Entropie de Kolmogorov-Sinai

T. de la Rue

Master 2 MFA Rouen 2024-2025

On s'intéresse ici à des processus stationnaires de la forme $(\xi_p)_{p\in\mathbb{Z}}$, où chaque ξ_p ne prend qu'un nombre fini de valeurs. On suppose sans perte de généralité que chaque processus considéré ici est défini sur un système dynamique mesuré (X, μ, T) : ses coordonnées sont toutes données par la relation

$$\forall p \in \mathbb{Z}, \quad \xi_p \; = \; \xi_0 \circ T^p.$$

Le processus est donc complètement déterminé par la donnée de T et de ξ_0 ; la variable ξ_0 est elle-même définie par une partition finie de X dont les atomes sont indexés par l'ensemble dans lequel le processus prend ses valeurs. On représente donc un processus sous la forme d'un couple $(\mathscr{P}, T)$, où $\mathscr{P}$ est une partition de l'espace de Lebesgue sur lequel agit la transformation T. Quelques propriétés élémentaires des partitions finies sur un espace de Lebesgue sont données dans la section 1.

L'objet de ce chapitre est d'introduire la théorie de l'entropie des processus stationnaires et des systèmes dynamiques mesurés développée par Kolmogorov et Sinai à la fin des années 1950. Cette théorie est basée sur l'*entropie de Shannon*, concept essentiel de la théorie de l'information introduit en 1948, qui est une quantité associée à une variable aléatoire discrète (ne prenant qu'un nombre fini de valeurs), et qui mesure la quantité moyenne d'information apportée par la connaissance de cette variable aléatoire. La définition et les propriétés essentielles de l'entropie de Shannon sont données dans la section 2.

En se basant sur l'entropie de Shannon, on définit dans la section 3.1 l'entropie d'un processus stationnaire comme un nombre réel positif qui permet de mesurer l'*imprévisibilité* de ce processus : si on connaît tout le passé $(\xi_p)_{p<0}$, est-il possible de prévoir *à coup sûr* la valeur de ξ_0 ? Si non, combien la connaissance de ce passé apporte-t-elle d'information pour la prévision de ξ_0 ?

Après avoir défini l'entropie d'un processus stationnaire, on peut introduire celle du système dynamique mesuré (X, μ, T) comme le supremum des entropies des processus réalisés dans le système (voir la section 3.2). On obtient ainsi un *invariant* du système, c'est-à-dire une quantité liée au système qui ne change pas lorsqu'on considère un autre système isomorphe. Elle permet donc de distinguer des systèmes dont on pourrait *a priori* imaginer qu'ils soient isomorphes, comme par exemple les décalages de Bernoulli $B(1/2, 1/2)$ et $B(1/3, 1/3, 1/3)$.

La dernière section de ce chapitre est consacrée au facteur de Pinsker d'un système dynamique mesuré (la plus grosse sous-tribu invariante par la dynamique sur laquelle l'entropie est nulle) et à la notion de K-système (système mesuré dont le facteur de Pinsker est trivial).

1 Partitions d'un espace de Lebesgue

On désigne toujours par $\big(X, \mathscr{B}(X), \mu\big)$ un espace de Lebesgue, qui est généralement supposé sans atome (donc isomorphe à l'intervalle $[0, 1]$ muni de la mesure de Lebesgue).

Définition 1.1. *On appelle* partition de X *une application $\mathscr{P}$ mesurable de X dans un ensemble fini I. (On dit alors que $\mathscr{P}$ est* indexée *par I.)*

Si $\mathscr{P} : X \to I$ est une partition (au sens défini ci-dessus), alors $\mathscr{P}$ définit une partition (au sens usuel) de X en un nombre fini de parties mesurables, qui sont les P_i, $i \in I$ donnés par

$$\forall i \in I, \quad P_i \ := \ \mathscr{P}^{-1}(\{i\}).$$

Les P_i sont appelés les *atomes de $\mathscr{P}$*. Notons qu'il n'est pas interdit que certains atomes P_i soient éventuellement vides.

On appelle *distribution* de la partition $\mathscr{P} = \{P_i,\ i \in I\}$ le vecteur $\Big(\mu(P_i)\Big)_{i\in I} \in [0, 1]^I$, que l'on note $\mathrm{dist}\mathscr{P}$. Cette distribution n'est pas autre chose, en fait, que la probabilité sur I image de μ par $\mathscr{P}$.

1.1 Distance entre deux partitions

Si $\mathscr{P} = \{P_i,\ i \in I\}$ et $\mathscr{Q} = \{Q_i,\ i \in I\}$ sont deux partitions indexées par le même ensemble I, on définit la *distance entre* $\mathscr{P}$ *et* $\mathscr{Q}$, notée $|\mathscr{P} - \mathscr{Q}|$, par

$$\begin{aligned} |\mathscr{P} - \mathscr{Q}| &:= \mu\Big(\{x \in X :\ \mathscr{P}(x) \neq \mathscr{Q}(x)\}\Big) \\ &= \sum_{i \in I} \mu(P_i \setminus Q_i) \\ &= \sum_{i \in I} \mu(Q_i \setminus P_i) \\ &= \frac{1}{2} \sum_{i \in I} \mu(P_i \vartriangle Q_i). \end{aligned}$$

On identifie toujours deux partitions $\mathscr{P}$ et $\mathscr{Q}$ telles que $|\mathscr{P} - \mathscr{Q}| = 0$. Alors, l'ensemble d'index I étant fixé, l'ensemble des partitions de X dans I muni de cette distance est un espace métrique complet (exercice...).

1.2 Sup de partitions

Si $\mathscr{P} :\ X \to I$ et $\mathscr{Q} :\ X \to J$ sont deux partitions de X, on note $\mathscr{P} \vee \mathscr{Q}$ la partition de X indexée par $I \times J$ définie par

$$\forall x \in X, \quad \mathscr{P} \vee \mathscr{Q}(x) \ := \ \Big(\mathscr{P}(x), \mathscr{Q}(x)\Big).$$

Ses atomes sont les $P_i \cap Q_j,\ i \in I, j \in J$. De même, si $(\mathscr{P}_k)_{1 \leq k \leq n}$ est une famille finie de partitions, où $\mathscr{P}_k$ est indexée par I_k, la partition $\bigvee_{k=1}^{n} \mathscr{P}_k$ est la partition indexée par $I_1 \times \cdots \times I_n$, définie par

$$\bigvee_{k=1}^{n} \mathscr{P}_k(x) \ := \ \Big(\mathscr{P}_1(x), \ldots, \mathscr{P}_n(x)\Big).$$

Par contre, si on considère une famille infinie dénombrable de partitions, par exemple $(\mathscr{P}_k)_{k \in \mathbb{N}}$, alors la notation $\bigvee_{k \in \mathbb{N}} \mathscr{P}_k$ désigne cette fois la *tribu* engendrée par la famille $(\mathscr{P}_k)_{k \in \mathbb{N}}$.

1.3 Ordre sur les partitions

On dit que la partition $\bar{\mathscr{P}}$: $X \to \bar{I}$ est *plus fine* que la partition $\mathscr{P}$: $X \to I$ si chaque atome de $\mathscr{P}$ est une réunion d'atomes de $\bar{\mathscr{P}}$. On note dans ce cas $\mathscr{P} \subset \bar{\mathscr{P}}$. On peut interpréter ce signe d'inclusion dans le sens « la tribu engendrée par $\mathscr{P}$ est contenue dans celle engendrée par $\bar{\mathscr{P}}$ ».

Il n'est pas difficile de vérifier l'équivalence entre $\mathscr{P} \subset \bar{\mathscr{P}}$, et chacune des deux propriétés suivantes.

1. Il existe une application $h : \bar{I} \to I$, telle que $\mathscr{P} = h \circ \bar{\mathscr{P}}$. (*i.e.* si on connaît $\bar{\mathscr{P}}(x)$, alors on connaît aussi $\mathscr{P}(x)$.)
2. Il existe une troisième partition $\mathscr{Q}$, indexée par un ensemble J, telle que les atomes non négligeables de $\mathscr{P} \vee \mathscr{Q}$ coïncident avec les atomes non négligeables de $\bar{\mathscr{P}}$.

Remarque 1.2. *Puisqu'on identifie deux partitions égales μ-presque partout, toutes les relations impliquant des partitions sont à comprendre « à un ensemble μ-négligeable près ». Par exemple, dans la première propriété de la liste ci-dessus, l'égalité $\mathscr{P} = h \circ \bar{\mathscr{P}}$ est implicitement à comprendre « μ-presque sûrement ».*

Soient $\mathscr{P}$: $X \to I$ et $\mathscr{Q}$: $X \to J$ deux partitions, et $\varepsilon > 0$. On dit que $\mathscr{Q}$ est *ε-mesurable par rapport à* $\mathscr{P}$, et on note $\mathscr{Q} \overset{\varepsilon}{\subset} \mathscr{P}$ si il existe une partition $\bar{\mathscr{Q}}$: $X \to J$, moins fine que $\mathscr{P}$, telle que $\left|\mathscr{Q} - \bar{\mathscr{Q}}\right| < \varepsilon$.

