Théorie Ergodique

Léo Gayral

Ces notes sont basées sur le cours de Sara Brofferio et Mélanie Guenais.

Table des matières

1 Généralités

1.1 Introduction

Définition 1 (Système dynamique) **:**

On considère un ensemble mesuré $(X, \mathcal{A}, \mu)$, avec $\mathcal{A}$ complète pour la mesure μ : si $B \subset A \in \mathcal{A}$ et $\mu(A) = 0$ alors $B \in \mathcal{A}$. La mesure μ est toujours σ-finie, et sauf mention explicite on considèrera $\mu(X) = 1$, donc un espace probabilisé.

On considère de plus une application $T : X \to X$ mesurable, qui préserve μ : pour tout $A \in \mathcal{A}$, on a $\mu(T^{-1}(A)) = \mu(A)$. En d'autres termes, pour toute fonction f positive ou L^1, on a :

$$\int_\Omega f(T(x))\,\mathrm{d}\mu(x) = \int_\Omega f(x)\,\mathrm{d}\mu(x)\,.$$

Proposition 2 (Lemme de classe monotone) **:**

Soit $\mathcal{C} \subset \mathcal{A}$ un π-système (stable par intersections finies) qui engendre $\mathcal{A}$.

Alors T préserve μ ssi, pour tout $C \in \mathcal{C}$, on a $\mu(T^{-1}(C)) = \mu(C)$.

Cette proposition reste valable si le π-système engendre une tribu qui se complète en $\mathcal{A}$.

Lemme 3 (Cas des espaces polonais) **:**

Si X est un espace polonais, alors les propriétés suivantes sont équivalentes :
— la transformation T préserve μ,
— pour toute fonction $f \in L^1(X)$, $\mu(f \circ T) = \mu(f)$,
— pour toute fonction $f \in \mathcal{C}_b(X)$, $\mu(f \circ T) = \mu(f)$,
— pour toute fonction f continue à support compact, $\mu(f \circ T) = \mu(f)$.

Définition 4 (Espace MPT et PPT) **:**

Si T préserve la mesure μ sur $(X, \mathcal{A})$, on dit que $(X, \mathcal{A}, \mu, T)$ est une *measure-preserving transformation*, un espace MPT.

Si de plus μ est une mesure de probabilité, $(X, \mathcal{A}, \mu, T)$ est une *probability-preserving transformation*, un espace PPT.

1.2 Exemples

Remarque 5 (Cas du Tore) **:**

Soit $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ le tore, muni de sa tribu borélienne complétée pour la mesure de Lebesgue λ. Dans ce cas, les translations $T_\alpha : x \mapsto x + \alpha$ préservent λ.

Soit désormais $T(x) = 2x$. On a :

$$\lambda(f \circ T) = \int_0^1 f(T(x))\,\mathrm{d}x = \int_0^{\frac{1}{2}} f(2x)\,\mathrm{d}x + \int_{\frac{1}{2}}^1 f(2x-1)\,\mathrm{d}x = 2 \times \frac{1}{2}\int_0^1 f(y)\,\mathrm{d}y = \lambda(f)$$

donc T préserve également λ.

Remarque 6 (Cas de variables aléatoires) **:**

Soit $(\Omega, \mathcal{B}, \rho)$ un espace probabilisé. Sur l'espace $(X, \mathcal{F}, \mu) = (\Omega, \mathcal{B}, \rho)^{\otimes \mathbb{N}}$, les projections canoniques $X_i(\omega) = \omega_i$ correspondent à une famille de variables aléatoires iid, de loi ρ.

Alors l'opérateur de shift $\Theta : (\omega_0, \omega_1, \ldots) \mapsto (\omega_1, \omega_2, \ldots)$ est une transformation qui préserve la mesure produit $\rho^{\otimes \mathbb{N}}$.

Lorsque Ω est fini, X est parfois appelé un espace de Bernouilli. En particulier, pour la mesure uniforme $\rho = \frac{\delta_0 + \delta_1}{2}$, on note cet espace $B\left(\frac{1}{2}, \frac{1}{2}\right)$.

Remarque 7 (Cas des chaînes de Markov) **:**

On considère ici une P-chaîne de Markov sur Ω dénombrable admettant une probabilité invariante ν. Alors il existe une mesure μ sur $\Omega^{\mathbb{Z}}$ pour laquelle les projections canoniques $(X_n)_{n \in \mathbb{Z}}$ forment une P-chaîne de Markov stationnaire.

Dans ce cas, à nouveau, l'opérateur de shift Θ préserve la mesure μ.

Théorème 8 :

Toute transformation T qui préserve une mesure μ sur un espace mesurable X induit un processus stationnaire $(X_n)_{n \in \mathbb{N}}$ à valeurs dans X, issu de la mesure :

$$\mathbb{P}(X_0 \in A_0, \ldots, X_n \in A_n) = \mu\left(\bigcap_{i=0}^n T^{-i}(A_i)\right) = \int_X \prod_{i=0}^n \mathbb{1}_{T^i(x) \in A_i}\,\mathrm{d}\mu(x)\,.$$

Comme $X_{n+1} = T(X_n)$, c'est en particulier une chaîne de Markov.

1.3 Théorème de récurrence de Poincaré

Théorème 9 (Théorème de récurrence de Poincaré) **:**

Soient $(X, \mathcal{A}, \mu, T)$ un PPT et A tel que $\mu(A) > 0$.

Pour μ-presque tout point $x \in A$, l'orbite $\{T^n(x),\, n \in \mathbb{N}\}$ du point x visite A infiniment souvent. Autrement dit :

$$\mu(\{x \in A,\, \forall n_0 \in \mathbb{N},\, \exists n \geq n_0,\, T^n(x) \in A\}) = \mu(A)\,.$$

Démonstration. Dire que $T^n(x) \in A$ équivaut à dire que $x \in T^{-n}(A)$. En conséquence, la trajectoire de x visite A après le rang n_0 ssi $x \in A_{n_0} := \bigcup\limits_{n \geq n_0} T^{-n}(A)$. En conséquence, l'ensemble des points qui visitent infiniment souvent A est l'ensemble $A_\infty = \overline{\lim}\, A_n = \bigcap\limits_{n \in \mathbb{N}} A_n$.

En outre, on constate que $T^{-1}(A_n) = A_{n+1}$, par compatibilité entre l'union ensembliste et l'image réciproque. Comme T préserve la mesure, la suite $\mu(A_n)$ est constante. En conséquence, $\mu(A_\infty) = \mu(A_0)$, donc $\mu(A_0 \backslash A_\infty) = 0$. Or $A \subset A_0$ d'où enfin $\mu(A \backslash A_\infty) = 0$. $\square$

Remarque 10 :
Si la mesure μ est infinie, le théorème n'est pas toujours vrai.

Ainsi, sur $\mathbb{R}$ muni de la mesure de Lebesgue, pour toute translation T_x ($x \neq 0$), et toute partie A bornée, on a un rang n_0 à partir duquel $T^n(A) \cap A = \emptyset$.

1.4 Temps de premier retour et théorème de Kac

Définition 11 (Temps de premier retour) **:**
Soit $A \in \mathcal{A}$. On pose $\tau_A(x) = \inf\{n > 0,\, T^n(x) \in A\} \in \overline{\mathbb{N}}$. Cette application est mesurable.

Corollaire 12 (Application du théorème de récurrence de Poincaré) **:**
Sur un espace PPT, on a $\mu(\{x \in A,\, \tau_A(x) < \infty\}) = \mu(A)$.

Théorème 13 (Théorème de Kac) **:**
Sur un PPT, on a $\int\limits_A \tau_A(x)\,\mathrm{d}\mu(x) = \mu(\{\tau_A < \infty\})$.

Démonstration. Soient $A_k = \{x \in A,\, \tau_A(x) = k\}$ et $E_k = \{x \notin A,\, \tau_A(x) = k\}$. On remarque que $T^{-1}(E_k) = \{x \in X,\, \tau_A(x) = k+1\} = E_{k+1} \sqcup A_{k+1}$. En conséquence, par induction :

$$\begin{aligned}
\mu(E_k) &= \mu(E_{k+1}) + \mu(A_{k+1}) \\
&= \mu(E_{n+1}) + \sum_{j=k+1}^{n} \mu(A_j) \\
&= \sum_{j=k+1}^{\infty} \mu(A_j)
\end{aligned}$$

car les parties E_n sont disjointes, donc $\mu(E_n) \underset{n\to\infty}{\longrightarrow} 0$. On a donc finalement :

$$\begin{aligned}
\int\limits_A \tau_A(x)\,\mathrm{d}\mu(x) &= \sum_{i=1}^{\infty} i \times \mu(A_i) \\
&= \sum_{k=1}^{\infty}\sum_{i=k}^{\infty} \mu(A_i) \\
&= \left(\sum_{i=1}^{\infty} \mu(A_i)\right) + \sum_{k=2}^{\infty} \mu(E_{k-1}) \\
&= \mu\left(\bigsqcup_{n\in\mathbb{N}} (A_n \sqcup E_n)\right) \\
&= \mu(\{\tau_A < \infty\})\,.
\end{aligned}$$

$\square$

Corollaire 14 :
Si $\mu(\{\tau_A < \infty\}) = 1$, alors $\int \tau_A(x)\,\mathrm{d}\mu_A(x) := \frac{1}{\mu(A)} \int\limits_A \tau_A(x)\,\mathrm{d}\mu(x) = \frac{1}{\mu(A)}$.

2 Théorie Ergodique

On s'intéresse désormais à la fréquence des retours en A, la limite de $\frac{1}{n}\sum\limits_{k=0}^{n-1} \mathbb{1}_A(T^n(x))$.

Remarque 15 (Cas de variables iid) **:**
Si $(X_n)_{n\in\mathbb{N}} \overset{\text{iid}}{\sim} \rho$ est une suite de variables iid, et $A = E_0 \times \Omega^{\mathbb{N}}$ est un évènement qui ne dépend que de X_0, on a alors $\frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{1}_A(T^n(\omega)) = \frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{1}_{E_0}(X_n) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X \in E_0) = \rho(E_0)$ par la loi des grands nombres.

Remarque 16 :
Soient $X = \mathbb{Z}/M\mathbb{Z}$ et $T : x \mapsto x+1$. Naturellement, la seule mesure invariante est $\mu = \mathcal{U}(X)$. Ici, $\sum\limits_{k=0}^{n-1} \mathbb{1}_A(T^n(x)) \in \left[n \times \frac{|A|}{M} \pm |A|\right]$ donc on a convergence vers $\frac{|A|}{M} = \mu(A)$ à vitesse $\frac{M}{n}$.

Remarque 17 :
On ne converge pas toujours vers une constante. Ainsi, si on juxtapose deux copies du système précédent, $Y = X \sqcup X$, et qu'on considère μ uniforme, alors la suite est stationnaire égale à 0 si on part d'une composante disjointe de A.

2.1 Théorème ergodique de Von Neumann dans L^2

On travaille sur l'espace de Hilbert $L^2(X,\mu)$ et on note $\langle f, g\rangle = \int f\overline{g}\,\mathrm{d}\mu$ son produit hermitien.

Définition 18 (Opérateur de Koopman) **:**
On pose $U_T : f \mapsto f \circ T$ un opérateur linéaire sur les fonctions mesurables.

Proposition 19 :
Si μ est T invariante, alors U_T est une isométrie.

Démonstration. Comme T préserve μ, on a $\int f \circ T\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$ pour toute fonction $f \in L^1$. C'est en particulier le cas pour $f\overline{g} \in L^1$ lorsque $f, g \in L^2$. □

Remarque 20 :

U_T est une isométrie donc une injection, mais n'est pas nécessairement surjective.

C'est par exemple le cas de l'opérateur de shift sur un espace produit.

Proposition 21 :

Soit U une application linéaire continue sur H un espace de Hilbert, et U^* son opérateur adjoint, tel que $\langle U^* f, g\rangle := \langle f, Ug\rangle$.

