Mémoire 2 : Théorie d'Ornstein et propriété de Pinsker faible

Séverin Benzoni

Encadrant : Thierry De La Rue

Du 11 Avril au 16 Septembre 2020

Suite à un problème technique, j'ai perdu le script d'une partie de mon mémoire (Chapire 1), donc j'ai du découper mon mémoire en deux chapitres. Pour cette raison, chaque chapitre a ses propres sommaire et bibliographie. Malheureusement il y a quelques erreurs dans le Chapitre 1 que je n'ai pas pu corriger et je m'en excuse. Je souhaiterait tout même donner ici la correction d'une faute de frappe : [Chapitre 1, Lemme 3.14, (iii)] devrait être :

$$\text{(iii) } \rho_0 \preccurlyeq (\pi' \vee \xi')_{[-N,N]}.$$

Chapitre 1 : Entropie et Systèmes dynamiques

Séverin Benzoni

Encadrant : Thierry De La Rue

Du 11 Avril au 25 Juin 2020

Sommaire

1 Carde formel du mémoire

La théorie ergodique est le cadre abstrait de l'étude des systèmes dynamiques mesurés : on se donne un ensemble X muni d'une tribu $\mathcal{F}$ et d'une mesure de probabilité μ sur $\mathcal{F}$ (dans tout ce mémoire, toutes les mesures sont des mesures de probabilité). Sur cet espace de probabilité, vient agir une transformation T qui est une application mesurable, inversible et d'inverse mesurable. On suppose que T préserve la mesure : pour tout $A \in \mathcal{F}$, $\mu(T^{-1}(A)) = \mu(A)$. Autrement dit, la mesure image de μ par T, $T_*\mu$, est égale à μ.

On appellera *système dynamique* ou simplement *automorphisme* un espace $(X, \mathcal{F}, \mu, T)$ tel que $(X, \mathcal{F}, \mu)$ soit un espace de Lebesgue, et que T soit une transformation inversible, mesurable, d'inverse mesurable et qui préserve la mesure. Par la suite, sauf risque de confusion, on écrira simplement (X, μ, T) pour désigner un système dynamique.

Définition 1.1. Soient $(X, \mathcal{F}, \mu, T)$ et $(Y, \mathcal{G}, \nu, S)$ deux systèmes dynamiques et $\pi : X \to Y$ vérifiant :

1. l'application π est mesurable,
2. la mesure sur Y est $\nu = \pi_*\mu$,
3. et $\pi \circ T = S \circ \pi$

On dit alors (Y, ν, S) est un *facteur* de (X, μ, T) et que π est une *application facteur*. On écrira souvent une application facteur sous la forme $\pi : (X, \mu, T) \longrightarrow (Y, \nu, S)$. Si π est une bijection et π^{-1} est aussi une application facteur, alors c'est un *isomorphisme* de systèmes dynamiques.

Comme il y a une bijection naturelle entre les sous-tribus T-invariantes de $\mathcal{F}$ et les facteurs de $(X, \mathcal{F}, \mu, T)$, on appellera aussi facteur toute tribu $\mathcal{H} \subset \mathcal{F}$ T-invariante. On notera $\sigma(\pi)$ la sous-tribu engendrée par l'application facteur π. On dira qu'une fonction sur X est π-mesurable si elle est $\sigma(\pi)$-mesurable.

Pour deux applications facteur π et φ, on définit l'application jointe par la formule :

$$\pi \vee \varphi(x) = (\pi(x), \varphi(x)),$$

(où $\pi \vee \varphi$ se lit "π *et* φ").

Présentons maintenant une classe de facteurs qui seront centraux pour la théorie de l'entropie. Pour un ensemble fini ou infini A, on appellera *processus stationnaire* sur l'*alphabet* A toute application facteur de la forme :

$$\xi : (X, \mathcal{F}, \mu, T) \to (A^{\mathbb{Z}}, \mathcal{C}, \xi_*\mu, T_A)$$

avec

$$T_A((a_n)_{n\in\mathbb{Z}}) = (a_{n+1})_{n\in\mathbb{Z}},$$

et où $\mathcal{C}$ est la tribu cylindrique. C'est-à-dire la tribu engendrée par les cylindres :

$$C_n^m(A_n, ..., A_m) = \{\mathbf{x} \in A^{\mathbb{Z}} | (x_k)_{k=n}^m \in A_n \times \cdots \times A_m\},$$

avec, pour $i \in \{n, ..., m\}$, $A_i \subset A$ et A_i est mesurable.

En écrivant $\xi = (\xi_n)_{n\in\mathbb{Z}}$, comme ξ est une application facteur, on a :

$$\forall n, \xi_{n+1} = \xi_n \circ T,$$

ce qui donne :

$$\forall n, \xi_n = \xi_0 \circ T^n$$

Le processus est donc complètement déterminé par la variable aléatoire ξ_0. Et réciproquement, toute variable aléatoire ξ_0 engendre un processus via la formule ci-dessus. On l'écrit alors :

$$\xi := \bigvee_{n=-\infty}^{\infty} \xi_n = \bigvee_{n=-\infty}^{\infty} \xi_0 \circ T^n.$$

Pour une partie $I \subset \mathbb{Z}$ on notera ξ_I la restriction de ξ à I, c'est-à-dire

$$\xi_I = \bigvee_{n\in I} \xi_i.$$

On remarque que si ξ est le processus engendré par ξ_0 et ζ est celui engendré par ζ_0, alors $\xi \vee \zeta$ est le processus engendré par $\xi_0 \vee \zeta_0$.

Lorsque l'on regarde la structure d'un processus stationnaire, on remarque que, une fois que l'alphabet A est fixé, celui-ci est totalement caractérisé par la mesure $\xi_*\mu$. Si l'on veut construire un tel processus, il suffit donc de définir une mesure sur $A^{\mathbb{Z}}$ invariante par T_A.

Donnons ici un exemple de construction. Fixons une mesure de probabilité sur A, que l'on appelle γ. On peut alors définir une mesure sur les cylindres via la formule

$$\nu(C_n^m(A_n, ..., A_m)) = \prod_{i=0}^{m-n+1} \gamma(A_i),$$

où $A_n \times \cdots \times A_m \subset A^{m-n+1}$. Le théorème d'extension de Kolmogorov affirme qu'il existe une unique mesure sur $A^{\mathbb{Z}}$ vérifiant la formule ci-dessus pour tous les cylindres. On l'appelle la *mesure produit* et on la note $\nu = \gamma^{\otimes \mathbb{Z}}$. On voit facilement que sur les cylindres, on a l'égalité

$$\nu(C_{n+1}^{m+1}(A_n, .., A_m)) = \nu(C_n^m(A_n, ..., A_m)),$$

ce qui, par l'unicité de l'extension, prouve que la mesure produit est T_A-invariante. On obtient donc le système dynamique : $(A^{\mathbb{Z}}, \mathcal{C}, \gamma^{\otimes \mathbb{Z}}, T_A)$. Tous les systèmes dynamiques construits ainsi sont appelés *schémas de Bernoulli*.

On remarque que la suite $(\xi_n : \mathbf{a} \mapsto a_n)$ est une suite de variable aléatoires $i.i.d.$ de loi γ. Donc si on prend, par exemple $\gamma = \mathcal{B}(\frac{1}{2}, \frac{1}{2})$, le schéma de Bernoulli est une représentation dynamique d'un objet bien connu en probabilité, un jeu de pile ou face. Le fait que tous les lancés de pièce soient indépendants implique que, à chaque étape, même en ayant connaissance de tous les lancés précédents, il est impossible de prédire le résultat que donnera le lancé suivant. Pour cette raison, les schémas de Bernoulli sont les systèmes les plus aléatoires que l'on puisse construire. Pour mesurer ce aspect "aléatoire" des systèmes dynamiques et donner un sens quantitatif à cette interprétation qualitative, à la fin des années soixante, Kolmogorov et Sinai, inspirés par la théorie de l'information développée par C. Shannon, ont introduit un nouveau concept qui a révolutionné la théorie ergodique : l'**entropie de Kolmogorov-Sinai**.

2 Introduction

2.1 L'entropie en théorie ergodique

2.1.1 L'entropie de Shannon

La notion d'entropie est apparue pour la première fois en mathématiques en 1948, introduite par Claude Shannon dans ses travaux fondateurs de la théorie de l'information [6]. On la définit ici sous une forme adaptée à son utilisation en théorie ergodique.

Soit (X, μ) un espace de probabilité, A un ensemble quelconque, et $\xi : X \to A$ une variable aléatoire à valeur dans A (en théorie de l'information, elle représenterait le signal à traiter). On définit alors l'entropie de Shannon de ξ :

$$H_\mu(\xi) = -\int_X \log(\mu(\{\xi = \xi(x)\}))d\mu(x).$$

Sur un ensemble A fini, cela devient :

$$H_\mu(\xi) = -\sum_{a \in A} \mu(\{\xi = a\}) \log(\mu(\{\xi = a\})).$$

(Lorsque le contexte est suffisament clair, on écrira $H := H_\mu$.) Il peut aussi être utile de définir l'entropie d'une partition finie $\mathcal{P}$ de X :

$$H_\mu(\mathcal{P}) = -\sum_{P \in \mathcal{P}} \mu(P) \log(\mu(P)).$$

Définissons aussi l'entropie conditionnelle : si $\mathcal{A}$ est une sous-tribu de $\mathcal{F}$

$$H_\mu(\xi|\mathcal{A}) = -\sum_{a \in A} \int \mathbb{E}[\mathbb{1}_{\{\xi=a\}}|\mathcal{A}] \log(\mathbb{E}[\mathbb{1}_{\{\xi=a\}}|\mathcal{A}])d\mu.$$

Cette définition permet de définir le conditionnement par rapport à une variable aléatoire ou part rapport à une partition, en prenant la tribu engendrée. Par exemple :

$$H_\mu(\xi \,|\, \pi) := H_\mu(\xi \,|\, \sigma(\pi)).$$

On introduit un ordre partiel sur les variables aléatoires : on note $\xi \preccurlyeq \eta$ et on dit que η est plus fine que ξ, si ξ est η-mesurable. On remarque alors que $\xi \mapsto H_\mu(\xi \,|\, \pi)$ est une fonction croissante et $\pi \mapsto H_\mu(\xi \,|\, \pi)$ est décroissante.

La propriété fondamentale est la formule suivante (appelée “chain rule” en anglais) :

$$H(\xi \vee \eta \,|\, \pi) = H(\xi \,|\, \pi) + H(\eta \,|\, \xi \vee \pi), \tag{1}$$

pour trois variables aléatoires ξ, η et π à valeurs dans des ensembles finis.

Notons ici un autre résultat utile : Soit $(\mathcal{F}_k)_{k \geq 0}$ une suite croissante de partitions, et notons $\mathcal{F}_\infty = \sigma((\mathcal{F}_k)_{k \geq 0})$. La théorie des martingales permet de conclure que :

$$H(\xi \,|\, \mathcal{F}_k) \underset{k \to \infty}{\longrightarrow} H(\xi \,|\, \mathcal{F}_\infty) \tag{2}$$

Remarque 2.1. On voit que, grâce à (2), on peut remplacer π dans (1) par une sous-tribu quelconque.

2.1.2 *Entropie d'un système dynamique*

En 1958, Kolmogorov et Sinai eurent l'intuition pour d'introduire le concept d'entropie en systèmes dynamiques. On présente ici une notion plus générale d'entropie relative qui servira plus tard.

Soit $(\xi_n)_{n\in\mathbb{Z}}$ un processus stationnaire sur (X,μ,T), et π une application facteur. On définit l'entropie de Kolmogorov-Sinai relative (ou KS-entropie relative) de ξ sachant π comme :

$$h(\xi,\mu,T\,|\,\pi) = \lim_{n\to\infty} \frac{1}{n} H_\mu \left(\bigvee_{i=0}^{n-1} \xi_i \,|\, \pi \right).$$

Ensuite, on prend une application facteur quelconque φ, on définit la KS-entropie relative de φ sachant π par

$$h(\varphi,\mu,T\,|\,\pi) = \sup\{h(\xi,\mu,T\,|\,\pi)\,;\,\xi \text{ processus sur un alphabet fini } \varphi\text{-mesurable}\}.$$

Donnons le théorème suivant, qui prouve que les deux définitions ci-dessus sont compatibles lorsqu'on prend $\varphi=\xi$. C'est une formulation, adaptée à l'entropie relative, du théorème de Kolmogorov-Sinai.

Théorème 2.1. *Soient π une application facteur, et ξ un processus stationnaire sur $(X,\mathcal{F},\mu)$. On a*

$$h(\xi,\mu,T\,|\,\pi) = \sup\{h(\eta,\mu,T\,|\,\pi)\,;\,\eta \text{ processus sur un alphabet fini } \xi\text{-mesurable}\}.$$

Démonstration. Pour tout $n\ge 0$, on introduit la variable aléatoire

$$\xi_0^{(n)} = \bigvee_{i=-n}^{n} \xi_i \text{ et } \xi^{(n)} \text{ le processus associé.}$$

On vérifie que $h(\xi^{(n)},\mu,T\,|\,\pi) = h(\xi,\mu,T\,|\,\pi)$, et on se donne un processus η ξ-mesurable. Alors (1) et la monotonicité de l'entropie donnent

$$h(\eta,\mu,T\,|\,\pi) \le h(\xi^{(n)},\mu,T\,|\,\pi) + H(\eta_0\,|\,\xi_0^{(n)}) = h(\xi,\mu,T\,|\,\pi) + H(\eta_0\,|\bigvee_{i=-n}^{n}\xi_i).$$

Et donc, grâce à (2), $n\to\infty$ donne

$$h(\eta,\mu,T\,|\,\pi) \le h(\xi,\mu,T\,|\,\pi),$$

car, η étant ξ-mesurable, $H(\eta_0\,|\,\xi)=0$. $\square$

Remarque 2.2. Si on prend $\varphi = id_X$ et π trivial, on retrouve la KS-entropie habituelle du système

$$h(\mu, T) := \sup\{h(\xi, \mu, T)\,;\, \xi \text{ processus sur } (X, \mu, T)\}.$$

Pour manipuler ce nouvel objet, il est utile d'avoir les résultats suivants, que l'on peut voir comme des propriétés de continuité de la KS-entropie.

Lemme 2.2. *Soit $(\mathcal{F}_k)_{k\in\mathbb{N}}$ une suite croissante de tribus T-invariantes, et notons $\mathcal{F}_\infty = \sigma((\mathcal{F}_k)_{k\geq 0})$. Alors :*

$$h(\mathcal{F}_k, \mu, T \,|\, \pi) \underset{k\to\infty}{\longrightarrow} h(\mathcal{F}, \mu, T \,|\, \pi).$$

Démonstration. Par définition, on a l'inégalité :

$$\forall k,\ h(\mathcal{F}_k \,|\, \pi) \leq h(\mathcal{F}_\infty \,|\, \pi).$$

Montrons la réciproque : soit ξ_0 un processus $\mathcal{F}_\infty$-mesurable, à valeur dans l'ensemble fini $\{0, ..., p\}$.

La suite $(\mathbb{E}\ [\xi_0|\mathcal{F}_k])_k$ est une martingale positive, donc elle converge p.s., et une application du théorème de convergence dominé nous donne une convergence de cette suite dans $L^1(\mu)$ vers ξ_0. On pose alors

$$\eta_0^k := \left\lfloor \mathbb{E}\,[\xi_0|\mathcal{F}_k] + \frac{1}{2} \right\rfloor,$$

et on remarque qu'en utilisant l'inégalité de Markov, on a

$$\begin{aligned} \mu(\eta_0^k \neq \xi_0) &= \mu(|\mathbb{E}\,[\xi_0|\mathcal{F}_k] - \xi_0| \geq \frac{1}{2}) \\ &\leq 2 \int |\mathbb{E}\,[\xi_0|\mathcal{F}_k] - \xi_0|\, d\mu \underset{k\to\infty}{\longrightarrow} 0 \end{aligned}$$

Soit $\varepsilon > 0$. Par continuité, il existe $\delta > 0$ tel que $x \in [0, \delta[\, \cup\,]1 - \delta] \Rightarrow x\log(x) \leq \varepsilon$. On fixe ensuite $k_0 \in \mathbb{N}$ tel que, pour tout $k \geq k_0$ et pour tout q, on a : $\mu(\eta_0^k \neq \xi_0) \leq \mu(\xi_0 = q)\,\delta$. On regarde alors le processus associé η^k.

En utilisant la remarque 2.1, on a

$$\begin{aligned}
h(\xi\,|\,\pi) - h(\eta^k\,|\,\pi) &\le h(\xi \vee \eta^k\,|\,\pi) - h(\eta^k\,|\,\pi) \\
&= \lim \frac{1}{n}\left[H\left(\xi_{[0,n[} \vee \eta^k_{[0,n[}\,|\,\pi\right) - H\left(\eta^k_{[0,n[}\,|\,\pi\right)\right] \\
&= \lim \frac{1}{n} H\left(\xi_{[0,n[}\,|\,\eta^k_{[0,n[} \vee \pi\right) \\
&\le \lim \frac{1}{n} H\left(\xi_{[0,n[}\,|\,\eta^k_{[0,n[}\right) \le H(\xi_0\,|\,\eta^k_0) \\
&= \sum_{q=0}^{p} \mu(\xi_0 = q) \sum_{q'=0}^{p} \mu_{(\xi_0=q)}(\eta^k_0 = q') \log\left(\mu_{(\xi_0=q)}(\eta^k_0 = q')\right).
\end{aligned}$$

Enfin, on voit que :

- $\forall q \neq q',\ \mu_{(\xi_0=q)}(\eta^k_0 = q') \le \frac{\mu(\xi_0 \neq \eta^k_0)}{\mu(\xi_0=q)} \le \delta$

- $\forall q \in \{0, ..., p\},\ \mu_{(\xi_0=q)}(\eta^k_0 = q) \ge 1 - \frac{\mu\left(\{\xi_0=q\}\Delta\{\eta^k_0=q\}\right)}{\mu(\xi_0=q)} \ge 1-\delta.$

Donc, on a

$$h(\xi\,|\,\pi) - h(\eta^k\,|\,\pi) \le (p+1)\varepsilon,$$

soit, pour tout $k \ge k_0$,

$$h(\xi\,|\,\pi) \le h(\eta^k\,|\,\pi) + (p+1)\varepsilon \le h(\mathcal{F}_k\,|\,\pi) + (p+1)\varepsilon.$$

On obtient le résultat souhaité en faisant $k \to \infty$ puis $\varepsilon \to 0$. □

On montre ensuite qu'une variante de (2) est aussi vraie pour la KS-entropie relative.

Lemme 2.3. *Soient ξ un processus, $(\mathcal{F}_k)_{k\in\mathbb{N}}$ une suite croissante de tribus T-invariantes, et $\mathcal{F}_\infty = \sigma((\mathcal{F}_k)_{k\ge 0})$:*

$$h(\xi, \mu, T\,|\,\mathcal{F}_k) \underset{k\to\infty}{\longrightarrow} h(\xi, \mu, T\,|\,\mathcal{F}_\infty).$$

Démonstration. On va simplement utiliser (2) :

$$\begin{aligned}
h(\xi, \mu, T \,|\, \mathcal{F}_\infty) &= \inf_n \frac{1}{n} H_\mu \left(\xi_{[0,n[} \,|\, \mathcal{F}_\infty\right) \\
&= \inf_n \frac{1}{n} \inf_k H_\mu \left(\xi_{[0,n[} \,|\, \mathcal{F}_k\right) \\
&= \inf_k \inf_n \frac{1}{n} H_\mu \left(\xi_{[0,n[} \,|\, \mathcal{F}_k\right) \\
&= \inf_k h(\xi, \mu, T \,|\, \mathcal{F}_k)
\end{aligned}$$

□

On reprend les résultats de [2, Section 10] en ajoutant les preuves, et celles-ci proviennent en grande partie de [4]. On commence par le [2, Lemma 10.2] :

Lemme 2.4. *Si $\pi : X \longrightarrow B^{\mathbb{Z}}$ est un processus, alors*

$$H \left(\xi_{[0,n[} \,|\, \pi_{[0,n[}\right) \geq H \left(\xi_{[0,n[} \,|\, \pi\right),$$

pour tout n, et on a

$$\lim_n \frac{1}{n} H \left(\xi_{[0,n[} \,|\, \pi_{[0,n[}\right) = h(\xi, \mu, T \,|\, \pi).$$

Si π est une application facteur quelconque, alors

$$h(\xi, \mu, T \,|\, \pi) = \inf\{h(\xi, \mu, T \,|\, \pi') \,;\, \pi' \textit{ processus } \pi\textit{-mesurable}\}.$$

Démonstration. Comme $\pi_{[0,n[}$ et π-mesurable, on a

$$H \left(\xi_{[0,n[} \,|\, \pi_{[0,n[}\right) \geq H \left(\xi_{[0,n[} \,|\, \pi\right).$$

Soit $m, n \in \mathbb{Z}$:

$$\begin{aligned}
H \left(\xi_{[0,n+2m[} \,|\, \pi_{[0,n+2m[}\right) &\leq H \left(\xi_{[0,m[} \,|\, \pi_{[0,m[}\right) + H \left(\xi_{[m,n+m[} \,|\, \pi_{[0,n+2m[}\right) \\
&\quad + H \left(\xi_{[n+m,n+2m[} \,|\, \pi_{[n+m,n+2m[}\right) \\
&\leq 2H \left(\xi_{[0,m[} \,|\, \pi_{[0,m[}\right) + H \left(\xi_{[0,n[} \,|\, \pi_{[-m,n+m[}\right)
\end{aligned}$$

Ainsi, pour tout $m \in \mathbb{Z}$

$$\begin{aligned}\lim_n \frac{1}{n} H\left(\xi_{[0,n[} \,|\, \pi_{[0,n[}\right) &= \lim_n \frac{1}{n} H\left(\xi_{[0,n+2m[} \,|\, \pi_{[0,n+2m[}\right) \\ &\leq \inf_n \frac{1}{n} H\left(\xi_{[0,n[} \,|\, \pi_{[-m,n+m[}\right)\end{aligned}$$

Finalement :

$$\begin{aligned}\lim_n \frac{1}{n} H\left(\xi_{[0,n[} \,|\, \pi_{[0,n[}\right) &\leq \inf_m \inf_n \frac{1}{n} H\left(\xi_{[0,n[} \,|\, \pi_{[-m,n+m[}\right) \\ &= \inf_n \frac{1}{n} \inf_m H\left(\xi_{[0,n[} \,|\, \pi_{[-m,n+m[}\right) \\ &= \inf_n \frac{1}{n} H\left(\xi_{[0,n[} \,|\, \pi\right) = h\left(\xi, \mu, T \,|\, \pi\right),\end{aligned}$$

l'avant dernière égalité découlant de (2).

Soit π une application facteur. Comme π est à valeurs dans un espace de Lebesgue, $\sigma(\pi)$ est engendrée par une famille dénombrable, et on peut donc construire une suite de variables aléatoires à valeur dans des ensembles finis $(\pi_0^k)_{k\geq 0}$ qui engendre $\sigma(\pi)$. Notons alors π^k les processus associés :

$$\pi^k := \bigvee_{n=-\infty}^{\infty} \pi_0^k \circ T^n.$$

Comme π engendre une tribu T-invariante, les processus $\{\pi^k; k \geq 0\}$ sont π-mesurables. De plus, $\forall k \geq 0$, π_0^k est π^k-mesurable, donc $\sigma(\{\pi^k\}_{k\geq 0}) = \sigma(\pi)$. En utilisant 2.3, il en découle

$$\begin{aligned}h(\xi, \mu, T \,|\, \pi) &= \inf_k h(\xi, \mu, T \,|\, \bigvee_{l=0}^{k-1} \pi^l) \\ &\geq \inf\{h(\xi, \mu, T \,|\, \pi')\,;\, \pi' \text{ processus } \pi\text{-mesurable}\}.\end{aligned}$$

□

Ensuite, [2, Lemma 10.3] donne un équivalent de (1) pour la KS-entropie relative.

Lemme 2.5 ("Chain rule")**.** *Pour des applications facteurs π, φ et ψ sur (X, μ, T), on a*

$$\mathrm{h}(\varphi \vee \psi, \mu, T \,|\, \pi) = \mathrm{h}(\varphi, \mu, T \,|\, \pi) + \mathrm{h}(\psi, \mu, T \,|\, \varphi \vee \pi). \tag{3}$$

Démonstration. Soit π une application facteur.

Commençons par traiter le cas où φ et ψ sont des processus. On utilise le lemme précédent et la remarque 2.1 :

$$\begin{aligned} h(\varphi \vee \psi \,|\, \pi) &= \lim \frac{1}{n} H\left(\varphi_{[0,n[} \vee \psi_{[0,n[} \,|\, \pi\right) \\ &= \lim \frac{1}{n}\left[H\left(\varphi_{[0,n[} \,|\, \pi\right) + H\left(\psi_{[0,n[} \,|\, \varphi_{[0,n[} \vee \pi\right)\right] \\ &= h(\varphi \,|\, \pi) + h(\psi \,|\, \varphi \vee \pi) \end{aligned}$$

Cas général : Soient φ et ψ des applications facteurs quelconques. On construit deux suites de processus $(\varphi_k)_{k\geq 0}$ et $(\psi_l)_{l\geq 0}$ qui engendrent $\sigma(\varphi)$ et $\sigma(\psi)$ respectivement. Alors la suite $(\varphi \vee \psi_l)_{l\geq 0}$ engendre $\sigma(\varphi \vee \psi)$ et ensuite $(\varphi_k \vee \psi_l)_{k\geq 0}$ engendre $\sigma(\varphi \vee \psi_l)$. On peut donc appliquer le lemme 2.2 :

$$\begin{aligned} h(\varphi \vee \psi \,|\, \pi) &= \lim_{l\to\infty} h(\varphi \vee \psi_l \,|\, \pi) \\ &= \lim_{l\to\infty} \lim_{k\to\infty} h(\varphi_k \vee \psi_l \,|\, \pi) \quad = \lim_{l\to\infty} \lim_{k\to\infty} \left[h(\varphi_k \,|\, \pi) + h(\psi_l \,|\, \varphi_k \vee \pi)\right] \\ &= h(\varphi \,|\, \pi) + h(\psi \,|\, \varphi \vee \pi) \end{aligned}$$

où on utilise la remarque 2.3 et le fait que $(\varphi_k \vee \pi)_{k\geq 0}$ engendre $\sigma(\varphi \vee \pi)$.

□

2.1.3 *Exemple : Schémas de Bernoulli*

Soient A un ensemble fini ou dénombrable, et γ une mesure de probabilité sur A. On veut calculer l'entropie du schéma de Bernoulli $(A^{\mathbb{Z}}, \mathcal{C}, \gamma^{\otimes\mathbb{Z}}, T_A)$. Notons α le processus canonique donné par

$$\alpha_n : \mathrm{a} \in A^{\mathbb{Z}} \mapsto a_n \in A.$$

On remarque que l'on peut alors écrire le cylindre $C_m^n(\mathbf{a})$ en fonction de α :

$$C_m^n(\mathbf{a}) = \bigcap_{i=m}^{n} \{\alpha_i = a_i\},$$

et donc α engendre la tribu cylindrique. Ainsi, le théorème 2.1 donne :

$$\begin{aligned} h(T_A) &:= h(\gamma^{\otimes \mathbb{Z}}, T_A) = h(\alpha, \gamma^{\otimes \mathbb{Z}}, T_A) \\ &= \lim_{n \longrightarrow \infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} \alpha_i) = \lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} H(\alpha_i) \\ &= H(\alpha_0) = -\sum_{a \in A} \gamma(a) \log(\gamma(a)). \end{aligned}$$

On remarque que pour tout autre processus ξ sur un espace quelconque (Y, ν, S), de même condition initiale que α (c'est-à-dire que la loi de ξ_0 est γ), on a

$$h(\xi, \nu, S) \leq H(\xi_0) = -\sum_{a \in A} \gamma(a) \log(\gamma(a)).$$

Les schémas de Bernoulli sont donc les processus les plus aléatoires, car ce sont ceux qui ont la plus grande entropie.

2.2 Divergence de Kullback-Leibler

On introduit un nouvel opérateur provenant aussi de la théorie de l'information : la divergence de Kullback-Leibler ou KL-divergence, qui sert à comparer deux mesures.

Définition 2.1. Soit $(X, \mathcal{F})$ un espace mesurable, et μ et ν deux mesures de probabilité sur X. La divergence de Kullback-Leibler de ν par rapport à μ est $+\infty$ si ν n'est pas absolument continue par rapport à μ, et si $\nu \ll \mu$ elle est donnée par la formule :

$$D(\nu \,\|\, \mu) = \int \log\left(\frac{d\nu}{d\mu}\right) d\nu = \int \log\left(\frac{d\nu}{d\mu}\right) \frac{d\nu}{d\mu} d\mu.$$

Remarques. Comme la fonction $x \mapsto x \log(x)$ est convexe, l'inégalité de Jensen donne

$$D(\nu \,\|\, \mu) = \int \log\left(\frac{d\nu}{d\mu}\right) \frac{d\nu}{d\mu} d\mu \geq \int \frac{d\nu}{d\mu} d\mu \log\left(\int \frac{d\nu}{d\mu} d\mu\right) = 0.$$

On voit que $D(\nu \,\|\, \mu) = 0 \Leftrightarrow \mu = \nu$. Clairement, si $\nu = \mu$ alors $d\nu d\mu = 1$ et donc $D(\nu \,\|\, \mu) = 0$. Réciproquement, si $D(\nu \,\|\, \mu) = 0$, alors $\nu \ll \mu$ et on a $\forall x, \frac{d\nu}{d\mu}(x) \in \{0, 1\}$, donc on peut écrire $\nu = \mathbb{1}_A \mu$, mais $\nu(X) = 1$, donc $\mu(A) = 1$, et donc $\nu = \mu$. Malgré cela, la KL-divergence n'est pas une distance car elle n'est pas symétrique. De plus, elle ne vérifie pas l'inégalité triangulaire.

En théorie de l'information la KL-divergence est souvent appelée *entropy relative* mais nous n'utiliserons pas ce nom ici pour éviter la confusion. Cependant, il y a un lien entre la KL-divergence et l'entropie, comme le montre [2, Lemma 2.2]. Ce résultat permet de donner un sens qualitatif à cet opérateur.

La propriété élémentaire par laquelle nous commençons est une forme de "chain rule" adaptée à la KL-divergence.

Lemme 2.6 ("Chain rule")**.** *Soient K et L deux espaces métriques compacts, et λ, $\lambda' \in Prob(K \times L)$. Notons λ_K (et λ'_k) la marginale selon la première variable et $x \mapsto \lambda_{L,x}$ (et $x \mapsto \lambda'_{L,x}$) une loi conditionnelle de λ (ou λ') sachant la première variable. Alors :*

$$D(\lambda \,\|\, \lambda') = D(\lambda_K \,\|\, \lambda'_K) + \int_K D(\lambda_{L,x} \,\|\, \lambda'_{L,x}) \, \mathrm{d}\lambda_K.$$

Démonstration. Supposons que λ est absolument continue par rapport à λ'. On voit facilement que $\lambda_K \ll \lambda'_K$. Traitons ensuite le cas des mesures conditionnelles : soient g une fonction continue sur L et f une fonction continue sur K, et on remarque que $\frac{1}{\frac{d\lambda_K}{d\lambda'_K}}$ est bien définie λ_K-p.s. car $\lambda_K(\frac{d\lambda_K}{d\lambda'_K} = 0) = 0$. On peut donc faire le calcul suivant :

$$\begin{aligned}\int_K f(x) \left[\int_L g(y) d\lambda_{L,x}(y)\right] d\lambda_K(x) &= \int_{K\times L} fg \; d\lambda = \int_{K\times L} fg \, \frac{d\lambda}{d\lambda'} \; d\lambda' \\ &= \int_K f(x) \left[\int_L g(y) \frac{d\lambda}{d\lambda'}(x,y) d\lambda'_{L,x}(y)\right] \frac{1}{\frac{d\lambda_K}{d\lambda'_K}(x)} d\lambda_K(x)\end{aligned}$$

On a donc montré que, pour tout g continue sur L, on a, pour λ_K-presque tout x :

$$\int_L g(y) d\lambda_{L,x}(y) = \frac{1}{\frac{d\lambda_K}{d\lambda'_K}(x)} \int_L g(y) \frac{d\lambda}{d\lambda'}(x,y) d\lambda'_{L,x}(y).$$

En utilisant la séparabilité de l'espace des fonctions continues sur L, on obtient, pour λ_K-presque tout x :

$$d\lambda_{L,x} = \frac{1}{\frac{d\lambda_K}{d\lambda'_K}(x)} \frac{d\lambda}{d\lambda'}(x,\cdot) d\lambda'_{L,x},$$

ce qui prouve que $\lambda_{L,x} \ll \lambda'_{L,x}$ et on obtient la formule naturelle :

$$\frac{d\lambda}{d\lambda'}(x,y) = \frac{d\lambda_{L,x}}{d\lambda'_{L,x}}(y)\frac{d\lambda_K}{d\lambda'_K}(x) \text{ pour } \lambda \text{ presque tout } x, y.$$

On peut désormais obtenir le résultat souhaité

$$\begin{aligned}
D(\lambda \,\|\, \lambda') &= \int \log\left(\frac{d\lambda}{d\lambda'}\right) d\lambda \\
&= \int_K \left[\int_L \log\left(\frac{d\lambda_{L,x}}{d\lambda'_{L,x}}(y)\frac{d\lambda_K}{d\lambda'_K}(x)\right) d\lambda_{L,x}(y)\right] d\lambda_K(x) \\
&= \int_K \log\left(\frac{d\lambda_K}{d\lambda'_K}(x)\right) d\lambda_K(x) + \int_K \left[\int_L \log\left(\frac{d\lambda_{L,x}}{d\lambda'_{L,x}}(y)\right) d\lambda_{L,x}(y)\right] d\lambda_K(x) \\
&= D(\lambda_K \,\|\, \lambda'_K) + \int_K D(\lambda_{L,x} \,\|\, \lambda'_{L,x}) \, \mathrm{d}\lambda_K.
\end{aligned}$$

Si λ n'est pas absolument continue par rapport a λ', alors le terme de gauche est infini. Si le terme de droite est fini, on a alors $D(\lambda_K \,\|\, \lambda'_K) < \infty$ et pour λ_K-presque x, on a $D(\lambda_{L,x} \,\|\, \lambda'_{L,x}) < \infty$, ce qui implique :

$$\lambda_K \ll \lambda'_K \text{ et } \lambda_{L,x} \ll \lambda'_{L,x} \text{ pour } \lambda_K\text{-presque tout } x.$$

On conclue avec un calcul comme ci-dessus : soit f une fonction continue sur $K \times L$

$$\begin{aligned}
\int f d\lambda &= \int_K \left[\int_L f(x,y) d\lambda_{L,x}\right] d\lambda_K \\
&= \int_K \left[\int_L f(x,y)\frac{d\lambda_{L,x}}{d\lambda'_{L,x}}(y) d\lambda'_{L,x}(y)\right] \frac{d\lambda_K}{d\lambda'_K}(x) d\lambda'_K(x) \\
&= \int f(x,y)\frac{d\lambda_{L,x}}{d\lambda'_{L,x}}(y)\frac{d\lambda_K}{d\lambda'_K}(x) d\lambda'(x,y).
\end{aligned}$$

Mais alors $\lambda \ll \lambda'$: absurde. Donc le terme de gauche est aussi infini et la formule est vraie. □

Corollaire 2.6.1. *Soit $\lambda_\bullet$ un noyau de mesures de K dans L et μ et ν deux mesures sur K. On a l'inégalité :*

$$D\left(\int_K \lambda_x d\nu(x) \,\|\, \int_K \lambda_x d\mu(x)\right) \leq D(\nu \,\|\, \mu).$$

Démonstration. On pose $\theta = \nu \ltimes \lambda_\bullet$ et $\theta' = \mu \ltimes \lambda_\bullet$, ce qui définit deux mesures sur $K \times L$. Leurs marginales sont :

$$\theta_K = \nu \text{ et } \theta'_K = \mu,$$

et

$$\theta_L = \int_K \lambda_x d\nu(x) \text{ et } \theta'_L = \int_K \lambda'_x d\mu(x).$$

On a aussi les mesures conditionnelles sachant la première variable

$$x \mapsto \theta_{L,x} = \lambda_x \text{ et } x \mapsto \theta'_{L,x} = \lambda_x.$$

On applique alors deux fois notre lemme :

$$D(\theta \,\|\, \theta') = D(\nu \,\|\, \mu) + \int_K D(\lambda_x \,\|\, \lambda_x)\, d\nu(x) = D(\nu \,\|\, \mu),$$

et

$$\begin{aligned} D(\theta \,\|\, \theta') &= D\left(\int_K \lambda_x d\nu(x) \,\|\, \int_K \lambda_x d\mu(x)\right) + \int_L D(\theta_{K,y} \,\|\, \theta'_{K,y}) d\theta_L(y) \\ &\geq D\left(\int_K \lambda_x d\nu(x) \,\|\, \int_K \lambda_x d\mu(x)\right). \end{aligned}$$

□

3 Théorie d'Ornstein relative

Rappelons que l'on appelle processus stationnaire la donnée d'un 5-uplet $(X, \mathcal{F}, \mu, T, \xi)$. On dit qu'un processus est *générateur* s'il vérifie la condition :

$$\mathcal{F} = \bigvee_{n=-\infty}^{+\infty} \xi_n = \sigma(\xi). \tag{4}$$

Cette condition implique l'isomorphisme

$$(X, \mathcal{F}, \mu, T) \cong (A^{\mathbb{Z}}, \mathcal{C}, \xi_* \mu, T_A), \tag{5}$$

mais on gardera la vision abstraite de ξ donnée précédemment car elle est très fructueuse. Par abus de notation, on désignera souvent un processus simplement par (T, ξ) et l'espace sur lequel T agit sera implicite. Enfin, dans cette partie, les processus seront à valeurs entières et prendront un nombre fini de valeurs, c'est-à-dire que l'alphabet A sera de la forme $\{1, ..., k\}$ pour $k \in \mathbb{N}$.