1.4 Restriction d'une partition

Si A est une partie mesurable non négligeable de X, et si $\mathscr{P}$ est une partition de X indexée par un ensemble I, la *restriction* de $\mathscr{P}$ à A, notée $\mathscr{P}|_A$ est vue comme une partition de l'espace de Lebesgue $(A, \mathscr{B}(X)_A, \mu_A)$. Sa distribution est

$$\mathrm{dist}\mathscr{P}|_A \;=\; \left(\frac{\mu(P_i \cap A)}{\mu(A)}\,,\; i \in I\right).$$

1.5 Partitions indépendantes

Soient deux partitions $\mathscr{P}$: $X \to I$ et $\mathscr{Q}$: $X \to J$. Elles sont dites *indépendantes* si pour tout $i \in I$ et tout $j \in J$, $\mu(P_i \cap Q_j) = \mu(P_i)\mu(Q_j)$. Une formulation équivalente consiste à dire que pour tout atome Q_j non négligeable de $\mathscr{Q}$, $\mathrm{dist}\mathscr{P}|_{Q_j} = \mathrm{dist}\mathscr{P}$.

2 Entropie de Shannon

Dans tout ce qui suit, la fonction logarithme joue un rôle prépondérant, mais la base dans laquelle on considère le logarithme n'a pas une grande importance (elle doit juste être fixée une fois pour toutes). Il est de tradition en théorie de l'information d'utiliser le logarithme en base 2 : la quantité d'information s'interprête comme un nombre de bits, où un bit est une variable qui peut prendre 2 valeurs.

On convient que $0 \log 0 := 0$, ce qui revient à prolonger par continuité la fonction $x \mapsto x \log x$ en 0.

2.1 Entropie d'une partition

Définition 2.1. *Soit $\mathscr{P}$ une partition de X, indexée par I. On appelle* entropie de la partition $\mathscr{P}$ *le nombre réel noté $H(\mathscr{P})$ défini par*

$$H(\mathscr{P}) \ := \ -\sum_{i \in I} \mu(P_i) \log \mu(P_i).$$

Notons que l'entropie de $\mathscr{P}$ ne dépend en fait que de la distribution de la partition $\mathscr{P}$. Les propriétés de la fonction $g \ : \ x \mapsto -x \log x$, qui apparaît dans la définition précédente, sont essentielles dans la suite. Il faut notamment remarquer que g est continue, positive et strictement concave sur $[0, 1]$, avec $g(x) = 0$ seulement si x vaut 0 ou 1. Une première conséquence est que l'entropie de $\mathscr{P}$ est toujours positive, et qu'elle est nulle uniquement dans le cas où $\mathscr{P}$ est triviale (c'est-à-dire lorsque $\mathscr{P}$ comporte un atome de mesure 1).

En fait, l'entropie d'une partition peut s'interpréter comme un nombre qui mesure la difficulté de prévoir dans quel atome de la partition tombe un point x pris "au hasard" dans X (avec la loi de probabilité μ). Cette difficulté est *a priori* maximum si tous les atomes de $\mathscr{P}$ sont équiprobables (le nombre d'atomes étant fixé) ; la proposition qui suit confirme cette idée intuitive.

Proposition 2.2. *Si on fixe le nombre d'atomes k de $\mathscr{P}$, le maximum de l'entropie est atteint lorsque tous les P_i ont la même mesure ; il vaut alors $\log k$.*

Démonstration. Il s'agit de déterminer

$$m(k) \ := \ \max_{(p_1,\ldots,p_k) \in S_k} \sum_{i=1}^{k} g(p_i),$$

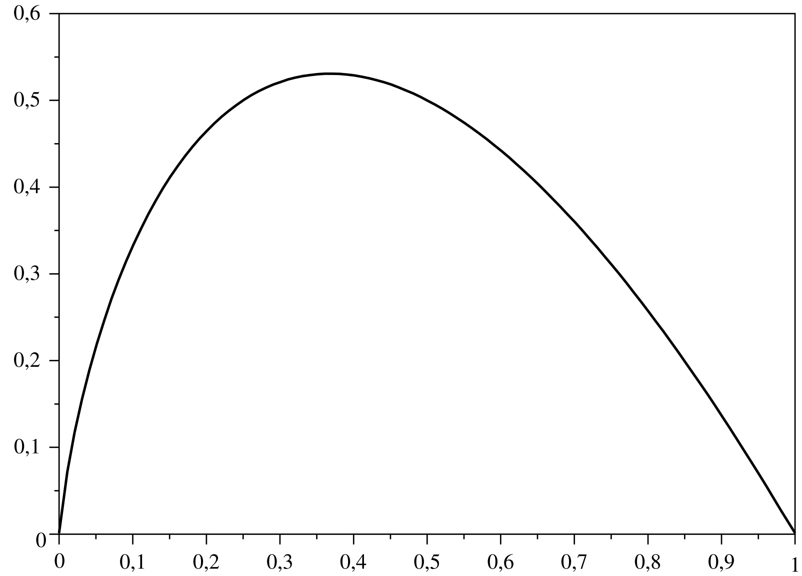

FIGURE 1 – Graphe de $g : x \mapsto -x \log x$

où S_k est l'ensemble compact constitué des k-uplets de réels positifs qui vérifient $\sum_{i=1}^{k} p_k = 1$. Soit $(p_1, \ldots, p_k)$ un point de S_k où ce maximum est atteint, et supposons par l'absurde $p_1 < p_2$. Comme g est strictement concave, on a

$$\frac{g(p_1) + g(p_2)}{2} \; < \; g\left(\frac{p_1 + p_2}{2}\right).$$

Posons alors $p_1' = p_2' := (p_1 + p_2)/2$, $p_3' := p_3, \ldots, p_k' := p_k$. Alors $(p_1', \ldots, p_k') \in S_k$, et

$$\sum_{i=1}^{k} g(p_i') \; > \; \sum_{i=1}^{k} g(p_i),$$

ce qui est impossible par choix de $(p_1, \ldots, p_k)$. On a donc finalement $p_1 = \cdots = p_k = 1/k$, et

$$m_k \; = \; -\frac{1}{k} \sum_{i=1}^{k} \log(1/k) \; = \; \log k.$$

□

Exercice 2.1. *Montrer que, étant donné $\varepsilon > 0$, on peut trouver $\delta > 0$ tel que, si $\mathscr{P}$ est une partition de X,*

$$H(\mathscr{P}) < \delta \implies \mathscr{P} \text{ a un atome de mesure supérieure à } 1 - \varepsilon.$$

Inversement, soit k un entier naturel, et $\varepsilon > 0$. Montrer qu'il existe $\delta > 0$ tel que, pour toute partition $\mathscr{P}$ comportant k atomes,

$$\mathscr{P} \text{ a un atome de mesure supérieure à } 1 - \delta \implies H(\mathscr{P}) < \varepsilon.$$

Pourquoi est-ce important de se limiter à des partitions ne comportant qu'un nombre borné d'atomes dans la question précédente ?

2.2 Entropie conditionnelle

Définition 2.3. *Soient $\mathscr{P}$ et $\mathscr{Q}$ deux partitions de X, indexées respectivement par I et J. On appelle* entropie conditionnelle de $\mathscr{P}$ sachant $\mathscr{Q}$ *le nombre réel, noté $H(\mathscr{P}|\mathscr{Q})$, défini par*

$$H(\mathscr{P}|\mathscr{Q}) \ := \ \sum_{j\in J} \mu(Q_j) H(\mathscr{P}|_{Q_j}).$$

Cette fois, on interprète l'entropie conditionnelle de $\mathscr{P}$ sachant $\mathscr{Q}$ comme la mesure de la difficulté de prévoir $\mathscr{P}(x)$ lorsque l'on connaît déjà $\mathscr{Q}(x)$: si on sait que $\mathscr{Q}(x) = j$, on mesure cette imprévisibilité par $H(\mathscr{P}|_{Q_j})$, puis on fait la moyenne sur tous les atomes Q_j de $\mathscr{Q}$. Clairement, plus on connaît de choses sur x, moins il doit être difficile de deviner $\mathscr{P}(x)$. Aussi, en gardant en tête cette interprétation, les résultats de la proposition suivante ne sont pas surprenants.