L'opérateur U est une isométrie ssi $UU^* = \mathrm{Id}$. Dans ce cas, en particulier, $Uf = f$ ssi $U^* f = f$.

Démonstration. L'application U est une isométrie si et seulement si $\forall f, g \in H$, $\langle Uf, Ug\rangle = \langle f, g\rangle$ ssi $\forall f, g \in H$, $\langle U^* U f, g\rangle = \langle f, g\rangle$ ssi $UU^* = \mathrm{Id}$.

Dans ce cas, si $Uf = f$ alors naturellement $U^* f = U^* U f = f$.

Réciproquement, si $U^* f = f$, alors :

$$\langle f, Uf\rangle = \langle U^* f, f\rangle = \langle f, f\rangle = \|f\|^2$$

et en passant au conjugué, $\langle Uf, f\rangle = \overline{\langle f, Uf\rangle} = \|f\|^2$. En développant le produit scalaire, on a $\|Uf - f\|^2 = 0$, d'où $Uf = f$. □

Remarque 22 :

En général, l'isométrie U n'est pas nécessairement surjective. En conséquence, même si on a $U^* U = \mathrm{Id}_H$ et que U^* est bien défini sur H tout entier, on ne peut pas les faire commuter directement, on n'a que $UU^* = \mathrm{Id}_{U(H)}$.

C'est pour cela que l'implication $U^* f = f \Rightarrow Uf = f$ est moins directe qu'il n'y parait.

Théorème 23 (Théorème de Von Neuman) **:**

Soit $I(U) = \ker(U - \mathrm{Id})$. C'est un H-sev fermé. On peut donc considérer $P : H \to I(U)$ la projection orthogonale.

On a alors $\frac{1}{n}\sum_{k=0}^{n-1} U^k(f) \xrightarrow[n\to\infty]{L^2} Pf$ pour tout $f \in H$.

Démonstration. Soit $L(U) = \mathrm{Im}(U - \mathrm{Id})$. On a alors $L(U)^\perp = I(U^*)$, En effet, $w = L(U)^\perp$ ssi pour tout $v \in H$ on a $\langle w, Uv\rangle = \langle w, v\rangle$ ssi $\langle U^*w - w, v\rangle = 0$ ssi $U^*w = w$.

D'après la proposition précédente, $I(U) = I(U^*)$. En conséquence, $H = I(U) \overset{\perp}{\oplus} \overline{L(U)}$. Si $w = Uv - v \in L(U)$ alors $\frac{1}{n}\sum_{k=0}^{n-1} U^k w = \frac{U^n v - v}{n} = O\left(\frac{1}{n}\right)$. Plus généralement, par densité, si $w \in \overline{L(U)}$, alors $\frac{1}{n}\sum_{k=0}^{n-1} U^k w \to 0$. On en déduit le résultat souhaité. □

Corollaire 24 :

Soit (X, μ, T) un espace MPT. Pour toute application $f \in L^2$, il existe $f^* \in L^2$ telle que :

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow{L^2} f^*$$

et $f^* \circ T = f^*$, autrement dit f^* est μ-presque-sûrement invariante par T.

Démonstration. On applique tout simplement le résultat précédent à l'espace $L^2(X, \mu, T)$ muni de l'isométrie U_T. □

Définition 25 (Système ergodique) **:**

On dit que le MPT (X, μ, T) lorsque, pour toute fonction f, si $f \circ T = f$ (presque-sûrement), alors f est constante (presque-sûrement).

Remarque 26 :

Si le MPT est ergodique, par Von Neumann, $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ converge vers une constante.

En particulier, si X est de mesure $\mu(X) = \infty$ infinie, la seule constante dans L^2 est 0.

2.2 Théorème ergodique ponctuel de Birkhoff

Lemme 27 :

Soit f mesurable sur un MPT (X,μ,T). On pose $S_n f = \sum\limits_{k=0} f\circ T^k$ les sommes de Birkhoff de f. On définit alors les fonctions $\overline{f} = \overline{\lim}\,\frac{1}{n}S_n f \geq \underline{\lim}\,\frac{1}{n}S_n f = \underline{f} \geq 0$.

Les fonctions $\overline{f}$ et $\underline{f}$ ainsi définies sont T-invariantes.

Démonstration. On a :

$$\overline{f}\circ T = \overline{\lim}\,\frac{1}{n}\sum_{k=1}^{n} f\circ T^k = \lim\frac{f}{n} + \lim\frac{n+1}{n}\times\overline{\lim}\left(\frac{1}{n+1}S_{n+1}f\right) = \overline{\lim}\,\frac{1}{n+1}S_{n+1}f = \overline{f}\,,$$

et il en va de même pour $\underline{f}$.

□

Lemme 28 :

Soit $f \geq 0$ sur un MPT sur (X,μ,T). On a $\int \overline{f}\,\mathrm{d}\mu \leq \int f\,\mathrm{d}\mu$.

Démonstration. Soient $R>0$ et $\varepsilon>0$. On définit $g(x) = \big((1-\varepsilon)\overline{f}(x)\big)\wedge R$, une troncature de $(1-\varepsilon)\overline{f}$. La fonction g est bornée donc intégrable, et T-invariante.

On pose $n(x) = \inf\big\{k\in\mathbb{N}^*,\ \frac{1}{k}S_k f(x) \geq g(x)\big\}$. Si $\overline{f}(x) = 0$ alors $f(x) = S_1 f(x) \geq 0 = g(x)$, d'où $n(x) = 1$. Sinon, si $\overline{f}(x) = \infty$, alors on a en particulier un rang k tel que $\frac{1}{k}S_k f(x) \geq R \geq g(x)$ donc $n(x)\leq k$. Enfin, si $0<\overline{f}(x)<\infty$, comme $g(x) \leq (1-\varepsilon)\overline{f}(x) \leq \sup\limits_{n>0}\frac{1}{n}S_n f(x)$, d'où k tel que $\frac{1}{k}S_k f(x) \geq (1-\varepsilon)\overline{f}(x)$, et $n(x)\leq k$. On a donc toujours $n(x)<\infty$.

On pose désormais $E_L = \{x\in X,\ n(x)<L\}$ pour $L\in\mathbb{N}$, et $\tau_L = \mathbb{1}_{E_L}\times n + \mathbb{1}_{E_L^c}\times 1$ la « restriction » de n à E_L. Si $x\in E$, alors comme g est T-invariante, on a presque-sûrement :

$$S_\tau f(x) = S_{n(x)}f(x) \geq n(x)g(x) = \sum_{k=0}^{n-1} g\circ T^k(x) \geq \sum_{k=0}^{n-1}(g\mathbb{1}_E)\circ T^k(x) = S_\tau(g\mathbb{1}_E)(x)\,.$$

Lorsque $x\notin E$, comme $\tau(x) = 1$, $S_\tau f(x) = S_1 f(x) = f(x) \geq 0 = g\mathbb{1}_E(x)$.

Par induction, on initialise $\tau_1 = \tau$ et on pose $\tau_{k+1} = \tau_k + \tau\circ T^{\tau_k}$. Alors :

$$S_{\tau_{k+1}}f(x) = S_{\tau_k}f(x) + S_\tau f(T^{\tau_k}(x)) \geq S_{\tau_k}(g\mathbb{1}_E)(x) + S_\tau(g\mathbb{1}_E)(T^{\tau_k}(x)) = S_{\tau_{k+1}}(g\mathbb{1}_E)(x)\,,$$

d'où l'inégalité souhaitée pour tout rang $k>0$.

Soit $N\in\mathbb{N}$. Par construction, comme $\tau\leq L$, pour tout $x\in X$, on au moins un $\tau_k(x)$ situé dans $[\![N, N+L]\!]$, d'où :

$$S_{N+L}f(x) \geq S_{\tau_k}f(x) \geq S_{\tau_k}(g\mathbb{1}_{E_L})(x) \geq S_N(g\mathbb{1}_{E_L})(x)\,.$$

En passant à l'intégrale, on a $(N+L)\mu(f) = \int S_{N+L}f\,\mathrm{d}\mu \geq \int S_N(g\mathbb{1}_{E_L})\,\mathrm{d}\mu = N\int g\mathbb{1}_{E_L}\,\mathrm{d}\mu$. Quitte à diviser l'inégalité par $N+L$, en passant à la limite $N\to\infty$, on a $\mu(f)\geq\mu(g\mathbb{1}_{E_L})$. Pour

$L \to \infty$, par convergence monotone, on a alors $\mu(f) \geq \mu(g)$. Pour $R \to \infty$, par convergence monotone, $\mu(f) \geq \mu\big((1-\varepsilon)\overline{f}\big) = (1-\varepsilon)\mu\big(\overline{f}\big)$. Enfin, pour $\varepsilon \to 0^+$, on obtient $\mu(f) \geq \mu\big(\overline{f}\big)$. $\square$

Théorème 29 (Théorème ergodique de Birkhoff) **:**
Soit (X, μ, T) un PPT. Si $f \in L^1$, alors $\frac{1}{n} S_n f \xrightarrow{L^1, \text{p.s.}} f^*$ où f^* est T-invariante, telle que $\int f^* \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$.

Démonstration. On utilise ici la démonstration de Katznelson et Weiss (1982). Par linéarité du problème, on peut se restreindre au cas f positive.

D'après les lemmes précédents, $\overline{f} \in L^1$ est T-invariante, et $\mu\big(\overline{f}\big) \leq \mu(f)$.

Si on montre que $\mu\big(\underline{f}\big) \geq \mu(f)$, comme $\underline{f} - \overline{f} \geq 0$, alors $0 \leq \mu\big(\overline{f} - \underline{f}\big) \leq \mu(f - f) = 0$, d'où $\underline{f} = \overline{f} = f^*$ p.s., et alors $\mu(f^*) \leq \mu(f) \leq \mu(f^*)$.

Soient $R > 0$ et $f_R = f \wedge R$. On pose alors $g = R - f_R \geq 0$. En particulier, comme on est sur un espace de masse finie, $g \in L^1$, et $\int g \, \mathrm{d}\mu = R + \int f_R \, \mathrm{d}\mu$. Par le lemme précédent, $\int \overline{g} \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$. Or $\overline{g} = R + \underline{f_R}$ donc $\int f_R \, \mathrm{d}\mu \leq \int \underline{f_R} \, \mathrm{d}\mu$. On en déduit par convergence monotone que $\mu(f) \leq \mu\big(\underline{f}\big)$.

On a donc montré que $\frac{1}{n} S_n f \xrightarrow{\text{p.s.}} f^*$, et que $f^* \in L^1$. Il reste à montrer la convergence de la suite. Si on tronque f positive par $R > 0$, on obtient $\mu\big(\big|f_R^* - \frac{1}{n} S_n f_R\big|\big) \underset{n \to \infty}{\longrightarrow} 0$ par convergence dominée par $2R$.

Par le lemme de Fatou, $\|f^* - f_R^*\|_1 \leq \underline{\lim} \int \frac{1}{n} S_n(f - f_R) \leq \|f - f_R\|_1$. Pour tout ε, il existe R tel que $\|f - f_R\|_1 \leq \varepsilon$. Par la propriété de convergence précédente appliquée à f_R :

$$\|S_n f - f^*\|_1 \leq \|f^* - f_R^*\|_1 + \|S_n f - S_n f_R\|_1 + \|S_n f_R - f_R^*\|_1 \leq 2\|f - f_R\|_1 + \|S_n f_R - f_R^*\|_1 \leq 3\varepsilon$$

à partir d'un rang n_0. On a donc la convergence L^1 souhaitée. □

Corollaire 30 :
Si de plus le PPT est ergodique, alors $f^* = \mu(f)$ est constante p.s.

3 Ergodicité et mélange

3.1 Ergodicité et ensembles invariants

Définition 31 (Ensemble invariant) **:**
On dit que $A \in \mathcal{A}$ est un ensemble T-invariant si $T^{-1}(A) = A$ (*vraie* égalité ensembliste), si $\mathbb{1}_A \circ T = \mathbb{1}_A$.

