Vrac sur la théorie ergodique

Introduction : Loi des grands nombres

Dans un cadre de physique statistique, dans la théorie cinétique des gaz, Ludwig Boltzmann introduit la notion de système ergodique. L'adjectif "ergodique" provient de contraction de deux mots grecs anciens : ἔργον (ergon), qu'on interprète comme *travail* et ὁδὸς (odos), qu'on interprète comme *chemin*. L'ambition était alors d'affirmer que l'étude d'une trajectoire permet en fait d'établir un comportement global.

La théorie ergodique est alors une théorie de systèmes dynamiques, et pourrait se limiter à cela. Néanmoins, il y a derrière cette théorie une vision probabiliste qui se dissimule assez souvent, une sorte d'uniformisation de la probabilité de présence qui fait qu'il s'agit aussi d'une sous-partie des Probabilités. Dans cet exposé, j'ai l'ambition de donner une trajectoire probabiliste aux résultats énoncés, en espérant apporter un éclairage différent que ceux provenant de la topologie ou encore de la théorie des représentations de groupe.

En tant que tel, il me semble qu'un point de départ judicieux est la loi forte des grands nombres. Un exposé concis, sur lequel je m'inspire, est [Kle20]. Énonçons alors ce théorème.

Théorème : Loi forte des grands nombres

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_n$ une suite de variables aléatoires L^1 indépendantes et identiquement distribuée, de moyenne m. Alors

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{} m\right) = 1.$$

Remarque 1. C'est exactement ce théorème qui nous donne l'interprétation qu'on donne souvent à l'espérance d'une variable aléatoire : en moyenne, une variable aléatoire X va être proche de cette moyenne. En fait, ce théorème nous même encore plus. Si on se donne une variable aléatoire X de référence, pas forcément intégrable, et on se donne $X_1, \cdots, X_n$ des copies indépendantes de X (c'est-à-dire que chaque X_i suit la même loi que X), et si on se donne un événement $A \in \mathcal{F}$, alors $\mathbb{P}$-presque sûrement, on a en fait :

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_A(X_k) \xrightarrow[n\to\infty]{\mathbb{P}-\text{p.s.}} \mathbb{P}(X \in A).$$

C'est encore la loi des grands nombres qui nous permet de dire que calculer une fraction "nombre de cas possibles" sur "nombre total de cas" est proche de la probabilité que la variable aléatoire d'intérêt vérifie une certaine propriété. ♣

Une question légitime serait alors de se demander à partir de quel entier n on peut affirmer que la moyenne empirique $\frac{1}{n}\sum_{k=1}^{n} X_k$ est proche de sa moyenne stochastique. La réponse est donnée par le Théorème Central Limite.

Théorème Central Limite

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_n$ une suite de variables aléatoires L^2, indépendantes et identiquement distribuées, de moyenne μ et d'écart-type σ. Alors, si $\overline{X}_n \overset{\text{def.}}{=} \frac{1}{n}\sum_{k=1}^{n} X_k$ est la moyenne empirique des X_i, on a :

$$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1).$$

En d'autres termes, on avait déjà $\bar{X}_n = \mu + o(1)$, le Théorème Central Limite le précise dans un sens plus faible, en affirmant qu'il est en fait $O\left(\frac{1}{\sqrt{n}}\right)$.

Le contexte de la théorie ergodique est plus large, et intègre plutôt l'esprit des systèmes dynamiques. Il s'agit d'étudier une transformation T d'un espace mesuré $(X, \mathcal{A}, \mu)$ (et non pas simplement d'un espace probabilisé), et d'étudier le comportement de $T^n(x)$ pour $x \in X$ et n grand en valeur absolue. L'objectif de cette présentation est d'introduire le vocabulaire spécifique à la théorie ergodique, et les principaux résultats. Nous allons notamment évoquer le théorème ergodique de Birkhoff, qui est une variante de la loi forte des grands nombres en théorie ergodique. J'ambitionne aussi d'évoquer d'autres thèmes moins classiques. Le premier est d'évoquer l'absence de vitesse de convergence dans le théorème ergodique de Birkhoff, ce qui est équivalent à affirmer qu'on *a priori* pas de Théorème Central Limite pour ce théorème, en suivant [JR79]. Un autre thème est l'ergodicité du Mouvement Brownien, qui nous permettra de l'introduire, d'étudier rapidement ses propriétés de récurrence, et de conclure sur son ergodicité, suivant notamment [KR53] et [Der54].

I Mesure invariante et récurrence de Poincaré

I.1 Mesure invariante

Définition I.1

Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. Soit $T : X \longrightarrow X$ une application mesurable. On dit que T *préserve* μ, ou que T est μ*-invariante* si pour tout $A \in \mathcal{A}$, $\mu(T^{-1}(A)) = \mu(A)$.

Remarque I.2. Ainsi, montrer qu'une application est invariante par une mesure revient exactement à montrer que deux mesures sont égales. ♣

Dans cet exposé, on s'autorise les notations du type $\{f \in A\}$ pour $f^{-1}(A)$. Ainsi, T préserve μ si et seulement si pour tout $A \in \mathcal{A}$, $\mu(A) = \mu(T \in A)$. En langage probabiliste, on dit que la loi de T est la même que la probabilité ambiante. Voyons quelques exemples de mesure invariante.

1. Soient $n \in \mathbb{N}^*$, $X = {}^{\mathbb{Z}}/_{n\mathbb{Z}}$, $\mathcal{A} = \mathcal{P}\left({}^{\mathbb{Z}}/_{n\mathbb{Z}}\right)$ et pour tout $A \in \mathcal{A}$, $\mu(A) = \sharp A$. Alors l'application $T : x \longmapsto x + 1$ est μ-invariante.

2. Un exemple d'application *mélangeante*, comme on verra un peu plus tard. Considérons l'*application doublante* (*doubling map*) définie sur $[0, 1[$ par

$$\forall x \in [0,1), T(x) \overset{\text{def.}}{=} \begin{cases} 2x & \text{si} \quad 0 \leqslant x < \frac{1}{2} \\ 2x - 1 & \text{si} \quad \frac{1}{2} \leqslant x \leqslant 1 \end{cases}.$$

Alors T est invariante pour la mesure de Lebesgue sur $[0, 1)$. En effet, si on considère $\mu \overset{\text{def.}}{=} \mathrm{Leb} \circ T^{-1}$, μ est alors la mesure image de la mesure de Lebesgue par T, on montre que $\mu = \mathrm{Leb}$ en comparant leurs transformées de Fourier. D'une part, pour tout $\xi \in \mathbb{R}$, on a par lemme de transfert :

$$\hat{\mu}(\xi) = \int_0^1 e^{-\mathrm{i}x\xi} \, \mathrm{d}\mu(x) = \int_0^1 e^{-\mathrm{i}\xi T(x)} \, \mathrm{d}x.$$

Ainsi, *via* une relation de Chasles du type $\int_0^1 = \int_0^{1/2} + \int_{1/2}^1$, on conclut que pour $\xi \neq 0$:

$$\hat{\mu}(\xi) = \frac{1 - e^{-\mathrm{i}\xi}}{\mathrm{i}\xi}.$$

Or, $\widehat{\mathrm{Leb}}(\xi)$ a exactement la même expression sur $\mathbb{R}^*$. De plus, on a aussi $\hat{\mu}(0) = 1 = \widehat{\mathrm{Leb}}(0)$, donc μ coïncide bien avec la mesure de Lebesgue par inversion de Fourier pour les mesures finies.

