.3. Il est remarquable que, sans aucune hypothèse algébrique sur le paramètre λ , Hochman arrive à une conclusion

pratiquement aussi forte que pour les nombres de Pisot, simplement sous une hypothèse géométrique de dimension non maximale sur la mesure.

Avant de parler dans la section 5 de la démonstration de ce théorème, montrons comment il implique plusieurs des résultats que nous avons évoqués dans l'introduction. Tout d'abord, on en déduit que l'inégalité dans le corollaire 3.2 est une égalité pour les nombres algébriques. On rappelle que $h(\lambda)$ est définie dans l'équation (3.2).

THÉORÈME 4.12. — *Soit $\lambda \in [1/2, 1[$ algébrique sur \mathbb{Q} . Alors on a*

$$(4.10) \quad \text{Dim}(\mu_\lambda) = \min\left(1, \frac{h(\lambda)}{|\log \lambda|}\right).$$

Démonstration. — On commence par montrer une version faible du lemme d'espace-ment 3.3 pour les nombres de Pisot. Si λ est algébrique sur \mathbb{Q} , montrons qu'il existe $B > 0$ et $c > 0$ tels que deux sommes $\sum_{n < N} \varepsilon_n \lambda^n$ et $\sum_{n < N} \varepsilon'_n \lambda^n$ (avec $\varepsilon_n, \varepsilon'_n \in \{-1, 1\}$) sont soit égales, soit distantes d'au moins ce^{-BN} . Quitte à remplacer λ par $k\lambda$ où k est entier, il suffit de montrer cet énoncé pour λ un entier algébrique. Posons $\eta_n = \varepsilon_n - \varepsilon'_n$. Si les sommes ne coïncident pas, alors en notant λ_k les conjugués de Galois de λ distincts de λ , le nombre

$$\left(\sum \eta_n \lambda^n\right) \cdot \prod_k \left(\sum \eta_n \lambda_k^n\right)$$

est un entier non nul, minoré en valeur absolue par 1. De plus, tous les termes $\sum \eta_n \lambda_k^n$ sont au plus exponentiellement grands en N . L'inégalité annoncée s'en déduit.

Prouvons maintenant l'égalité (4.10). L'inégalité $\text{Dim}(\mu_\lambda) \leq \min\left(1, \frac{h(\lambda)}{|\log \lambda|}\right)$ est toujours vraie d'après le corollaire 3.2. Supposons que cette inégalité soit stricte. On a alors $\text{Dim}(\mu_\lambda) < 1$, ce qui permet d'appliquer le théorème 4.11. Ce théorème montre que $H(\sum_{n < N} \xi_n \lambda^n, \lambda^{AN})/N|\log \lambda| \rightarrow \text{Dim}(\mu_\lambda)$ pour tout A . Prenons A assez grand pour que les différentes valeurs de ces sommes soient toutes séparées de plus de λ^{AN} , c'est possible d'après le résultat de séparation précédent. Alors $H(\sum_{n < N} \xi_n \lambda^n, \lambda^{AN}) = H(\sum_{n < N} \xi_n \lambda^n)$ car il n'y a aucun phénomène de regroupement par paquets. Finalement, $H(\sum_{n < N} \xi_n \lambda^n)/(N|\log \lambda|) \rightarrow \text{Dim}(\mu_\lambda)$. Comme $H(\sum_{n < N} \xi_n \lambda^n)/N \rightarrow h(\lambda)$ par définition de $h(\lambda)$ (voir la proposition 3.1), on trouve $\text{Dim}(\mu_\lambda) = h(\lambda)/|\log \lambda|$. Cela contredit notre hypothèse que l'inégalité dans (4.10) était stricte. \square

Dans le cas extrême où les valeurs prises par les sommes de Bernoulli sont toutes distinctes, on obtient le résultat suivant.

PROPOSITION 4.13 (Hochman [Hoc14]). — *Soit $\lambda \in [1/2, 1[$ algébrique sur \mathbb{Q} , mais qui n'est pas racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$. Alors $\text{Dim}(\mu_\lambda) = 1$.*

Démonstration. — Si λ n'est pas racine d'un polynôme à coefficients dans $\{-1, 0, 1\}$, alors $X_N = \sum_{n < N} \xi_n \lambda^n$ prend 2^N valeurs distinctes, de manière équiprobable. Par conséquent, $H(X_N) = N \log 2$, puis $h(\lambda) = \log 2$. Finalement, grâce au corollaire 4.12,

$$\text{Dim}(\mu_\lambda) = \min\left(1, \frac{h(\lambda)}{|\log \lambda|}\right) = \min\left(1, \frac{\log 2}{|\log \lambda|}\right) = 1,$$

la dernière égalité provenant du fait que $\lambda \in [1/2, 1[$. \square

Lorsqu'il y a des coïncidences entre les valeurs prises par les convolutions de Bernoulli, on peut quand même exploiter le théorème 4.12 si l'on est capable de quantifier la fréquence de ces coïncidences. C'est ce que font Breuillard et Varjú dans [BV15]. Ils en déduisent le résultat suivant. On rappelle que la mesure de Mahler M_λ a été définie en (2.1).