Dans ce contexte, on a une nouvelle vision des schémas de Bernoulli : l'isomorphisme (5) nous montre qu'un système (X, μ, T) est isomorphe à un schéma de Bernoulli si et seulement s'il existe un processus ξ sur X générateur tel que les variables aléatoires $\{\xi_n\}_{n\in\mathbb{Z}}$ sont deux-à-deux indépendantes.

3.1 Notations

Introduisons d'abord les notations nécessaires. On a choisi de regarder des processus à valeurs entières, donc on en fera autant pour toutes les variables aléatoires que l'on considérera : une variable aléatoire sera donc une variable aléatoire ξ_0 à valeur dans $[k] = \{1, ..., k\}$, $k \in \mathbb{N}$. On notera $d(\xi_0)$ la loi de ξ_0, qui est simplement un vecteur de $\mathbb{R}^k$. On a donc une distance entre les lois

$$|d(\xi_0) - d(\eta_0)| = \sum_{j=1}^{k} |\mu(\{\xi_0 = j\}) - \nu(\{\eta_0 = j\})|,$$

où ξ_0 et η_0 sont deux variables à valeurs dans $[k]$. On remarque qu'elles ne sont pas nécessairement définies sur le même espace. Si, de plus elles sont définies sur le même espace, on a une distance plus fine entre ξ_0 et η_0 donnée par

$$|\xi_0 - \eta_0| = \mu(\{\xi_0 \neq \eta_0\}).$$

On a toujours

$$|d(\xi_0) - d(\eta_0)| \leq 2|\xi_0 - \eta_0|.$$

On notera, pour $\varepsilon > 0$, $\xi_0 \preccurlyeq_\varepsilon \eta_0$ s'il existe ξ_0' telle que $|\xi_0 - \xi_0'| \leq \varepsilon$, et $\xi_0' \preccurlyeq \eta_0$. On pourra utiliser les mêmes notations dans le cas où η_0 est remplacée par une tribu.

Soit $F \subset X$ mesurable. On note $\xi_0|_F$ la restriction de ξ_0 à F. La loi de $\xi_0|_F$ est alors

$$d(\xi_0|_F) = (\mu_F(\{\xi_0 = 1\}), ..., \mu_F(\{\xi_0 = k\})) = \left(\frac{\mu(\{\xi_0 = 1\} \cap F)}{\mu(F)}, ..., \frac{\mu(\{\xi_0 = k\} \cap F)}{\mu(F)}\right).$$

Si ξ est un processus, on appellera, pour $x \in X$, ξ-nom de x la suite bi-infinie $(\xi_n(x))_{n\in\mathbb{Z}}$, et pour $n \geq 0$ fixé, on appellera (ξ, n)-nom de x le suite finie $(\xi_j(x))_{0\leq j\leq n-1}$.

On dira que deux processus (T,ξ) et (S,ζ) sont isomorphes, et notera $(T,\xi)\sim(S,\zeta)$, si pour tout $n\geq 0$, on a

$$d(\xi_{[0,n[}) = d(\zeta_{[0,n[}).$$

3.2 Distance entre deux processus

Soient (T,ξ) et (S,ζ) deux processus engendrés par (X,μ,T) et (Y,ν,S) respectivement, ainsi que π et π' tels que $(T,\pi)\sim(S,\pi')$. On souhaite définir une distance qui permet de comparer ces deux processus relativement à π et π'. Pour cela, on se fixe un espace de Lebesgue (Z,λ) et on regarde les ensembles de bijections suivants :

$$\mathcal{E}_X = \{\varphi : X \longrightarrow Z\,;\ \varphi \text{ bijection telle que } \varphi_*\mu = \lambda\},$$

et

$$\mathcal{E}_Y = \{\psi : Y \longrightarrow Z\,;\ \psi \text{ bijection telle que } \psi_*\nu = \lambda\}.$$

On ajoute ensuite une condition associée à π et π' :

$$\mathcal{E}_n = \{\varphi\in\mathcal{E}_X, \psi\in\mathcal{E}_Y \,|\, \varphi_*\pi_{[0,n[} = \psi_*\pi'_{[0,n[} \bmod 0\}.$$

On peut alors définir la “distance relative” suivante :

$$\bar{d}((T,\xi),(S,\zeta)\,|\,\pi,\pi') := \sup_{n\in\mathbb{N}} \inf_{(\varphi,\psi)\in\mathcal{E}_n} \frac{1}{n}\sum_{j=0}^{n-1} |\varphi_*\xi_j - \psi_*\zeta_j|,$$

où $\varphi_* f$ est le “poussé en avant” de f par φ : $\varphi_* f = f\circ\varphi^{-1}$. On remarquera que $\mathcal{E}_n$ est non vide grâce à l'hypothèse d'isomorphisme entre π et π'.

Dans la théorie d'Ornstein, la distance $\bar{d}$ s'avère très pratique. Cependant, ce n'est pas celle utilisée par Austin dans son article. Pour que les résultats présentés dans cette partie puissent être appliqués, il faut vérifier que les deux notions sont équivalentes. Soient (X,μ,T,ξ) et (Y,ν,T,ζ) deux processus. On définit alors la distance $\bar{d}_1$ par la formule

$$\bar{d}_1((T,\xi),(S,\zeta)) := \bar{d}((\pi\vee\xi)_*\mu, (\pi'\vee\zeta)_*\nu\,|\,\beta),$$

où $\bar{d}$ est la distance qu'Austin utilise dans son article ([2, Definition 11.1]). La discussion ci-dessous permet d'obtenir le résultat énoncé par Austin ([2, Theorem 11.5]) à partir du corollaire démontré dans la suite de cette section (Corollaire 3.3.1).

Vérifions que cette nouvelle nouvelle définition nous définit bien le même objet :

Lemme 3.1. $\bar{d} = \bar{d}_1$.

Démonstration. Cette démonstration est inspirée du travail d'Ornstein ([5, Appendix C]) car la définition qu'il donne de $\bar{d}$ se rapproche de la notre. Avant de commencer la preuve, on voit que l'on peut supposer que $\pi \vee \xi$ et $\pi' \vee \zeta$ sont générateurs. De plus, on se donne $k, r \geq 1$ tels que ξ et ζ soient à valeurs dans $[k]$ et π et π' à valeurs dans $[r]$.

- $\bar{d} \leq \bar{d}_1$:

Pour les définitions utilisées dans la suite, on se réfère à [2, Section]. Soit λ un élément de $\text{Join}(\xi_*\mu, \zeta_*\nu \,|\, \pi_*\mu)$. Par définition de λ, $(Z, \lambda, R) = ([r]^{\mathbb{Z}} \times [k]^{\mathbb{Z}} \times [k]^{\mathbb{Z}}, \lambda, T_{[r]\times[k]\times[k]})$ est un système dynamique. Sur Z, on a alors les processus $\tilde{\pi}$, $\tilde{\xi}$ et $\tilde{\zeta}$ (obtenus par projection sur chaque coordonnée), tels que $\tilde{\pi} \vee \tilde{\xi}$ et $\tilde{\pi} \vee \tilde{\zeta}$ ont les mêmes lois que $\pi \vee \xi$ et $\pi' \vee \zeta$. Donc, pour tout n, on peut utiliser le lemme 4.1 pour construire des bijections $\varphi_n : X \longrightarrow Z$ et $\psi_n : Y \longrightarrow Z$ qui préservent la mesure et qui envoient

$$(\pi \vee \xi)_{[0,n[} \text{ sur } (\tilde{\pi} \vee \tilde{\xi})_{[0,n[}, \text{ et } (\pi' \vee \zeta)_{[0,n[} \text{ sur } (\tilde{\pi} \vee \tilde{\zeta})_{[0,n[}.$$

Alors $(\varphi_n, \psi_n) \in \mathcal{E}_n$. Finalement, on obtient :

$$\begin{aligned}
\frac{1}{n}\sum_{l=0}^{n-1} |\varphi_{n*}\xi_l - \psi_{n*}\zeta_l| &= \frac{1}{n}\sum_{l=0}^{n-1} |\tilde{\xi}_l - \tilde{\zeta}_l| \\
&= \frac{1}{n}\sum_{l=0}^{n-1} |\tilde{\xi}_0 \circ R^l - \tilde{\zeta}_0 \circ R^l| \\
&= \frac{1}{n}\sum_{l=0}^{n-1} |\tilde{\xi}_0 - \tilde{\zeta}_0| \\
&= \lambda\big(\tilde{\xi}_0(\mathbf{i},\mathbf{j},\mathbf{j}') \neq \tilde{\zeta}_0(\mathbf{i},\mathbf{j},\mathbf{j}')\big) = \lambda\big(\{(\mathbf{i},\mathbf{j},\mathbf{j}') \,|\, j_0 \neq j'_0\}\big).
\end{aligned}$$

Donc

$$\bar{d}((T,\xi),(S,\zeta)) \leq \lambda\big(\{(\mathbf{i},\mathbf{j},\mathbf{j}') \,|\, j_0 \neq j'_0\}\big).$$

Comme λ était un élément de $\text{Join}(\xi_*\mu, \zeta_*\nu \,|\, \pi_*\mu)$ quelconque, on a

$$\bar{d}((T,\xi),(S,\zeta) \,|\, \pi, \pi') \leq \bar{d}_1((T,\xi),(S,\zeta) \,|\, \pi, \pi').$$

- $\bar{d} \geq \bar{d}_1$:

On va devoir construire un couplage relatif de ξ et ζ, et pour cela, on va construire un processus sur l'alphabet $A = [r] \times [k] \times [k]$, c'est-à-dire, concrètement que l'on doit construire une mesure invariante par translation sur $A^{\mathbb{Z}}$.

On rappelle la définition des cylindres : pour $\mathbf{a} \in A^{n-m+1}$:

$$C_m^n(\mathbf{a}) = \{\mathbf{x} \in A^{\mathbb{Z}} | (x_k)_{k=m}^n = \mathbf{a}\},$$

et les tribus associées :

$$\mathcal{C}_m^n = \sigma\big(C_m^n(\mathbf{a}); \mathbf{a} \in A^{n-m+1}\big).$$

Et, par définition, la tribu cylindrique est $\mathcal{C} = \sigma\big(\mathcal{C}_m^n; m, n \in \mathbb{Z}\big)$.

Soit $(\varepsilon_n)_{n\geq 0}$ qui tend vers 0 et pour tout n, fixons deux bijections $\varphi_n : X \longrightarrow Z$ et $\psi_n : Y \longrightarrow Z$ telles que

$$\frac{1}{n}\sum_{l=0}^{n-1} |\varphi_{n*}\xi_l - \psi_{n*}\zeta_l| \leq \bar{d}((T,\xi),(S,\zeta)\,|\,\pi,\pi') + \varepsilon_n. \tag{6}$$

A chaque suite finie $\mathbf{a} \in A^n$, il correspond un unique atome de

$$(\varphi_{n*}\pi \vee \varphi_{n*}\xi \vee \psi_{n*}\zeta)_{[0,n[},$$

et on en déduit donc une mesure $\tilde{\gamma}_n$ sur $\mathcal{C}_0^{n-1}$. On va la modifier un peu comme suit : pour $r < n$, on définit sur $\mathcal{C}_0^{r-1}$:

$$\gamma_n^r\big((a_0, ..., a_{r-1})\big) = \frac{1}{n-r}\sum_{l=0}^{n-r-1} \tilde{\gamma}_n\big(T_A^{-l}(a_0, ..., a_{r-1})\big), \tag{7}$$

qui est bien défini car $T_A^{-l}(\mathcal{C}_0^{r-1}) = \mathcal{C}_l^{l+r-1} \subset \mathcal{C}_0^{n-1}$.

On utilise l'argument diagonal de Cantor : comme l'espace des mesures sur $A^{\mathbb{Z}}$ est compact pour la topologie faible, il existe une extractrice α_0 et une mesure γ_∞^0 telles que

$$\gamma_{\alpha_0(n)}^0 \rightharpoonup \gamma_\infty^0.$$

Ensuite, on itère pour construire deux suites $(\alpha_r)_{r\geq 0}$ et $(\gamma_\infty^r)_{r\geq 0}$ telles que

$$\gamma_{\alpha_0\circ\cdots\circ\alpha_r(n)}^r \rightharpoonup \gamma_\infty^r.$$

On définit alors α via : $\alpha(n) = \alpha_0 \circ \cdots \circ \alpha_n(n)$ et alors, pour tout $r \geq 0$, on a

$$\gamma_{\alpha(n)}^r \rightharpoonup \gamma_\infty^r.$$

Pour tout r, γ_∞^r est définie sur $\mathcal{C}_0^{r-1}$, et donc $T_{A*}^l \gamma_\infty^r$ est une mesure définit sur $\mathcal{C}_l^{r+l-1}$. Pour appliquer le théorème d'extension de Kolmogorov, il faut vérifier que ces mesures sont compatibles. Or, on voit en utilisant (7) que $\gamma_\infty^{r+1}\big((a_0, ..., a_{r-1})\big) = \gamma_\infty^r\big((a_0, ..., a_{r-1})\big)$. Soit γ l'unique extension ainsi obtenue. Elle est clairement invariante par T_A. Enfin, on utilise (7) et (6) pour voir que

$$\begin{aligned}
\gamma\big(\{(i_n, j_n, j'_n)_{n\in\mathbb{Z}} \in A^{\mathbb{Z}} \,|\, j_0 \neq j'_0\}\big) &= \gamma_\infty^0(\{j \neq j'\}) \\
&= \lim_{n\longrightarrow\infty} \frac{1}{\alpha(n)} \sum_{l=0}^{\alpha(n)-1} \tilde{\gamma}_{\alpha(n)}\big(T_A^{-l}(\{j \neq j'\})\big) \\
&= \lim_{n\longrightarrow\infty} \frac{1}{\alpha(n)} \sum_{l=0}^{\alpha(n)-1} |\varphi_{\alpha(n)_*}\xi_l - \psi_{\alpha(n)_*}\zeta_l| \\
&\leq \bar{d}((T,\xi),(S,\zeta))
\end{aligned}$$

Or, on a construit γ de sorte que $\gamma \in \mathrm{Join}(\xi_*\mu, \zeta_*\nu \,|\, \pi_*\mu)$. Donc

$$\bar{d}_1((T,\xi),(S,\zeta) \,|\, \pi, \pi') \leq \bar{d}((T,\xi),(S,\zeta) \,|\, \pi, \pi').$$

$\square$

3.3 Processus relativement finiment déterminés

Équipés de cette distance, on peut introduire le concept central de cette partie :

Définition 3.1. Soit (Y, ν, S) un système ergodique. Un processus (S, ζ) est *relativement finiment déterminé* par rapport à un processus (S, π) si pour tout $\varepsilon > 0$, il existe $\delta > 0$ et $n_1 \geq 0$, tels que, pour tout système ergodique (X, μ, T) muni de ξ et π' tels que $(T, \pi') \sim (S, \pi)$ et vérifiant

(i) $|d((\pi' \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| \leq \delta$

(ii) $|h(\pi' \vee \xi, \mu, T) - h(\pi \vee \zeta, \nu, S)| \leq \delta,$

on a : $\bar{d}((T,\xi),(S,\zeta) \,|\, \pi', \pi) \leq \varepsilon$.

On dit que le système (Y, ν, S) est *relativement finiment déterminé* par rapport à π s'il existe ζ sur Y tel que $\pi \vee \zeta$ est générateur et ζ est relativement finiment déterminé par rapport à π.

On remarque que, grâce au lemme 3.1, cette définition équivaut à [2, Definition 11.3].

Le théorème que l'on va vouloir prouver est

Théorème 3.2. *Soient (X, μ, T) et (Y, ν, S) deux systèmes ergodiques munis de processus ξ, π' et ζ, π tels que $(T, \pi') \sim (S, \pi)$ et $\pi' \vee \xi$, $\pi \vee \zeta$ sont générateurs et ξ (resp. ζ) est relativement finiment déterminé par rapport à π' (resp. π).*

Si $h(\pi' \vee \xi, T) = h(\pi \vee \zeta, S)$, alors (X, μ, T) et (Y, ν, S) sont isomorphes. Plus précisément, il existe ζ^* sur X tel que $(T, \pi' \vee \zeta^*) \sim (S, \pi \vee \zeta)$ et $\pi' \vee \zeta^*$ est générateur.

Théorème 3.3. *Soit (X, μ, T) un système ergodique muni d'un processus π. Si (X, μ, T) est relativement Bernoulli par rapport à π, alors (X, μ, T) est relativement finiment déterminé par rapport à π.*

Ces deux théorèmes donnent alors une caractérisation des systèmes relativement Bernoulli. En effet, en combinant ces deux théorèmes, on obtient le réciproque du théorème 3.3

Corollaire 3.3.1. *Si (X, μ, T) est relativement finiment déterminé par rapport à π, alors (X, μ, T) est relativement Bernoulli par rapport à π.*

Ce corollaire nous donne alors [2, Theorem 11.5].

Donnons ici une preuve du théorème 3.3. Pour cela on introduit le définition suivante :

Définition 3.2. Soient deux variables aléatoires α et β à valeurs dans A et B respectivement. On dit que α est ε-indépendante de β s'il existe $\mathcal{E} \subset B$ tel que

$$\mu(\beta \in \mathcal{E}) \geq 1 - \varepsilon \text{ et } \forall b \in \mathcal{E},\ |d(\alpha_{|\{\beta = b\}}) - d(\alpha)| \leq \varepsilon.$$

Un processus ξ est ε-indépendant si pour tout $n \in \mathbb{N}$, ξ_n est ε-indépendant de $\xi_{[0,n[}$.

L'intérêt de cette nouvelle notion est donnée par le lemme suivant :

Lemme 3.4. *Soit $\varepsilon > 0$. Soient $(T, \pi' \vee \xi)$ et $(S, \pi \vee \zeta)$ deux processus tels que*

(i) ζ est un processus indépendant et ζ et π sont indépendants,

(ii) pour tout $n \in \mathbb{N}$, ξ_n est ε-indépendant de $(\pi' \vee \xi)_{[0,n[}$,

(iii) et $|d(\xi_0) - d(\zeta_0)| < \varepsilon$.

Alors : $\bar{d}((T,\xi),(S,\zeta)\,|\,\pi',\pi) \leq 2\varepsilon$.

Démonstration. Soit $k \geq 1$ tel que ξ_0 et ζ_0 sont à valeurs dans $[k]$. Pour $n \in \mathbb{N}$, notons

$$\bar{d}_n(\xi_{[0,n[}, \zeta_{[0,n[}\,|\,\pi',\pi) := \inf_{(\varphi,\psi)\in\mathcal{E}_n} \frac{1}{n}\sum_{j=0}^{n-1} |\varphi_*\xi_j - \psi_*\zeta_j|.$$

Il suffit lors de majorer strictement $\bar{d}_n$ par 2ε, pour chaque n. Par un découpage judicieux de Z et le lemme 4.1, on obtient l'affirmation suivante :

Affirmation 3.4.1. *Si α et β sont deux variables aléatoire sur X et Y respectivement, il existe $\varphi \in \mathcal{E}_X$ et $\psi \in \mathcal{E}_Y$ telles que $|\varphi_*\alpha - \psi_*\beta| = \frac{|d(\alpha)-d(\beta)|}{2}$.*

En utilisant (iii), l'affirmation donne le résultat pour $n = 1$. Soit $n \geq 1$, et supposons la majoration de $\bar{d}_n$ vraie, et montrons qu'elle est vraie pour $n+1$. On prend alors $(\varphi_n, \psi_n) \in \mathcal{E}_n$ telles que

$$\frac{1}{n}\sum_{j=0}^{n-1} |\varphi_{n*}\xi_j - \psi_{n*}\zeta_j| < 2\varepsilon \tag{8}$$

Construisons maintenant φ_{n+1} et ψ_{n+1} telles que (8) soit encore valable au rang $n+1$. Par l'hypothèse (ii), on peut fixer $\mathcal{E}_n \subset ([r]\times[k])^n$ tel que

$$\mu((\pi' \vee \xi)_{[0,n[} \in \mathcal{E}_n) \geq 1 - \varepsilon$$

et

$$\forall \mathbf{i} \in \mathcal{E}_n,\ |d(\xi_n|_{\{(\pi'\vee\xi)_{[0,n[}=\mathbf{i}\}}) - d(\xi_n)| \leq \varepsilon.$$

En combinant cela avec les hypothèses (i) et (iii), on obtient

$$\forall\ \mathbf{i} \in ([r]\times[k])^n,\ \mathbf{i}' \in \mathcal{E}_n,\ |d(\xi_n|_{\{(\pi'\vee\xi)_{[0,n[}=\mathbf{i}\}}) - d(\zeta_n|_{\{(\pi\vee\zeta)_{[0,n[}=\mathbf{i}'\}})| \leq 2\varepsilon.$$

Fixons $\mathbf{i} \in ([r]\times[k])^n$ et $\mathbf{i}' \in \mathcal{E}_n$, et notons

$$Z_{\mathbf{i},\mathbf{i}'} = \{\varphi_{n*}((\pi'\vee\xi)_{[0,n[}) = \mathbf{i}\} \cap \{\psi_{n*}((\pi\vee\zeta)_{[0,n[}) = \mathbf{i}'\}$$

ainsi que

$$X_{\mathbf{i},\mathbf{i}'} = \varphi_n^{-1} Z_{\mathbf{i},\mathbf{i}'} \ \text{ et } \ Y_{\mathbf{i},\mathbf{i}'} = \psi_n^{-1} Z_{\mathbf{i},\mathbf{i}'}.$$

L'affirmation (3.4.1) énoncé ci-dessus appliqué à $(\pi' \vee \xi)_{[0,n[}$ et $(\pi \vee \zeta)_{[0,n[}$ nous donne $\varphi_{n+1}\big|_{X_{\mathbf{i},\mathbf{i}'}} : X_{\mathbf{i},\mathbf{i}'} \longrightarrow Z_{\mathbf{i},\mathbf{i}'}$ et $\psi_{n+1}\big|_{Y_{\mathbf{i},\mathbf{i}'}} : Y_{\mathbf{i},\mathbf{i}'} \longrightarrow Z_{\mathbf{i},\mathbf{i}'}$ qui préservent la mesure et telles que

$$\begin{aligned}
|(\varphi_{n+1})_*(\xi_n\big|_{\{(\pi'\vee\xi)_{[0,n[}=\mathbf{i}\}}) - (\psi_{n+1})_*(\zeta_n\big|_{\{(\pi\vee\zeta)_{[0,n[}=\mathbf{i}'\}})| &= \frac{|d(\xi_n\big|_{\{(\pi'\vee\xi)_{[0,n[}=\mathbf{i}\}}) - d(\zeta_n\big|_{\{(\pi\vee\zeta)_{[0,n[}=\mathbf{i}'\}})|}{2} \\
&\leq \varepsilon.
\end{aligned}$$

On étend ensuite φ_{n+1} et ψ_{n+1} au cas $\mathbf{i}' \notin \mathcal{E}_n$ en des bijections préservant la mesure en s'assurant que $(\varphi_{n+1}, \psi_{n+1}) \in \mathcal{E}_{n+1}$ (grâce au lemme 4.1). Cette construction donne alors :

$$\begin{aligned}
|\varphi_{n+1*}\xi_n - \psi_{n*}\zeta_n| &= \sum_{\mathbf{i},\mathbf{i}'\in[k]^n} \lambda(Z_{\mathbf{i},\mathbf{i}'})\ |(\varphi_{n+1})_*(\xi_n\big|_{\{(\pi'\vee\xi)_{[0,n[}=\mathbf{i}\}}) - (\psi_{n+1})_*(\zeta_n\big|_{\{(\pi\vee\zeta)_{[0,n[}=\mathbf{i}'\}})| \\
&\leq \sum_{\mathbf{i}\in[k]^n,\mathbf{i}'\in\mathcal{E}_n} \lambda(Z_{\mathbf{i},\mathbf{i}'})\ |(\varphi_{n+1})_*(\xi_n\big|_{\{(\pi'\vee\xi)_{[0,n[}=\mathbf{i}\}}) - (\psi_{n+1})_*(\zeta_n\big|_{\{(\pi\vee\zeta)_{[0,n[}=\mathbf{i}'\}})| + \varepsilon \\
&\leq 2\varepsilon.
\end{aligned}$$

De plus, comme on observe que $X_{\mathbf{i},\mathbf{i}'} \subset \{(\pi' \vee \xi)_{[0,n[} = \mathbf{i}\}$ et $Y_{\mathbf{i},\mathbf{i}'} \subset \{(\pi \vee \zeta)_{[0,n[} = \mathbf{i}'\}$, pour tout $l \leq n-1$, on a

$$\varphi_{n+1*}\xi_l = \varphi_{n*}\xi_l \ \text{ et } \ \psi_{n+1*}\zeta_l = \psi_{n*}\zeta_l.$$

Finalement, en utilisant ces résultats et (8), on a

$$\begin{aligned}
\frac{1}{n+1}\sum_{j=0}^{n} |\varphi_{n+1*}\xi_j - \psi_{n+1*}\zeta_j| &= \frac{n}{n+1}\frac{1}{n}\sum_{j=0}^{n-1} |\varphi_{n*}\xi_j - \psi_{n*}\zeta_j| + \frac{1}{n+1}|\varphi_{n+1*}\xi_n - \psi_{n*}\zeta_n| \\
&< 2\varepsilon.
\end{aligned}$$

Donc, pour tout n, on a

$$\bar{d}_{n+1}(\xi_{[0,n+1[}, \zeta_{[0,n+1[}\,|\,\pi', \pi) < 2\varepsilon,$$

et donc

$$\bar{d}((T,\xi),(S,\zeta)\,|\,\pi',\pi) \leq 2\varepsilon.$$

□

Complétons ce lemme avec le résultat suivant, qui établit le lien entre l'entropie et l'ε-indépendance d'un processus.

Lemme 3.5. *Pour tout $\varepsilon > 0$ et $k \geq$, il existe $\delta > 0$ tel que, si α est une variable aléatoire à valeurs dans $[k]$ et $H(\alpha \,|\, \beta) \geq H(\alpha) - \delta$, alors α est ε-indépendant de β.*

Démonstration. Soient $\varepsilon > 0$ et $k \geq 1$. Notons $\mathcal{A}$ l'ensemble des couples (α, β) tels que α est à valeurs dans $[k]$ et n'est pas ε-indépendante de β. On va vouloir alors minorer $H(\alpha) - H(\alpha \,|\, \beta)$ sur $\mathcal{A}$. Montrons que l'on peut se ramener au cas où β est à valeurs dans $\{0, 1\}$: notons $\mathcal{E}$ l'ensemble des entiers i tels que

$$|d(\alpha_{|\{\beta=i\}} - d(\alpha)| > \varepsilon.$$

Par hypothèse, $\mu(\beta \in \mathcal{E}) > \varepsilon$, donc

$$\sum_{i \in \mathcal{E}} \mu(\beta = i) \; |d(\alpha_{|\{\beta=i\}}) - d(\alpha)| > \varepsilon^2.$$

Par le principe des tiroirs, on peut choisir $j \in [k]$ tel que

$$\sum_{i \in \mathcal{E}} \mu(\beta = i) \; |\mu_{|\{\beta=i\}}(\alpha = j) - \mu(\alpha = j)| > \frac{\varepsilon^2}{k},$$

puis $\mathcal{E}' \subset \mathcal{E}$ tel que

$$\left| \sum_{i \in \mathcal{E}'} \mu(\beta = i) \; \big(\mu_{|\{\beta=i\}}(\alpha = j) - \mu(\alpha = j)\big) \right| > \frac{\varepsilon^2}{2k}.$$

Enfin, on pose $\rho(\beta) = \mathbb{1}_{\beta \in \mathcal{E}'} \preccurlyeq \beta$, et on a

$$H(\alpha) - H(\alpha \,|\, \beta) \geq H(\alpha) - H(\alpha \,|\, \rho(\beta)).$$

De plus, les estimations précédentes donnent :

$$\mu(\rho = 0) \geq \varepsilon, \;\; \mu(\rho = 1) \geq \frac{\varepsilon^2}{2k}, \text{ et } |d(\alpha_{|\rho=1}) - d(\alpha)| \geq \frac{\varepsilon^2}{2k}. \tag{9}$$

Or, on peut voir $H(\alpha) - H(\alpha \,|\, \rho)$ comme une fonction continue de $d(\alpha \vee \rho)$, et (9) montre qu'il existe une partie compacte K de $\mathbb{R}^{2k}$ sur laquelle $H(\alpha) - H(\alpha \,|\, \rho)$ ne s'annule pas, et pour tout $\alpha, \beta \in \mathcal{A}$, $d(\alpha \vee \rho(\beta)) \in K$.

Il existe alors $\delta > 0$ tel que, pour tout $\alpha, \beta \in \mathcal{A}$, $H(\alpha) - H(\alpha \,|\, \beta) \geq \delta$. $\square$

Démonstration du théorème 3.3. Soient (S, ζ) processus indépendant et π indépendant de ζ. Soit $\varepsilon > 0$. Prenons un processus (T, ξ) à valeurs dans $[k]$ et π' tel que $(T, \pi') \sim (S, \pi)$, vérifiant

(i) $|d((\pi' \vee \xi)_0) - d((\pi \vee \zeta)_0)| \leq \delta$

(ii) $|h(\pi' \vee \xi, T) - h(\pi \vee \zeta, S)| \leq \delta,$

pour $\delta > 0$ à fixer plus tard. Soit $\delta' > 0$ donné par le lemme 3.5. Par (i), en prenant δ assez petit, on a $|H(\xi_0) - H(\zeta_0)| \leq \frac{\delta'}{2}$, et alors, pour tout $n \geq 0$

$$\begin{aligned} H(\xi_n \,|\, (\pi' \vee \xi)_{[0,n[}) &= H(\xi_0 \,|\, (\pi' \vee \xi)_{[-n,0[}) \\ &\geq h(\xi, T \,|\, \pi') \geq h(\zeta, S \,|\, \pi) - \delta \quad \text{par (ii)} \\ &= H(\zeta_0) - \delta \geq H(\xi_0) - \frac{\delta'}{2} - \delta \\ &\geq H(\xi_n) - \delta', \end{aligned}$$

si on prend $\delta \leq \frac{\delta'}{2}$. Et donc, par le lemme 3.5, ξ_n est ε-indépendante de $(\pi' \vee \xi)_{[0,n[}$. Enfin, si $\delta \leq \varepsilon$, le lemme 3.4 et (i) nous donnent : $\bar{d}((T, \xi), (S, \zeta)) \leq 2\varepsilon$. Quitte à remplacer ε par $\frac{\varepsilon}{2}$, on a montré que (S, ζ) est finiment déterminé. □

Remarque 3.1. On voit dans la preuve ci-dessus que, pour les systèmes relativement Bernoulli, l'entier n_1 qui apparaît dans la définition 3.1 vaut 1.

3.4 Gadgets

Dans cette section, nous allons introduire un nouvel objet : les gadgets. Commençons par mentionner un théorème, dû à Rohlin.

