Théorèmes ergodiques, ergodicité et composantes ergodiques

T. de la Rue

Master 2 MFA Rouen 2024-2025

Nous considérons dans ce chapitre un système dynamique mesuré (X, μ, T), où μ est une mesure de probabilité sur l'espace Borel-standard $\big(X, \mathscr{B}(X)\big)$, et T est une transformation [1] mesurable de X dans lui-même [2] 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, Pour une fonction f mesurable sur X à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$, on voudrait étudier l'éventuelle limite (en un sens à préciser) des moyennes temporelles

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k.$$

1 La sous-tribu des invariants

Comme nous le verrons dans ce chapitre, une sous-tribu particulière de $\mathscr{B}(X)$ joue un rôle essentiel dans l'identification de la limite des moyennes

1. On suppose en général que T est aussi inversible, mais l'inversibilité ne joue aucun rôle dans ce chapitre.

2. On suppose en général que T est aussi inversible, mais l'inversibilité ne joue aucun rôle dans ce chapitre.

temporelles : il s'agit de la sous-tribu des ensembles T-invariants :

$$\mathscr{I}_T := \{A \in \mathscr{B}(X) : A = T^{-1}(A)\}.$$

Il est facile de vérifier que $\mathscr{I}_T$ est bien une sous-tribu de $\mathscr{B}(X)$, et que si $f : X \to \mathbb{C}$ est une fonction mesurable, alors f est $\mathscr{I}_T$-mesurable si et seulement si pour *tout* $x \in X$, $f(Tx) = f(x)$.

Invariant, ou presque invariant ?

On pourrait avoir envie de considérer la notion d'invariance dans un sens un peu plus large, qui ignore les ensembles de mesure nulle. On peut ainsi définir la sous-tribu des ensembles *presque* T-invariants :

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

De même, le soin est laissé au lecteur de vérifier que $\mathscr{I}_T^\mu$ est également une sous-tribu de $\mathscr{B}(X)$, et que si $f : X \to \mathbb{C}$ est une fonction mesurable, alors f est $\mathscr{I}_T^\mu$-mesurable si et seulement si pour μ-presque tout $x \in X$ on a $f(Tx) = f(x)$. De plus, on a évidemment $\mathscr{I}_T \subset \mathscr{I}_T^\mu$.

Exercice 1.1. *Prouver que pour tout $A \in \mathscr{B}(X)$, les ensembles $\bigcap_{n\geq 0} T^{-n}A$ et $\bigcup_{n\geq 0} T^{-n}A$ sont presque T-invariants.*

Une apparente bonne raison de considérer $\mathscr{I}_T^\mu$ au lieu de $\mathscr{I}_T$ est la suivante : on travaille souvent avec des observables dans $L^1(\mu)$ ou $L^2(\mu)$, qui ne sont donc définies qu'à un ensemble de mesure μ nulle près. Et quand on dit qu'une fonction f dans $L^1(\mu)$ est T-*invariante*, cela signifie que $f \circ T = f$ dans $L^1(\mu)$, ce qui ne correspond qu'à une égalité μ-presque sûre.

Néanmoins, la proposition suivante prouve qu'il n'est pas si important de distinguer entre $\mathscr{I}_T$ et $\mathscr{I}_T^\mu$.

Proposition 1.1. *Si $A \in \mathscr{I}_T^\mu$, alors il existe $\tilde{A} \in \mathscr{I}_T$ tel que $\mu(\tilde{A} \vartriangle A) = 0$. Si $f : X \to \mathbb{C}$ est $\mathscr{I}_T^\mu$-mesurable, alors il existe $\tilde{f}$, $\mathscr{I}_T$-mesurable, telle que $f = \tilde{f}$ μ-p.s.*

Démonstration. Traitons le cas d'une application $f : X \to \mathbb{C}$ qui est $\mathscr{I}_T^\mu$-mesurable (ceci donnera le résultat également pour $A \in \mathscr{I}_T^\mu$ en prenant $f = \mathbb{1}_A$). Puisque $f(x) = f(Tx)$ μ-presque sûrement, on a

$$\mu\left(\left\{x \in X : f(x) = \lim_{n\to\infty} f(T^n x)\right\}\right) \;=\; 1.$$

On obtient donc le résultat annoncé en posant

$$\tilde{f}(x) \;:=\; \begin{cases} \lim_{n\to\infty} f(T^n x) & \text{si cette limite existe,} \\ 0 & \text{sinon.} \end{cases}$$

□

Comme on le verra dans la suite, c'est en particulier l'espérance conditionnelle par rapport à la sous-tribu des invariants qui joue un rôle central dans les théorèmes ergodiques. Or le résultat suivant découle immédiatement de la proposition précédente.