THÉORÈME 4.14 (Breuillard-Varjú [BV15]). — *Il existe c_0 avec la propriété suivante ($c_0 = 0,44$ fonctionne). Soit $\lambda \in [1/2, 1[$ algébrique sur \mathbb{Q} avec*

$$\lambda \geq \min(2, M_\lambda)^{-c_0}.$$

Alors $\text{Dim}(\mu_\lambda) = 1$.

Esquisse de démonstration. — Le point clé de la preuve est l'estimée suivante [BV15, Theorem 5], qui montre comment l'entropie $h(\lambda)$ définie en (3.2) est liée à la mesure de Mahler de λ lorsque λ est algébrique : on a

$$(4.11) \quad c_0 \min(\log 2, \log M_\lambda) \leq h(\lambda) \leq \min(\log 2, \log M_\lambda),$$

pour $c_0 = 0,44$. Nous admettrons cette estimée, qui constitue la contribution majeure de [BV15].

D'après le théorème 4.12, pour démontrer que $\text{Dim}(\mu_\lambda) = 1$, il s'agit de voir que $h(\lambda) \geq |\log \lambda| = -\log \lambda$. Grâce à (4.11), c'est vrai si $c_0 \min(\log 2, \log M_\lambda) \geq -\log \lambda$, ce qui est exactement l'hypothèse du théorème. \square

Enfin, on en déduit que tout paramètre avec $\text{Dim}(\mu_\lambda) < 1$ est extrêmement bien approché par des zéros de polynômes à coefficients dans $\{-1, 0, 1\}$: cela traduit le fait que des sommes $\sum \xi_n \lambda^n$ sont parfois extrêmement proches (sans quoi l'entropie serait maximale et la dimension serait donc 1), si bien que λ est presque annulé par un polynôme à coefficients dans $\{-1, 0, 1\}$. On rappelle que l'ensemble de Hochman \mathcal{H} a été défini en (2.2).

THÉORÈME 4.15 (Hochman [Hoc14]). — *Soit λ tel que $\text{Dim}(\mu_\lambda) < 1$. Alors $\lambda \in \mathcal{H}$.*

Démonstration. — Fixons $A > 0$. Le théorème 4.11 donne $H(X_N, e^{-AN}) / (N|\log \lambda|) \rightarrow \text{Dim}(\mu_\lambda)$ puisque $\text{Dim}(\mu_\lambda) < 1$. En particulier, pour tout N assez grand, les valeurs prises par les sommes $X_N = \sum_{n < N} \xi_n \lambda^n$ ne sont pas toutes séparées de e^{-AN} , sans quoi l'entropie précédente serait égale à $N \log 2$, ce qui donnerait une contradiction. En faisant la différence de deux sommes très proches, on obtient le fait suivant : il existe un polynôme P non nul, de degré au plus N , à coefficients dans $\{-1, 0, 1\}$, tel que $|P(\lambda)| \leq e^{-AN}$. On peut sans perte de généralité prendre P unitaire.

Soit $\varepsilon < 1/4$ assez petit pour que $|\lambda| < 1 - 2\varepsilon$. Admettons pour l'instant le fait qu'il existe $k = k(\varepsilon)$, indépendant de N et de P , tel que

$$(4.12) \quad \text{Card}\{z : P(z) = 0 \text{ et } \varepsilon \leq |z| \leq 1 - \varepsilon\} \leq k.$$

On écrit $P(z) = \prod_i (z - \lambda_i) \cdot \prod_j (z - \mu_j)$, où les λ_i sont les racines de P de module dans $[\varepsilon, 1 - \varepsilon]$ (il y en a au plus k d'après (4.12)) et les μ_j sont les autres racines (il y en a au plus N , et elles sont toutes distantes de λ d'au moins ε). On a

$$\prod_i |\lambda - \lambda_i| = |P(\lambda)| / \prod_j |\lambda - \mu_j| \leq e^{-AN} / \varepsilon^N = e^{-(A-C)N},$$

pour $C = |\log \varepsilon|$. En particulier, l'un des λ_i vérifie $|\lambda - \lambda_i| \leq e^{-(A-C)N/k}$. Cela donne $d(\lambda, \mathcal{Z}_N) \leq e^{-(A-C)N/k}$ pour N assez grand. Comme le coefficient de N dans l'exponentielle peut être rendu arbitrairement grand en augmentant A , cela montre que λ appartient à \mathcal{H} par définition de cet ensemble.