Proposition 32 :

Les propriétés suivantes sont équivalentes :

1. le système est ergodique,
2. si A est T-invariant, alors $\mu(A) \in \{0, 1\}$,
3. si $\mu(A \Delta T^{-1}(A)) = 0$, alors $\mu(A) \in \{0, 1\}$.

Démonstration. L'implication $(1 \Rightarrow 2)$ découle de la définition d'ergodicité appliquée à $f = \mathbb{1}_A$.

L'implication $(2 \Rightarrow 3)$ a été vue en exercice. En effet, si A est presque-sûrement T-invariant, alors $B = \overline{\lim}\, T^{-n}(A)$ est T-invariant, et $\mu(A \Delta B) = 0$. On applique alors la seconde propriété, d'où $\mu(A) = \mu(B) \in \{0, 1\}$.

L'implication $(3 \Rightarrow 1)$ se déduit en considérant les évènements $A_c = \{f \leq c\}$. La fonction $F : c \mapsto \mu(A_c)$ est càdlàg, croissante, et par T-invariance de f, elle est à valeurs dans $\{0, 1\}$. On a donc un seuil $c = \inf F^{-1}(1)$ tel que $F(c^-) = 0$ et $F(c) = 1$, d'où $f = c$ presque-sûrement. □

Proposition 33 :

Soit $\mathbb{T}$ muni de la mesure de Lebesgue λ. On considère la translation $T_\alpha : x \mapsto x + \alpha$. Si $\alpha \in \mathbb{R} \backslash \mathbb{Q}$ alors T_α est ergodique.

Démonstration. Il suffit de vérifier l'ergodicité sur les fonctions L^2 pour conclure, car on aura en particulier le résultat sur les fonctions indicatrices, égales à 0 ou 1.

Posons $(e_n : x \mapsto e^{inx})_{n \in \mathbb{Z}}$. Toute fonction $f \in L^2(\mathbb{T})$ est caractérisée par ses coefficients de Fourier $\left(\widehat{f}(n) := \langle e_n, f \rangle\right)_{n \in \mathbb{Z}}$.

Or $e_n \circ T_\alpha = e_n(\alpha) \times e_n$. Si $\alpha \notin \mathbb{Q}$, alors pour tout $n \neq 0$, $e_n(\alpha) \neq 1$. En outre, on a $\widehat{f \circ T}(n) = \widehat{f}(n) \times e_n(\alpha)$. Si f est T-invariante, alors pour tout $n \neq 0$ on a nécessairement $\widehat{f}(n) = \widehat{f \circ T}(n) = 0$, d'où finalement $f = \widehat{f}(0)$ constante. □

3.2 Ergodicité et récurrence

Proposition 34 :

Sur un PPT, les propriétés suivantes sont équivalentes :

1. le système est ergodique,
2. si $\mu(A) > 0$ alors $\mu\left(\bigcup_{n \in \mathbb{N}} T^{-n}(A)\right) = 1$,
3. Si $\mu(A) > 0$ et $\mu(B) > 0$ alors il existe $n \in \mathbb{N}$ tel que $\mu(B \cap T^{-n}(A)) > 0$.

Démonstration. Pour l'implication $(1 \Rightarrow 2)$, on pose $B = \bigcup_{n \in \mathbb{N}} T^{-n}(A)$. Alors $T^{-1}(B) \subset B$, donc $\mu(B \Delta T^{-1}(B)) = \mu(B) - \mu(T^{-1}(B)) = 0$. On en déduit $\mu(B) \in \{0, 1\}$, et $\mu(B) \geq \mu(A) > 0$, d'où $\mu(B) = 1$.

Pour l'implication ($2 \Rightarrow 3$), naturellement, $\mu(B) = \mu\left(\bigcup_{n\in\mathbb{N}} T^{-1}(A) \cap B\right) > 0$, d'où un certain n pour lequel la mesure est non nulle.

Enfin, pour ($3 \Rightarrow 1$), on considère $A \in \mathcal{A}$ et $B = A^c$. Si $T^{-1}(A) = A$, alors pour tout $n \in \mathbb{N}$, $T^{-n}(A) \cap B = A \cap A^c = \emptyset$. En conséquence, par contraposée de la troisième propriété, nécessairement $\mu(A) = 0$ ou bien $\mu(A) = 1 - \mu(B) = 1$. □

Remarque 35 :
La proposition précédente reste vrai si on exige un seuil inférieur $n \geq n_0$ au lieu de $n \in \mathbb{N}$.

3.3 Moyenne temporelle et spatiale

Proposition 36 :
Les propriétés suivantes sont équivalentes :

1. le système est ergodique,
2. pour toute fonction $f \in L^1$, $\frac{1}{n} S_n(f) \xrightarrow{\text{p.s.}} \mu(f)$,
3. pour tous $A, B \in \mathcal{A}$, on a $\frac{1}{n} \sum_{k=0}^{n-1} \mu\big(B \cap T^{-k}(A)\big) \to \mu(A)\mu(B)$.

Démonstration. La première implication découle du théorème de Birkhoff.

Pour la seconde implication, par convergence dominée, on a :

$$\frac{1}{n} \sum_{k=0}^{n-1} \mu\big(B \cap T^{-k}(A)\big) = \int \mathbb{1}_B \times \frac{1}{n} S_n \mathbb{1}_A \,\mathrm{d}\mu \underset{n\to\infty}{\longrightarrow} \int \mathbb{1}_B \mathbb{1}_A^* \,\mathrm{d}\mu = \mu(A) \int \mathbb{1}_B \,\mathrm{d}\mu = \mu(A)\mu(B)\,.$$

Pour la troisième implication, on utilise la caractérisation précédente avec $B = A$. Si on a $T^{-1}(A) = A$, alors pour tout rang k, on a $T^{-k}(A) \cap B = A$. En conséquence, à la limite, $\mu(A) = \frac{1}{n} \sum_{k=0}^{n-1} \mu\big(T^{-k}(A) \cap B\big) \to \mu(A)^2$, d'où enfin $\mu(A) \in \{0, 1\}$. □

Corollaire 37 :
Soit un π-système $\mathcal{C}$ qui engendre la tribu $\mathcal{A}$.

Si pour tous $A, B \in \mathcal{C}$, on a $\frac{1}{n} \sum_{k=0}^{n-1} \mu\big(B \cap T^{-k}(A)\big) \to \mu(A)\mu(B)$, alors le système est ergodique.

Démonstration. Posons $\mathcal{M}$ l'ensemble des éléments $A \in \mathcal{A}$ pour lesquels la propriété précédente est vraie pour *tout* choix de $B \in \mathcal{C}$. Naturellement, $\emptyset \in \mathcal{M}$ et, par hypothèse, $\mathcal{C} \subset \mathcal{M}$.

On vérifie aisément que $\mathcal{M}$ est stable par passage au complémentaire. En effet, on a toujours $\mu(B \cap A^c) = \mu(B) - \mu(B \cap A)$. En outre, $T^{-1}(A^c) = T^{-1}(A)^c$. Si $A \in \mathcal{M}$, alors on a l'égalité

$\frac{1}{n}\sum_{k=0}^{n-1}\mu\big(B\cap T^{-k}(A^c)\big)=\mu(B)-\frac{1}{n}\sum_{k=0}^{n-1}\mu\big(B\cap T^{-k}(A)\big)\to\mu(B)-\mu(A)\mu(B)=\mu(A^c)\mu(B)$, d'où $A^c\in\mathcal{M}$.

En outre, pour une famille croissante $A_j\in\mathcal{M}^{\mathbb{N}}$, de limite $A\in\mathcal{A}$, on a donc l'inégalité $\left|\frac{1}{n}\sum_{k=0}^{n-1}\big(B\cap T^{-k}(A)\big)-\mu(A)\mu(B)\right|\leq\left|\frac{1}{n}\sum_{k=0}^{n-1}\big(B\cap T^{-k}(A_j)\big)-\mu(A_j)\mu(B)\right|+2\mu\big(A\cap A_j^c\big)$, donc on a bien l'égalité souhaitée en limite.

On peut dans un second temps suivre le même raisonnement pour les ensembles B, pour tout $A\in\mathcal{M}=\mathcal{A}$ cette fois-ci, et conclure. □

Définition 38 (Système fortement mélangeant) **:**
Un PPT est mélangeant si, pour tous $A,B\in\mathcal{A}$, on a $\mu(T^{-n}(A)\cap B)\to\mu(A)\mu(B)$.

Corollaire 39 :
Si un système est fortement mélangeant, alors il est ergodique.

Démonstration. Par le théorème de Cesàro, on obtient facilement la troisième caractérisation de l'ergodicité, issue du théorème précédent. □

La réciproque est fausse, et on en verra des exemples ci-après.

4 Cas des chaînes de Markov à espace d'états fini

Soient $\Omega = [d]$ et $X = \Omega^{\mathbb{N}}$ muni de la tribu cylindrique. On pose X_n les coordonnées canoniques du processus.

On considère une matrice de stochastique P telle que $P_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i)$, et une loi initiale ν sur Ω. Alors il existe une unique mesure $\mathbb{P}_\nu$ sur X telle que (X_n) est une P-chaîne de Markov et $X_0 \sim \nu$. Plus généralement, $X_n \sim \nu P^n$.

Notons θ l'opérateur de décalage, tel que $\theta(x_0, \ldots) = (x_1, \ldots)$. Si $\nu P = \nu$ est P-invariante, alors $(X, \mathbb{P}_\nu, \theta)$ est un PPT.

Pour tout $n \in \mathbb{N}^*$, on peut définir $\overline{P_n} = \frac{1}{n} \sum_{k=0}^{n-1} P^k$ une matrice stochastique également.

Proposition 40 :
La suite $\overline{P_n}$ est bornée dans $M_n(\mathbb{R})$, donc admet une valeur d'adhérence Q. La matrice Q est stochastique, et $PQ = QP = Q$.

En conséquence, chaque ligne de Q correspond à une mesure P-invariante.

Démonstration. L'ensemble des matrices stochastiques est un compact de $M_d(\mathbb{R})$, donc toute valeur d'adhérence est aussi stochastique. Il est clair que $QP = PQ$ car Q est la limite de polynômes en P, qui commutent naturellement avec P. Pour conclure sur $PQ = Q$, remarquons que $PQ = \lim \frac{1}{n} \sum_{k=1}^{n} P^k$, donc $\|PQ - Q\| = \lim \left\| \frac{P^n - I}{n} \right\| = 0$. □

Proposition 41 :
Si P est irréductible ($\forall i, j \in \Omega$, $\exists n > 0$, $P_{i,j}^n > 0$) alors la mesure invariante π est unique, et $(X, \mathbb{P}_\pi, \theta)$ est ergodique.

Démonstration. Soit $v \geq 0$ un vecteur propre de P, tel que $Pv = v$. Alors v est constant. En effet, soit $k = \operatorname{argmax}_{i \in [d]} v_i$ (en particulier $v_k > 0$). Alors $v_k = P^n v_k = \sum_{j=1}^{d} P_{k,j}^n v_j \leq \sum_{j=1}^{d} P_{k,j}^n v_k = v_k$. Pour tout j, on a n tel que $P_{k,j}^n > 0$, ce qui impose naturellement $v_j = v_k$.

Pour toute valeur d'adhérence Q de $\overline{P_n}$, comme $PQ = Q$, chaque colonne de Q est constante, donc les lignes de Q sont identiques, égales à π. Si ν est une probabilité P-invariante (vecteur propre à *gauche*), alors ν est Q-invariante, et $\nu_j = (\nu Q)_j = \sum_{i=1}^{d} \nu_i Q_{i,j} = \sum_{i=1}^{d} \nu_i \pi_j = \pi_j$.

La mesure P-invariante π est unique. En conséquence, la suite $\overline{P_n}$ admet une unique valeur d'adhérence, $\frac{1}{n} \sum_{k=0}^{n-1} P_{i,j}^k \xrightarrow[n \to \infty]{} \pi_j$.