Corollaire 1.2. *Pour toute observable $f \in L^1(\mu)$, on a*

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

Dans la suite, on utilisera de préférence la sous-tribu $\mathscr{I}_T$, dont l'avantage est qu'elle ne dépend pas de la mesure sous-jacente. On peut donc garder la même notation pour les invariants, même lorsque l'on considère d'autres mesures T-invariantes sur l'espace d'état (voir par exemple l'équation (19) dans la démonstration de la proposition 5.7).

2 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 $L^2(\mu)$ (en prenant évidemment f dans $L^2(\mu)$ au départ).

2.1 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 $\big(X, \mathscr{B}(X), \mu\big)$, on définit sur $L^2(\mu)$ l'opérateur U_T par

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

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 $L^2(\mu)$. Si de plus T est inversible et bimesurable, alors U_T est lui-même inversible, c'est donc un opérateur unitaire de $L^2(\mu)$. 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.

2.2 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 2.1. *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 $f = 0$. $\square$

Théorème 2.2 (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 2.1, 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 $L^2(\mu)$, où $\big(X, \mathscr{B}(X), \mu\big)$ est un espace de Lebesgue, et U est l'isométrie U_T associée à la transformation T préservant la mesure sur $\big(X, \mathscr{B}(X), \mu\big)$.

Une fonction $f \in L^2$ est invariante par U_T si et seulement si pour toute partie mesurable B de $\mathbb{C}$, $f^{-1}(B) \in \mathscr{I}_T^\mu$. Ainsi, I est le sous-espace de L^2 constitué des fonctions $\mathscr{I}_T^\mu$-mesurables. De plus, pour tout $f \in L^2$ on a $P_I(f) = \mathbb{E}_\mu\left[f \mid \mathscr{I}_T^\mu\right] = \mathbb{E}_\mu\left[f \mid \mathscr{I}_T\right]$ (pour la dernière égalité, on a appliqué le corollaire 1.2). Le théorème ergodique en moyenne de Von Neumann se formule alors de la façon suivante.

Théorème 2.3 (Théorème ergodique en moyenne)**.** *Pour tout $f \in L^2(\mu)$, on a*

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

3 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.

Théorème 3.1 (Birkhoff, 1931 [3])**.** *Soit $f \in L^1(\mu)$, à valeurs réelles.*

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

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. *On a μ-p.s.*

$$f^* \;=\; \mathbb{E}_\mu \left[f \mid \mathscr{I}_T \right].$$

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 3.2. *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 $x \in B_{b,N,\varepsilon}$

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

Preuve du lemme 3.2 – 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

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

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

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

Le but est de montrer que pour un tel entier N, avec une grande probabilité la moyenne des valeurs de f sur les N premiers points de l'orbite d'un point $x \in S_b$ est au moins $b - \varepsilon$. Dans un premier temps, on va montrer que sur un ensemble $B_{b,N,\varepsilon}$ qui remplit une grosse proportion de S_b, la contribution des mauvais points n'est pas importante, et de même pour la contribution du terme non borné f_2.

Par (6), on a

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

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

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

d'où également

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

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

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

et

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

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^kx, $k\in\mathbb{N}$ 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 $k \in \{N-n, \dots, N-1\}$.

Par (10) et (7), la proportion de $\{0, \dots, N-1\}$ 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 $\{j_i, j_i + 1, \dots, j_i + \ell_i - 1\}$ 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 $k \in \{0, \dots, N-1\}$ 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 3.2.

Terminons maintenant la preuve du théorème de Birkhoff. Par un argument analogue, on obtient un résultat équivalent au lemme 3.2, 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

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

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 $\mathbb{E}_\mu[|f_2|] < \varepsilon$. 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

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

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

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

Comme ε est arbitraire on obtient aussi (2).