Il reste à vérifier (4.12). Quitte à diviser $P(z)$ par la plus grande puissance de z qui le divise, on peut sans perte de généralité supposer que P ne s'annule pas en 0. On a alors $|P(0)| = 1$. Prenons $r \in [1 - \varepsilon/2, 1 - \varepsilon/4]$ tel que P ne s'annule pas sur le cercle de rayon r . Notons $\lambda_1, \dots, \lambda_I$ les zéros de P de module $< r$. La formule de Jensen donne alors

$$\log|P(0)| + \sum_i \log\left(\frac{r}{|\lambda_i|}\right) = \frac{1}{2\pi} \int_0^{2\pi} \log|P(re^{i\theta})| d\theta.$$

Le terme de droite de cette équation est majoré indépendamment de P , puisque $|P(re^{i\theta})| \leq 1/(1-r) \leq 4/\varepsilon$ car les coefficients de P sont tous bornés par 1. Il en va donc de même du terme de gauche. Puisque $|P(0)| = 1$, la somme sur i est donc également majorée. C'est une somme de termes positifs. De plus, chaque zéro de module au plus $1 - \varepsilon$ donne une contribution minorée par $\log(r/(1-\varepsilon)) \geq \log((1-\varepsilon/2)/(1-\varepsilon)) > 0$. Ainsi, le nombre de ces zéros est uniformément borné, ce qui prouve (4.12). \square

5. CROISSANCE DE L'ENTROPIE PAR CONVOLUTION

5.1. Un résultat qualitatif de croissance de l'entropie

Quand on convole deux mesures, on a vu que l'entropie ne pouvait pas décroître dans les lemmes 4.8 et 4.9. Pour démontrer le théorème 4.11, la clé va être de voir que l'entropie a même tendance à augmenter de manière quantitative par convolution. Cette assertion ne peut pas être vraie en général. Par exemple, si on convole une mesure μ avec un Dirac, on retrouve un translaté de μ et l'entropie n'augmente donc pas. À l'opposé, si on convole la mesure de Lebesgue (d'entropie maximale) avec n'importe quelle mesure, on conserve la mesure de Lebesgue, et l'entropie n'augmente pas non plus. Plus généralement, si on regarde l'entropie entre les échelles r et s et qu'on a affaire à des mesures qui, à une certaine échelle entre r et s , ressemblent à un Dirac ou à la mesure de Lebesgue (par exemple une mesure supportée par une progression arithmétique de raison r), alors l'entropie ne va pas augmenter par convolution.

Le théorème suivant affirme que les contre-exemples simplistes qui précèdent forment essentiellement le seul cas de figure où l'entropie ne croît pas : si on considère une mesure μ qui a des trous à toutes les échelles (ce qui exclut la mesure de Lebesgue à

n'importe quelle échelle) et qu'on la convole par une mesure d'entropie positive (pour exclure les masses de Dirac), alors l'entropie va augmenter. Le fait d'avoir des trous à toutes les échelles, même localement, est formalisé par la définition suivante : si l'on conditionne sur les intervalles de longueur $2t$ où μ a beaucoup de masse, alors les restrictions de μ à ces intervalles sont loin d'avoir entropie $\log 2$ à l'échelle t , i.e., elles sont loin d'être uniformes.

DÉFINITION 5.1. — Soit $\delta > 0$. Une mesure de probabilité réelle à support compact μ est dite δ -poreuse si, pour tout t , elle satisfait $H(\mu, t \mid 2t) \leq (1 - \delta) \log 2$.

THÉORÈME 5.2. — Soient $\delta_1, \delta_2 > 0$. Il existe alors $\varepsilon = \varepsilon(\delta_1, \delta_2) > 0$ avec la propriété suivante. Considérons deux échelles $r \leq s$ avec s/r suffisamment grand. Soient μ et ν deux mesures réelles à support compact. On suppose que :

- La mesure μ est δ_1 -poreuse.
- On a $H(\nu, r \mid s) \geq \delta_2 \log(s/r)$.

Alors $H(\mu * \nu, r \mid s) \geq H(\mu, r \mid s) + \varepsilon \log(s/r)$.

L'hypothèse sur ν indique qu'elle ne ressemble pas du tout à un Dirac entre les échelles r et s : elle a une entropie qui n'est pas ridicule (rappelons que l'entropie maximale entre les échelles r et s est $\log(s/r)$). Alors le théorème affirme que l'entropie de $\mu * \nu$ est quantitativement plus grande que celle de μ .

Cet énoncé est démontré (sous une forme légèrement différente) par Hochman dans [Hoc14, Theorem 8]. Sa preuve, délicate, occupe la majeure partie de son article. Nous utiliserons cet énoncé comme une boîte noire, et allons voir qu'il implique le théorème 4.11.

