Autour des théorèmes ergodiques

1. Quelques définitions et rappels

On considère un espace de probabilité $(X, \mathcal{F}, \mu)$ et une application mesurable $T : X \to X$ qui préserve la mesure, autrement dit pour tout $A \in \mathcal{F}$ on a $\mu(T^{-1}A) = \mu(\{x \in X : Tx \in A\}) = \mu(A)$. On désigne par T^n la $n^{\text{ième}}$ itérée de la transformation T, i.e. $T^0 = \mathrm{id}_X$ et $T^{n+1}(x) = T(T^n(x))$ pour $n \geq 0$.

On définit la *tribu invariante* associée à l'application T comme la sous-tribu de la tribu ambiante $\mathcal{F}$ définie par $\mathcal{I}_T := \{A \in \mathcal{F},\ T^{-1}A = A\}$. Notons qu'une fonction $\mathcal{F}-$mesurable $\varphi : X \to \mathbb{R}$ est une fonction $\mathcal{I}_T-$mesurable si et seulement elle est $T-$invariante [1], i.e. $\varphi \circ T = \varphi$.

Étant donnée une fonction intégrable $f : X \to \mathbb{R}$, on peut alors considérer son espérance conditionnelle $\mathbb{E}[f|\mathcal{I}_T]$, qui est par définition l'unique (aux négligeables près) fonction $\mathcal{I}_T-$mesurable telle que pour tout fonction $\mathcal{I}_T-$mesurable bornée $h : X \to \mathbb{R}$, on a

$$\int_X f(x)h(x)\mu(dx) = \int \left(\mathbb{E}[f|\mathcal{I}_T](x)\right) h(x)\mu(dx).$$

On dit que la transformation T est *ergodique* lorsque la tribu $\mathcal{I}_T$ est triviale, autrement dit lorsque tout $A \in \mathcal{I}_T$ vérifie l'alternative $\mu(A) = 0$ ou $\mu(A) = 1$, ce qui revient à dire que les seules fonctions $f \in L^1(X, \mu)$ telles que $f \circ T = f$ sont les fonctions constantes. Ainsi, lorsque T est ergodique, les fonctions $\mathcal{I}_T-$mesurables sont constantes, de sorte que si f est intégrable

$$\mathbb{E}[f|\mathcal{I}_T] = \int_X f(x)\mu(dx).$$

2. Théorèmes ergodiques

On cherche à comprendre les propriété des orbites $\{T^n(x), n \geq 1\}$, du point de vue géométrique mais aussi de point de vue de la théorie de la mesure. Le théorème fondamental suivant a été établi par Birkhoff en 1931. Il existe de nombreuses variantes de la démonstration de ce résultat, voir par exemple [Kre85, HK95, EW11].

Théorème ergodique de Birkhoff. *Soient $(X, \mathcal{F}, \mu)$ un espace de probabilité et $T : X \to X$ une application qui préserve la mesure μ. Alors si $f \in L^1(X, \mu)$, pour $\mu-$presque tout $x \in X$ on a*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \mathbb{E}[f|\mathcal{I}_T](x).$$

En particulier, si T est ergodique, pour $\mu-$presque tout $x \in X$ on a

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

Démonstration. On suit la preuve de [HK95] (Théorème 4.1.2 p.133). À toute fonction $g \in L^1(X, \mu)$, on associe la suite croissante de fonctions $G_n : X \to \mathbb{R}$

$$G_n(x) := \max_{1\leq k\leq n} \sum_{k=0}^{k-1} g(T^\ell x) = \max_{1\leq k\leq n} \left\{ g(x), g(x) + g(T(x)), \ldots, \sum_{k=0}^{n-1} g(T^\ell x) \right\}.$$