Si on note $[i] = \{X_0 = i\}$, alors $\mathbb{P}\big([i] \cap \theta^{-k}([j])\big) = \pi_i P^k_{i,j}$. En conséquence :

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}\big([i] \cap \theta^{-k}([j])\big) = \pi_i \times \frac{1}{n}\sum_{k=0}^{n-1} P^k_{i,j} \to \pi_i \pi_j \,.$$

Plus généralement, si on considère deux cylindres $[a] = \{X_0 = a_0, \ldots, X_N = a_N\}$ et $[b]$, alors pour $k > N$ on a :

$$\mathbb{P}\big([a] \cap \theta^{-k}([b])\big) = \mathbb{P}([a]) P^{k-N}_{a_N,b_0} \prod_{j=0}^{M-1} P_{b_j,b_{j+1}} \,,$$

donc en passant à la limite on a :

$$\begin{aligned}
\tfrac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}\big([a] \cap \theta^{-k}([b])\big) &= \mathbb{P}([a]) \times \left(\tfrac{1}{n}\sum_{k'=0}^{n-1-N} P^{k'}_{a_N,b_0}\right) \times \prod_{j=0}^{M-1} P_{b_j,b_{j+1}} + O\big(\tfrac{1}{n}\big) \\
&\underset{n\to\infty}{\longrightarrow} \mathbb{P}([a]) \times \pi_{b_0} \prod_{j=0}^{M-1} P_{b_j,b_{j+1}} \\
&= \mathbb{P}([a])\mathbb{P}([b]) \,.
\end{aligned}$$

Comme les cylindres forment un π-système qui engendre la tribu, on en déduit l'ergodicité du PPT. □

Lorsque P est irréductible, on a donc un système ergodique.

Lemme 42 :

Si la chaîne P est irréductible périodique, de période $d \geq 2$, alors le PPT ergodique précédent n'est *pas* fortement mélangeant.

Remarque 43 :
Il suffit de vérifier que $\mu\big(T^{-k}(A)\cap B\big) \to \mu(A)\mu(B)$ sur un π-système qui engendre la tribu pour conclure que le système est mélangeant.

5 Produits, facteurs et extensions

5.1 Produit direct

Définition 44 (Produit direct) **:**
Deux MPT $\mathcal{X}_i = (X_i, \mathcal{A}_i, \mu_i, T_i)$ ($i \in \{0,1\}$) induisent un MPT produit $\mathcal{X}_0 \otimes \mathcal{X}_1$.

Si les espaces sont des PPT, alors leur produit aussi.

Remarque 45 :
Si le produit est ergodique, alors les $\mathcal{X}_i$ le sont.

La réciproque n'est pas forcément vraie. Ainsi, $x \mapsto x+1$ sur $\mathbb{F}_2$ muni de la mesure uniforme est ergodique, mais le produit de cet espace avec lui-même a deux 2-orbites disjointes.

Lemme 46 :
Si $\mathcal{X}_0$ est ergodique et $\mathcal{X}_1$ est mélangeant, alors le produit est ergodique.

Démonstration. Il suffit de vérifier la propriété sur le π-système des ensembles produits $A_0 \times A_1$, ce qui se fait sans problèmes. □

5.2 Produit croisé

Soit T le shift sur $X = \{0,1\}^{\mathbb{N}}$ muni de la mesure uniforme. On pose $\Phi(\omega) = (-1)^{\omega_0}$. Alors $X \times \mathbb{Z}$ muni de $T_\varphi(x,k) = (T(x), k + \varphi(x))$ correspond à la marche aléatoire usuelle sur $\mathbb{Z}$. Comme $k \mapsto k + \varphi(x)$ préserve la mesure de comptage λ pour tout choix de x, T_φ préserve $\mu \times \lambda$.

Définition 47 (Produit croisé) **:**
Soient $\mathcal{X} = (X, T, \mu)$ un MPT, et (Y, λ) un espace mesuré.

Si $\varphi : X \times Y \to Y$ est mesurable, et pour tout $x \in X$, λ est φ_x-invariante, alors $\mu \otimes \lambda$ est T_φ-invariante, avec $\mathbb{T}_\varphi(x,y) := (T(x), \varphi(x,y))$.

En conséquence, $(X \times Y, T_\varphi, \lambda \otimes \mu)$ est un MPT.

Si $\mathcal{X}$ est un PPT et μ une mesure de probabilité, alors le produit croisé est un PPT.

Remarque 48 :
Soit $Y = (G, *)$ un groupe topologique localement compact, à base dénombrable d'ouverts, et λ_G une mesure de Haar à gauche sur G, telle que $\int_G f(g * h)\,\mathrm{d}h = \int_G f(h)\,\mathrm{d}h$ pour tout $g \in G$.

Pour toute application $\varphi : X \to G$ mesurable, la transformation $T_\varphi(x, g) = (T(x), \varphi(x) * g)$ induit un produit croisé.

Corollaire 49 :
Si de plus $X = G^{\mathbb{N}}$ muni d'une mesure produit $\rho^{\mathbb{N}}$ et du shift, alors $\varphi(g) = g_0$ convient.

Le produit croisé obtenu correspond à la marche aléatoire stationnaire sur G, avec des pas iid de loi ρ, et de mesure stationnaire λ_G.

5.3 Facteurs, extensions, isomorphismes

Définition 50 (Facteur, extension) **:**
Soient $\mathcal{X}$ et $\mathcal{Y}$ deux MPT et $\pi : X \to Y$ mesurable.

On dit que π est un morphisme lorsqu'il existe $X' \in \mathcal{A}_X$ et $Y' \in \mathcal{A}_Y$ de mesures pleines, tels que $T_X(X') \subset X'$ et $T_Y(Y') \subset Y'$, $\pi : X' \to Y'$ est surjective et $\pi \circ T_X = T_Y \circ \pi$ sur X'. On veut enfin que $\pi * \mu_X = \mu_Y$.

On dit alors que Y est un facteur de X, et que X est une extension de Y.

Remarque 51 :
Sur un produit direct, les projections sont des morphismes, donc $\mathcal{X}_1$ et $\mathcal{X}_2$ sont des facteurs de $\mathcal{X}_1 \otimes \mathcal{X}_2$.

Si $(X \times G, T_\varphi)$ est un produit croisé, alors X est un facteur de $X \times G$. A priori, on n'a pas forcément G facteur du produit croisé.

Si $\mathcal{X} = (X, \mathcal{A}, \mu, T)$ et que T est $\mathcal{B}$-mesurable avec $\mathcal{B} \subset \mathcal{A}$, alors $\mathcal{Y} = (X, \mathcal{B}, \mu, T)$ est un facteur de $\mathcal{X}$ via l'injection canonique.

Définition 52 (Isomorphisme) **:**
Si $\pi : X' \to Y'$ est une bijection bimesurable, alors X et Y sont isomorphes.

Remarque 53 :
La bimesurabilité est essentielle. Ainsi, en reprenant l'exemple précédent, en réduisant la

tribu $\mathcal{A}$ à la tribu $\mathcal{B}$, T est l'identité donc est inversible, mais l'image directe d'un borélien de $\mathcal{A}\backslash\mathcal{B}$ n'est pas mesurable à l'arrivée, donc T^{-1} n'est pas mesurable.

Proposition 54 :

Soient $(\mathbb{T}, T_2 : x \mapsto 2x, \lambda)$ le tore muni de la mesure de Lebesgue, et $\left(\{0,1\}^{\mathbb{N}}, \left(\frac{\delta_0+\delta_1}{2}\right)^{\mathbb{N}}, S\right)$ une suite de variables $\mathcal{B}\left(\frac{1}{2}\right)$ iid munie du shift S.

Ces deux PPT sont isomorphes.

Démonstration. Quite à identifier le tore $\mathbb{T}$ à l'intervalle $[0,1[$, considérons $f(x) = \mathbb{1}_{x<\frac{1}{2}}$. Posons $\pi(x) = (f \circ T_2^n(x))_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$. Naturellement, par construction, $\pi \circ T_2 = S \circ \pi$.

En outre, pour tout intervalle $\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[$ dyadique, on a clairement :

$$\lambda\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\right) = \frac{1}{2^n} = \mu\left(T_2\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\right)\right),$$

puisqu'on impose la valeur des n premières coordonnées de la suite dans ce cas.

Enfin, on vérifie que l'image $Y' = \pi(X)$ est l'ensemble des suites qui ne sont pas égales à 1 à partir d'un rang, et que cet espace est de mesure pleine.

L'application $\pi : X \to Y'$ est bien une bijection bimesurable, puisqu'elle envoie les intervalles dyadiques sur des cylindres, et que ces mesurables engendrent les tribus associées à X et Y. $\square$

6 Entropie métrique

6.1 Entropie d'une partition

Soit $(X, \mathcal{A}, \mu)$ un système probabilisé.

Définition 55 (Information d'un évènement) **:**
On pose $I(A) = -\log_2(\mu(A))$.

Cette application est positive, décroissante pour l'inclusion, telle que $I(X) = 0$. Lorsque A et B sont indépendants, on a $I(A \cap B) = I(A) + I(B)$.

Définition 56 (Entropie d'une partition) **:**
Soit $P = (P_i)_{i \in I}$ une partition finie (ou dénombrable) de X.

Si on note $P(x)$ comme la classe de x dans la partition, c'est une variable aléatoire sur X. On considère alors $H(P) = \mathbb{E}[I(P)] = \sum_{i \in I} \mu(P_i) I(P_i)$.

Remarque 57 :
Soit $\Phi : p \mapsto -p \log_2(p)$ définie sur $[0, 1]$. Cette application est strictement concave, positive.

Lemme 58 :
Si P est une partition en k ensembles, alors $H(P) \leq \log_2(k)$, avec égalité ssi $\mu(P_i) = \frac{1}{k}$ pour tout $1 \leq i \leq k$.

Démonstration. On utilise l'inégalité de Jensen : $H(P) = k \times \sum_{i=1}^{k} \frac{1}{k} \Phi(\mu(P_i)) \leq k\Phi\left(\frac{1}{k}\right) = \log_2(k)$, avec égalité ssi $\mu(P_i) = \mu(P_j)$ pour tous $1 \leq i, j \leq k$. $\square$

Définition 59 (Information conditionnelle) **:**
On pose $I(P|Q) = I(P \cap Q) - I(Q) \geq 0$ pour des évènements P et Q.

De même, pour deux partitions P et Q, on a :

$$\begin{array}{rcl} H(P|Q) & := & \mathbb{E}[I(P(x)|Q(x))] \\ & = & -\sum_{i\in I}\sum_{j\in J}\mu(P_i\cap Q_j)\log_2(\mu(P_i|Q_j)) \\ & = & -\sum_{i\in I}\sum_{j\in J}\mu(Q_j)\Phi(\mu(P_i|Q_j)) \\ & = & \sum_{j\in J}\mu(Q_j)H_{\mu(\cdot|Q_j)}(P) \\ & \geq & 0 \end{array}$$

Définition 60 (Intersection de partitions) **:**
On pose $P\vee Q=\{P_i\cap Q_j,\, i\in I,\, j\in J\}$.

Lemme 61 :
On a les propriétés suivantes :

1. $H(P\vee Q)=H(Q)+H(P|Q)$,
2. si Q raffine P ($P\vee Q=Q$), alors $H(P)\leq H(Q)$,
3. $H(P\vee Q|R)=H(Q|R)+H(P|Q\vee R)$,
4. $H(P\vee Q|Q)=H(P|Q)$,
5. $H(P|P)=0$.

Définition 62 :
Les partitions P et Q sont indépendantes lorsque, pour tous $i\in I$ et $j\in J$, on a P_i et Q_j indépendants.

Proposition 63 :
On a P et Q indépendantes ssi $H(P\vee Q)=H(P)+H(Q)$ ssi $H(P|Q)=H(P)$.