Proposition 2.4. *On a toujours $0 \leq H(\mathscr{P}|\mathscr{Q}) \leq H(\mathscr{P})$, avec*

$$\begin{aligned} H(\mathscr{P}|\mathscr{Q}) = 0 &\iff \mathscr{Q} \text{ est plus fine que } \mathscr{P}, \\ H(\mathscr{P}|\mathscr{Q}) = H(\mathscr{P}) &\iff \mathscr{P} \text{ et } \mathscr{Q} \text{ sont indépendantes.} \end{aligned}$$

Démonstration. Puisque l'entropie d'une partition est toujours positive, on a $\forall j,\ H(\mathscr{P}|_{Q_j}) \geq 0$, et donc $H(\mathscr{P}|\mathscr{Q}) \geq 0$, avec égalité si et seulement si pour tout atome Q_j (non négligeable), la partition $\mathscr{P}|_{Q_j}$ est triviale, c'est-à-dire lorsque $\mathscr{Q}$ est plus fine que $\mathscr{P}$. Puis, en utilisant la stricte concavité de g, on obtient

$$\begin{aligned}
H(\mathscr{P}|\mathscr{Q}) &= \sum_{j\in J}\mu(Q_j)\sum_{i\in I} g\left(\mu_{Q_j}(P_i)\right) \\
&= \sum_{i\in I}\sum_{j\in J}\mu(Q_j)\, g\left(\mu_{Q_j}(P_i)\right) \\
&\leq \sum_{i\in I} g\left(\sum_{j\in J}\mu(Q_j)\,\mu_{Q_j}(P_i)\right) \\
&= \sum_{i\in I} g\left(\sum_{j\in J}\mu(P_i\cap Q_j)\right) \\
&= \sum_{i\in I} g(\mu(P_i)) \;=\; H(\mathscr{P}),
\end{aligned}$$

où l'inégalité est une égalité si et seulement si pour tout i, $\mu_{Q_j}(P_i)$ ne dépend pas de j, c'est-à-dire si $\mathscr{P}$ et $\mathscr{Q}$ sont indépendantes. (*Remarque* : dans tous les calculs effectués, on se restreint implicitement aux atomes non négligeables des partitions.)

□

Proposition 2.5. *Soient $\mathscr{P}$, $\mathscr{Q}$ et $\bar{\mathscr{Q}}$ trois partitions de X, indexées par I, J et $\bar{J}$ respectivement. On a les égalités et inégalités suivantes.*

$$H(\mathscr{P}\vee\mathscr{Q}) \;=\; H(\mathscr{P}) + H(\mathscr{Q}|\mathscr{P}) \;\leq\; H(\mathscr{P}) + H(\mathscr{Q}), \tag{1}$$

avec égalité si et seulement si $\mathscr{P}$ et $\mathscr{Q}$ sont indépendantes, et

$$H(\mathscr{P}\vee\mathscr{Q}|\overline{\mathscr{Q}}) \;=\; H(\mathscr{Q}|\bar{\mathscr{Q}}) + H(\mathscr{P}|\mathscr{Q}\vee\bar{\mathscr{Q}}) \leq\; H(\mathscr{P}|\overline{\mathscr{Q}}) + H(\mathscr{Q}|\overline{\mathscr{Q}}). \tag{2}$$

Démonstration.

$$\begin{aligned}
H(\mathscr{P}\vee\mathscr{Q}) &= -\sum_{i,j}\mu(P_i\cap Q_j)\log\mu(P_i\cap Q_j) \\
&= -\sum_{i\in I}\mu(P_i)\sum_{j\in J}\mu_{P_i}(Q_j)\log\mu(P_i\cap Q_j) \\
&= -\sum_{i\in I}\mu(P_i)\sum_{j\in J}\mu_{P_i}(Q_j)\log\mu_{P_i}(Q_j) \\
&\quad -\sum_{i\in I}\mu(P_i)\log\mu(P_i)\sum_{j\in J}\mu_{P_i}(Q_j) \\
&= H(\mathscr{Q}|\mathscr{P}) + H(\mathscr{P}).
\end{aligned}$$

Comme $H(\mathscr{Q}|\mathscr{P}) \leq H(\mathscr{Q})$, avec égalité si et seulement si $\mathscr{P}$ et $\mathscr{Q}$ sont indépendantes, la première partie de la proposition est prouvée. Puis, en conditionnant par rapport à la troisième partition $\bar{\mathscr{Q}}$, on peut écrire

$$H(\mathscr{P} \vee \mathscr{Q}|\bar{\mathscr{Q}}) = \sum_{j \in \bar{J}} \mu(\bar{Q}_j) H\Big((\mathscr{P} \vee \mathscr{Q})|_{\bar{Q}_j}\Big).$$

Comme $(\mathscr{P} \vee \mathscr{Q})|_{\bar{Q}_j} = \mathscr{P}|_{\bar{Q}_j} \vee \mathscr{Q}|_{\bar{Q}_j}$, en effectuant le calcul précédent dans chaque atome $\bar{Q}_j$, on obtient

$$\begin{aligned} H(\mathscr{P} \vee \mathscr{Q}|\bar{\mathscr{Q}}) &= \sum_{j \in \bar{J}} \mu(\bar{Q}_j) \left(H\left(\mathscr{Q}|_{\bar{Q}_j}\right) + H\left(\mathscr{P}|_{\bar{Q}_j} \,\middle|\, \mathscr{Q}|_{\bar{Q}_j}\right)\right) \\ &= H(\mathscr{Q}|\bar{\mathscr{Q}}) + H(\mathscr{P}|\mathscr{Q} \vee \bar{\mathscr{Q}}) \\ &\leq \sum_{j \in \bar{J}} \mu(\bar{Q}_j) \left(H\left(\mathscr{Q}|_{\bar{Q}_j}\right) + H\left(\mathscr{P}|_{\bar{Q}_j}\right)\right) \\ &= H(\mathscr{Q}|\bar{\mathscr{Q}}) + H(\mathscr{P}|\bar{\mathscr{Q}}). \end{aligned}$$

$\square$

Proposition 2.6 (Monotonie de l'entropie)**.** *Si $\bar{\mathscr{Q}}$ est plus fine que $\mathscr{Q}$, on a*

$$\begin{aligned} H(\overline{\mathscr{Q}}) &\geq H(\mathscr{Q}), \\ H(\overline{\mathscr{Q}}|\mathscr{P}) &\geq H(\mathscr{Q}|\mathscr{P}), \\ \textit{et} \qquad H(\mathscr{P}|\overline{\mathscr{Q}}) &\leq H(\mathscr{P}|\mathscr{Q}). \end{aligned}$$

Démonstration. Comme $\bar{\mathscr{Q}}$ est plus fine que $\mathscr{Q}$, on a

$$H(\bar{\mathscr{Q}}) = H(\mathscr{Q} \vee \bar{\mathscr{Q}}) = H(\mathscr{Q}) + H(\bar{\mathscr{Q}}|\mathscr{Q}) \geq H(\mathscr{Q}).$$

En conséquence, on a pour chaque atome P_j non négligeable de $\mathscr{P}$:

$$H(\bar{\mathscr{Q}}|_{P_j}) \geq H(\mathscr{Q}|_{P_j}),$$

et la seconde inégalité s'en déduit. Enfin, on a

$$H(\mathscr{P}|\bar{\mathscr{Q}}) = H(\mathscr{P}|\mathscr{Q} \vee \bar{\mathscr{Q}}) \leq H(\mathscr{P}|\mathscr{Q}),$$

grâce à l'inégalité dans (2), en inversant les rôles de $\mathscr{Q}$ et $\bar{\mathscr{Q}}$. $\square$

Les résultats à prouver dans les deux exercices qui suivent seront utilisés dans la suite.

