

# Théorie Ergodique - Introduction et premiers exemples

T. de la Rue

Master 2 MFA Rouen 2024-2025

## 1 Systèmes dynamiques, moyennes temporelles et mesures invariantes

### 1.1 Très courte introduction aux systèmes dynamiques

L'expression *système dynamique* est utilisée en mathématiques dans des contextes variés, qui en général font référence à la modélisation de l'évolution dans le temps d'un système dont l'état change selon certains principes. Dans ce cours nous nous intéresserons à des systèmes dynamiques dits *déterministes*, pour lesquels l'état du système à un instant donné détermine entièrement l'évolution future du système. Les systèmes étudiés seront le plus souvent observés à intervalles de temps réguliers (à des instants qui sont des multiples entiers d'une unité de temps) : on parle de systèmes « à temps discret » (par opposition à des systèmes en temps continu, où on peut considérer l'état du système à tout instant  $t \in \mathbb{R}$  ou  $\mathbb{R}_+$ ).

On considère l'ensemble  $X$  de tous les états possibles du système ( $X$  est l'espace d'états). Si à l'instant  $n = 0$  on sait que le système est dans l'état  $x \in X$ , alors on peut théoriquement prédire tous les états futurs du système. En particulier, à l'instant  $n = 1$ , le nouvel état du système est entièrement déterminé par  $x$  : c'est donc une fonction de  $x$  que l'on va noter  $T(x)$ , ou  $Tx$ . Ici  $T : X \rightarrow X$  est ainsi une transformation de l'espace d'états, qui modélise l'évolution du système entre deux unités de temps. On suppose bien sûr que le système continue d'évoluer suivant les mêmes règles, et donc à l'instant  $n = 2$  il sera dans l'état  $T(T(x)) = T^2x$ . Plus généralement,  $T^n x$  représente

l'état du système à l'instant  $n$  pour tout entier naturel  $n$ . Le plus souvent on supposera également que la transformation  $T$  est *inversible* (autrement dit, que c'est une bijection de  $X$  sur lui-même), et donc on pourra considérer l'état  $T^n x$  du système pour tout instant  $n \in \mathbb{Z}$ .

## 1.2 Moyennes temporelles, mesures invariantes

L'objet central de la *théorie ergodique* est l'étude des propriétés *statistiques* de tels systèmes, c'est-à-dire les propriétés liées au comportement des *moyennes temporelles* de fonctions (« observables ») le long des trajectoires, ou *orbites*, du système. Étant donnée une observable  $f$ , qui est une fonction numérique définie sur l'espace d'états  $X$  et à valeurs dans  $\mathbb{R}$  ou  $\mathbb{C}$ , on s'intéresse aux moyennes temporelles de la forme

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x), \quad (1)$$

pour  $x \in X$  et  $N$  entier naturel non nul. Par exemple, si  $A$  est une partie de l'espace d'états  $X$  on peut considérer l'observable  $f = \mathbb{1}_A$  donnée par l'indicatrice de  $A$ . Dans ce cas l'expression (1) ci-dessus représente la proportion du temps passé dans  $A$  sur les  $N$  premiers points de l'orbite d'un état initial  $x$ .

Comme on va le voir dans la section qui suit, le comportement à long terme de ces moyennes temporelles est intimement lié à la nature des mesures de probabilités sur  $X$  qui sont *invariantes* par la transformation  $T$ . Afin de bien préciser cette notion, rappelons tout d'abord celle de *mesure image* : si  $\Phi$  est une application mesurable de  $(X, \mathcal{B}(X))$  dans  $(Y, \mathcal{B}(Y))$ , et si  $\mu$  est une mesure sur  $(X, \mathcal{B}(X))$ , on définit la mesure image  $\Phi_*(\mu)$  comme étant la mesure sur  $(Y, \mathcal{B}(Y))$  donnée, pour tout  $B \in \mathcal{B}(Y)$ , par

$$\Phi_*(\mu)(B) := \mu\left(\Phi^{-1}(B)\right).$$

Cette mesure image est également caractérisée par le fait que pour toute fonction  $f : Y \rightarrow \mathbb{R}$  mesurable et bornée, l'égalité suivante a lieu :

$$\int_Y f d\Phi_*(\mu) = \int_X f \circ \Phi d\mu.$$

Une mesure  $\mu$  sur  $(X, \mathcal{B}(X))$  est dite *T-invariante* si  $T_*(\mu) = \mu$ , autrement dit si pour tout  $A \in \mathcal{B}(X)$ ,

$$\mu(T^{-1}A) = \mu(A),$$

ou encore : pour toute  $f : X \rightarrow \mathbb{R}$  mesurable bornée,

$$\int_X f \circ T d\mu = \int_X f d\mu.$$

### 1.3 Systèmes dynamiques topologiques, existence de mesures invariantes

On se place ici dans la situation où l'espace d'états  $X$  est un espace métrisable compact, et  $T$  est une transformation *continue* sur  $X$ . On dit dans ce cas que  $(X, T)$  est un *système dynamique topologique*.

On étudie les moyennes temporelles de type (1) pour des observables continues  $f : X \rightarrow \mathbb{C}$ . Remarquons que l'on peut écrire une telle moyenne temporelle sous la forme

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f dm_N^x,$$

où  $m_N^x$  est la mesure de probabilité sur  $X$  définie par

$$m_N^x := \frac{1}{N} \sum_{n=0}^{N-1} \delta_{T^n x}.$$

La notation  $\delta_y$  désigne la masse de Dirac au point  $y$ , ainsi  $m_N^x$  est la mesure uniforme sur les  $N$  premiers points de l'orbite de  $x$  pour la transformation  $T$ . On désigne une telle mesure par le terme de *mesure empirique* partant de  $x$ . Il est naturel de tenter de relier la convergence des moyennes temporelles (1) à une éventuelle convergence, dans un sens à préciser, des moyennes empiriques partant de  $x$ .

Or, l'espace  $\mathcal{M}_1(X)$  des mesures de probabilité sur l'espace métrisable compact  $X$  est muni d'une topologie parfaitement adaptée dans ce contexte : celle associée à la *convergence faible* : rappelons qu'une suite de probabilités

$(\mu_N)$  converge faiblement vers  $\mu \in \mathcal{M}_1(X)$  si pour toute fonction continue  $f : X \rightarrow \mathbb{R}$ , on a

$$\int_X f d\mu_N \xrightarrow{N \rightarrow \infty} \int_X f d\mu.$$

Pour cette topologie,  $\mathcal{M}_1(X)$  est lui-même métrisable compact. En effet, la métrisabilité est une conséquence du fait que l'espace  $C(X)$  des fonctions continues de  $X$  dans  $\mathbb{R}$  est séparable. En fixant une famille dénombrable  $(f_j)_{j \in \mathbb{N}}$  dense dans  $C(X)$ , on peut définir une distance sur  $\mathcal{M}_1(X)$  induisant la topologie de la convergence faible en posant

$$d(\mu, \nu) := \sum_{j \in \mathbb{N}} 2^{-j} \left| \int_X f_j d\mu - \int_X f_j d\nu \right|.$$

La compacité séquentielle se montre alors par un procédé diagonal classique : de toute suite  $(\mu_N)$  dans  $\mathcal{M}_1(X)$  on peut extraire une sous-suite  $(\mu_{N_k})$  telle que pour tout  $j$ ,  $\int_X f_j d\mu_{N_k}$  converge vers un réel  $\Lambda(f_j)$ . Puis par densité, pour toute  $f \in C(X)$ ,  $\int_X f d\mu_{N_k}$  converge vers un réel  $\Lambda(f)$ . On peut alors invoquer le théorème de représentation de Riesz pour trouver  $\mu \in \mathcal{M}_1(X)$  telle que  $\Lambda(f) = \int_X f d\mu$  pour toute  $f \in C(X)$ .

