Théorèmes ergodiques, ergodicité

T. de la Rue

août 2016

Nous considérons dans ce chapitre un système dynamique $(X, \mathscr{A}, \mu, T)$, où $(X, \mathscr{A}, \mu)$ est un espace probabilisé de Lebesgue, et T est une transformation mesurable de X dans lui-même préservant la mesure de probabilité μ. Pour une partie mesurable A de X, on cherche à estimer la proportion du temps passé dans A le long d'une orbite de T, c'est-à-dire étudier quand $n \to \infty$ le comportement des moyennes temporelles

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_A(T^k x).$$

Plus généralement, si f est une fonction mesurable sur X à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$, on définit pour tout entier $n \geq 1$ les sommes partielles

$$S_n(f) \; := \; \sum_{k=0}^{n-1} f(T^k x),$$

et on voudrait étudier l'éventuelle limite (en un sens à préciser) des moyennes temporelles $(1/n)S_n(f)$.

La sous-tribu des invariants

Comme nous le verrons dans ce chapitre, une sous-tribu particulière de $\mathscr{A}$ joue un rôle essentiel dans l'identification de la limite des moyennes temporelles : il s'agit de la sous-tribu des invariants, notée $\mathscr{I}$.

Il y a *a priori* plusieurs notions d'invariance possibles. Une partie mesurable A de X est dite *strictement (T-)invariante* si $T^{-1}A = A$. On dit également qu'elle est *presque (T-)invariante* si $\mu(T^{-1}A \vartriangle A) = 0$.

Exercice 0.1. *Prouver que pour tout $A \in \mathscr{A}$, les ensembles $\bigcap_{n\geq 0} T^{-n}A$ et $\bigcup_{n\geq 0} T^{-n}A$ sont presque T-invariants.*

Le lemme suivant prouve qu'il n'est pas important de distinguer entre les deux notions d'invariance stricte et de presque-invariance.

Lemme 0.1. *Si A est presque T-invariant, alors il existe $\tilde{A}$ qui est strictement T-invariant et tel que $\mu(\tilde{A} \vartriangle A) = 0$.*

Idée de la preuve. On commence par remarquer que $A_1 := A \cap T^{-1}A$ est inclus dans A et a même mesure que A. Puis par récurrence, $A_n := A \cap T^{-1}A \cap \cdots \cap T^{-n}A$ a toujours la même mesure que A. À la limite, $B := \bigcap_{n\geq 1} A_n$ est contenu dans A et a toujours la même mesure que A. En particulier, $\mu(B \vartriangle A) = 0$. De plus, on a $B \subset T^{-1}B$. On en déduit que la suite des $T^{-n}B$, $n \geq 1$, est croissante, et constituée de parties qui on toutes même mesure que B et A. En posant $\tilde{A} := \bigcup_{n\geq 1} T^{-n}B$, on obtient l'ensemble invariant recherché. $\square$

Dans la suite, la notion d'invariance est à comprendre au sens le plus large de presque-invariance. En particulier, la sous-tribu $\mathscr{I}$ des invariants est définie par

$$\mathscr{I} := \{A \in \mathscr{A} : \mu(T^{-1}A \vartriangle A) = 0\}.$$

(Vérifier en exercice que cet ensemble de parties forme bien une tribu !)

1 Le théorème ergodique en moyenne

Le premier résultat important dans ce domaine est dû à John Von Neumann (1932), et concerne la convergence de ces moyennes dans L^2 (en prenant évidemment f dans L^2 au départ).

1.1 Opérateur unitaire associé à une transformation préservant la mesure

Introduisons ici un outil essentiel dans l'étude des systèmes dynamiques. Si T est une transformation préservant la mesure sur un espace mesuré $(X, \mathscr{A}, \mu)$, on définit sur $L^2(X, \mathscr{A}, \mu)$ l'opérateur U_T par

$$\forall f \in L^2, \quad U_T f \ := \ f \circ T.$$

On vérifie facilement que U_T est linéaire, et puisque T préserve μ, on a

$$\forall f \in L^2, \quad \|U_T f\|_{L^2}^2 \ = \ \|f\|_{L^2}^2 \,,$$

ainsi U_T est une isométrie de $L^2(X, \mathscr{A}, \mu)$. Si de plus T est inversible et bimesurable, alors U_T est lui-même inversible, c'est donc un opérateur unitaire de $L^2(X, \mathscr{A}, \mu)$. Comme on le verra plus tard, de nombreuses propriétés de la transformation T peuvent s'exprimer en termes de l'opérateur U_T.

Si maintenant on s'intéresse à la convergence dans L^2 des moyennes $(1/n)S_n(f)$ pour $f \in L^2$, on est ramené à l'étude, dans un espace de Hilbert H, des moyennes $(1/n)\sum_{k=0}^{n-1} U^k f$, où U est une isométrie de H.

1.2 Le théorème ergodique dans L^2

On se place donc ici dans le cadre d'une isométrie U agissant sur un espace de Hilbert H. Il existe alors un sous-espace de H qui joue un rôle important dans l'étude des moyennes $(1/n)\sum_{k=0}^{n-1} U^k f$: c'est le sous-espace I des vecteurs U-invariants, c'est-à-dire des f tels que $Uf = f$. Il est immédiat que I est un sous-espace fermé de H, et que si $f \in I$, toutes les moyennes $(1/n)\sum_{k=0}^{n-1} U^k f$ sont égales à f.