Exercice 2.2. *Soient $\mathscr{P}$, $\mathscr{Q}$ et $\mathscr{R}$ trois partitions de X. Prouver « l'inégalité triangulaire »*

$$H(\mathscr{P}|\mathscr{Q}) \ \leq \ H(\mathscr{P}|\mathscr{R}) + H(\mathscr{R}|\mathscr{Q}).$$

Exercice 2.3. *Montrer que, étant donné $\varepsilon > 0$, on peut trouver $\delta > 0$ tel que, si $\mathscr{P}$ et $\mathscr{Q}$ sont deux partitions de X,*

$$H(\mathscr{P}|\mathscr{Q}) < \delta \implies \mathscr{P} \overset{\varepsilon}{\subset} \mathscr{Q}.$$

Inversement, soit k un entier naturel, et $\varepsilon > 0$. Montrer qu'il existe $\delta > 0$ tel que, pour toute partition $\mathscr{P}$ comportant k atomes, et toute partition $\mathscr{Q}$,

$$\mathscr{P} \overset{\delta}{\subset} \mathscr{Q} \implies H(\mathscr{P}|\mathscr{Q}) < \varepsilon.$$

2.3 Entropie conditionnelle sachant une sous-tribu

Soit $(\mathscr{P}_k)_{k\geq 1}$ une suite de partitions, et $\mathscr{B} := \bigvee_{k\geq 1} \mathscr{P}_k$ la tribu engendrée par ces partitions. Pour une partition $\mathscr{P}$ quelconque, les propriétés de monotonie de l'entropie conditionnelle assurent que la suite

$$H\left(\mathscr{P}\,\middle|\,\bigvee_{k=1}^{n} \mathscr{P}_k\right), \quad n \geq 1,$$

est décroissante. Comme elle est positive, elle admet une limite, que l'on aimerait appeler *entropie conditionnelle de $\mathscr{P}$ sachant la sous-tribu $\mathscr{B}$*. Mais, pour que ceci ait un sens, il faut auparavant vérifier que la limite ne dépend pas de la suite $(\mathscr{P}_k)_{k\geq 1}$ choisie pour engendrer $\mathscr{B}$. Soit donc $(\mathscr{P}'_k)_{k\geq 1}$ une autre suite de partitions engendrant $\mathscr{B}$. Posons

$$\ell := \lim_{n\to\infty} H\left(\mathscr{P}\,\middle|\,\bigvee_{k=1}^{n} \mathscr{P}_k\right), \text{ et } \quad \ell' := \lim_{n\to\infty} H\left(\mathscr{P}\,\middle|\,\bigvee_{k=1}^{n} \mathscr{P}'_k\right).$$

Soit $\varepsilon > 0$, et n assez grand pour que

$$H\left(\mathscr{P}\,\middle|\,\bigvee_{k=1}^{n} \mathscr{P}_k\right) < \ell + \varepsilon.$$

Grâce au résultat de l'exercice 2.3, on peut ensuite trouver $\delta > 0$ tel que, pour toute partition $\mathscr{Q}$ de X,

$$\bigvee_{k=1}^{n} \mathscr{P}_k \overset{\delta}{\subset} \mathscr{Q} \implies H\left(\bigvee_{k=1}^{n} \mathscr{P}_k\,\middle|\,\mathscr{Q}\right) < \varepsilon.$$

Puisque les partitions $\mathscr{P}_k$ sont mesurables par rapport à la tribu engendrée par les $\mathscr{P}'_k$, on peut alors trouver $N \geq 1$ tel que

$$\bigvee_{k=1}^{n} \mathscr{P}_k \overset{\delta}{\subset} \bigvee_{k=1}^{N} \mathscr{P}'_k,$$

et on a alors en utilisant le résultat de l'exercice 2.2

$$H\left(\mathscr{P}\middle|\bigvee_{k=1}^{N}\mathscr{P}'_k\right) \leq H\left(\mathscr{P}\middle|\bigvee_{k=1}^{n}\mathscr{P}_k\right) + H\left(\bigvee_{k=1}^{n}\mathscr{P}_k\middle|\bigvee_{k=1}^{N}\mathscr{P}'_k\right) < \ell + 2\varepsilon.$$

Comme ε est arbitraire, on en déduit que $\ell' \leq \ell$. (Notons que jusqu'ici, on n'a utilisé que l'inclusion $\bigvee_{k\geq 1} \mathscr{P}'_k \subset \bigvee_{k\geq 1} \mathscr{P}_k$.) Avec l'inclusion inverse, on montre de même que $\ell \leq \ell'$, d'où l'égalité des deux limites. On peut donc maintenant définir de façon cohérente l'entropie de $\mathscr{P}$ sachant une sous-tribu.

Définition 2.7. *Soit $\mathscr{P}$ une partition de X, et $\mathscr{B}$ une sous-tribu de $\mathscr{B}(X)$. On appelle* entropie conditionnelle de $\mathscr{P}$ sachant la sous-tribu $\mathscr{B}$ *le nombre réel positif noté $H(\mathscr{P}|\mathscr{B})$ défini par*

$$H(\mathscr{P}|\mathscr{B}) := \lim_{n\to\infty} H\left(\mathscr{P}\middle|\bigvee_{k=1}^{n}\mathscr{P}_k\right),$$

où $(\mathscr{P}_k)_{k\geq 1}$ est une suite de partitions engendrant $\mathscr{B}$.

Le lecteur est invité à vérifier en exercice que les résultats de monotonie vus dans la proposition 2.6 s'étendent au cas de l'entropie conditionnelle sachant une sous-tribu :

Proposition 2.8. *Si $\mathscr{B} \subset \mathscr{B}'$, alors $H(\mathscr{P}|\mathscr{B}) \geq H(\mathscr{P}|\mathscr{B}')$, et si $\mathscr{Q} \subset \mathscr{P}$, alors $H(\mathscr{P}|\mathscr{B}) \geq H(\mathscr{Q}|\mathscr{B})$.*

On peut également caractériser la mesurabilité par rapport à une sous-tribu par l'entropie conditionnelle sachant cette sous-tribu.

Proposition 2.9. *Soit $\mathscr{P}$ une partition de X, et $\mathscr{B}$ une sous-tribu engendrée par une suite de partitions $(\mathscr{P}_k)_{k\geq 1}$. Les trois propriétés suivantes sont équivalentes.*

1. *$\mathscr{P}$ est $\mathscr{B}$-mesurable.*
2. *Pour tout $\varepsilon > 0$, il existe n tel que $\mathscr{P} \overset{\varepsilon}{\subset} \bigvee_{k=1}^{n} \mathscr{P}_k$.*
3. *$H(\mathscr{P}|\mathscr{B}) = 0$.*

La preuve est laissée en exercice...

3 Entropie avec transformation

Jusqu'ici, la transformation T de $(X, \mathscr{B}(X), \mu)$ n'a encore joué aucun rôle, mais tout ce qui vient d'être exposé va maintenant servir pour définir l'entropie d'un processus $(\mathscr{P}, T)$. Pour tout entier $k \in \mathbb{Z}$, on note $T^{-k}\mathscr{P}$ la partition, indexée aussi par I, et dont les atomes sont les $T^{-k}P_i,\ i \in I$; c'est donc la partition $\mathscr{P} \circ T^k$.

3.1 Entropie d'un processus $(\mathscr{P}, T)$

Étant donné le processus $(\mathscr{P}, T)$, puisque la suite de partitions $\bigvee_{k=1}^n T^k\mathscr{P}$ devient de plus en plus fine quand n grandit, la proposition 2.6 prouve que la suite

$$\left(H\left(\mathscr{P} \,\middle|\, \bigvee_1^n T^k\mathscr{P} \right) \right)_{n\geq 1}$$

est décroissante (et positive !) : sa limite est par définition *l'entropie du processus* $(\mathscr{P}, T)$, notée $h(\mathscr{P}, T)$. L'entropie du processus $(\mathscr{P}, T)$ est donc l'entropie de $\mathscr{P}$ sachant le *passé de* $\mathscr{P}$, où le passé de $\mathscr{P}$ est la sous-tribu engendrée par les partitions $T^k\mathscr{P}$, $k > 0$ (qui correspondent aux coordonnées X_p, $p < 0$).

Proposition 3.1. *On a aussi*

$$h(\mathscr{P}, T) \;=\; \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_0^{n-1} T^k\mathscr{P} \right).$$

Démonstration. En appliquant $(n-1)$ fois les égalités de la proposition 2.5, on obtient pour tout $n \geq 1$

$$\begin{aligned} H\left(\bigvee_0^{n-1} T^k\mathscr{P} \right) \;&=\; H(T^{n-1}\mathscr{P}) + H(T^{n-2}\mathscr{P} \,|\, T^{n-1}\mathscr{P}) + \\ &\quad H(T^{n-3}\mathscr{P} \,|\, T^{n-1}\mathscr{P} \vee T^{n-2}\mathscr{P}) + \cdots + H\left(\mathscr{P} \,\middle|\, \bigvee_1^{n-1} T^k\mathscr{P} \right). \end{aligned}$$

Or, pour tout $j \in \{0, \dots, n-1\}$ on a par conservation de la mesure

$$H\left(T^j\mathscr{P} \,\middle|\, \bigvee_{j+1}^{n-1} T^k\mathscr{P} \right) \;=\; H\left(\mathscr{P} \,\middle|\, \bigvee_1^{n-j-1} T^k\mathscr{P} \right),$$

ainsi

$$\frac{1}{n} H\left(\bigvee_0^{n-1} T^k \mathscr{P}\right) \;=\; \frac{1}{n}\sum_{\ell=0}^{n-1} H\left(\mathscr{P} \;\middle|\; \bigvee_1^{\ell} T^k \mathscr{P}\right),$$

et en appliquant le théorème de convergence de Cesàro, on obtient

$$\frac{1}{n} H\left(\bigvee_0^{n-1} T^k \mathscr{P}\right) \xrightarrow[n\to\infty]{} h(\mathscr{P}, T) \;=\; \lim_{\ell\to\infty} H\left(\mathscr{P} \;\middle|\; \bigvee_1^{\ell} T^k \mathscr{P}\right).$$

□

Proposition 3.2. *On a toujours $h(\mathscr{P}, T) \leq H(\mathscr{P})$, avec égalité si et seulement si les partitions $T^k\mathscr{P},\ k \in \mathbb{Z}$, sont indépendantes.*

Démonstration. D'après la proposition 2.4, on a pour tout n

$$H\left(\mathscr{P} \;\middle|\; \bigvee_1^{n} T^k \mathscr{P}\right) \;\leq\; H(\mathscr{P}),$$

d'où par passage à la limite, $h(\mathscr{P}, T) \leq H(\mathscr{P})$. De plus, comme $H\left(\mathscr{P} \mid \bigvee_1^{n-1} T^k \mathscr{P}\right)$ décroît, on a l'égalité à la limite si et seulement si pour tout n,

$$H\left(\mathscr{P} \;\middle|\; \bigvee_1^{n} T^k \mathscr{P}\right) \;=\; H(\mathscr{P}),$$

c'est-à-dire si pour tout n, $\mathscr{P}$ est indépendante de $\bigvee_1^n T^k \mathscr{P}$.