Démonstration. L'équivalence entre les deux propriétés de droite découle du lemme précédent. D'autre part, par Jensen, on a :

$$H(P|Q)=\sum_{i,j}\mu(Q_j)\Phi(\mu(P_i|Q_j))\leq\sum_i\Phi\left(\sum_j\mu(Q_j)\mu(P_i|Q_j)\right)=H(P)$$

avec égalité ssi pour tous $i\in I$ et $j,j'\in J$ on a $\mu(P_i|Q_j)=\mu(P_i|Q_{j'})$ donc en particulier égaux à $\mu(P_i)$, autrement dit on a l'indépendance entre P_i et les Q_j. $\square$

Proposition 64 :

On a les propriétés :

1. $H(P|Q) \geq H(P|Q \vee R)$,
2. $H(P \vee Q) \leq H(P) + H(Q)$,
3. $H(P \vee Q|R) \leq H(P|R) + H(Q|R)$,
4. si R raffine Q, alors $H(P|Q) \geq H(P|R)$.

Remarque 65 (Entropie conditionnelle par rapport à une tribu) **:**

Étant donnée une sous-σ-algèbre $\mathcal{G} \subset \mathcal{A}$, on peut définit une mesure conditionnelle $\mu(A|\mathcal{G})$, une variable $\mathcal{G}$-mesurable à valeurs dans $[0, 1]$.

On peut alors définir l'information conditionnelle $I(A|\mathcal{G})$ (aléatoire), et l'entropie conditionnelle $H(P|\mathcal{G})$ en passant à l'espérance. Lorsque $\mathcal{G}$ est la tribu engendrée par une partition finie, on retrouve l'entropie définie précédemment.

Lorsque $\mathcal{G}$ est la tribu engendrée par une partition Q, on retrouve l'entropie précédemment introduite.

Plus généralement, soient (Q_n) une suite de partitions de plus en plus fines, et $\mathcal{G} = \sigma\left(\bigcup_{n\in\mathbb{N}} Q_i\right)$. Dans ce cas, on a $\mu(A|Q_n) \xrightarrow{L^1,\text{p.s.}} \mu(A|\mathcal{G})$, et donc convergence presque-sûre de l'information conditionnelle. Ceci ne garantit pas directement la convergence L^1 de l'information, autrement dit la convergence de l'entropie.

6.2 Entropie d'une PPT

Soit $(X, \mathcal{A}, \mu, T)$ un PPT.

6.2.1 Entropie d'une PPT par rapport à une partition

Définition 66 (Tiré en arrière d'une partition) **:**

Toute parition $P = (P_i)_{i\in I}$ induit une partition $T^{-1}(P) = (T^{-1}(P_i))_{i\in I}$.

On vérifie aisément que $H(T^{-1}(P)) = H(P)$. Plus généralement, les propriétés précédentes restent vraies si on remplace toutes les partitions P (resp. Q et R) par $T^{-1}(P)$ (resp. $T^{-1}(Q)$ et $T^{-1}(R)$).

Définition 67 (Partition itérée) **:**

Soit P une partition. On pose $P_0^n = \bigvee_{k=0}^{n-1} T^{-k}(P)$. Plus généralement, on note $P_i^j = \bigvee_{k=i}^{j-1} T^{-k}(P)$.

Définition 68 :

Entropie de T par rapport à P On pose $h(T,P) = \overline{\lim_{n\to\infty}} \frac{1}{n} H(P_0^n)$.

Remarque 69 :

Dans le cas d'un 5-cycle muni de la mesure uniforme, avec la partition P complète, on $P_0^n = P$ pour tout n, et $H(P) = \log_2(5) < \infty$, on a à limite $h(T,P) = 0$.

Dans le cas d'une suite de variables iid sur Ω fini, munie du shift, si P est une partition selon la valeur de la première variable, on a $H(P_0^n) = nH(P)$, d'où $h(T,P) = H(P)$.

Théorème 70 :

On a en fait :

$$h(T,P) = \lim_{n\to\infty} \frac{1}{n} H(P_0^n) = \inf_{n\geq 1} \frac{1}{n} H(P_0^n) = \lim_{n\to\infty} H(P|P_1^n)\,.$$

En particulier, le dernier point garantit $h(T,P) \leq H(P)$.

Démonstration. On a $P_0^n = P_0^i \vee P_i^n$. En conséquence, $H\left(P_0^{n+m}\right) \leq H(P_0^n) + H(P_n^{n+m})$ et en outre $P_n^{n+m} = T^{-n}(P_0^m)$. Par T-invariance, on en déduit que la suite $(H(P_0^n))_{n\in\mathbb{N}}$ est sous-additive, donc par le lemme de Fekete, on a bien la convergence souhaitée, $h(T,P)$ est une vraie limite.

D'autre part, on a $H(P_0^n) = H(P|P_1^n) + H(P_1^n) = H(P|P_1^n) + H\left(P_0^{n-1}\right) = \sum_{i=1}^{n} H(P|P_1^n)$. La suite $(H(P|P_1^n))_{n\geq 1}$ est décroissante, positive, donc convergente vers $h_\infty \geq 0$. Par convergence de Cesàro, $\frac{1}{n} H(P_0^n) \to h_\infty$ est également une limite décroissante, ce qui conclut la preuve. $\square$

Remarque 71 :

On pourrait montrer que, en notant $P_1^\infty = \sigma\big(T^{-k}(P),\, k \geq 1\big)$, on a $h(T,P) = H(P|P_1^\infty)$.

Proposition 72 :

Soient P et Q deux partitions finies. On a les propriétés suivantes :

1. $h(T,P) \leq h(T, P\vee Q) \leq h(T,P) + h(T,Q)$,
2. $h(T,P) \leq h(T,Q) + H(P|Q)$,
3. $h(T, T^{-1}(P)) = h(T,P)$,
4. $h(T, P_0^n) = h(T,P)$.

Démonstration. Le premier point découle directement de :

$$H(P_0^n) \leq H(P_0^n \vee Q_0^n) = H((P\vee Q)_0^n) \leq H(P_0^n) + H(Q_0^n)\,.$$

Le second point découle de :

$$H(P_0^n) \leq H(P_0^n \vee Q_0^n) = H(Q_0^n) + H(P_0^n | Q_0^n),$$

avec en particulier :

$$H(P_0^n|Q_0^n) \leq \sum_{k=0}^{n-1} H\big(T^{-k}(P)\big|Q_0^n\big) \leq \sum_{k=0}^{n-1} H\big(T^{-k}(P)\big|T^{-k}(Q)\big) = nH(P|Q).$$

Le troisième point est direct, et le dernier découle de $(P_0^n)_0^m = P_0^{n+m-1}$. □

Définition 73 (Entropie d'une transformation) **:**
On pose $h(T) = \sup\{h(T,P),\ P \text{ partition finie}\}$.

Définition 74 :
Deux sous-tribus $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$ coïncident modulo 0 si pour tout $A \in \mathcal{F}$, il existe $B \in \mathcal{G}$ tel que $\mu(A\Delta B) = 0$ et inversement, pour tout $B \in \mathcal{G}$, il existe $A \in \mathcal{F}$ tel que $\mu(A\Delta B) = 0$.

Définition 75 (Partition génératrice) **:**
On dit que P est une partition génératrice si P_0^∞ coïncide avec $\mathcal{A}$.

Remarque 76 :
C'est par exemple le cas de la partition selon la valeur de X_1, pour une suite de variables iid munie du shift.

Lemme 77 :
Soit $k \in \mathbb{N}^*$. Il existe $\psi_k : [0,1] \to \mathbb{R}^+$ croissante, nulle et continue en 0, de sorte que pour toutes k-partitions $(P_i)_{1\leq i\leq k}$ et $(Q_i)_{1\leq i\leq k}$, si pour tout $1 \leq i \leq k$ on a $\mu(P_i\Delta Q_i) \leq \delta$ alors on a $H(P|Q) \leq \psi_k(\delta)$.

Démonstration. On considère $R = \{R_{i,j} := P_i \cap Q_j,\ i \neq j\} \sqcup \{R_0\}$ où $R_0 = \bigsqcup_{i=1}^k (P_i \cap Q_i)$. Par construction, $P \vee Q = R \vee Q$, donc $H(P|Q) = H(R|Q) \leq H(R)$.

On vérifie sans soucis que $\mu(R_{i,j}) \leq \delta$, et donc $\mu(R_0) \leq 1-k\delta$. Par convexité, on majore $H(R)$ par $\psi_k(\delta)$ l'entropie associée à la distribution $(\delta, \ldots, \delta, 1-k\delta)$ lorsque $\delta \leq \frac{1}{k}$. Par continuité de l'entropie, pour $\delta \to 0$, $\psi_k(\delta) \to 0$. □

Lemme 78 :
Soient P une k-partition et Q une partition quelconque. Si on a des $A_i \in \sigma(Q)$ (pour tout $1 \leq i \leq k$) tels que $\mu(A_i\Delta P_i) \leq \delta$, alors $H(P|Q) \leq \psi_k(2k^2\delta)$.

Démonstration. La famille (A_i) n'est pas nécessairement une partition ici. Si on peut construire une partition R à partir des A_i, telle que $\sigma(R) \subset \sigma(Q)$, alors on aura $H(P|R) \geq H(P|Q)$.

On veut donc une partition R, telle que $\mu(P_i \Delta R_i) \leq 2k^2\delta$. Posons $R_i = A_i \backslash R_{i-1}$ pour $1 \leq i < k$, et finalement $R_k = \left(\bigcup_{i=1}^{k-1} A_i\right)^c = \left(\bigsqcup_{i=1}^{k-1} R_i\right)^c$. Ce faisant, on a bien une k-partition. $\square$

Théorème 79 (Kolmogorov-Sinai) **:**
Si P est une partition génératrice, alors $h(T) = h(T, P)$.

Démonstration. On a naturellement $h(T, P) \leq h(T)$.

Soit Q une k-partition. Comme P est génératrice, on peut approcher Q à ε près par une partition R, raffinée par P_0^n à partir d'un rang. On a alors $H(Q|R) \leq \psi_k(2k^2\delta)$. En particulier :

$$h(T, Q) \leq h(T, R) + \psi_k\big(2k^2\delta\big) \leq h(T, P_0^n) + \psi_k\big(2k^2\delta\big) = h(T, P) + \psi_k\big(2k^2\delta\big) .$$

Quitte à prendre $\delta \to 0$, on a donc $h(T, Q) \leq h(T, P)$, d'où l'égalité. □

Remarque 80 :
Si deux PPT sont isomorphes, alors ils ont la même entropie $h(T)$.

Considérons ainsi les suites sur l'alphabet $\Omega_d = [\![1, d]\!]$, munies du shift S et de la mesure uniforme.

Soit P une partition selon le premier terme de la suite. Par définition de la tribu cylindrique, P est génératrice. En conséquence $h(S) = h(S, P)$.

Or $H(P_0^n)$ est l'entropie d'une mesure uniforme sur d^n états, donc est égale à $n \log_2(d)$. On a donc $h(S) = \log_2(d)$. On en déduit donc que, si $d \neq d'$, les PPT ne sont pas isomorphes.

Comme $T_d : x \mapsto dx$ sur $\mathbb{T}$ est isomorphe à une suite de variables uniformes iid sur Ω_d, on en déduit que les transformations de l'espace probabilisé $(\mathbb{T}, \lambda)$ n'ont pas la même entropie, donc ne sont pas isomorphes.

6.3 Entropie locale

Théorème 81 (Shannon, Mc Millan, Breimann) **:**
Soit P une partition. On a la convergence $\frac{1}{n} I(P_0^n(x)) \overset{\mu-\text{p.s.}}{\longrightarrow} h(T, P)$.