Appelons maintenant un *cobord* tout élément de H qui s'écrit $Uh - h$, avec $h \in H$. Là encore, l'ensemble C des cobords est un sous-espace de H, mais qui n'est pas fermé en général.

Les moyennes $(1/n)\sum_{k=0}^{n-1} U^k f$ sont également simples à étudier lorsque f est un cobord. En effet, si $f = Uh - h$, on a

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k f \;=\; \frac{1}{n}(U^n h - h) \;\xrightarrow[n\to\infty]{}\; 0.$$

Notons que, C étant inclus dans l'orthogonal de I (c'est facile à vérifier !), 0 est la projection orthogonale du cobord f sur I. La généralisation de ce résultat à tous les vecteurs f de H se déduira facilement du lemme suivant.

Lemme 1.1. *Le sous-espace C des cobords est dense dans l'orthogonal du sous-espace I des vecteurs invariants.*

Démonstration. Il suffit de montrer que si un vecteur f est orthogonal à la fois à I et à C, il est nul. Or, si f est un tel vecteur, comme $Uf - f$ est un cobord, on a

$$\langle f\,,\,Uf - f\rangle \;=\; 0,$$

d'où

$$\langle f\,,\,Uf\rangle \;=\; \|f\|^2\,.$$

Ceci entraîne $f = Uf$, c'est-à-dire $f \in I$. Et comme $f \perp I$, on en déduit $f = 0$. □

Théorème 1.2 (Von Neumann, 1932)**.** *Soit U une isométrie d'un espace de Hilbert H, et soit P_I la projection sur le sous-espace fermé I de H constitué des vecteurs invariants par U. Alors, pour tout $f \in H$, on a*

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k f \;\xrightarrow[n\to\infty]{}\; P_I(f). \tag{1}$$

Démonstration. On a déjà vu que si $f \in I$ ou si $f \in C$, f vérifie (1). Par ailleurs, il est facile de voir que l'ensemble des $f \in H$ pour lesquels la convergence (1) a lieu est un sous-espace fermé de H. Or, d'après le lemme 1.1, le plus petit sous-espace fermé contenant à la fois C et I est H tout entier. □

Voyons maintenant comment s'applique ce résultat aux systèmes dynamiques. Cette fois, H est l'espace $L^2(X,\mathscr{A},\mu)$, où $(X,\mathscr{A},\mu)$ est un espace de Lebesgue, et U est l'isométrie U_T associée à la transformation T préservant la mesure sur $(X,\mathscr{A},\mu)$.

Une fonction $f \in L^2$ est invariante par U_T si et seulement si pour toute partie mesurable B de $\mathbb{C}$, $f^{-1}(B) \in \mathscr{I}$. Ainsi, I est le sous-espace de L^2 constitué des fonctions $\mathscr{I}$-mesurables. De plus, pour tout $f \in L^2$ on a $P_I(f) = \mathbb{E}\,[f \mid \mathscr{I}\,]$. Le théorème de Von Neumann se formule alors de la façon suivante.

Théorème 1.3 (Théorème ergodique en moyenne)**.** *Pour tout $f \in L^2(X,\mathscr{A},\mu)$, on a*

$$\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k \;\xrightarrow[n\to\infty]{L^2}\; \mathbb{E}\,[f \mid \mathscr{I}\,]\,.$$

2 Le théorème ergodique ponctuel

Le théorème ergodique en moyenne, s'il renseigne sur le comportement global des moyennes $(1/n)S_n(f)$, ne donne en revanche aucune précision sur la convergence, en un point x donné, de la suite $(1/n)\sum_{k=0}^{n-1} f(T^k x)$. Cette convergence ponctuelle a été obtenue par Birkhoff, à qui l'on doit le théorème suivant [1].

Théorème 2.1 (Birkhoff, 1931)**.** *Soit $f \in L^1(\mu)$, à valeurs réelles.*

1. *Pour μ-presque tout x, la suite $(1/n)\sum_{k=0}^{n-1} f(T^k x)$ a une limite quand $n \to \infty$, notée $f^*(x)$.*
2. *La fonction f^* ainsi définie est T-invariante, dans L^1, et vérifie*
$$\int_X f^*\, d\mu \;=\; \int_X f\, d\mu. \tag{2}$$
3. *Soit $\mathscr{I}$ la tribu des événements T-invariants. On a μ-p.s.*
$$f^* \;=\; \mathbb{E}\left[\, f \mid \mathscr{I}\,\right],$$

Démonstration. Il existe plusieurs façons de démontrer ce théorème. La preuve qui suit est très fortement inspirée de l'esquisse donnée par Kalikow et McCutcheon dans [2]. On définit

$$f^*(x) \;:=\; \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x),$$

et

$$f_*(x) \;:=\; \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x).$$

Il est immédiat que f^* et f_* sont invariantes par T. On a $-\infty \le f_* \le f^* \le +\infty$, et l'essentiel de la suite consiste à vérifier l'égalité (presque sûre) de f_* et f^*. L'idée pour y parvenir consiste à établir que si l'entier N est suffisamment grand, la moyenne $(1/N)\sum_{k=0}^{N-1} f(T^k x)$ est avec forte probabilité presque aussi grande que $f^*(x)$ et presque aussi petite que $f_*(x)$. Plus précisément, on va montrer le lemme suivant.