Revenons à la suite  $(m_N^x)_{N \geq 1}$  des moyennes empiriques partant d'un point quelconque  $x \in X$ . Par compacité de  $\mathcal{M}_1(X)$  pour la convergence faible, on peut toujours en extraire une sous-suite  $(m_{N_k}^x)_{k \geq 1}$  qui converge vers une mesure de probabilité  $\mu \in \mathcal{M}_1(X)$ . Comparons maintenant avec les moyennes empiriques partant de  $Tx$  : pour  $f \in C(X)$ , on a

$$\left| \int_X f dm_N^x - \int_X f dm_N^{Tx} \right| = \frac{1}{N} |f(x) - f(T^N x)| \leq \frac{2}{N} \|f\|_\infty.$$

Comme  $\int_X f dm_{N_k}^x$  converge vers  $\int_X f d\mu$ , on en déduit que  $\int_X f dm_{N_k}^{Tx}$  converge également vers  $\int_X f d\mu$ . Mais par ailleurs,

$$\int_X f dm_{N_k}^{Tx} = \int_X f \circ T dm_{N_k}^x \xrightarrow{k \rightarrow \infty} \int_X f \circ T d\mu,$$

la convergence étant garantie par le fait que  $T$ , et donc  $f \circ T$ , sont continues. On en déduit ainsi que toute valeur d'adhérence  $\mu$  de la suite des mesures empiriques  $(m_N^x)_{N \geq 1}$  satisfait, pour toute  $f \in C(X)$

$$\int_X f d\mu = \int_X f \circ T d\mu.$$

Il s'en suit que  $\mu$  est forcément  $T$ -invariante :  $\mu = T_*(\mu)$ .

Avec l'argument ci-dessus, on peut se convaincre d'une part que les mesures de probabilité sur  $X$  qui sont invariantes par  $T$  contrôlent le comportement asymptotique des moyennes temporelles le long des orbites de la transformation. D'autre part, on obtient comme corollaire immédiat le théorème suivant qui assure de l'existence d'au moins une mesure de probabilité invariante dans un système dynamique topologique.

**Théorème 1.1** (Kryloff-Bogoliouboff, [3]). *Si  $T$  est une transformation continue d'un espace métrique compact  $X$ , alors il existe au moins une mesure de probabilité sur  $X$  qui est invariante par  $T$ .*

Notons que si on retire l'hypothèse de compacité de  $X$ , la conclusion peut ne plus avoir lieu : par exemple on peut prendre comme espace d'états  $X = \mathbb{Z}$ , et considérer la transformation  $T : \mathbb{Z} \rightarrow \mathbb{Z}$  définie par

$$\forall x \in \mathbb{Z}, Tx := x + 1.$$

Cette transformation  $T$  est bien continue sur  $\mathbb{Z}$ , mais une mesure sur  $\mathbb{Z}$  est  $T$ -invariante si et seulement si elle donne la même masse à tous les points de  $\mathbb{Z}$ . Soit cette masse est nulle, et alors la mesure est nulle, soit elle est strictement positive et alors la masse totale de l'espace est infinie. Il n'y a donc pas de mesure de probabilité  $T$ -invariante sur  $\mathbb{Z}$ .

Dans certains systèmes dynamiques topologiques, il n'existe qu'une seule mesure de probabilité invariante. On dit dans ce cas que ce système est *uniquement ergodique*. Dans d'autre cas l'ensemble des mesures de probabilités invariantes peut être extrêmement riche. L'identification de toutes les mesures de probabilité  $T$ -invariantes dans un système topologique donné est une question fondamentale qui peut parfois s'avérer très difficile. Mais dans ce cours nous nous concentrerons principalement sur l'étude de systèmes dynamiques pour lesquels une mesure invariante est déjà fixée.

## 2 Système dynamique mesuré

### 2.1 Espace Borel-standard, espace de Lebesgue

Dans la suite, on se place dans un cadre un peu plus général que celui d'un espace d'états métrisable compact : on suppose que  $X$  est un espace

mesurable dit *Borel-standard*. Cela signifie que  $X$  est muni d'une tribu  $\mathcal{B}(X)$  qui est la tribu borélienne (c'est-à-dire : engendrée par les ouverts) d'une topologie sur  $X$  qui en fait un espace *polonais* (c'est-à-dire : métrisable, séparable, complet).

Cette hypothèse structurelle sur l'espace d'états  $X$  constitue un excellent compromis entre la généralité qu'elle englobe et les avantages qu'elle procure. En effet, la très grande majorité des espaces couramment utilisés sont Borel-standard lorsqu'ils sont munis de leur tribu des boréliens pour leur topologie usuelle : c'est le cas bien sûr de  $\mathbb{R}$ ,  $\mathbb{R}^n$ , tout sous-intervalle de  $\mathbb{R}$ , tout espace de Banach séparable (par exemple l'espace des fonctions continues de  $[0, 1]$  dans  $\mathbb{C}$ ). De plus, tout produit dénombrable d'espaces Borel-standard (avec sa tribu produit) est lui-même Borel-standard. Enfin, si  $(X, \mathcal{B}(X))$  est Borel-standard, pour toute partie  $Y \in \mathcal{B}(X)$  l'espace  $(Y, \mathcal{B}(Y))$ , où  $\mathcal{B}(Y) := \{A \in \mathcal{B}(X) : A \subset Y\}$  est lui aussi Borel-standard. D'un autre côté, tout espace Borel-standard  $(X, \mathcal{B}(X))$  bénéficie de propriétés pratiques, parmi lesquelles on peut citer dès maintenant les deux qui suivent (voir par exemple [2]).

**Propriété 1.** *Si  $(X, \mathcal{B}(X))$  et  $(Y, \mathcal{B}(Y))$  sont deux espaces Borel-standard et  $f : X \rightarrow Y$  est une application mesurable injective, alors  $f(X) \in \mathcal{B}(Y)$ . Plus généralement, si  $f : X \rightarrow Y$  est une application mesurable et si  $A \in \mathcal{B}(X)$  est tel que  $f|_A$  est injective, alors  $f(A) \in \mathcal{B}(Y)$ .*

En particulier, il en découle que si  $f : (X, \mathcal{B}(X)) \rightarrow (Y, \mathcal{B}(Y))$  est une bijection mesurable entre deux espaces Borel-standard, alors son inverse est mesurable également.