3. Soit $\mathbb{S}^1 = \{z \in \mathbb{C}, |z| = 1\}$. Considérons $\Theta : [0, 1) \longrightarrow \mathbb{S}^1$ donnée par $\Theta(x) = e^{2\mathrm{i}\pi x}$. Sur $\mathbb{S}^1$, on définit la mesure $\mu \overset{\text{def.}}{=} \mathrm{Leb} \circ \Theta^{-1}$ la mesure image de Θ par la mesure de Lebesgue. Notons que μ est invariante par rotation *id est* si pour $\theta \in [0, 1)$, on définit $R_\theta(x) \overset{\text{def.}}{=} e^{2\mathrm{i}\pi\theta}x$, pour tout $x \in \mathbb{S}^1$, alors $\mu = \mu \circ R_\theta^{-1}$. Pour prouver cela, il suffit de calculer leurs transformées de Fourier respectives :

$$\widehat{\mu \circ R_\theta}(\xi) = \int_{\mathbb{S}^1} e^{\mathrm{i}R_\theta(x)\xi} \, \mathrm{d}\mu(x) = \int_0^1 \exp\left(\mathrm{i}e^{2\mathrm{i}\pi(x+\theta)}\xi\right) \, \mathrm{d}x.$$

On conclut par périodicité. Si pour tout $z \in \mathbb{S}^1$, on définit $T(z) = z^n$, avec $n \in \mathbb{Z}\backslash\{0\}$, alors T est μ-invariante. Pour cela, on utilise le fait que la mesure de Lebesgue est l'unique mesure de probabilité sur $[0, 1)$ invariante par translation, donc que μ est l'unique mesure de probabilité sur $\mathbb{S}^1$ invariante par rotation. Or, si $\nu = \mu \circ T^{-1}$, $e^{2\mathrm{i}\pi\theta} \in \mathbb{S}^1$ et $A \in \mathcal{B}(\mathbb{S}^1)$,

$$T^{-1}\left(e^{2\mathrm{i}\pi\theta}A\right) = R_{\frac{\theta}{n}}^{-1} \circ T^{-1}(A),$$

d'où par invariance $\nu \circ R_\theta(A) = \mu \circ T^{-1}\left(e^{2\mathrm{i}\pi\theta}A\right) = \mu \circ T^{-1}(A) = \nu(A)$. Donc ν est invariante par rotation. Par unicité, $\mu = \nu$, donc T est invariante pour μ.

4. Soit $\mathfrak{A} \overset{\text{def.}}{=} [\![1, N]\!]$, avec $N \geqslant 1$ et définissons $X \overset{\text{def.}}{=} \mathfrak{A}^{\mathbb{Z}}$. L'ensemble X représente l'ensemble des suites indexées par $\mathbb{Z}$ à valeurs dans $[\![1, N]\!]$. On définit alors dans la tribu borélienne associée à la topologie discrète l'ensemble des tous les *cylindres* : on dit que $B \subset X$ est un ensemble cylindrique s'il existe $d \geqslant 1$, des entiers $n_1, \cdots, n_d \in Z$ et $x_1, \cdots, x_d \in \mathfrak{A}$ tels que $B = \{u \in X, \forall 1 \leqslant i \leqslant d, u_{n_i} = x_i\}$. On note Cyl l'ensemble de tous les ensembles cylindriques :

$$\mathrm{Cyl} \overset{\text{def.}}{=} \bigcup_{d \geqslant 1} \bigcup_{n_1, \cdots, n_d \in \mathbb{Z}} \bigcup_{x_1, \cdots, x_d \in \mathfrak{A}} \left(\bigcap_{i=1}^{d} \{\pi_{n_i} = x_i\} \right).$$

On a ici noté $\pi_i : X \longrightarrow \mathfrak{A}$ la projection sur la coordonnée i : $\pi_i(u) = u_i$. La tribu engendrée par les cylindres est alors exactement la tribu borélienne pour la topologie discrète. On définit alors la mesure suivante : si B est un ensemble cylindrique, on définit

$$\mu(B) \overset{\text{def.}}{=} \prod_{i=1}^{d} p_{x_i},$$

où l'on se donne $\{p_1, \cdots, p_N\}$ tels que $p_i \in (0,1)$ et $\sum_{i=1}^{N} p_i = 1$. Notons que $\mu(B)$ ne dépend pas des n_i, mais uniquement des x_i. Une autre manière savante de l'écrire est :

$$\mu \overset{\text{def.}}{=} \bigotimes_{i\in\mathbb{N}} \left(\sum_{x\in\mathfrak{A}} p_x \delta_x \right).$$

Cette mesure μ modélise une suite indépendante et identiquement distribuée de variables aléatoires. Plus précisément, si on se donne sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ une suite de variables aléatoires $(X_n)_{n\in\mathbb{Z}}$ telle que $\mathcal{L}\left((X_n)_{n\in\mathbb{Z}}\right) = \mu$, alors chaque X_i a la même loi $\sum_{x\in\mathfrak{A}} p_x \delta_x$ et sont indépendants, car dans ce cas

$$\mathbb{P}\left(\bigcap_{i=1}^{d} X_{n_i} = x_i \right) = \mu\left(\left\{ u \in X, \bigcap_{i=1}^{d} u_{n_i} = x_i \right\} \right) = \prod_{i=1}^{d} p_{x_i},$$

soit la probabilité de l'intersection est égal au produit des probabilités. Considérons enfin l'application $\sigma : X \longrightarrow X$, dit de *décalage*, donné par

$$\forall u \in X, \sigma\left((u_n)_{n\in\mathbb{Z}}\right) \overset{\text{def.}}{=} (u_{n+1})_{n\in\mathbb{Z}}.$$

Alors, puisque μ ne voit pas les indices n_i mais seulement les valeurs x_i, l'application σ préserve la mesure μ. Le système (X, μ, σ) est appelé *décalage de* BERNOUILLI. On le note parfois $\mathcal{B}(p_1, \cdots, p_N)$.

Remarque I.3. Provenant de [Pet83]. Considérons un processus stochastique discret $(X_n, n \in \mathbb{Z})$ sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, et l'application Φ_X donnée par

$$\Phi_X : \begin{pmatrix} (\Omega, \mathcal{F}) & \longrightarrow & \left(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{Z}}\right)\right) \\ \omega & \longmapsto & (X_n(\omega))_{n\in\mathbb{Z}} \end{pmatrix},$$

on considère alors l'espace mesuré $\left(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{Z}}\right), \mu\right)$, où $\mu = \mathbb{P} \circ \Phi_X^{-1}$. Considérons l'opérateur de décalage : pour toute suite $(u_n)_{n\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$, $T\left((u_n)_{n\in\mathbb{Z}}\right) = (u_{n+1})_{n\in\mathbb{Z}}$. Alors exiger que T est μ-invariante est équivalent à exiger que la suite $(X_n)_n$ soit stationnaire. En effet, puisque $\mathcal{B}\left(\mathbb{R}^{\mathbb{Z}}\right)$ est engendré par les cylindres

$$\text{Cyl} = \bigcup_{d\in\mathbb{N}^*} \bigcup_{n_1,\cdots,n_d\in\mathbb{Z}} \bigcup_{A_1,\cdots,A_d\in\mathcal{B}(\mathbb{R})} \left\{ u \in \mathbb{R}^{\mathbb{Z}}, u_{n_1} \in A_1, \cdots, u_{n_d} \in A_d \right\},$$

alors on a $\mu \circ T^{-1} = \mu$ si et seulement si pour tout $d \in \mathbb{N}^*$, pour tout $A_1, \cdots, A_d \in \mathcal{B}(\mathbb{R})$ et $n_1, \cdots, n_d \in \mathbb{Z}$:

$$\mathbb{P} \circ \Phi_X^{-1}\left(\left\{ u \in \mathbb{R}^{\mathbb{Z}} \;\middle|\; u_{n_1} \in A_1, \cdots, u_{n_d} \in A_d \right\} \right) = \mathbb{P} \circ \Phi_X^{-1} \circ T^{-1}\left(\left\{ u \in \mathbb{R}^{\mathbb{Z}} \;\middle|\; u_{n_1} \in A_1, \cdots, u_{n_d} \in A_d \right\} \right),$$

si et seulement si pour tout $d \in \mathbb{N}^*$, pour tout $A_1, \cdots, A_d \in \mathcal{B}(\mathbb{R})$ et $n_1, \cdots, n_d \in \mathbb{Z}$:

$$\mathbb{P}\left(X_{n_1} \in A_1, \cdots, X_{n_d} \in A_d\right) = \mathbb{P}\left(X_{n_1+1} \in A_1, \cdots, X_{n_d+1} \in A_d\right),$$

si et seulement si $(X_n)_n$ est stationnaire. ♣

Un dernier exemple.