Enfin, pour tout $A \in \mathscr{I}_T$ 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 $\mathscr{I}_T$-mesurable (car T-invariante), (15) dit exactement que $f^* = \mathbb{E}_\mu [f \mid \mathscr{I}_T]$ (p.s.). □

4 Ergodicité

Comme on le voit dans les énoncés des théorèmes ergodiques, la σ-algèbre $\mathscr{I}_T$ 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 4.1. *Le système dynamique (X, μ, T) est dit* ergodique *si pour tout $A \in \mathscr{I}_T$, $\mu(A) = 0$ ou $\mu(A) = 1$.*

Lorsque la transformation T est clairement fixée par le contexte, on dit aussi simplement que *μ est ergodique* (pour T).

En appliquant la proposition 1.1, il est immédiat que cette définition est équivalente au fait que les ensembles presque T-invariants sont ceux dont la mesure vaut 0 ou 1.

4.1 Quelques caractérisations de l'ergodicité

Proposition 4.2. *Les propriétés suivantes sont équivalentes.*

1. *Le système dynamique (X, μ, T) est ergodique.*
2. *Pour tout $A \in \mathscr{B}(X)$ avec $\mu(A) > 0$, $\mu\left(\bigcup_{n\geq 1} T^{-n}A\right) = 1$.*
3. *Pour tous A, B dans $\mathscr{B}(X)$ 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 4.1. *Montrer comment la loi forte des grands nombres peut être vue comme un corollaire du théorème ergodique ponctuel. (Penser aux schémas de Bernoulli, dont l'ergodicité a été vue au chapitre précédent.)*

Exercice 4.2. *Trouver un système dynamique ergodique* (X, μ, T) *et une fonction mesurable* $f : X \to \mathbb{R}$*, non intégrable, pour laquelle la suite*

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

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

(Indication : s'inspirer de l'exercice précédent, et utiliser la loi d'une variable aléatoire non intégrable pour laquelle la loi des grands nombres ne s'applique pas, par exemple une loi de Cauchy.)

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 4.3. *Soit* $(B_j)_{j\in\mathbb{N}}$ *une algèbre dénombrable séparant les points de* X*. Le système* (X, μ, T) *est ergodique si et seulement si*

$$\forall j \in \mathbb{N}, \quad \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{B_j}(T^k x) \xrightarrow[n\to\infty]{\mu\text{-p.s.}} \mu(B_j). \tag{16}$$

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

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

Si l'hypothèse 16 est satisfaite, alors pour tout j,

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

On en déduit que

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

Comme c'est valable pour tout $A \in \mathscr{B}(X)$, cela signifie que $\mathscr{I}_T$ est la tribu triviale, ce qui caractérise l'ergodicité du système.

Réciproquement, il est clair que si le système est ergodique, alors (16) est satisfaite (conséquence du théorème ergodique ponctuel). □

4.2 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. Remarquons que les fonctions constantes, qui sont dans $L^2(\mu)$ 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 donc 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 dans $L^2(\mu)$ étant T-invariante si et seulement si elle est $\mathscr{I}_T^\mu$-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 4.4. *Les propriétés suivantes sont équivalentes.*

1. *Le système dynamique (X, μ, 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.*

5 Décomposition en composantes ergodiques

Si le système dynamique $(X, \mathscr{B}(X), \mu, T)$ est ergodique, il existe une partie $X_0 \in \mathscr{B}(X)$ de probabilité $\mu(X_0) = 1$ tel que l'observation de la trajectoire d'un point quelconque $x \in X_0$ permet de retrouver la mesure μ. En effet, ayant fixé une algèbre dénombrable $(B_j)_{j\in\mathbb{N}}$ séparant les points de X, le théorème ergodique ponctuel nous permet d'affirmer qu'avec probabilité 1, on a pour tout $j \in \mathbb{N}$

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

Or, la donnée des $\mu(B_j)$, $j \in \mathbb{N}$ fixe la mesure μ sur toute la tribu $\mathscr{B}(X)$ qui est engendrée par les (B_j) (application du théorème d'extension de Carathéodory).