**Propriété 2.** *Par la séparabilité de la topologie qui engendre la tribu, il existe dans tout espace Borel-standard  $(X, \mathcal{B}(X))$  une famille dénombrable  $(B_n)_{n \in \mathbb{N}} \subset \mathcal{B}(X)$  qui sépare les points : si  $x$  et  $y$  sont deux points distincts dans  $X$ , alors il existe  $n \in \mathbb{N}$  tel que  $\mathbb{1}_{B_n}(x) \neq \mathbb{1}_{B_n}(y)$ . Toute famille dénombrable vérifiant cette propriété engendre la tribu  $\mathcal{B}(X)$ .*

Si on considère une mesure de probabilité  $\mu$  sur  $(X, \mathcal{B}(X))$ , le triplet  $(X, \mathcal{B}(X), \mu)$  est alors un *espace probabilisé Borel-standard*, aussi appelé *espace de Lebesgue*. Ce type d'espaces a été introduit par V.A. Rokhlin, qui les a étudiés dans un article fondamental [4]. Le terme « espace de Lebesgue » (on dit aussi « espace de Lebesgue-Rokhlin ») se justifie par le résultat qui suit.

**Théorème 2.1** (Theorem 17.41 dans [2]). *Si la mesure de probabilité  $\mu$  sur l'espace Borel-standard  $(X, \mathcal{B}(X))$  est non-atomique, alors  $(X, \mathcal{B}(X), \mu)$  est isomorphe à l'espace probabilisé  $([0, 1], \mathcal{B}([0, 1]), \lambda)$  (où  $\lambda$  désigne la mesure de Lebesgue sur l'intervalle  $[0, 1]$ ) : il existe une bijection mesurable  $\Phi : X \rightarrow [0, 1]$  telle que  $\Phi_*(\mu) = \lambda$ .*

Notre espace étant à présent muni d'une mesure de probabilité, nous nous intéresserons à des propriétés pour lesquelles ce qui se passe sur des ensembles de mesure nulle ne joue aucun rôle. Nous introduisons notamment les notations suivantes qui concernent les sous-tribus de  $\mathcal{B}(X)$  : si  $\mathcal{C}$  et  $\mathcal{D}$  sont deux sous-tribus de  $\mathcal{B}(X)$ , on dira que «  $\mathcal{C}$  est contenue dans  $\mathcal{D}$  modulo  $\mu$  », et on notera  $\mathcal{C} \stackrel{\mu}{\subset} \mathcal{D}$  si pour tout  $C \in \mathcal{C}$ , il existe  $D \in \mathcal{D}$  tel que  $\mu(C \Delta D) = 0$ . De même on dira que «  $\mathcal{C}$  est égale à  $\mathcal{D}$  modulo  $\mu$  », et on notera  $\mathcal{C} \stackrel{\mu}{=} \mathcal{D}$  si on a à la fois  $\mathcal{C} \stackrel{\mu}{\subset} \mathcal{D}$  et  $\mathcal{D} \stackrel{\mu}{\subset} \mathcal{C}$ . En pratique, on ne distinguera pas deux sous-tribus qui sont égales modulo la mesure de probabilité sous-jacente.

La propriété ci-dessous sera souvent utile lorsqu'il faudra approcher des ensembles boréliens quelconques par des ensembles dont la structure est plus simple.

**Propriété 3.** *Soit  $(X, \mathcal{B}(X), \mu)$  un espace de Lebesgue, et soit  $(B_n)_{n \in \mathbb{N}} \subset \mathcal{B}(X)$  une famille dénombrable de parties mesurables de  $X$ . On note*

$$\mathcal{A} := \bigcup_{N \in \mathbb{N}} \sigma(B_0, \dots, B_N)$$

*l'algèbre engendrée par les  $B_n$  (qui reste dénombrable).*

*Les propriétés suivantes sont équivalentes :*

- (i) *Il existe  $X_0 \in \mathcal{B}(X)$  avec  $\mu(X_0) = 1$ , tel que la famille  $(B_n)_{n \in \mathbb{N}}$  sépare les points sur  $X_0$ .*
- (ii) *Pour tout  $A \in \mathcal{B}(X)$  et tout  $\varepsilon > 0$ , il existe  $\tilde{A} \in \mathcal{A}$  tel que  $\mu(A \Delta \tilde{A}) < \varepsilon$ .*
- (iii) *On a  $\sigma\{B_n : n \in \mathbb{N}\} \stackrel{\mu}{=} \mathcal{B}(X)$ .*

*Idées pour la preuve.* L'implication (i)  $\Rightarrow$  (iii) est une conséquence de la propriété 2.

Pour l'implication réciproque (iii)  $\Rightarrow$  (i), on suppose que la famille  $(B_n)$  satisfait (iii) et on introduit la relation d'équivalence sur  $X$  :

$$x \sim y \iff \forall n \in \mathbb{N}, \mathbb{1}_{B_n}(x) = \mathbb{1}_{B_n}(y).$$

La partition en classes d'équivalences est ce que l'on appelle une *partition mesurable* de  $(X, \mathcal{B}(X))$ . On prouve ensuite que tout  $C \in \sigma\{B_n : n \in \mathbb{N}\}$  est saturé pour cette relation d'équivalence, c'est-à-dire que si  $x \in C$  et si  $x \sim y$ , alors  $y \in C$  (en effet, les parties saturées forment une tribu qui contient tous les  $B_n$ ). Soit alors  $(D_n)_{n \in \mathbb{N}}$  une famille dénombrable dans  $\mathcal{B}(X)$  qui sépare les points. D'après la propriété (iii), pour tout  $n$  il existe  $C_n \in \sigma\{B_n : n \in \mathbb{N}\}$  tel que  $\mu(C_n \Delta D_n) = 0$ . Posons

$$X_0 := X \setminus \bigcup_{n \in \mathbb{N}} (C_n \Delta D_n).$$

On a alors  $\mu(X_0) = 1$ , et pour tout  $n \in \mathbb{N}$ ,  $C_n \cap X_0 = D_n \cap X_0$ . Chaque  $(C_n \cap X_0)$  étant saturé pour la restriction de la relation d'équivalence à  $X_0$ , il en va de même pour chaque  $(D_n \cap X_0)$ . Mais comme les  $D_n$  séparent les points, on a pour tous  $x, y \in X_0$

$$x \sim y \Rightarrow [\forall n \in \mathbb{N}, \mathbb{1}_{D_n}(x) = \mathbb{1}_{D_n}(y)] \Rightarrow x = y.$$

Ce qui prouve que les  $B_n$  séparent les points sur  $X_0$ .

L'implication (ii)  $\Rightarrow$  (iii) peut se montrer ainsi : soit  $C \in \mathcal{B}(X)$ , alors si (ii) est satisfait, pour tout entier  $k \geq 1$  on peut trouver  $C_k \in \sigma\{B_n : n \in \mathbb{N}\}$  tel que  $\mu(C \Delta C_k) < 2^{-k}$ . Par Borel-Cantelli,  $\mu$ -presque tout  $x$  n'appartient qu'à un nombre fini d'ensembles  $C \Delta C_k$ . En posant

$$\tilde{C} := \bigcup_{K \geq 1} \bigcap_{k \geq K} C_k \in \sigma\{B_n : n \in \mathbb{N}\},$$

on a alors  $\mu(C \Delta \tilde{C}) = 0$ .