5. Reprenons le contexte de l'exemple précédent : on considère toujours $\mathfrak{A} = [\![1, N]\!]$ et $X = \mathfrak{A}^{\mathbb{Z}}$. On considère toujours σ l'opérateur de décalage. On change cette fois la mesure ambiante. Considérons $\mathbf{p} \overset{\text{def.}}{=} {}^t(p_1, \cdots, p_N)$ comme précédemment et $A \in \mathcal{M}_N(\mathbb{R})$ telle que $A\mathbf{p} = \mathbf{p}$. On impose de plus que la matrice A soit *stochastique* : pour tout $x, y \in \mathfrak{A}$, $[A]_{x,y} \overset{\text{def.}}{=} a_{x,y} \in (0,1)$ avec pour tout $x \in \mathfrak{A}$, $\sum_{y\in\mathfrak{A}} a_{x,y} = 1$ et pour tout $y \in \mathfrak{A}$, $\sum_{x\in\mathfrak{A}} a_{x,y} = 1$. On dit alors que $\mathbf{p}$ est une *probabilité invariante* pour A. Soit $B \subset X$ de la forme, où $i \in \mathbb{Z}$, $d \geqslant 1$ et pour tout $j \in [\![1, d]\!]$, $x_j \in \mathfrak{A}$:

$$B = \left\{ u \in X, \bigcap_{j=1}^{d} u_{i+j} = x_j \right\}.$$

On définit pour ces ensembles B la mesure (de probabilité) μ_A suivante :

$$\mu_A(B) \overset{\text{def.}}{=} p_{x_1} \prod_{j=1}^{d-1} a_{x_j, x_{j+1}}.$$

Alors on appelle (X, μ_A, σ) le *décalage de* MARKOV. En effet, si sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ on se donne une suite $(X_n)_{n\in\mathbb{Z}}$ telle que $\mathcal{L}((X_n)_{n\in\mathbb{Z}}) = \mu_A$ alors $(X_n)_n$ est une chaîne de MARKOV homogène de matrice de transition A, *id est* pour tout $d \geqslant 1$, pour tout $i \in \mathbb{Z}$ et $x_1, \cdots, x_d \in \mathfrak{A}$:

$$\mathbb{P}\left(\bigcap_{j=1}^{d} X_{i+j} = x_j\right) = \mathbb{P}(X_{i+1} = x_{i+1}) \prod_{j=1}^{d-1} a_{x_j, x_{j+1}}.$$

Le terme *homogène* signifie que le terme de droite ne dépend pas des i, mais uniquement des x_j. Cette égalité traduit ce qu'est une chaîne de MARKOV : $\mathcal{L}(X_{n+1}|X_1, \cdots, X_n) = \mathcal{L}(X_{n+1}|X_n)$: *le futur ne dépend du passé que par le présent.* L'opérateur de décalage se retrouve alors être encore invariante pour μ_A, puisque cette dernière ne dépend pas du temps observé mais uniquement des valeurs prises.

I.2 Récurrence

Nous cherchons à définir une notion de *récurrence* au sens de la théorie ergodique. De manière générale, étudier la récurrence d'une application revient à chercher les x tels qu'il existe un entier $n \in \mathbb{N}$ tel que $T^n x = x$, de sorte à ce qu'on ait une propriété du type $T^{n+p}x = T^p x$, et utiliser une division euclidienne pour étudier T. Dans le cadre de la théorie ergodique, et donc de la théorie de la mesure, on peut se confronter au problème de négligeabilité. Ainsi, plutôt que d'étudier une *orbite* ponctuelle, on étudie l'orbite d'un ensemble mesurable.

Définition I.4

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré et $T : X \longrightarrow X$ une application mesurable.

1. On dit que T est *récurrente* pour μ si pour tout $A \in \mathcal{A}$ et pour μ-presque tout $x \in A$, il existe un entier $n = n(x) \geqslant 1$ tel que $T^n(x) \in A$. En d'autres termes,

$$\forall A \in \mathcal{A}, \mu\left(A \Bigg\backslash \left(\bigcup_{n\geqslant 1} \{T^n \in A\}\right)\right) = 0.$$

2. On dit que T est *infiniment récurrente* pour μ si pour tout $A \in \mathcal{A}$, et pour μ-presque tout $x \in A$, il existe une infinité d'entiers $(n_k, k \geqslant 1) = (n_k(x), k \geqslant 1)$ tels que pour tout $k \geqslant 1$, $T^{n_k}(x) \in A$. En d'autres termes,

$$\forall A \in \mathcal{A}, \mu\left(A \Bigg\backslash \left(\bigcap_{N\geqslant 1} \bigcup_{n\geqslant N} \{T^n \in A\}\right)\right) = 0.$$

Le théorème de récurrence de POINCARÉ peut alors s'énoncer directement avec ces définitions.

Théorème I.5 : Théorème de récurrence de POINCARÉ (1890)

Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, avec $\mu(X) < \infty$. Soit $T : X \longrightarrow X$ une application μ-invariante. Alors T est infiniment récurrente.

Démonstration : Soit $A \in \mathcal{A}$. Puisque $\mu(X) < \infty$, il est équivalent de montrer que

$$\mu\left(A \cap \left(\bigcap_{N\geqslant 1} \bigcup_{n\geqslant N} \{T^n \in A\}\right)\right) = \mu(A).$$

L'idée est d'utiliser l'invariance de T pour l'union sur n et de conclure par continuité décroissante de la mesure finie μ. Notons

$$\forall n \in \mathbb{N}, E_n \overset{\text{def.}}{=} \bigcup_{k\geqslant n} \{T^k \in A\}.$$

Le plan de preuve est le suivant.

1. Pour tout $n \in \mathbb{N}$, $\mu(E_n) = \mu(E_0)$ car T est invariante ;
2. Par décroissance des $(E_n)_n$, on en déduit le théorème car μ est finie.

1. Soit $n \in \mathbb{N}$. Par définition,

$$\{T^n \in E_0\} = \left\{ T^n \in \bigcup_{k \geqslant 0} \{T^k \in A\} \right\}.$$

Par propriété des images réciproques, on a

$$\{T^n \in E_0\} = \bigcup_{k \geqslant 0} \left\{ T^n \in \{T^k \in A\} \right\} = \bigcup_{k \geqslant 0} \left\{ T^{n+k} \in A \right\}.$$

Ainsi, par un changement de variables, on conclut que

$$\{T^n \in E_0\} = E_n.$$

Ainsi,

$$\mu(E_n) = \mu(T^n \in E_0) = \mu(E_0),$$

par invariance de T.