Disons seulement que les techniques de preuve trouvent leur origine dans un domaine *a priori* complètement distinct de celui des convolutions de Bernoulli, la combinatoire additive. Le thème général de ce domaine est, étant données deux parties A et B d'un groupe additif, de comprendre dans quelle mesure $A + B = \{a + b : a \in A, b \in B\}$ est plus grand que A et B . Par exemple, si A est une progression arithmétique $\{kx : k \in \{0, \dots, n-1\}\}$ de cardinal n , alors $A + A$ est de cardinal $2n - 1$, comparable à $\text{Card } A$. Dans le cas où A n'a pas une telle structure, on s'attend plutôt à ce que $A + A$ ait un cardinal de l'ordre de $(\text{Card } A)^2$. Il existe de nombreux énoncés et théorèmes profonds dans cette direction, affirmant que si $\text{Card}(A + A) \leq K \text{Card } A$ où K est une constante fixée, ou si $\text{Card}(A + A) \leq (\text{Card } A)^{1+\delta}$ avec δ assez petit, alors A doit ressembler à une progression arithmétique, ou à une de ses variantes (des progressions arithmétiques à différentes échelles, par exemple). Le théorème 5.2 relève de ce domaine si l'on pense à l'analogie comme quoi l'entropie compte le (logarithme du) cardinal des valeurs typiques de la mesure. En effet, si A représente l'ensemble des valeurs typiques de μ et B celles de ν , les valeurs typiques de $\mu * \nu$ sont essentiellement données par $A + B$. Le théorème affirme bien que $A + B$ est quantitativement plus gros que A , sous certaines hypothèses qui excluent que μ ressemble à une progression arithmétique à quelque échelle que ce soit (c'est l'hypothèse de porosité) et que ν ait trop peu de valeurs typiques (tous les

énoncés de combinatoire additive n’ont de sens que si les ensembles sont suffisamment gros).

L’apport de Hochman est double :

- D’une part, comprendre que des techniques de combinatoire additive pouvaient avoir un rôle à jouer dans l’étude des convolutions de Bernoulli, et que le théorème 5.2 était donc un énoncé pertinent dans ce contexte (on va voir plus bas qu’il implique facilement le théorème 4.11, et donc toutes les conséquences qu’on a décrites dans le paragraphe 4.3).
- D’autre part, démontrer ce théorème 4.11 : les résultats de combinatoire additive classique ne s’appliquent pas directement, et la preuve est difficile (bien qu’élémentaire).

5.2. Démonstration du théorème de Hochman

Pour montrer le théorème 4.11 à partir du théorème 5.2, la première étape est de voir qu’une mesure de dimension < 1 est δ_1 -poreuse pour un certain $\delta_1 > 0$.

LEMME 5.3 (Breuillard-Varjú [BV16]). — *Soit $\lambda \in [1/2, 1[$ tel que $\text{Dim}(\mu_\lambda) < 1$. Alors il existe $\delta > 0$ tel que, pour toute échelle t , on ait $H(\mu_\lambda, t \mid 2t) \leq (1 - \delta) \log 2$. Autrement dit, μ_λ est δ -poreuse.*

Démonstration. — On admet l’estimée suivante sur l’entropie [BV16, Lemma 15] : pour toute mesure réelle μ à support compact, on a

$$(5.1) \quad \log 2 - H(\mu, 2r \mid 4r) \leq 4(\log 2 - H(\mu, r \mid 2r)).$$

Autrement dit, si μ est presque d’entropie maximale entre les échelles r et $2r$, i.e., elle est presque uniforme, alors elle est aussi presque uniforme entre les échelles $2r$ et $4r$ (avec une légère perte).

Supposons que μ_λ n’est pas poreuse, on va voir que $\text{Dim}(\mu_\lambda) = 1$. Fixons un grand entier K . Par hypothèse de non-porosité, on peut choisir t tel que $H(\mu_\lambda, t \mid 2t)$ est très proche de $\log 2$. D’après (5.1), il en va de même de $H(\mu_\lambda, 2^k t \mid 2^{k+1} t)$ pour k fixé. En sommant, on obtient que $H(\mu_\lambda, t \mid 2^K t)$ est proche de $K \log 2$, et donc au moins $K \log 2 - \log 2$. Approchons t et $2^K t$ par deux puissances de λ , soient λ^{a+b} et λ^a , avec une erreur multiplicative au plus 2. En particulier, $K \log 2 = b|\log \lambda| \pm 2 \log 2$. Grâce au lemme 4.7, on obtient

$$H(\mu_\lambda, \lambda^{a+b} \mid \lambda^a) \geq H(\mu_\lambda, t \mid 2^K t) - 4 \log 2 \geq K \log 2 - 5 \log 2 \geq b|\log \lambda| - 7 \log 2.$$