Enfin, pour la dernière implication (iii)  $\Rightarrow$  (ii), on peut remarquer que l'ensemble  $\mathcal{C}$  des  $A \in \mathcal{B}(X)$  qui vérifient la conclusion de (ii) (c'est-à-dire qui peuvent être arbitrairement bien approchés par un ensemble dans l'algèbre engendrée les  $B_n$ ) est une classe monotone, qui contient le  $\pi$ -système  $\mathcal{A}$ . Par le théorème des classes monotones de Sierpiński (voir par exemple [1]),  $\mathcal{C}$  contient la tribu  $\sigma\{B_n : n \in \mathbb{N}\}$ .  $\square$



## 2.2 Système dynamique mesuré

Dans la suite, on s'intéressera donc à des systèmes dynamiques pour lesquels l'espace d'états  $(X, \mathcal{B}(X))$  est Borel-standard, muni d'une mesure de probabilité  $\mu$  qui en fait un espace de Lebesgue, et d'une transformation mesurable  $T : X \rightarrow X$ , bijective (d'inverse automatiquement mesurable par la propriété 1), et qui préserve la mesure  $\mu$  : pour tout  $A \in \mathcal{B}(X)$ , on a  $\mu(T^{-1}A) = \mu(A) = \mu(TA)$ . Un tel système est appelé *système dynamique mesuré*. On le désigne par le quadruplet  $(X, \mathcal{B}(X), \mu, T)$ , ou plus simplement par le triplet  $(X, \mu, T)$  (la tribu borélienne étant implicite).

## 3 Premiers exemples

### 3.1 Permutations sur un ensemble fini

Le premier exemple présenté ici, bien qu'assez trivial du point de vue de la dynamique, permet néanmoins d'illustrer de manière élémentaire des concepts essentiels, mais en général plus subtils, dans la théorie ergodique. L'espace d'états  $X$  est ici un ensemble fini (par exemple,  $X = \{1, \dots, k\}$  pour un certain entier naturel  $k \geq 1$ ), et  $T : X \rightarrow X$  est une bijection, autrement dit une permutation des éléments de  $X$ . La mesure de probabilité uniforme  $\mu$  sur  $X$ , qui donne une masse  $1/k$  à chacun des  $k$  états dans  $X$ , est clairement invariante par  $T$ .

L'orbite d'un état quelconque  $x \in X$ , c'est-à-dire l'ensemble  $\{T^k x : k \in \mathbb{Z}\}$  est évidemment une partie finie de  $X$ , et  $T$  agit de manière cyclique sur cette orbite : si on note  $\ell$  le nombre d'éléments dans l'orbite de  $x$ , on peut écrire les éléments de cette orbite sous la forme  $\{x_1, \dots, x_\ell\}$ , où pour  $i = 1, \dots, \ell - 1$ , on a  $Tx_i = x_{i+1}$ , et  $Tx_\ell = x_1$ . Si on fixe une observable  $f : X \rightarrow \mathbb{C}$ , il est alors facile de se convaincre que les moyennes temporelles de  $f$  en partant de  $x$  convergent en temps long vers l'intégrale de  $f$  pour la mesure uniforme sur l'orbite de  $x$  :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \frac{1}{\ell} \sum_{j=1}^{\ell} f(x_j).$$

La relation définie par

$$x \sim y \iff y \text{ est dans l'orbite de } x$$

étant une relation d'équivalence, les différentes orbites de la permutation  $T$  partitionnent l'espace d'états. On peut alors distinguer deux cas (voir Figure 1).

**Cas 1.** Si tous les éléments de  $X$  sont dans la même orbite, c'est-à-dire si  $T$  agit cycliquement sur  $X$  tout entier, alors pour tout  $x \in X$  on aura, pour toute observable  $f : X \rightarrow \mathbb{C}$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f d\mu.$$

Dans ce premier cas, la permutation  $T$  est dite *ergodique* (voir la définition 3.2 dans le cas général).

**Cas 2.** En revanche, si il existe plusieurs orbites disjointes, la limite des moyennes temporelles dépendra de l'état initial  $x$  choisi. Cependant cette limite s'écrit toujours comme l'intégrale de  $f$  par rapport à une mesure uniforme, dont le support (l'orbite de  $x$ ) est une partie de  $X$  sur laquelle  $T$  agit cycliquement. Dans ce deuxième cas,  $T$  n'est pas ergodique, mais la partition en cycles correspond à ce que l'on appellera en général la *décomposition en composantes ergodiques* du système.

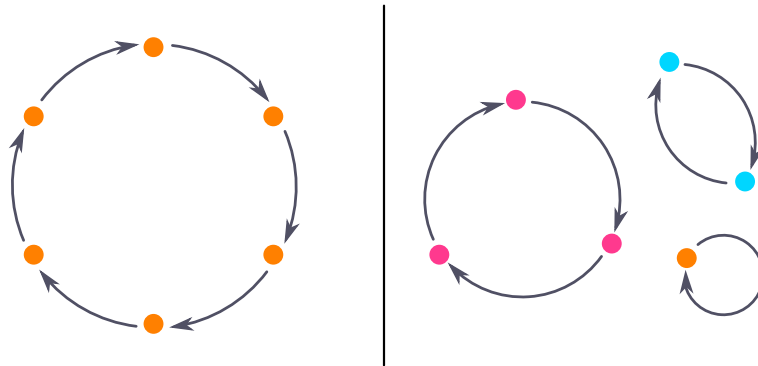


FIGURE 1 – Deux exemples de permutations sur un ensemble d'états à six éléments, illustrant le cas 1 (ergodique, à gauche) et le cas 2 (non ergodique, à droite).

### 3.2 Rotations du cercle

Soit  $\alpha$  un nombre réel. Sur le cercle unité du plan complexe, on considère la rotation  $R_\alpha$  d'angle  $2\pi\alpha$  :  $z \mapsto e^{i2\pi\alpha}z$ . Cette transformation préserve clairement la mesure de Lebesgue normalisée sur le cercle (notée  $\lambda$ ), mais on peut deviner que les propriétés du système dynamique ainsi obtenu vont dépendre de la valeur de  $\alpha$ . En particulier, le fait que  $\alpha$  soit rationnel ou non joue un rôle fondamental : si  $\alpha \in \mathbb{Q}$ , la rotation  $R_\alpha$  est périodique et l'orbite de tout point  $z$  du cercle est formée des sommets d'un polygone régulier. Au contraire, si  $\alpha \notin \mathbb{Q}$ , on peut imaginer que les points de l'orbite de  $z$  se répartissent uniformément sur le cercle (voir Figure 2). Le théorème qui suit précise cette idée.

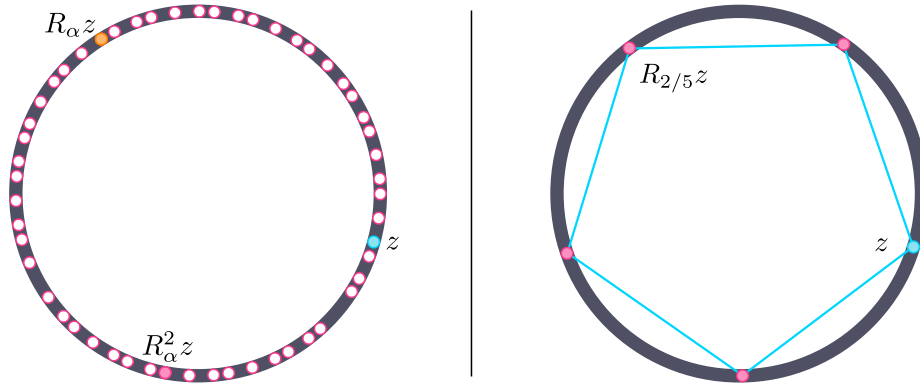


FIGURE 2 – Un point  $z$  sur le cercle unité, et ses premières images successives par une rotation  $R_\alpha$ . À gauche,  $\alpha = \frac{3-\sqrt{5}}{2} \notin \mathbb{Q}$ . À droite,  $\alpha = 2/5 \in \mathbb{Q}$ .