On considère également l'ensemble $A \in \mathcal{I}_T$ défini par $A := \{x \in X,\ \sup_{n\geq 1} G_n(x) = +\infty\}$. Si $x \in A$, alors la nouvelle suite

$$G_{n+1}(x) - G_n(T(x)) = g(x) - \min\{0, G_n(T(x))\} \geq g(x)$$

1. En effet, si la fonction φ est $\mathcal{I}_T-$mesurable, pour tout $y \in \mathrm{Im}(\varphi)$ on a $\varphi^{-1}(\{y\}) = \{x \in X,\ \varphi(x) = y\} \in \mathcal{I}_T$. Ainsi, par définition de $\mathcal{I}_T$, si $x \in T^{-1}\left(\varphi^{-1}(\{y\})\right) = \varphi^{-1}(\{y\})$ alors $\varphi(T(x)) = y = \varphi(x)$ et φ est donc $T-$invariante. Réciproquement, si on suppose que la fonction mesurable φ est $T-$invariante, pour tout borélien $B \in \mathcal{B}(\mathbb{R})$, on a $\varphi^{-1}(B) = \{x \in X, \varphi(x) \in B\} = \{x \in X, \varphi(T(x)) \in B\} = \{x \in X, T(x) \in \varphi^{-1}(B)\} = T^{-1}(\varphi^{-1}(B))$.

est décroissante et converge vers $g(x)$ lorsque n tend vers l'infini. Par le théorème de convergence dominée, comme μ est $T-$invariante

$$0 \leq \int_A (G_{n+1} - G_n)\, d\mu = \int_A (G_{n+1} - G_n \circ T)\, d\mu \xrightarrow[n\to+\infty]{} \int_A g d\mu. \tag{1}$$

Lorsque $x \notin A$, la suite croissante $G_n(x)$ est majorée de sorte que

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k x) \leq \limsup_{n\to+\infty} \frac{G_n(x)}{n} = 0. \tag{2}$$

Supposons que l'espérance conditionnelle de g sachant $\mathcal{I}_T$ vérifie l'inégalité $\mathbb{E}[g|\mathcal{I}_T] < 0$. Alors, pour tout ensemble C mesurable par rapport à $\mathcal{I}_T$ et tel que $\mu(C) > 0$, on a

$$\int_C g d\mu = \int_X \mathbb{1}_C g d\mu \overset{\text{déf}}{=} \int_X \mathbb{1}_C \mathbb{E}[g|\mathcal{I}_T] d\mu = \int_C \mathbb{E}[g|\mathcal{I}_T] d\mu < 0.$$

Ainsi par l'équation (1), on a nécessairement $\mu(A) = 0$, et par (2) on déduit que pour μ presque tout $x \in X$

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k x) \leq 0. \tag{3}$$

Pour $f \in L^1(X, \mu)$ et $\varepsilon \in \mathbb{R}_+^* \cap \mathbb{Q}$ et on pose alors $g_\varepsilon = f - \mathbb{E}[f|\mathcal{I}_T] - \varepsilon$. Comme $\mathbb{E}[f|\mathcal{I}_T]$ est une fonction $\mathcal{I}_T-$mesurable, elle est T invariante, i.e. $\mathbb{E}[f|\mathcal{I}_T] \circ T = \mathbb{E}[f|\mathcal{I}_T]$. Aussi

$$\frac{1}{n} \sum_{k=0}^{n-1} g_\varepsilon(T^k x) = \frac{1}{n} \sum_{k=0}^{n-1} (f - \mathbb{E}[f|\mathcal{I}_T] - \varepsilon)\,(T^k(x)) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) - \mathbb{E}[f|\mathcal{I}_T](x) - \varepsilon. \tag{4}$$

Par ailleurs, on a $\mathbb{E}[g_\varepsilon|\mathcal{I}_T] = -\varepsilon < 0$. Aussi, en appliquant l'inégalité (3) à g_ε, on obtient que pour μ presque tout $x \in X$ (i.e. sur un ensemble de mesure pleine qui dépend à priori de ε)

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} g_\varepsilon(T^k x) \leq 0,$$

et par (4), on déduit donc que pour μ presque tout $x \in X$ (i.e. sur le même ensemble de mesure pleine qui dépend à priori de ε)

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \leq \mathbb{E}[f|\mathcal{I}_T](x) + \varepsilon.$$