D’après (4.7), si l’on contracte tout d’un facteur λ^{nb} , on obtient la même borne pour $H(\mu_\lambda^{[0, \lambda^{nb}]}, \lambda^{a+(n+1)b} \mid \lambda^{a+nb})$, puisque $\mu_\lambda = \mu_\lambda^{[0, 1]}$. Avec (4.6), cela donne

$$H(\mu_\lambda, \lambda^{a+(n+1)b} \mid \lambda^{a+nb}) \geq b|\log \lambda| - 9 \log 2.$$

En sommant ces inégalités de 0 à $N - 1$, on obtient

$$H(\mu_\lambda, \lambda^{a+Nb} \mid \lambda^a) \geq Nb|\log \lambda| - 9N \log 2.$$

Le terme de gauche de cette équation est équivalent à $Nb|\log \lambda| \text{Dim}(\mu_\lambda)$ lorsque N tend vers l'infini, d'après la proposition 4.5. Ainsi, $\text{Dim}(\mu_\lambda) \geq 1 - 9 \log 2 / (b|\log \lambda|)$. Comme K (et donc b) est arbitrairement grand, cela donne $\text{Dim}(\mu_\lambda) = 1$. \square

On peut maintenant démontrer le théorème 4.11.

Démonstration du théorème 4.11 à partir du théorème 5.2. — Soit A un entier strictement positif. Il s'agit de voir que

$$(5.2) \quad H(\mu_\lambda^{[\lambda^N, 1]}, \lambda^{AN} \mid \lambda^N) = o(N),$$

puisqu'on sait déjà par la proposition 4.6 que $H(\mu_\lambda^{[\lambda^N, 1]}, \lambda^N) / N|\log \lambda| \rightarrow \text{Dim}(\mu_\lambda)$. On décompose l'entropie dans (5.2) comme $\sum_{a=1}^{A-1} H(\mu_\lambda^{[\lambda^N, 1]}, \lambda^{(a+1)N} \mid \lambda^{aN})$, il suffit de voir que chaque terme est $o(N)$.

D'après la proposition 4.5, on a

$$(5.3) \quad \sum_{a=0}^{A-1} H(\mu_\lambda, \lambda^{(a+1)N} \mid \lambda^{aN}) = H(\mu_\lambda, \lambda^{AN}) \sim AN|\log \lambda| \text{Dim}(\mu_\lambda).$$

Comme μ_λ s'écrit comme la convolution de $\mu_\lambda^{[0, \lambda^{aN}]}$ et d'une autre mesure, on a d'après (4.6)

$$(5.4) \quad H(\mu_\lambda, \lambda^{(a+1)N} \mid \lambda^{aN}) \geq H(\mu_\lambda^{[0, \lambda^{aN}]}, \lambda^{(a+1)N} \mid \lambda^{aN}) - 2 \log 2 \\ = H(\mu_\lambda, \lambda^N \mid 1) - 2 \log 2 \sim N|\log \lambda| \text{Dim}(\mu_\lambda),$$

où l'égalité s'obtient en dilatant tout d'un facteur λ^{-aN} comme en (4.7), ce qui envoie $\mu_\lambda^{[0, \lambda^{aN}]}$ sur μ_λ par auto-similarité, et où le dernier équivalent vient de la proposition 4.5. En sommant les termes extrêmes de (5.4), on retrouve les termes de (5.3). Cela montre que l'inégalité dans (5.4) ne peut pas être trop stricte : on a

$$(5.5) \quad H(\mu_\lambda, \lambda^{(a+1)N} \mid \lambda^{aN}) = H(\mu_\lambda^{[0, \lambda^{aN}]}, \lambda^{(a+1)N} \mid \lambda^{aN}) + o(N).$$

Autrement dit, l'entropie de μ_λ entre les échelles $\lambda^{(a+1)N}$ et λ^{aN} vient de son facteur $\mu_\lambda^{[0, \lambda^{aN}]}$, et les autres facteurs ne contribuent donc pas à l'entropie. Cela va permettre de conclure que $\mu_\lambda^{[\lambda^N, 1]}$ n'a presque pas d'entropie entre ces échelles, comme annoncé, grâce au théorème 5.2.

Écrivons μ_λ comme $\mu_\lambda^{[0, \lambda^{aN}]} * \mu_\lambda^{[\lambda^{aN}, \lambda^N]} * \mu_\lambda^{[\lambda^N, 1]}$. Notons que la mesure $\mu_\lambda^{[0, \lambda^{aN}]}$ (similaire par dilatation à la mesure μ_λ) est δ_1 -poreuse d'après le lemme 5.3. Supposons par l'absurde que $H(\mu_\lambda^{[\lambda^N, 1]}, \lambda^{(a+1)N} \mid \lambda^{aN})$ ne soit pas $o(N)$, elle est donc $\geq \delta_2 N$ pour une infinité de valeurs de N , pour un certain $\delta_2 > 0$. Soit $\varepsilon = \varepsilon(\delta_1, \delta_2)$ donné par le théorème 5.2. On a alors