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 qu'à x fixé, la limite ci-dessus est une fonction additive de A : si A_1 et A_2 sont disjoints, on a toujours $\mathbb{1}_{A_1\cup A_2}(T^k x) = \mathbb{1}_{A_1}(T^k x) + \mathbb{1}_{A_2}(T^k x)$, et donc $m(x, A_1 \cup A_2) = m(x, A_1) + m(x, A_2)$. De plus, on a évidemment $m(x, X) = 1$. La *désintégration en composantes ergodiques*, qui est présentée dans la section suivante, précise la nature de cette limite : pour μ-presque tout x, $A \longmapsto m(x, A)$ est en fait une mesure de probabilité qui est T-invariante et *ergodique* ! Ainsi, même dans un système dynamique mesuré non ergodique, l'observation d'une seule trajectoire ne révèle (avec probabilité 1) qu'une mesure ergodique. On montre de plus que μ s'écrit toujours comme une moyenne des mesures ergodiques qui apparaissent ainsi, et cela permet dans de nombreux cas de ramener l'étude d'un système mesuré quelconque au cas où le système est ergodique.

5.1 Désintégration des mesures et composantes ergodiques

On utilise le théorème suivant, qui assure l'existence dans un espace de Lebesgue $(X, \mathscr{B}(X), \mu)$ d'une version régulière de l'espérance conditionnelle sachant une sous-tribu de $\mathscr{B}(X)$. Ce résultat se trouve dans l'article de référence sur les espaces de Lebesgue écrit par Rokhlin [5], mais sa démonstration peut aussi se trouver dans de nombreux ouvrages de théorie de la mesure ou des probabilités (par exemple [1, Cor. 10.4.6 p.361]).

Théorème 5.1 (Désintégration d'une probabilité par rapport à une sous–tribu)**.** *Dans un espace de Lebesgue $(X, \mathscr{B}(X), \mu)$, soit $\mathscr{F}$ une sous-tribu de $\mathscr{B}(X)$. Alors il existe une application $(x, A) \longmapsto \mu_x(A \,|\, \mathscr{F})$ de $X \times \mathscr{B}(X)$ dans $[0,1]$ telle que*

— *pour tout A fixé dans $\mathscr{B}(X)$, $x \longmapsto \mu_x(A \,|\, \mathscr{F})$ est $\mathscr{F}$-mesurable ;*
— *on a $\mu_x(A \,|\, \mathscr{F}) \;=\; \mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{F}\right](x)$ (μ-p.s.) ;*
— *pour μ-presque tout x, $A \longmapsto \mu_x(A \,|\, \mathscr{F})$ est une mesure de probabilité sur $(X, \mathscr{B}(X))$.*

L'idée consiste maintenant à appliquer ce théorème avec $\mathscr{F} = \mathscr{I}_T$, la sous-tribu des événements T-invariants, puisque l'on sait d'après le théorème ergodique que la limite presque sûre de $(1/n)\sum_{k=0}^{n-1} \mathbb{1}_A(T^k x)$ est justement $\mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{I}_T\right]$. 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 $A \in \mathscr{B}(X)$

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) \xrightarrow[n\to\infty]{\mu\text{-p.s.}} \mu_x(A). \tag{17}$$

Théorème 5.2. *Pour μ-presque tout x, la probabilité μ_x est T-invariante, et le système dynamique mesuré (X, μ_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

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

Soit X_0 l'ensemble des $x \in X$ pour lesquels la convergence (17) a lieu pour tout A dans l'ensemble dénombrable $\{T^{-\ell}B_j,\ j \in \mathbb{N},\ \ell \in \mathbb{N}\} \cup \{A_{\alpha,j},\ \alpha \in \mathbb{Q} \cap [0,1]\ , j \in \mathbb{N}\}$. 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 $j \in \mathbb{N}$, (17) 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 $A \in \mathscr{B}(X)$. Puis, pour prouver l'ergodicité du système $(X, \mathscr{B}(X), \mu_x, T)$, il suffit d'après le lemme 4.3 de vérifier que pour tout j,

$$\mu_x \left(\left\{ y \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{B_j}(T^k y) = \mu_x(B_j) \right\} \right) = 1. \tag{18}$$

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 (17) 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 (18). □

La mesure μ_x définie pour μ-presque tout x dans l'énoncé du théorème précédent est appelée la *composante ergodique du point x*. 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 5.3. *Soient μ et ν deux mesures de probabilité T-invariantes tels que les systèmes (X, μ, T) et (X, ν, T) soient ergodiques. Si $\mu \neq \nu$, alors $\mu \perp \nu$.*

Démonstration. Soioent μ et ν T-invariantes et ergodiques, et supposons qu'elles ne sont 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, μ, T) donne, pour tout $A \in \mathscr{B}(X)$,

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

alors que dans le système (X, ν, 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 $\mu = \nu$. □

La section suivante explique comment on reconstitue la mesure invariante μ à partir de ces composantes ergodiques.