En faisant tendre ε vers zéro (intersection dénombrable d'ensembles de mesure pleine), on conclut que pour μ presque tout $x \in X$

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \leq \mathbb{E}[f|\mathcal{I}_T](x).$$

En remplaçant f par $-f$ dans le raisonnement ci-dessus, on obtient de la même façon que pour μ presque tout $x \in X$

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \geq \mathbb{E}[f|\mathcal{I}_T](x).$$

□

Pour des histoires de calendrier serré et de rivalités scientifiques, la publication du théorème de Birkhoff en 1931 précède celle du théorème de von Neumann en 1932, qui lui est pourtant antérieur (von Neumann avait communiqué sa preuve à Birkhoff avec que ce dernier n'obtienne son résultat).

Théorème ergodique de von Neumann. *Soient $(X, \mathcal{F}, \mu)$ un espace de probabilité et $T : X \to X$ une application qui préserve la mesure μ. Alors si $f \in L^2(X, \mu)$, on a la convergence dans $L^2(X, \mu)$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \cdot) = \mathbb{E}[f|\mathcal{I}_T](\cdot).$$

La preuve dans le cas L^2 est de nature hilbertienne et le théorème ci-dessus est en fait un cas particulier du théorème suivant, également appelé théorème ergodique de von Neumann. On peut trouver sa preuve dans de très nombreux textes, par exemple [BMP04] (exercice 3.6 p. 137) ou [EW11]. La preuve ci-après est due à Riez.

Théorème ergodique de von Neumann bis. *Soit $\mathcal{H}$ un espace de Hilbert et $U : \mathcal{H} \to \mathcal{H}$ une contraction faible, c'est à dire un opérateur linéaire continu de norme inférieure ou égale à 1, autrement dit pour tout $x \in \mathcal{H}, \|Ux\| \leqslant \|x\|$, alors*

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k \underset{n\to\infty}{\longrightarrow} \Pi,$$

où Π est la projection orthogonale sur $\mathrm{Ker}(U-\mathrm{id})$.

Avant de donner la preuve du dernier théorème, voyons comment il implique la version en haut de page. On choisit pour $\mathcal{H}$ l'espace $L^2(X,\mu)$, avec le produit scalaire usuel $\langle \varphi, \psi\rangle_{\mathcal{H}} = \int \varphi\psi d\mu$, i.e. $||\varphi||^2 = \langle \varphi, \varphi\rangle_{\mathcal{H}}$. L'opérateur U n'est autre que l'opérateur de composition $U(\varphi) = \varphi \circ T$ sur $L^2(X,\mu)$, qui est bien linéaire continu avec $||U(\varphi)|| = ||\varphi||$ car μ est supposée $T-$invariante. Le noyau $\mathrm{Ker}(U-\mathrm{id})$ est composé des fonctions $\varphi \in L^2(X,\mu)$ telles que $\varphi\circ T = \varphi$, autrement dit des fonctions $T-$invariantes, i.e. les fonctions $\mathcal{I}_T-$mesurables. Dans ce cadre L^2, la projection orthogonale Π sur $\mathrm{Ker}(U-\mathrm{id})$ est donc l'espérance conditionnelle $\mathbb{E}[\varphi|\mathcal{I}_T]$, d'où le résultat.

Pour clarifier la preuve du théorème bis, isolons le lemme suivant.

Lemme 1. *Soit $\mathcal{H}$ un espace de Hilbert.*

1. *Soit F est un sous-espace vectoriel de H, dont l'orthogonal est noté $F^\perp$. Alors $F^\perp$ est fermé, $F^\perp = \bar{F}^\perp$ et $\bar{F}\oplus F^\perp = \mathcal{H}$.*
2. *Si U est une application linéaire continue, alors $(\mathrm{Im}(U))^\perp = \mathrm{Ker}\,(U^*)$.*
3. *Si U est une contraction faible, alors U^* également et $\mathrm{Ker}(U-\mathrm{id}) = \mathrm{Ker}\,(U^*-\mathrm{id})$.*