Démonstration. Voir Viana, section 9.3. □

Remarque 82 :
Dans le cas d'une suite de variables $(X_i)_{i \in \mathbb{N}}$ iid, d'un espace $\big(\Omega^{\mathbb{N}}, \rho^{\mathbb{N}}, S\big)$ muni du shift, lorsque la partition P ne dépend que de la première valeur, alors la loi des grands nombres donne :

$$\frac{1}{n} I(P_0^n(x)) = \frac{1}{n} \sum_{i=0}^{n-1} I(P(X_i)) \longrightarrow \mathbb{E}(I(P(X))) = H(P) = h(S, P) .$$

7 Théorie spectrale

Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé standard, dont la tribu est en particulier engendrée par une famille dénombrable de parties.

Dans ce chapitre, on s'intéresse à $U_T : f \mapsto f \circ T$ en tant qu'isométrie sur l'espace de Hilbert $L^2(X, \mu, \mathbb{C})$ complexe, muni du produit hermitien $\langle f, g\rangle = \int f\overline{g}\,\mathrm{d}\mu$.

7.1 Spectre ponctuel

Définition 83 (Spectre ponctuel) **:**
On pose $\sigma(U)$ le spectre de l'opérateur U, l'ensemble de ses valeurs propres.

Lemme 84 :
On a les propriétés suivantes :

1. Si U est une isométrie, alors $\sigma(U) \subset \mathbb{U}$,
2. la fonction constante $\mathbb{1}$ est une 1-fonction propre de U_T,
3. si f est une fonction propre de U_T associée à $\lambda \neq 1$, alors $\mu(f) = 0$.

Proposition 85 :
Si le PPT est ergodique, les seules 1-fonctions propres sont les constantes. Autrement dit, l'espace propre associé à 1 est de dimension 1.

Si le PPT est fortement mélangeant, alors $\sigma(U_T) = \{1\}$, les seules fonctions propres sont les constantes.

Démonstration. Le premier point a déjà été vu. Pour le second point, si f est une λ-fonction propre, alors par la caractérisation L^2 de la propriété de mélange :

$$\lambda^n \|f\|^2 = \langle U_T^n(f), f\rangle = \mu\big(f \circ T^n \times \overline{f}\big) \longrightarrow \mu(f)\mu\big(\overline{f}\big) = |\mu(f)|^2,$$

donc λ^n est une suite convergente, nécessairement $\lambda = 1$. □

Proposition 86 :
Si deux PPT sont isomorphes, alors ils ont le même spectre.

Remarque 87 (Cas des translations du Tore) **:**
Soit $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ et $T_\alpha : \mathbb{T} \to \mathbb{T}$ la translation sur le tore associée.

Alors par analyse de Fourier, on constate que les valeurs propres sont les $\{e^{2i\pi\alpha},\ k \in \mathbb{Z}\}$ et les fonctions propres associées sont les $\left\{x \mapsto e^{2i\pi k\alpha},\ k \in \mathbb{Z}\right\}$.

En particulier, si $\beta \notin \pm\alpha + \mathbb{Z}$, alors les PPT associés aux deux translations ne sont pas isomorphes.

7.2 Mesure spectrale

Proposition 88 :

Soit $f \in L^2$. Alors il existe une unique mesure σ_f positive finie sur $\mathbb{U}$, telle que :

— si $n \in \mathbb{N}$, alors $\widehat{\sigma_f}(n) := \int_{\mathbb{U}} z^n \,\mathrm{d}\sigma_f(z) = \langle U_T^n(f), f\rangle$,

— si $n \in -\mathbb{N}$, alors $\widehat{\sigma_f}(n) := \left\langle f, U_T^{-n}(f)\right\rangle$.

On appelle cette mesure la mesure spectrale de f.

Démonstration. On utilise le théorème de Herglotz : si $(r_n)_{n\in\mathbb{Z}}$ est de type positif (autrement dit, pour tout $n \in \mathbb{Z}$, $r_{-n} = \overline{r_n}$) et telle que, pour toute suite $(a_n) \in \mathbb{C}^{\mathbb{Z}}$ et pour tout $N \in \mathbb{N}$, on a $\sum\limits_{-N\leq m,n\leq N} r_{n-m}a_n\overline{a_m} \in \mathbb{R}^+$, alors il existe une unique mesure positive finie telle que $\widehat{\sigma}(n) = r_n$.

Dans notre cas, la suite est $r_n = \langle U_T^n(f), f\rangle$ pour $n \geq 0$ et $r_n = \langle f, U_T^n(f)\rangle$ lorsque $n \leq 0$. Lorsque $-N \leq m \leq n \leq N$, on a naturellement :

$$r_{n-m} = \left\langle U_T^{n-m}(f), f\right\rangle = \left\langle U_T^{N+n-m}(f), U_T^N(f)\right\rangle = \left\langle U_T^{N+n}(f), U_T^{N+m}(f)\right\rangle,$$

et on a le même résultat si $n \leq m$. On vérifie alors que :

$$\begin{aligned}
\sum_{-N\leq m,n\leq N} r_{n-m}a_n\overline{a_m} &= \sum_{-N\leq m,n\leq N} \left\langle a_n U_T^{N+n}(f), a_m U_T^{N+m}(f)\right\rangle \\
&= \left\| \sum_{n=0}^{2N} a_{n-N} U_T^n(f) \right\|_2^2 \\
&\geq 0\,.
\end{aligned}$$

□

Remarque 89 :

Si $U_T(f) = \lambda f$ est une fonction propre, alors on a $\int_{\mathbb{U}} z^n \,\mathrm{d}\sigma_f(z) = \lambda^n \|f\|^2$ pour tout $n \in \mathbb{N}$, d'où $\sigma_f = \delta_\lambda \|f\|^2$ par unicité.

Proposition 90 :
Soit $f \in L^2(\mu)$ telle que $\mu(f) = 0$. Si σ_f a un atome en λ, alors il existe une λ-fonction propre non nulle (non constante).

Démonstration. Soit $F_N = \frac{1}{N} \sum\limits_{n=0}^{N-1} \frac{1}{\lambda^n} U_T^n(f)$. Cette suite est uniformément bornée dans L^2, donc elle a une sous-suite faiblement convergence vers une limite F, telle que $\langle h, F_{\theta(n)} \rangle \longrightarrow \langle h, F \rangle$ pour toute fonction $h \in L^2$, par Banach-Alaoglu.

On a $U_T(F) = \lambda F$ ssi la propriété est vraie en prenant le produit scalaire contre toute fonction $h \in L^2(\mu)$. En outre, $U_T(F_N) = \lambda F_N + \frac{1}{N}\left(f - \frac{1}{\lambda^N} U_T^N(f)\right) = \lambda F_N + O\left(\frac{1}{N}\right)$. On a donc :

$$\langle h, U_T(F) \rangle = \langle U_T^*(h), F \rangle = \lim_{n\to\infty} \langle U_T^*(h), F_{\theta(n)} \rangle = \lim_{n\to\infty} \langle h, U_T(F_{\theta(n)}) \rangle = \lambda \lim_{n\to\infty} \langle h, F_{\theta(n)} \rangle = \lambda \langle h, F \rangle \,,$$

donc F est bien une λ-fonction propre.

Reste à justifier que $F \neq 0$. Pour ce faire, remarquons que :

$$\begin{aligned}
\langle F, f \rangle = \lim_{n\to\infty} \langle F_{\theta(n)}, f \rangle &= \lim_{n\to\infty} \frac{1}{\theta(n)} \sum_{k=0}^{n-1} \frac{1}{\lambda^k} \langle U_T^k(f), f \rangle \\
&= \sigma_f(\{\lambda\}) + \lim_{n\to\infty} \int_{\mathbb{U}} \mathbb{1}_{z\neq\lambda} \frac{1}{\theta(n)} \sum_{k=0}^{\theta(n)-1} \left(\frac{z}{\lambda}\right)^k \mathrm{d}\sigma_f(z) \\
&= \sigma_f(\{\lambda\}) + \lim_{n\to\infty} \int \mathbb{1}_{z\neq\lambda} \frac{1}{\theta(n)} \times \frac{\lambda^k - z^k}{\lambda - z} \mathrm{d}\sigma_f(z) \,.
\end{aligned}$$

Pour $z \neq \lambda$ fixé, la fonction intégrée converge vers 0. En outre, on peut majorer cette moyenne de $\theta(n)$ termes par 1. Par convergence dominée, $\langle F, f \rangle = \sigma_f(\{\lambda\}) \neq 0$. En conséquence, $F \neq 0$ est une λ-fonction propre. $\square$

7.3 Mélange faible et spectre continu

Définition 91 (Spectre continu) **:**
On dit que U_T a un spectre continu lorsque, pour toute fonction $f \in L^2$ telle que $\mu(f) = 0$, on a σ_f sans atomes.

Définition 92 (Mélange faible) **:**
On dit qu'un PPT admet la propriété de mélange faible lorsque, pour tous évènements A et B, on a $\frac{1}{n} \sum\limits_{i=0}^{n} \left| \mu\left(A \cap T^{-k}(B)\right) - \mu(A)\mu(B) \right| \underset{n\to\infty}{\longrightarrow} 0$.

Remarque 93 :

Par convergence de Cesàro, le mélange fort implique le mélange faible.

Par inégalité triangulaire, le mélange faible implique l'ergodicité.

Lemme 94 :

De façon équivalente, le mélange faible est caractérisé par :

$$\forall f,g \in L^2,\ \frac{1}{n}\sum_{k=0}^{n-1}\left|\left\langle U_T^k(f),g\right\rangle - \langle f,\mathbb{1}\rangle\langle \mathbb{1},g\rangle\right| \underset{n\to\infty}{\longrightarrow} 0\,.$$

Théorème 95 :

On a l'équivalence entre les propriétés :

1. le PPT admet le mélange faible,
2. les seules fonctions propres sont les constantes,
3. le spectre est continu.

Démonstration. Pour l'implication ($2 \Rightarrow 3$), on a vu que si le spectre n'est pas continu, on a une λ-fonction propre pour un certain $\lambda \neq 1$, non constante.

Pour l'implication ($1 \Rightarrow 2$), considérons $\lambda \neq 1$, et f telle que $U_T(f) = \lambda f$. Comme le système est faiblement mélangeant et $\mu(f) = 0$, on a $\frac{1}{n}\sum\limits_{k=0}^{n-1}\left|\left\langle U_T^k(f),f\right\rangle\right| = \|f\|^2 \longrightarrow 0$, d'où $f = 0$.

Enfin, considérons l'implication ($3 \Rightarrow 1$). Commençons par montrer la propriété de mélange faible lorsque $f = g$ et $\mu(f) = 0$. On a :

$$\left|\left\langle U_T^k(f),f\right\rangle\right|^2 = \int_{\mathbb{U}} z^k \,\mathrm{d}\sigma_f(z) \int_{\mathbb{U}} w^{-k}\,\mathrm{d}\sigma_f(w) = \int_{\mathbb{U}^2}\left(\frac{z}{w}\right)^k \mathrm{d}\sigma_f(z)\,\mathrm{d}\sigma_f(w)\,.$$

Quitte à passer à la somme :

$$\frac{1}{n}\sum_{k=0}^{n-1}\left|\left\langle U_T^k(f),f\right\rangle\right|^2 = \int_{\mathbb{U}^2}\frac{1}{n}\times\frac{w^n - z^n}{w - z}\,\mathrm{d}\sigma_f(z)\,\mathrm{d}\sigma_f(w)\,,$$

avec le terme dans l'intégrale égal à 1 lorsque $w = z$, par continuité. La mesure σ_f est sans atomes, la fonction est majorée par 1 et converge simplement vers 0 dès que $w \neq z$, d'où la limite souhaitée par convergence dominée.

On utilise alors le lemme de Koopmann-Von Neuman : une suite positive (a_n) converge vers 0 au sens de Cesàro ssi il existe $N \subset \mathbb{N}$, de densité pleine, ($\frac{1}{n}\#(N \cap [\![0,n-1]\!]) \longrightarrow 1$), tel que l'extraction de a selon N tend vers 0.