**Théorème 3.1** (Équirépartition des points de l'orbite d'une rotation irrationnelle). *Soit  $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ . Pour tout point  $z$  du cercle, l'ensemble des points  $R_\alpha^k(z)$ ,  $k \in \mathbb{N}$  est dense dans le cercle. De plus, pour tout arc de cercle  $A$ , on a*

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_A(R_\alpha^n z) \xrightarrow{N \rightarrow \infty} \lambda(A). \quad (2)$$

*Plan d'une preuve.* Voici quelques indications pour prouver ce théorème de manière complètement élémentaire :

1. Vérifier que,  $\alpha$  étant irrationnel, les points  $R_\alpha^k(z)$ ,  $k \in \mathbb{N}$  sont deux à deux distincts.

2. En déduire que, étant donné  $\varepsilon > 0$ , il existe un réel  $\beta$  avec  $0 < |\beta| < \varepsilon$ , et  $m \geq 1$  tel que  $R_\alpha^m = R_\beta$ . (Utiliser le principe des tiroirs, en partitionnant le cercle en arcs de longueurs inférieures à  $\varepsilon$ .)
3. Montrer que, pour le  $\beta$  trouvé ci-dessus, pour tout arc de cercle  $A$  et tout point  $z$  du cercle,

$$\lambda(A) - \varepsilon < \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_A(R_\beta^n z) \leq \limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_A(R_\beta^n z) < \lambda(A) + \varepsilon. \quad (3)$$

Pour cela, vérifier que le nombre de points consécutifs de l'orbite de  $z$  pour  $R_\beta$  qui tombent dans  $A$  est toujours compris entre  $\lambda(A)/|\beta| - 1$  et  $\lambda(A)/|\beta| + 1$ .

4. Montrer que l'orbite d'un point  $z$  sous l'action de  $R_\alpha$  est la réunion des orbites de  $m$  points du cercle sous l'action de  $R_\beta$ , en regroupant les points  $R_\alpha^k(z)$  suivant le reste de la division euclidienne de  $k$  par  $m$ . En déduire que (3) reste vrai si on remplace  $\beta$  par  $\alpha$ , et conclure.

□

La convergence (2) reste évidemment vraie si  $A$  est une union finie d'arcs de cercle. Mais que se passe-t-il pour un borélien quelconque ? On ne peut certainement pas espérer la validité de (2) pour *tout* point  $z$  sans imposer de condition sur  $A$ . En effet, prenons pour  $A$  un arc de cercle d'intérieur non vide privé de tous les points de l'orbite d'un  $z_0$  fixé. Alors la mesure de  $A$  est la même que celle de l'arc de cercle, donc strictement positive, car on n'a retiré qu'un nombre dénombrable de points. Mais par définition de  $A$ ,  $R_\alpha^n z_0$  ne tombe jamais dans  $A$  et donc on a toujours

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}_A(R_\alpha^n z_0) = 0.$$

En fait, la convergence (2) continue à être valide pour un borélien quelconque, mais seulement pour  $\lambda$ -presque tout  $z$  dans le cercle, le négligeable en-dehors duquel la convergence a lieu dépendant de l'ensemble  $A$  considéré. Ce résultat est un cas particulier du théorème ergodique ponctuel de Birkhoff, qui sera abordé un peu plus tard.

Que se passe-t-il dans le cas où  $\alpha$  est rationnel ? Clairement, (2) n'est plus vrai même à un ensemble négligeable près, puisqu'alors il existe des

arcs de cercle  $A$  de mesure  $\lambda(A) > 0$  et un ensemble de points  $z$  de mesure strictement positive dont l'orbite ne rencontre jamais  $A$ . Cela est dû à l'existence de partitions du cercle en deux parties de mesure strictement positive *invariantes* par  $R_\alpha$ , c'est-à-dire telles que  $A = R_\alpha^{-1}(A)$ .

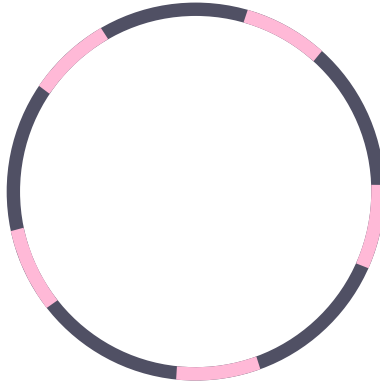


FIGURE 3 – Une partition du cercle en deux ensembles invariants pour  $R_{2/5}$

Ceci mène à la définition de la première propriété importante d'un système dynamique.

**Définition 3.2** (Ergodicité). *Le système  $(X, \mu, T)$  est dit ergodique si les seules parties  $A \in \mathcal{B}(X)$  qui sont invariantes par  $T$  sont de mesure 0 ou 1.*

**Théorème 3.3.** *Si  $\alpha$  est irrationnel, la rotation  $R_\alpha$  est ergodique sur le cercle muni de la mesure de Lebesgue.*

*Idée d'une preuve.* (Voir [5], p. 22.) Il s'agit de montrer que si  $A$  est une partie mesurable  $R_\alpha$ -invariante telle que  $\lambda(A) > 0$ , alors  $\lambda(A) = 1$ . Or, étant donné  $\varepsilon > 0$ , puisque  $\lambda(A) > 0$  on peut toujours trouver un arc de cercle  $I$  de mesure  $\lambda(I) < \varepsilon$  et tel que  $\lambda(A \cap I) > (1 - \varepsilon)\lambda(I)$ . Puis, en utilisant le fait que  $A$  est  $R_\alpha$ -invariant, on vérifie que  $A$  remplit toujours une proportion au moins  $1 - \varepsilon$  de n'importe quel  $R_\alpha^k I$  pour  $k \in \mathbb{Z}$ . Enfin, en utilisant la densité de l'orbite d'une des bornes de  $I$  sous l'action de  $R_\alpha$ , il est facile de trouver des entiers  $k_1 < k_2 < \dots < k_\ell$  tels que les arcs de cercle  $R_\alpha^{k_j} I$ ,  $1 \leq j \leq \ell$  soient deux à deux disjoints et recouvrent une proportion au moins  $1 - 2\varepsilon$  du cercle. On en déduit

$$\lambda(A) > (1 - \varepsilon)(1 - 2\varepsilon),$$

et puisque  $\varepsilon$  est arbitraire,  $\lambda(A) = 1$ . □

### 3.3 La transformation du boulanger

L'espace d'états est ici constitué du carré  $[0, 1) \times [0, 1)$ , et la transformation évoque le pétrissage de la pâte par le boulanger. On la décrit en trois étapes :

1. on « étale » le carré pour en faire un rectangle de côtés 2 et 1/2
2. on coupe le rectangle en deux,
3. on replace le second morceau au-dessus du premier pour reconstituer un carré de côté 1.