Théorème 3.6. *Soit (X, μ, T) un système ergodique et ξ_0 une variable aléatoire. Pour tout $\varepsilon > 0$ et pour tout $n \in \mathbb{N}$, il existe un ensemble F tel que les ensembles $T^j F$, $0 \leq j \leq n-1$, sont deux-à-deux disjoints, $\mu\left(\bigcup_{j=0}^{n-1} T^j F\right) \geq 1 - \varepsilon$, et $d(\xi_0|_F) = d(\xi_0)$.*

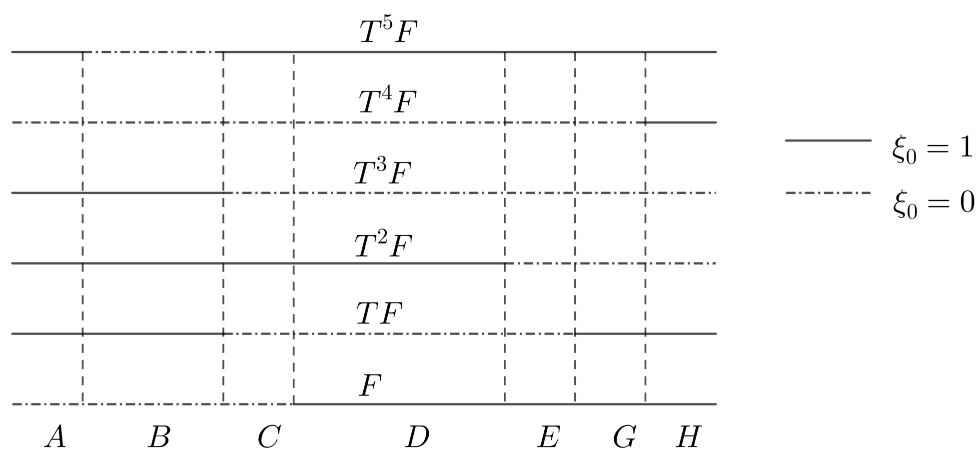
On se fixe alors un ensemble F donné par le théorème. On appellera alors $\{F, TF, ..., T^{n-1}F\}$ une *pile*, et F la base de la pile. Chaque ensemble $T^j F$ est appelé "étage" de la pile.

Définition 3.3. Un *gadget* est un quadruplet (T, F, n, ξ_0), où T est une transformation ergodique, F est la base d'une pile de hauteur n et ξ_0 est une variable aléatoire définie sur la pile.

Tâchons de donner une meilleure compréhension de la structure des gadgets. Sur les $n-1$ premiers étages de la pile, l'action de T est très simple : lorsque l'on applique T, on passe à l'étage supérieur. Mais le gadget ne donne aucune information sur l'action de T sur le dernier étage $T^{n-1}F$ et sur le complémentaire de la pile. Pour cette raison, dans cette partie, on ne s'intéressera pas au processus complet associé à ξ_0, car, pour $x \in F$, la suite $(T^j(x))_{j\in\mathbb{Z}}$ sort de la pile. On va donc se contenter de l'information donnée par les valeurs prises par ξ_0 sur la pile : pour chaque $j \leq n-1$, on regarde la restriction de ξ_0 à l'étage T^jF de la pile, puis on applique T pour obtenir une variable sur F : $\xi_0|_{T^jF} \circ T^j$. Ainsi, la structure du gadget est contenu dans la variable $\bigvee_{j=0}^{n-1} \xi_0|_{T^jF} \circ T^j$. Mais on voit que

$$\bigvee_{j=0}^{n-1} \xi_0|_{T^jF} \circ T^j = \bigvee_{j=0}^{n-1} \xi_j|_F = \left(\bigvee_{j=0}^{n-1} \xi_0\right)\Bigg|_F = \xi_{[0,n[}|_F.$$

Ainsi, connaître la structure du gadget revient à connaître les (ξ, n)-noms des éléments de F. On obtient alors une partition de F composée d'atomes de la forme $A = \{\xi_{[0,n[}|_F = \mathbf{i}\}$, avec $\mathbf{i} \in [k]^n$. Pour chaque A de cette forme, la pile associée $C_A = \{A, TA, ..., T^{n-1}A\}$ est appelée une *colonne* du gadget. Par exemple, dans la figure ci-dessous, qui donne une représentation visuelle du gadget, $A = \{\xi_{[0,5]}|_F = (0,1,1,1,0,1)\}$, et la colonne associée est au dessus de A (délimitée par les pointillés verticaux).



On voit que, une fois les colonnes construite, on obtient l'injection suivante, en leur associant leurs (ξ, n)-noms :

$$\begin{aligned} \Phi : \textit{Ens. des colonnes} &\longrightarrow [k]^n \\ C_A &\longmapsto (\xi_0(A), \xi_0(TA), ..., \xi_0(T^{n-1}A)) \end{aligned}$$

Réciproquement, si on se donne un gadget, un ensemble de colonnes $\mathcal{C}$, $l \geq 1$ et une injection quelconque :

$$\Psi : \textit{Ens. des colonnes} \longrightarrow [l]^n,$$

alors on peut trouver une nouvelle variable aléatoire ζ à valeurs dans $[l]$ qui engendre les mêmes colonnes et telle que Ψ donne le (ζ, n)-nom des colonnes. Pour cela, il suffit de voir que, comme Ψ associe une suite de n éléments à chaque colonne, elle associe à chaque étage de la colonne un élément de $[l]$. Appelons cela "l'étiquette" de l'étage. On peut ensuite définir ζ via :

$$\{\zeta = i\} = \bigcup_{C \in \mathcal{C}} \{\text{étages de C dont l'étiquette est } i\}$$

Ainsi, plusieurs variables aléatoires différentes peuvent donner les mêmes colonnes, mais il y a certaines contraintes. Le lemme suivant en donne une :

Lemme 3.7. *Supposons que (T, F, n, ξ_0) et (T, F, n, ζ_0) ont les mêmes colonnes. Posons $G = \bigcup_{j=0}^{n-1} T^j F$ et $\rho = \mathbb{1}_F|_G$. Alors :*

$$\xi_0|_G \preccurlyeq \left. \left(\bigvee_{j=-(n-1)}^{n-1} (\zeta_0 \vee \rho) \circ T^j \right) \right|_G$$

Intéressons nous à la notion d'isomorphisme de gadgets : on a vu que la structure du gadget était contenue dans la variable aléatoire $\xi_{[0,n[}|_F$, donc on dit que deux gadgets (T, F, n, ξ_0) et (S, E, n, ζ_0) sont isomorphes s'ils vérifient :

$$d(\xi_{[0,n[}|_F) = d(\zeta_{[0,n[}|_E).$$

Remarque 3.2. Grâce au lemme 4.1, on voit que (T, F, n, ξ_0) et (S, E, n, ζ_0) sont isomorphe si et seulement s'il existe une bijection $\varphi : F \longrightarrow E$ telle que $\nu_F = \varphi_*(\mu_F)$ et $\xi_{[0,n[}|_F = \zeta_{[0,n[}|_E \circ \varphi$.

Remarque 3.3. On a vu que sur un gadget, on ne peut pas calculer tout le processus associé à une v.a.. Cependant, si une variable aléatoire ξ_0 est définie sur tout l'espace, le processus ξ est bien défini, et par abus de notation, on notera parfois le gadget (T, F, n, ξ) (au lieu de (T, F, n, ξ_0)).

Lemme 3.8. *Soit (T, F, n, ξ_0) un gadget et (Y, ν, S) un système ergodique. Pour tout $\varepsilon > 0$, il existe $E \subset Y$ mesurable et une variable aléatoire ζ_0 tels que (S, E, n, ζ_0) est isomorphe à (T, F, n, ξ_0), et $\nu(\bigcup_{j=0}^{n-1} S^j E) \geq 0$.*

Ce lemme nous montre que la donnée d'un gadget donne très peu d'information sur le système sous-jacent. En revanche, l'extension suivante s'avérera très intéressante :

Lemme 3.9. *Supposons que (T, F, n, ξ_0) et (S, E, n, ζ_0) sont isomorphes. Soit ξ_0' une variable aléatoire sur $\bigcup_{j=0}^{n-1} T^j F$. Alors, il existe une variable aléatoire ζ_0' sur $\bigcup_{j=0}^{n-1} S^j E$ telle que $(T, F, n, \xi_0 \vee \xi_0')$ et $(S, E, n, \zeta_0 \vee \zeta_0')$ sont isomorphes.*

Comme pour les processus, on veut être en mesure de comparer deux différents gadgets, et va donc construire une distance sur l'ensemble des gadgets. On se fixe n et on prend deux gadgets de taille n : (T, F, n, ξ_0) et (S, E, n, ζ_0). Notons $G = \bigcup_{j=0}^{n-1}$ et introduisons

$$\mathcal{H} = \{\zeta_0' \text{ v. a. sur } G \,|\, (T, F, n, \zeta_0') \cong (S, E, n, \zeta_0)\}.$$

On définit alors

$$\begin{aligned} \bar{d}\big((T, F, n, \xi_0), (S, E, n, \zeta_0)\big) :&= \inf_{\zeta' \in \mathcal{H}} \frac{1}{n} \sum_{j=0}^{n-1} |\xi_0|_{T^j F} - \zeta_0'|_{T^j F}| \\ &= \inf_{\zeta' \in \mathcal{H}} \frac{1}{n} \sum_{j=0}^{n-1} |\xi_j|_F - \zeta_j'|_F|. \end{aligned}$$

Le lemme suivant nous dit ce qu'il se passe lorsque cette distance est petite :

Lemme 3.10. *Soit E l'ensemble des points de F dont le (ξ, n)-nom et le (ζ, n)-nom diffèrent à au moins $n\varepsilon$ endroits. Si (T, F, n, ξ_0) et (T, F, n, ζ_0) vérifient :*

$$\frac{1}{n} \sum_{j=0}^{n-1} |\xi_0|_F - \zeta_0|_F| \leq \varepsilon^2,$$

alors : $\mu(E) \leq \varepsilon \mu(F)$.

Démonstration. Notons $E_j = \{x \in F \,|\, \xi_j(x) \neq \zeta_j(x)\}$, et on peut alors écrire

$$E = \{x \in F \,|\, \sum_{j=0}^{n-1} \mathbb{1}_{E_j}(x) \geq \varepsilon n\}.$$

Le résultat du lemme découle alors de l'inégalité de Markov :

$$\begin{aligned}
\mu(E) &\leq \frac{1}{n\varepsilon} \int \sum_{j=0}^{n-1} \mathbb{1}_{E_j}(x) d\mu \\
&= \frac{1}{n\varepsilon} \sum_{j=0}^{n-1} \mu(E_j) = \frac{1}{\varepsilon}\frac{1}{n} \sum_{j=0}^{n-1} |\xi_j|_F - \zeta'_j|_F| \\
&\leq \frac{1}{\varepsilon}\varepsilon^2 = \varepsilon.
\end{aligned}$$

□

Voyons maintenant le lien avec la distance $\bar{d}$ définie pour les processus :

Lemme 3.11. *Soit $\varepsilon > 0$. Soient T et S deux transformations ergodiques. Si $\bar{d}\big((T,\xi),(S,\zeta) \,|\, \pi', \pi\big) < \varepsilon$, alors, pour tout n et pour tout $\delta > 0$, il existe des gadgets $(T, F, n, \pi'_0 \vee \xi_0)$ et $(S, E, n, \pi_0 \vee \zeta_0)$ tels que $\mu(\bigcup_{j=0}^{n-1} T^j F) \geq 1 - \delta$, $\nu(\bigcup_{j=0}^{n-1} S^j E) \geq 1 - \delta$, et*

$$\bar{d}\big((T, F, n, \pi'_0 \vee \xi_0), (S, E, n, \pi_0 \vee \zeta_0)\big) < \varepsilon.$$

Remarque 3.4. Il est facile de voir que la réciproque est aussi vraie, mais cela ne nous sera pas utile, donc on omet la preuve.

Démonstration. Soient $n \in \mathbb{N}$ et $\delta > 0$. On voit facilement que l'hypothèse du lemme nous permet de construire une bijection $\varphi : X \longrightarrow Y$ telle que $\varphi_*\mu = \nu$, $\varphi_*\pi'_{[0,n[} = \pi_{[0,n[}$ et

$$\frac{1}{n} \sum_{j=0}^{n-1} |\varphi_*\xi_j - \zeta_j| < \varepsilon.$$

Par le théorème 3.6, on peut contruire deux gadgets (T, F, n, ξ_0) et (S, E, n, ζ_0) tels que $\mu(\bigcup_{j=0}^{n-1} T^j F) \geq 1 - \delta$ et $\nu(\bigcup_{j=0}^{n-1} S^j E) \geq 1 - \delta$, et

$$d((\pi' \vee \xi)_{[0,n[}|_F) = d((\pi' \vee \xi)_{[0,n[}),$$

et
$$d\big(\big(\varphi_*(\xi_{[0,n[}) \vee (\pi \vee \zeta)_{[0,n[}\big|_E\big) = d(\varphi_*(\xi_{[0,n[}) \vee (\pi \vee \zeta)_{[0,n[}).$$
De plus, $\varphi_*(\xi_{[0,n[})$ induit une structure de colonnes sur le gadget, donc il existe une v.a. ξ'_0 sur $\bigcup_{j=0}^{n-1} S^j E$ telle que $\varphi_*(\xi_{[0,n[})\big|_E = \xi'_{[0,n[}\big|_E$. On a alors

$$\begin{aligned} d((\pi \vee \xi')_{[0,n[}\big|_E) &= d(\varphi_*(\pi'_{[0,n[}) \vee \varphi_*(\xi_{[0,n[})\big|_E) = d(\varphi_*(\pi'_{[0,n[}) \vee \varphi_*(\xi_{[0,n[})) \\ &= d(\pi'_{[0,n[} \vee \xi_{[0,n[}) = d((\pi' \vee \xi)_{[0,n[}\big|_F). \end{aligned}$$

Donc $(S, E, n, \pi_0 \vee \xi'_0)$ est isomorphe à $(T, F, n, \pi'_0 \vee \xi_0)$. De plus, on a

$$\begin{aligned} \frac{1}{n}\sum_{j=0}^{n-1} |\xi'_j|_E - \zeta_j|_E| &= \frac{1}{n}\sum_{j=0}^{n-1} |(\varphi_*\xi_j)|_E - \zeta_j|_E| \\ &= \frac{1}{n}\sum_{j=0}^{n-1} |\varphi_*\xi_j - \zeta_j| < \varepsilon. \end{aligned}$$

Et alors
$$\bar{d}\big((T, F, n, \pi'_0 \vee \xi_0), (S, E, n, \pi_0 \vee \zeta_0)\big) < \varepsilon.$$

□

3.5 Une version relative du lemme fondamental d'Ornstein

Le but de cette section est de prouver le résultat suivant :

Lemme 3.12 (Lemme fondamental d'Ornstein)**.** *Soient ζ et π des processus sur Y tels que ζ est relativement finiment déterminé par rapport à π. Soit $\varepsilon > 0$: il existe $\delta > 0$ et $n_1 \geq 1$ tels que, pour tout système (X, μ, T) ergodique tel que $h(T) \geq h(\pi \vee \zeta, S)$, muni de ξ et π' tels que $(T, \pi') \sim (S, \pi)$ et vérifiant*

(i) $|d((\pi' \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| \leq \delta$,

(ii) et $0 \leq h(\pi \vee \zeta, S) - h(\pi' \vee \xi, T) \leq \delta$,

il existe un processus ξ' sur X tel que

a) $(T, \pi' \vee \xi') \sim (S, \pi \vee \zeta)$, *b)* *et* $|\xi_0 - \xi'_0| \leq \varepsilon$.

Commençons par un résultat plus simple, qui sera utile plus tard, car il montre qu'il existe un processus qui satisfait (i) et (ii) du lemme fondamental. De plus, on verra que sa preuve contient déjà plusieurs éléments importants de la preuve du lemme fondamental.

Lemme 3.13. *Soient ζ et π deux processus sur Y tels que $h(\pi \vee \zeta, S) > h(\pi, S)$. Soit (X, μ, T) un système ergodique tel que $h(T) \geq h(\pi \vee \zeta, S)$, muni de π' tel que $(T, \pi') \sim (S, \pi)$. Pour tous $\delta > 0$ et $n_1 \geq 1$, il existe un processus ξ sur X tel que*

(i) $|d((\pi' \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| \leq \delta$,

(ii) *et* $0 \leq h(\pi \vee \zeta, S) - h(\pi' \vee \xi, T) \leq \delta$.

Dans le cas où $h(\pi \vee \zeta, S) = h(\pi, S)$, le résultat n'est vrai que pour T vérifiant $h(T) = h(\pi \vee \zeta, S)$.

Remarque 3.5. On va travailler sur l'espace des fonction mesurables prenant un nombre fini de valeurs entières muni de la distance :

$$|f - g| = \mu(f \neq g).$$

On laisse au lecteur de montrer que c'est un espace connexe et que $\xi_0 \mapsto h(\xi, T)$ est continue. Et, pour tout $k \geq 1$, l'espace des v.a. à valeurs dans $[k]$ est complet.

Démonstration. Soient $\delta > 0$ et $n_1 \geq 1$. Prenons de plus $\alpha > 0$ et $\beta >$ à fixer plus tard.

Supposons que $h(\pi \vee \zeta, S) > h(\pi, S)$. Alors, par la remarque 3.5 et le fait que $h(T) \geq h(\pi \vee \zeta, S) > h(\pi, S)$, il existe θ tel que

$$0 < h(\pi \vee \zeta, S) - h(\pi' \vee \theta, T) < \alpha. \tag{10}$$

On se fixe k, l et r tels que ζ est a valeurs dans $[k]$, θ est à valeurs dans $[l]$ et π et π' sont à valeurs dans $[r]$. On applique le théorème de Shannon-McMillan-Breiman (4.3) pour trouver n assez grand, $\mathcal{E}_n \subset ([r] \times [l])^n$ et $\mathcal{E}'_n \subset ([r] \times [k])^n$ tels que

$$\mu((\pi' \vee \theta)_{[0,n[} \in \mathcal{E}_n), \nu((\pi \vee \zeta)_{[0,n[} \in \mathcal{E}'_n) \geq 1 - \beta, \tag{11}$$

et $\forall \mathbf{i} \in \mathcal{E}_n$:

$$2^{-(h(\pi'\vee\theta,T)+\beta)n} \leq \mu((\pi' \vee \theta)_{[0,n[} = \mathbf{i}) \leq 2^{-(h(\pi'\vee\theta,T)-\beta)n}, \tag{12}$$

et $\forall \mathbf{i}' \in \mathcal{E}'_n$:

$$2^{-(h(\pi\vee\zeta,T)+\beta)n} \leq \nu((\pi \vee \zeta)_{[0,n[} = \mathbf{i}') \leq 2^{-(h(\pi\vee\zeta,T)+\beta)n}. \tag{13}$$

En particulier, par (10), pour β assez petit puis n assez grand, on a

$$\#\mathcal{E}_n \leq \#\mathcal{E}'_n. \tag{14}$$

Le théorème ergodique va ensuite nous donner de l'information sur les éléments de $\mathcal{E}'_n$: pour deux suites $\mathbf{i} \in ([r] \times [k])^n$ et $\mathbf{i}' \in ([r] \times [k])^{n_1}$, on note

$$f_n(\mathbf{i}, \mathbf{i}') := \frac{1}{n - n_1} \sum_{j=0}^{n-n_1} \mathbb{1}_{\{(\pi\vee\zeta)_{[0,n_1[}=\mathbf{i}'\}}(S^j(x)),\ \forall x \in \{(\pi \vee \zeta)_{[0,n[} = \mathbf{i}\}.$$

Plus simplement, $f_n(\mathbf{i}, \mathbf{i}')$ est le nombre moyen d'apparition de $\mathbf{i}'$ dans la suite $\mathbf{i}$. Bien que plus compliquée, la formule ci-dessus permet de voir que, grâce au théorème ergodique et au théorème d'Egorov, on aurait pu choisir n assez grand et $\mathcal{E}'_n$ tel que, pour tout $\mathbf{i} \in \mathcal{E}'_n$:

$$\sum_{\mathbf{i}'\in([r]\times[k])^{n_1}} |f_n(\mathbf{i}, \mathbf{i}') - \nu((\pi \vee \zeta)_{[0,n_1[} = \mathbf{i}')| \leq \beta. \tag{15}$$

On applique ensuite le théorème de Rohlin 3.6 pour construire F' tel que $\mu(\bigcup_{j=0}^{n-1} T^j F') \geq 1 - \beta$ et

$$d((\pi' \vee \theta)_{[0,n[}\big|_{F'}) = d((\pi' \vee \theta)_{[0,n[}). \tag{16}$$

On pose ensuite $F = F' \cap \{(\pi' \vee \theta)_{[0,n[} \in \mathcal{E}_n\}$. On remarque que, en notant $G = \bigcup_{j=0}^{n-1} T^j F$, on a $\mu(G) \geq 1 - 2\beta$.

Pour $\mathbf{b} \in [r]^n$, notons $\mathcal{E}_n(\mathbf{b})$ (resp. $\mathcal{E}'_n(\mathbf{b})$) l'ensemble des éléments de $\mathcal{E}_n$ (resp. $\mathcal{E}'_n$) de la forme $\mathbf{b} \times \mathbf{a}$, avec $\mathbf{a} \in [l]^n$ (resp. $\mathbf{a} \in [k]^n$). On pose alors

$$\mathcal{B}_n = \{\mathbf{b} \in [r]^n \mid \#\mathcal{E}_n(\mathbf{b}) \leq \#\mathcal{E}'_n(\mathbf{b})\}.$$

Par (11), (12) et (13), on peut voir que

$$\begin{aligned} \mu(\pi'_{[0,n[} \notin \mathcal{B}_n) &\leq \beta \frac{2^{(h(\pi\vee\zeta,S)-h(\pi'\vee\theta,T)-2\beta)n}}{2^{(h(\pi\vee\zeta,S)-h(\pi'\vee\theta,T)-2\beta)n} - 1} \\ &\leq 2\beta, \end{aligned} \tag{17}$$

pour β assez petit et n assez grand.

Comme on a l'inégalité (14), il existe une injection $\varphi : \mathcal{E}_n \longrightarrow \mathcal{E}'_n$ telle que $\forall \mathbf{b} \in \mathcal{B}_n$,

$$\varphi(\mathcal{E}_n(\mathbf{b})) \subset \mathcal{E}'_n(\mathbf{b}). \tag{18}$$

On peut donc regarder la fonction $\varphi \circ (\pi' \vee \theta)_{[0,n[}\big|_F$ définie sur F, à valeurs dans $([r] \times [k])^n$. Il existe donc une variable aléatoire $\pi_0^* \vee \xi_0$ sur $\bigcup_{j=0}^{n-1} T^j F$ à valeurs dans $[r] \times [k]$ telle que $(\pi^* \vee \xi)_{[0,n[}\big|_F = \varphi \circ (\pi' \vee \theta)_{[0,n[}\big|_F$. C'est-à-dire, comme φ est injective, que $\pi_0^* \vee \xi_0$ engendre les mêmes colonne que $\pi'_0 \vee \theta_0$. On étend ensuite $\pi_0^* \vee \xi_0$ à l'espace entier de manière quelconque. On a définit φ de sorte que π' ne soit pas trop modifié : en effet, par (17) et (18), on a

$$\begin{aligned} |\pi'_0 - \pi_0^*| &\leq \mu(\{\pi'_0 \neq \pi_0^*\} \cap G) + 2\beta \\ &\leq n\mu(\pi'_{[0,n[}\big|_F \neq \pi^*_{[0,n[}\big|_F) + 2\beta \\ &\leq n\mu(\pi'_{[0,n[}\big|_{F'} \notin \mathcal{B}_n) + 2\beta \\ &= n\mu(F')\ \mu(\pi'_{[0,n[} \notin \mathcal{B}_n) + 2\beta \\ &\leq 4\beta, \end{aligned} \tag{19}$$

où l'avant-dernière ligne s'obtient par la condition d'indépendance (16).

- Montrons que $\pi' \vee \xi$ vérifie (i) :

Pour tout $A \subset F$ de la forme $A = \{(\pi^* \vee \xi)_{[0,n[} = \mathbf{i}\}$, on note $C_A^{n_1}$ la colonne associée à A mais tronquée de ses $n_1 - 1$ derniers étages, c'est-à-dire

$$C_A^{n_1} = \bigcup_{j=0}^{n-n_1} T^j A.$$

Notons $G_{n_1} = \bigcup_{j=0}^{n-n_1} T^j F = \bigcup_{\mathbf{i} \in \mathcal{E}'_n} C^{n_1}_{\{(\pi^* \vee \xi)_{[0,n[} = \mathbf{i}\}}$. Avec ces notations, (i) va découler de (15) :

$$|d((\pi^* \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| = \sum_{\mathbf{i}' \in [k]^{n_1}} \left|\mu\left((\pi^* \vee \xi)_{[0,n_1[} = \mathbf{i}'\right) - \nu\left((\pi \vee \zeta)_{[0,n_1[} = \mathbf{i}'\right)\right|$$

$$\leq \sum_{\mathbf{i}' \in [k]^{n_1}} \left|\mu\left(\{(\pi^* \vee \xi)_{[0,n_1[} = \mathbf{i}'\} \cap G_{n_1}\right) - \nu\left((\pi \vee \zeta)_{[0,n_1[} = \mathbf{i}'\right)\mu\left(G_{n_1}\right)\right| + 2\mu\left(G_{n_1}^c\right)$$

$$\leq \sum_{\mathbf{i}' \in [k]^{n_1}} \sum_{\mathbf{i} \in \mathcal{E}'_n} \Big|\mu\left(\{(\pi^* \vee \xi)_{[0,n_1[} = \mathbf{i}'\} \cap C^{n_1}_{\{(\pi^* \vee \xi)_{[0,n[} = \mathbf{i}\}}\right)$$

$$- \nu\left((\pi \vee \zeta)_{[0,n_1[} = \mathbf{i}'\right)\mu\left(C^{n_1}_{\{(\pi^* \vee \xi)_{[0,n[} = \mathbf{i}\}}\right)\Big| + 2(2\beta + \frac{n_1}{n}).$$

Il suffit alors de remarquer que

$$\mu\left(\{(\pi^*\vee\xi)_{[0,n_1[}=\mathbf{i}'\}\cap C^{n_1}_{\{(\pi^*\vee\xi)_{[0,n[}=\mathbf{i}\}}\right)=f_n(\mathbf{i},\mathbf{i}')\cdot\mu\left(C^{n_1}_{\{(\pi^*\vee\xi)_{[0,n[}=\mathbf{i}\}}\right),$$

pour obtenir

$$|d((\pi^*\vee\xi)_{[0,n_1[})-d((\pi\vee\zeta)_{[0,n_1[})|\leq 5\beta+2\frac{n_1}{n}.$$

Donc, en ajoutant à cela l'estimation (19), pour β assez petit puis n assez grand, $\pi'\vee\xi$ vérifie la condition (i).

- Montrons que $\pi'\vee\xi$ vérifie (ii) :

On pose $\rho_0=\mathbb{1}_F|_G$. Par le lemme 3.7, vu que $\pi'\vee\theta$ et $\pi^*\vee\xi$ engendrent les mêmes colonnes sur G, on a

$$\pi'_0\vee\theta_0\preccurlyeq\bigvee_{j=-n}^{n}(\pi^*_0\vee\xi_0\vee\rho_0)\circ T^j,\quad\text{sur } G.$$

Il existe alors χ tel que l'identité ci-dessus soit vraie sur l'espace entier et $|(\pi'\vee\theta)_0-\chi_0|\leq 2\beta$. Soit $\gamma>0$. Pour β assez petit, on a alors

$$\begin{aligned}h(\pi'\vee\theta,T)&\leq h(\chi,T)+\gamma\leq h(\pi^*\vee\xi\vee\rho,T)+\gamma\\&\leq h(\pi^*\vee\xi,T)+H(\rho_0)+\gamma.\end{aligned}$$

En répétant ce raisonnement en inversant $\pi'\vee\theta$ et $\pi^*\vee\xi$, on obtient

$$|h(\pi'\vee\theta,T)-h(\pi^*\vee\xi,T)|\leq H(\rho_0)+\gamma.$$

Alors, en ajoutant à cela les estimations (10) et (19), pour α, γ, β assez petits et n assez grand, $\pi'\vee\xi$ vérifie la condition (ii).

Il nous reste à traiter le cas $h(\pi\vee\zeta,S)=h(\pi,S)$, mais il est beaucoup plus simple : comme on suppose que $h(T)=h(\pi,S)$, la condition (ii) est automatiquement vérifiée. Alors une application du théorème 3.6 puis du lemme 3.9 permet de construire ξ vérifiant (i). □

Ensuite, on poursuit le travail avec le lemme suivant, qui reprend les idées de la construction ci-dessus en utilisant les propriétés de finie détermination pour mieux choisir φ, et constitue l'étape principale vers le lemme fondamental.

Lemme 3.14. *Soient ζ et π des processus sur Y tels que ζ est relativement finiment déterminé par rapport à π. Soit $\varepsilon > 0$: il existe $\delta > 0$ et $n_1 \geq 1$ tels que, pour tout système (X, μ, T) ergodique tel que $h(T) \geq h(\pi \vee \zeta, S)$, muni de ξ et π' tels que $(T, \pi') \sim (S, \pi)$ et vérifiant*

(i) $|d((\pi' \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| \leq \delta$,

(ii) *et* $0 \leq h(\pi \vee \zeta, S) - h(\pi' \vee \xi, T) \leq \delta$,

pour tout $\delta' > 0$ et pour tout $p \geq 1$, il existe un processus ξ' sur X tel que

(a) $|d((\pi' \vee \xi')_{[0,p[}) - d((\pi \vee \zeta)_{[0,p[})| \leq \delta'$,

(b) $0 \leq h(\pi \vee \zeta, S) - h(\pi' \vee \xi', T) \leq \delta'$,

(c) *et* $|\xi_0 - \xi'_0| \leq \varepsilon$.

Démonstration. Fixons k et r tels que ζ_0 soit à valeurs dans $[k]$ et π et π' soient à valeurs dans $[r]$.

Soit ε', à fixer plus tard, et δ et n_1 associés par la finie détermination relative de (S, ζ). Fixons un processus ξ vérifiant (i) et (ii). On a choisi δ et n_1 pour avoir

$$d\big((S, \zeta), (T, \xi) \,|\, \pi', \pi\big) < \varepsilon'. \tag{20}$$

Soient $\delta' > 0$ et $p \geq 1$.

- Commençons par éliminer un cas simple : supposons que pour tout processus ξ' sur l'alphabet $[k]$ vérifiant $|\xi_0 - \xi'_0| \leq 2\varepsilon'$, on aie $h(\pi' \vee \xi', T) \geq h(\pi \vee \zeta, S)$.

On se donne $\alpha > 0$, et par le lemme 3.11 on construit deux gadgets $(T, F, n, \pi'_0 \vee \xi_0)$ et $(S, E, n, \pi_0 \vee \zeta_0)$ tels que $\mu(\bigcup_{j=0}^{n-1} T^j F), \nu(\bigcup_{j=0}^{n-1} S^j E) \geq 1 - \alpha$ et

$$\bar{d}\big((T, F, n, \pi'_0 \vee \xi_0), (S, E, n, \pi_0 \vee \zeta_0)\big) < \varepsilon'.$$

Par définition, il existe alors ξ' sur X tel que

$$(T, F, n, \pi'_0 \vee \xi'_0) \cong (S, E, n, \pi_0 \vee \zeta_0), \tag{21}$$

et

$$\frac{1}{n}\sum_{j=0}^{n-1}|\xi'_0|_{T^jF}-\zeta_0|_{T^jF}|\leq\varepsilon'. \tag{22}$$

De plus, il nous est possible de choisir ξ' tel qu'il soit $\pi'\vee\xi$-mesurable. On pose $G=\bigcup_{j=0}^{n-1}T^jF$, et par (22), on a

$$|\xi_0-\xi'_0|\leq\mu(\{\xi_0\neq\xi'_0\}\cap G)+\mu(G^c)\leq\varepsilon'+\alpha\leq 2\varepsilon',$$

si $\alpha\leq\varepsilon'$. Alors, par notre hypothèse, on a $h(\pi'\vee\xi',T)\geq h(\pi\vee\zeta,S)$. Mais, comme ξ' est $\pi'\vee\xi$-mesurable, on a aussi

$$h(\pi'\vee\xi',T)\leq h(\pi'\vee\xi,T)\leq h(\pi\vee\zeta,S),$$

donc ξ' vérifie (b). De plus, par (21), on a

$$|d((\pi'\vee\xi')_{[0,p[})-d((\pi\vee\zeta)_{[0,p[})|\leq 2\alpha+2\frac{p}{n},$$

et dons, si $\alpha<\frac{\delta'}{2}$ et n est assez grand, on obtient que ξ' vérifie (a).

- On finit la preuve en traitant le cas où l'on a :

$$0<h(\pi\vee\zeta,S)-h(\pi'\vee\xi,T)\leq\delta. \tag{23}$$

Soient $\alpha>0$, $\beta>0$, et $n\geq 1$ à fixer plus tard. Par (23), on trouve θ tel que $\xi_0\preccurlyeq\theta_0$ et

$$0<h(\pi\vee\zeta,S)-h(\pi'\vee\theta,T)<\alpha. \tag{24}$$

On fixe $l\geq 1$ tel que θ soit à valeurs dans $[l]$. Comme dans la preuve du lemme 3.13, on trouve $\mathcal{E}_n\subset([r]\times[l])^n$ et $\mathcal{E}'_n\subset([r]\times[k])^n$ tels que $\mu((\pi'\vee\theta)_{[0,n[}\in\mathcal{E}_n),\nu((\pi\vee\zeta)_{[0,n[}\in\mathcal{E}'_n)\geq 1-\beta$, et $\#\mathcal{E}_n\leq\#\mathcal{E}'_n$, ainsi que

$$\forall\mathbf{i}\in\mathcal{E}'_n,\quad\sum_{\mathbf{i}'\in([r]\times[k])^{n_1}}|f_n(\mathbf{i},\mathbf{i}')-\nu((\pi\vee\zeta)_{[0,n_1[}=\mathbf{i}')|\leq\beta. \tag{25}$$

Et on construit un gadget $(T,F',n,\pi'\vee\theta_0)$ tel que $\mu(\bigcup_{j=0}^{n-1}T^jF')\geq 1-\beta$ et

$$d((\pi'\vee\theta)_{[0,n[}\big|_{F'})=d((\pi'\vee\theta)_{[0,n[}). \tag{26}$$

On note aussi $F=F'\cap\{(\pi'\vee\theta)_{[0,n[}\in\mathcal{E}_n\}$, et en posant $G=\bigcup_{j=0}^{n-1}T^jF$, on a $\mu(G)\geq 1-2\beta$.