2. Ainsi, par continuité décroissante de μ (μ est finie) :

$$\mu\left(\bigcap_{n \geqslant 1} E_n \right) = \lim_{n \to \infty} \mu(E_n) = \mu(E_0).$$

Cela ne s'applique pas *a priori* pour la suite $(A \cap E_n)_n$, mais essayons tout de même de l'écrire, en utilisant l'égalité, vraie pour toute mesure μ (finie ou non) et tout mesurables A, B :

$$\mu(A) = \mu(A \backslash B) + \mu(A \cap B).$$

Remarquons d'abord que $E_0 \cap A = A$, puisque $A \subset E_0$. Ainsi, on a

$$\mu(A) = \mu\left(A \cap E_0 \cap \left(\bigcap_{n \geqslant 1} E_n \right) \right) + \mu\left(A \cap \left(E_0 \Big\backslash \left(\bigcap_{n \geqslant 1} E_n \right) \right) \right).$$

Alors le deuxième terme est nul, car puisque μ est finie :

$$\mu\left(A \cap \left(E_0 \Big\backslash \left(\bigcap_{n \geqslant 1} E_n \right) \right) \right) \leqslant \mu\left(E_0 \Big\backslash \left(\bigcap_{n \geqslant 1} E_n \right) \right) = 0.$$

Quand au premier, en utilisant encore que $A \cap E_0 = A$, on obtient ce qu'on veut montrer. On a bien montré que T était infiniment récurrent pour μ. □

Remarque I.6. Le théorème peut être mis en défaut en mesure infinie. Par exemple, pour $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}), \sharp)$ muni de la mesure de comptage $\sharp$, l'application $T : x \longmapsto x + 1$ est bien invariante pour $\sharp$, mais n'est pas récurrente. En effet, si on regarde l'orbite de $\{0\}$, on a en fait $\{T^n = 0\} = \{-n\}$. Ainsi, pour tout $n \in \mathbb{N}$, $\bigcup_{n \geqslant 1} \{T^n = 0\} = \mathbb{Z}_-^*$ est l'ensemble de tous les entiers négatifs, qui est de mesure $\sharp$ infinie. Or,

$$\{0\} \Big\backslash \left(\bigcup_{n \geqslant 1} \{T^n = 0\} \right) = \{0\} \backslash \mathbb{Z}_-^* = \{0\},$$

qui est de mesure 1. Ainsi, alors que T est invariante, on a tout de même $\sharp\left(\{0\} \Big\backslash \left(\bigcup_{n \geqslant 1} \{T^n = 0\} \right) \right) = 1$, donc T n'est pas récurrente pour la mesure de comptage $\sharp$ sur $\mathbb{Z}$. ♣

II Théorèmes ergodiques

Après avoir étudié la notion de récurrence, attardons-nous sur le cœur du sujet, les théorèmes ergodiques. Nous avons vu en introduction qu'on pouvait les interpréter comme une loi des grands nombres, une convergence de la moyenne empirique vers la moyenne stochastique, ou encore, une convergence de la moyenne dite *temporelle* et de la moyenne *spatiale*. Notez qu'il en existe plusieurs, développés dans la même tranche d'années 1931-1934, notamment par Birkhoff et von Neumann qui ont leurs propres théorèmes ergodiques, mais aussi pensons à Hopf et Koopman, ce dernier ayant son nom à un opérateur présent pour prouver ces théorèmes ergodiques.

Avant de plonger dans les énoncés, remarquons que les théorèmes ergodiques ne feront pas intervenir l'hypothèse d'ergodicité (et pour cause, on l'évoquera qu'en fin de section) !

II.1 Théorème ergodique de von Neumann (*Mean Ergodic Theorem*)

La première preuve de ce théorème a été publié en 1932 dans [Neu32] par von Neumann. On ne suivra pas cette piste-là dans la suite, puisqu'elle utilise la notion de réduction spectrale. La preuve ici est plus élémentaire, dans le sens où elle ne fait appel qu'à des notions d'analyse hilbertienne.

Théorème II.1 : Théorème ergodique de von Neumann (1932)

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et $T : X \longrightarrow X$ une application μ-invariante. Alors, pour tout $f \in L^2(\mu)$, il existe $f^* \in L^2(\mu)$ telle que

$$\lim_{n \to \infty} \int_X \left(\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) - f^*(x) \right)^2 \mathrm{d}\mu(x) = 0.$$

Plus précisément, si $P : L^2(\mu) \longrightarrow L^2(\mu)$ est la projection sur $\{f \in L^2(\mu), f \circ T = f\}$, alors $f^* = Pf$.

Remarque II.2. On nomme ce théorème aussi par *Théorème Ergodique en Moyenne*, car il s'agit bien sûr de constater que ce théorème annonce une convergence quadratique. ♣

Remarque II.3. En termes de loi des grands nombres, considère une suite de variables aléatoires $(X_n, n \in \mathbb{N})$, $\mu = \mathbb{P} \circ \Phi_X^{-1}$ et $T : \mathbb{R}^{\mathbb{Z}} \longrightarrow \mathbb{R}^{\mathbb{Z}}$ l'opérateur de décalage définis comme dans la Remarque I.2. Alors, considérer $f \in L^2(\mu)$ signifie qu'on considère (la classe de) $f : \mathbb{R}^{\mathbb{Z}} \longrightarrow \mathbb{R}$ vérifiant

$$\int_{\mathbb{R}^{\mathbb{Z}}} f(u)^2 \mathrm{d}\mathbb{P} \circ \Phi_X^{-1}(u) = \int_\Omega f\left((X_n(\omega))_{n\in\mathbb{N}}\right)^2 \ \mathrm{d}\mathbb{P}(\omega) = \mathbb{E}\left[f\left((X_n)_{n\in\mathbb{N}}\right)^2\right] < \infty.$$

Autrement dit, on considère une fonctionnelle de suites, et on l'applique à la suite $(X_n)_n$. Alors le théorème ergodique en moyenne annonce que pour tout $f \in L^2(\mu)$, il existe $f^* \in L^2(\mu)$ telle que

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_{k=0}^{N-1} f\left((X_{n+k})_{n\geqslant 0}\right) - f^*\left((X_n)_{n\geqslant 0}\right)\right)^2\right] \xrightarrow[n\to\infty]{} 0.$$

On en déduit la loi des grands nombres L^2 affirmée en introduction en prenant $f = \pi_0 : (u_n)_n \longmapsto u_0$. De plus, la limite $\pi_0^*\left((X_n)_{n\geqslant 0}\right)$ vérifie alors $\pi_0^* = \pi_0^* \circ T$, ou encore que pour toute suite $u \in \mathbb{R}^{\mathbb{Z}}$, $\pi_0^*\left((u_n)_{n\geqslant 0}\right) = \pi_0^*\left((u_{n+1})_{n\geqslant 0}\right)$. ♣

Remarque II.4. Le théorème ergodique de von Neumann peut être généralisé en prenant un espace de Hilbert général au lieu de $L^2(\mu)$ de la sorte. Soient $\mathcal{H}$ un espace de Hilbert et $U : \mathcal{H} \longrightarrow \mathcal{H}$ une contraction linéaire : pour tout $f \in \mathcal{H}$, $\|Uf\|_{\mathcal{H}} \leqslant \|f\|_{\mathcal{H}}$. Alors si P est la projection de $\mathcal{H}$ sur $\ker(U - \mathrm{id}_{\mathcal{H}})$, on a pour tout $f \in \mathcal{H}$:

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} U^k f - Pf\right\|_{\mathcal{H}} \xrightarrow[n\to\infty]{} 0.$$

Ce qui joue le rôle de U dans le théorème est la précomposition par T : $Uf = f \circ T$. On l'appelle *opérateur de Koopman*. ♣

Prouvons ce théorème, en suivant [Pet83].

Démonstration : L'idée de départ est de bien savoir dans quel espace la limite éventuelle vit. Le théorème donne la réponse : dans $\{f = f \circ T\}$. Notons alors U_T l'opérateur de Koopman : $U_T f = f \circ T$. Alors l'image de P est exactement $\ker(U_T - \mathrm{id})$. Le plan de preuve est alors le suivant.

1. Soit $\mathcal{N} = \ker(U_T - \mathrm{id})^\perp$. On explicite $\mathcal{N}$ en montrant que $\mathcal{N}$ est la fermeture dans $L^2(\mu)$ de l'image de $U_T - \mathrm{id}$.
2. On montre alors que pour tout $f \in \mathcal{N}$, $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ tend vers 0.
3. En utilisant la décomposition $L^2(\mu) = \mathrm{Im}(P) \oplus \mathcal{N}$, on conclut sur le théorème.

Rentrons dans les détails.