5.2 Moyenne de composantes ergodiques

Notons $\mathscr{M}_1(X)$ l'ensemble des mesures de probabilité sur $\big(X, \mathscr{B}(X)\big)$, puis définissons $\mathscr{M}_1(X,T)$ comme le sous-ensemble de $\mathscr{M}_1(X)$ formé par les mesures T-invariantes, et $\mathscr{M}_{1,e}(X,T) \subset \mathscr{M}_1(X,T)$ comme le sous-ensemble des mesures de probabilités ν qui sont T-invariantes et telles que le système (X,ν,T) soit ergodique.

Une première conséquence du théorème 5.2 est la suivante : si $\mathscr{M}_1(X,T) \neq \emptyset$, alors $\mathscr{M}_{1,e}(X,T) \neq \emptyset$. En particulier, dans le cas où T est un homéomorphisme d'un espace métrique compact X, le théorème de Kryloff-Bogoliouboff nous assure qu'il existe au moins une mesure de probabilité T-invariante, donc il existe au moins une mesure de probabilité T-invariante et ergodique. De plus, si il n'existe qu'une *unique* mesure de probabilité T-invariante, alors celle-ci est automatiquement ergodique.

On munit $\mathscr{M}_1(X)$ de la tribu $\mathscr{B}\big(\mathscr{M}_1(X)\big)$ engendrée par les « évaluations » $\phi_A : \mathscr{M}_1(X) \to \mathbb{R}$, $\phi_A(\mu) := \mu(A)$ $(A \in \mathscr{B}(X))$. Alors l'espace mesurable $\Big(\mathscr{M}_1(X), \mathscr{B}\big(\mathscr{M}_1(X)\big)\Big)$ est aussi un espace Borel-standard (voir par exemple le théorème 1.5 dans [3]). Par ailleurs, $\mathscr{M}_1(X,T)$ et $\mathscr{M}_{1,e}(X,T)$ sont des parties mesurables dans $\mathscr{B}\big(\mathscr{M}_1(X)\big)$ [4]. Alors l'application $C : x \mapsto \mu_x$ est mesurable de X dans $\mathscr{M}_{1,e}(X,T)$. On peut donc définir la mesure de probabilité image $P := C_*(\mu)$ sur $\mathscr{M}_{1,e}(X,T)$. On remarque ensuite que pour tout $A \in \mathscr{B}(X)$,

$$\mu(A) = \mathbb{E}_\mu\Big[\mathbb{E}_\mu\left[\mathbb{1}_A \mid \mathscr{I}_T\right]\Big] = \int_X \mu_x(A)\, d\mu(x) = \int_{\mathscr{M}_{1,e}(X,T)} \nu(A)\, dP(\nu).$$

De plus, la mesure de probabilité P sur $\mathscr{M}_{1,e}(X,T)$ qui satisfait la relation ci-dessus est unique. En effet, supposons que Q soit une mesure de probabilité sur $\mathscr{M}_{1,e}(X,T)$ telle que pour tout $A \in \mathscr{B}(X)$, on ait

$$\mu(A) = \int_{\mathscr{M}_{1,e}(X,T)} \nu(A)\, dQ(\nu).$$

4. Le vérifier en exercice, on peut utiliser le lemme 4.3 pour traiter $\mathscr{M}_{1,e}(X,T)$.

L'application $x \mapsto \mu_x$ étant bien définie μ-presque partout, elle est bien définie ν-presque partout pour Q-presque toute mesure $\nu \in \mathscr{M}_{1,e}(X,T)$. De plus, comme une telle mesure ν est ergodique, on a $\mu_x = \nu$ pour ν-presque tout x. On veut maintenant établir que $Q = C_*(\mu)$. Pour cela, considérons un ensemble mesurable $E \subset \mathscr{M}_{1,e}(X,T)$. On a alors

$$\begin{aligned} C_*(\mu)(\nu \in E) &= \mu\left(\{x \in X : \mu_x \in E\}\right) \\ &= \int_{\mathscr{M}_{1,e}(X,T)} \nu\left(\{x \in X : \mu_x \in E\}\right) \, dQ(\nu) \\ &= \int_{\mathscr{M}_{1,e}(X,T)} \mathbb{1}_E(\nu) \, dQ(\nu) \\ &= Q(E). \end{aligned}$$

On a donc démontré le théorème ci-dessous.