Lemme 2.2. *Soit $b \in \mathbb{R}$. Définissons*

$$S_b \;:=\; \{x \in X :\; f^*(x) > b\}\,.$$

Pour tout $\varepsilon > 0$, pour tout N assez grand, il existe $B_{b,N,\varepsilon} \subset S_b$ avec $\mu(S_b \setminus B_{b,N,\varepsilon}) < \varepsilon$, tel que pour tout $x \in B_{b,N,\varepsilon}$

$$\frac{1}{N}\sum_{k=0}^{N-1} f(T^k x) \;>\; b - \varepsilon. \tag{3}$$

1. Bien qu'ayant démontré son théorème après Von Neumann, Birkhoff a publié son résultat un peu plus tôt.

Preuve du lemme 2.2 – Fixons b et $\varepsilon > 0$. Soit $\delta > 0$, assez petit (à préciser dans la suite). Puisque f est intégrable, on peut trouver $m \geq 1$ tel que $f = f_1 + f_2$ où $|f_1| \leq m$ et

$$\mathbb{E}\left[|f_2|\right] < \delta^2. \tag{4}$$

Par définition de S_b, pour presque tout $x \in S_b$ il existe un premier entier $\ell(x)$ tel que

$$\frac{1}{\ell(x)} \sum_{k=0}^{\ell(x)-1} f(T^k x) \; > \; b. \tag{5}$$

Puis, il existe un entier n pour lequel

$$\mu\left(\{x \in S_b : \; \ell(x) > n\}\right) \; < \; \delta^2/m. \tag{6}$$

On dit qu'un point x est *mauvais* si $x \in S_b$ et $\ell(x) > n$, et on note M l'ensemble des mauvais points. Soit maintenant un entier N assez grand pour que

$$\frac{n}{N} \; < \; \delta/m. \tag{7}$$

Par (6), on a

$$\mathbb{E}\left[\frac{1}{N}\sum_{k=0}^{N-1} \mathbb{1}_M(T^k x)\right] \; = \; \mu(M) \; < \; \delta^2/m,$$

d'où

$$\mu\left(\frac{1}{N}\sum_{k=0}^{N-1} \mathbb{1}_M(T^k x) > \delta/m\right) \; < \; \delta. \tag{8}$$

Par (4), on a aussi

$$\mathbb{E}\left[\frac{1}{N}\sum_{k=0}^{N-1} |f_2(T^k x)|\right] \; = \; \mathbb{E}\left[|f_2|\right] \; < \; \delta^2,$$

d'où également

$$\mu\left(\frac{1}{N}\sum_{k=0}^{N-1} |f_2(T^k x)| > \delta\right) \; < \; \delta. \tag{9}$$

Définissons maintenant $B_{b,N,\varepsilon}$ comme l'ensemble des $x \in S_b$ qui vérifient à la fois

$$\frac{1}{N}\sum_{k=0}^{N-1} \mathbb{1}_M(T^k x) \; \leq \; \delta/m, \tag{10}$$

et

$$\frac{1}{N}\sum_{k=0}^{N-1} |f_2(T^k x)| \; \leq \; \delta. \tag{11}$$

Par (8) et (9), on a bien pour δ assez petit

$$\mu(S_b \setminus B_{b,N,\varepsilon}) \; < \; \varepsilon.$$

Fixons maintenant un x dans $B_{b,N,\varepsilon}$. Remarquons que tous les points $T^k x$, $k \in \mathbb{N}$ sont également dans S_b à cause de l'invariance par T de f^*. Notons j_1 le plus petit entier positif tel que $T^{j_1}x \notin M$. On a donc $\ell_1 := \ell(T^{j_1}x) \leq n$, et par définition de $\ell(T^{j_1}x)$

$$\frac{1}{\ell_1}\sum_{k=0}^{\ell_1-1} f(T^{j_1+k}x) \ > \ b.$$

Puis, on définit récursivement les suites $(j_i)_{i\geq 1}$ et $(\ell_i)_{i\geq 1}$ par

$$\begin{array}{rrcl} & j_{i+1} & := & \text{le plus petit entier supérieur à } j_i+\ell_i \\ & & & \text{tel que } T^{j_{i+1}}x \text{ ne soit pas mauvais,} \\ \text{et} & \ell_{i+1} & := & \ell(T^{j_{i+1}}x) \ \leq n. \end{array}$$

On a toujours

$$\frac{1}{\ell_i}\sum_{k=0}^{\ell_i-1} f(T^{j_i+k}x) \ > \ b. \tag{12}$$

Soit r le plus grand indice tel que $j_r+\ell_r \leq N-1$. Les r intervalles d'entiers $\{j_i, j_i+1, \ldots, j_i+\ell_i-1\}$, $1 \leq i \leq r$ sont deux à deux disjoints, et si un entier k dans $\{0,\ldots,N-1\}$ n'appartient pas à la réunion de ces r intervalles, c'est que

— ou bien $T^k x$ est mauvais,

— ou bien $k \in \{N-n,\ldots,N-1\}$.