1. Pour montrer que $\mathcal{N} = \overline{\mathrm{Im}(U_T - \mathrm{id})}$, on va montrer en fait que $\mathcal{N}^\perp = \mathrm{Im}(U_T - \mathrm{id})^\perp$ ou encore que $\ker(U_T - \mathrm{id}) = \mathrm{Im}(U_T - \mathrm{id})^\perp$. Pour cela, on va utiliser le fait que U_T est une contraction. C'est en effet le cas car pour tout $f \in L^2(\mu)$, on a par lemme de transport

$$\|U_T f\|^2 = \int_X f(T(x))\ \mathrm{d}\mu(x) = \int_X f(x)\ \mathrm{d}\mu \circ T^{-1}(x) = \|f\|^2.$$

Donc U_T est en fait mieux qu'une contraction, c'est une isométrie.

Soit $g \in \mathrm{Im}(U_T - \mathrm{id})^\perp$. Montrons que $U_T g = g$. Pour tout $f \in L^2(\mu)$, on a

$$\langle g, f - f \circ T\rangle = 0.$$

Or, U_T est une contraction, donc est un opérateur borné. Par théorème de représentation de Riesz, U_T admet un adjoint U_T^*. Ainsi, pour tout $f \in L^2(\mu)$, on a

$$\langle g - U_T^* g, f\rangle = 0,$$

si bien que $g = U_T^* g$. Il ne nous reste plus qu'à montrer que $g = U_T g$. Pour cela, il suffit de développer

$$\|U_T g - g\|^2 = \|U_T g\|^2 - 2\langle U_T g, g\rangle + \|g\|^2 = 0,$$

en utilisant l'adjonction et le fait que U_T soit une isométrie. D'où $g = U_T g$, donc on a montré que $\mathrm{Im}(U_T - \mathrm{id})^\perp \subset \mathcal{N}^\perp$.

Réciproquement, si on considère $g \in \ker(U_T - \mathrm{id})$ alors $U_T g = g$. Or, U_T^* est une contraction car

$$\|U_T g\|^2 = \langle g, U_T U_T^* g\rangle \leqslant \|g\|\|U_T^* g\|.$$

Ainsi, puisque $U_T g = g$, on a alors $U_T^* g = g$, toujours en développant $\left\|U_T^* g - g\right\|^2$. Montrons alors que $g \in \mathrm{Im}(U_T - \mathrm{id})^\perp$. Soit $f \in L^2(\mu)$. Alors

$$\langle g, U_T f - f\rangle = \langle U_T^* g - g, f\rangle = 0.$$

D'où l'inclusion.

On a alors montré que $\mathcal{N} = \mathrm{Im}(P)^\perp$ était alors égal à la fermeture de $\mathrm{Im}(U_T - \mathrm{id})$.

2. Notons $\mathcal{N}_0 \overset{\text{def.}}{=} \mathrm{Im}(U_T - \mathrm{id})$. Soit $f \in \mathcal{N}_0$. Montrons que $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ tend vers 0. Il existe $g \in L^2(\mu)$ telle que $f = g \circ T - g$. Ainsi, on a en fait par télescopage

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

Ainsi,

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\right\| \leqslant \frac{2\|g\|}{n},$$

donc la somme $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$ tend bien vers 0 pour tout $f \in \mathcal{N}_0$.

Pour $f \in \mathcal{N}$, on l'approche par un élément de $\mathcal{N}_0$. Soit $\varepsilon > 0$. Il existe $f^\varepsilon \in \mathcal{N}_0$ telle que $\|f - f^\varepsilon\| \leqslant \varepsilon$. Alors, par inégalité triangulaire

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\right\| \leqslant \left\|\frac{1}{n}\sum_{k=0}^{n-1} f^\varepsilon \circ T^k\right\| + \left\|\frac{1}{n}\sum_{k=0}^{n-1}(f - f^\varepsilon) \circ T^k\right\|.$$

D'où

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k\right\| \leqslant \left\|\frac{1}{n}\sum_{k=0}^{n-1} f^\varepsilon \circ T^k\right\| + \varepsilon.$$

Le premier terme tend vers 0, donc en se fixant $n \in \mathbb{N}$ assez grand, on conclut sur la convergence vers 0 de la somme pour tout $f \in \mathcal{N}$.

3. Pour tout $f \in L^2(\mu)$, on a alors

$$f = Pf + f_0,$$

avec $f_0 \in \mathcal{N}$. Ainsi, puisque $Pf \circ T = Pf$:

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = Pf + \frac{1}{n}\sum_{k=0}^{n-1} f_0 \circ T^k.$$

Le deuxième terme tend vers 0 par **2.**, d'où la convergence annoncée dans le théorème. □

Remarque II.5. Avec cette preuve, il n'y a aucune obstruction à considérer une version continue du théorème ergodique en moyenne : si on dispose d'un semi-groupe $(T_t)_{t\geqslant 0}$ (vérifiant $T_0 = \mathrm{id}$ et $T_{t+s} = T_t \circ T_s$) qui préserve μ ($\mu = \mu \circ T_t^{-1}$ pour tout $t \geqslant 0$), alors pour toute fonction $f \in L^2(\mu)$:

$$\int_X \left(\frac{1}{T} \int_0^T f(T_t(x))\, \mathrm{d}t - f^*(x) \right)^2 \mathrm{d}\mu(x) \xrightarrow[T\to\infty]{} 0.$$

C'est d'ailleurs dans cette optique qu'a été prouvé le théorème par von NEUMANN dans [Neu32]. C'est dans cette version continue qu'on comprend le sens de *moyenne temporelle*. L'ergodicité nous aidera à comprendre le terme *moyenne spatiale*. ♣

II.2 Théorème ergodique de Birkhoff (*Pointwise Ergodic Theorem*)

Théorème II.6 : Théorème ergodique de BIRKHOFF (1931)

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré σ-fini et $T : X \longrightarrow X$ une application μ-invariante. Alors, pour tout $f \in L^1(\mu)$, il existe $f^* \in L^1(\mu)$ telle que pour μ-presque tout $x \in X$:

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

De plus, pour μ-presque tout $x \in X$, $f^*(T(x)) = f^*(x)$.

Les preuves de ce théorème reposent toutes sur un même théorème annexe, le théorème ergodique *maximal*.

Théorème II.7 : Théorème ergodique maximal (1939)

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré et $T : X \longrightarrow X$ une application mesurable. Soit $f \in L^1(\mu)$. On définit pour tout $x \in X$: $\bar{f}(x) \overset{\text{def.}}{=} \sup_{n\geqslant 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \in \mathbb{R} \cup \{+\infty\}$. Alors

$$\int_{\{\bar{f}>0\}} f \,\mathrm{d}\mu \geqslant 0.$$

Avant de discuter du théorème ergodique maximal, prouvons alors le théorème ergodique de BIRKHOFF. Dans le reste de ce papier, on va noter, lorsque le contexte est posé (c'est-à-dire si l'on dispose d'un espace mesuré $(X, \mathcal{A}, \mu)$, d'une application $T : X \longrightarrow X$ mesurable et d'une fonction test $f \in L^1(\mu)$) :

$$B_n(f) \overset{\text{def.}}{=} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k.$$

On nommera parfois $B_n(f)$ la n-ème somme de BIKHOFF. La preuve présentée est la plus classique, et provient ici de [Wal82] et de [Pet83]. Une preuve moins classique, mais utilisant tout de même le théorème ergodique maximal est présenté dans [EW11].

Démonstration : (du théorème ergodique de Birkhoff). Voici un plan de démonstration.

1. Supposons d'abord que μ soit une mesure finie. Soient $\alpha < \beta$ deux rationnels. Définissons alors
$$E_{\alpha,\beta} \overset{\text{def.}}{=} \left\{ \liminf_{n\to\infty} B_n(f) < \alpha < \beta < \limsup_{n\to\infty} B_n(f) \right\}.$$
Le but est alors de montrer que $\mu(E_{\alpha,\beta}) = 0$ pour conclure en la convergence presque sûre de $B_n(f)$. La stratégie est alors de montrer que
$$\beta\mu(E_{\alpha,\beta}) \leqslant \int_{E_{\alpha,\beta}} f \,\mathrm{d}\mu \leqslant \alpha\mu(E_{\alpha,\beta}).$$
Or, prouver cette double inégalité est une conséquence astucieuse du théorème ergodique maximal, et définira alors f^* μ-presque partout.
2. On montre que $f^* \in L^1(\mu)$ en utilisant l'invariance de T par μ et le lemme de Fatou. On montre aussi que μ-presque partout $f^* \circ T = f^*$.
3. On conclut alors dans le cas général où μ est de masse infinie, en montrant, *via* une autre application astucieuse du théorème ergodique maximal, que $\mu(E_{\alpha,\beta}) < \infty$ et en appliquant ce qu'on fait en **1.**.