Aucune de ces opérations ne change la surface d'une partie  $A$  quelconque du carré, et la mesure de Lebesgue est donc préservée par  $T$ .

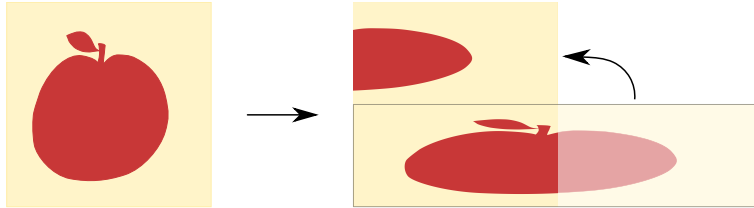


FIGURE 4 – La transformation du boulanger agissant sur le carré

Plus formellement, la transformation du boulanger peut être décrite par la formule qui suit : pour tout  $x = (u, v) \in [0, 1) \times [0, 1)$ ,

$$T(x) := \begin{cases} (2u, v/2) & \text{si } u < 1/2, \\ (2u - 1, v/2 + 1/2) & \text{si } u \geq 1/2. \end{cases} \quad (4)$$

En itérant un grand nombre de fois la transformation du boulanger, on observe que les points d'une partie  $B$  donnée au départ semblent se répartir uniformément dans tout le carré : si  $B$  est de surface  $2/5$ , et si on se fixe une fenêtre  $A$  dans le carré, les points de  $T^n B$  qui tombent dans  $A$  occupent environ  $2/5$  de  $A$  si  $n$  est grand. Cette propriété porte le nom de *mélange*.

**Définition 3.4** (Mélange). *Le système  $(X, \mu, T)$  est mélangeant si pour tous  $A$  et  $B$  dans  $\mathcal{B}(X)$ ,*

$$\mu(A \cap T^n B) \xrightarrow[n \rightarrow \infty]{} \mu(A)\mu(B). \quad (5)$$

**Exercice 3.1.** *Montrer que le mélange entraîne l'ergodicité.*

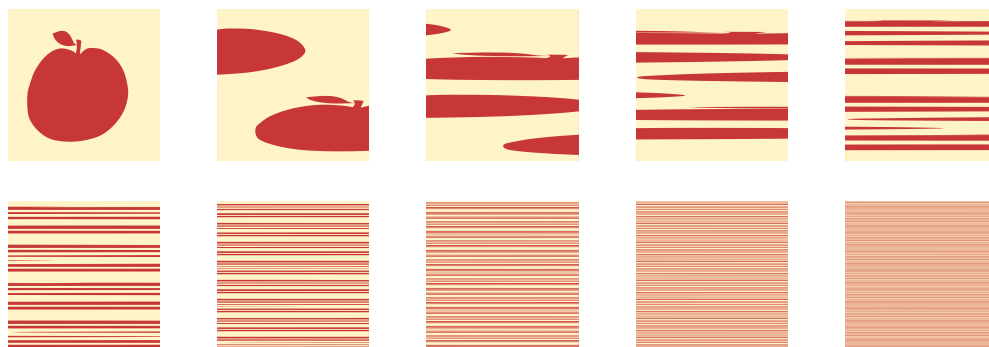


FIGURE 5 – Une partie  $B$  du carré et ses images successives  $T(B), \dots, T^9(B)$  par la transformation du boulanger

Le mélange pour la transformation du boulanger sera justifié dans la suite. Notons que les rotations du cercle ne sont certainement pas mélangeantes, car après un nombre quelconque d'itérations, un arc de cercle reste toujours un arc de cercle. En fait, les rotations du cercle vérifient une propriété très forte qui empêche le mélange.

**Définition 3.5** (Rigidité). *Le système  $(X, \mu, T)$  est rigide s'il existe une suite strictement croissante d'entiers  $(n_k)$  telle que pour  $A \in \mathcal{B}(X)$ ,*

$$\mu(A \Delta T^{n_k} A) \xrightarrow[k \rightarrow \infty]{} 0.$$

**Exercice 3.2.** *Vérifier que la rigidité empêche le mélange.*

**Exercice 3.3.** *Montrer qu'une rotation du cercle est toujours rigide.*

### 3.4 Système construit à partir d'un processus stochastique stationnaire, schéma de Bernoulli

Sur un espace probabilisé  $(\Omega, \mathbb{P})$ , on considère un processus stochastique stationnaire  $\xi = (\xi_n)_{n \in \mathbb{Z}}$ , où les variables aléatoires  $\xi_n$  sont à valeurs dans un espace Borel-standard  $(E, \mathcal{B}(E))$ . Cet espace  $E$  est souvent un ensemble fini, mais il peut être aussi (une partie de)  $\mathbb{R}$  ou  $\mathbb{C}$ , ou un espace plus abstrait. La *stationnarité* du processus s'exprime par la propriété suivante : pour tout  $n \geq 1$  et tout  $m \in \mathbb{Z}$ ,

$$\mathcal{L}(\xi_m, \xi_{m+1}, \dots, \xi_{m+n-1}) = \mathcal{L}(\xi_0, \xi_1, \dots, \xi_{n-1}),$$

où la notation  $\mathcal{L}(\xi_m, \dots, \xi_{m+n-1})$  désigne la loi du  $n$ -uplet  $(\xi_m, \dots, \xi_{m+n-1})$ , c'est-à-dire la mesure de probabilité définie sur  $E^n$  comme la mesure image de  $\mathbb{P}$  par l'application

$$\omega \mapsto (\xi_m(\omega), \dots, \xi_{m+n-1}(\omega)).$$

On peut de même définir la loi  $\mu$  du processus global  $\xi$ , qui est cette fois une mesure de probabilité sur  $E^{\mathbb{Z}}$ . Une autre façon d'exprimer la stationnarité du processus est de dire que la mesure  $\mu$  est invariante par la transformation  $S : E^{\mathbb{Z}} \rightarrow E^{\mathbb{Z}}$  qui consiste à décaler toutes les coordonnées d'une suite d'un cran vers la gauche : si  $x = (x_n)_{n \in \mathbb{Z}} \in E^{\mathbb{Z}}$ ,  $Sx$  est la suite  $y = (y_n)_{n \in \mathbb{Z}}$  définie par  $y_n := x_{n+1}$  pour tout  $n$ . Ainsi, à partir du processus stationnaire  $\xi$  on peut construire un système dynamique mesuré de la forme  $(E^{\mathbb{Z}}, \mu, S)$  où  $\mu = \mathcal{L}(\xi)$ .

Inversement, si  $(X, \mu, T)$  est un système dynamique mesuré, et si on se donne une fonction mesurable  $f : X \rightarrow E$  où  $(E, \mathcal{B}(E))$  est un espace Borel-standard, on peut définir sur l'espace probabilisé  $(X, \mu)$  un processus  $\xi = (\xi_n)_{n \in \mathbb{Z}}$  en posant pour tout  $n \in \mathbb{Z}$ ,  $\xi_n := \xi_0 \circ T^n$ . De l'invariance de  $\mu$  par  $T$ , on déduit que tout processus  $\xi$  ainsi construit est un système stationnaire.