□

Proposition 3.3. *Soient $\mathscr{P}$ et $\mathscr{Q}$ deux partitions de X. On a*

$$h(\mathscr{P}, T) \;\leq\; h(\mathscr{P} \vee \mathscr{Q}, T) \;\leq\; h(\mathscr{P}, T) + h(\mathscr{Q}, T).$$

En particulier, si $\mathscr{Q}$ est plus fine que $\mathscr{P}$, alors $h(\mathscr{Q}, T) \geq h(\mathscr{P}, T)$.

Démonstration. Utiliser les propositions 2.5 et 3.1.

□

Proposition 3.4 (Continuité de $\mathscr{P} \mapsto h(\mathscr{P}, T)$)**.** *Soient I un ensemble fini et $\varepsilon > 0$ fixés. Alors il existe $\delta > 0$ tel que, si $\mathscr{P}$ et $\overline{\mathscr{P}}$ sont deux partitions de X indexées par I et telles que $|\mathscr{P} - \overline{\mathscr{P}}| < \delta$, alors $|h(\mathscr{P}, T) - h(\overline{\mathscr{P}}, T)| < \varepsilon$.*

Démonstration. Comme la fonction g utilisée pour la définition de l'entropie est continue sur $[0,1]$, on peut prendre $\delta > 0$ tel que

$$\begin{aligned} 0 \le x < \delta &\implies g(x) < \varepsilon/(k^2), \\ \text{et} \quad 0 \le 1 - x < \delta &\implies g(x) < \varepsilon/(k^2). \end{aligned}$$

Soient maintenant $\mathscr{P}$ et $\bar{\mathscr{P}}$ deux partitions de X, indexées par un ensemble I comportant k éléments. On construit alors une partition $\mathscr{Q}$ de X en posant $\mathscr{Q}(x) := 0$ si $\mathscr{P}(x) = \bar{\mathscr{P}}(x)$, et $\mathscr{Q}(x) := (i,j)$ si $\mathscr{P}(x) = i \neq \bar{\mathscr{P}}(x) = j$. Si $\left|\mathscr{P} - \bar{\mathscr{P}}\right| < \delta$, on a

$$H(\mathscr{Q}) \;=\; g\left(\mu\left(\bigcup_{i\in I} P_i \cap \bar{P}_i\right)\right) + \sum_{i\neq j} g\left(\mu\left(P_i \cap \bar{P}_j\right)\right) \;<\; \varepsilon.$$

De plus, la partition $\mathscr{Q}$ est construite de telle sorte que les atomes de $\mathscr{P} \vee \bar{\mathscr{P}}$ sont les mêmes que ceux de $\mathscr{P} \vee \mathscr{Q}$. On a alors

$$\begin{aligned} h(\mathscr{P}, T) &\le h(\mathscr{P} \vee \bar{\mathscr{P}}, T) \;=\; h(\mathscr{P} \vee \mathscr{Q}, T) \\ &\le h(\mathscr{P}, T) + h(\mathscr{Q}, T) \\ &\le h(\mathscr{P}, T) + H(\mathscr{Q}) \\ &< h(\mathscr{P}, T) + \varepsilon, \end{aligned}$$

d'où

$$h(\mathscr{P} \vee \bar{\mathscr{P}}, T) - \varepsilon \;<\; h(\mathscr{P}, T) \;\le\; h(\mathscr{P} \vee \bar{\mathscr{P}}, T).$$

On obtient le même résultat en échangeant $\mathscr{P}$ et $\bar{\mathscr{P}}$, ce qui donne finalement

$$\left|h(\mathscr{P}, T) - H(\bar{\mathscr{P}}, T)\right| \;<\; \varepsilon.$$

□

Proposition 3.5. *Soit $\mathscr{P}$ une partition de X, K un entier naturel, et soit $\mathscr{Q} := \bigvee_{i=-K}^{K} T^k \mathscr{P}$. Alors*

$$h(\mathscr{Q}, T) \;=\; h(\mathscr{P}, T).$$

Démonstration. En utilisant la proposition 3.1, on a

$$h(\mathscr{Q}, T) \;=\; \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{0}^{n-1} T^k \mathscr{Q}\right) \;=\; \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{-K}^{K+n-1} T^k \mathscr{P}\right) \;=\; h(\mathscr{P}, T).$$

□

3.2 Entropie de T

Ayant défini l'entropie d'un processus $(\mathscr{P}, T)$, on peut maintenant introduire l'entropie du système dynamique (X, μ, T).

Définition 3.6. *L'entropie du système dynamique (X, μ, T), notée $h(T)$, est définie par*

$$h(T) \;:=\; \sup_{\mathscr{P}} h(\mathscr{P}, T),$$

où le sup est pris sur toutes les partitions finies de X.

Par cette définition, il est clair que

$$0 \;\leq\; h(T) \;\leq\; +\infty,$$

et que $h(T)$ est un invariant du système (si S est isomorphe à T, alors $h(S) = h(T)$). Le théorème qui suit permet dans certains cas de calculer effectivement $h(T)$. Rappelons qu'une partition $\mathscr{P}$ est génératrice du système (X, μ, T) si

$$\bigvee_{k\in\mathbb{Z}} T^{-k}\mathscr{P} \;=\; \mathscr{B}(X) \mod \mu.$$

Théorème 3.7 (Kolmogorov–Sinai)**.** *Si $\mathscr{P}$ est une partition génératrice du système dynamique (X, μ, T), alors*

$$h(T) \;=\; h(\mathscr{P}, T).$$

Démonstration. Par définition de $h(T)$, on a toujours $h(\mathscr{P}, T) \leq h(T)$. Supposons $\mathscr{P}$ génératrice, et prenons $\varepsilon > 0$. On peut trouver une partition $\mathscr{Q}$ telle que $h(\mathscr{Q}, T) > h(T) - \varepsilon/100$. Puis, soit $\delta > 0$ donné par la proposition 3.4 pour que $\left|\mathscr{Q} - \bar{\mathscr{Q}}\right| < \delta$ entraîne $\left|h(\mathscr{Q}, T) - h(\bar{\mathscr{Q}}, T)\right| < \varepsilon/100$. Comme $\mathscr{P}$ est génératrice, pour K assez grand on peut trouver une partition $\bar{\mathscr{Q}} \subset \bigvee_{-K}^{K} T^i\mathscr{P}$, avec $\left|\mathscr{Q} - \bar{\mathscr{Q}}\right| < \delta$. On a alors, en utilisant les propositions 3.5 et 3.3

$$h(\mathscr{P}, T) \;=\; H\left(\bigvee_{-K}^{K} T^i\mathscr{P}, T\right) \;\geq\; h(\bar{\mathscr{Q}}, T) \;\geq\; h(\mathscr{Q}, T) - \varepsilon/100 \;>\; h(T) - \varepsilon.$$

Comme ε est arbitraire, on a aussi $h(\mathscr{P}, T) \geq h(T)$. $\square$

Remarque 3.8. *Par un argument similaire, on montre que si $(\mathscr{P}_n)$ est une suite croissante de partitions de X telle que*

$$\bigvee_n \mathscr{P}_n \;=\; \mathscr{B}(X) \mod \mu,$$

alors

$$\lim_{n\to\infty} h(\mathscr{P}_n, T) \;=\; h(T).$$

Comme application immédiate du théorème de Kolmogorov-Sinai, on peut maintenant calculer l'entropie d'un décalage de Bernoulli. Soit T le décalage de Bernoulli $B(p_1,\dots,p_k)$, défini sur l'espace $X := \{1,\dots,k\}^{\mathbb{Z}}$. La partition $\mathscr{P}$ de X définie par la coordonnée x_0, indexée par $\{1,\dots,k\}$, est ici génératrice. On a donc $h(T) = h(\mathscr{P}, T)$. De plus, les partitions $T^j\mathscr{P}$, $j \in \mathbb{Z}$ sont indépendantes, car la mesure considérée sur X est ici $(p_1,\dots,p_k)^{\otimes\mathbb{Z}}$. On a donc, par la proposition 3.2,

$$h(T) \;=\; h(\mathscr{P}, T) \;=\; H(\mathscr{P}) \;=\; -\sum_{i=1}^{k} p_i \log p_i.$$

Corollaire 3.9. *Pour que les décalages de Bernoulli $B(p_1,\dots,p_k)$ et $B(q_1,\dots,q_l)$ soient isomorphes, il est nécessaire d'avoir*

$$\sum_{i=1}^{k} p_i \log p_i \;=\; \sum_{i=1}^{k} q_i \log q_i.$$

La question de savoir si cette condition est aussi suffisante est restée ouverte pendant une bonne dizaine d'années, jusqu'à ce que D.S. Ornstein prouve en 1970 [4] que deux décalages de Bernoulli de même entropie sont toujours isomorphes (pour un exposé complet de la théorie d'Ornstein, lire par exemple [5]).