Par (10) et (7), la proportion de $\{0,\ldots,N-1\}$ recouverte par ces r intervalles vérifie donc

$$\frac{\ell_1+\ell_2+\cdots+\ell_r}{N} \ > \ 1-2\delta/m. \tag{13}$$

Considérons alors la moyenne $(1/N)\sum_{k=0}^{N-1} f(T^k x)$, et séparons les contributions des entiers k qui sont dans la réunion U des r intervalles $\{j_i, j_i+1,\ldots,j_i+\ell_i-1\}$ et de ceux qui ne sont pas dans U. Pour les premiers, elle s'écrit

$$\frac{1}{N}\sum_{i=1}^{r}\sum_{k=0}^{\ell_i-1} f(T^{j_i+k}x)$$
$$= \ \frac{\sum \ell_i}{N}\frac{1}{\sum \ell_i}\left(\ell_1\frac{1}{\ell_1}\sum_{k=0}^{\ell_1-1} f(T^{j_1+k}x)+\cdots+\ell_r\frac{1}{\ell_r}\sum_{k=0}^{\ell_r-1} f(T^{j_r+k}x)\right).$$

Grâce à (12) et à (13), on voit que cette contribution est supérieure à $b-\varepsilon/2$ si δ est assez petit. Puis, pour la part des entiers $k \in \{0,\ldots,N-1\}$ en dehors de U, on écrit

$$\left|\frac{1}{N}\sum_{k\notin U} f(T^k x)\right| \ \leq \ \frac{1}{N}\sum_{k\notin U}|f_1(T^k x)| + \frac{1}{N}\sum_{k=0}^{N-1}|f_2(T^k x)|.$$

Puisque $|f_1| \leq m$, on peut en utilisant (13) et (11) majorer cette contribution par $\varepsilon/2$ si δ est assez petit, ce qui achève la preuve du lemme 2.2.

Terminons maintenant la preuve du théorème de Birkhoff. Par une preuve analogue, on obtient un résultat équivalent au lemme 2.2, concernant cette fois f_* : pour tout $a \in \mathbb{R}$, si on définit

$$I_a \ := \ \{x \in X : \ f_*(x) < a\},$$

alors pour tout $\varepsilon > 0$, pour tout N assez grand, il existe $C_{a,N,\varepsilon} \subset I_a$ avec $\mu(I_a \setminus C_{a,N,\varepsilon}) < \varepsilon$, tel que pour tout $x \in C_{a,N,\varepsilon}$

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) \ < \ a + \varepsilon. \tag{14}$$

On en déduit que si $a < b$, $\mu(I_a \cap S_b) = 0$. En effet, si ce n'était pas le cas, pour tout $\varepsilon < \mu(I_a \cap S_b)/2$ et pour N assez grand les ensembles $C_{a,N,\varepsilon}$ et $B_{b,N,\varepsilon}$ seraient d'intersection non vide, ce qui entraînerait l'existence de points vérifiant à la fois (3) et (14). C'est évidemment impossible si ε est assez petit. L'ensemble

$$\bigcup_{a<b,\ (a,b)\in\mathbb{Q}\times\mathbb{Q}} I_a \cap S_b,$$

est donc un négligeable en dehors duquel on a $f^*(x) = f_*(x)$, ce qui prouve la première partie du théorème.

La limite, toujours notée $f^*(x)$, de la suite $(1/n)\sum_{k=0}^{n-1} f(T^k x)$ est évidemment T-invariante. Puisque, pour tout $n \geq 1$, on a

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)\right] \ = \ \mathbb{E}\left[f\right],$$

le théorème de convergence dominée donne (2) dans le cas où f est bornée (f^* est évidemment intégrable dans ce cas). Si on suppose seulement $f \in L^1(\mu)$, on écrit comme précédemment $f = f_1 + f_2$ où f_1 est bornée et $\mathbb{E}\left[|f_2|\right] < \varepsilon$. On a alors $f^* = f_1^* + f_2^*$, où

$$\int_X f_1^* \, d\mu \ = \ \int_X f_1 \, d\mu,$$

et par le lemme de Fatou

$$\begin{aligned}
\int_X |f_2^*| \, d\mu \ &= \ \int_X \left| \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f_2(T^k x) \right| d\mu \\
&\leq \ \int_X \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |f_2(T^k x)| \, d\mu \\
&\leq \ \lim_{n\to\infty} \int_X \frac{1}{n} \sum_{k=0}^{n-1} |f_2(T^k x)| \, d\mu \\
&= \ \mathbb{E}\left[|f_2|\right] \ < \ \varepsilon.
\end{aligned}$$

Ceci prouve en particulier que f_2^*, donc aussi f^*, est dans $L^1(\mu)$. De plus,

$$\left| \int_X f_1 \, d\mu - \int_X f \, d\mu \right| \leq \mathbb{E}\left[|f_2|\right] < \varepsilon.$$

Comme ε est arbitraire on obtient aussi (2).