Revenons au système  $(E^{\mathbb{Z}}, \mu, S)$  construit à partir d'un processus stationnaire abstrait  $\xi$ . Lorsque les coordonnées  $\xi_n$  de ce processus sont indépendantes et identiquement distribuées (i.i.d.), on a  $\mu = \mu_0^{\otimes \mathbb{Z}}$ , où  $\mu_0 = \mathcal{L}(\xi_0)$ . Le système dynamique  $(E^{\mathbb{Z}}, \mu_0^{\otimes \mathbb{Z}}, S)$  est alors appelé un *schéma de Bernoulli*.

Si de plus  $E$  est un ensemble fini, disons  $E = \{e_1, \dots, e_k\}$  pour un certain entier  $k \geq 1$ , on peut noter  $p_j := \mathbb{P}(\xi_0 = e_j)$  pour tout  $j = 1, \dots, k$ , et la loi  $\mu_0$  s'écrit alors comme une moyenne pondérée de masses de Dirac :  $\mu_0 = \sum_{j=1}^k p_j \delta_{e_j}$ . Le schéma de Bernoulli

$$\left( E^{\mathbb{Z}}, \left( \sum_{j=1}^k p_j \delta_{e_j} \right)^{\otimes \mathbb{Z}}, S \right)$$

sera alors noté  $\mathcal{B}(p_1, \dots, p_k)$ .

**Remarque 3.6.** *Le système dynamique mesuré construit de cette façon est associé à l'évolution d'un processus stochastique  $\xi$ , pourtant il rentre bien dans le cadre des systèmes dynamiques dits déterministes, tels qu'on les a définis au début de ce chapitre. Ce qui peut paraître comme un paradoxe*



s'explique par le choix de l'espace d'états : ici un « état » du système encode l'intégralité de la trajectoire du processus stochastique.

Ainsi, même dans un système dynamique déterministe, les trajectoires observées peuvent être totalement imprévisibles, car l'observation n'apporte en général qu'une information très partielle sur l'état du système (par exemple : seulement la coordonnée 0 du processus).

Le théorème qui suit est le premier résultat important sur les schémas de Bernoulli.

**Théorème 3.7.** *Un schéma de Bernoulli est toujours mélangeant (et, par conséquent, ergodique).*

*Idée d'une preuve.* On doit vérifier la propriété (5) pour tous  $A$  et  $B$  dans  $\mathcal{B}(E^{\mathbb{Z}})$ . Considérons d'abord le cas où  $A$  et  $B$  sont mesurables par rapport à un nombre fini de coordonnées dans  $E^{\mathbb{Z}}$ , disons  $\sigma(x_{-k}, \dots, x_k)$ -mesurables. Alors  $T^n A$  est mesurable par rapport à  $\sigma(x_{-k-n}, \dots, x_{k-n})$ . Alors, si  $n$  est assez grand ( $n > k$ ),  $T^n A$  est indépendant de  $B$ . Puis, on utilise le fait que tout  $A$  dans  $\mathcal{B}(E^{\mathbb{Z}})$  peut être approché arbitrairement bien par un ensemble  $\tilde{A}$  mesurable par rapport à un nombre fini de coordonnées, au sens où  $\mu(A \Delta \tilde{A}) < \varepsilon$  pour un certain  $\varepsilon > 0$  fixé à l'avance (cf. *Propriété 3*). En approchant de même  $B$  par  $\tilde{B}$ , on peut écrire

$$\begin{aligned} \left| \mu(A \cap T^n B) - \mu(A)\mu(B) \right| &\leq \left| \mu(A \cap T^n B) - \mu(\tilde{A} \cap T^n B) \right| \\ &+ \left| \mu(\tilde{A} \cap T^n B) - \mu(\tilde{A} \cap T^n \tilde{B}) \right| \\ &+ \left| \mu(\tilde{A} \cap T^n \tilde{B}) - \mu(\tilde{A})\mu(\tilde{B}) \right| \\ &+ \left| \mu(\tilde{A})\mu(\tilde{B}) - \mu(\tilde{A})\mu(B) \right| \\ &+ \left| \mu(\tilde{A})\mu(B) - \mu(A)\mu(B) \right|. \end{aligned}$$

Si  $n$  est assez grand pour que  $\tilde{B}$  et  $T^n \tilde{A}$  soient indépendants, le troisième terme de la somme ci-dessus est nul, et les 4 autres sont majorés chacun par  $\varepsilon$  par choix de  $\tilde{A}$  et  $\tilde{B}$ .  $\square$

## 4 Isomorphisme et codage de systèmes dynamiques mesurés

### 4.1 De la transformation du boulanger au schéma de Bernoulli

Reprenons l'exemple de la transformation du boulanger, définie sur  $X := [0, 1) \times [0, 1)$  (voir la section 3.3). À un point  $x = (u, v)$  de  $X$ , faisons correspondre la suite doublement infinie

$$\varphi(x) := (\dots, y_{-3}, y_{-2}, y_{-1}, y_0, y_1, y_2 \dots) \in Y := \{0, 1\}^{\mathbb{Z}},$$

où  $0, y_0 y_1 y_2 \dots$  et  $0, y_{-1} y_{-2} y_{-3} \dots$  sont les développements en base 2 de  $u$  et  $v$  respectivement. On vérifie assez facilement que, par construction du développement en base 2, si  $x = (u, v)$  est distribué suivant la mesure de Lebesgue 2-dimensionnelle sur  $X$  notée  $\mu$ , les coordonnées  $y_j$  sont toutes indépendantes et équidistribuées, elles valent 0 ou 1 avec probabilité  $1/2$ . Autrement dit, si  $\nu := \varphi_*(\mu)$ ,  $\nu$  est la mesure produit  $(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes \mathbb{Z}}$ . En notant  $S$  le décalage des coordonnées sur  $Y$ ,  $(Y, \nu, S)$  est donc le schéma de Bernoulli  $\mathcal{B}(1/2, 1/2)$ . Par ailleurs, on obtient facilement à partir de (4) que pour tout  $x \in X$ ,

$$\varphi \circ T(x) = S \circ \varphi(x),$$

*i.e.* le diagramme suivant est commutatif :

$$\begin{array}{ccc} (X, \mu) & \xrightarrow{T} & (X, \mu) \\ \varphi \downarrow & & \downarrow \varphi \\ (Y, \nu) & \xrightarrow{S} & (Y, \nu) \end{array}$$

Cette application  $\varphi : X \rightarrow Y$  réalise ce que l'on appelle un *isomorphisme* entre les systèmes mesurés  $(X, \mu, T)$  et  $(Y, \nu, S)$ , au sens précisé dans la définition ci-dessous.

**Définition 4.1** (Isomorphisme de systèmes mesurés). *Soient  $(X, \mu, T)$  et  $(Y, \nu, S)$  deux systèmes dynamiques mesurés. On dit qu'ils sont isomorphes si il existe  $X_0 \in \mathcal{B}(X)$  stable par  $T$ ,  $Y_0 \in \mathcal{B}(Y)$  stable par  $S$ , tels que  $\mu(X_0) = \nu(Y_0) = 1$ , et une bijection mesurable  $\varphi : X_0 \rightarrow Y_0$  qui vérifie*

- $\varphi_*(\mu) = \nu$ ;
- $\varphi \circ T = S \circ \varphi$ .