1. Supposons que μ soit finie. Soient $\alpha < \beta$ deux rationnels. D'une part, si $\mu(E_{\alpha,\beta}) = 0$, on écrit simplement $0 = 0 = 0$. Supposons alors par l'absurde que $\mu(E_{\alpha,\beta}) > 0$.

Définissons sur $(X, \mathcal{A})$ la mesure trace sur E_α de μ :

$$\forall A \in \mathcal{A}, \mu_{\alpha,\beta}(A) \overset{\text{def.}}{=} \frac{\mu(A \cap E_{\alpha,\beta})}{\mu(E_{\alpha,\beta})}.$$

Il s'agit alors d'une nouvelle mesure sur $(X, \mathcal{A})$, qui est alors probabilisé, et sur lequel il est tout à fait possible d'utiliser le théorème ergodique maximal (à $x \longmapsto f(x) - \beta$ qui est intégrable car μ est finie) :

$$\int_{\{\bar{f}>\beta\}} f \,\mathrm{d}\mu_{\alpha,\beta} \geqslant \beta\mu_{\alpha,\beta}(\bar{f} > \beta).$$

Par définition de $\mu_{\alpha,\beta}$, on obtient alors

$$\int_{\{\bar f>\beta\}} f\ \mathrm{d}\mu \geqslant \beta\mu_{\alpha,\beta}(\bar f>\beta)\mu(E_{\alpha,\beta}).$$

Or, μ-presque sûrement, $\bar f \geqslant \limsup_{n\to\infty} B_n(f)$. Ainsi, on a montré l'inclusion $E_{\alpha,\beta}\subset\{\bar f>\beta\}$. D'où

$$1=\mu_{\alpha,\beta}(E_{\alpha,\beta})\leqslant \mu_{\alpha,\beta}(\bar f>\beta)\leqslant 1.$$

Ainsi, $\mu_{\alpha,\beta}(\{\bar f>\beta\}\ \Delta\ E_{\alpha,\beta})=0$ et donc

$$\int_{\{\bar f>\beta\}} f\ \mathrm{d}\mu_{\alpha,\beta}=\int_{E_{\alpha,\beta}} f\ \mathrm{d}\mu_{\alpha,\beta}.$$

On a alors montré que

$$\int_{E_{\alpha,\beta}} f\ \mathrm{d}\mu \geqslant \beta\mu_{\alpha,\beta}(E_{\alpha,\beta})\mu(E_{\alpha,\beta})=\beta\mu(E_{\alpha,\beta}).$$

Pour l'autre inégalité, même routine mais avec $-f$, en utilisant $\limsup -B_n(f)=-\liminf B_n(f)>-\alpha$. On aura alors

$$\int_{E_{\alpha,\beta}} f\ \mathrm{d}\mu \geqslant -\alpha\mu(E_{\alpha,\beta}).$$

On conclut en multipliant par -1 de par et d'autre. On a alors montré que

$$\beta\mu(E_{\alpha,\beta})\leqslant\int_{E_{\alpha,\beta}} f\ \mathrm{d}\mu\leqslant\alpha\mu(E_{\alpha,\beta}).$$

Or, $\alpha<\beta$, donc $\mu(E_{\alpha,\beta})=0$, pour tout $\alpha<\beta$ rationnels, d'où

$$\mu\left(\bigcup_{\alpha,\beta\in\mathbb{Q}}\left\{\liminf_{n\to\infty} B_n(f)<\alpha<\beta<\limsup_{n\to\infty} B_n(f)\right\}\right)=0.$$

Soit encore

$$\mu\left(\liminf_{n\to\infty} B_n(f)<\limsup_{n\to\infty} B_n(f)\right)=0.$$

On a alors prouvé que $(B_n(f))_{n\geqslant1}$ converge μ-presque partout vers un certain f^*.

2. Montrons que $f^*\in L^1(\mu)$. Pour cela, on utilise le lemme de Fatou :

$$\int_X |f^*(x)|\ \mathrm{d}\mu(x)\leqslant\liminf_{n\to\infty}\frac1n\sum_{k=0}^{n-1}\int_X |f\circ T^k(x)|\ \mathrm{d}\mu(x).$$

Puisque T est μ-invariante, il suit que $\|f^*\|_{L^1(\mu)}\leqslant\|f\|_{L^1(\mu)}<\infty$.

Montrons que f^* est T-invariante. C'est en fait aussi le cas de $\limsup B_n(f)$ et de $\liminf B_n(f)$. Pour tout $n\geqslant1$

$$B_n(f)\circ T-B_n(f)=\frac{f\circ T^n-f}{n}$$

Donc, puisque $f\in L^1(\mu)$ et T est μ-invariante,

$$B_n(f)\circ T-B_n(f)\xrightarrow[n\to\infty]{L^1(\mu)}0.$$

Par théorème de Riesz-Fischer, la convergence est aussi μ-presque partout pour une sous-suite. Or,

$$B_n(f)\circ T-B_n(f)\xrightarrow[n\to\infty]{\mu-\text{p.p.}}f^*\circ T-f^*.$$

D'où $f^*\circ T=f^*$ μ-presque partout.

3. Supposons que $\mu(X)=+\infty$. Soient $\alpha<\beta$ deux rationnels. Montrons que $\mu(E_{\alpha,\beta})<\infty$. Puisque X est σ-finie, il existe une suite croissante $(X_n)_n$ de mesurables tels que $X=\bigcup_{n\geqslant1}^{\uparrow}X_n$ et $\mu(X_n)<\infty$ pour tout $n\geqslant1$.

Fixons $n\geqslant1$ et considérons $C\overset{\text{def.}}{=}E_{\alpha,\beta}\cap X_n$. Alors $\mu(C)<\infty$. Montrons que

$$\mu(C)\leqslant\frac1\beta\int_X |f|\ \mathrm{d}\mu.$$

Considérons pour cela $h\overset{\text{def.}}{=}f-\beta\mathbf{1}_C\in L^1(\mu)$. Alors $C\subset\{\bar h>0\}$. En effet, dans C, $\limsup_{n\to\infty}B_n(f)>\beta$, si bien qu'il existe une sous-suite $(\phi(n))_n$ telle que pour tout $n\geqslant1$, $B_{\phi(n)}(f)>\beta$ et donc que

$$B_{\phi(n)}(h)>\beta B_{\phi(n)}(1-\mathbf{1}_C)=0,$$

puisqu'on se place dans C. Il suit que $C\subset\{\bar h>0\}$. Or, par le théorème ergodique maximal,

$$\int_{\{\bar h>0\}}(f-\beta\mathbf{1}_C)\ \mathrm{d}\mu\geqslant0.$$

Ainsi,

$$\int_X |f|\ \mathrm{d}\mu\geqslant\int_{\{\bar h>0\}} f\ \mathrm{d}\mu\geqslant\beta\mu\left(C\cap\left\{\bar h>0\right\}\right)=\beta\mu(C).$$

Ainsi, on a montré que pour $n\geqslant1$, avec $C=E_{\alpha,\beta}\cap X_n$:

$$\mu(E_{\alpha,\beta}\cap X_n)\leqslant\frac1\beta\int_X |f|\ \mathrm{d}\mu.$$

On conclut en passant à la limite que $\mu(E_{\alpha,\beta})\leqslant\frac1\beta\int_X|f|\ \mathrm{d}\mu<\infty$. Le raisonnement en **1.** permet alors de conclure au théorème ergodique de Birkhoff, lorsque $\beta>0$. Si $\beta\leqslant0$, alors $\alpha<0$, et le même raisonnement s'applique à $-f$. □

Remarque II.8. Si A,B sont deux ensembles, on appelle *différence symétrique de A et B*, noté $A\Delta B$, l'ensemble $A\backslash B\sqcup B\backslash A$. Pour une mesure finie μ et deux ensembles mesurables A,B, si $\mu(A\Delta B)=0$ alors $\mu(A)=\mu(B)$. La réciproque est vraie si l'un est inclus dans l'autre. ♣

Pour être complet, donnons alors une démonstration du théorème ergodique maximal. On va en fait en présenter deux. La première fonctionne dans un contexte plus général de contraction positive, et est la preuve donnée par [Wal82]. La deuxième fait apparaître des gratte-ciels et donneront l'occasion de voir d'autres raisonnements de théorie ergodique. Elle provient de [Pet83].