Comme dans la preuve du lemme précédent, on va construire une injection $\varphi : \mathcal{E}_n \longrightarrow \mathcal{E}'_n$ et une variable aléatoire ξ'_0 selon le même procédé. Cependant, cette fois on va construire φ de sorte que ξ' vérifie (c).

On construit un gadget $(S, E, n, \pi_0 \vee \zeta_0)$ tel que $\nu(\bigcup_{j=0}^{n-1} S^j E) \geq 1 - \beta$ et

$$d((\pi \vee \zeta)_{[0,n[}\big|_E) = d((\pi \vee \zeta)_{[0,n[}). \tag{27}$$

Par (20), (quitte à modifier un peu F') il existe un nouveau processus ζ^* sur X telle que

$$(T, F', n, \pi'_0 \vee \zeta^*_0) \cong (S, E, n, \pi_0 \vee \zeta_0),$$

et

$$\frac{1}{n} \sum_{j=0}^{n-1} |\xi_0|_{T^j F'} - \zeta^*_0|_{T^j F'}| \leq \varepsilon'.$$

Lorsque l'on remplace F' par F, on a les estimations suivantes :

$$\begin{aligned} |d((\pi \vee \zeta)_{[0,n[}\big|_E) - d((\pi' \vee \zeta^*)_{[0,n[}\big|_F)| &\leq |d((\pi' \vee \zeta^*)_{[0,n[}\big|_{F'}) - d((\pi' \vee \zeta^*)_{[0,n[}\big|_F)| \\ &\leq 2\frac{\mu(F \Delta F')}{\mu(F')} = 2\mu((\pi' \vee \theta)_{[0,n[} \notin \mathcal{E}_n) \leq 2\beta \end{aligned} \tag{28}$$

et

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} |\zeta^*_0|_{T^j F} - \xi_0|_{T^j F}| &\leq \frac{1}{n} \sum_{j=0}^{n-1} |\zeta^*_0|_{T^j F'} - \xi_0|_{T^j F'}| \; \frac{\mu(F')}{\mu(F)} \\ &\leq \frac{\varepsilon'}{1-\beta} \leq 2\varepsilon', \end{aligned} \tag{29}$$

si $\beta \leq 1/2$. Pour obtenir ces estimations, on a utilisé le fait que F' et $\{(\pi' \vee \theta)_{[0,n[} \in \mathcal{E}_n\}$ sont indépendants, ce qui découle de (26).

Par l'estimation fournie par le corollaire 4.3.1, on voit que, pour n assez grand, on a, pour tous $\mathbf{i} \in \mathcal{E}_n$ et $\mathbf{i}' \in \mathcal{E}'_n$

$$4\,\nu((\pi \vee \zeta)_{[0,n[} = \mathbf{i}') \leq \mu((\pi' \vee \theta)_{[0,n[} = \mathbf{i}).$$

En utilisant la définition de ζ^* et les propriété d'indépendance (26) et (27), il en découle que, pour tous $\mathbf{i} \in \mathcal{E}_n$ et $\mathbf{i}' \in \mathcal{E}'_n$

$$4\,\mu(\{(\pi' \vee \zeta^*)_{[0,n[} = \mathbf{i}'\} \cap F) \leq \mu(\{(\pi' \vee \theta)_{[0,n[} = \mathbf{i}\} \cap F). \tag{30}$$

Introduisons une nouvelle notation : si γ et χ sont deux processus, et $\mathbf{i}$ et $\mathbf{j}$ sont deux suites telles que $\{\gamma_{[0,n[} = \mathbf{i}\} \subset \{\chi_{[0,n[} = \mathbf{j}\}$, on dit que $\mathbf{j}$ est le (χ, n)-nom de $\mathbf{i}$. Comme $\xi_0 \preccurlyeq \theta_0$, pour chaque $\mathbf{i} \in \mathcal{E}_n$, il existe un unique $\mathbf{j} \in [k]^n$ tel que $\mathbf{j}$ est le (ξ, n)-nom de $\mathbf{i}$.

Pour tout $\mathbf{i} \in \mathcal{E}_n$, on pose $\mathcal{J}_{\mathbf{i}}$ l'ensemble des mots $\mathbf{i}' \in \mathcal{E}'_n$ tels que le (ζ, n)-nom de $\mathbf{i}'$ diffère du (ξ, n)-nom de $\mathbf{i}$ en moins de $n\sqrt{2\varepsilon'}$ endroits, et le (π, n)-nom de $\mathbf{i}'$ est le (π', n)-nom de $\mathbf{i}$. Cela nous permet de poser :

$$\mathcal{I} = \left\{ \mathbf{i} \in \mathcal{E}_n \,|\, \mu\left(\{(\pi' \vee \theta)_{[0,n[}\big|_F = \mathbf{i}\} \cap \{(\pi' \vee \zeta^*)_{[0,n[}\big|_F \in \mathcal{J}_{\mathbf{i}}\}\right) \geq \frac{\mu((\pi' \vee \theta)_{[0,n[}\big|_F = \mathbf{i})}{2} \right\}.$$

Enfin, on pose $\Omega = \{x \in F \,|\, \xi_{[0,n[}(x)$ et $\zeta^*_{[0,n[}(x)$ diffèrent en au moins $n\sqrt{2\varepsilon'}$ endroits$\}$, et, grâce à (28), on a l'estimation

$$\begin{aligned}
\mu((\pi' \vee \theta)_{[0,n[}\big|_F \notin \mathcal{I}) &= \sum_{\mathbf{i} \notin \mathcal{I}} \mu((\pi' \vee \theta)_{[0,n[}\big|_F = \mathbf{i}) \\
&\leq \sum_{\mathbf{i} \notin \mathcal{I}} 2\mu(\{(\pi' \vee \theta)_{[0,n[}\big|_F = \mathbf{i}\} \cap \{(\pi' \vee \zeta^*)_{[0,n[}\big|_F \notin \mathcal{J}_{\mathbf{i}}\}) \\
&\leq \sum_{\mathbf{i} \notin \mathcal{I}} 2[\mu\big(\{(\pi' \vee \theta)_{[0,n[}\big|_F = \mathbf{i}\} \cap \Omega\big) \\
&\qquad\qquad + \mu\big(\{(\pi' \vee \theta)_{[0,n[}\big|_F = \mathbf{i}\} \cap \{(\pi' \vee \zeta^*)_{[0,n[}\big|_F \notin \mathcal{E}'_n\}\big)] \\
&\leq 2[\mu(\Omega) + \mu((\pi' \vee \zeta^*)_{[0,n[}\big|_F \notin \mathcal{E}'_n)] \\
&= 2[\mu(\Omega) + \nu((\pi \vee \zeta)_{[0,n[} \notin \mathcal{E}'_n)\mu(F) + 2\beta\mu(F)] \leq 2[\mu(\Omega) + 3\beta\mu(F)].
\end{aligned}$$

Or, d'après le lemme 3.10 et (29), on a $\mu(\Omega) \leq \sqrt{2\varepsilon'}\mu(F)$. Donc pour ε' et β assez petits, on a

$$\mu((\pi' \vee \theta)_{[0,n[}\big|_F \in \mathcal{I}) \geq (1 - \varepsilon/3)\mu(F). \tag{31}$$

On voit aussi que

$$\mu(\pi'_{[0,n[}\big|_F \neq \pi^*_{[0,n[}\big|_F) \leq \mu((\pi' \vee \theta)_{[0,n[} \notin \mathcal{I}) \leq 2(\sqrt{2\varepsilon'} + 3\beta)\mu(F). \tag{32}$$

On veut maintenant définir une injection φ sur $\mathcal{I}$ telle que $\varphi(\mathbf{i}) \in \mathcal{J}_{\mathbf{i}}$. Pour cela, on va faire appel au lemme des mariages. Soit $\mathcal{I}' \subset \mathcal{I}$. En utilisant la définition de

$\mathcal{I}$ et (30), on a

$$\begin{aligned}
\mu\Bigg(\bigcup_{\mathbf{i}\in\mathcal{I}'} \{(\pi'\vee\theta)_{[0,n[}\big|_F=\mathbf{i}\}\cap\{(\pi'\vee\zeta^*)_{[0,n[}\big|_F\in\mathcal{J}_{\mathbf{i}}\}\Bigg)& \\
&= \sum_{\mathbf{i}\in\mathcal{I}'}\mu\left(\{(\pi'\vee\theta)_{[0,n[}\big|_F=\mathbf{i}\}\cap\{\pi'\vee\zeta^*)_{[0,n[}\big|_F\in\mathcal{J}_{\mathbf{i}}\}\right)\\
&\geq \sum_{\mathbf{i}\in\mathcal{I}'}\frac{\mu((\pi'\vee\theta)_{[0,n[}\big|_F=\mathbf{i})}{2}\\
&\geq 2\ \sup_{\mathbf{j}\in\mathcal{E}'_n}\{\mu((\pi'\vee\zeta^*)_{[0,n[}\big|_F=\mathbf{j})\}\ \#\mathcal{I}'
\end{aligned}$$

Mais on a aussi,

$$\begin{aligned}
\mu\Bigg(\bigcup_{\mathbf{i}\in\mathcal{I}'} \{(\pi'\vee\theta)_{[0,n[}\big|_F=\mathbf{i}\}\cap\{(\pi'\vee\zeta^*)_{[0,n[}\big|_F\in\mathcal{J}_{\mathbf{i}}\}\Bigg)& \\
&= \mu\left(\bigcup_{\mathbf{i}\in\mathcal{I}'}\bigcup_{\mathbf{j}\in\mathcal{J}_{\mathbf{i}}}\{(\pi'\vee\theta)_{[0,n[}\big|_F=\mathbf{i}\}\cap\{(\pi'\vee\zeta^*)_{[0,n[}\big|_F=\mathbf{j}\}\right)\\
&\leq \mu\left(\bigcup_{\mathbf{i}\in\mathcal{I}'}\bigcup_{\mathbf{j}\in\mathcal{J}_{\mathbf{i}}}\{(\pi'\vee\zeta^*)_{[0,n[}\big|_F=\mathbf{j}\}\right)\\
&\leq \sup_{\mathbf{j}\in\mathcal{E}'_n}\{\mu((\pi'\vee\zeta^*)_{[0,n[}\big|_F=\mathbf{j})\}\ \#\bigcup_{\mathbf{i}\in\mathcal{I}'}\mathcal{J}_{\mathbf{i}}
\end{aligned}$$

Cela établit la “condition de mariage” :

$$\#\mathcal{I}'\leq 2\#\mathcal{I}'\leq\#\bigcup_{\mathbf{i}\in\mathcal{I}'}\mathcal{J}_{\mathbf{i}}.$$

On peut alors appliquer le lemme des mariages pour construire une injection φ : $\mathcal{I}\longrightarrow\mathcal{E}'_n$ telle que $\varphi(\mathbf{i})\in\mathcal{J}_{\mathbf{i}}$. Comme on a $\mathcal{E}_n\leq\#\mathcal{E}'_n$, on peut étendre φ en une injection de $\mathcal{E}_n$ dans $\mathcal{E}'_n$. On a vu dans la preuve du lemme précédent (3.13), qu'on peut alors prendre deux processus sur X π^*et ξ' tels que

$$(\pi^*\vee\xi')_{[0,n[}\big|_F=\varphi\circ(\pi^*\vee\theta)_{[0,n[}\big|_F,$$

et que, pour α, β assez petits et n assez grand, $\pi'\vee\xi'$ vérifie (a) et (b) (on obtient

(19) grâce à (32)). Montrons que ξ' vérifie (c) :

$$\begin{aligned}
|\xi_0 - \xi_0'| &\leq \mu(G)|\xi_0|_G - \xi_0'|_G| + \mu(G^c) \\
&\leq \sum_{\mathbf{i}\in\mathcal{I}} \mu\left(\{\xi_0 \neq \xi_0'\} \cap C_{\{(\pi'\vee\theta)_{[0,n[}|_F=\mathbf{i}\}}\right) + \mu((\pi'\vee\theta)_{[0,n[}|_F \notin \mathcal{I}) + 2\beta \\
&\leq \sum_{\mathbf{i}\in\mathcal{I}} \sqrt{2\varepsilon'}\mu\left(C_{\{(\pi'\vee\theta)_{[0,n[}|_F=\mathbf{i}\}}\right) + \frac{\varepsilon}{3} + 2\beta \\
&\leq \frac{2\varepsilon}{3} + 2\beta, \quad \text{si } \varepsilon' \text{ est assez petit.}
\end{aligned}$$

Et alors, si $\beta \leq \varepsilon/3$, on a prouvé que ξ' vérifie (c), et la preuve est complète. □

Démonstration du lemme fondamental. Soient (S,ζ) un processus relativement finiment déterminé par rapport à π, à valeurs dans $[k]$ et $\varepsilon > 0$. Fixons une suite $(\varepsilon_r)_{r\geq 1}$ telle que $\sum_{r\geq 1}\varepsilon_r = \varepsilon$. Pour tout r, on prend $\delta_r > 0$ et n_r associés à ε_r, donnés par le lemme 3.14.

Soit (X,μ,Ti) un système ergodique tel que $h(T) \geq h(\pi\vee\zeta, S)$, et ξ, π' qui vérifient $(T,\pi') \sim (S,\pi)$, et les conditions (i) et (ii) pour δ_1 et n_1. En utilisant le lemme 3.14, on construit par récurrence une suite de processus $(\xi^{(r)})_{r\geq 1}$ telle que $\xi^{(1)} = \xi$ et :

(a) $|d((\pi'\vee\xi^{(r+1)})_{[0,n_{r+1}[}) - d((\pi\vee\zeta)_{[0,n_{r+1}[})| \leq \delta_{r+1}$,

(b) $0 \leq h(\pi\vee\zeta, S) - h(\pi'\vee\xi^{(r+1)}, T) \leq \delta_{r+1}$,

(c) et $|\xi_0^{(r+1)} - \xi_0^{(r)}| \leq \varepsilon_r$.

De plus, on peut supposer que $\delta_r \to 0$ et $n_r \to \infty$. Comme chaque $\xi^{(r)}$ est à valeurs dans $[k]$, (c) et la remarque 3.5 implique que la suite $(\xi_0^{(r)})_{r\geq 1}$ converge. Notons ξ_0' sa limite et ξ' le processus associé. Alors (a) implique que

$$\forall p \geq 1,\ d((\pi'\vee\xi')_{[0,p[}) = d((\pi\vee\zeta)_{[0,p[}),$$

et (c) implique que

$$|\xi_0 - \xi_0'| \leq \varepsilon.$$

□

3.6 Preuve du théorème d'isomorphisme

Dans cette section, nous allons utiliser le lemme fondamental pour prouver le théorème 3.2. Soient (X, μ, T) et (Y, ν, S) deux systèmes ergodiques munis de processus ξ, π' et ζ, π tels que $(T, \pi') \sim (S, \pi)$ et $\pi' \vee \xi$, $\pi \vee \zeta$ sont générateurs et ξ (resp. ζ) est relativement finiment déterminé par rapport à π' (resp. π). Pour prouver le théorème 3.2, il va nous suffire de construire un processus η sur X tel que $\pi' \vee \eta$ soit générateur et

$$(T, \pi' \vee \eta) \sim (S, \pi \vee \zeta). \tag{33}$$

Par le lemme fondamental démontré dans la partie précédente, et le lemme 3.13, on peut déjà construire un processus vérifiant (33), mais le processus ainsi obtenu ne sera pas générateur. Nous allons voir maintenant que le lemme fondamental peut être utilisé pour prouver des résultats plus précis.

Commençons par une nouvelle notation : dans ce qui suit, on va souvent s'intéresser aux variables aléatoires $\xi_{[-n,n]}$-mesurables. Or un résultat classique de probabilité nous dit que toute v.a. $\xi_{[-n,n]}$-mesurable à valeurs dans un alphabet quelconque A, s'écrit sous la forme $\Pi \circ \xi_{[-n,n]}$, avec Π une fonction de $[k]^{[-n,n]}$ dans A. On notera alors

$$\Pi(\xi) = \Pi \circ \xi_{[-n,n]}.$$

Lemme 3.15. *Fixons k tel que ξ soit à valeurs dans $[k]$. Supposons qu'il existe un processus ρ sur X tel que $(T, \pi' \vee \rho) \sim (S, \pi \vee \zeta)$. Soit $\varepsilon > 0$, et, comme $\pi' \vee \xi$ est générateur, on peut fixer $N \geq 1$ tel que $\rho_0 \preccurlyeq_\varepsilon (\pi' \vee \xi)_{[-N,N]}$.*

Alors il existe ξ' sur X tel que

(i) $(T, \pi' \vee \xi') \sim (T, \pi' \vee \xi)$,

(ii) ξ' *est* $\pi' \vee \rho$*-mesurable,*

(iii) et $\rho_0 \preccurlyeq_{2\varepsilon} (\pi' \vee \xi)'_{[-N,N]}$.

Remarque 3.6. La preuve du lemme donne un résultat un peu plus précis que (iii) : si Π est une fonction sur $([r] \times [k])^{[-N,N]}$ telle que $|\Pi(\pi' \vee \xi) - \rho_0| \leq \varepsilon$, alors on a $|\Pi(\pi' \vee \xi') - \rho_0| \leq 2\varepsilon$.

Démonstration. Fixons $l \geq 1$ et $r \geq 1$ tels que ρ et ζ soient à valeurs dans $[l]$, et π, π' soient à valeurs dans $[r]$.

Notons $\mathcal{F}$ la tribu du système (X, μ, T), et $\mathcal{F}_{\pi' \vee \rho}$ la sous-tribu invariante engendrée par $\pi' \vee \rho$. On notera $(X_{\pi' \vee \rho}, \mu_{\pi' \vee \rho}, T_{\pi' \vee \rho})$ le facteur associé, $p : X \longrightarrow$

$X_{\pi'\vee\rho}$ la projection, et on notera $\bar{\rho}$ et $\bar{\pi}'$ les processus induits sur le facteur $(X_{\pi'\vee\rho}, \mu_{\pi'\vee\rho}, T_{\pi'\vee\rho})$ par ρ et π'. Le fait important à garder en tête est que la tribu engendrée par p est $\mathcal{F}_{\pi'\vee\rho}$ (en fait, $p = \pi' \vee \rho$, mais on garde deux notations distinctes pour éviter la confusion).

Soient $\varepsilon > 0$ et $N \geq 0$ tels que

$$\rho_0 \preccurlyeq_\varepsilon (\pi' \vee \xi)_{[-N,N]}.$$

On introduit $\alpha > 0$, $\beta > 0$, et $n \geq 1$, à fixer plus tard. Il existe $N_1 \geq N$ tel que

$$\rho_0 \preccurlyeq_\alpha \xi_{[-N_1,N_1]}.$$

On applique le théorème de Rohlin (3.6) sur $(X_{\pi'\vee\rho}, \mu_{\pi'\vee\rho}, T_{\pi'\vee\rho})$ pour construire un gadget $(T_{\pi'\vee\rho}, F_{\pi'\vee\rho}, n, \bar{\pi}'_0 \vee \bar{\rho}_0)$ tel que $\mu_{\pi'\vee\rho}(\bigcup_{j=0}^{n-1} T^j_{\pi'\vee\rho} F_{\pi'\vee\rho}) \geq 1 - \beta$, et

$$d((\bar{\pi}' \vee \bar{\rho})_{[0,n[}) = d((\bar{\pi}' \vee \bar{\rho})_{[0,n[}\Big|_{F_{\pi'\vee\rho}}).$$

En prenant $F = p^{-1}(F_{\pi'\vee\rho})$, on obtient un gadget sur X qui vérifie

$$(T, F, n, \pi' \vee \rho) \cong (T_{\pi'\vee\rho}, F_{\pi'\vee\rho}, n, \bar{\pi}' \vee \bar{\rho}).$$

Alors, par le lemme 3.9, il existe $\tilde{\xi}_{\pi'\vee\rho}$ sur $X_{\pi'\vee\rho}$ tel que

$$(T, F, n, \pi' \vee \rho \vee \xi) \cong (T_{\pi'\vee\rho}, F_{\pi'\vee\rho}, n, \bar{\pi}' \vee \bar{\rho} \vee \tilde{\xi}_{\pi'\vee\rho}). \tag{34}$$

On veut maintenant appliquer le lemme fondamental à $(T, \pi' \vee \xi)$ pour modifier $(T_{\pi'\vee\rho}, \bar{\pi}' \vee \tilde{\xi}_{\pi'\vee\rho})$; soit $\varepsilon' > 0$ et prenons $\delta > 0$, $n_1 \geq 1$ donnés par le lemme fondamental appliqué à $(T, \pi' \vee \xi)$.

Pour montrer que $(T_{\pi'\vee\rho}, \bar{\pi}' \vee \tilde{\xi}_{\pi'\vee\rho})$ vérifie la condition (i) du lemme fondamental, on note $G^{n_1} = \bigcup_{j=0}^{n-n_1} T^j F$ et $G^{n_1}_{\pi'\vee\rho} = \bigcup_{j=0}^{n-n_1} T^j_{\pi'\vee\rho} F_{\pi'\vee\rho}$, et l'isomorphisme ci-dessus donne

$$d((\pi' \vee \xi)_{[0,n_1[}\big|_{G^{n_1}}) = d((\bar{\pi}' \vee \tilde{\xi}_{\pi'\vee\rho})_{[0,n_1[}\Big|_{G^{n_1}_{\pi'\vee\rho}}).$$

Et alors

$$|d((\pi' \vee \xi)_{[0,n_1[}) - d((\bar{\pi}' \vee \tilde{\xi}_{\pi'\vee\rho})_{[0,n_1[})| \leq 2(\beta + \frac{n_1}{n}) \leq \delta, \tag{35}$$

si β est assez petit et n est assez grand.

Montrons que $(T_{\pi'\vee\rho}, \bar{\pi}' \vee \tilde{\xi}_{\pi'\vee\rho})$ vérifie (ii) du lemme fondamental : soit Π' une fonction sur $([r] \times [k])^{[-N_1,N_1]}$ telle que $|\Pi'(\pi' \vee \xi) - \rho_0| \leq \alpha$. On se donne aussi $K = \bigcup_{j=N_1}^{n-N_1-1} T^j F$ et $K_{\pi' \vee \rho} = \bigcup_{j=N_1}^{n-N_1-1} T_{\pi' \vee \rho}{}^j F_{\pi' \vee \rho}$, c'est-à-dire les piles privées de leurs N_1 premiers et N_1 derniers étages. Alors (34) nous donne :

$$d((\Pi'(\pi' \vee \xi) \vee \rho_0)\,|_K) = d\left(\left(\Pi'(\bar{\pi'} \vee \tilde{\xi}_{\pi' \vee \rho}) \vee \bar{\rho}_0\right)\Big|_{K_{\pi' \vee \rho}}\right).$$

Si, de plus, $\frac{2N_1}{n} \leq \beta$, alors $\mu(K), \mu_{\pi' \vee \rho}(K_{\pi' \vee \rho}) \geq 1 - 2\beta$, et on voit que

$$|\Pi'(\bar{\pi'} \vee \tilde{\xi}_{\pi' \vee \rho}) - \bar{\rho}_0| \leq \frac{3}{2}|\Pi'(\pi' \vee \xi) - \rho_0| + \mu(K^c_{\pi' \vee \rho}) \leq \frac{3}{2}\alpha + 2\beta \leq 2\alpha, \quad (36)$$

si $\beta \leq \alpha/4$. En particulier, on a

$$\bar{\rho}_0 \preccurlyeq_{2\alpha} (\bar{\pi'} \vee \tilde{\xi}_{\pi' \vee \rho})_{[-N_1, N_1]},$$

et pour α assez petit, cela donne

$$0 \leq h(\bar{\pi'} \vee \bar{\rho}, T_{\pi' \vee \rho}) - h(\bar{\pi'} \vee \tilde{\xi}_{\pi' \vee \rho}, T_{\pi' \vee \rho}) \leq \delta,$$

et, comme $h(\bar{\pi'} \vee \bar{\rho}, T_{\pi' \vee \rho}) = h(\pi' \vee \rho, T) = h(\pi \vee \zeta, S) = h(\pi' \vee \xi, T)$, on a

$$0 \leq h(\pi' \vee \xi, T) - h(\bar{\pi'} \vee \tilde{\xi}_{\pi' \vee \rho}, T_{\pi' \vee \rho}) \leq \delta. \quad (37)$$

Donc (35) et (37) nous permettent d'appliquer le lemme fondamental et de construire $\xi'_{\pi' \vee \rho}$ sur $X_{\pi' \vee \rho}$ tel que $(T, \pi' \vee \xi) \sim (T_{\pi' \vee \rho}, \bar{\pi'} \vee \xi'_{\pi' \vee \rho})$ et $|(\xi'_{\pi' \vee \rho})_0 - (\tilde{\xi}_{\pi' \vee \rho})_0| \leq \varepsilon'$. Il nous reste à poser $\xi' = \xi'_{\pi' \vee \rho} \circ p$, et alors ξ' est $\pi' \vee \rho$-mesurable et

$$(T, \pi' \vee \xi') \sim (T, \pi' \vee \xi),$$

et

$$|\xi'_0 - \tilde{\xi}_0| \leq \varepsilon',$$

en posant $\tilde{\xi} = \tilde{\xi}_{\pi' \vee \rho} \circ p$.

Enfin, (34) implique

$$(T, F, n, \pi' \vee \rho \vee \xi) \cong (T, F, n, \pi' \vee \rho \vee \tilde{\xi}),$$

et alors, avec le même argument que (36), on voit que, si Π est une fonction sur $([r] \times [k])^{[-N,N]}$ telle que

$$|\Pi(\pi' \vee \xi) - \rho_0| \leq \varepsilon,$$

alors

$$|\Pi(\pi' \vee \tilde{\xi}) - \rho_0| \leq \frac{3}{2}\varepsilon + 2\beta, \text{ (car } N \leq N_1 \leq \frac{n\beta}{2}).$$

Finalement, comme $|\xi'_0 - \tilde{\xi}_0| \leq \varepsilon'$, pour ε' et β assez petits, cela implique

$$|\Pi(\pi' \vee \xi') - \rho_0| \leq 2\varepsilon.$$

□

Le lemme suivant nous dit qu'on va pouvoir utiliser le lemme 3.15 pour trouver une nouvelle "copie" de ζ proche de ρ, et qui sera presque génératrice :

Lemme 3.16. *Fixons k tel que ξ soit à valeurs dans $[k]$. Supposons qu'il existe un processus ρ sur X tel que $(T, \pi' \vee \rho) \sim (S, \pi \vee \zeta)$. Soit $\varepsilon > 0$: il existe un processus ζ^* sur X et $K \in \mathbb{N}$ tels que*

(i) $(T, \pi' \vee \zeta^*) \sim (S, \pi \vee \zeta)$,

(ii) $\xi_0 \preccurlyeq_\varepsilon (\pi' \vee \zeta^*)_{[-K,K]}$,

(iii) et $|\zeta^*_0 - \rho_0| \leq \varepsilon$.

Démonstration. On fixe $l, r \geq 1$ tel que ζ et ρ soient à valeurs dans $[l]$ et π, π' soient à valeurs dans $[r]$. Soit $\varepsilon > 0$. On introduit $\alpha > 0$, $\beta > 0$ et $n \geq 1$ à fixer plus tard. Il existe N_1 tel que

$$\rho_0 \preccurlyeq_\alpha (\pi' \vee \xi)_{[-N_1,N_1]}.$$

On prend alors une fonction Π_1 sur $([r] \times [k])^{[-N_1,N_1]}$ telle que

$$|\Pi_1(\pi' \vee \xi) - \rho_0| \leq \alpha.$$

Appliquons le lemme 3.15 : il existe ξ' $\pi' \vee \rho$-mesurable tel que

$$\text{(i) } (T, \pi' \vee \xi') \sim (T, \pi' \vee \xi) \text{ et (ii) } |\Pi_1(\pi' \vee \xi') - \rho_0| \leq 2\alpha. \tag{38}$$

On peut choisir $N_2 \geq N_1$ et une fonction Π_2 sur $([r] \times [l])^{[-N_2,N_2]}$ tels que

$$|\Pi_2(\pi' \vee \rho) - \xi'_0| \leq \alpha. \tag{39}$$

Pour des raisons qui apparaîtront plus tard, on se donne une nouvelle approximation : on choisit $N_3 \geq N_2$ et Π_3 sur $([r] \times [l])^{[-N_3,N_3]}$ tels que

$$|\Pi_3(\pi' \vee \rho) - \xi'_0| \leq \beta. \tag{40}$$

Par le théorème de Rohlin (3.6), on prend $F, E \subset X$ tels que

$$d(\xi_{[0,n[}|_F) = d(\xi_{[0,n[}) \text{ et } d(\xi'_{[0,n[}|_E) = d(\xi'_{[0,n[}),$$

ce qui donne

$$(T, F, n, \pi' \vee \xi) \cong (T, E, n, \pi' \vee \xi').$$

De plus, comme d'habitude, on a $\mu(G_{n_1}), \mu(G'_{n_1}) \geq 1 - 2\beta$ pour n assez grand, avec $G_{n_1} = \bigcup_{j=0}^{n-n_1} T^j F$ et $G'_{n_1} = \bigcup_{j=0}^{n-n_1} T^j E$.

Par le lemme 3.9, il existe ρ^* sur X tel que

$$(T, F, n, \pi' \vee \xi \vee \rho^*) \cong (T, E, n, \pi' \vee \xi' \vee \rho).$$

Notons $K = \bigcup_{j=N_3}^{n-N_3-1} T^j F$ et $K' = \bigcup_{j=N_3}^{n-N_3-1} T^j E$, et l'isomorphisme donne

$$d(\Pi_3(\pi' \vee \rho^*) \vee \xi_0|_K) = d(\Pi_3(\pi' \vee \rho) \vee \xi'_0|_{K'}).$$

On peut supposer que $\frac{2N_3}{n} \leq \beta$ de sorte que $\mu(K), \mu(K') \geq 1 - 2\beta$, et alors, comme pour 36, on a

$$|\Pi_3(\pi' \vee \rho^*) - \xi_0| \leq 2|\Pi_3(\pi' \vee \rho) - \xi'_0| + 2\beta \leq 4\beta. \tag{41}$$

De plus, par définition de ρ^*, on a

$$|d((\pi' \vee \rho)_{[0,n_1[}) - d((\pi' \vee \rho^*)_{[0,n_1[})| \leq \mu(G_{n_1}) + \mu(G'_{n_1}) \leq 4\beta. \tag{42}$$

On va maintenant rectifier ρ^* en utilisant le lemme fondamental : soit $\gamma > 0$ et $\delta > 0$ et $n_1 \geq 1$ associés donnés par le lemme fondamental appliqué à $(T, \pi' \vee \rho)$. Par (42) et (41), pour β assez petit, on obtient

$$|d((\pi' \vee \rho)_{[0,n_1[}) - d((\pi' \vee \rho^*)_{[0,n_1[})| \leq \delta,$$

et

$$0 \leq h(\pi' \vee \rho, T) - h(\pi' \vee \rho^*, T) \leq \delta,$$

en utilisant que $h(\pi' \vee \rho, T) = h(\pi' \vee \xi, T)$. Par le lemme fondamental, il existe ζ^* sur X tel que

$$\text{(i) } (T, \pi' \vee \zeta^*) \sim (T, \pi' \vee \rho) \text{ et (ii) } |\zeta^*_0 - \rho^*_0| \leq \gamma.$$

Par définition de ρ^* et par le (ii) de (38), on a

$$|\Pi_1(\pi' \vee \xi) - \rho^*_0| \leq 2|\Pi_1(\pi' \vee \xi') - \rho_0| + 2\beta \leq 4\alpha + 2\beta,$$

donc
$$|\Pi_1(\pi' \vee \xi) - \zeta_0^*| \leq 4\alpha + 2\beta + \gamma \leq 5\alpha,$$
si β et γ sont assez petits. Alors $|\zeta_0^* - \rho_0| \leq 6\alpha \leq \varepsilon$, si $\alpha \leq \varepsilon/6$.

On fixe maintenant α et donc N_2 est aussi fixé. On estime ensuite
$$|\Pi_2(\pi' \vee \rho^*) - \xi_0| \leq 2|\Pi_2(\pi' \vee \rho) - \xi_0'| + 2\beta \leq 2\alpha + 2\beta.$$
Donc, on peut prendre β et γ assez petits tels que
$$|\Pi_2(\pi' \vee \zeta^*) - \xi_0| \leq 3\alpha + 2\beta \leq 4\alpha \leq \varepsilon,$$
ce qui termine la preuve avec $K = N_2$. □

On est maintenant en mesure de terminer la preuve du théorème d'isomorphisme :

Démonstration du théorème 3.2. On rappelle que (T, ξ) et (S, ζ) sont deux processus relativement finiment déterminés par rapport à π' et π avec $(T, \pi') \sim (S, \pi)$, et tels que $\pi \vee \zeta$ et $\pi' \vee \xi$ sont générateurs, et on suppose que $h(\pi' \vee \xi, T) = h(\pi \vee \zeta, S)$. Par le lemme 3.13 et le lemme fondamental, on peut construire un premier processus $\eta^{(0)}$ sur X qui vérifie $(T, \pi' \vee \eta^{(0)}) \sim (S, \pi \vee \zeta)$.

Ensuite, grâce au lemme 3.16, on construit par récurrence deux suites $(\varepsilon_r)_{r\geq 1}$ et $(K_r)_{r\geq 1}$ ainsi qu'une suite de processus processus $(\eta^{(r)})_{r\geq 0}$ telles que, pour chaque r, on ait

(i) $(T, \pi' \vee \eta^{(r)}) \sim (S, \pi \vee \zeta)$,

(ii) $\xi_0 \preccurlyeq_{\varepsilon_r} (\pi' \vee \eta^{(r)})_{[-K_r, K_r]}$,

(iii) $|\eta_0^{(r)} - \eta_0^{(r-1)}| \leq \varepsilon_r$,

(iv) et $\forall j \leq r-1,\ \xi_0 \preccurlyeq_{\varepsilon_j(\sum_{0\leq l\leq r-j} 2^{-l})} (\pi' \vee \eta^{(r)})_{[-K_j, K_j]}$.