$$\begin{aligned} H(\mu_\lambda, \lambda^{(a+1)N} \mid \lambda^{aN}) &= H(\mu_\lambda^{[0, \lambda^{aN}]} * \mu_\lambda^{[\lambda^{aN}, \lambda^N]} * \mu_\lambda^{[\lambda^N, 1]}, \lambda^{(a+1)N} \mid \lambda^{aN}) \\ &\geq H(\mu_\lambda^{[0, \lambda^{aN}]} * \mu_\lambda^{[\lambda^N, 1]}, \lambda^{(a+1)N} \mid \lambda^{aN}) - 2 \log 2 \\ &\geq H(\mu_\lambda^{[0, \lambda^{aN}]}, \lambda^{(a+1)N} \mid \lambda^{aN}) - 2 \log 2 + \varepsilon N|\log \lambda|, \end{aligned}$$

où la première inégalité provient de (4.6) et la seconde du théorème 5.2. C'est absurde car cela contredit (5.5). \square

5.3. Versions quantitatives

Si l'on veut obtenir des résultats plus forts, on a besoin de raffiner le théorème 5.2, et en particulier de comprendre comment ε dépend de δ_1 et δ_2 . Dans cette direction, Varjú a démontré dans [Var16] les deux théorèmes quantitatifs suivants :

THÉORÈME 5.4. — *Soient $\alpha \in]0, 1/2]$ et $r > 0$. On considère deux mesures réelles μ et ν à support compact, telles que $H(\mu, s \mid 2s) \geq \log 2 - \alpha$ et $H(\nu, s \mid 2s) \geq \log 2 - \alpha$, pour tout s avec $|\log s - \log r| < 3|\log \alpha|$. Alors*

$$H(\mu * \nu, r \mid 2r) \geq \log 2 - 10^9 |\log \alpha|^3 \alpha^2.$$

Autrement dit, si deux mesures ont des entropies presque maximales, leur convolée est encore plus uniforme, le défaut d'entropie se comportant presque multiplicativement.

THÉORÈME 5.5. — *Soient $\delta_1, \delta_2 \in]0, 1/2]$ et $r < s$. On considère deux mesures réelles μ et ν à support compact. On suppose que $H(\mu, t \mid 2t) \leq \log 2(1 - \delta_1)$ pour tout $t \in [r, s]$ (c'est une hypothèse de porosité sur μ) et que $H(\nu, r \mid s) \geq \delta_2 \log(s/r)$ (c'est une hypothèse d'entropie non négligeable sur ν). Alors*

$$H(\mu * \nu, r \mid s) \geq H(\mu, r \mid s) + 10^{-8} \frac{\delta_1}{|\log \delta_1|} \frac{\delta_2}{|\log \delta_2|} \log(s/r) - 3 \log 2.$$

Ainsi, dans le théorème 5.2, on peut prendre $\varepsilon(\delta_1, \delta_2)$ presque linéaire en δ_1 et δ_2 .

Ces théorèmes permettent d'aller plus loin que le théorème 4.11 de Hochman. Ils sont utilisés dans la démonstration des résultats suivants :

THÉORÈME 5.6 (Varjú [Var16]). — *Pour tout $\varepsilon > 0$, il existe $c(\varepsilon)$ avec la propriété suivante. Soit $\lambda \in [1/2, 1[$ algébrique sur \mathbb{Q} avec*

$$\lambda > 1 - c(\varepsilon) \min(\log M_\lambda, (\log M_\lambda)^{-1-\varepsilon}).$$

Alors μ_λ est absolument continue.

THÉORÈME 5.7 (Breuillard-Varjú [BV16]). — *Soit $\lambda \in [1/2, 1[$ tel que $\text{Dim}(\mu_\lambda) < 1$. Il existe alors une suite $n_i \rightarrow \infty$ et une suite d'approximations $\lambda_{n_i} \in \mathcal{Z}_{n_i} \cap [1/2, 1[$ telles que $|\lambda - \lambda_{n_i}| \leq e^{-n_i^{\log \log \log n_i}}$ et $\text{Dim}(\mu_{\lambda_{n_i}}) \rightarrow \text{Dim}(\mu_\lambda)$.*

On rappelle que l'ensemble \mathcal{Z}_n est formé des zéros des polynômes de degré au plus n à coefficients dans $\{-1, 0, 1\}$.

Les preuves de ces résultats reposent fondamentalement sur des idées similaires à ce qu'on a déjà exposé : écrire μ_λ comme le produit de convolution d'un certain nombre de mesures $\mu_\lambda^{I_k}$ (où les I_k forment une partition de $[0, 1]$), et contrôler l'entropie des $\mu_\lambda^{I_k}$ et/ou de μ_λ en utilisant l'auto-similarité d'une part, et les contrôles de croissance de l'entropie d'autre part. Les décompositions à utiliser sont toutefois beaucoup plus

sophistiquées que celles que nous avons expliquées plus haut. Nous ne rentrerons pas dans les détails, et renvoyons le lecteur vers les articles originaux.