Démonstration : (du théorème ergodique maximal).
[Démo 1] Soit pour tout $N\geqslant1$:

$$\bar S_N(x)\overset{\text{def.}}{=}\sup_{1\leqslant n\leqslant N}\sum_{k=0}^{n-1}f(T^k(x)).$$

Le plan de démonstration est le suivant.

1. Montrons qu'il suffit de montrer que

$$\forall N\geqslant1,\int_{\{\bar S_N>0\}} f\ \mathrm{d}\mu\geqslant0.$$

2. On montre, *via* la croissance de $\left(\bar S_N\right)_N$ l'inégalité

$$f\geqslant\bar S_N-\bar S_N^+\circ T.$$

3. On intègre alors cette inégalité sur $\{\bar S_N>0\}$ pour conclure.

Démarrons la preuve de ce théorème.

1. La suite de fonctions $\left(\bar S_N\right)_{N\geqslant1}$ s'agit d'une suite croissante de fonctions. Alors, on a

$$\{\bar f>0\}=\bigcup_{N\geqslant1}^{\uparrow}\{\bar S_N>0\},$$

de sorte que par convergence monotone, on a μ-presque partout :

$$\int_{\{\bar f>0\}} f\ \mathrm{d}\mu=\lim_{N\to\infty}\int_{\{\bar S_N>0\}} f\ \mathrm{d}\mu.$$

Ainsi, le but de cette démonstration est de prouver que

$$\forall N\geqslant1,\int_{\{\bar S_N>0\}} f\ \mathrm{d}\mu\geqslant0.$$

2. Montrons que pour tout $N \geqslant 1$,
$$f \geqslant \bar{S}_N - \bar{S}_N^+ \circ T,$$
où $\bar{S}_N^+ = \max\left\{\bar{S}_N, 0\right\}$. L'ingrédient principal pour cela la croissance de la suite de fonctions $\left(\bar{S}_N\right)_N$. Pour $N = 1$, on écrit $f \geqslant f - f^+ \circ T$, avec $f^+ \circ T$ qui est une fonction positive. Il suit que le cas $N = 1$ est vrai. Pour $N \geqslant 1$ et $1 \leqslant n \leqslant N$:
$$\sum_{k=0}^{n-1} f \circ T^k = f + \left(\sum_{k=0}^{n-2} f \circ T^k\right) \circ T.$$
Ainsi, par définition de $\bar{S}_n$:
$$\sum_{k=0}^{n-1} f \circ T^k \leqslant f + \bar{S}_{n-1} \circ T.$$
Par croissance de la suite $(\bar{S}_n)_n$:
$$\sum_{k=0}^{n-1} f \circ T^k \leqslant f + \bar{S}_N \circ T \leqslant f + \bar{S}_N^+ \circ T.$$
Ainsi, le membre de droite est bien indépendant de n, donc il suit
$$\bar{S}_N \leqslant f + \bar{S}_N^+ \circ T,$$
ce qui est l'inégalité affirmée.

3. Intégrons alors cette inégalité :
$$\int_{\{\bar{S}_N > 0\}} f \, \mathrm{d}\mu \geqslant \int_{\{\bar{S}_N > 0\}} \left(\bar{S}_N - \bar{S}_N^+ \circ T\right) \, \mathrm{d}\mu.$$
Sur $\left\{\bar{S}_N > 0\right\}$, $\bar{S}_N^+ = \bar{S}_N$:
$$\int_{\{\bar{S}_N > 0\}} f \, \mathrm{d}\mu \geqslant \int_{\{\bar{S}_N > 0\}} \left(\bar{S}_N - \bar{S}_N \circ T\right) \, \mathrm{d}\mu.$$
L'invariance de T par μ conclut alors que le membre de droite est nul, ce qui conclut sur le théorème ergodique maximal.

[Démo 2] □

II.3 Ergodicité

Supposons que T soit une application μ-invariante sur X. S'il existe $B \in \mathcal{A}$ tel que $B = \{T \in B\}$, alors $X \backslash B = \{T \in X \backslash B\}$. De plus, si $0 < \mu(B) < \mu(X)$, cela signifie qu'on peut se contenter d'étudier $T_{|_B}$ ou $T_{|_{X\backslash B}}$.

L'ergodicité établit alors si on se situe dans un cadre où on peut simplifier l'étude de T dans le sens qu'on vient d'expliquer. Une formulation équivalente est *métriquement transitif*. Un autre intérêt de l'ergodicité est de pouvoir qualitativement situer le temps moyen de retour dans un mesurable donné *via* le lemme de Kac. Enfin, les théorèmes ergodiques dans le cadre ergodique sont explicites, la limite $f^*(x)$ ne dépendra pas de x et vaudra alors la moyenne dite spatiale $\int_X f(x) \, \mathrm{d}\mu(x)$. Commençons par définir l'ergodicité.

Définition II.9

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et $T : X \longrightarrow X$ une application mesurable μ-invariante. On dit que T est *ergodique* si pour tout mesurable $A \in \mathcal{A}$ tel que $\mu(A \,\Delta\, \{T \in A\}) = 0$, on a $\mu(A) = 0$ ou $\mu(X \backslash A) = 0$.

Si A vérifie $\mu(A \,\Delta\, \{T \in A\}) = 0$, on dira que A est invariant par T. Donnons une caractérisation pratique de l'ergodicité.

Proposition II.10 : Caractérisation en termes de fonctions constantes de l'ergodicité

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré et $T : X \longrightarrow X$ une application μ-invariante. Alors les assertions suivantes sont équivalentes :

1. L'application T est ergodique ;
2. Pour toute application mesurable $f \in L^0(\mu)$, si $f \circ T = f$, f est μ-presque partout constante (il existe $c \in \mathbb{R}$ tel que $\mu(X \backslash \{f = c\}) = 0$).

Remarque II.11. Lorsque μ est finie, la même preuve de l'implication directe montre que l'ergodicité de T est équivalente au fait que toute application $f \in L^2(\mu)$ vérifiant $f = f \circ T$ μ-presque partout est constante μ-presque partout. L'intérêt est alors d'utiliser le fait que toute fonction $f \in L^2(\mu)$ est développable en série de Fourier dans L^2 pour montrer effectivement que $f = f \circ T$. ♣

Démonstration : [$\Longrightarrow$] Si T est ergodique, considérons $f \in L^0(X)$ vérifiant $f = f \circ T$, et considérons sa fonction de répartition : pour tout $x \in \mathbb{R}$
$$F(x) \overset{\text{def.}}{=} \mu(f \leqslant x).$$
Alors F est croissante, tend vers 0 en $-\infty$ et vers $\mu(X)$ en $+\infty$. De plus, puisque $f = f \circ T$, on a
$$\{T \in \{f \leqslant x\}\} = \{f \circ T \leqslant x\} = \{f \leqslant x\}.$$
Ainsi, $\{f \leqslant x\}$ est invariant par T. Puisque T est ergodique, il suit que $F(x) = 0$ ou $F(x) = \mu(X)$. Puisque F est croissante, il existe alors $x_0 \in \mathbb{R}$ tel que $F(x) = 0$ pour tout $x < x_0$ et $F(x) = \mu(X)$ pour tout $x > x_0$. Or,
$$\mu(X \backslash \{f = x_0\}) \leqslant \sum_{n=1}^{\infty} \left(\mu\left(f < x_0 - \frac{1}{n}\right) + \mu\left(f > x_0 + \frac{1}{n}\right)\right) = 0.$$
donc f est constante, égale à x_0 μ-presque partout.