La condition (iv) est obtenue à chaque étape de la récurrence à partir de (iii) en s'assurant de prendre ε_r assez petit. De plus, il clair que l'on peut supposer que $\sum_{r\leq 1} \varepsilon_r < \infty$, et $(\eta_0^{(r)})_{r\geq 0}$ est alors une suite de Cauchy, et par la remarque 3.5, elle converge. Notons η_0 sa limite et η le processus associé. Il vérifie encore la condition (i) :
$$(T, \pi' \vee \eta) \sim (S, \pi \vee \zeta).$$

Il reste à vérifier que η est générateur : la condition (iv) donne, pour tout couple r, j tel que $j < r$,

$$\xi_0 \preccurlyeq_{2\varepsilon_j} (\pi' \vee \eta^{(r)})_{[-K_j, K_j]},$$

et on peut faire $r \longrightarrow \infty$, ce qui nous donne

$$\forall j \geq 1,\ \xi_0 \preccurlyeq_{2\varepsilon_j} \pi' \vee \eta.$$

Mais alors, comme $\varepsilon_j \longrightarrow 0$, ξ_0 est $\pi' \vee \eta$-mesurable, et vu que $\pi' \vee \xi$ est générateur, $\pi' \vee \eta$ l'est aussi. La preuve du théorème d'isomorphisme est donc complète. □

4 Annexe

4.1 Espaces de Lebesque

Lemme 4.1. *Notons $\mathcal{L}$ la mesure de Lebesgue sur* $[0, 1]$. *Soient $\alpha \in]0, 1[$, et F, F' deux boréliens tels que $\mathcal{L}(F) = \alpha = \mathcal{L}(F')$. Il existe $E \subset F$ et $E' \subset F'$ tels que $\mathcal{L}(E) = \alpha$ et $\mathcal{L}(E') = \alpha$, ainsi que $\varphi : E \longrightarrow E'$ une bijection qui préserve la mesure.*

Démonstration. Par transitivité, on peut supposer que $F' = [0, \alpha[$. Comme la mesure de Lebesgue n'a pas d'atomes, la fonction

$$x \mapsto \frac{\mathcal{L}(F \cap [0, x[)}{\mathcal{L}(F)}$$

est continue de $[0, 1]$ dans $[0, 1]$ et est surjective (car elle vaut 0 en 0 et 1 en 1).

On choisit x_0 tel que $\mathcal{L}(F \cap [0, x[) = \frac{\mathcal{L}(F)}{2}$, et on décompose F en

$$F_0 = F \cap [0, x_0[$$
$$F_1 = F \cap [x_0, 1[$$

Puis de même, on décompose F_0 et F_1 pour obtenir $F_{0,0}, F_{0,1}, F_{1,0}, F_{1,1}$, et on itère ce procédé pour construire, pour tout $n \in \mathbb{N}$, une partition de F par la famille

$$\{F_{(\epsilon_k)}; (\epsilon_k)_{k=0}^{n-1} \in \{0, 1\}^n\}.$$

Ensuite, pour chaque $x \in F$, la "suite associée à x" sera simplement la suite $(\epsilon_k)_{k\geq 0}$ telle que pour tout n, $x \in F_{(\epsilon_k)_{k=0}^{n-1}} \in \{0, 1\}$. On voit alors facilement que la fonction

$$\varphi : F \longrightarrow \{0, 1\}^{\mathbb{N}}$$
$$x \mapsto \textit{suite associée à } x$$

est injective sur un ensemble F_1 de mesure totale, et $\varphi_* \mathcal{L}_{F_1} = \mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right)$. De plus, dans l'image de φ il n'y a aucune suite constante égale à 1 a partir d'un certain rang, donc on peut composer par

$$: \ (\epsilon_k)_{k \leq 0} \mapsto \alpha \sum_{k \leq 0} \frac{\epsilon_k}{2^{k+1}}$$

et $\psi \circ \varphi : \ F_1 \longrightarrow [0, \alpha[$ est encore injective. Elle n'est pas nécessairement surjective, mais son image est de mesure pleine. Cela découle du fait que $\psi_* \mathcal{B}\left(\frac{1}{2}, \frac{1}{2}\right) = \mathcal{L}_{[0,\alpha[}$ et donc $(\psi \circ \varphi)_* \mathcal{L}_{F_1} = \mathcal{L}_{[0,\alpha[}$. □

4.2 Convergence presque uniforme et théorème d'Egorov

Le théorème de Birkhoff, central en théorie ergodique et souvent utilisé dans ce mémoire, donne convergence ponctuelle, mais ce n'est pas toujours ce dont on a besoin en pratique. Avec cette idée en tête, on donne la reformulation suivante :

Théorème 4.2. *Soit (X, μ) un espace de probabilité quelconque. Soit $(f_n)_{n \geq 0}$ un suite de fonctions, et supposons qu'elle converge presque partout vers f. Alors, pour tout $\varepsilon > 0$, il existe $E \subset X$ tel que $\mu(E) \geq 1 - \varepsilon$ et $(f_n)_{n \geq 0}$ converge uniformément sur E. On dit alors que $(f_n)_{n \geq 0}$ converge presque uniformément vers f.*

Remarque 4.1. La réciproque est aussi vraie, mais nous omettons la preuve car c'est un fait que nous n'utiliseront pas.

Démonstration. Soit $\varepsilon > 0$. Posons, pour $k, m \geq 1$,

$$E_k^m = \left\{ x \in X \mid \max_{l \geq k} \{ |f_l(x) - f(x)| \} \leq \frac{1}{m} \right\}.$$

Comme $(f_n)_{n \geq 0}$ converge presque partout, on $\mu\left(\bigcup_{k \geq 0} E_k^m\right) = 1$. De plus on remarque que pour tout m, la suite $(E_k^m)_{k \geq 1}$ est croissante, donc on a $\lim_{k \longrightarrow \infty} \mu(E_k^m) = 1$. Pour chaque m, on peut donc choisir $k(m)$ tel que $\mu\left(E_{k(m)}^m\right) \geq 1 - \frac{\varepsilon}{2^m}$. Il nous reste alors simplement à poser $E = \bigcap_{m \geq 1} E_{k(m)}^m$ car :

$$\mu\left(E^c\right) = \mu\left(\bigcup_{m \geq 1} \left(E_{k(m)}^m\right)^c\right) \leq \sum_{m \geq 1} \frac{\varepsilon}{2^m} = \varepsilon,$$

et on a construit E de sorte que $(f_n)_{n \geq 0}$ converge uniformément sur E. □

4.3 Théorème de Shannon-McMillan-Breiman

Cette partie est principalement tirée de [3].

Soit (T, ξ) un processus stationnaire à valeurs dans $\{1, ..., k\}$. Pour tout $n \in \mathbb{N}$ et pour $\mathbf{i} = (i_0, ..., i_{n-1}) \in \{1, ..., k\}^n$, notons

$$p_n(i_0, ..., i_{n-1}) = \mu(\xi_0 = i_0, ..., \xi_{n-1} = i_{n-1}) = \mu(\xi_{[0,n[} = \mathbf{i}).$$

On rappelle que $\xi_{[0;n[} = \bigvee_{l=0}^{n-1} \xi_l$. L'intérêt de cette fonction vient de la formule suivante, qui se vérifie facilement :

$$H\left(\bigvee_{l=0}^{n-1} \xi_l\right) = -\int \log(p_n(\xi_0(x), ..., \xi_{n-1}(x)))d\mu(x) = \mathbb{E}[\log(p_n(\xi_0, ..., \xi_{n-1}))].$$

En particulier, on peut réécrire la définition de la KS-entropie :

$$h(\xi, T) = \lim_{n \longrightarrow \infty} \mathbb{E}\left[-\frac{1}{n}\log(p_n(\xi_0, ..., \xi_{n-1}))\right]. \tag{43}$$

Le théorème de Shannon-McMillan-Breiman montre que l'intégrande ci-dessus converge en fait presque partout vers $h(\xi, T)$:

Théorème 4.3. *Pour μ-presque tout x, on a*

$$\lim_{n \longrightarrow \infty} -\frac{1}{n}\log(p_n(\xi_0(x), ..., \xi_{n-1}(x))) = h(\xi, T).$$

En combinant le théorème ci-dessus au théorème d'Egorov (4.2), on obtient

Corollaire 4.3.1. *Pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que, pour tout $n \geq N$, il existe $\mathcal{E}_n \subset \{1, ..., k\}^n$ tel que $\mu\left(\bigcup_{i \in \mathcal{E}_n} \{\xi_{[0,n[} = \boldsymbol{i}\}\right) \geq 1 - \varepsilon$, et*

(i) $\forall \boldsymbol{i} \in \mathcal{E}_n$, $2^{-(h(\xi,T)+\varepsilon)n} \leq \mu(\xi_{[0,n[} = \boldsymbol{i}) \leq 2^{-(h(\xi,T)-\varepsilon)n}$

(ii) et $(1-\varepsilon)2^{(h(\xi,T)-\varepsilon)n} \leq \#\mathcal{E}_n \leq 2^{(h(\xi,T)+\varepsilon)n}$

Remarque 4.2. Dans l'énoncé du corollaire, il est facile de voir que (ii) découle de (i).

Démonstration du théorème 4.3. On introduits les nouvelles fonctions suivantes :

$$g_0(x) = -\log(p_0(\xi_0(x)))$$
$$g_n(x) = -\log\left(\frac{p_{n+1}(\xi_{-n}(x), ..., \xi_0(x))}{p_n(\xi_{-n}(x), ..., \xi_{-1}(x))}\right)$$

On voit alors qu'on peut écrire

$$-\frac{1}{n}\log(p_n(\xi_0(x), ..., \xi_{n-1}(x))) = \frac{1}{n}\sum_{l=0}^{n-1} g_l(T^l(x)).$$

On va vouloir ensuite applique le théorème ergodique de Birkhoff. Pour cela, intéressons nous au comportement asymptotique de la suite $(g_n)_{n\geq 0}$. On a

$$\begin{aligned} g_n(x) &= \sum_{i=1}^{k} -\log\left(\frac{p_{n+1}(\xi_{-n}(x), ..., i)}{p_n(\xi_{-n}(x), ..., \xi_{-1}(x))}\right)\mathbb{1}_{\xi_0=i}(x) \\ &= \sum_{i=1}^{k} -\log(\mu(\xi_0 = i \,|\, \xi_{-n}, ..., \xi_{-1})(x))\mathbb{1}_{\xi_0=i}(x) \end{aligned}$$

Or, par le théorème de convergence des martingales, $\mu(\xi_0 = i \,|\, \xi_{-n}, ..., \xi_{-1})(x)$ converge presque sûrement, et donc la suite $(g_n(x))_{n\geq 0}$ converge presque sûrement. Notons g la limite, et

$$g_n^{(i)}(x) = -\log(\mu(\xi_0 = i \,|\, \xi_{-n}, ..., \xi_{-1})(x)).$$

Soit $\lambda > 0$. Notons $E_n = \{x \,|\, \max_{0\leq l\leq n-1} g_l(x) \leq \lambda < g_n(x)\}$, de sorte que

$$\mu(\{x \,|\, \sup_n g_n(x) > \lambda\}) = \sum_{n\geq 0} \mu(E_n).$$

Ensuite

$$\begin{aligned} \mu(E_n) &= \sum_{i=1}^{k} \mu(E_n \cap \{\xi_0 = i\}) \\ &= \sum_{i=1}^{k} \mu(F_n^{(i)} \cap \{\xi_0 = i\}) \end{aligned}$$

où $F_n^{(i)} = \{x \,|\, \max_{0\leq l\leq n-1} g_l^{(i)}(x) \leq \lambda < g_n^{(i)}(x)\}$.

Mais alors, $F_n^{(i)} \in \sigma(\{\xi_l \,;\, -n \leq l \leq -1\})$, donc

$$\begin{aligned}\mu(F_n^{(i)} \cap \{\xi_0 = i\}) &= \int_{F_n^{(i)}} \mu(\xi_0 = i \,|\, \xi_{-n}, ..., \xi_{-1})(x) d\mu(x) \\ &= \int_{F_n^{(i)}} 2^{-g_n^{(i)}(x)} d\mu(x) \\ &\leq 2^{-\lambda} \mu(F_n^{(i)}).\end{aligned}$$

Et alors

$$\sum_{n \geq 0} \mu(E_n) \leq 2^{-\lambda} \sum_{i=1}^{k} \sum_{n \geq 0} \mu(F_n^{(i)}) \leq k 2^{-\lambda}.$$

Et donc

$$\begin{aligned}\int \sup_{n \geq 0} g_n(x) d\mu(x) &= \sum_{m \in \mathbb{N}} \int_{\{m \leq \sup_n g_n < m+1\}} \sup_n g_n(x) d\mu(x) \\ &\leq \sum_{m \in \mathbb{N}} (m+1) \mu(\{x \mid \sup_n g_n(x) > m\}) \\ &\leq k \sum_{m \in \mathbb{N}} (m+1) 2^{-m} < \infty\end{aligned}$$

On applique ensuite le théorème de convergence dominée ainsi que (43)

$$\begin{aligned}\mathbb{E}[g] &= \lim_{n \longrightarrow \infty} \mathbb{E}[g_n] = \lim_{n \longrightarrow \infty} \mathbb{E}\left[\frac{1}{n} \sum_{l=0}^{n-1} g_l \circ T^l\right] \\ &= \lim_{n \longrightarrow \infty} \mathbb{E}\left[-\frac{1}{n} \log(p_n(\xi_0, ..., \xi_{n-1}))\right] = h(\xi, T).\end{aligned}$$

Posons $G_N(x) = \sup_{l \geq N} |g_l(x) - g(x)|$. On va maintenant vouloir "remplacer" g_l par g. Pour cela on vérifie que, pour tout $N \geq 0$:

$$\begin{aligned}\limsup_{n \longrightarrow \infty} \frac{1}{n} \sum_{l=0}^{n-1} |g_l(T^l(x)) - g(T^l(x))| &= \limsup_{n \longrightarrow \infty} \frac{1}{n} \sum_{l=N}^{n-1} |g_l(T^l(x)) - g(T^l(x))| \\ &\leq \limsup_{n \longrightarrow \infty} \frac{1}{n} \sum_{l=N}^{n-1} G_N(T^l(x)) \\ &\leq \limsup_{n \longrightarrow \infty} \frac{1}{n} \sum_{l=0}^{n-1} G_N(T^l(x)) \\ &= \mathbb{E}[G_N] \ \mu\text{-presque partout,}\end{aligned}$$

la dernière égalité provenant du théorème ergodique de Birkhoff. Mais on sait que G_N tend vers 0 presque partout, et la majoration obtenue plus haut permet d'appliquer le théorème de convergence dominée : $\mathbb{E}[G_N] \longrightarrow 0$. Enfin, on a

$$\begin{aligned}
\lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{l=0}^{n-1} g_l(T^l(x)) &= \lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{l=0}^{n-1} g(T^l(x)) + \lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{l=0}^{n-1} g_l(T^l(x)) - g(T^l(x)) \\
&= \mathbb{E}[g] \ \mu\text{-presque partout} \\
&= h(\xi, T),
\end{aligned}$$

en utilisant de nouveau le théorème de Birkhoff. □

4.4 Lemme des mariages

Lemme 4.4. *Soit X un ensemble fini et S une famille finie de parties de X comptées avec multiplicité. Il existe une injection $f : S \longrightarrow X$ telle que $f(s) \in s$ si, et seulement si,*

$$\forall W \subset S,\ \#W \leq \# \bigcup_{A\in W} A.$$

Remarques.

- Si f existe, $f(S)$ est un *système de représentants disjoints* de S.
- La condition énoncée dans le théorème est appelée "condition de mariage".

On trouve une présentation et une preuve de ce résultat sur le site [1].

Références

[1] Hall's mariage theorem. https://brilliant.org/wiki/hall-marriage-theorem/. Visité le 06/06/2020.

[2] Tim Austin. Measure concentration and the weak Pinsker property. *Publ. Math., Inst. Hautes Étud. Sci.*, 128 :1–119, 2018.

[3] P. Billingsley. *Ergodic theory and information.* John Wiley & Sons, Hoboken, NJ, 1965.

[4] Tomasz Downarowicz. *Entropy in dynamical systems.*, volume 18. Cambridge : Cambridge University Press, 2011.

[5] Donald S. Ornstein. Ergodic theory, randomness, and dynamical systems. Yale Mathematical Monographs. 5. New Haven - London : Yale University Press. VII, 141 p. £2.50 (1974)., 1974.

[6] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27 :379–423, 623–656, 1948.

Chapitre 2 : Propriété de Pinsker faible

Séverin Benzoni

Encadrant : Thierry De La Rue

Du 25 Juin au 16 Septembre 2020

Sommaire

1 Histoire du problème

En 1958, Kolmogorov et Sinaï ont introduit la notion d'entropie en théorie ergodique, que l'on a présentée dans le Chapitre 1 ([Chapitre 1, Remarque 2.2] pour retrouver la définition sous sa forme originale). La même année, Kolmogorov introduit une deuxième notion importante : les K-systèmes. Il définit un K-système comme système dynamique $(X, \mathcal{F}, \mu, T)$ qui possède un processus générateur ξ dont la tribu de queue $\bigcap_{n\geq 1} \sigma(\xi_{]-\infty,-n]})$ est triviale. Il existe une définition équivalente plus intrinsèque au système : (X, μ, T) est un K-système si, et seulement si, tout observable non-trivial ξ_0 vérifie $h(\xi, \mu, T) > 0$ (on peut trouver une preuve de cette équivalence, ainsi qu'une présentation plus complète de cette notion dans [2]). Il est aussi équivalent de supposer que le facteur de Pinsker du système est trivial, le facteur de Pinsker étant la tribu

$$\Pi_T = \{A \in \mathcal{F} \mid h(\mathbb{1}_A, \mu, T) = 0\}.$$

Le facteur de Pinsker est en fait simplement le plus grand facteur de (X, μ, T) d'entropie nulle. Un K-système ne possède donc aucun facteur non trivial d'entropie nulle : il est entièrement non-déterministe. Par exemple, les K-systèmes les plus élémentaires sont les schémas de Bernoulli, car ils vérifient la loi du $0-1$ de Kolmogorov.

L'entropie est un invariant permettant de quantifier le "chaos" d'un système dynamique, et de nombreuses questions apparues après son invention avaient pour but de comprendre la structure de ce "chaos". La première question, que Kolmogorov pose après avoir prouvé que les schémas de Bernoulli sont des K-systèmes, était de savoir si tous les K-systèmes sont des schémas de Bernoulli, ce qui impliquerait que ces systèmes chaotiques sont de structure très simple. Des questions plus générales ont ensuite émergé, et nous reviendront dessus dans les paragraphes qui suivent, car elles sont l'objet de l'article étudié dans ce mémoire.

La découverte de l'entropie a d'abord permis d'obtenir des résultats de non-isomorphisme, et notamment pour les schémas de Bernoulli : deux schémas schémas de Bernoulli isomorphes doivent avoir la même entropie. La réciproque de ce résultat, montrée par Ornstein ([5], [6]), constitue l'un des succès les plus notables de la KS-entropie. Mais les ramifications de la théorie d'Ornstein vont bien au delà des schémas de Bernoulli, et ont profondément impacté l'évolution de la théorie ergodique. Nous nous contenterons ici de raconter l'histoire de la conjecture de Pinsker.

Au début des années 1960, Pinsker, qui travaille alors à Moscou avec Kolmogorov, montre que tout K-facteur de (X, μ, T) est indépendant du facteur de

Pinsker Π_T (je ne donne pas de référence ici car elles sont toutes en Russe). Suite à ce résultat, bien que l'existence d'un K-facteur ne soit pas encore prouvée, il émet une conjecture (appelée plus tard "conjecture de Pinsker") : tout système d'entropie non-nulle est isomorphe au produit direct de son facteur de Pinsker et d'un K-système. Quelques années plus tard, Sinaï publie [11] qui semble conforter cette conjecture : il prouve l'existence d'un facteur de (X, μ, T) isomorphe à un schéma de Bernoulli de même entropie que (X, μ, T). Étant donné le résultat d'indépendance de Pinsker, il aurait suffit de prouver que le facteur construit par Sinaï et le facteur de Pinsker engendrent la tribu entière pour obtenir un résultat encore plus fort que la conjecture de Pinsker : (X, μ, T) serait alors isomorphe au produit direct de son facteur de Pinsker et d'un schéma de Bernoulli. Cette "conjecture de Pinsker forte" aurait aussi prouvé que tout K-système est isomorphe à un schéma de Bernoulli.

Mais cela s'est avéré faux : Ornstein publie un premier exemple de K-système non-Bernoulli ([9]) qui contredit donc la conjecture de Pinsker forte. Par la suite, de nombreux autres contre-exemples ont été construits, et il s'avère que la classe des K-système est très vaste, laissant peu d'espoir pour une classification complète de ces systèmes. Parmi tous ces contre-exemples, on peut trouver une construction d'Ornstein ([7]) qui permet de contredire la conjecture de Pinsker. De plus, il affine ensuite ce résultat en construisant un système *mélangeant* qui ne vérifie pas la conjecture de Pinsker ([8]). Ainsi, toutes les hypothèses formulés dans les premières années de l'étude de l'entropie se sont avérées fausses, révélant une grande variété de phénomènes possibles.

Une des ramifications des travaux d'Ornstein peut être trouvée dans les travaux de Thouvenot, qui, à partir de 1975, s'intéresse aux systèmes relativement Bernoulli et développe une version "relative" de la théorie d'Ornstein. Suite à ces travaux, dans son article [12] de 1977, il introduit la propriété de Pinsker faible : pour tout $\varepsilon > 0$, (X, μ, T) est isomorphe au produit direct d'un schéma de Bernoulli et d'un système d'entropie inférieure à ε. Cependant, il se montre prudent et ne formule pas de conjecture concernant cette propriété. Ensuite, un certain nombre de résultats sont prouvés sur les systèmes vérifiant cette propriété. Austin présente une partie de ces résultats dans la [1, Section 16] de son article, et on en expose un en détails à la fin de ce mémoire (Section 3).

Cependant, durant quatre décennies, on ne sait pas si tous les systèmes vérifient la propriété de Pinsker faible. Thouvenot pensait que ce n'était pas le cas, et aucun résultat ne semblait lui donner tort. Mais, en 2018, Tim Austin publie un long article sur le sujet ([1]) dans lequel il prouve que tous les systèmes ergodiques vérifient la propriété de Pinsker faible. La lecture de cet article a été le point de départ de ce mémoire, et je tâche de donner une présentation partielle de la preuve dans la section suivante.

2 Une présentation de la preuve de Tim Austin

La preuve donnée par Tim Austin regroupe beaucoup de travail de sa part, et je ne serai donc pas en mesure de tout présenter dans ce mémoire. De plus, son article est très bien rédigé donc je me permettrai de me reposer sur les preuves qu'il donne d'un certain nombre de résultats, afin de plutôt détailler la partie finale de la preuve. Ainsi, pour un certain nombre de résultats qui seront utiles dans les parties que je souhaite détailler, je me contenterai de recopier les énoncés pour faciliter la lecture, mais sans fournir de commentaire nouveau. Je tâcherai tout de même de signaler les résultats dont la preuve n'est pas élémentaire.

2.1 Énoncé du théorème

Tim Austin montre le théorème suivant, un peu plus fort que la propriété de Pinsker faible :

Théorème 2.1. *[Théorème d'Austin] Soit* $\pi : (X, \mu, T) \longrightarrow (Y, \nu, S)$ *une application facteur entre systèmes ergodiques. Pour tout* $\varepsilon > 0$*, il existe une deuxième application facteur* φ *telle que* (X, μ, T) *est relativement Bernoulli par rapport à* $\pi \vee \varphi$*, et* $h(\varphi, \mu, T) \leq \varepsilon$*.*

La définition de "$h(\varphi, \mu, T)$" est celle donnée dans le Chapiter 1 ([Chapitre 1, Remarque 2.2]).

Austin prouve même une version de ce théorème pour des actions de groupes autres que $\mathbb{Z}$, mais je ne me suis pas intéressé à ces notions, donc on n'en parlera pas dans ce mémoire.

On remarque que ce théorème implique bien la conjecture de Pinsker faible : il suffit de se donner un facteur π d'entropie inférieure à $\frac{\varepsilon}{2}$ et d'appliquer le théorème avec $\frac{\varepsilon}{2}$ comme paramètre d'erreur. Il serait encore plus simple de prendre le facteur trivial pour π, mais il ne me semble pas que la preuve soit valable dans ce cas : on utilise le fait que (Y, ν, S) est un système ergodique et que (Y, ν) est un espace de Lebesgue sans atomes, ce qui ne peut être vérifié pour le système trivial. En revanche, pour n'importe quel processus sur (X, μ, T) à valeurs dans un alphabet avec au moins deux éléments, les conditions ci-dessus sont vérifiées, et on peut appliquer le théorème d'Austin.

2.2 Résultat de concentration de la mesure

Le résultat novateur qui a permis à T. Austin de prouver le Théorème 2.1 concerne les phénomènes de concentration de la mesure. J'en donne ici une présentation rapide, sans preuves, car ce n'est pas la partie de l'article sur laquelle je me suis concentré.

On se donne un alphabet fini A et $n \geq 1$. Les phénomènes de concentrations de la mesure s'observent sur les mesures de probabilité sur A^n, lorsque n devient grand. Pour les quantifier, introduisons la distance de Hamming

$$\forall \mathbf{a}, \mathbf{a}' \in A^n,\ d_n(\mathbf{a}, \mathbf{a}') = \frac{1}{n} \#\{i \in [n] | a_i \neq a'_i\}, \tag{1}$$

et la distance de transport associée : $\forall \mu, \nu \in Prob(A^n)$,

$$\bar{d_n}(\mu, \nu) = \inf \left\{ \int d_n(\mathbf{a}, \mathbf{a}') d\lambda(\mathbf{a}, \mathbf{a}'); \lambda \text{ est un couplage de } \mu \text{ et } \nu \right\}. \tag{2}$$

On dira alors qu'une mesure $\mu \in Prob(A^n)$ satisfait $T(\kappa, r)$ (ou la T-inégalité de paramètres κ et r) si pour tout $\nu \in Prob(A^n)$

$$\bar{d_n}(\mu, \nu) \leq \frac{1}{\kappa} D(\nu \,||\, \mu) + r.$$

Le résultat utilisé par la suite est [1, Theorem C], que l'on énonce ici

Théorème 2.2. *Pour tous $\varepsilon, r > 0$, il existe $c > 0$ et $\kappa > 0$ tels que, pour tout alphabet A et pour n assez grand, on ait : soit $\mu \in Prob(A^n)$, notons, pour $i \in [n]$, $\mu_{\{i\}} \in Prob(A)$ les lois marginales de μ et posons*

$$E(\mu) = \sum_{i=1}^{n} H(\mu_{\{i\}}) - H(\mu).$$

Alors il existe une partition

$$A^n = U_1 \cup \cdots \cup U_m,$$

telle que

(i) $m \leq c e^{cE}$,

(ii) $\mu(U_1) \leq \varepsilon$,

(iii) et pour tout $j \in \{2, ..., m\}$, $\mu(U_j) > 0$ et $\mu_{|U_j}$ satisfait $T(\kappa n, r)$.

Il est crucial que c et κ soient indépendants de l'alphabet A, comme nous le verrons au moment de l'application de ce théorème. [1, Part I] est dédiée à la preuve de ce théorème. Cette preuve ne fait appel à aucune notion de théorie ergodique, il s'agit d'un résultat de probabilités discrètes.

Le contrôle de la partition construite dans le théorème ci-dessus est principalement donné par la grandeur E. Il s'agit d'une grandeur venue de la théorie de l'information, où on l'appelle corrélation totale et on la note $TC(\mu)$ (et cette notation se retrouve dans la preuve du Théorème 2.2, mais nous nous contenteront de la noter $E(\mu)$). Étant donné que l'on admet le Théorème 2.2, on n'a pas besoin de connaître en détail ses propriétés, car il nous suffira de la majorer de manière élémentaire lorsque que nous appliqueront ce théorème (Étape 3 de la preuve de la Proposition 2.14).

2.3 Extrémalité

Le but de cette partie est de donner les outils pour faire le lien entre le résultat de probabilités discrètes ci-dessus et le résultat de théorie ergodique que l'on veut prouver.

2.3.1 Distance de transport

On introduit une généralisation de la définition (2) : soit (K, d) un espace métrique compact et $Prob(K)$ l'ensemble des mesures de probabilité boréliennes sur K, que l'on munit de la distance de transport

$$\bar{d}(\mu, \nu) = \inf \left\{ \int d(x, y) d\lambda(x, y) \,;\, \lambda \text{ est couplage de } \mu \text{ et } \nu \right\}. \tag{3}$$

Une remarque importante est de voir que dans le cas particulier de l'équation (2), la distance obtenue est la même que celle utilisée dans la théorie d'Ornstein, définie dans une version relative dans la preuve du [Chapitre 1, Lemme 3.4]. L'utilisation des distances de transport permet simplement de donner une définition plus moderne.

Le résultat classique à connaitre sur cette distance est le théorème de dualité de Monge-Kantorovic-Rubinstein, dont on peut trouver une preuve dans [3, Theorem 11.8.2] :

Théorème 2.3. *[Dualité de Monge-Kantorovic-Rubinstein] Pour tous* $\mu, \nu \in Prob(K)$

$$\bar{d}(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu \, ; \, f \in Lip_1(K) \right\}.$$

Nous donnons aussi l'énoncé d'un résultat simple, mais qui sera très utile ([1, Lemma 4.3] dans l'article de T. Austin) :

Lemme 2.4. *Soient* (K, d_K) *et* (L, d_L) *deux espaces métriques compacts,* $0 < \alpha < 1$*, et définissons la distance suivante sur* $K \times L$ *:*

$$d((x, y), (x', y')) := \alpha d_K(x, x') + (1 - \alpha) d_L(y, y'). \tag{4}$$

Soient $\mu \in \mathrm{Prob}(K)$*,* $\nu \in \mathrm{Prob}(L)$ *et* $\lambda \in \mathrm{Prob}(K \times L)$*. Enfin, notons* λ_K *la marginale de* λ *sur* K*, et prenons* $x \mapsto \lambda_{L,x}$ *une loi conditionnelle de* λ *sachant la coordonnée sur* K*.*

Alors

$$\overline{d}(\lambda, \mu \times \nu) \leq \alpha \overline{d_K}(\lambda_K, \mu) + (1 - \alpha) \int_K \overline{d_L}(\lambda_{L,x}, \nu) \, \lambda_K(\mathrm{d}x).$$

2.3.2 L'extrémalité en théorie de la mesure

Dans cette partie, on donne un résumé de [1, Section 9]. On y définit l'extrémalité comme une notion portant sur des objets de théorie de la mesure (sans aspects dynamiques). On introduit la notion d'extrémalité, qui permet de reformuler le Théorème 2.2, pour l'appliquer dans la preuve finale.

Soient $(\Omega, \mathbb{P})$ un espace de probabilité et X un espace métrique complet et séparable. On note $Prob(X)$ l'ensemble des mesures de probabilité boréliennes sur X. Un noyau de Ω vers X est une application mesurable pour la tribu engendrée par la topologie faible sur $Prob(X)$:

$$\mu_\bullet : \omega \in \Omega \mapsto \mu_\omega \in Prob(X).$$

On peut alors définir le *croisement* d'un noyau $\mu_\bullet$ et de la mesure de probabilité $\mathbb{P}$, noté $\mathbb{P} \ltimes \mu_\bullet$, comme une mesure de probabilité sur $\Omega \times X$ via la formule

$$\mathbb{P} \ltimes \mu_\bullet(A \times B) = \int_A \mu_\omega(B) d\mathbb{P}(\omega).$$

J'ai inventé le terme "croisement" pour traduire le terme anglais "hook up", mais je ne sais pas s'il existe un autre terme français pour désigner cet objet.

Si on a
$$\mu = \int \mu_\omega d\mathbb{P}(\omega),$$
on dit que $(\mathbb{P}, \mu_\bullet)$ est une *décomposition* de μ. On remarque que le croisement $\mathbb{P} \ltimes \mu_\bullet$ est alors un couplage de $\mathbb{P}$ et de μ.

Par exemple, si on se donne un espace de probabilité (X, μ), toute partition $(P_i)_{1\leq i\leq k}$ de X engendre une décomposition de μ : il suffit de poser $\Omega = \{1, ..., k\}$ et $\mathbb{P} = (\mu(P_i))_{1\leq i\leq k}$ et $\mu_i = \mu_{|P_i}$.

Définition 2.1. Soit (K, d) un espace métrique compact et μ une mesure de probabilité borélienne. On dit que μ est (κ, r)-extrémale si, pour toute décomposition $(\mathbb{P}, \mu_\bullet)$ de μ, on a
$$\int \bar{d}(\mu, \mu_\omega) d\mathbb{P}(\omega) \leq \frac{1}{\kappa} \int D(\mu_\omega \,||\, \mu) d\mathbb{P}(\omega) + r.$$

On rappelle que la KL-divergence de ν par rapport à μ, $D(\nu \,||\, \mu)$, vaut ∞ si ν n'est pas absolument coninue par rapport à μ, et sinon, elle est donnée par la formule
$$D(\nu \,||\, \mu) = \int \frac{d\nu}{d\mu} \log\left(\frac{d\nu}{d\mu}\right) d\mu.$$

Les propriétés précises de l'opérateur de KL-divergence D jouent un rôle important dans la preuve du Théorème 2.2, mais comme nous ne donnons pas ici la preuve de ce théorème, on pourra ce contenter d'en avoir une compréhension élémentaire. La fait important dans la Définition 2.1 est que le terme de gauche est petit pour toutes les décomposition pas trop complexes. Cette idée de décomposition "pas trop complexe" apparaît plus clairement dans la section suivante, lorsqu'on introduit l'extrémalité de Thouvenot.