Théorème 5.4 (Décomposition en composantes ergodiques)**.** *Il existe une unique mesure de probabilité P sur l'ensemble $\mathscr{M}_{1,e}(X,T)$ des mesures de probabilité T-invariantes et ergodiques, telle que pour tout $A \in \mathscr{B}(X)$,*

$$\mu(A) \; = \; \int_{\mathscr{M}_{1,e}(X,T)} \nu(A) \, dP(\nu).$$

On écrit généralement la conclusion du théorème précédent sous la forme condensée

$$\mu \; = \; \int_X \mu_x \, d\mu(x) \; = \; \int_{\mathscr{M}_{1,e}(X,T)} \nu \, dP(\nu),$$

et cette façon d'exprimer μ comme une moyenne de mesures invariantes ergodiques pour T est appelée la *décomposition en composantes ergodiques* de la mesure T-invariante μ.

Exercice 5.1. *Décrire la décomposition en composantes ergodiques d'une rotation rationnelle.*

Exercice 5.2. *Considérons le système dynamique $(X, \mathscr{B}(X), \mu, T)$, où $X = [0,1[\times[0,1[$, μ est la mesure de Lebesgue 2-dimensionnelle sur le carré, et*

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

Décrire sa décomposition en composantes ergodiques.

5.3 Mesures ergodiques et points extrémaux

La décomposition en composantes ergodiques expliquée ci-dessus peut se voir comme un cas particuler du *théorème de représentation intégrale* de Choquet. Avant d'énoncer ce théorème, on introduit la notion de *point extrémal* d'une partie convexe.

Définition 5.5. *Soit K une partie convexe d'un espace vectoriel. On dit que $M \in K$ est un* point extrémal de K *si*

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

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

Théorème 5.6 (Théorème de représentation intégrale de Choquet [4])**.** *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 tout système dynamique mesuré 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 $\mathscr{M}_1(X,T)$ des mesures de probabilité T-invariantes est lui-même un compact convexe métrisable, non vide par le théorème de Kryloff-Bogoliouboff). L'application du théorème de Choquet à $\mathscr{M}_1(X,T)$ permet d'affirmer que toute probabilité $\mu \in \mathscr{M}_1(X,T)$ peut être représentée sous une forme intégrale

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

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

Proposition 5.7. *Soit $\mu \in \mathscr{M}_1(X,T)$; μ est un point extrémal de $\mathscr{M}_1(X,T)$ si et seulement si $\mu \in \mathscr{M}_{1,e}(X,T)$.*

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

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) \xrightarrow[n\to\infty]{\nu_i\text{-p.s.}} \mathbb{E}_{\nu_i}\left[\mathbb{1}_A \,|\, \mathscr{I}_T\right], \tag{19}$$

tandis que le même théorème appliqué cette fois dans (X,μ,T) ergodique dit que la convergence a lieu μ-presque sûrement vers $\mu(A)$. Mais puisque $\nu_i \ll \mu$, on a aussi pour $i = 1,2$

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

En comparant avec (19), on obtient

$$\mathbb{E}_{\nu_i}\left[\mathbb{1}_A \,|\, \mathscr{I}_T\right] \;=\; \mu(A) \quad (\nu_i\text{-p.s.}).$$

En intégrant par rapport à ν_i, on prouve alors que $\nu_i(A) = \mu(A)$, et ceci pour tout $A \in \mathscr{B}(X)$. Ainsi $\nu_1 = \nu_2 = \mu$, ce qui prouve que μ est un point extrémal de $\mathscr{M}_1(X,T)$.

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 $\mathscr{M}_1(X,T)$, 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 $\mathscr{M}_1(X,T)$. □

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

Proposition 5.8. *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] V. I. Bogachev, *Measure theory. Vol. II*, Berlin : Springer, 2007.

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

[3] Olav Kallenberg, *Random measures, theory and applications*, Probab. Theory Stoch. Model., vol. 77, Cham : Springer, 2017.

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

[5] V.A. Rohlin, *On the fundamental ideas of measure theory*, AMS Translations Serie 1 **10** (1963), 2–53, (Première publication en russe : 1949.).