Preuve du lemme. Les deux premiers points sont classiques. Pour le troisième item, pour $x, y \in \mathcal{H}$ on a

$$|\langle U^*x, y\rangle| = |\langle x, Uy\rangle| \leqslant \|x\|\|Uy\| \leqslant \|x\|\|y\|.$$

En maximisant sur y puis sur x on en déduit que $\|U^*\| \leqslant 1$. Ensuite, si $U^*x = x$ on a

$$\|x\|^2 = \langle x, x\rangle = \langle U^*x, x\rangle = \langle x, Ux\rangle \leqslant \|x\|^2,$$

ainsi on est dans le cas d'égalité de l'inégalité de Cauchy–Schwarz, et on déduit que $Ux = x$. On a donc $\mathrm{Ker}\,(U^*-\mathrm{id}) \subset \mathrm{Ker}(U-\mathrm{id})$, et comme $U^{**} = U$, l'inclusion réciproque a également lieu. □

Preuve du théorème bis. On donne maintenant la preuve du théorème de von Neumann dans sa version hilbertienne. D'après le lemme ci-dessus, on a la décomposition de $\mathcal{H}$ suivante.

$$\begin{aligned}\mathcal{H} &= \overline{\mathrm{Im}(\mathrm{id}-U)} \oplus (\mathrm{Im}(\mathrm{id}-U))^\perp\\ &= \overline{\mathrm{Im}(\mathrm{id}-U)} \oplus \mathrm{Ker}\,(\mathrm{id}-U^*)\\ &= \overline{\mathrm{Im}(\mathrm{id}-U)} \oplus \mathrm{Ker}(\mathrm{id}-U).\end{aligned}$$

Pour $x \in \mathrm{Ker}(\mathrm{id}-U)$, on a naturellement

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k x = x \underset{n\to\infty}{\longrightarrow} \Pi(x).$$

Par ailleurs, si $x \in \mathrm{Im}(\mathrm{id}-U)$, en écrivant $x = y - Uy$ il vient

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k x = \frac{1}{n}\sum_{k=0}^{n-1} U^k(y-Uy) = \frac{1}{n}\left(y - U^n y\right) \underset{n\to\infty}{\longrightarrow} 0,$$

car $\|y - U^n y\| \leq \|y\| + \|U^n y\| \leq 2\|y\|$. Pour conclure, il reste à voir que le résultat est vrai si on passe à l'adhérence. Soit donc $x \in \overline{\mathrm{Im}(\mathrm{id}-U)}$ avec $x = \lim_{q\to+\infty} x_q$ et $x_q \in \mathrm{Im}(\mathrm{id}-U)$. On remarque que $S_n := \frac{1}{n}\sum_{k=0}^{n-1}$ est une contraction faible de sorte que par l'inégalité triangulaire

$$\|S_n(x)\| \leq \|S_n(x_q)\| + \|S_n(x-x_q)\| \leq \|S_n(x_q)\| + \|x - x_q\|.$$

Pour $\varepsilon > 0$ donné, il existe q assez grand de sorte $\|x - x_q\| \leq \varepsilon$. À q fixé, comme $x_q \in \mathrm{Im}(\mathrm{id}-U)$, d'après ci-dessus on a $\|S_n(x_q)\| \leq \varepsilon$ pour n assez grand, ainsi $\limsup_n \|S_n(x)\| \leq \varepsilon$, d'où le résultat.

□

3. Deux applications classiques

Loi forte des grands nombres. La loi forte des grands nombres est un cas particulier du théorème ergodique de Birkhoff. En effet, soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires réelles i.i.d de loi commune μ. Sans perdre en généralité, on peut supposer les variables X_n sont les applications coordonnées sur l'espace de probabilité canonique $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\mathbb{R}^{\mathbb{N}}, \otimes_{\mathbb{N}}\mathcal{B}(\mathbb{R}), \otimes_{\mathbb{N}}\mu\right)$, autrement dit si $\omega = (\omega_1, \omega_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$ alors $X_n(\omega) := \omega_n$. On considère alors l'opérateur de décalage $T((\omega_n)_{n\geq 0}) := (\omega_{n+1})_{n\geq 0}$. Naturellement T préserve la mesure produit $\otimes_{\mathbb{N}}\mu$ et par ailleurs, par la loi du $0-1$ de Kolmogorov, la tribu invariante $\mathcal{I}_T$ est triviale. Autrement dit, T est ergodique. Considérons l'application première coordonnée $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ telle que $f(\omega) = \omega_0$. D'après le théorème ergodique de Birkhoff, pour $(\otimes_{\mathbb{N}}\mu)-$presque tout $\omega \in \mathbb{R}^{\mathbb{N}}$