Pour ce qui concerne la KL-divergence, les preuves que l'on détaille dans ce mémoire reposent uniquement sur la "chain rule" présentée dans le Chapitre 1 ([Chapitre 1, Lemme 2.6]), et sur un résultat très simple rappelé par T. Austin ([1, Lemma 2.2]) qui affirme que

Lemme 2.5. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et X un ensemble fini. Soit ξ : $\Omega \longrightarrow X$ une v.a. de loi μ. On se donne deux sous tribus, $\mathcal{G}$ et $\mathcal{H}$, de $\mathcal{F}$ telles que $\mathcal{G} \subset \mathcal{H}$, et $\mu_\bullet$ et $\nu_\bullet$ des lois conditionnelles de ξ sachant $\mathcal{G}$ et $\mathcal{H}$ respectivement. Alors*
$$\int D(\nu_\omega \,||\, \mu_\omega) d\mathbb{P}(\omega) = H(\xi \,|\, \mathcal{G}) - H(\xi \,|\, \mathcal{H}).$$

En particulier, si ξ et ζ sont deux variables aléatoires telles que μ est la loi de ξ et $\mu_\bullet$ est la loi conditionnelle de ξ sachant ζ, l'extrémalité de μ se réécrit

$$\int \bar{d}(\mu_\omega, \mu) d\mathbb{P}(\omega) \leq \frac{1}{\kappa}(H(\xi) - H(\xi \,|\, \zeta)) + r. \tag{5}$$

Pour être en mesure de définir la notion d'extrémalité *relative*, il faut étendre la Définition 2.1 aux noyaux de mesures.

Définition 2.2. Soient (Y, ν) un espace de probabilité, (K, d) un espace métrique compact et $\mu_\bullet$ un noyau de Y vers K. On dit que la paire $(\nu, \mu_\bullet)$ est (κ, r)-extrémale s'il existe une fonction mesurable

$$y \in Y \mapsto r_y \in]0, \infty[,$$

telle que, pour tout y, μ_y est (κ, r_y)-extrémale et

$$\int r_y d\nu(y) \leq r.$$

Une formulation qui sera plus agréable par la suite, est de réécrire la définition ci-dessus dans le cas où $\mu_\bullet$ est une loi conditionnelle d'une v.a. à valeurs dans K :

Définition 2.3. Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, (K, d) un espace métrique compact et $F : \Omega \longrightarrow K$ une fonction mesurable. On se donne une sous-tribu $\mathcal{G}$ et $\mu_\bullet$ une loi conditionnelle de F par rapport à $\mathcal{G}$. F est relativement (κ, r)-extrémale par rapport à $\mathcal{G}$ si $(\mathbb{P}, \mu_\bullet)$ est (κ, r)-extrémale.

Le résultat clé sera le lemme suivant ([1, Lemma 9.10] dans l'article d'Austin). Il permet aussi de donner une meilleure compréhension de ces définitions, et c'est ici que le Lemme 2.5 entre en jeu.

Lemme 2.6. *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $\mathcal{G} \subset \mathcal{H}$ deux sous-tribus de $\mathcal{F}$, (K, d) un espace métrique fini, et $F : \Omega \longrightarrow K$ mesurable et relativement (κ, r)-extrémale par rapport à $\mathcal{G}$. On se donne $\mu_\bullet$ et $\nu_\bullet$ des lois conditionnelles de F par rapport à $\mathcal{G}$ et $\mathcal{H}$ respectivement. On suppose que $\mathcal{G}$ et $\mathcal{H}$ sont chacune engendrée par une famille dénombrable de parties de Ω. Alors on a*

$$\int \bar{d}(\nu_\omega, \mu_\omega) d\mathbb{P}(\omega) \leq \frac{1}{\kappa}(H(F \,|\, \mathcal{G}) - H(F \,|\, \mathcal{H})) + r.$$

Ce lemme possède aussi un corollaire intéressant :

Corollaire 2.6.1. *Dans le contexte du lemme précédent, on pose*

$$a := H(F \,|\, \mathcal{G}) - H(F \,|\, \mathcal{H}).$$

Alors F est $(\kappa, 2a/\kappa + 3r)$-extrémale par rapport à $\mathcal{H}$.

Il nous faut encore mentionner un résultat additionnel concernant l'extrémalité :

Lemme 2.7. *Soient A un ensemble fini et d_n la distance de Hamming sur A^n, (Y, ν) un espace de probabilité et $\mu_\bullet$ un noyau de probabilité de Y vers A^n. Soient $\delta > 0$, et $S \subset [n]$ tel que $\#S \geq (1-\delta)n$ et, pour tout $y \in Y$, $\mu_{S,y}$ la projection de μ_y sur A^S. Si $(\nu, \mu_{S,\bullet})$ est (κ, r)-extrémale, alors $(\nu, \mu_\bullet)$ est $(n\kappa/\#S, r + a)$-extrémale.*

2.3.3 L'introduction de l'extrémalité en théorie ergodique

On résume ici [1, Section 12], dont le but est de définir l'extrémalité relative en théorie ergodique. On rappelle que, sauf contre-indication, (X, μ, T) désigne un système dynamique ergodique.

Définition 2.4. Soient $r, \kappa > 0$. Soient ξ et π deux processus sur (X, μ, T). On dit que ξ est *relativement* (κ, r)*-extrémal par rapport* à π si, pour tout n assez grand, $\xi_{[0,n[}$ est relativement $(\kappa n, r)$-extrémale par rapport à $\pi_{[0,n[}$ (sur l'espace de probabilité (X, μ)).

Austin n'est pas le premier à introduire la notion extrémalité d'un processus. Tâchons de voir ce que les définitions antérieures peuvent nous apprendre sur la définition ci-dessus.

L'extrémalité en théorie ergodique est due, à l'origine, à Thouvenot ([4, Definition 6.3, p.220]) et elle donne une caractérisation des schémas de Bernoulli. Cependant, la formulation donnée par Austin est légèrement différente (en plus du fait qu'il donne version relative). Dans le reste de cette section, nous allons tâcher de comparer ces deux formulations.

On dira qu'un processus (T, ξ) est extrémal au sens d'Austin si pour tout $r > 0$, il existe $\kappa > 0$ et $n_1 \geq 1$ tels que pour tout $n \geq n_1$, $\xi_{[0,n[}$ est $(\kappa n, r)$-extrémale (c'est-à-dire que la loi de $\xi_{[0,n[}$ est $(\kappa n, r)$-extrémale).

On dira qu'un processus (T, ξ) est extrémal au sens de Thouvenot si pour tout $\varepsilon > 0$, il existe $\delta > 0$ et $n_1 \geq 1$ tels que pour tout $n \geq n_1$ et toute v.a. F à valeurs dans $\{1, ..., \lfloor 2^{n\delta} \rfloor\}$, il existe $\mathcal{E} \subset \{1, ..., \lfloor 2^{n\delta} \rfloor\}$ tel que $\mu(F \in \mathcal{E}) \geq 1 - \varepsilon$ et pour tout $k \in \mathcal{E}$

$$\bar{d}(\lambda, \lambda_k) \leq \varepsilon,$$

où λ est la loi de $\xi_{[0,n[}$ et λ_k est la loi de $\xi_{[0,n[}$ sachant que F vaut k.

Proposition 2.8. *Un processus extrémal au sens d'Austin est extrémal au sens de Thouvenot.*

Démonstration. Soient (T, ξ) un processus extrémal au sens d'Austin et $\varepsilon > 0$. Il existe $\kappa > 0$ et $n_1 \geq 1$ tels que pour tout $n \geq n_1$, $\xi_{[0,n[}$ est $(\kappa n, \frac{\varepsilon^2}{2})$-extrémale. On se donne ensuite $\delta > 0$ à fixer plus tard et une v.a. F à valeurs dans $\{1, ..., \lfloor 2^{n\delta} \rfloor\}$. La donnée d'une loi conditionnelle de $\xi_{[0,n[}$ sachant F, $(\lambda_k)_{1 \leq k \leq 2^{n\delta}}$, nous donne une décomposition de la loi de $\xi_{[0,n[}$, λ, et on lui applique alors l'hypothèse d'extrémalité ainsi que (5) :

$$\begin{aligned}
\sum_{1 \leq k \leq 2^{n\delta}} \mu(F = k)\bar{d}(\lambda, \lambda_k) &\leq \frac{1}{\kappa n} \sum_{1 \leq k \leq 2^{n\delta}} \mu(F = k) D(\lambda_k \,||\, \lambda) + \frac{\varepsilon^2}{2} \\
&= \frac{1}{\kappa n}(H(\xi) - H(\xi \,|\, F)) + \frac{\varepsilon^2}{2} \\
&\leq \frac{1}{\kappa n} H(F) + \frac{\varepsilon^2}{2} \leq \frac{1}{\kappa n} \log(2^{n\delta}) + \frac{\varepsilon^2}{2} \\
&\leq \frac{\log(2)\delta}{\kappa} + \frac{\varepsilon^2}{2} = \varepsilon^2, \quad \text{si } \delta = \frac{\kappa \varepsilon^2}{2\log(2)}.
\end{aligned}$$

On obtient enfin le résultat par l'inégalité de Markov. □

Je ne vois pas de preuve simple pour la réciproque, et Austin n'affirme pas qu'il en existe. Cependant, par [1, Proposition 13.1], l'extrémalité au sens d'Austin caractérise aussi les schémas de Bernoulli.

2.3.4 Lien avec la finie détermination

Nous avons déjà étudié en détails la notion de finie détermination dans le Chapitre 1, car elle est au cœur de la théorie d'Ornstein. Il sera agréable d'utiliser par la suite une définition un peu plus spécifique dans laquelle le "paramètre d'erreur" ε est fixé :

Définition 2.5. Soit (Y, ν, S) un système ergodique. Un processus (S, ζ) est relativement ε-finiment déterminé par rapport à un processus (S, π) s'il existe $\delta > 0$

et $n_1 \geq 1$, tels que, pour tout système ergodique (X, μ, T) muni de ξ et π' tels que $(T, \pi') \sim (S, \pi)$ et vérifiant

(i) $|d((\pi' \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| \leq \delta$,

(ii) et $|h(\pi' \vee \xi, \mu, T) - h(\pi \vee \zeta, \nu, S)| \leq \delta$,

on a : $\bar{d}((T, \xi), (S, \zeta) \,|\, \pi', \pi) \leq \varepsilon$.

Clairement, un processus (S, ζ) est relativement finiment déterminé si, et seulement si, pour tout $\varepsilon > 0$, il est relativement ε-finiment déterminé. On utilise cependant cette nouvelle notation car il sera utile dans la suite d'expliciter ε dans les étapes intermédiaires de la preuve du Théorème d'Austin.

Grâce au résultat [Chapitre 1, Lemme 3.1], la Définition 2.5 est complètement équivalente à celle donnée par Austin ([1, Definition 11.4]).

Dans la section précédente, j'ai tâché de donner quelques explications et quelques résultats pour donner au lecteur une compréhension de l'extrémalité, mais les informations les plus éclairantes se trouvent dans [1, Section 13] où T. Austin prouve qu'un processus relativement extrémal est relativement finiment déterminé ([1, Proposition 13.1]). Plus précisément, il prouve que :

Proposition 2.9. *Soit $r > 0$. S'il existe κ tel que ξ est relativement (κ, r)-extrémal par rapport π, alors ξ est relativement $(7r)$-finiment déterminé par rapport à π.*

Remarque 2.1. Il est facile de voir dans la preuve que, au lieu de supposer que ξ est relativement (κ, r)-extrémal par rapport π, on peut se contenter de supposer : Il existe $t \geq 1$ tel que

(i) $\xi_{[0,t[}$ est relativement $(\kappa t, r)$-extrémale par rapport à $\pi_{[0,t[}$,

(ii) et on a l'estimation

$$H(\xi_{[0,t[} \,|\, \pi_{[0,t[}) < (h(\xi, \mu, T \,|\, \pi) + \kappa r) \cdot t.$$

Il s'agit d'une très petite modification, mais elle nous sera pratique par la suite.

Le résultat particulièrement éclairant est [1, Proposition 13.3], qui donne la conséquence structurelle fondamentale de l'extrémalité relative. Prenons le temps de l'énoncer ici. Pour l'énoncer en toute généralité, il faut travailler sur un triplet (X, μ, T) tel que μ est T^n-invariante (alors que dans le reste du mémoire, toutes les mesures sont supposées T-invariantes). Avec cette remarque prise en compte, on a le résultat suivant :

Proposition 2.10. *On se donne deux processus ξ et π sur (X, μ, T) tels que*

(i) $\xi_{[0,n[}$ *est relativement $(n\kappa, r)$-extrémale par rapport à $\pi_{[0,n[}$,*

(ii) et $H(\xi_{[0,n[} \,|\, \pi_{[0,n[}) < h(\xi_{[0,n[}, \mu, T^n \,|\, \pi) + \kappa r n$.

Alors, pour tout $k \in \mathbb{N}$,

$$\int \bar{d}_{kn} \left(\mu_{|\pi_{[0,kn[}=b_{[0,kn[}}(\xi_{[0,kn[} = \cdot), \prod_{i=0}^{k-1} \mu_{|\pi_{[0,n[}=b_{[in,(i+1)n[}}(\xi_{[0,n[} = \cdot) \right) d\nu(\boldsymbol{b}) < 2r,$$

où ν est la loi de π.

Ce résultat s'obtient par k applications du Lemme 2.4, et en utilisant le Lemme 2.6 à chaque étape pour contrôler le terme d'erreur.

Si on analyse ce résultat, on voit qu'il permet de déterminer la loi du processus sur une durée arbitrairement grande à partir de la loi des n premières coordonnées, avec un terme d'erreur contrôlé par une estimation de l'entropie du processus : c'est précisément l'idée de la finie détermination. Pour les détails de ce raisonnement, on se référera à [1, Subsection 13.2].

2.4 Une finalisation alternative de la preuve

À l'aide des notions introduites ci-dessus, Austin applique ensuite son résultat de concentration de la mesure (Théorème 2.2) pour prouver le Théorème d'Austin (Théorème 2.1). Cette construction constitue la majeure partie de [1, Part III]. Dans sa preuve, Austin prouve d'abord son théorème en supposant que (Y, ν, S) est d'entropie finie et possède un ensemble N-périodique pour tout $N \geq 1$ (ce qu'il appelle Assumption (A)), et étend le résultat au cas général par des arguments d'équivalence orbitale.

Un ensemble F est dit N-*périodique* si les ensembles $T^j F, j \in \{0, ..., N-1\}$, forment une partition de $X \mod \mu$. L'appellation "périodique" est due au fait que l'on a alors : $T^N F = F \mod \mu$.

On donne ici une preuve alternative, en remplaçant l'hypothèse (A) par une application du Théorème de Rohlin (Théorème 2.11 pour être précis). L'idée de cette preuve vient d'une remarque dans l'article d'Austin ([1, p. 109]) dans laquelle il assure qu'un tel raisonnement est possible.

On définit une nouvelle notion, pour remplacer celle d'ensemble périodique :

Définition 2.6. Soit (X, μ, T) un système ergodique. Un ensemble F est dit (α, β, N)-presque périodique si :

(a) les ensembles $T^j F$, $j \in \{0, ..., N-1\}$, sont deux-à-deux disjoints,

(b) $\mu(\bigcup_{j=0}^{N-1} T^j F) \geq 1 - \beta$,

(c) et $\mu(F \Delta T^N F) \leq \alpha$.

Théorème 2.11. *Soit (X, μ, T) un système ergodique. Pour tous $\alpha, \beta > 0$ et $N \geq 1$, il existe un ensemble F qui est (α, β, N)-presque périodique.*

Le Théorème de Rohlin permet de prouver ce résultat car (c) découle de (b), si β est assez petit. Ainsi, on a pas besoin d'émettre d'hypothèse particulière pour utiliser un ensemble presque périodique : on peut directement donner une preuve valable pour tous les systèmes ergodiques.

On commence par un lemme élémentaire, mais bien utile lorsqu'on manipule des ensembles presque périodiques :

Lemme 2.12. *Soient (X, μ, T) un système ergodique, ξ et π deux v.a. sur X, et F et G deux parties π-mesurables de X telles que $\mu(F \Delta G) \leq 1/2$. On note A l'alphabet de ξ, et on a :*

$$|H_{\mu_{|G}}(\xi \,|\, \pi) - H_{\mu_{|F}}(\xi \,|\, \pi)| \leq \left(\frac{1}{\mu(F)\mu(G)} + \frac{1}{\mu(G)} \right) \log(\#A) \mu(F \Delta G).$$

De plus, si $\mu(G) = \mu(F)$, on a une meilleure borne :

$$|H_{\mu_{|G}}(\xi \,|\, \pi) - H_{\mu_{|F}}(\xi \,|\, \pi)| \leq \frac{1}{\mu(G)} \log(\#A) \mu(F \Delta G).$$

Démonstration. On utilise le fait que F et G soient π-mesurables pour voir que, pour μ-presque tout x dans F, on a

$$\mathbb{E}_{\mu_{|F}}[\mathbb{1}_{\xi_0 = a} \,|\, \pi \vee \xi_{]-\infty,0[}](x) = \mathbb{E}_{\mu}[\mathbb{1}_{\xi_0 = a} \,|\, \pi \vee \xi_{]-\infty,0[}](x),$$

et pour μ-presque tout x dans G, on a

$$\mathbb{E}_{\mu_{|G}}[\mathbb{1}_{\xi_0 = a} \,|\, \pi \vee \xi_{]-\infty,0[}](x).$$

Donc :

$$\begin{aligned}
|H_{\mu_{|G}}(\xi \,|\, \pi) - H_{\mu_{|F}}(\xi \,|\, \pi)| &= \left| \sum_{a \in A} \int_{\{\xi_0 = a\}} \log(\mathbb{E}_{\mu}[\mathbb{1}_{\xi_0 = a} \,|\, \pi \vee \xi_{]-\infty,0[}]) \, d(\mu_{|F} - \mu_{|G}) \right| \\
&\leq \left| \frac{1}{\mu(F)} - \frac{1}{\mu(G)} \right| \log(\#A) + \frac{1}{\mu(G)} \mu(F \Delta G) \log(\#A) \\
&\leq \left(\frac{1}{\mu(F)\mu(G)} + \frac{1}{\mu(G)} \right) \mu(F \Delta G) \log(\#A).
\end{aligned}$$

$\square$

Dans les manipulations des ensembles périodiques, Austin utilise souvent [1, Lemma 10.6]. On se donne donc une version adaptée aux ensembles presque périodiques :

Lemme 2.13. *Soient π une application facteur de (X, μ, T), ξ un processus à valeurs dans A et F un ensemble (α, β, N)-presque périodique et π-mesurable. Alors*

$$H(\xi_{[0,m[} \,|\, \pi\,;\, F) \geq H(\xi_{[0,m[} \,|\, \pi \vee \xi_{]-\infty,0[}\,;\, F) \geq (h(\xi, \mu, T \,|\, \pi) - 3\beta \log(\#A)) \cdot m,$$

où $H(\xi \,|\, \pi\,; F) = H_{\mu_{|F}}(\xi \,|\, \pi)$.

Je donne une preuve qui me semble un peu moins astucieuse que celle d'Austin.

Démonstration. L'inégalité de gauche est triviale. On se concentre donc sur celle de droite. On ne va pas s'inspirer de la preuve de [1, Lemme 10.6], on va plutôt utiliser la définition de l'entropie conditionnelle donnée dans le Chapitre 1 :

$$\begin{aligned} H_\mu(\xi \,|\, \mathcal{A}) &= -\sum_{a \in A} \int \mathbb{E}_\mu[\mathbb{1}_{\{\xi=a\}} | \mathcal{A}] \log(\mathbb{E}_\mu[\mathbb{1}_{\{\xi=a\}} | \mathcal{A}]) d\mu \\ &= -\sum_{a \in A} \int_{\{\xi=a\}} \log(\mathbb{E}_\mu[\mathbb{1}_{\{\xi=a\}} | \mathcal{A}]) d\mu \end{aligned}$$

On commence par utiliser la "chain rule" de l'entropie de Shannon :

$$\begin{aligned} H(\xi_{[0,m[} \,|\, \pi \vee \xi_{]-\infty,0[}\,;\, F) &= \sum_{i=0}^{m-1} H(\xi_i \,|\, \pi \vee \xi_{]-\infty,i[}\,; F) \\ &= \sum_{i=0}^{m-1} H(\xi_0 \,|\, \pi \vee \xi_{]-\infty,0[}\,; T^i F) \end{aligned}$$

Ensuite on utilise le fait que, pour tout $i \in \{0, ..., m-1\}$, $T^i F$ soit π-mesurable pour voir que, pour μ-presque tout x dans $T^i F$, on a

$$\mathbb{E}_{\mu_{|T^i F}}[\mathbb{1}_{\xi_0 = a} \,|\, \pi \vee \xi_{]-\infty,0[}](x) = \mathbb{E}_\mu[\mathbb{1}_{\xi_0=a} \,|\, \pi \vee \xi_{]-\infty,0[}](x).$$

Enfin, on pose $G = \bigcup_{j=0}^{m-1} T^j F$, et on a

$$\begin{aligned}
H(\xi_{[0,m[} \,|\, \pi \vee \xi_{]-\infty,0[}\,;\, F) &= \sum_{i=0}^{m-1} -\sum_{a\in A} \int_{\{\xi_0=a\}} \log(\mathbb{E}_\mu[\mathbb{1}_{\xi_0=a} \,|\, \pi \vee \xi_{]-\infty,0[}]) d\mu_{|T^iF} \\
&= -\sum_{a\in A} \int_{\{\xi_0=a\}} \log(\mathbb{E}_\mu[\mathbb{1}_{\xi_0=a} \,|\, \pi \vee \xi_{]-\infty,0[}])\; m \cdot d\mu_{|G} \\
&= m \cdot H_{\mu_{|G}}(\xi_0 \,|\, \pi \vee \xi_{]-\infty,0[}),
\end{aligned}$$

car $\mu_{|G} = \frac{1}{m}\sum_{i=0}^{N-1} \mu_{|T^iF}$. Mais alors, par le Lemme 2.12, on a :

$$H(\xi_{[0,m[} \,|\, \pi \vee \xi_{]-\infty,0[}\,;\, F) \geq m \cdot (H(\xi_0 \,|\, \pi \vee \xi_{]-\infty,0[}) - 3\beta \log(\#A)).$$

Par une formule classique de la KS-entropie, on conclue

$$H(\xi_{[0,m[} \,|\, \pi \vee \xi_{]-\infty,0[}\,;\, F) \geq m \cdot (h(\xi, \mu, T \,|\, \pi) - 3\beta \log(\#A)).$$

□

On est maintenant équipés pour passer à la preuve. Cette preuve étant technique et longue, on ne donnera pas tous les détails, et on référera le lecteur à l'article d'Austin pour certains passages.

2.4.1 Application du résultat de concentration de la mesure

On se donne un processus ξ sur (X, μ, T) à valeurs dans un alphabet fini A. L'élément clé de la construction est le résultat suivant (à rapprocher de [1, Proposition 14.2]), qui fait usage du Théorème 2.2 :

Proposition 2.14. *Pour tous $\varepsilon > 0, r > 0$, il existe un processus φ sur (X, μ, T), $\alpha, \beta > 0$, $N \geq 1$, $\kappa > 0$, et un ensemble (α, β, N)-presque périodique F φ_0-mesurable tels que*

(a) $h(\varphi, \mu, T \,|\, \pi) < \varepsilon$,

(b) $\xi_{[0,N[}$ *est relativement $(\kappa N, r)$-extrémal par rapport à $(\pi \vee \varphi)_{[0,N[}$ sur l'espace de probabilité $(X, \mu_{|F})$,*

(c) $H(\xi_{[0,N[} \,|\, (\pi \vee \varphi)_{[0,N[}\,;\, F) < (h(\xi, \mu, T \,|\, \pi \vee \varphi) + r\kappa) \cdot N$

(d) et on a les estimations :

$$\textit{(i)}\ \alpha \leq \min\left\{\frac{\kappa r^2}{10(H(\pi_0)+\log(2))}, \frac{r^2}{10}\right\} \ \ \textit{et (ii)}\ \beta \leq \frac{\kappa r}{3\log(\#A)}$$

.

Étant donnée la longueur de la preuve, Austin découpe la preuve en plusieurs étapes, et nous allons faire de même :

- *Étape 0 : Schéma de la preuve*

On décompose A^N : on écrit N comme $N = nl$ et alors

$$A^N = \underbrace{A^l \times \cdots \times A^l}_{n \text{ fois}} = \tilde{A}^n.$$

On voit donc A^N comme le produit de n copies du bloc $\tilde{A} = A^l$. On utilise le fait que l'entropie $l \mapsto H(\xi_{[0,l[} \,|\, \pi_{[0,l[})$ est asymptotiquement linéaire pour choisir l tel que

$$H(\xi_{[0,l[} \,|\, \pi_{[0,l[}) \approx h(\xi, \mu, T \,|\, \pi) \cdot l.$$

Cette approximation a pour conséquence :

$$H(\xi_{[0,nl[} \,|\, \pi_{[0,nl[}) \approx \sum_{j=0}^{n-1} H(\xi_{[jl,(j+1)\ell[} \,|\, \pi_{[jl,(j+1)l[}).$$

On choisit ensuite un ensemble F presque-périodique tel que cette estimation soit aussi vraie lorsqu'on se restreint à F :

$$H(\xi_{[0,nl[} \,|\, \pi_{[0,nl[}\,;\, F) \approx \sum_{j=0}^{n-1} H(\xi_{[jl,(j+1)l[} \,|\, \pi_{[jl,(j+1)l[}\,;\, F). \tag{6}$$

On travaille à partir de maintenant sur l'espace de probabilité $(X, \mu_{|F})$. On décompose alors la loi de $\xi_{[0,N[}$ par rapport à $\pi_{[0,N[}$:

$$\tilde{\lambda}(\mathbf{a} \,|\, \mathbf{b}) = \mu_{|\{\pi_{[0,N[}=\mathbf{b}\}\cap F}(\xi_{[0,N[} = \mathbf{a}),$$

pour $\mathbf{a} \in A^N$ et $\mathbf{b} \in B^N$. Par (6), on voit que

$$\int E(\tilde{\lambda}(\cdot \,|\, \mathbf{b}))d\tilde{\nu}(\mathbf{b}) \leq \delta^2 N, \tag{7}$$

pour un petit paramètre δ.

On applique ensuite le Théorème 2.2 à la famille de mesures $(\tilde{\lambda}(\cdot \,|\, \mathbf{b}))_{\mathbf{b}\in B^N}$, pour obtenir une famille de partitions de $\tilde{A}^n$

$$\forall \mathbf{b} \in B^N,\ \tilde{A}^n = U_{\mathbf{b},1} \cup \cdots \cup U_{\mathbf{b},m_{\mathbf{b}}}.$$

Grâce à (7) et un bon choix de δ, on sait que, pour la plupart des $\mathbf{b} \in B^N$,

$$m_{\mathbf{b}} \leq \exp(\varepsilon N). \tag{8}$$

En transposant ces partitions à travers ξ et π, on construit un nouveau processus ψ sur X. Et alors, par construction,

$$(\mathbf{b}, j) \mapsto \tilde{\lambda}(\cdot \,|\, \mathbf{b})\Big|_{U_{\mathbf{b},j}}$$

est un loi conditionnelle de $\xi_{[0,N[}$ sachant $(\psi \vee \pi)_{[0,N[}$ sur $(X, \mu_{|F})$. Donc, par le Théorème 2.2, $\xi_{[0,N[}$ *est relativement* $(\kappa N, 3r)$*-extrémale par rapport à* $(\psi \vee \pi)_{[0,N[}$ sur $(X, \mu_{|F})$.

Ensuite, l'estimation de l'entropie découle de (8) et du Lemme 2.13 :

$$h(\psi, \mu, T) \leq \frac{1}{N} H(\psi_{[0,N[}\,;\, F) \leq \frac{1}{N} \log(m) < \varepsilon,$$

(cette ligne étant un peu trop simplifiée pour être rigoureuse).

Pour finir, on pose τ le processus engendré par F et $\varphi := \tau \vee \psi$. Les conditions (c) et (d) s'obtiennent par un choix judicieux des paramètres au cours de cette preuve.

- *Étape 1 : Choix des paramètres*

On peut supposer que ε est inférieur à $\log(2)$. On se donne ensuite un certain nombre de paramètres, qui ne seront plus modifiés ensuite :

(P1) (Constantes données par le Théorème 2.2) Soient c_0 et κ_0 les constantes fournies par le Théorème 2.2 lorsque les deux les deux termes d'erreurs valent r. On voit ici qu'il est important de pourvoir les choisir indépendamment de l'alphabet auquel on va appliquer le Théorème 2.2. On introduit alors un nouveau terme d'erreur $\delta < \min\{\varepsilon/c_0, r/2\}$.

(P2) Grâce à [Chapitre 1, Lemme 2.4], on fixe une première échelle de temps $l \in \mathbb{N}$ suffisamment grande pour avoir

$$H(\xi_{[0,l[} \,|\, \pi_{[0,l[}) < (h(\xi, \mu, T \,|\, \pi) + \delta^2/4) \cdot l.$$

On fixe ensuite $\kappa = \kappa_0/l$. On pose aussi $\tilde{A} = A^l$, et c'est à cet alphabet que l'on appliquera le Théorème 2.2.

(P3) On choisit une nouvelle échelle de temps $n \in \mathbb{N}$ suffisamment grande pour avoir :

(P3.1) une deuxième estimation de l'entropie

$$H(\xi_{[0,nl[} \,|\, \pi_{[0;nl[}) < (h(\xi, \mu, T \,|\, \pi) + r\kappa/4) \cdot nl,$$

(P3.2) n est assez grand pour appliquer le Théorème 2.2 à $\tilde{A}^n$,

(P3.3) comme $\delta < \varepsilon/c_0$, on peut demander :

$$c_0 \exp(c_0 \delta nl) < \exp(\varepsilon nl),$$

(P3.4) et on prend n assez grand pour avoir

$$\frac{\log(nl+1)}{nl} < \varepsilon.$$

On pose alors $N = nl$.

(P4) On fixe les termes d'erreurs α et β :

(P4.1)

$$\alpha \leq \min\left\{ \frac{\kappa r^2}{10(H(\pi_0) + \log(2))}, \frac{r^2}{10} \right\},$$

(P4.2) on veut que $\beta \leq \min\{1/2, r\}$ ainsi que

$$\text{(i) } \beta \leq \frac{\min\{\delta^2, \kappa r\}}{12 \log(\#A)} \;\text{ et }\; \text{(ii) } c_0 \exp(c_0 \delta N) \cdot 2^{3\beta N} < \exp(\varepsilon N),$$

(P4.3) et on veut avoir

$$\log(N+1)\beta < \varepsilon.$$

Le choix fait ci-dessus de se donner deux échelles de temps l et n est très important : la partition donnée par le Théorème 2.2 sera la base de la construction du nouveau processus, et le fait que ce processus ait une petite entropie découlera du fait que cette partition comporte peu d'éléments. Or, le nombre d'éléments dans cette partition est contrôlée par la corrélation totale E , qui évalue la corrélation entre les projections sur chaque coordonnée de $\tilde{A}^n$. La majoration de E découle de (P2), donc si on avait appliqué ce théorème directement à A^N, on aurait eu aucun moyen de majorer E. Ainsi, il est plus intéressant de voir A^N comme le produit de n copies de A^l que de le voir comme le produit de N copies de A (bien que $A^N = A^{nl} = \tilde{A}^n$). Les détails de cette explication qualitative apparaissent dans les étapes 3 et 4 de cette preuve.

Bien que l'on ne donne pas tous les détails de la preuve, on fera en sorte de signaler les endroits où chacune de ces estimations est utilisée.

- *Étape 2 : Choix d'un ensemble presque périodique*

C'est ici que le Théorème 2.11 intervient : on se donne un ensemble F_0 sur X, (α, β, N)-presque périodique et π-mesurable. Comme les ensembles $T^i F_0$, $i \in \{0, ..., N-1\}$ forment une partition de X, l'entropie de Shannon sur X peut être obtenue comme la moyenne des entropies sur chaque $T^i F_0$. Ainsi, par l'inégalité de Markov et le lemme des tiroirs, les estimations (P2) et (P3.1) montrent qu'il existe $i \in \{0, ..., N-1\}$ tel que

$$H(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, T^i F_0) < (h(\xi, \mu, T \,|\, \pi) + r\kappa) \cdot N, \tag{9}$$

et

$$\frac{1}{n} \sum_{j=0}^{n-1} H(\xi_{[0,l[} \,|\, \pi_{[0;l[}\,;\; T^{i+jl} F_0) < (h(\xi, \mu, T \,|\, \pi) + \delta^2) \cdot l. \tag{10}$$

Ce résultat correspond à [1, Lemma 14.3]. Austin le fait dans le cas d'un ensemble périodique, mais dans le cas de notre ensemble presque périodique, son raisonnement marche aussi, en remplaçant [1, Lemma 10.6] par le Lemme 2.13 et en faisant apppel à (P4.2)(i). On pose alors $F = T^i F_0$, qui est encore un ensemble (α, β, N)-presque périodique.

- *Étape 3 : Contrôle des paramètres du Théorème 2.2*

Cette étape me semble particulièrement importante, et est auto-contenue, je vais donc la présenter en détails. C'est ici que l'on introduit les mesures auxquelles on appliquera le Théorème 2.2. Il s'agit du noyau de mesures suivant :

$$\tilde{\lambda}(\mathbf{a} \,|\, \mathbf{b}) = \mu_{|\{\pi_{[0,N[}=\mathbf{b}\}\cap F}(\xi_{[0,N[} = \mathbf{a}),$$

pour $\mathbf{a} \in \tilde{A}^n$ et $\mathbf{b} \in B^N$. C'est un noyau de B^N vers $\tilde{A}^n$. On lui associe la mesure de probabilité sur B^N suivante :

$$\tilde{\nu}(\mathbf{b}) = \mu_{|F}(\pi_{[0,N[} = \mathbf{b}).$$

Le but de cette étape est de montrer le résultat suivant :

Lemme 2.15. *On pose $W \subset B^N$ l'ensemble des mots $\boldsymbol{b} \in B^N$ tels que :*

$$\tilde{\nu}(\boldsymbol{b}) > 0 \ \text{ et } \ E(\tilde{\lambda}(\cdot \,|\, \boldsymbol{b})) \leq N\delta.$$

Alors $\tilde{\nu}(W) \geq 1 - 2\delta$.

Il est important de rappeler que l'on regarde $\tilde{\lambda}(\cdot \,|\, \mathbf{b})$ comme une mesure sur $\tilde{A}^n$, et que les lois marginales qui apparaissent dans la définition de E sont les projections de $\tilde{\lambda}(\cdot \,|\, \mathbf{b})$ sur chaque coordonnée de $\tilde{A}^n$, et non pas sur celles de A^N.