On déduit directement de ce lemme que, dans le cas où a est positive, on a a qui converge vers 0 au sens de Cesàro ssi a^2 converge de même. En appliquant cette propriété à la convergence établie, on a finalement la propriété de mélange faible pour le couple (f,f).

Plus généralement, avec $g \in L^2$ et $f = g - \mu(g)$, on vérifie que :

$$\left\langle U_T^k(g), g \right\rangle - \langle g, \mathbb{1} \rangle \langle \mathbb{1}, g \rangle = \left\langle U_k^T(f), f \right\rangle,$$

dont on déduit la propriété de mélange faible pour (g, g).

Soient désormais $f \in L^2$ fixée. On pose F_f le sous-ensemble de L^2 des fonctions g qui vérifient la propriété de mélange faible contre f. C'est clairement un sous-espace vectoriel, qui contient f et $\mathbb{1}$. On peut vérifier que F_f est stable par U_T, et fermé. En conséquence, F_f contient l'espace engendré par la suite $\left(U_T^k(f)\right)$. Si $g \in F_f^\perp$, alors tous les termes utilisés dans la propriété de mélange faible sont nuls (sauf $\mu(f)$, qu'on multiplie par $\overline{\mu(g)} = 0$). En conséquence, g vérifie de fait la propriété de mélange faible contre f, donc $g \in F_f$, $g = 0$. Autrement dit, $F_f = L^2$ pour toute fonction f, d'où un PPT qui a le mélange faible. □

8 Existence de mesures invariantes

Soit (X, d) métrique compact, muni de sa tribu borélienne. On considère $\mathcal{M}^1(X)$ l'ensemble des probabilités sur cet espace.

8.1 Topologie préfaible

On considère l'espace $(\mathcal{C}^0(X, \mathbb{R}), \|\cdot\|_\infty)$. C'est un espace polonais.

Définition 96 (Topologie préfaible) **:**

On peut voir $\mathcal{M}^1(X)$ comme un ensemble de formes linéaires 1-lipschitziennes sur $\mathcal{C}^0(X)$, convexe.

On dit que (μ_n) converge vers μ pour la topologie préfaible (faible-*) lorsque, pour toute fonction $f \in \mathcal{C}^0(X)$, on a $\mu_n(f) \to \mu(f)$. On note $\mu_n \xrightarrow{w} \mu$.

Cette topologie est métrisable, en fixant une suite (f_n) dense dans $\mathcal{C}^0(X)$, via la distance $d(\mu, \nu) = \sum_{n=0}^{\infty} \frac{|\mu(f_n) - \nu(f_n)|}{2^n \|f_n\|_\infty}$. En général, la suite et donc la distance associée ne sont pas canoniques.

Théorème 97 :

L'espace $\mathcal{M}^1(X)$ est compact pour la topologie préfaible.

Démonstration. Ce résultat découle du théorème de représentation de Riesz-Markov-Kakutani, qui dit que toute forme linéaire continue L sur $\mathcal{C}^0(X)$, positive (si $f \geq 0$ alors $L(f) \geq 0$) telle que $L(\mathbb{1}) = \mathbb{1}$, L peut être vue comme une mesure de probabilité.

Ainsi, pour toute suite de mesures (μ_n), par extraction diagonale, on obtient une sous-suite telle que $(\mu_n(f_j))_n$ converge vers une limite $L(f_j)$ pour tout $j \in \mathbb{N}$, avec $(f_j)_j$ une suite de fonctions dénombrable dense. Par densité de la suite, on vérifie que L se prolonge bien de façon unique en une forme linéaire positive, et donc en une mesure μ, telle que $\mu(f)$ est bien limite de $(\mu_n(f))_n$ pour toute fonction f. $\square$

8.2 Existence de mesures invariantes

Proposition 98 :

Soient X métrique compact et $T \in \mathcal{C}^0(X,X)$.

Pour une mesure $\nu \in \mathcal{M}^1(X)$, on note $T * \nu(f) = \int f \circ T \,\mathrm{d}\nu$. Posons $\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} T^k * \nu$.

Toute valeur d'adhérence de la suite $(\mu_n)_n$ est T-invariante. Par compacité, il existe au moins une telle valeur d'adhérence.

Démonstration. Soit μ une valeur d'adhérence, limite préfaible de $\left(\mu_{\theta(n)}\right)$. Dans ce cas, par continuité de T, pour toute fonction f continue, on a :

$$T * \mu(f) = \lim_{n\to\infty} \mu_{\theta(n)}(f \circ T) = \lim_{n\to\infty} \left(\mu_{\theta(n)}(f) + \nu\left(\frac{f \circ T^n - f}{n} \right) \right) = \mu(f)\,,$$

car $\|f\|_\infty < \infty$. □

Remarque 99 (Contre-exemple dans le cas X non compact) **:**

Soit $X =]0,1]$ muni de la transformation $T(x) = \frac{x}{2}$ continue. Indépendamment de la mesure ν initiale, on a $\mu_n\left(\left]\frac{1}{2^{n-1}}, 1\right]\right) = 0$. Pour toute limite préfaible, on en déduit que cet ensemble est de masse nulle, et donc que $\mu(X) = 0$, ce n'est donc pas une probabilité.

Remarque 100 (Contre-exemple dans le cas T discontinue) **:**

Soit $T(x) = \frac{x}{2}$ si $x > 0$, et $T(0) = 1$, sur $X = [0,1]$. Pour toute mesure initiale ν, on a directement convergence préfaible des moyennes vers δ_0. Cependant, $T * \delta_0 = \delta_1$. La mesure limite n'est donc pas T-invariante.

Remarque 101 (Système dynamique stochastique) **:**

Soit Y compact. On considère des transformations continues $(\varphi_i)_{1 \le i \le d}$ sur Y. On considère une loi ρ sur $[\![1,d]\!]$, et $\Phi_n \overset{\text{iid}}{\sim} \varphi_\rho$.

On en déduit une chaîne de Markov $(Y_n)_n$, tel que $Y_0 = y$ et $Y_{n+1} = \Phi_{n+1}(Y_n)$.

Considérons l'espace $\{\varphi_1, \dots, \varphi_d\}^{\mathbb{N}}$ muni de la mesure $\rho^{\otimes \mathbb{N}}$, et du shift S. On peut définir le produit croisé $X = \{\varphi_1, \dots, \varphi_d\}^{\mathbb{N}} \times Y$, via $T(\Phi, y) = (S(\Phi), \Phi_0(y))$. On vérifie que T est continue, et X est compact.

En partant de $\rho^{\otimes \mathbb{N}} \otimes \delta_y$, quitte à extraire, on obtient une mesure T-invariante μ. En projetant sur la première coordonnée, pour tout rang fini (et donc à la limite) on a la mesure S-invariante

$\rho^{\otimes\mathbb{N}}$. En projetant sur la seconde coordonnée, on obtient une mesure sur Y invariante pour la chaîne de Markov décrite précédemment.

8.3 Décomposition ergodique

Théorème 102 :

Soit (X, T, μ) un PPT métrique compact. Alors il existe X_0 de mesure pleine tel que :

1. pour tout $x \in X_0$, $\frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}$ admet une limite préfaible μ_x,
2. pour toute fonction $f \in L^1(X, \mu)$, il existe $X_f \subset X_0$ de mesure pleine tel que pour tout $x \in X_f$, $\frac{1}{n}\sum_{k=0}^{n-1} f\big(T^k(x)\big) \to \mu_x(f)$, et $\int \mu_x(f)\,\mathrm{d}\mu(x) = \mu(f)$,
3. Pour tout $x \in X_0$, μ_x est T-invariante, ergodique.

Démonstration. Par séparabilité, considérons une famille dénombrable dense (f_j) dans l'espace $(\mathcal{C}^0(X), \|\cdot\|_\infty)$.

Pour chaque telle fonction, on a un ensemble de masse pleine dans lequel $\frac{1}{n}S_n f_j(x) \to f_j^*(x)$ pour tout x. Notons X_1 l'ensemble des points pour lesquels on a la convergence précédente pour tout j, de mesure pleine.

Par densité de la famille (f_j), pour tout $x \in X_1$ et toute fonction $f \in L^1(X)$, on a convergence de $\frac{1}{n}S_n f(x)$ vers une limite $f^*(x)$. L'application $f \mapsto f^*(x)$ est alors une forme linéaire positive, et l'image de $\mathbb{1}$ est 1, donc par le théorème de représentation de Riesz, on a bien une mesure de probabilité μ_x associée, telle que $f^*(x) = \mu_x(f)$ pour toute fonction continue. On satisfait ainsi le premier point.

Soit (φ_n) une suite dans $L^1(X, \mu)$, qui converge vers f de façon croissante. Montrons qu'on a la convergence p.s. croissante de (φ_n^*) vers f^*. Quitte à considérer $g_n = f - f_n$, on se ramène à une suite positive décroissante, qui converge vers 0 dans L^1. On a donc décroissance de la suite (g_n^*) sur X_1. Par positivité, il existe une limite p.s. $\overline{g}$, et par le lemme de Fatou, $\mu(\overline{g}) \leq \lim \mu(g_n^*) = 0$, d'où $\overline{g} = 0$. On a donc la convergence croissante p.s. (et L^1) souhaitée.

Soit $M = \{A \in \mathcal{A},\ \mathbb{1}_A^*(x) = \mu_x(A) \text{ p.s.}\}$. Si U est ouvert, alors $U \in M$. En effet, on peut approcher $\mathbb{1}_U$ par en dessous de façon croissante, par des fonctions continues. On a les égalités p.s. $\mathbb{1}_U^* = \lim \varphi_n^* = \lim \mu_x(f_n) = \mu_x(U)$, en utilisant ce qui précède. On vérifie que M est stable par complémentaire, et par union dénombrable disjointe, c'est donc une sous-tribu de $\mathcal{A}$ qui contient une partie génératrice, $\mathcal{M} = A$.

Avec un raisonnement analogue, on étend le résultat à toute fonction $f \in L^1$, de sorte que $f^*(x) = \mu_x(f)$ presque-sûrement, sur un ensemble X_f. Par le théorème de Birkhoff, on a donc $\mu(f) = \int f^*(x)\,\mathrm{d}\mu(x) = \int \mu_x(f)\,\mathrm{d}\mu(x)$, ce qui conclut le second point.

Pour le troisième point, considérons $X_2 = \bigcap_{j\in\mathbb{N}} X_{f_j\circ T} \subset X_1$, de mesure pleine. On a alors

$\mu_x(f_j \circ T) = \lim \frac{1}{n} S_n(f_j \circ T)(x) = \lim \frac{1}{n} S_n f_j(x) = \mu_x(f_j)$ puisque $x \in X_1$. Par densité, le résultat s'étend à toute fonction f continue, et donc les mesures sont T-invariantes.

Ainsi, pour $x \in X_2$, on a $f_j^*(T(x)) = f_j^*(x) = \mu_x(f_j)$. Pour $j \in \mathbb{N}$ et $q \in \mathbb{Q}$, posons alors $A_{j,q} = \left\{ y \in X_1,\ f_j^*(y) \leq q \right\}$. Ces ensembles de niveau sont T-invariants. On a donc l'égalité $\mathbb{1}_{A_{j,q}}(x) = \mathbb{1}^*_{A_{j,q}}(x) = \mu_x(A_{j,q})$ sur $X_2 \cap X_{\mathbb{1}_{A_{j,q}}}$, μ-presque-sûrement.

Considérons finalement $X_0 \subset X_2$ l'ensemble des points pour lesquels on a les égalités ci-avant pour tous j et q. Comme $\mu(X_2) = 1 = \int \mu_x(X_2)\,\mathrm{d}\mu(x) \leq 1$, on peut restreindre X_0 de sorte que, pour tout $x \in X_0$, on a $\mu_x(X_2) = 1$. Cet ensemble est encore de mesure pleine, et inclus dans X_1 donc le premier point y est vérifié à nouveau. Quitte à restreindre les $X_f \subset X_1$ précédemment obtenus à X_0, on vérifie encore le seconde point.