Enfin, pour tout $A \in \mathscr{I}$ avec $\mu(A) > 0$, on a également

$$\int_A f^* \, d\mu \ = \ \int_A f \, d\mu. \tag{15}$$

En effet, il suffit dans tout ce qui précède de remplacer l'espace X par A, sur lequel agit la restriction $T|_A$, qui préserve clairement la probabilité sur A

$$\mu_A := \mu(\cdot)/\mu(A).$$

Or, puisque f^* est $\mathscr{I}$-mesurable (car T-invariante), (15) dit exactement que $f^* = \mathbb{E}[f \mid \mathscr{I}]$ (p.s.). □

3 Ergodicité

Comme on le voit dans les énoncés des théorèmes ergodiques, la σ-algèbre $\mathscr{I}$ des événements T-invariants joue un rôle essentiel dans l'étude de la convergence des moyennes $(1/n)S_n(f)$. Rappelons maintenant la notion d'*ergodicité*, qui concerne précisément cette tribu des invariants.

Définition 3.1. *Le système dynamique $(X, \mathscr{A}, \mu, T)$ est dit* ergodique *si, pour tout $A \in \mathscr{A}$, A est T-invariants implique $\mu(A) = 0$ ou $\mu(A) = 1$.*

3.1 Quelques caractérisations de l'ergodicité

Proposition 3.2. *Les propriétés suivantes sont équivalentes.*

1. *Le système dynamique $(X, \mathscr{A}, \mu, T)$ est ergodique.*
2. *Pour tout $A \in \mathscr{A}$ avec $\mu(A) > 0$, $\mu\left(\bigcup_{n\geq 1} T^{-n}A\right) = 1$.*
3. *Pour tous A, B dans $\mathscr{A}$ avec $\mu(A) > 0$, $\mu(B) > 0$, il existe un entier $n \geq 1$ tel que $\mu(T^{-n}A \cap B) > 0$.*
4. *Pour toute fonction $f \in L^1(X)$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \xrightarrow[n\to\infty]{\mu\text{-p.s.}} \int_X f\,d\mu.$$

La preuve est laissée en exercice...

Exercice 3.1. *Montrer comment la loi forte des grands nombres peut être vue comme un corollaire du théorème ergodique ponctuel.*

Exercice 3.2. *Trouver un système dynamique ergodique $(X, \mathscr{A}, \mu, T)$ et une fonction $f : X \to \mathbb{R}$ (non intégrable !) pour laquelle la suite*

$$(1/n)\sum_{k=0}^{n-1} f(T^k x)$$

n'a presque sûrement pas de limite (même infinie).

Il est parfois utile de pouvoir tester l'ergodicité du système dynamique en n'ayant recours qu'à une famille dénombrable de fonctions tests. On dispose pour cela du lemme qui suit.

Lemme 3.3. *Soit $(B_j)_{j\in\mathbb{N}}$ une algèbre dénombrable séparant les points de X. Si, pour tout $j \in \mathbb{N}$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{B_j}(T^k x) \xrightarrow[n\to\infty]{\mu\text{-p.s.}} \mu(B_j),$$

alors $(X, \mathscr{A}, \mu, T)$ est ergodique.

Démonstration. Si $(B_j)_{j\in\mathbb{N}}$ est une algèbre dénombrable séparant les points, cette algèbre engendre la tribu $\mathscr{A}$ et la famille (B_j) est automatiquement dense dans $\mathscr{A}$, au sens où pour tout $A \in \mathscr{A}$ et tout $\varepsilon > 0$ il existe j tel que $\mu(B_j \vartriangle A) < \varepsilon$. Autrement dit, pour tout $A \in \mathscr{A}$ il existe une sous-suite (j_k) telle que

$$\mathbb{1}_{B_{j_k}} \xrightarrow[k\to\infty]{L^1} \mathbb{1}_A.$$

Or l'hypothèse de l'énoncé dit que pour tout j,

$$\mathbb{E}\left[\mathbb{1}_{B_j} \mid \mathscr{I}\right] \;=\; \mu(B_j) \quad \text{(p.s.)}.$$

On en déduit

$$\mathbb{E}\left[\mathbb{1}_A \mid \mathscr{I}\right] \;=\; \lim_{k\to\infty} \mu(B_{j_k}) \;=\; \mu(A) \quad (\mu\text{-p.s.}).,$$

ce qui entraîne l'ergodicité du système. □

Remarque – L'intérêt de prendre l'hypothèse "(B_j) sépare les points" au lieu de simplement supposer que (B_j) engendre $\mathscr{A}$ tient dans le fait pas toujours clairement mis en évidence que la notion de *tribu engendrée par une famille d'ensembles* sous-entend bien souvent "modulo la mesure μ". Lorsque l'on dit "(B_j) engendre $\mathscr{A}$" sur un espace probabilisé fixé $(X, \mathscr{A}, \mu)$, cela signifie précisément : pour tout A dans $\mathscr{A}$, il existe $\tilde{A} \in \sigma(B_j,\ j \in \mathbb{N})$ tel que $\mu(A \vartriangle \tilde{A}) = 0$, et cette propriété d'être famille génératrice de $\mathscr{A}$ dépend de la mesure μ sous-jacente. En supposant que $(B_j)_{j\in\mathbb{N}}$ sépare les points de X, on est assuré que (B_j) engendre $\mathscr{A}$ *quelle que soit la mesure sous-jacente* !