Exercice 3.1. *Soit $(X_p)_{p\in\mathbb{Z}}$ un processus de Markov stationnaire à valeurs dans un alphabet fini I, dont les probabilités de transition sont données par*

$$P(X_0 = j | X_{-1} = j) \;=\; p_{i,j} \quad (i, j \in I).$$

Calculer l'entropie du système dynamique engendré par ce processus en fonction des $p_{i,j}$ et des $\mu_i := P(X_0 = i)$ $(i \in I)$.

3.3 Un exemple de transformations d'entropie nulle

Rappelons que le système dynamique mesuré (X, μ, T) est dit *rigide* s'il existe une sous-suite d'entiers $(j_n)_{n\geq 1}$ tendant vers $+\infty$, telle que pour tout $A \in \mathscr{B}(X)$,

$$\mu(A \bigtriangleup T^{-j_n} A) \xrightarrow[n\to\infty]{} 0. \tag{3}$$

Un exemple simple de transformation rigide est donné par une rotation sur le cercle.

Proposition 3.10. *Si T est rigide, alors $h(T) = 0$.*

Démonstration. Il s'agit de montrer que pour toute partition $\mathscr{P}$ de X, $h(\mathscr{P}, T) = 0$. Soit donc $\mathscr{P}$ une partition, et $(j_n)_{n\geq 1}$ une suite de rigidité pour T. On voit facilement que

$$\left|\mathscr{P} - T^{j_n}\mathscr{P}\right| \xrightarrow[n\to\infty]{} 0.$$

En utilisant la seconde partie de l'exercice 2.3, on en déduit que

$$H\left(\mathscr{P} \,\middle|\, \bigvee_1^n T^k\mathscr{P}\right) \xrightarrow[n\to\infty]{} 0,$$

d'où $h(\mathscr{P}, T) = 0$. □

3.4 Entropie d'un facteur de T

Rappelons qu'un système (Y, ν, S) est un facteur de (X, μ, T) si il existe une application mesurable $\pi : X \to Y$ telle que $\pi_*(\mu) = \nu$ et vérifiant $\pi \circ T = S \circ \pi$. La démonstration de la proposition suivante est immédiate ; elle est laissée en exercice.

Proposition 3.11. *Si le système dynamique (Y, ν, S) est un facteur de (X, μ, T), alors*

$$h(S) \ \leq\ h(T).$$

Si $\mathscr{F} \subset \mathscr{B}(X)$ est une sous-tribu facteur de T (c'est-à-dire une sous-tribu de $\mathscr{B}(X)$ qui est T-invariante, on définit également l'entropie de T sur $\mathscr{F}$ par

$$h(\mathscr{F}, T) \ := \ \sup\left\{h(\mathscr{P}, T) : \ \mathscr{P} \text{ partition } \mathscr{F}\text{-mesurable de } X\right\}.$$

On a de même

$$h(\mathscr{F}, T) \ \leq\ h(T).$$

3.5 Entropie des puissances de T

Proposition 3.12. *Soit T un automorphisme de $\big(X, \mathscr{B}(X), \mu\big)$, et $p \in \mathbb{Z}$. Alors*

$$h(T^p) \;=\; |p|\, h(T).$$

Démonstration. Montrons d'abord que $H(T^{-1}) = h(T)$. Pour cela, il suffit de remarquer que pour toute partition $\mathscr{P}$,

$$h(\mathscr{P}, T) \;=\; \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k \mathscr{P}\right) \;=\; \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k} \mathscr{P}\right) \;=\; H(\mathscr{P}, T^{-1}).$$

On peut ensuite supposer $p > 0$. (Pour $p = 0$, c'est évident.) Soit $\mathscr{P}$ une partition de X, et soit $\mathscr{Q} := \bigvee_{k=0}^{p-1} T^k \mathscr{P}$.

$$\begin{aligned}
H(\mathscr{Q}, T^p) &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{pk} \mathscr{Q}\right) \\
&= \lim_{n\to\infty} \frac{p}{np} H\left(\bigvee_{k=0}^{np-1} T^{k} \mathscr{P}\right) \\
&= p\, h(\mathscr{P}, T).
\end{aligned}$$

Puis

$$\begin{aligned}
h(T^p) &= \sup_{\mathscr{P}} h(\mathscr{P}, T^p) \\
&= \sup_{\mathscr{P}} h\left(\bigvee_{k=0}^{p-1} T^k \mathscr{P}, T^p\right) \\
&= p \sup_{\mathscr{P}} h(\mathscr{P}, T) \\
&= p\, h(T).
\end{aligned}$$

□

3.6 Entropie d'un produit

Proposition 3.13. *Soient (X, μ, T) et (Y, ν, S) deux systèmes dynamiques mesurés. On considère le système produit $(X \times Y, \mu \otimes \nu, T \times S)$. Alors*

$$h(T \times S) \;=\; h(T) + h(S).$$

Démonstration. On considère une suite croissante $(\mathscr{P}_n)$ de partitions de X telle que

$$\bigvee_n \mathscr{P}_n \ = \ \mathscr{B}(X) \mod \mu,$$

et de même une suite croissante $(\mathscr{Q}_n)$ de partitions de Y telle que

$$\bigvee_n \mathscr{Q}_n \ = \ \mathscr{B}(Y) \mod \nu.$$

Par la remarque 3.8, on a $h(T) = \lim_n h(\mathscr{P}_n, T)$, et $h(S) = \lim_n h(\mathscr{Q}_n, S)$. Pour chaque n, on forme les partitions $\bar{\mathscr{P}}_n$ et $\bar{\mathscr{Q}}_n$ et $\bar{\mathscr{P}}_n \vee \bar{\mathscr{Q}}_n$ de $X \times Y$, définies par $\bar{\mathscr{P}}_n(x,y) := \mathscr{P}_n(x)$, et $\bar{\mathscr{Q}}_n(x,y) := \mathscr{Q}_n(y)$. On vérifie immédiatement que $h(\bar{\mathscr{P}}_n, T \times S) = h(\mathscr{P}_n, T)$, et $h(\bar{\mathscr{Q}}_n, T \times S) = h(\mathscr{Q}_n, S)$. On considère alors la suite de partitions $(\bar{\mathscr{P}}_n \vee \bar{\mathscr{Q}}_n)$ de $X \times Y$: c'est une suite croissante de partitions telle que

$$\bigvee_n \bar{\mathscr{P}}_n \vee \bar{\mathscr{Q}}_n \ = \ \mathscr{B}(X \times Y) \mod \mu \otimes \nu.$$

On donc de même par la remarque 3.8

$$h(T \times S) = \lim_n h(\bar{\mathscr{P}}_n \vee \bar{\mathscr{Q}}_n, T \times S).$$

Or, les sous-tribus $\bigvee_{k \in \mathbb{Z}} (T \times S)^{-k} \bar{\mathscr{P}}_n$ et $\bigvee_{k \in \mathbb{Z}} (T \times S)^{-k} \bar{\mathscr{Q}}_n$ étant indépendantes sous la mesure produit $\mu \otimes \nu$, on obtient facilement que pour tout n,

$$h(\bar{\mathscr{P}}_n \vee \bar{\mathscr{Q}}_n, T \times S) = h(\bar{\mathscr{P}}_n, T \times S) + h(\bar{\mathscr{Q}}_n, T \times S),$$

et on a alors tous les éléments pour conclure la preuve. □

4 Facteur de Pinsker et K-systèmes

Si A est une partie mesurable de X, la fonction indicatrice $\mathbb{1}_A$ peut être vue comme une partition de X, indexée par $\{0,1\}$, et dont les atomes sont A et $X \setminus A$.