Démonstration. Pour tout $\mathbf{b}$ tel que $\tilde{\nu}(\mathbf{b}) > 0$, $\tilde{\lambda}(\cdot \,|\, \mathbf{b})$ est la loi de $\xi_{[0,N[}$ sachant l'événement $\{\pi_{[0,N[} = \mathbf{b}\} \cap F$. Les marginales qui apparaissent dans la définition de $E(\tilde{\lambda}(\cdot \,|\, \mathbf{b}))$ sont alors les lois des variables aléatoires

$$\xi_{[0,l[}, \ldots, \xi_{[(n-1)l,nl[},$$

sachant l'événement $\{\pi_{[0,N[} = \mathbf{b}\} \cap F$. Ainsi la corrélation totale s'écrit :

$$E(\tilde{\lambda}(\cdot \,|\, \mathbf{b})) = \sum_{i=0}^{n-1} H_\mu(\xi_{[il,(i+1)l[}\,; \{\pi_{[0,N[} = \mathbf{b}\} \cap F) - H_\mu(\xi_{[0,N[}\,; \{\pi_{[0,N[} = \mathbf{b}\} \cap F).$$

On voit facilement, par la définition de l'entropie conditionnelle, qu'en intégrant l'égalité ci-dessus, on obtient

$$\int E(\tilde{\lambda}(\cdot \,|\, \mathbf{b})) d\tilde{\nu}(\mathbf{b}) = \sum_{i=0}^{n-1} H_\mu(\xi_{[il,(i+1)l[} \,|\, \pi_{[0,N[}\,; F) - H_\mu(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,; F).$$

On peut ensuite majorer le terme de droite pour avoir :

$$\begin{aligned}
\int E(\tilde{\lambda}(\cdot \,|\, \mathbf{b})) d\tilde{\nu}(\mathbf{b}) &\leq \sum_{i=0}^{n-1} H_\mu(\xi_{[il,(i+1)l[} \,|\, \pi_{[il,(i+1)l[}\,; F) - H_\mu(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,; F) \\
&= \sum_{i=0}^{n-1} H_\mu(\xi_{[0,l[} \,|\, \pi_{[0,l[}\,; T^{il}F) - H_\mu(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,; F) \\
&\leq (h(\xi, \mu, T \,|\, \pi) + \delta^2) \cdot N - H_\mu(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,; F),
\end{aligned}$$

où la dernière inégalité provient de (10). Enfin, le Lemme 2.13 donne

$$\int E(\tilde{\lambda}(\cdot \,|\, \mathbf{b}))d\tilde{\nu}(\mathbf{b}) \leq \delta^2 N + 3\beta \log(\#A)N \leq 2\delta^2 N,$$

en utilisant de nouveau (P4.2)(i). Le résultat du lemme découle alors par l'inégalité de Markov. □

On remarque dans cette preuve l'application cruciale de (10), qui prouve l'intérêt d'utiliser l'alphabet $\tilde{A}$ plutôt que A.

Grâce à ce résultat, on sait que l'on pourra appliquer le Théorème 2.2 aux mesures choisies et avoir un bon contrôle sur la plus part des partitions obtenues. C'est ce que l'on fait dans l'étape suivante.

- *Étape 4 : Application du Théorème 2.2*

Soit $\mathbf{b} \in B^N$. On considère alors deux cas

- Si $\mathbf{b} \in W$:

Par le Théorème 2.2, il existe $m_{\mathbf{b}}$ et une partition de $\tilde{A}^n$ en $m_{\mathbf{b}}$ éléments :

$$\tilde{A}^n = U_{\mathbf{b},1} \cup \cdots \cup U_{\mathbf{b},m_{\mathbf{b}}},$$

telle que

(i) $m_{\mathbf{b}} \leq c_0 e^{c_0 \delta N} \leq e^{\varepsilon N}$, la deuxième inégalité provenant de (P3.3),

(ii) $\tilde{\lambda}(U_{\mathbf{b},1} \,|\, \mathbf{b}) \leq r$,

(iii) et pour tout $j \in \{2, ..., m_{\mathbf{b}}\}$, $\tilde{\lambda}(U_j \,|\, \mathbf{b}) > 0$ et $\tilde{\lambda}(\cdot \,|\, \mathbf{b})\big|_{U_{\mathbf{b},j}}$ satisfait $T(\kappa_0 n, r)$.

Ici, on a utilisé (P3.2).

On va maintenant cesser d'utiliser $\tilde{A}$ et considérer A^N naturellement comme le produit de N copies de A. Cependant, une discussion est nécessaire ici : la distance de Hamming, définie par la formule (1), n'est pas la même sur A^N et sur $\tilde{A}^n$. Notons d_N la distance de Hamming sur A^N et d_n celle sur $\tilde{A}^n$, et vérifions rapidement que $d_N \leq d_n$: soient $\mathbf{a}, \mathbf{a}' \in A^N$

$$\begin{aligned} d_N(\mathbf{a}, \mathbf{a}') &= \frac{1}{N} \#\{i \in [0, N[\,|\, a_i \neq a'_i\} \\ &\leq \frac{1}{N} l \, \#\{j \in [0, n[\,|\, (a_{jl}, ..., a_{jl+l-1}) \neq (a'_{jl}, ..., a'_{jl+l-1})\} \\ &= \frac{1}{n} \#\{j \in [0, n[\,|\, (a_{jl}, ..., a_{jl+l-1}) \neq (a'_{jl}, ..., a'_{jl+l-1})\} = d_n(\mathbf{a}, \mathbf{a}'). \end{aligned}$$

Donc on aura aussi $\bar{d_N} \leq \bar{d_n}$, donc si $\tilde{\lambda}(\cdot\,|\,\mathbf{b})\big|_{U_{\mathbf{b},j}}$ satisfait $T(\kappa_0 n, r)$ pour la distance d_n, elle la satisfait aussi pour la distance d_N. De plus, comme on a $\kappa = \kappa_0/l$, on peut alors dire que $\tilde{\lambda}(\cdot\,|\,\mathbf{b})\big|_{U_{\mathbf{b},j}}$ satisfait $T(\kappa N, r)$. Enfin, quitte à rajouter un crtain nombre de copies de l'ensemble vide, on peut supposer qu'il existe m vérifiant (i) et tel que $m_{\mathbf{b}} = m$, pour tout $\mathbf{b} \in W$.

- Si $\mathbf{b} \notin W$:

Dans ce cas, on se contente de choisir une partition de $\tilde{A}^n$ en m éléments quelconque.

On peut alors définir $\Psi : B^N \times A^N \longrightarrow \{1, ..., m\}$ par

$$\Psi(\mathbf{b}, \mathbf{a}) = j \;\text{ si } \mathbf{a} \in U_{\mathbf{b},j},$$

puis $\Phi : F \longrightarrow \{1, ..., m\}$ par la formule

$$\Phi(x) := \psi(\pi_{[0,N[}(x), \xi_{[0,N[}(x)).$$

Cette construction terminée, on a le résultat suivant (qui correspond à [1, Lemma 14.5] dans l'article d'Austin) :

Affirmation 2.15.1. *On a construit Φ de sorte que $\xi_{[0,N[}$ soit relativement $(\kappa N, 3r)$-extrémale par rapport à $\pi_{[0,N[} \vee \Phi$ sur l'espace de probabilité $(X, \mu_{|F})$.*

En effet, par construction, la famille

$$\lambda_{\mathbf{b},j} = \tilde{\lambda}(\cdot\,|\,\mathbf{b})\big|_{U_{\mathbf{b},j}},\ \mathbf{b} \in B^N \;\text{ et }\; j \in \{1, ..., m\},$$

est une loi conditionnelle de $\xi_{[0,N[}$ sachant $\pi_{[0,N[} \vee \Phi$ (ce fait est détaillé par Austin dans la preuve de [1, Lemma 14.5]). Par le Théorème 2.2, on sait que la plus part de ces mesures satisfont $T(\kappa N, r)$. Plus précisément, on a

$$\begin{aligned}
&\mu_{|F}(\pi_{[0,N[} \vee \Phi \in \{(\mathbf{b}, j)\,|\,\lambda_{\mathbf{b},j} \text{ satisfait } T(\kappa N, r)\}) \\
&\qquad \geq 1 - \mu_{|F}(\pi_{[0,N[} \notin W) - \mu_{|F}(\pi_{[0,N[} \in W \text{ et } \Phi \in U_{\pi_{[0,N[},1}) \\
&\qquad \geq 1 - 2\delta - r \geq 1 - 2r,
\end{aligned}$$

où l'avant-dernière inégalité s'obtient par le Théorème 2.2 et le Lemme 2.15. La dernière inégalité découle du choix de δ fait dans (P1).

On laisse le lecteur se convaincre que toute mesure sur A^N satisfait $T(\kappa', 1)$ pour tout κ', et en particulier elle satisfait $T(\kappa N, 1)$. En remarquant que toute mesure satisfaisant $T(\kappa N, r)$ est $(\kappa N, r)$-extrémale, on conclue que $\xi_{[0,N[}$ est bien relativement $(\kappa N, 3r)$-extrémal par rapport à $\pi_{[0,N[} \vee \Phi$ sur l'espace de probabilité $(X, \mu_{|F})$, et l'affirmation est donc démontrée.

- *Étape 5 : Construction du nouveau processus*

À ce stade, on a construit une variable aléatoire Φ sur F, et pas encore un processus. Pour construire le processus, on va utiliser un procédé auquel on a déjà eu recours de nombreuses fois dans la théorie d'Ornstein ([Chapitre 1, Section 3]). Pour cela, on va ré-indexer Φ : comme $m \leq e^{\varepsilon N}$ et $\varepsilon \leq \log(2)$, on peu choisir un ensemble $S \subset \{0,1\}^{[0,N[}$ de cardinal m. On peut alors supposer que Φ est à valeurs dans S. Maintenant que Φ est indexée par des suites de N éléments, on peut construire une variable aléatoire ψ_0 sur $\bigcup_{j=0}^{N-1} T^j F$ telle que :

$$\psi_{[0,N[}\big|_F = \Phi. \tag{11}$$

On étend alors ψ_0 à X tout entier de manière arbitraire, en s'assurant simplement que ψ_0 est $\xi \vee \pi$-mesurable. Une dernière modification est nécessaire : on introduit le processus engendré par $\tau_0 : X \longrightarrow \{0, ..., N-1, N\}$ engendrée par F, que l'on définit par

$$\tau(x) = j \;\text{ si }\; x \in T^j F, \text{ pour } j \in \{1, ..., N-1\}, \tag{12}$$

et

$$\tau(x) = N \;\text{ si }\; x \notin \bigcup_{j=0}^{N-1} T^j F. \tag{13}$$

Enfin, on pose $\varphi := \psi \vee \tau$. On remarque bien que c'est un changement mineur, car $\varphi \vee \pi$ et $\psi \vee \pi$ engendrent le même facteur (car F est π-mesurable), mais il permet d'assurer que F est φ_0-mesurable.

Démonstration de la Proposition 2.14. Il nous reste à montrer que le processus φ construit ci-desssus vérifie les conditions (a), (b) et (c) de la proposition (la condition (d) est déjà vérifiée par (P4)).

On commence par majorer l'entropie de τ :

$$\begin{aligned}
h(\tau,\mu,T) &= H(\tau_0 \,|\, \tau_{]-\infty,0[}) \\
&= \int -\sum_{j=0}^{N} \mathbb{E}[\mathbb{1}_{\tau_0=j} \,|\, \tau_{]-\infty,0[} = b_{]-\infty,0[}] \log(\mathbb{E}[\mathbb{1}_{\tau_0=j} \,|\, \tau_{]-\infty,0[} = b_{]-\infty,0[}]) d\theta(\mathbf{b}) \\
&\leq \log(N+1)\left(\frac{1}{N}+\beta\right) \\
&+ \int_{b_{-1}\in\{0,\ldots,N-2\}} -\sum_{j=0}^{N} \mathbb{1}_{\tau_0=j} \log(\mathbb{E}[\mathbb{1}_{\tau_0=j} \,|\, \tau_{]-\infty,0[} = b_{]-\infty,0[}]) d\theta(\mathbf{b}) \\
&= \log(N+1)\left(\frac{1}{N}+\beta\right),
\end{aligned}$$

car si $\tau_{-1} \in \{0, N-2\}$, on a $\tau_0 = \tau_{-1}+1$ μ-p.s.. Et alors, par (P3.4) et (P4.3), on a

$$h(\tau,\mu,T) < 2\varepsilon.$$

Ensuite, on a

$$h(\varphi,\mu,T) \leq h(\psi,\mu,T \,|\, \tau) + h(\tau,\mu,T).$$

De plus, en utilisant le Lemme 2.13 et (11) :

$$\begin{aligned}
h(\psi,\mu,T \,|\, \tau) &\leq \frac{1}{N} H(\psi_{[0,N[}\,;\, F) + 3\beta\log(2) \\
&= \frac{1}{N} H(\Phi\,;\, F) + 3\beta\log(2) \leq \frac{1}{N}\log(\#S) + 3\beta\log(2) < \varepsilon,
\end{aligned}$$

la dernière inégalité provenant de (P4.2)(ii). Donc

$$h(\varphi,\mu,T) < 3\varepsilon.$$

Les lois conditionnelles de $\xi_{[0,N[}$ sachant $(\pi\vee\varphi)_{[0,N[}$ et celles sachant $(\pi\vee\psi)_{[0,N[}$ sur l'espace de probabilité $(X,\mu_{|F})$ sont les mêmes, car τ est constante sur F. En effet, si $\mu_\bullet$ est une loi conditionnelle de $\xi_{[0,N[}$ sachant $(\pi\vee\psi)_{[0,N[}$, alors elle est caractérisée par : pour tout $(\mathbf{b},\mathbf{c}) \in (B\times\{0,1\})^{[0,N[}$

$$\int_{(\pi\vee\psi)_{[0,N[}=(\mathbf{b},\mathbf{c})} \mu_x(\cdot) d\mu_{|F}(x) = \mu_{|F}(\{(\pi\vee\psi)_{[0,N[} = (\mathbf{b},\mathbf{c})\} \cap \{\xi_{[0,N[} \in \cdot\}).$$

Il est alors clair que, comme τ est constante sur F, on peut écrire : pour tout $(\mathbf{b},\mathbf{c}) \in (B \times \{0,1\})^{[0,N[}$ et pour tout $j \in \{0,...,N\}$

$$\int_{(\pi\vee\psi)_{[0,N[}=(\mathbf{b},\mathbf{c}),\tau=j} \mu_x(\cdot)d\mu_{|F}(x) = \mu_{|F}(\{(\pi\vee\psi)_{[0,N[} = (\mathbf{b},\mathbf{c})\}\cap\{\tau = j\}\cap\{\xi_{[0,N[} \in \cdot\}),$$

ce qui prouve que $\mu_\bullet$ est une loi conditionnelle de $\xi_{[0,N[}$ sachant $(\pi \vee \varphi)_{[0,N[}$. Ensuite, l'Affirmation 2.15.1 nous permet de conclure que $\xi_{[0,N[}$ est relativement $(\kappa N, 3r)$-extrémale par rapport à $(\pi \vee \varphi)_{[0,N[}$ sur l'espace $(X, \mu_{|F})$.

Vérifions que l'inégalité (c) est vérifiée. Par construction, sur F, $\psi_{[0,N[}$ est $(\xi \vee \pi)_{[0,N[}$-mesurable, et par conséquent, on a par la "chain rule"

$$\begin{aligned} H(\xi_{[0,N[} \vee \psi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F) &= H(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F) + H(\psi_{[0,N[} \,|\, \xi_{[0,N[} \vee \pi_{[0,N[}\,;\, F) \\ &= H(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F). \end{aligned}$$

En utilisant ce résultat, encore la "chain rule" et le fait que $\tau_{[0,N[}$ est constant sur F, on a

$$\begin{aligned} H(\xi_{[0,N[} \,|\, (\pi \vee \varphi)_{[0,N[}\,;\, F) &= H(\xi_{[0,N[} \,|\, (\pi \vee \psi)_{[0,N[}\,;\, F) \\ &= H(\xi_{[0,N[} \vee \psi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F) - H(\psi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F) \\ &= H(\xi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F) - H(\psi_{[0,N[} \,|\, \pi_{[0,N[}\,;\, F) \end{aligned}$$

Alors, par (9) et le Lemme 2.13, on a

$$\begin{aligned} H(\xi_{[0,N[} \,|\, (\pi \vee \varphi)_{[0,N[}\,;\, F) &\leq (h(\xi,\mu,T \,|\, \pi) + \kappa r) \cdot N - h(\psi,\mu,T \,|\, \pi) \cdot N + 3\beta \log(2) N \\ &= (h(\xi,\mu,T \,|\, \pi) - h(\psi,\mu,T \,|\, \pi) + 2\kappa r) \cdot N, \end{aligned}$$

en utilisant une fois de plus (P4.2)(i). En appliquant la "chain rule" de la KS-entropie, on a

$$h(\xi,\mu,T \,|\, \pi) - h(\psi,\mu,T \,|\, \pi) \leq h(\xi\vee\psi,\mu,T \,|\, \pi) - h(\psi,\mu,T \,|\, \pi) = h(\xi,\mu,T \,|\, \pi\vee\psi),$$

et donc

$$\begin{aligned} H(\xi_{[0,N[} \,|\, (\pi \vee \varphi)_{[0,N[}\,;\, F) &\leq (h(\xi,\mu,T \,|\, \pi \vee \psi) + 2\kappa r) \cdot N \\ &= (h(\xi,\mu,T \,|\, \pi \vee \varphi) + 2\kappa r) \cdot N. \end{aligned}$$

Et comme $2r < 3r$, on a prouvé la proposition avec les termes d'erreur de 3ε et $3r$, mais comme ε et r sont arbitrairement petits, cela prouve la proposition. $\square$

2.4.2 Un ajustement de la Proposition 2.10

La construction du processus φ dans la section précédente contient tous les arguments importants de la preuve de la propriété de Pinsker faible, mais malheureusement, le résultat n'est pas suffisant à cause des conditions (b) et (c) dans lesquelles on travail sur l'espace de probabilité $(X, \mu_{|F})$ et non (X, μ). Le but de cette section est de corriger ce problème.

La première étape est de prouver une variante de la Proposition 2.10 qu'Austin utilise dans le cas où F est un ensemble périodique. Plus précisément :

Proposition 2.16. *Dans le contexte de la Proposition 2.14, on pose $\tilde{\pi} = \pi \vee \varphi$ et on a*

$$\int \bar{d}_{kN}\left(\mu_{|F\cap\{\tilde{\pi}_{[0,kN[}=b_{[0,kN[}\}}(\xi_{[0,kN[} = \cdot), \prod_{i=0}^{k-1} \mu_{|F\cap\{\tilde{\pi}_{[0,N[}=b_{[iN,(i+1)N[}\}}(\xi_{[0,N[} = \cdot)\right) d\nu(\boldsymbol{b}) < 3r + 8\alpha k,$$

avec $\nu = \tilde{\pi}_* \mu_{|F}$.

Démonstration. La preuve que l'on donne ici va principalement suivre la preuve de [1, Proposistion 13.3], en ajoutant quelques estimations supplémentaires. On prouve par récurrence sur k l'assertion suivante : pour tout $S \subset \mathbb{Z}$ tel que $[0, kN[\subset S$, on a

$$I_{k,S} := \int \bar{d}_{kN}\left(\mu_{|F\cap\{\tilde{\pi}_S=b_S\}}(\xi_{[0,kN[} = \cdot), \prod_{i=0}^{k-1} \mu_{|F\cap\{\tilde{\pi}_{[0,N[}=b_{[iN,(i+1)N[}\}}(\xi_{[0,N[} = \cdot)\right) d\nu(\mathbf{b}) < 3r + 8\alpha k.$$

Soit $k \geq 1$. On traite le cas $k = 1$ en même temps que les autres. On suppose le résultat vrai pour $k - 1$ et on va montrer qu'il est vrai pour k, et si $k = 1$, on ne suppose rien du tout.

Soit S tel que $[0, kN[\subset S$. Pour $\mathbf{b} \in B^S$ et $\mathbf{a} \in (A^N)^{(k-1)}$, on pose

$$\lambda(\cdot \,|\, \mathbf{b}, \mathbf{a}) := \mu_{|F\cap\{\tilde{\pi}_S=\mathbf{b}\}\cap\{\xi_{[0,(k-1)N[}=\mathbf{a}\}}(\xi_{[(k-1)N,kN[} = \cdot).$$

Si $k = 1$, on ignore $\mathbf{a}$, et on pose $\lambda(\cdot \,|\, \mathbf{b}) := \mu_{|F\cap\{\tilde{\pi}_S=\mathbf{b}\}}(\xi_{[0,N[} = \cdot)$. On a construit $\lambda(\cdot \,|\, \cdot, \cdot)$ de sorte que, pour tout $\mathbf{b}$,

$$\mathbf{a} \in (A^N)^{(k-1)} \mapsto \lambda(\cdot \,|\, \mathbf{b}, \mathbf{a}) \in Prob(A^N)$$

soit une loi conditionnelle de $\xi_{[(k-1)N,kN[}$ sachant $\xi_{[0,(k-1)N[}$ sur l'espace de probabilité $(X,\mu_{|F\cap\{\tilde{\pi}_S=\mathbf{b}\}})$. Alors le Lemme 2.4 donne l'estimation

$$\begin{aligned} I_{k,S} \leq {} & \frac{k-1}{k} I_{k-1,S} \\ & + \frac{1}{k}\int_{B^{\mathbb{Z}}}\int_{(A^N)^{(k-1)}} \bar{d}_N\Big(\lambda(\cdot\,|\,b_S,\mathbf{a}),\mu_{|F\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\}}(\xi_{[0,N[}=\cdot)\Big) \\ & \qquad\qquad d\mu_{|F\cap\{\tilde{\pi}_S=b_S\}}(\xi_{[0,(k-1)N[}=\mathbf{a})d\nu(\mathbf{b}) \end{aligned}$$

Par l'hypothèse de récurrence, on peut majorer $I_{k-1,S}$:

$$I_{k-1,S} \leq 3r + 8\alpha(k-1). \tag{14}$$

Il nous reste à traiter le second terme. Par l'inégalité triangulaire, il est inférieur à

$$\begin{aligned} & \frac{1}{k}\int_{B^S}\int_{(A^N)^{(k-1)}} \bar{d}_N\Big(\lambda(\cdot\,|\,\mathbf{b},\mathbf{a}),\mu_{|F\cap\{\tilde{\pi}_{[(k-1)N,kN[}=b_{[(k-1)N,kN[}\}}(\xi_{[(k-1)N,kN[}=\cdot)\Big) \\ & \qquad\qquad d\mu_{|F}(\{\xi_{[0,(k-1)N[}=\mathbf{a}\}\cap\{\pi_S=\mathbf{b}\}) \\ & + \frac{1}{k}\int_{B^{\mathbb{Z}}}\int_{(A^N)^{(k-1)}} \bar{d}_N\Big(\mu_{|F\cap\{\tilde{\pi}_{[(k-1)N,kN[}=b_{[(k-1)N,kN[}\}}(\xi_{[(k-1)N,kN[}=\cdot), \\ & \mu_{|F\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\}}(\xi_{[0,N[}=\cdot)\Big)d\mu_{|F\cap\{\tilde{\pi}_S=b_S\}}(\xi_{[0,(k-1)N[}=\mathbf{a})d\nu(\mathbf{b}) \\ & \qquad\qquad = \frac{1}{k}(J_1+J_2). \end{aligned}$$

Majorons d'abord J_2 :

$$\begin{aligned} J_2 &= \int_{B^{\mathbb{Z}}} \bar{d}_N\Big(\mu_{|F\cap\{\tilde{\pi}_{[(k-1)N,kN[}=b_{[(k-1)N,kN[}\}}(\xi_{[(k-1)N,kN[}=\cdot), \\ & \qquad \mu_{|F\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\}}(\xi_{[0,N[}=\cdot)\Big)d\nu(\mathbf{b}) \\ &= \int_{B^{\mathbb{Z}}} \bar{d}_N\Big(\mu_{|T^{(k-1)N}F\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\}}(\xi_{[0,N[}=\cdot), \\ & \qquad \mu_{|F\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\}}(\xi_{[0,N[}=\cdot)\Big)d\nu(\mathbf{b}) \\ &\leq \int_{B^{\mathbb{Z}}} 2\frac{\mu((F\Delta T^{(k-1)N}F)\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\})}{\mu(F\cap\{\tilde{\pi}_{[0,N[}=b_{[(k-1)N,kN[}\})}d\nu(\mathbf{b}) \\ &= 2\frac{\mu(F\Delta T^{(k-1)N})}{\mu(F)} \leq 2(k-1)\alpha. \end{aligned}$$

Ensuite, J_1 peut être majoré grâce au Lemme 2.6. Il faut tout de même faire attention lorsqu'on fait cela : $\xi_{[0,N[}$ est relativement $(\kappa N, r)$-extrémale par rapport à $\tilde{\pi}_{[0,N[}$ sur $(X, \mu_{|F})$, donc, par T-invariance de μ, $\xi_{[(k-1)N,kN[}$ est relativement $(\kappa N, r)$-extrémale par rapport à $\tilde{\pi}_{[(k-1)N,kN[}$ sur $(X, \mu_{|T^{-(k-1)N}F})$. Il faut donc s'assurer qu'on utilise les bonnes mesures de probabilité. Pour cela, on écrit :

$$\begin{aligned} J_1 \leq \int_{B^S} \int_{(A^N)^{(k-1)}} \bar{d_N} \Big(& \mu_{|T^{-(k-1)N}F \cap \{\tilde{\pi}_S = \mathbf{b}\} \cap \{\xi_{[0,(k-1)N[} = \mathbf{a}\}}(\xi_{[(k-1)N,kN[} = \cdot), \\ & \mu_{|T^{-(k-1)N}F \cap \{\tilde{\pi}_{[(k-1)N,kN[} = b_{[(k-1)N,kN[}\}}(\xi_{[(k-1)N,kN[} = \cdot) \Big) \\ & \qquad d\mu_{|T^{-(k-1)N}F}(\{\xi_{[0,(k-1)N[} = \mathbf{a}\} \cap \{\pi_S = \mathbf{b}\}) \\ & \qquad\qquad + 6\alpha(k-1), \end{aligned}$$

où on a utilisé des estimations similaires à celle utilisée pour majorer J_2.

Mais alors, les deux noyaux de mesures qui apparaissent dans $\bar{d_N}$ sont des lois conditionnelles de $\xi_{[(k-1)N,kN[}$ sachant

$$\tilde{\pi}_S \vee \xi_{[0,(k-1)N[} \text{ et } \tilde{\pi}_{[(k-1)N,kN[}$$

respectivement sur l'espace de probabilité $(X, \mu_{|T^{-(k-1)N}F})$. Pour appliquer le Lemme 2.6, il nous faut l'estimation :

$$\begin{aligned} H(\xi_{[(k-1)N,kN[} \,|\, \tilde{\pi}_S \vee \xi_{[0,(k-1)N[}\,;\, T^{-(k-1)N}F) & \\ \geq H(\xi_{[(k-1)N,kN[} \,|\, \tilde{\pi} \vee \xi_{]-\infty,(k-1)N[}\,;\, T^{-(k-1)N}F) & \\ = H(\xi_{[0,N[} \,|\, \tilde{\pi} \vee \xi_{]-\infty,0[}\,;\, F) & \\ \geq (h(\xi, \mu, T \,|\, \tilde{\pi}) - 3\beta \log(\#A)) \cdot N & \\ \geq (H(\xi_{[0,N[} \,|\, \tilde{\pi}_{[0,N[}\,;\, F) - \kappa r - 3\beta \log(\#A)) \cdot N & \\ \geq (H(\xi_{[(k-1)N,kN[} \,|\, \tilde{\pi}_{[(k-1)N,kN[}\,;\, T^{-(k-1)N}) - 2\kappa r) \cdot N, & \end{aligned}$$

pour laquelle on a utilisé le Lemme 2.13 et (c) et (d)(ii) de la Proposition 2.14. Le Lemme 2.6 donne alors :

$$J_1 \leq 3r + 6\alpha(k-1).$$

En utilisant les majorations de J_1, J_2 et (14), on a

$$I_{k,S} \leq 3r + \frac{k-1}{k}(8\alpha(k-1) + 8\alpha) \leq 3r + 8\alpha k.$$

□

En utilisant les résultats de la Section 2.3.2, on peut alors montrer que :

Proposition 2.17. *Soient (X, μ, T) un système ergodique et ξ et π deux processus sur X. Pour tous $\varepsilon > 0$, $r > 0$, il existe un processus φ sur (X, μ, T), $t \geq 1$ et $\kappa > 0$ tels que*

(i) $h(\varphi, \mu, T) < \varepsilon$,

(ii) $\xi_{[0,t[}$ *est relativement* $(\kappa t, r)$*-extrémale par rapport à* $(\pi \vee \varphi)_{[0,t[}$,

(iii) et $H(\xi_{[0,t[} \,|\, (\pi \vee \varphi)_{[0,t[}) < (h(\xi, \mu, T \,|\, \pi \vee \varphi) + \kappa r) \cdot t$.

Démonstration. Posons $r' = r/72$, et on considère le processus φ donné par la Proposition 2.14 pour les paramètres d'erreur ε et r'. On pose de nouveau $\tilde{\pi} = \pi \vee \varphi$. Par la condition (d)(i) de la Proposition 2.14, il existe un entier k tel que

$$\text{(A) } k \geq \max\left\{\frac{2(H(\pi_0) + \log(2))}{\kappa r'}, \frac{1}{r'}\right\} \text{ et (B) } \alpha k \leq \min\{1, \kappa\} r'.$$

On pose alors $t = (k + 1)N$.

Par les lemmes [1, Lemma 9.7, Lemma 9.8] énoncés par Austin, l'estimation de la Proposition 2.16 et la condition (c) de la Proposition 2.14 implique que $\xi_{[0,kN[}$ est relativement $(\kappa kN, 7r' + 16\alpha k)$-extrémale par rapport à $\tilde{\pi}_{[0,kN[}$ sur l'espace de probabilité $(X, \mu_{|F})$. Mais alors, par T-invariance de μ et (B), pour $j \in \{0, ..., N - 1\}$, $\xi_{[j,kN+j[}$ est relativement $(\kappa kN, 23r')$-extrémale par rapport à $\tilde{\pi}_{[j,kN+j[}$ sur l'espace de probabilité $(X, \mu_{|T^{-j}F})$. Fixons $j \in \{0, ..., N - 1\}$.

Ensuite, on va appliquer le Corollaire 2.6.1. Pour cela, on a l'estimation :

$$H(\xi_{[j,kN+j[} \,|\, \pi_{[j,kN+j[}) - H(\xi_{[j,kN+j[} \,|\, \pi_{[0,t[}) \leq H(\tilde{\pi}_{[0,j[} \vee \tilde{\pi}_{[kN+j,t[}) \leq NH(\tilde{\pi}_0).$$

Par le Corollaire 2.6.1, on voit que $\xi_{[j,kN+j[}$ est relativement $(\kappa kN, \frac{2NH(\tilde{\pi}_0)}{\kappa kN} + 69r')$-extrémale par rapport à $\tilde{\pi}_{[0,t[}$. Par (A), ces paramètres sont plus forts que

$$(\kappa kN, 70r') \,.$$

Enfin, on peut utiliser le Lemme 2.7 pour voir que $\xi_{[0,t[}$ est relativement $(\frac{t \cdot \kappa kN}{kN} = \kappa t, 69r' + \frac{1}{k+1})$-extrémale par rapport à $\tilde{\pi}_{[0,t[}$.

Donc, par (A), on a montré que, pour tout $j \in \{0, ..., N - 1\}$, $\xi_{[0,t[}$ est relativement $(\kappa t, 71r')$-extrémale par rapport à $\tilde{\pi}_{[0,t[}$ sur l'espace de probabilité $(X, \mu_{|T^{-j}F})$.

Mais cela implique que $\xi_{[0,t[}$ est relativement $(\kappa t, 71r' + \beta)$-extrémale par rapport à $\tilde{\pi}_{[0,t[} \vee \tau_0$ sur l'espace de probabilité (X, μ), où τ_0 est la v.a. engendrée par F

(voir (12) et (13)). Mais τ_0 est φ_0-mesurable, donc elle est $\tilde{\pi}_{[0,t[}$-mesurable, donc $\tilde{\pi}_{[0,t[}$ et $\tilde{\pi}_{[0,t[} \vee \tau_0$ engendrent la même tribu.

Finalement, comme $\beta \leq r'$, $\xi_{[0,t[}$ est relativement $(\kappa t, 72r' = r)$-extrémale par rapport à $\tilde{\pi}_{[0,t[}$ sur l'espace de probabilité (X, μ).