3.2 Ergodicité et action de U_T

Comme on va le voir maintenant, l'ergodicité est un premier exemple de propriété d'un système dynamique qui peut s'observer sur l'action de l'opérateur U_T. Notons qu'un événement A est T-invariant si et seulement si la fonction $\mathbb{1}_A$ est T-invariante. Remarquons que les fonctions constantes, qui sont dans L^2 puisque $\mu(X)$ est finie, sont toujours T-invariantes : ce sont donc toujours des fonctions propres de U_T associées à la valeur propre 1. Si 1 est une valeur propre simple de U_T, la fonction $\mathbb{1}_A$ ne peut être T-invariante que si elle est constante, c'est-à-dire si $\mu(A) = 0$ ou 1. Dans ce cas, T est donc ergodique. Réciproquement, une fonction étant T-invariante si et seulement si elle est $\mathscr{I}$-mesurable, l'ergodicité de T entraîne que les seules fonctions T-invariantes sont les constantes. On en déduit la caractérisation suivante de l'ergodicité.

Proposition 3.4. *Les propriétés suivantes sont équivalentes.*

1. *Le système dynamique $(X, \mathscr{A}, \mu, T)$ est ergodique.*
2. *$U_T f = f$ si et seulement si f est une fonction constante.*
3. *La valeur propre 1 est une valeur propre simple de U_T.*

4 Décomposition en composantes ergodiques

Si le système dynamique $(X, \mathscr{A}, \mu, T)$ est ergodique, l'observation de la trajectoire d'un seul point x permet de retrouver la mesure μ. En effet, ayant fixé une algèbre dénombrable $(B_j)_{j \in \mathbb{N}}$ séparant les points de X, on sait qu'en choisissant x dans une partie de probabilité 1, on aura pour tout j

$$\mu(B_j) \ = \ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{B_j}(T^k x),$$

et cela fixe la mesure μ sur toute la tribu $\mathscr{A}$.

Mais que dire de la limite lorsque T n'est pas ergodique ? Le théorème ergodique ponctuel assure de toute façon que pour tout A et μ-presque tout $x \in X$,

$$m(x, A) \ := \ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{A}(T^k x)$$

existe. Il est évident que la fonction l ainsi définie est, à x fixé, une fonction additive de A : si A_1 et A_2 sont disjoints, $m(x, A_1 \cup A_2) = m(x, A_1) + m(x, A_2)$. De plus, on a évidemment $m(x, X) = 1$. La théorie de la décomposition en composantes ergodiques précise cette limite $m(x, A)$: pour μ-presque tout x, $A \longmapsto m(x, A)$ est en fait une mesure de probabilité qui est T-invariante et *ergodique* ! Ainsi, dans un système dynamique quelconque, pas nécessairement ergodique, l'observation d'une seule trajectoire ne pourra de toute façon que révéler une mesure ergodique. On montre de plus que μ s'écrit toujours comme une moyenne des mesures ergodiques qui apparaissent ainsi. La moralité est que l'étude des systèmes dynamiques peut, pour l'essentiel, se résumer à celle des systèmes dynamiques ergodiques.

On présente dans la suite deux façons différentes d'aborder le problème de la décomposition en composantes ergodiques.

4.1 Décomposition ergodique via la désintégration des mesures

On utilise le théorème suivant, qui précise l'existence dans un espace de Lebesgue d'une version assez régulière de l'espérance conditionnelle sachant une sous-tribu. Ce résultat découle de l'étude de la théorie des espaces de Lebesgue effectuée par Rokhlin ([4]), mais sa démonstration peut se trouver dans de nombreux ouvrages de théorie de la mesure ou des probabilités (par exemple [1], p.460).

Théorème 4.1 (Désintégration d'une probabilité par rapport à une sous-tribu)**.** *Soit* $(X, \mathscr{A}, \mu)$ *un espace de Lebesgue, et* $\mathscr{F}$ *une sous-tribu de* $\mathscr{A}$. *Il existe une application* $(x, A) \longmapsto \mu_x(A|\mathscr{F})$ *de* $X \times A$ *dans* $[0, 1]$ *telle que*

— *pour tout* A *fixé dans* $\mathscr{A}$, $x \longmapsto \mu_x(A|\mathscr{F})$ *est* $\mathscr{F}$*-mesurable, et*

$$\mu_x(A, \mathscr{F}) \ = \ \mathbb{E}\left[\mathbb{1}_A \mid \mathscr{F}\right] \quad (\mu\text{-p.s.});$$

— *pour* μ*-presque tout* x, $A \longmapsto \mu_x(A|\mathscr{F})$ *est une mesure de probabilité sur* $(X, \mathscr{A})$.

L'idée consiste maintenant à appliquer ce théorème avec $\mathscr{F} = \mathscr{I}$, la tribu des événements T-invariants, puisque le théorème ergodique précise que la limite de $(1/n) \sum_{k=0}^{n-1} \mathbb{1}_A(T^k x)$ est justement $\mathbb{E}\left[\mathbb{1}_A \mid \mathscr{I}\right]$. Le théorème de désintégration assure l'existence d'une application que l'on notera simplement $(x, A) \longmapsto \mu_x(A)$ telle que pour μ-presque tout x, $A \longmapsto \mu_x(A)$ est une mesure de probabilité sur X, et vérifiant pour tout $A \in \mathscr{A}$

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) \xrightarrow[n\to\infty]{\mu\text{-p.s.}} \mu_x(A). \tag{16}$$