Une telle application  $\varphi$  est appelée isomorphisme de  $(X, \mu, T)$  vers  $(Y, \nu, S)$ .

**Remarque 4.2.** Par la propriété 1,  $\varphi^{-1} : Y_0 \rightarrow X_0$  est aussi mesurable et réalise donc un isomorphisme de  $(Y, \nu, S)$  vers  $(X, \mu, T)$ .

Les propriétés des systèmes dynamiques mesurés étudiées en théorie ergodique (par exemple, le mélange) sont insensibles à ce qui se passe sur une partie de mesure nulle de l'espace d'états. C'est la raison pour laquelle l'application  $\varphi$  n'a pas besoin d'être une bijection entre l'intégralité des ensembles  $X$  et  $Y$ . Et lorsqu'un isomorphisme existe entre deux systèmes, toute propriété vérifiée par l'un est automatiquement vérifiée par l'autre.

Les arguments développés juste avant permettent donc d'obtenir les résultats suivants.

**Théorème 4.3.** La transformation du boulanger et le schéma de Bernoulli  $\mathcal{B}(1/2, 1/2)$  sont isomorphes.

**Corollaire 4.4.** La transformation du boulanger est mélangeante.

**Exercice 4.1.** Soient  $p_1, \dots, p_k$  des réels positifs tels que  $\sum_{j=1}^k p_j = 1$ . Construire une généralisation de la transformation du boulanger isomorphe au schéma de Bernoulli  $\mathcal{B}(p_1, \dots, p_k)$ .

## 4.2 Codage symbolique d'un système dynamique

L'isomorphisme entre la transformation du boulanger et le schéma de Bernoulli décrit au paragraphe précédent est un cas particulier d'une procédure assez commune en théorie ergodique. Dans un système dynamique  $(X, \mu, T)$ , on se donne une partition mesurable, finie ou dénombrable,  $\mathcal{P} = (P_i)_{i \in E}$  de  $X$ , indexée par un ensemble fini ou dénombrable  $E$  (souvent appelé « alphabet »). Cette partition peut être identifiée à l'application mesurable (notée aussi  $\mathcal{P}$ ), qui à  $x \in X$  fait correspondre l'indice  $i \in E$  tel que  $x \in P_i$ . Dans l'exemple précédent de la transformation du boulanger, la partition finie utilisée est celle obtenue en coupant le carré en deux dans le sens vertical, qui correspond à l'application  $x = (u, v) \mapsto y_0 \in \{0, 1\}$ , où  $y_0$  est le premier chiffre de l'écriture en base 2 de  $u$ .

Appelons  $\mathcal{P}$ -nom d'un point  $x \in X$  la suite doublement infinie

$$\mathcal{P}|_{-\infty}^{+\infty}(x) := (\dots, \mathcal{P}(T^{-2}x), \mathcal{P}(T^{-1}x), \mathcal{P}(x), \mathcal{P}(Tx), \mathcal{P}(T^2x), \dots) \in E^{\mathbb{Z}}.$$

On peut alors considérer l'application  $\varphi^{\mathcal{P}} : X \rightarrow E^{\mathbb{Z}}$ , qui à  $x \in X$  fait correspondre son  $\mathcal{P}$ -nom. Il est immédiat que, tout comme l'application  $\phi$  définie dans le cas de la transformation du boulanger (qui en est un cas particulier), cette application vérifie

$$\varphi^{\mathcal{P}} \circ T = S \circ \varphi^{\mathcal{P}},$$

où  $S$  est le décalage des coordonnées sur  $E^{\mathbb{Z}}$ . Comme  $\varphi^{\mathcal{P}}$  est mesurable, on peut également définir la mesure image  $\nu^{\mathcal{P}} := \varphi_*^{\mathcal{P}}(\mu)$  sur  $E^{\mathbb{Z}}$  (c'est la loi du  $\mathcal{P}$ -nom de  $x$  lorsque  $x$  suit la loi  $\mu$ ). Puisque  $\mu$  est  $T$ -invariante, on obtient facilement que  $\nu^{\mathcal{P}}$  est  $S$ -invariante. On obtient ainsi un système dynamique mesuré dit *symbolique*  $(E^{\mathbb{Z}}, \nu^{\mathcal{P}}, S)$ , qui est vu dans le système  $(X, \mu, T)$  lorsque l'on observe  $\mathcal{P}$ . Mais, à la différence de ce que l'on a constaté dans l'exemple de  $\varphi$  pour la transformation du boulanger, rien ne dit *a priori* que l'application  $\varphi^{\mathcal{P}}$  est injective : en général, deux points différents de  $X$  peuvent avoir le même  $\mathcal{P}$ -nom. On dit que l'application  $\varphi^{\mathcal{P}}$  est une *application facteur* de  $(X, \mu, T)$  vers  $(E^{\mathbb{Z}}, \nu^{\mathcal{P}}, S)$ , au sens précisé dans la définition suivante.

**Définition 4.5** (Application facteur, facteur et extension). *Soient  $(X, \mu, T)$  et  $(Y, \nu, S)$  deux systèmes dynamiques mesurés. Une application facteur de  $(X, \mu, T)$  vers  $(Y, \nu, S)$  est une application mesurable  $\varphi$  à valeurs dans  $Y$ , définie sur un ensemble  $X_0 \in \mathcal{B}(X)$  stable par  $T$  et de mesure 1, et qui vérifie*

- $\varphi_*(\mu) = \nu$  ;
- $\varphi \circ T = S \circ \varphi$ .

*Lorsqu'une telle application existe, on dit que  $(Y, \nu, S)$  est un facteur de  $(X, \mu, T)$ , et que  $(X, \mu, T)$  est une extension de  $(Y, \nu, S)$ .*

Si il existe une partie  $X_0 \in \mathcal{B}(X)$  de mesure 1 sur laquelle l'application  $\varphi^{\mathcal{P}}$  est injective, alors  $\varphi^{\mathcal{P}}$  devient un isomorphisme entre  $(X, \mu, T)$  et le système symbolique  $(E^{\mathbb{Z}}, \nu^{\mathcal{P}}, S)$ . On dit dans ce cas que la partition  $\mathcal{P}$  est *génératrice* du système.

**Exercice 4.2.** *Montrer que dans le cas d'une rotation irrationnelle, une partition du cercle en deux arcs d'intérieur non vide est toujours génératrice.*

## Références

- [1] Olav Kallenberg, *Foundations of modern probability*, Probab. Appl., New York, NY : Springer, 1997.

- [2] A.S KECHRIS, *Classical descriptive set theory*, Springer-Verlag, 1995.
- [3] Nicolas Kryloff and Nicolas Bogoliouboff, *La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire*, Ann. Math. (2) **38** (1937), 65–113.
- [4] V.A. ROHLIN, *On the fundamental ideas of measure theory*, AMS Translations Serie 1 **10** (1963), 2–53, (Première publication en russe : 1949.).
- [5] P.C. SHIELDS, *The ergodic theory of discrete sample path*, American Mathematical Society, 1996.