Il nous reste à montrer (iii). Pour cela, on commence par poser $G = \bigcup_{j=0}^{N-1} T^{-j}F$, $\rho = \xi \vee \pi$ et par estimer

$$\begin{aligned}
H(\varphi_{[0,t[} \,|\, \rho_{[0,t[}) &\leq H(\varphi_{[0,t[} \,|\, \rho_{[0,t[}\,;\, G) + 3\beta \log(2) t \\
&\leq \frac{1}{N} \sum_{j=0}^{N-1} H(\varphi_{[0,t[} \,|\, \rho_{[0,t[}\,;\, T^{-j}F) + \kappa r' t \\
&= \frac{1}{N} \sum_{j=0}^{N-1} H(\varphi_{[-j,t-j[} \,|\, \rho_{[-j,t-j[}\,;\, F) + \kappa r' t \\
&\leq H(\varphi_{[0,kN[} \,|\, \rho_{[0,kN[}\,;\, F) + \log(2) + \kappa r' t \\
&\leq \sum_{i=0}^{k-1} H(\varphi_{[i,(i+1)N[} \,|\, \rho_{[i,(i+1)N[}\,;\, F) + \log(2) + \kappa r' t \\
&= \sum_{i=0}^{k-1} H(\varphi_{[0,N[} \,|\, \rho_{[0,N[}\,;\, T^{iN}F) + \log(2) + \kappa r' t \\
&\leq \sum_{i=0}^{k-1} H(\varphi_{[0,N[} \,|\, \rho_{[0,N[}\,;\, F) + \frac{1}{\mu(F)} N \log(2) \mu(F \Delta T^{iN} F) + \log(2) + \kappa r' t \\
&\leq N \log(2) \alpha k^2 + \log(2) + \kappa r' t \leq (\log(2)\alpha k + \log(2)/t + \kappa r') \cdot t,
\end{aligned}$$

car, par construction, on a

$$H(\varphi_{[0,N[} \,|\, \xi_{[0,N[} \vee \pi_{[0,N[}\,;\, F) = 0.$$

L'avant dernière ligne est obtenue par le Lemme 2.12. De plus, par (A) et (B), on a

$$\log(2)\alpha k + \log(2)/t \leq \log(2)\kappa r' + \kappa r'/2 \leq \kappa r'.$$

Donc

$$H(\varphi_{[0,N[} \,|\, \xi_{[0,N[} \vee \pi_{[0,N[}) \leq 2\kappa r' t.$$

On conclue alors

$$\begin{aligned}
H(\xi_{[0,t[}\,|\,(\pi\vee\varphi)_{[0,t[}) &= H((\xi\vee\varphi)_{[0,t[}\,|\,\pi_{[0,t[}) - H(\varphi_{[0,t[}\,|\,\pi_{[0,t[})\\
&= H(\xi_{[0,t[}\,|\,\pi_{[0,t[}) + H(\varphi_{[0,t[}\,|\,(\xi\vee\pi)_{[0,t[}) - H(\varphi_{[0,t[}\,|\,\pi_{[0,t[})\\
&\leq H(\xi_{[0,t[}\,|\,\pi_{[0,t[}) - H(\varphi_{[0,t[}\,|\,\pi_{[0,t[}) + 2\kappa r' t\\
&< (h(\xi,\mu,T\,|\,\pi) - h(\varphi,\mu,T\,|\,\pi) + 3\kappa r')\cdot t,
\end{aligned}$$

la dernière inégalité s'obtient par (P3.1) et par le fait que t est un multiple de N. On obtient alors (iii)

$$\begin{aligned}
H(\xi_{[0,t[}\,|\,(\pi\vee\varphi)_{[0,t[}) &< (h(\xi\vee\varphi,\mu,T\,|\,\pi) - h(\varphi,\mu,T\,|\,\pi) + 3\kappa r')\cdot t\\
&= (h(\xi,\mu,T\,|\,\pi\vee\varphi) + \kappa r)\cdot t
\end{aligned}$$

par la "chain rule" de la KS-entropie. □

2.4.3 Preuve du Théorème d'Austin via la Proposition 2.17

Pour appliquer le Proposition 2.17, on a besoin d'un résultat sur la finie détermination relative :

Lemme 2.18. *Si ξ, π et π' sont trois processus sur (X,μ,T) tels que ξ est relativement finiment déterminé par rapport à π et π et π' engendrent le même facteur* $\mathrm{mod}\ \mu$, *alors ξ est relativement finiment déterminé par rapport à π'.*

La preuve de ce résultat se trouve aussi dans l'article d'Austin (voir [1, Lemma 11.6]). On peut maintenant prouver le Théorème d'Austin.

Soit $(\xi^{(k)})_{k\geq 0}$ une suite de processus qui engendre la tribu $\mathcal{F}$ du système (X,μ,T). Quitte à considérer $(\xi^{(0)}\vee\cdots\vee\xi^{(k)})_{k\geq 0}$, on peut supposer que $(\xi^{(k)})_{k\geq 0}$ est croissante, c'est-à-dire que $\xi_0^{(k)} \preccurlyeq \xi_0^{(k+1)}$. On fait aussi en sorte que $\xi^{(0)}$ soit le processus trivial. On fera remarquer au lecteur qu'il est possible de choisir une suite génératrice car nous travaillons sur des espaces de Lebesgue.

On va vouloir appliquer la Proposition 2.17 une infinité de fois à chaque processus de la suite $(\xi^{(k)})_{k\geq 0}$. Pour cette raison, on choisit une suite de nombres entiers $(k_j)_{j\geq 0}$ qui passe par tous les éléments de $\mathbb{N}$ une infinité de fois. Pour construire une telle suite, on peut, par exemple, concaténer les mots $\{0\}$, $\{0,1\}$, $\{0,1,2\}$ *etc* ... . Une fois cette suite construite, on utilise la Proposition 2.17 pour construire par récurrence une suite de processus $(\varphi^{(j)})_{j\geq 0}$ et une suite $(\kappa_j)_{j\geq 0}$ telles que

(i) $h(\varphi^{(j)}, \mu, T) < 2^{-(j+1)} \min\{\varepsilon, \kappa_0, ..., \kappa_{j-1}\}$,

(ii) $\xi^{(k_j)}_{[0,t_j[}$ est relativement $(\kappa_j t_j, 2^{-(j+1)})$-extrémale par rapport à

$$\psi^{(j)}_{[0,t_j[} := (\pi \vee \xi^{(k_j-1)} \vee \varphi^{(0)} \vee \cdots \vee \varphi^{(j)})_{[0,t_j[},$$

(iii) et $H(\xi^{(k_j)}_{[0,t_j[} \,|\, \psi^{(j)}_{[0,t_j[}) < (h(\xi^{(k_j)}, \mu, T \,|\, \psi^{(j)}) + \kappa_j 2^{-(j+1)}) \cdot t_j$.

Lorsque $k_j = 0$, on se contente de prendre $\kappa_j = 1$ et un processus $\varphi^{(j)}$ satisfaisant seulement (i).

Démonstration du Théorème d'Austin. Il nous reste à montrer que (X, μ, T) est relativement Bernoulli par rapport à $\pi \vee \bigvee_{j=0}^{\infty} \varphi^{(j)}$ et à estimer l'entropie de $\bigvee_{j=0}^{\infty} \varphi^{(j)}$. Tout d'abord, par (i), on voit que

$$\begin{aligned} h(\pi \vee \bigvee_{j=0}^{\infty} \varphi^{(j)}, \mu, T \,|\, \pi) &\leq h(\pi, \mu, T \,|\, \pi) + \sum_{j=0}^{\infty} h(\varphi^{(j)}, \mu, T \,|\, \pi) \\ &\leq \sum_{j=0}^{\infty} h(\varphi^{(j)}, \mu, T) \leq \varepsilon. \end{aligned} \tag{15}$$

Donc on a $h(\pi \vee \bigvee_{j=0}^{\infty} \varphi^{(j)}, \mu, T) \leq h(\pi, \mu, T) + \varepsilon < \infty$, donc, par le Théorème de Krieger, il existe un processus $\tilde{\pi}$ sur un alphabet fini qui engendre le facteur $\pi \vee \bigvee_{j=0}^{\infty} \varphi^{(j)}$. De plus, par l'estimation (15), on a

$$h(\tilde{\pi}, \mu, T \,|\, \pi) \leq \varepsilon. \tag{16}$$

Soit $k \in \mathbb{N}$. Soit $j \geq 0$ tel que $k_j = k$. Par (ii), on sait que $\xi^{(k)}_{[0,t_j[}$ est $(\kappa_j t_j, 2^{-(j+1)})$-extrémale par rapport à $\psi^{(j)}_{[0,t_j[}$. Or, par la "chain rule" de la KS-entropie, on a

$$\begin{aligned} h(\xi^{(k)}, \mu, T \,|\, \psi^{(j)}) - h(\xi^{(k)}, \mu, T \,|\, \psi^{(j)} \vee \tilde{\pi}) = h(\xi^{(k)}, \mu, T \,|\, \psi^{(j)}) \\ - h(\xi^{(k)} \vee \tilde{\pi}, \mu, T \,|\, \psi^{(j)}) + h(\tilde{\pi}, \mu, T \,|\, \psi^{(j)}) \\ \leq h(\tilde{\pi}, \mu, T \,|\, \psi^{(j)}). \end{aligned}$$

De plus, on peut obtenir une estimation similaire à (15), en utilisant (i) de nouveau :

$$\begin{aligned} h(\tilde{\pi}, \mu, T \,|\, \psi^{(j)}) &= h(\bigvee_{i=j+1}^{\infty} \varphi^{(i)}, \mu, T \,|\, \psi^{(j)}) \\ &\leq \sum_{i=j+1}^{\infty} h(\varphi^{(i)}, \mu, T) \leq \kappa_j 2^{-(j+1)}. \end{aligned}$$

Enfin, par (iii), on a

$$\begin{aligned}
&H(\xi^{(k)}_{[0,t_j[}\,|\,\psi^{(j)}_{[0,t_j[}) - H(\xi^{(k)}_{[0,t_j[}\,|\,\psi^{(j)}_{[0,t_j[} \vee \tilde{\pi}_{[0,t_j[}) \\
&\qquad\qquad\leq (h(\xi^{(k)}, \mu, T\,|\,\psi^{(j)}) - h(\xi^{(k)}, \mu, T\,|\,\psi^{(j)} \vee \tilde{\pi}) + \kappa_j 2^{-(j+1)}) \cdot t_j \\
&\qquad\qquad\leq \kappa_j 2^{-j} t_j
\end{aligned}$$

Donc, par le Corollaire 2.6.1, $\xi^{(k)}_{[0,t_j[}$ est relativement $(\kappa_j t_j, 7 \cdot 2^{-(j+1)})$-extrémale par rapport à $(\tilde{\pi} \vee \psi^{(j)})_{[0,t_j[}$. De plus, on peut obtenir une version modifiée de (iii) :

$$\begin{aligned}
H(\xi^{(k)}_{[0,t_j[}\,|\,\psi^{(j)}_{[0,t_j[} \vee \tilde{\pi}_{[0,t_j[}) &\leq H(\xi^{(k)}_{[0,t_j[}\,|\,\psi^{(j)}_{[0,t_j[}) \\
&< (h(\xi^{(k)}, \mu, T\,|\,\psi^{(j)}) + \kappa_j 2^{-(j+1)}) \cdot t_j \\
&\leq (h(\xi^{(k)}, \mu, T\,|\,\tilde{\pi} \vee \psi^{(j)}) + 2 \cdot \kappa_j 2^{-(j+1)}) \cdot t_j \\
&< (h(\xi^{(k)}, \mu, T\,|\,\tilde{\pi} \vee \psi^{(j)}) + 7 \cdot \kappa_j 2^{-(j+1)}) \cdot t_j
\end{aligned}$$

Mais alors, par la Remarque 2.1, $\xi^{(k)}$ est relativement $(49 \cdot 2^{-(j+1)})$-finiment déterminé par rapport à $\tilde{\pi} \vee \psi^{(j)}$. Et, par le Lemme 2.18, $\xi^{(k)}$ est relativement $(49 \cdot 2^{-(j+1)})$-finiment déterminé par rapport à $\tilde{\pi} \vee \xi^{(k-1)}$. Or, on a choisit la suite $(k_j)_{j\geq 0}$ de sorte que l'on puisse prendre j arbitrairement grand, donc $\xi^{(k)}$ est *relativement finiment déterminé par rapport* à $\tilde{\pi} \vee \xi^{(k-1)}$.

On fait alors appel au résultat du Chapitre 1 ([Chapitre 1, Théorème 3.2]) pour dire qu'il existe un schéma de Bernoulli $\zeta^{(k)}$ indépendant de $\tilde{\pi} \vee \xi^{(k-1)}$ et tel que

$$\tilde{\pi} \vee \xi^{(k-1)} \vee \zeta^{(k)} \ \text{ et } \ \tilde{\pi} \vee \xi^{(k-1)} \vee \xi^{(k)} = \tilde{\pi} \vee \xi^{(k)}$$

engendrent le même facteur. Mais alors, par récurrence, on voit que

$$\tilde{\pi} \vee \bigvee_{k=0}^{\infty} \zeta^{(k)} \ \text{ et } \ \tilde{\pi} \vee \bigvee_{k=0}^{\infty} \xi^{(k)}$$

engendrent le même facteur. De plus, on avait choisit les processus $\xi^{(k)}$ de sorte qu'ils engendrent le tribu entière, donc $\tilde{\pi} \vee \bigvee_{k=0}^{\infty} \zeta^{(k)}$ est générateur. Enfin, $\bigvee_{k=0}^{\infty} \zeta^{(k)}$ est un schéma de Bernoulli indépendant de $\tilde{\pi}$. Ainsi, (X, μ, T) est relativement Bernoulli par rapport à $\tilde{\pi}$. Cela combiné avec l'estimation (16) prouve le Théorème d'Austin. □

Le schéma de Bernoulli $\bigvee_{k=0}^{\infty} \zeta^{(k)}$ que l'on a construit, contrairement à la majorité de ceux que l'on a manipulé dans ce mémoire, est à valeurs sur un alphabet infini, mais il vérifie bien la définition des schémas de Bernoulli que l'on avait donnée dans le Chapitre 1.

3 Conséquences du théorème d'Austin

Parmi les différents résultats que l'on peut obtenir à partir de la propriété de Pinsker faible (voir [1, Section 16]), on se contentera ici d'en présenter un seul en détails. Montrons le théorème suivant, dont la preuve est due à Thouvenot (voir [13]).

Théorème 3.1. *Soient (X, μ, T) et (Y, ν, S) deux systèmes ergodiques et $\mathcal{B}$ un schéma de Bernoulli tels que $(X, \mu, T) \times \mathcal{B}$ et $(Y, \nu, S) \times \mathcal{B}$ sont isomorphes. Alors (X, μ, T) et (Y, ν, S) sont isomorphes.*

Nous avons choisi ce résultat car il permet de nous éclairer sur le résultat donné par la propriété de Pinsker faible. En effet, celle-ci promet l'existence d'un facteur d'entropie arbitrairement faible par rapport auquel le système initial est relativement Bernoulli, mais ce facteur n'a rien de canonique. On peut seulement dire qu'il contient le facteur de Pinsker du système. Le théorème ci-dessus nous donne tout de même une information supplémentaire : si $\mathbf{X}_\varepsilon$ et $\mathbf{Y}_\varepsilon$ sont deux facteurs d'entropie ε donnés par la propriété de Pinsker faible, ils sont isomorphes. Cependant, il est important de noter que, bien que ces deux systèmes soient isomorphes, les deux applications facteurs qui leur sont associées ne sont pas nécessairement les mêmes, et ces deux facteurs n'engendrent donc pas nécessairement la même sous-tribu sur (X, μ, T).

Passons maintenant à la preuve du Théorème 3.1. On aura besoin du résultat suivant, qui correspond à [14, Proposition 4], et qui s'obtient à partir des résultats de [Chapitre 1, Sous-section 3.5] :

Théorème 3.2. *Soient (Y, ν, S) un système ergodique d'entropie finie et π un processus tels que (Y, ν, S) est relativement Bernoulli par rapport à π.*

Tout processus ζ sur Y tel que ζ_0 prenne un nombre fini de valeurs est relativement finiment déterminé par rapport à π.

Démonstration. Soit η un schéma de Bernoulli sur Y tel que $\pi \vee \eta$ soit générateur. Comme (Y, ν, S) est d'entropie finie, on peut supposer que η est à valeurs dans alphabet fini. En effet, il suffit de prendre un schéma de Bernoulli sur un alphabet fini de même entropie que η et d'utiliser le fait que deux schémas de Bernoulli de même entropie sont isomorphes.

Soient ζ sur Y à valeurs dans l'alphabet $[k]$ et $\varepsilon > 0$. On se donne $\delta > 0$ et $n_1 \geq 1$ à fixer plus tard. Soit (X, μ, T) un système ergodique tel que $h(T) \geq h(\zeta \vee \pi, S)$, munit de deux processus ξ et π' tels que $(T, \pi') \sim (S, \pi)$ et vérifiant :

(i) $|d((\pi' \vee \xi)_{[0,n_1[}) - d((\pi \vee \zeta)_{[0,n_1[})| \leq \delta$,

(ii) $0 \leq h(\pi' \vee \xi, T) - h(\pi \vee \zeta, S) \leq \delta$.

Remarque 3.1. Dans la définition de la finie détermination, on mettait la condition (ii)$'$: $|h(\pi' \vee \xi, T) - h(\pi \vee \zeta, S)| \leq \delta$, mais effectuer cette preuve avec la condition (ii) plus forte donnée ci-dessus est suffisant pour appliquer les résultats montrés dans le Chapitre 1.

Nous allons montrer que, si δ est assez petit et n_1 est assez grand, on a

$$\bar{d}\big((T, \xi), (S, \zeta) \,|\, \pi', \pi\big) \leq \varepsilon. \tag{17}$$

On se donne un alphabet fini A et une mesure γ_1 dessus telle que $H(\gamma_1) = h(S) - h(\pi \vee \zeta, S)$. En utilisant le lemme fondamental d'Ornstein relatif prouvé dans le premier chapitre ([Chapitre 1, Lemme 3.12]) ainsi que [Chapitre 1, Lemme 3.13], on construit un schéma de Bernoulli ρ sur Y à valeurs dans A, indépendant de $\pi \vee \zeta$ et tel que $d(\rho_0) = \gamma_1$. On a donc

$$h(\pi \vee \zeta \vee \rho, S) = h(\pi \vee \zeta, S) + H(\rho_0) = h(S).$$

Si $h(\pi' \vee \xi, T) < h(\pi \vee \zeta, S)$, on se donne un nouvel alphabet fini B et γ_2 telle que $H(\gamma_2) = h(\pi \vee \zeta, S) - h(\pi' \vee \xi, T)$. On travaillera alors avec le système

$$(Z, \lambda, R) = (X \times A^{\mathbb{Z}} \times B^{\mathbb{Z}}, \mu \otimes \gamma_1^{\otimes \mathbb{Z}} \otimes \gamma_2^{\otimes \mathbb{Z}}, T \times T_A \times T_B),$$

et on notera ρ' la projection sur la deuxième coordonnée (clairement, on a $(S, \rho) \sim (R, \rho')$). En ajoutant B dans la construction de Z, on s'est assuré que $h(R) = h(S)$.

Notre but va être d'utiliser le lemme fondamental relatif affin de copier η sur Z. On se donne $\alpha > 0$, $\beta > 0$ et $\gamma > 0$ que l'on fixera plus tard. On se donne une première estimation : il existe $K \geq 1$ tel que

$$\zeta_0 \preccurlyeq_\gamma (\pi \vee \eta)_{[-K,K]}. \tag{18}$$

On peut aussi prendre une deuxième estimation plus précise : il existe $N_1 \geq K$ assez grand pour avoir

$$\zeta_0 \preccurlyeq_\alpha (\pi \vee \eta)_{[-N_1,N_1]}. \tag{19}$$

Par le théorème de Rohlin ([Chapitre 1, Théorème 3.4]), il existe des bases de tours de Rohlin $F' \subset Z$ et $E' \subset Y$ telles que $\lambda(\bigcup_{j=0}^{n_1-1} R^j F'), \nu(\bigcup_{j=0}^{n_1-1} S^j E') \geq 1-\beta$ et

$$d((\pi' \vee \xi \vee \rho')_{[0,n_1[}\big|_{F'}) = d((\pi' \vee \xi \vee \rho')_{[0,n_1[}), \tag{20}$$

et

$$d((\pi \vee \zeta \vee \rho)_{[0,n_1[}\big|_{E'}) = d((\pi \vee \zeta \vee \rho)_{[0,n_1[}). \tag{21}$$

On remarque que, comme ρ et ρ' sont de même loi et indépendants de $\pi \vee \zeta$ et $\pi' \vee \xi$ respectivement, on a

$$|d((\pi \vee \zeta \vee \rho)_{[0,n_1[}) - d((\pi' \vee \xi \vee \rho')_{[0,n_1[})| = |d((\pi \vee \zeta)_{[0,n_1[}) - d((\pi' \vee \xi \vee)_{[0,n_1[})|.$$

En utilisant cette remarque ainsi que (20), (21) et (i), si δ est assez petit, on peut trouver $F \subset F'$ et $E \subset E'$ tels que $\lambda(\bigcup_{j=0}^{n_1-1} R^j F'), \nu(\bigcup_{j=0}^{n_1-1} S^j E') \geq 1-2\beta$ et

$$d((\pi' \vee \xi \vee \rho')_{[0,n_1[}\big|_{F}) = d((\pi \vee \zeta \vee \rho)_{[0,n_1[}\big|_{E}).$$

Donc les gadgets $(S, E, n_1, \pi_0 \vee \zeta_0 \vee \rho_0)$ et $(R, F, n_1, \pi'_0 \vee \xi_0 \vee \rho'_0)$ sont isomorphes. Par [Chapitre 1, Lemme 3.9], il existe η^* sur Z tel qu'on ait l'isomorphisme :

$$(S, E, n_1, \pi_0 \vee \zeta_0 \vee \rho_0 \vee \eta_0) \cong (R, F, n_1, \pi'_0 \vee \xi_0 \vee \rho'_0 \vee \eta^*_0). \tag{22}$$

Soit $\varepsilon' > 0$ et on prend $\delta' > 0$ et $n'_1 \geq 1$ donnés par le lemme fondamental appliqué à π et η. On commence par voir que, pour β assez petit et n_1 assez grand, l'isomorphisme de gadgets ci-dessus implique

$$|d((\pi' \vee \eta^*)_{[0,n'_1[}) - d((\pi \vee \eta)_{[0,n'_1[})| \leq \delta'. \tag{23}$$

Prenons $N_2 \geq N_1$ tel que

$$\pi_0 \vee \zeta_0 \vee \rho_0 \preccurlyeq_\beta (\pi \vee \eta)_{[-N_2,N_2]}.$$

Alors, par le même argument que pour [Chapitre 1, (36)], on voit que cette estimation et (22) implique

$$\pi'_0 \vee \xi_0 \vee \rho'_0 \preccurlyeq_{5\beta} (\pi' \vee \eta^*)_{[-N_2,N_2]},$$

si $\frac{2N_2}{n_1} \leq \beta$. Alors, pour β assez petit, en notant $(\pi' \vee \eta^*)^{(N_2)}$ le processus vérifiant $(\pi' \vee \eta^*)_0^{(N_2)} = (\pi' \vee \eta^*)_{[-N_2,N_2]}$, on a

$$\begin{aligned} h(\pi' \vee \eta^*, R) = h((\pi' \vee \eta^*)^{(N_2)}, R) &\geq h(\pi' \vee \xi \vee \rho', R) - \alpha \\ &\geq h(\pi \vee \zeta \vee \rho) - \alpha - \delta \\ &= h(\pi \vee \eta, S). \end{aligned}$$

Comme on a aussi

$$h(\pi' \vee \eta^*, R) \leq h(R) = h(S) = h(\pi \vee \eta, S),$$

on a, pour α puis δ assez petits :

$$0 \leq h(\pi \vee \eta, S) - h(\pi' \vee \eta^*, R) \leq \delta'. \tag{24}$$

Les estimations (23) et (24) nous permettent alors d'appliquer le lemme fondamental : il existe η' sur Z tel que

$$(a)\ (R, \pi' \vee \eta') \sim (S, \pi \vee \eta) \ \text{ et } \ (b)\ |\eta_0^* - \eta_0'| \leq \varepsilon'.$$

Comme $\pi \vee \eta$ est générateur, (Y, ν, S) est alors un facteur de (Z, λ, R), et l'application facteur associée envoie $\pi' \vee \eta'$ sur $\pi \vee \eta$. Posons ζ' le relevé sur (Z, λ, R) de ζ. Alors $(S, \pi \vee \zeta) \sim (R, \pi' \vee \zeta')$ et il reste à voir que ζ_0' est proche de ξ_0 : grâce à (18), on peut trouver une fonction Π sur $([r] \times [l])^{[-K,K]}$ telle que

$$|\zeta_0 - \Pi(\pi \vee \eta)| \leq \gamma. \tag{25}$$

On peut transporter cette estimation via l'application facteur construite ci-dessus :

$$|\zeta_0' - \Pi(\pi' \vee \eta')| \leq \gamma.$$

On peut aussi utiliser une nouvelle fois l'argument utilisé pour obtenir [Chapitre 1, (36)], et (25) devient alors

$$|\xi_0 - \Pi(\pi' \vee \eta^*)| \leq 2\gamma + 3\beta,$$

car $N_2 \geq K$ et donc $\frac{2K}{n_1} \leq \beta$. Alors, par (b), pour ε' assez petit

$$|\xi_0 - \Pi(\pi' \vee \eta')| \leq 3\gamma + 3\beta,$$

(pour cela on utilise le fait que γ et K ne dépendent pas de ε'). Enfin, on a

$$|\xi_0 - \zeta_0'| \leq 4\gamma + 3\beta.$$

On choisit alors γ puis β tels que $4\gamma + 3\beta \leq \varepsilon$, et la preuve est complète. □

Remarque 3.2. Il est important de voir dans la preuve ci-dessus que les paramètres δ et n_1 ne dépendent que des processus portés par Y (π, η et ζ) et de ε. On peut aussi remarquer que le paramètre n_1' vaut 1, mais cela simplifie très peu la preuve.

Cela permet de prouver le Théorème 3.1, en utilisant la propriété de Pinsker faible :

Démonstration du Théorème 3.1. Tout d'abord, on voit que, grâce à la conjecture de Pinsker faible, on peut supposer T et S sont d'entropie finie.

Pour toute application facteur φ, on notera $\mathbf{X}_\varphi$ le système engendré par φ.

On commence par remarquer que l'isomorphisme implique que $h(T) = h(S)$, et note $a = h(T) = h(S)$. Ensuite, on applique la propriété de Pinsker faible à T et S pour construire des applications facteurs π_1, ψ_1 sur X et π_2, ψ_2 sur Y telles que

(i) $(X, \mu, T) \cong \mathbf{X}_{\pi_1} \times \mathbf{X}_{\psi_1}$ et $(Y, \nu, S) \cong \mathbf{X}_{\pi_2} \times \mathbf{X}_{\psi_2}$,

(ii) $\mathbf{X}_{\psi_1}$ et $\mathbf{X}_{\psi_2}$ sont des schémas de Bernoulli,

(iii) et $h(\pi_1, T), h(\pi_2, S) \leq \frac{a}{3}$.

L'isomorphisme permet aussi de construire un système (Z, λ, R) sur lequel on a quatre applications facteurs φ_1, φ_2, χ_1 et χ_2 teles que

(a) $\mathbf{X}_{\varphi_1} = \mathbf{X}_{\varphi_2} = \mathcal{B}$,

(b) $\mathbf{X}_{\chi_1} = (X, \mu, T)$ et $\mathbf{X}_{\chi_2} = (Y, \nu, S)$,

(c) et $(Z, \lambda, R) \cong \mathbf{X}_{\varphi_1} \times \mathbf{X}_{\chi_1} \cong \mathbf{X}_{\varphi_2} \times \mathbf{X}_{\chi_2}$.

Mais on peut relever π_1, π_2, ψ_1 et ψ_2 en des applications sur Z, et alors, en combinant (i), (iii), (a) et (c), on voit que (Z, λ, R) est relativement Bernoulli par rapport à π_1, et aussi par rapport à π_2. De plus, π_1 et π_2 sont d'entropie finie, et donc par le théorème de Krieger, on peut supposer que ce sont des processus sur des alphabets finis. Le Théorème 3.2 nous dit alors que π_2 est relativement finiment déterminé par rapport à π_1, et inversement, que π_1 est relativement finiment déterminé par rapport à π_2. De plus, il est facile de voir que c'est encore vrai sur $\mathbf{X}_{\pi_1 \vee \pi_2}$. Il existe alors deux nouveaux schémas de Bernoulli $\mathcal{B}_1$ et $\mathcal{B}_2$, tels que

$$\mathbf{X}_{\pi_1 \vee \pi_2} \cong \mathbf{X}_{\pi_1} \times \mathcal{B}_1 \text{ et } \mathbf{X}_{\pi_1 \vee \pi_2} \cong \mathbf{X}_{\pi_2} \times \mathcal{B}_2.$$

Prenons $\mathcal{B}'$ un schéma de Bernoulli d'entropie $a - h(\pi_1 \vee \pi_2, R)$. La construction du paragraphe précédent implique que

$$\mathbf{X}_{\pi_1} \times \mathcal{B}_1 \times \mathcal{B}' \cong \mathbf{X}_{\pi_2} \times \mathcal{B}_2 \times \mathcal{B}'.$$

Mais comme deux schémas de Bernoulli de même entropie sont isomorphes, on voit que

$$\mathcal{B}_1 \times \mathcal{B}' \cong \mathbf{X}_{\psi_1} \text{ et } \mathcal{B}_2 \times \mathcal{B}' \cong \mathbf{X}_{\psi_2}.$$

On en conclut

$$(X, \mu, T) \cong \mathbf{X}_{\pi_1} \times \mathbf{X}_{\psi_1} \cong \mathbf{X}_{\pi_2} \times \mathbf{X}_{\psi_2} \cong (Y, \nu, S).$$

□

4 Annexe

Cette section a pour but de compléter l'annexe du Chapitre 1. Ce découpage de l'annexe en deux parties n'est pas idéal, et je m'en excuse.

4.1 Espaces de Lebesgue

Tous les espaces de probabilité considérés dans ce mémoire sont des espaces de Lebesgue, donc il me semble utile de les définir précisément. Dans l'annexe du Chapitre 1, il y avait déjà une partie intitulée "Espaces de Lebesgue", mais elle ne contenait qu'un résultat élémentaire ([Chapitre 1, Lemme 4.1]) sur la mesure de Lebesgue, pas très intéressant à démontrer, mais qui est bien utile dans le manipulation de gadgets.

L'espace probabilisable fondamental que l'on considère est le segment $[0, 1]$ munit de la tribu de Lebesgue Σ. On rappelle que la tribu de Lebesgue est la complétion de la tribu Borélienne pour la mesure de Lebesgue. De plus, on notera $\mathcal{L}$ la mesure de Lebesgue sur cet espace.

On dit que deux espaces de probabilité (X, μ) et (Y, ν) sont isomorphes s'il existe $X_0 \subset X$ et $Y_0 \subset Y$ de mesure totale et une bijection $\Phi : X_0 \longrightarrow Y_0$ mesurable, d'inverse mesurable et telle que $\Phi_* \mu = \nu$.

Définition 4.1. Un espace de probabilité (X, μ) est un *espace de Lebesgue* s'il existe $0 \leq \alpha \leq 1$, un ensemble $I \subset [0, 1]$ au plus dénombrable et une famille de nombres positifs $(a_i)_{i \in I}$ telle que $\sum_{i \in I} a_i = 1$ tels que (X, μ) est isomorphe à l'espace probabilisable $([0, 1], \Sigma)$ munit de la mesure

$$\alpha \mathcal{L} + (1 - \alpha) \sum_{i \in I} a_i \delta_i,$$

où δ_i est la mesure de Dirac en i.

Cette définition peut sembler trop restrictive, mais elle englobe en fait la très grande majorité des espaces de probabilité utilisés (voir [10, Chapter 1] pour voir d'autres caractérisation équivalentes des espaces de Lebesgue). Cependant une propriété importante est préservée par isomorphisme : la présence d'atomes. On appelle *atome* d'un espace de probabilité tout point dont le singleton est de mesure non-nulle. Dans la définition des espaces de Lebesgue, les atomes sont les éléments de I. On trouve souvent une définition des espaces de Lebesgue pour laquelle on n'autorise pas la présence d'atomes, et on suppose qu'un espace de Lebesgue est isomorphes au segment $[0, 1]$ munit de la mesure de Lebesgue. En particulier, c'est sous ces conditions que l'on peut appliquer [Chapitre 1, Lemme 4.1]. Étant donné que l'on a utilisé ce lemme à de nombreuses reprises dans la partie concernant la théorie d'Ornstein, il est important de vérifier que cela était correct. Cela est dû au résultat suivant :

Proposition 4.1. *Soit (X, μ, T) un système dynamique ergodique. Ou bien (X, μ) ne possède pas d'atomes, ou bien il existe $n \geq 1$ tel que (X, μ, T) est isomorphe à la rotation sur $\mathbb{Z}/_{n\mathbb{Z}}$ munit de la mesure de probabilité uniforme.*

Démonstration. Notons $Atom(X)$ l'ensemble des atomes de X. Comme T préserve la mesure μ, $Atom(X)$ est est ensemble T-invariant, et, comme (X, μ, T) est ergodique, $Atom(X)$ est de mesure 0 ou 1. Supposons que $\mu(Atom(X)) = 1$ et montrons que (X, μ, T) est isomorphe à une rotation. Fixons $x_0 \in Atom(x)$. Comme, pour tout $k \geq 0$, $\mu(\{T^k x_0\}) = \mu\{x_0\}) > 0$, l'orbite de x_0 est finie. De plus, T est une bijection sur $Atom(X)$, donc x_0 est un point périodique et on note n sa période. Enfin, l'orbite de x_0 est un ensemble T-invariant et de mesure non-nulle, donc de mesure 1. Vu que la restriction de T à l'orbite de x_0 est isomorphe à la rotation sur $\mathbb{Z}/_{n\mathbb{Z}}$, la preuve est terminée. □

Ainsi, on a une description complète des systèmes ergodiques qui possèdent des atomes. Par conséquent, vu que tous les systèmes auxquelles on a appliqué [Chapitre 1, Lemme 4.1] étaient ergodiques, on peut se contenter d'étudier ces systèmes dans le cas où ils n'ont pas d'atomes.

Références

[1] Tim Austin. Measure concentration and the weak Pinsker property. *Publ. Math., Inst. Hautes Étud. Sci.*, 128 :1–119, 2018.

[2] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic theory. Transl. from the Russian by A. B. Sossinskii.*, volume 245. Springer, Berlin, 1982.

[3] R. M. Dudley. *Real analysis and probability. Repr.*, volume 74. Cambridge : Cambridge University Press, repr. edition, 2002.

[4] B. Hasselblatt and A. Katok, editors. *Handbook of dynamical systems. Volume 1A.* Amsterdam : North-Holland, 2002.

[5] D. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.*, 4 :337–352, 1970.

[6] Donald Ornstein. Two Bernoulli shifts with infinite entropy are isomorphic. *Adv. Math.*, 5 :339–348, 1971.

[7] Donald S. Ornstein. A K-automorphism with no square root and Pinsker's conjecture. *Adv. Math.*, 10 :89–102, 1973.

[8] Donald S. Ornstein. A mixing transformation for which Pinsker's conjecture fails. *Adv. Math.*, 10 :103–123, 1973.

[9] Donald S. Ornstein. An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Adv. Math.*, 10 :49–62, 1973.

[10] Paul Shields. The theory of Bernoulli shifts. Chicago Lectures in Mathematics. Chicago - London : The University of Chicago Press. X, 118 p. £1.35 ; cloth £3.55 (1973)., 1973.

[11] Ja. G. Sinaĭ. On a weak isomorphism of transformations with invariant measure. *Mat. Sb. (N.S.)*, 63 (105) :23–42, 1964.

[12] J.-P. Thouvenot. On the stability of the weak Pinsker property. *Isr. J. Math.*, 27 :150–162, 1977.

[13] J.-P. Thouvenot. Two facts concerning the transformations which satisfy the weak Pinsker property. *Ergodic Theory Dyn. Syst.*, 28(2) :689–695, 2008.

[14] Jean-Paul Thouvenot. Quelques propriétés des systèmes dynamiques qui se decomposent en un produit de deux systèmes dont l'un est un schema de Bernoulli. *Isr. J. Math.*, 21 :177–207, 1975.