4.1 Définition du facteur de Pinsker

Proposition 4.1. *Soit (X, μ, T)s un système dynamique mesuré. Alors*

$$\Pi(T) \ := \ \{A \in \mathscr{B}(X) : \ h(\mathbb{1}_A, T) = 0\}$$

est une sous-tribu T-invariante, appelée facteur de Pinsker *de T. C'est la plus grande sous-tribu facteur d'entropie nulle dans le système.*

Démonstration. Clairement, $\Pi(T)$ contient X et $\emptyset$, et est stable par passage au complémentaire. Vérifions la stabilité de $\Pi(T)$ par réunion finie. Prenons A et B dans $\Pi(T)$. On a alors

$$h(\mathbb{1}_A \vee \mathbb{1}_B, T) \ \leq \ h(\mathbb{1}_A, T) + h(\mathbb{1}_B, T) \ = \ 0.$$

Mais la partition $\mathbb{1}_{A \cup B}$ est moins fine que $\mathbb{1}_A \vee \mathbb{1}_B$. On a donc

$$h(\mathbb{1}_{A \cup B}, T) \ \leq \ h(\mathbb{1}_A \vee \mathbb{1}_B, T) \ = \ 0,$$

d'où $A \cup B \in \Pi(T)$.

Il reste maintenant à vérifier la stabilité de $\Pi(T)$ par limite croissante. Soit $(A_n)_{n \in \mathbb{N}}$ une suite croissante d'éléments de $\Pi(T)$, et $A := \bigcup_{n \in \mathbb{N}} A_n$. On a

$$|\mathbb{1}_A - \mathbb{1}_{A_n}| \xrightarrow[n \to \infty]{} 0,$$

et donc par continuité de $\mathscr{P} \mapsto h(\mathscr{P}, T)$,

$$h(\mathbb{1}_A, T) \ = \ \lim_{n \to \infty} h(\mathbb{1}_{A_n}, T) \ = \ 0,$$

et donc $A \in \Pi(T)$. Ainsi, $\Pi(T)$ est bien une sous-tribu de $\mathscr{B}(X)$. On vérifie immédiatement que $\Pi(T)$ est T-invariante, et donc $\Pi(T)$ est un facteur de T. Par ailleurs, si $\mathscr{P}$ est une partition de X qui est $\Pi(T)$-mesurable, alors

$$h(\mathscr{P}, T) = h\left(\bigvee_{A \text{ atome de } \mathscr{P}} \mathbb{1}_A, T \right) \leq \sum_{A \text{ atome de } \mathscr{P}} h(\mathbb{1}_A, T) = 0,$$

et par définition il est clair que $\Pi(T)$ est la plus grande sous-tribu facteur d'entropie nulle de T. $\square$

4.2 Passé lointain d'une partition

Définition 4.2. *Soit $\mathscr{P}$ une partition de X. On appelle* passé lointain de $\mathscr{P}$ *la sous-tribu de $\mathscr{B}(X)$, notée $\Pi(\mathscr{P},T)$, définie par*

$$\Pi(\mathscr{P},T) \ := \ \bigcap_{n\geq 1} \bigvee_{k=n}^{+\infty} T^k\mathscr{P}.$$

Il est facile de voir que $\Pi(\mathscr{P},T)$ est encore un facteur de T. On souhaite maintenant établir que ce facteur est d'entropie nulle, et donc inclus dans $\Pi(T)$. Pour cela, on aura besoin du lemme suivant.

Lemme 4.3. *Soient $\mathscr{P}$ et $\mathscr{Q}$ deux partitions de X. Alors*

$$h(\mathscr{P},T) \ = \ H\left(\mathscr{P} \,\middle|\, \bigvee_{k\geq 1} T^k\mathscr{P} \vee \Pi(\mathscr{Q},T)\right).$$

Démonstration. On vérifie aisément par récurrence que pour toute partition $\mathscr{R}$, et tout entier $n\geq 1$, on a

$$H\left(\bigvee_0^{n-1} T^{-j}\mathscr{R} \,\middle|\, \bigvee_{k\geq 1} T^k\mathscr{R}\right) \ = \ n\, h(\mathscr{R},T). \tag{4}$$

On applique (4) avec $\mathscr{R} := \mathscr{P}\vee\mathscr{Q}$. On obtient

$$\begin{aligned}
h(\mathscr{P}\vee\mathscr{Q},T) \quad &= \quad \underbrace{\frac{1}{n}H\left(\bigvee_0^{n-1} T^{-j}(\mathscr{P}\vee\mathscr{Q}) \,\middle|\, \bigvee_{k\geq 1} T^k(\mathscr{P}\vee\mathscr{Q})\right)}_{A_n} \\
&\leq \quad \underbrace{\frac{1}{n}H\left(\bigvee_0^{n-1} T^{-j}(\mathscr{P}\vee\mathscr{Q}) \,\middle|\, \bigvee_{k\geq 1} T^k\mathscr{P}\right)}_{B_n} \\
&\leq \quad \frac{1}{n}H\left(\bigvee_0^{n-1} T^{-j}(\mathscr{P}\vee\mathscr{Q})\right) \\
&\xrightarrow[n\to\infty]{} \quad h(\mathscr{P}\vee\mathscr{Q},T).
\end{aligned}$$

On en déduit que A_n et B_n ont même limite quand $n \to \infty$. On peut ensuite écrire

$$A_n = \underbrace{\frac{1}{n} H\left(\bigvee_0^{n-1} T^{-j}\mathscr{P} \,\middle|\, \bigvee_{k\geq 1} T^k(\mathscr{P}\vee\mathscr{Q})\right)}_{A'_n} + \underbrace{\frac{1}{n} H\left(\bigvee_0^{n-1} T^{-j}\mathscr{Q} \,\middle|\, \bigvee_{k\geq 1} T^k\mathscr{Q} \vee \bigvee_{-(n-1)}^{+\infty} T^k\mathscr{P}\right)}_{A''_n},$$

et

$$B_n = \underbrace{\frac{1}{n} H\left(\bigvee_0^{n-1} T^{-j}\mathscr{P} \,\middle|\, \bigvee_{k\geq 1} T^k\mathscr{P}\right)}_{B'_n} + \underbrace{\frac{1}{n} H\left(\bigvee_0^{n-1} T^{-j}\mathscr{Q} \,\middle|\, \bigvee_{-(n-1)}^{+\infty} T^k\mathscr{P}\right)}_{B''_n}.$$

On a toujours $A'_n \leq B'_n$, et $A''_n \leq B''_n$. Forcément, A'_n et B'_n ont donc aussi même limite. Mais B'_n est toujours égal à $h(\mathscr{P}, T)$. Ainsi,

$$\begin{aligned}
h(\mathscr{P},T) &= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_0^{n-1} T^{-j}\mathscr{P} \,\middle|\, \bigvee_{k\geq 1} T^k(\mathscr{P}\vee\mathscr{Q})\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} H\left(T^{-j}\mathscr{P} \,\middle|\, \bigvee_{k\geq 1} T^k\mathscr{Q} \vee \bigvee_{-(j-1)}^{+\infty} T^k\mathscr{P}\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} H\left(\mathscr{P} \,\middle|\, \bigvee_{k\geq j+1} T^k\mathscr{Q} \vee \bigvee_1^{+\infty} T^k\mathscr{P}\right) \\
&= \lim_{j\to\infty} H\left(\mathscr{P} \,\middle|\, \bigvee_1^{+\infty} T^k\mathscr{P} \vee \bigvee_{k\geq j+1} T^k\mathscr{Q}\right) \\
&\leq H\left(\mathscr{P} \,\middle|\, \bigvee_1^{+\infty} T^k\mathscr{P} \vee \Pi(\mathscr{Q},T)\right).
\end{aligned}$$

L'autre inégalité

$$H\left(\mathscr{P} \,\middle|\, \bigvee_1^{+\infty} T^k\mathscr{P} \vee \Pi(\mathscr{Q},T)\right) \leq h(\mathscr{P},T)$$

étant évidente, le lemme est démontré. □

On peut maintenant prouver que pour toute partition $\mathscr{Q}$, $\Pi(\mathscr{Q},T) \subset \Pi(T)$. En effet, prenons A dans $\Pi(\mathscr{Q},T)$, et soit $\mathscr{P} := \mathbb{1}_A$. D'après le lemme précédent, on a

$$h(\mathscr{P},T) \;=\; H\left(\mathscr{P}\,\middle|\,\bigvee_{1}^{+\infty} T^k\mathscr{P} \vee \Pi(\mathscr{Q},T)\right) \;=\; 0$$

car $\mathscr{P}$ est $\Pi(\mathscr{Q},T)$-mesurable. Ce qui prouve que $A \in \Pi(T)$. On a donc le résultat suivant.