Fixons par la suite un élément $x \in X_0$.

Pour tout $j \in \mathbb{N}$, les ensembles de niveau $\left(A_{j,q}\right)_q$ sont de mesure 0 ou 1 sous la loi μ_x, donc f_j^* est μ_x-p.s. constante. Notons que, pour tout $y \in X_2$ on a $f_j^*(y) = \lim \frac{1}{n} S_n f_j(y)$. Comme $\mu_x(X_2) = 1$, f_j^* est en fait la limite obtenue en appliquant le théorème ergodique de Birkhoff au PPT (X, μ_x), donc $\mu_x\left(f_j^*\right) = \mu_x(f_j)$. Comme f_j^* est elle-même constante, on en déduit que $\mu_y(f_j) = f_j^*(y) = \mu_x(f_j)$ pour μ_x-presque-tout y.

Plus largement, on vérifie que si $f \in \mathcal{C}^0(X, \mathbb{R})$, alors $\left\| f^* - f_j^* \right\|_\infty \leq \left\| f - f_j \right\|_\infty$ dans l'espace $\left\{ f^*,\ f \in \mathcal{C}^0 \right\} \subset L^\infty(X, \mu_x)$. Comme les fonctions f_j^* sont toutes μ_x-p.s. constantes, on en déduit plus largement que f^* est elle-même μ_x-p.s. constante, égale à $\mu_x(f)$ sur cet espace.

Pour un ouvert U, quitte à approcher inférieurement $\mathbb{1}_U$ par des fonctions (φ_n) continues, on vérifie alors que $\mathbb{1}_U^*$ est μ_x-p.s. constante, égale à $\mu_x(U)$. En outre, $\mathbb{1}_U^*$ est limite des $\frac{1}{n} S_n \mathbb{1}_U$ dans $L^1(X, \mu_x)$. Pour deux ouverts U et V, on a donc :

$$\frac{1}{n} \sum_{k=0}^{n-1} \mu_x\left(T^{-k}(U) \cap V\right) - \mu_x(U)\mu_x(V) = \mu_x\left(\left(\frac{1}{n} S_n \mathbb{1}_U - \mathbb{1}_U^* \right) \times \mathbb{1}_V \right) \underset{n \to \infty}{\longrightarrow} 0\,,$$

par convergence dominée. Par une des propriétés du cours, on en déduit que μ_x est ergodique. $\square$

9 Théorèmes ergodiques sous-additif, multiplicatif

9.1 Cocycles dans un groupe

9.1.1 Marche aléatoire sur un groupe discret finiment engendré

Remarque 103 :
On peut par exemple considérer le groupe $\mathbb{Z}^d$, engendré par l'ensemble $S = \{\pm e_i,\ 1 \leq i \leq d\}$.

On peut également considérer un groupe libre $[a, b]$, engendré par $\{a, b, a^{-1}, b^{-1}\}$.

Définition 104 (Marche aléatoire droite) **:**
Soit $(G, \times)$ un groupe discret, engendré par S fini. Soit ρ une loi sur S. On considère alors le PPT $\left(S^{\mathbb{N}}, \rho^{\otimes \mathbb{N}}, T\right)$ muni du shift. Autrement dit, on a une suite de variables (g_n) iid.

Une marche aléatoire droite est alors donnée par la suite de variables (r_n), qui vérifie la relation $r_{n+1} = r_n g_{n+1}$. Autrement dit, si $g \in S^{\mathbb{N}}$, et $f(g) = g_0$, alors $r_{n+1} = r_n \times f \circ T^{n+1}(g)$.

9.1.2 Cocycle différentiel

Remarque 105 :
Soient $V \subset \mathbb{R}^d$, et $T : V \to V$ un difféomorphisme qui préserve la mesure μ.

Dans ce cas, on a $d(T^n) = dT \circ T^{n-1} \times \cdots \times dT \circ T \times dT$.

Définition 106 :
Soit (X, μ, T) un PPT et $f : X \to (G, \times)$.

Le cocycle droit est alors $f_n(x) = f(x) \times \cdots \times f \circ T^{n-1}(x)$.

Remarque 107 :
Si $G = (\mathbb{R}, +)$ et que le PPT est ergodique, on sait que $f_n = S_n f$, donc que $\frac{f_n}{n}$ converge vers $\mu(f)$.

On aimerait arriver à étendre ces résultats à un cadre plus général, pour étudier le comportement de la suite (f_n).

9.2 Théorème ergodique sous-additif de Kingman

Remarque 108 :
On suppose que $(G, \times)$ est muni d'une norme telle que $\|gh\| \leq \|g\| \times \|h\|$.

Si G est engendré par S fini, on peut par exemple considérer $\|g\| = e^{d_S(g,1_G)}$, où d_S est la distance d'édition entre deux éléments de G dans le système de générateurs S.

Si $G \subset GL_d(\mathbb{R})$, la norme opérateur convient.

Dans ce cas, on a $\ln(\|f_{n+m}(x)\|) \leq \ln(\|f_n(x)\|) + \ln(\|f_m \circ T^n(x)\|)$.

Théorème 109 (Théorème ergodique sous-additif de Kingman) **:**
Soit (X, μ, T) un PPT.

On considère une suite sous-additive (a_n), autrement dit $a_{n+m}(x) \leq a_n(x) + a_m \circ T^n(x)$ pour tout $x \in X$.

Si $a_1 \in L^1(X, \mu)$, alors $\frac{a_n}{n}$ converge presque-sûrement vers une limite T-invariante a^*.

Démonstration. Soit $b_n = \sum_{j=0}^{n-1} a_1 \circ T^j - a_n = S_n a_1 - a_n$. On vérifie facilement que $b_n \geq 0$, et que b_n est sur-additive. Si on prouve le résultat voulu sur la suite (b_n), alors en appliquant le théorème ergodique de Birkhoff sur a_1, on aura le résultat souhaité. Ce point-ci est une application directe du lemme ci-après. □

Lemme 110 :
Si (a_n) est une suite positive sur-additive, alors $\frac{a_n}{n}$ converge presque-sûrement.

Démonstration. On ira ici assez vite car la preuve est très similaire à celle de Birkhoff.

Soit $\overline{a} = \overline{\lim} \frac{a_n}{n} \geq 0$. Pour $R > 0$ et $\varepsilon > 0$, fixons $\overline{g} = ((1-\varepsilon)\overline{a}) \wedge R$.

Soit $\overline{n}(x) = \inf\{k \in \mathbb{N},\, a_k(x) \geq k\overline{g}(x)\}$. Naturellement, $\overline{n}(x) < \infty$ par construction. Notons $E_L = \{\overline{n} \leq L\}$. On pose alors $\tau = \overline{h}$ sur E_L, et 1 en dehors.

Remarquons que $\overline{a}$ est T-invariante. En effet, $\frac{a_n(T(x))}{n} \leq \frac{a_{n+1}(x)}{n+1} \times \frac{n+1}{n} - \frac{a_1(x)}{n}$, et donc à la limite $\overline{a} \circ T \leq \overline{a}$. Par croissance, on a donc $\overline{g} \circ T \leq \overline{g}$. Comme $\overline{g}$ est bornée donc intégrable, on vérifie que $\int \overline{g} - \overline{g} \circ T = 0$, donc la fonction est T-invariante. Par convergence monotone, on a alors la T-invariance de $\overline{a}$ aussi.

Comme dans le cas de la preuve de Birkhoff, on montre alors que $a_\tau \geq S_\tau(\overline{g}\mathbb{1}_E)$. Plus largement, avec $\tau_{k+1} = \tau_k + \tau \circ T^{\tau_k}$, on vérifie que $a_{\tau_k} \geq \overline{g} \times S_{\tau_k}(\mathbb{1}_E)$.

Notons que $\tau_k \to \infty$ presque-sûrement, et que $|\tau_{k+1} - \tau_k| \leq L$. On en dédyut donc plus largement que, pour tout rang N, on a $a_{N+L} \geq \overline{g} S_N(\mathbb{1}_E)$. Finalement, pour $N \to \infty$, on a $\underline{a} \geq \overline{g} \times \mathbb{1}_E^*$.

Pour $L \to \infty$, par convergence monotone, $\mathbb{1}_{E_L} to 1$, et donc $\mathbb{1}_{E_L^*}$ aussi. Ainsi, $\underline{a} \geq \overline{g}$.

Enfin, avec $\varepsilon \to 0$, et $R \to \infty$, on a la convergence de $\overline{g}$ vers $\overline{a}$, d'où enfin $\underline{a} \geq \overline{a}$, donc les deux limites sont *égales* à une fonction a^*, T-invariante. □

Corollaire 111 :

Soit $A : X \to GL_d(\mathbb{R})$, avec (X, T, μ) un PPT. On pose $A_n = A \circ T^{n-1} \times \cdots \times A$.

Alors il existe un exposant de Lyapunov T-invariant χ, tel que $\frac{1}{n} \ln(\|A\|) \xrightarrow{\text{p.s.}} \chi$.

9.3 Théorème multiplicatif

Remarque 112 :

On utilise ici la décomposition polaire de la matrice $A_n(x)$.

Autrement dit, $A_n = W_n \mathrm{Diag}\big(\lambda_n^1, \ldots, \lambda_n^d\big) V_n$, avec $W_n, V_n \in O_d(\mathbb{R})$, et $\lambda_n^1 \geq \ldots \lambda_n^d$.

Dans ce cas, on pose $A_n^T A_n = V_n^T D_n^2 V_n$. On pose alors $\Lambda_n := V_n^T \sqrt[n]{D_n} V_n = \sqrt[2n]{A_n^T A_n}$. Avec cette matrice, pour tout vecteur $v \in \mathbb{R}^d$, on a $\|\Lambda_n^n v\| = \|A_n v\|$.

Théorème 113 (Théorème ergodique multiplicatif d'Osseledets) **:**

Si $\ln(\|A_n^{\pm 1}\|) \in L^1(X, \mu)$ pour tout n, alors on a $\Lambda_n \to \Lambda$ presque-sûrement.

De plus, $\frac{1}{n} \ln(\|A_n \Lambda^{-n}\|) \to 0$, et $\frac{1}{n} \ln(\|\Lambda^n A_n^{-1}\|) \to 0$.

Enfin, on a $\Lambda(x) = V^T(x) \mathrm{Diag}(\lambda_1(x), \ldots, \lambda_d(x)) V(x)$, avec $\lambda_j = \lim (\lambda_n^j)^{\frac{1}{n}}$ est T-invariante.

Démonstration. Admis. □

Remarque 114 :

On a $\|A_n v\| \leq \|A_n \Lambda^{-n}\| \|\Lambda^n v\|$, et $\|A_n \Lambda^{-n}\| \to 1$. On a de même l'autre inégalité, en majorant $\|\Lambda^n v\| \leq \|\Lambda^n A_n^{-1}\| \|A_n v\|$, avec $\|\Lambda^n A_n^{-1}\| \to 1$.

En conséquence, $\|A_n v\|$ et $\|\Lambda^n v\|$ ont le même comportement asymptotique.

De façon quantitative, $\frac{1}{n} \ln(\|A_n v\|)$ et $\frac{1}{n} \ln(\|\Lambda^n v\|)$ ont la même limite lorsqu'elle existe.

En particulier, en $v_j = V^{-1} e_j$, on a $\lim \frac{1}{n} \ln(\|A_n v_j\|) = \ln(\lambda_j) = \chi_j$.

Corollaire 115 :

Dans le cas où le PPT est ergodique, il existe $\overline{\chi_1} < \cdots < \overline{\chi_k}$, et un drapeau $(E_k(x))_{0 \leq j \leq k}$ avec $k \leq d$, qui satisfont les propriétés suivantes. On a $\lim \frac{1}{n} \ln(\|A_n(x) v\|) = \overline{\chi_j}(x)$ si et seulement si $v \in E_j(x) \backslash E_{j-1}(x)$. De plus, $E_j(T(x)) = A(x) E_j(x)$.