Théorème 4.2. *Pour μ-presque tout x, la probabilité μ_x est T-invariante, et le système dynamique $(X, \mathscr{A}, \mu_x, T)$ est ergodique.*

Démonstration. On commence par fixer une algèbre dénombrable $(B_j)_{j\in\mathbb{N}}$ qui sépare les points sur X. On considère également la famille dénombrable de parties $A_{\alpha,j}$, $\alpha \in \mathbb{Q}\cap[0,1]$ et $j \in \mathbb{N}$, définie par

$$A_{\alpha,j} \; := \; \left\{ y \in X : \; \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{B_j}(T^k y) \leq \alpha \right\}.$$

Soit X_0 l'ensemble des $x \in X$ pour lesquels la convergence (16) a lieu pour tout A dans l'ensemble dénombrable $\{T^{-l}B_j,\ j \in \mathbb{N},\ l \in \mathbb{N}\} \cup \{A_{\alpha,j},\ \alpha \in \mathbb{Q}\cap[0,1]\ ,j \in \mathbb{N}\}$. Alors $\mu(X_0) = 1$, et on va montrer maintenant que si $x \in X_0$, μ_x vérifie les propriétés annoncées.

Fixons donc $x \in X_0$. Il est facile de voir que μ_x est T-invariante, en remarquant que pour tout $j \in \mathbb{N}$, (16) est valable à la fois pour B_j et pour $T^{-1}B_j$, ce qui permet d'en déduire que $\mu_x(T^{-1}B_j) = \mu_x(B_j)$ pour tout j ; cela entraîne $\mu_x(T^{-1}A) = \mu_x(A)$ pour tout $A \in \mathscr{A}$. Puis, pour prouver l'ergodicité du système $(X, \mathscr{A}, \mu_x, T)$, il suffit d'après le lemme 3.3 de vérifier que pour tout j,

$$\mu_x\left(\left\{ y \in X : \; \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{B_j}(T^k y) = \mu_x(B_j) \right\}\right) \; = \; 1. \tag{17}$$

Or, chaque $A_{\alpha,j}$ étant clairement invariant par T, on s'aperçoit que

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{A_{\alpha,j}}(T^k x) \; = \; \begin{cases} 1 & \text{si } \mu_x(B_j) \leq \alpha, \\ 0 & \text{sinon.} \end{cases}$$

Grâce à la validité de (16) pour tous les $A_{\alpha,j}$, on en déduit que

$$\mu_x(A_{\alpha,j}) \; = \; \begin{cases} 1 & \text{si } \mu_x(B_j) \leq \alpha, \\ 0 & \text{sinon,} \end{cases}$$

ce qui suffit à prouver (17). □

Corollaire 4.3. *Il existe une mesure de probabilité P sur l'ensemble $\mathscr{M}_e(T)$ des mesures de probabilité T-invariantes et ergodiques, telle que pour tout $A \in \mathscr{A}$,*

$$\mu(A) \; = \; \int_{\mathscr{M}_e(T)} \nu(A)\, dP(\nu).$$

Idée de la preuve. On a défini une application $x \longmapsto \mu_x$ de X dans $\mathscr{M}_e(T)$. Soit P l'image par μ de cette application (munir auparavant $\mathscr{M}_e(T)$ d'une tribu convenable !). On utilise ensuite le fait que pour tout $A \in \mathscr{A}$,

$$\mu(A) \; = \; \int_X \mu_x(A)\, d\mu(x).$$

□

Une remarque sur les composantes ergodiques

Les mesures ergodiques μ_x apparaissant dans la désintégration de μ par rapport à la tribu $\mathscr{I}$ sont appelées les *composantes ergodiques* de μ. Si μ_x et μ_y sont deux telles composantes ergodiques, on peut remarquer qu'elles sont soit identiques, soit mutuellement singulières. En effet, on a le résultat suivant.

Proposition 4.4. *Soient μ et ν dans $\mathscr{M}_e(T)$. Si $\mu \neq \nu$, alors $\mu \perp \nu$.*

Démonstration. Supposons que μ et ν ne soient pas mutuellement singulières. Alors on peut trouver une probabilité η à la fois absolument continue par rapport à μ et à ν. Le théorème ergodique appliqué dans le système $(X, \mathscr{A}, \mu, T)$ donne, pour tout $A \in \mathscr{A}$,

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) \xrightarrow[n\to\infty]{\mu\text{-p.s.}} \mu(A),$$

alors que dans le système $(X, \mathscr{A}, \nu, T)$, on obtient

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) \xrightarrow[n\to\infty]{\nu\text{-p.s.}} \nu(A).$$

Puisqu'on a $\eta \ll \mu$ et $\eta \ll \nu$, cette même moyenne converge η-presque sûrement à la fois vers $\mu(A)$ et $\nu(A)$, d'où forcément $\mu = \nu$. □

Exercice 4.1. *Décrire la décomposition en composantes ergodiques d'une rotation rationnelle.*

Exercice 4.2. *Considérons le système dynamique $(X, \mathscr{A}, \mu, T)$, où $X = [0, 1[\times[0, 1[$, μ est la mesure de Lebesgue 2-dimensionnelle sur le carré, et*

$$T(x, y) := (x + y \mod 1, y).$$

Décrire sa décomposition en composantes ergodiques.