$$\frac{1}{n}\sum_{k=0}^{n} f(T^k(\omega)) \xrightarrow[n\to+\infty]{} \int_{\Omega} f(\omega)\mathbb{P}(d\omega),$$

c'est-à-dire comme $f(T^k(\omega)) = \omega_k = X_k(\omega)$

$$\frac{1}{n}\sum_{k=0}^{n} X_k(\omega) \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}} \omega_0 d\mu(\omega_0) = \mathbb{E}[X_0].$$

Nombres normaux. Soient $k \geq 2$ un nombre entier, $x \in [0,1[$ et $x = 0.a_1a_2a_3\ldots$ sa décomposition en base k, i.e. $a_i \in \{0,1,\ldots,k-1\}$ et $x = \sum_{j\geq 1} a_j k^{-j}$. On s'intéresse aux occurences de suites prescrites de nombres dans la suite des décimales, autrement dit, étant donnée une suite de nombres $b_1 \ldots b_j$ dans $\{0,1,\ldots,k-1\}$, on cherche à dénombrer les indices n pour lesquels $a_{n+1} = b_1, \ldots, a_{n+j} = b_j$. On dit que x est normal en base k si

$$\lim_{N\to\infty} \frac{1}{N}\#\left\{n \leqslant N : a_{n+1} = b_1, \ldots, a_{n+j} = b_j\right\} = k^{-j},$$

pour tout j et pour tout choix de nombres $b_1, \ldots, b_j$ dans $\{0,1,\ldots,k-1\}$. Si T désigne l'application de multiplication par k dans $\mathbb{R}/\mathbb{Z}$, i.e. $Tx = kx \mod 1$, et si on munit $[0,1[$ de la mesure de Lebesgue λ, on peut montrer que T est ergodique. Par ailleurs $a_{n+1} = b_1, \ldots, a_{n+j} = b_j$ si et seulement si $T^n x$ appartient à l'intervalle

$$I := \frac{b_1}{k} + \cdots + \frac{b_j}{k^j} + \left[0, \frac{1}{k^j}\right[.$$

Par le théorème ergodique de Birkhoff, appliqué à la fonction $f = \mathbb{1}_I$, on obtient que pour Lebesgue presque tout $x \in [0,1[$ (i.e. sur un ensemble de mesure pleine qui dépend de j et de $b_1, \ldots, b_j$)

$$\lim_{N\to\infty} \frac{1}{N}\#\left\{n \leqslant N : a_{n+1} = b_1, \ldots, a_{n+j} = b_j\right\} = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_I\left(T^n x\right) = \lambda(I) = k^{-j}.$$

Puisque les paramètres j et $b_1, \ldots, b_j$ sont au plus dénombrables, on déduit que pour Lebesgue presque tout $x \in [0,1[$, x est normal en base k. Enfin, comme k vit lui-même dans un ensemble au plus dénombrable, on conclut que pour Lebesgue presque tout $x \in [0,1[$, x est normal dans toute base, un tel nombre est dit absolument normal.

Références

[BMP04] Beck, V. and Malick, J. and Peyré, G., Mathématiques Objectif Agrégation, seconde édition, 2004.

[EW11] Einsiedler, M. and Ward, T., Ergodic Theory with a view towards Number Theory, Graduate Texts in Mathematics, Springer, 2011.

[HK95] Hasselblatt, B., Katok, A., Introduction to the Modern Theory of Dynamical Systems. Cambridge University Press, 1995.

[Kre85] Krengel, U., Ergodic Theoems, de Gruyter, 1985.