6. UNE BORNE INFÉRIEURE RIGOUREUSE SUR LA DIMENSION DES CONVOLUTIONS DE BERNOULLI

La proposition suivante utilise toute la force des énoncés précédents, en particulier les théorèmes 4.12 et 5.7.

PROPOSITION 6.1 (Hare et Sidorov [HS16]). — *Pour tout $\lambda \in [1/2, 1[$, on a $\text{Dim}(\mu_\lambda) \geq 0,82$.*

Esquisse de démonstration. — L'entropie provient d'un phénomène sous-additif comme énoncé dans la proposition 3.1. Obtenir des bornes supérieures sur l'entropie (et la dimension) est donc facile (il suffit de regarder ce qui se passe en un temps fixé), mais il est beaucoup plus délicat d'obtenir des bornes *inférieures*, comme dans cette proposition.

Notons tout d'abord qu'il suffit de montrer, pour tout $\lambda \in [1/2, 1[$, l'inégalité

$$(6.1) \quad h(\lambda)/|\log \lambda| \geq 0,82.$$

En effet, si cette inégalité est vraie, on déduit que pour λ algébrique on a $\text{Dim}(\mu_\lambda) \geq 0,82$ grâce à la formule $\text{Dim}(\mu_\lambda) = \min(1, h(\lambda)/|\log \lambda|)$ donnée dans le théorème 4.12 pour les paramètres algébriques. L'inégalité $\text{Dim}(\mu_\lambda) \geq 0,82$ s'étend ensuite aux paramètres non algébriques en les approchant par des paramètres algébriques de dimension proche, grâce au théorème 5.7.

Expliquons maintenant comment montrer (6.1). Notons $M_N(\lambda)$ le suprémum en x de

$$\text{Card} \left\{ (\varepsilon_0, \dots, \varepsilon_{N-1}) \in \{-1, 1\}^N : \sum_{i < N} \varepsilon_i \lambda^i \in [x - \lambda^N/(1 - \lambda), x + \lambda^N/(1 - \lambda)] \right\}.$$

Si N et λ sont fixés, cette quantité peut être calculée numériquement. Avec de l'arithmétique d'intervalles, on peut même estimer numériquement $\sup_\lambda M_N(\lambda)$.

On peut aussi utiliser cette quantité pour borner inférieurement l'entropie. Montrons d'abord par récurrence sur n que, pour tout x ,

$$(6.2) \quad \text{Card}\{(\varepsilon_i)_{i < nN} \in \{-1, 1\}^{nN} : \sum_{i < nN} \varepsilon_i \lambda^i = x\} \leq M_N(\lambda)^n.$$

Cette inégalité est claire pour $n = 0$. Montrons-la pour $n > 0$. Si $\sum_{i < nN} \varepsilon_i \lambda^i = x$, alors $\sum_{i < N} \varepsilon_i \lambda^i \in [x - \lambda^N/(1 - \lambda), x + \lambda^N/(1 - \lambda)]$ puisque les termes restants donnent une erreur de au plus $\lambda^N/(1 - \lambda)$. Il y a donc au maximum $M_N(\lambda)$ préfixes possibles de longueur N . Un tel préfixe $(\varepsilon_0, \dots, \varepsilon_{N-1})$ étant fixé, ses prolongements vérifient

$$\sum_{N \leq i < nN} \varepsilon_i \lambda^i = x - \sum_{i < N} \varepsilon_i \lambda^i.$$

D'après l'hypothèse de récurrence appliquée à $\lambda^{-N}(x - \sum_{i < N} \varepsilon_i \lambda^i)$, il y a au plus $M_N(\lambda)^{n-1}$ prolongements possibles. Cela termine la démonstration de (6.2).

On en déduit que les atomes de la partition \mathcal{Q}_{nN} où X_{nN} est constante ont tous mesure au plus $2^{-nN} M_N(\lambda)^n = e^{-nN(\log 2 - \log M_N(\lambda)/N)}$. Comme un atome typique a mesure $e^{-nNh(\lambda)}$ d'après la proposition 3.1, on obtient

$$h(\lambda) \geq \log 2 - \log M_N(\lambda)/N.$$

Vu la définition de $M_N(\lambda)$, on a trivialement $M_N(\lambda) \leq 2^N$, ce qui donne la borne triviale $h(\lambda) \geq 0$. Si l'on arrive à obtenir de meilleures estimées sur $M_N(\lambda)$, on va améliorer cette borne. Comme $M_N(\lambda)$ peut être estimé numériquement, cela donne une stratégie implémentable sur ordinateur pour démontrer l'estimée (6.1). C'est ainsi que Hare et Sidorov procèdent dans [HS16]. Il faut encore être soigneux et astucieux pour que la complexité de l'estimation de $M_N(\lambda)$ reste raisonnable, je renvoie à leur article pour les détails de leur implémentation (ils ont besoin d'aller jusqu'à $N = 14$ pour certains intervalles de paramètres). \square