4.2 Décomposition ergodique via le théorème de Choquet

Cet autre point de vue sur la décomposition en composantes ergodiques utilise un résultat important dû à Choquet (voir par exemple [3]). Avant de l'énoncer, il est nécessaire de définir la notion de *point extrémal* d'un ensemble convexe.

Définition 4.5. *Soit K une partie convexe d'un espace vectoriel. On dit que $M \in K$ est un* point extrémal de *K si*

$$M = tA + (1 - t)B \text{ avec } 0 < t < 1 \text{ et } A,\ B \text{ dans } K \implies A = B = M.$$

On notera $\mathscr{E}(K)$ l'ensemble des points extrémaux de K.

Théorème 4.6 (Théorème de représentation intégrale de Choquet)**.** *Soit E un espace topologique localement convexe, et $K \subset E$ une partie compacte, convexe et métrisable. Pour tout $M \in K$, il existe une probabilité P sur $\mathscr{E}(K)$ telle que*

$$M = \int_{\mathscr{E}(K)} A\, dP(A).$$

(Cette intégrale signifiant précisément que pour toute forme linéaire φ continue sur E, $\varphi(M) = \int_{\mathscr{E}(K)} \varphi(A)\, dP(A)$.)

Pour pouvoir appliquer ce théorème au problème de la décomposition ergodique, on suppose que X est un espace métrique compact et que T est une transformation continue de X (cette hypothèse n'est en fait pas restrictive car on peut montrer que tout système dynamique peut être représenté à isomorphisme près de cette façon). On considère l'espace E des mesures signées sur X, que l'on munit de la topologie de la convergence faible (engendrée par les applications $\mu \longmapsto \int_X f d\mu$ pour f fonction continue sur X). La partie de E formée par les mesures de probabilité est alors une partie compacte métrisable de E. Puis, le sous-ensemble $\mathscr{M}_1(T)$ des mesures de probabilité T-invariantes est lui-même un compact convexe métrisable (non vide par le théorème de Krylov-Bogolioubov). L'application du théorème de Choquet à $\mathscr{M}_1(T)$ permet d'affirmer que toute probabilité $\mu \in \mathscr{M}_1(T)$ peut être représentée sous une forme intégrale

$$\mu \;=\; \int_{\mathscr{E}(\mathscr{M}_1(T))} \nu \, dP(\nu),$$

où P est une probabilité portée par l'ensemble des points extrémaux de $\mathscr{M}_1(T)$. Cette écriture de μ n'est autre que la décomposition du système en composantes ergodiques, ce qui apparaît grâce au résultat qui suit.

Proposition 4.7. *Soit $\mu \in \mathscr{M}_1(T)$; μ est un point extrémal de $\mathscr{M}_1(T)$ si et seulement si le système $(X, \mathscr{A}, \mu, T)$ est ergodique.*

Démonstration. Supposons d'abord $(X, \mathscr{A}, \mu, T)$ ergodique, et écrivons $\mu = t\nu_1 + (1-t)\nu_2$ où $0 < t < 1$ et ν_1 et ν_2 sont dans $\mathscr{M}_1(T)$. Alors ν_1 et ν_2 sont toutes les deux absolument continues par rapport à μ. Le théorème ergodique dans $(X, \mathscr{A}, \nu_i, T)$ $(i = 1, 2)$ donne alors, pour tout A dans $\mathscr{A}$

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) \;\xrightarrow[n\to\infty]{\nu_i\text{-p.s.}}\; \nu_i(A),$$

tandis que le même théorème appliqué cette fois dans $(X, \mathscr{A}, \mu, T)$ dit que la convergence a lieu μ-presque sûrement, donc ν_i-presque sûrement, vers $\mu(A)$. On en déduit que pour tout $A \in \mathscr{A}$, $\nu_1(A) = \nu_2(A) = \mu(A)$, ce qui prouve que μ est un point extrémal de $\mathscr{M}_1(T)$.

Inversement, supposons que μ n'est pas ergodique. Alors il existe A invariant par T avec $0 < \mu(A) < 1$. Les deux mesures conditionnelles $\mu(\,\cdot\,|A)$ et $\mu(\,\cdot\,|X \setminus A)$ sont aussi dans $\mathscr{M}_1(T)$, et l'écriture

$$\mu \;=\; \mu(A)\mu(\,\cdot\,|A) + \big(1 - \mu(A)\big)\mu(\,\cdot\,|X \setminus A)$$

montre que μ n'est pas un point extrémal de $\mathscr{M}_1(T)$. □

Remarque – On a montré au passage un résultat qui est parfois utile, énoncé dans la proposition qui suit.

Proposition 4.8. *Si μ est ergodique pour T, et si ν est une mesure T-invariante absolument continue par rapport à μ, alors $\nu = \mu$.*

Références

[1] P. Billingsley, *Probability and measure, second edition*, John Wiley & Sons, 1986.

[2] S. Kalikow and R. McCutcheon, *An outline of ergodic theory*, 2010.

[3] R. Phelps, *Lectures on Choquet's theorem, second edition*, Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, 2001.

[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] M. Valadier, *What differentiates stationary stochastic processes from ergodic ones : a survey*, Sūrikaisekikenkyūsho Kōkyūroku (2001), no. 1215, 33–52, Mathematical economics (Japanese) (Kyoto, 2000).