Proposition 4.4. *Pour toute partition $\mathscr{Q}$, on a*

$$\Pi(\mathscr{Q},T) \;\subset\; \Pi(T).$$

Proposition 4.5. *Soient $\mathscr{P}$ et $\mathscr{Q}$ deux partitions telles que $\mathscr{Q}$ soit $\bigvee_{-\infty}^{+\infty} T^k\mathscr{P}$-mesurable. Alors*

$$\Pi(\mathscr{Q},T) \;\subset\; \Pi(\mathscr{P},T).$$

Démonstration. Soit $\mathscr{R}$ une partition mesurable par rapport à $\Pi(\mathscr{Q},T)$. Par une technique analogue à celle utilisée pour prouver le lemme 4.3, on montre que si $\overline{\mathscr{P}}$ est mesurable par rapport à $\bigvee_{-N}^{N} T^k\mathscr{P}$, alors

$$H(\overline{\mathscr{P}}|\Pi(\mathscr{P},T)\vee\mathscr{R}) \;=\; H(\overline{\mathscr{P}}|\Pi(\mathscr{P},T)). \tag{5}$$

(Pour les détails, on peut regarder [1] p. 287.) L'hypothèse du lemme assure alors que $\mathscr{R}$ est mesurable par rapport à $\bigvee_{-\infty}^{+\infty} T^k\mathscr{P}$, donc peut être approchée arbitrairement bien par une partition $\overline{\mathscr{P}} \subset \bigvee_{-N}^{N} T^k\mathscr{P}$ pourvu que N soit assez grand. Par continuité de l'entropie, on obtient alors grâce à (5)

$$0 \;=\; H(\mathscr{R}|\Pi(\mathscr{P},T)\vee\mathscr{R}) \;=\; H(\mathscr{R}|\Pi(\mathscr{P},T)),$$

et donc $\mathscr{R}$ est $\Pi(\mathscr{P},T)$-mesurable. □

Corollaire 4.6. *Si $\mathscr{P}$ est une partition génératrice du système, alors*

$$\Pi(T) \;=\; \Pi(\mathscr{P},T).$$

4.3 K-systèmes

On note ici $\mathcal{T}$ la sous-tribu triviale de $\mathscr{B}(X)$, c'est-à-dire l'ensemble des parties mesurables de X de probabilité 0 ou 1.

Définition 4.7. *On dit que le système dynamique* (X, μ, T) *est un* K-système *si le facteur de Pinsker* $\Pi(T)$ *est réduit à* $\mathcal{T}$*.*

Autrement dit, T est un K-système si $h(\mathscr{P}, T) > 0$ pour toute partition $\mathscr{P}$ non triviale de X, ou encore si le seul facteur d'entropie nulle de T est le facteur trivial. On dit aussi dans ce cas que T *est d'entropie totalement positive.*

Théorème 4.8. *Les trois propriétés suivantes sont équivalentes.*

1. *T est un K-système.*
2. *Pour toute partition $\mathscr{P}$ de X, $\Pi(\mathscr{P}, T) = \mathcal{T}$.*
3. *Pour tout $A \in \mathscr{B}(X)$, et toute partition $\mathscr{P}$,*

$$\sup_{B \in \bigvee_n^{+\infty} T^k \mathscr{P}} |\mu(A \cap B) - \mu(A)\mu(B)| \xrightarrow[n \to +\infty]{} 0.$$

Remarquons que la propriété 3 est plus forte que la propriété de mélange : elle impose une certaine uniformité dans la convergence de $\mu(T^n B \cap A)$ vers $\mu(B)\mu(A)$. Les transformations vérifiant cette propriété sont dites *K-mélangeantes.*

Démonstration. 1 $\Longrightarrow$ 2 se déduit immédiatement de la proposition 4.4.

2 $\Longrightarrow$ 1. Soit $\mathscr{P}$ une partition telle que $h(\mathscr{P}, T) = 0$. On a donc

$$H\left(\mathscr{P} \,\middle|\, \bigvee_{k \geq 1} T^k \mathscr{P}\right) = 0,$$

et donc $\mathscr{P}$ est $\bigvee_{k \geq 1} T^k \mathscr{P}$-mesurable. On voit alors facilement par récurrence que pour tout $n \geq 1$, $\mathscr{P}$ est $\bigvee_{k \geq n} T^k \mathscr{P}$-mesurable, et donc $\mathscr{P}$ est $\Pi(\mathscr{P}, T)$-mesurable. Si on suppose 2, $\Pi(\mathscr{P}, T) = \mathcal{T}$ et donc $\mathscr{P}$ ne peut être que la tribu triviale.

2 $\Longrightarrow$ 3. Soit $A \in \mathscr{B}(X)$, et $\mathscr{P}$ une partition de X. Pour $n \geq 1$, soit $\mathscr{B}_n := \bigvee_{k \geq n} T^k \mathscr{P}$, et $\mathscr{B}_\infty := \bigcap_{n \geq 1} \mathscr{B}_n = \Pi(\mathscr{P}, T)$. En supposant 2, $\mathscr{B}_\infty$

est donc la tribu triviale $\mathcal{T}$. Par un théorème classique de convergence de martingale, on a alors

$$\mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{B}_n\right] \xrightarrow[n\to+\infty]{L^1} \mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{B}_\infty\right] = \mu(A).$$

Soit maintenant $n \geq 1$. Pour tout $B \in \mathscr{B}_n$, on a

$$\begin{aligned}
|\mu(A\cap B) - \mu(A)\mu(B)| &= \left|\int_X \Big(\mathbb{1}_A - \mu(A)\Big)\mathbb{1}_B\, d\mu\right| \\
&= \left|\int_B \Big(\mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{B}_n\right] - \mu(A)\Big)\, d\mu\right| \\
&\leq \int_X \left|\mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{B}_n\right] - \mu(A)\right| d\mu \xrightarrow[n\to+\infty]{} 0.
\end{aligned}$$

3 $\Longrightarrow$ 2. Soit $\mathscr{P}$ une partition, et $A \in \Pi(\mathscr{P}, T)$. On a donc $A \in \bigvee_{k\geq n} T^k\mathscr{P}$ pour tout $n \geq 1$, et donc en supposant 3,

$$\left|\mu(A) - \mu(A)^2\right| \leq \sup_{B\in\bigvee_n^{+\infty} T^k\mathscr{P}} |\mu(A\cap B) - \mu(A)\mu(B)| \xrightarrow[n\to+\infty]{} 0.$$

On a donc $\mu(A) = \mu(A)^2$, *i.e.* $\mu(A) = 0$ ou 1. D'où $\Pi(\mathscr{P}, T) = \mathcal{T}$.

□

Proposition 4.9. *Les décalages de Bernoulli sont des K-systèmes.*

Démonstration. Si (X, μ, T) est un décalage de Bernoulli, il existe une partition $\mathscr{P}$ génératrice telle que les partitions $T^k\mathscr{P}$, $k \in \mathbb{Z}$, soient indépendantes. On a alors par le corollaire 4.6

$$\Pi(T) = \Pi(\mathscr{P}, T).$$

Mais la loi du 0-1 de Kolmogorov prouve dans ce cas que $\Pi(\mathscr{P}, T)$ est la tribu triviale, et donc $\Pi(T) = \mathcal{T}$.

□

La réciproque n'est pas vraie : il existe une multitude d'exemples de K-systèmes non Bernoulli. Le plus simple à décrire est le système couramment désigné par « T, T^{-1} » : l'espace d'états est $Y = X \times X$, où $X := \{-1, 1\}^{\mathbb{Z}}$.

On désigne par T le décalage vers la gauche des coordonnées sur X, et on considère sur X la mesure

$$\mu := \left(\frac{\delta_{-1} + \delta_1}{2}\right)^{\otimes \mathbb{Z}}.$$

On définit alors sur Y la transformation S par

$$S(x_1, x_2) \;:=\; \left(Tx_1, T^{x_1(0)}x_2\right).$$

Il a été démontré en 1974 par Meilijson [3] que (Y, ν, S) est un K-système, puis par Kalikow en 1982 [2] que ce système n'est pas isomorphe à un schéma de Bernoulli.

Références

[1] I.P. Cornfeld, S.V. Fomin, and Ya.G. Sinai, *Ergodic theory*, Springer-Verlag, 1982.

[2] S.A. Kalikow, *T, T^{-1} transformation is not loosely bernoulli*, Annals of Math. **115** (1982), 393–409.

[3] I. Meilijson, *Mixing properties of a class of skew-products*, Israel Journal of Mathematics **19** (1974), 266–270.

[4] D.S. Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Advances in Math. **4** (1970), 337–352.

[5] ______, *Ergodic theory, randomness and dynamical systems*, Yale Univ. Press, New Haven, 1974.