Remarque 6.2. — En théorie, on pourrait aussi étudier la fonction f_k définie en (4.8) pour trouver une borne inférieure pour l'entropie, puisque f_k converge en croissant vers l'entropie comme on l'a vu dans la preuve de la proposition 4.8. À ma connaissance, cela n'a pas été fait.

RÉFÉRENCES

- [BDGGH⁺92] M.-J. BERTIN, A. DECOMPS-GUILLOUX, M. GRANDET-HUGOT, M. PATHIAUX-DELEFOSSE & J.-P. SCHREIBER – *Pisot and Salem numbers*, Birkhäuser Verlag, Basel, 1992, avec une préface de David W. Boyd.
- [BV15] E. BREUILLARD & P. P. VARJÚ – « Entropy of Bernoulli convolutions and uniform exponential growth for linear groups », prépublication, 2015.
- [BV16] ———, « On the dimension of Bernoulli convolutions », prépublication, 2016.
- [Erd39] P. ERDŐS – « On a family of symmetric Bernoulli convolutions », *Amer. J. Math.* **61** (1939), p. 974–976.
- [Erd40] ———, « On the smoothness properties of a family of Bernoulli convolutions », *Amer. J. Math.* **62** (1940), p. 180–186.
- [Fal90] K. FALCONER – *Fractal geometry*, John Wiley & Sons, Ltd., Chichester, 1990, Mathematical foundations and applications.
- [FH09] D.-J. FENG & H. HU – « Dimension theory of iterated function systems », *Comm. Pure Appl. Math.* **62** (2009), no. 11, p. 1435–1500.
- [Gar63] A. M. GARSIA – « Entropy and singularity of infinite convolutions », *Pacific J. Math.* **13** (1963), p. 1159–1169.

- [Hoc14] M. HOCHMAN – « On self-similar sets with overlaps and inverse theorems for entropy », *Ann. of Math. (2)* **180** (2014), no. 2, p. 773–822.
- [HS16] K. G. HARE & N. SIDOROV – « A lower bound for the dimension of Bernoulli convolutions », prépublication, à paraître dans *Exp. Math.*, 2016.
- [JW35] B. JESSEN & A. WINTNER – « Distribution functions and the Riemann zeta function », *Trans. Amer. Math. Soc.* **38** (1935), no. 1, p. 48–88.
- [Kah71] J.-P. KAHANE – « Sur la distribution de certaines séries aléatoires », *Bull. Soc. Math. France* **25** (1971), p. 119–122.
- [Led92] F. LEDRAPPIER – « On the dimension of some graphs », in *Symbolic dynamics and its applications (New Haven, CT, 1991)*, *Contemp. Math.*, vol. 135, Amer. Math. Soc., Providence, RI, 1992, p. 285–293.
- [Mah64] K. MAHLER – « An inequality for the discriminant of a polynomial », *Michigan Math. J.* **11** (1964), p. 257–262.
- [PS00] Y. PERES & W. SCHLAG – « Smoothness of projections, Bernoulli convolutions, and the dimension of exceptions », *Duke Math. J.* **102** (2000), no. 2, p. 193–251.
- [PSS00] Y. PERES, W. SCHLAG & B. SOLOMYAK – « Sixty years of Bernoulli convolutions », in *Fractal geometry and stochastics, II (Greifswald/Koserow, 1998)*, *Progr. Probab.*, vol. 46, Birkhäuser, Basel, 2000, p. 39–65.
- [Shm14] P. SHMERKIN – « On the exceptional set for absolute continuity of Bernoulli convolutions », *Geom. Funct. Anal.* **24** (2014), no. 3, p. 946–958.
- [Shm16] P. SHMERKIN – « On Furstenberg’s intersection conjecture, self-similar measures, and the l^q norms of convolutions », prépublication, 2016.
- [Sol95] B. SOLOMYAK – « On the random series $\sum \pm \lambda^n$ (an Erdős problem) », *Ann. of Math. (2)* **142** (1995), no. 3, p. 611–625.
- [Var16] P. P. VARJÚ – « Absolute continuity of bernoulli convolutions for algebraic parameters », prépublication, 2016.

Sébastien GOUËZEL

Université de Nantes
 Laboratoire Jean Leray
 2 rue de la Houssinière
 B.P. 92208

44322 Nantes Cedex 3

E-mail : sebastien.gouezel@univ-nantes.fr