[$\Longleftarrow$] Supposons que tout $f \in L^0(X)$ tel que $f = f \circ T$ est constante presque partout. Considérons $A \in \mathcal{A}$ un ensemble mesurable T-invariant. Alors
$$\mathbf{1}_A \circ T = \mathbf{1}_{\{T \in A\}} = \mathbf{1}_A,$$
si bien que $\mathbf{1}_A$ se retrouve constant presque partout, égal à 0 ou 1, donc $\mu(A) = 0$ ou $\mu(X \backslash A) = 0$, donc T est ergodique. □

Un aspect fondamental de l'ergodicité est de pouvoir identifier la limite dans le théorème ergodique : les moyennes temporelles et les moyennes spatiales coïncident. Attention, on se place ici dans le cas où la mesure est finie.

Pour commencer, évoquons dans le prochain une conséquence du théorème ergodique lorsqu'on se place en mesure finie.

Lemme II.12

Soient $(X, \mathcal{A}, \mu)$ est un espace mesuré fini et $T : X \longrightarrow X$ une application μ-invariante. Alors, pour tout $f \in L^1(\mu)$, le théorème ergodique de BIRKHOFF a aussi lieu dans $L^1(\mu)$

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow[n\to\infty]{L^1(\mu)} f^*.$$

De plus, on a alors

$$\int_X f \, \mathrm{d}\mu = \int_X f^* \, \mathrm{d}\mu.$$

Démonstration du lemme : Soit $f \in L^1(\mu)$. Alors, puisque μ est finie, pour tout $\varepsilon > 0$, il existe une application $f^\varepsilon \in L^\infty(\mu)$ telle que

$$\int_X |f - f^\varepsilon| \, \mathrm{d}\mu \leqslant \varepsilon.$$

Rappelons que $B_n(f) \overset{\text{def.}}{=} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k$. Alors

$$\int_X |f^* - B_n(f)| \, \mathrm{d}\mu \leqslant \int_X |f^* - f^{\varepsilon *}| \, \mathrm{d}\mu + \int_X |f^{\varepsilon *} - B_n(f^\varepsilon)| \, \mathrm{d}\mu + \int_X |B_n(f^\varepsilon) - B_n(f)| \, \mathrm{d}\mu.$$

Les premier et troisième termes sont inférieurs à ε. Quand au deuxième, le théorème ergodique donne la convergence μ-presque partout de $B_n(f^\varepsilon)$ vers $f^{\varepsilon *}$, avec la domination $|B_n(f^\varepsilon)| \leqslant \|f^\varepsilon\|_\infty$, qui est bien intégrable car μ est finie. Par convergence dominée, le deuxième terme tend vers 0 lorsque $[n \to \infty]$. Il suit alors que pour n assez grand, $\|f^* - B_n(f)\|_{L^1(\mu)} \leqslant 3\varepsilon$, si bien qu'on conclut sur la convergence L^1 dans le théorème ergodique.

Ainsi, on conclut aussi sur $\int_X f^* \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu$ *via* le fait que pour tout $n \geqslant 1$, $\int_X f \, \mathrm{d}\mu = \int_X B_n(f) \, \mathrm{d}\mu$ et donc

$$\left| \int_X f \, \mathrm{d}\mu - \int_X f^* \, \mathrm{d}\mu \right| \leqslant \int_X |B_n(f) - f^*| \, \mathrm{d}\mu.$$

□

Nous pouvons alors enfin évoquer le théorème ergodique dans le cadre ergodique.

Théorème II.13 : Théorème ergodique et ergodicité

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré fini et $T : X \longrightarrow X$ une application μ-invariante. Alors les assertions suivantes sont équivalentes.

1. L'application T est ergodique ;
2. Pour tout $f \in L^1(\mu)$,

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

Démontrons ce théorème avant quelques remarques.

Démonstration : **[1. $\Longrightarrow$ 2.]** Supposons que T soit ergodique. Par le théorème ergodique de Birkhoff, les sommes $B_n(f)$ convergent vers f^* quand $[n \to \infty]$, μ-presque partout, et dans $L^1(\mu)$ par le lemme précédent.. Or, l'application f^* vérifie $f^* \circ T = f^*$. Puisque T est ergodique, il suit que f^* est constante *via* la caractérisation avec les fonctions constantes. De plus, par le lemme précédent, la constante f^* vérifie alors

$$\mu(X) f^* = \int_X f \, \mathrm{d}\mu.$$

Il suit qu'on obtient **2.**.

[2. $\Longrightarrow$ 1.] Supposons qu'on soit dans le cadre du théorème ergodique de Birkhoff, avec pour tout $f \in L^1(\mu)$, $f^* = \frac{1}{\mu(X)} \int_X f \, \mathrm{d}\mu$. Montrons que T est ergodique. Soit $f \in L^1(\mu)$ telle que $f \circ T = f$. Alors pour tout $n \geqslant 1$, $B_n(f) = f$. Ainsi, en passant à la limite, par hypothèse **2.**, on conclut que $f = \frac{1}{\mu(X)} \int_X f \, \mathrm{d}\mu$ est en fait constante. On ne peut pas encore conclure *via* la caractérisation en termes de fonctions constantes, puisqu'on l'a énoncé pour $f \in L^0(\mu)$. Or, la preuve fonctionne de la même manière, imitons-la brièvement. On vient de prouver que pour tout $f \in L^1(\mu)$ telle que $f \circ T = f$, f est constante. Or, si $A \in \mathcal{A}$ est T-invariante, $\mathbf{1}_A \in L^1(\mu)$, donc est constante, donc vérifie $\mu(A) \in \{0, \mu(X)\}$, donc T est ergodique. □

Remarque II.14. Le fil rouge de cet exposé revient à la charge ! En termes de loi des grands nombres, les théorèmes ergodiques s'écrivent sur $\mathbb{R}^{\mathbb{Z}}$ muni de la mesure de probabilité $\mu = \mathbb{P} \circ \Phi_X^{-1}$ (invariante si et seulement si $(X_n)_n$ est stationnaire) :

$$\forall f \in L^1(\mu), \frac{1}{n}\sum_{k=0}^{n-1} f\left((u_{n+k})_{n\in\mathbb{Z}}\right) \xrightarrow[n\to\infty]{\mu-\text{p.s.}, L^1(\mu)} \int_{\mathbb{R}^{\mathbb{Z}}} f(u) \, \mathrm{d}\mu(u).$$

Cela revient alors à dire, par définition de $\mu = \mathbb{P} \circ \Phi_X^{-1}$ que pour tout $f : \mathbb{R}^{\mathbb{Z}} \longrightarrow \mathbb{R}$ telle que $\mathbb{E}\left[\left|f\left((X_n)_{n\in\mathbb{Z}}\right)\right|\right] < \infty$:

$$\frac{1}{n}\sum_{k=0}^{n-1} f\left((X_{n+k})_{n\in\mathbb{Z}}\right) \xrightarrow[n\to\infty]{\mathbb{P}-\text{p.s.}, L^1(\mathbb{P})} \mathbb{E}\left[f\left((X_n)_{n\in\mathbb{Z}}\right)\right].$$

Si on applique ceci à $f = \pi_0 : (u_n)_n \longmapsto u_0$ alors en supposant que $X_0 \in L^1(\mathbb{P})$, on a bien $\pi_0((X_n)_n) \in L^1(\mathbb{P})$ avec

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow[n\to\infty]{\mathbb{P}-\text{p.s.}, L^1(\mathbb{P})} \mathbb{E}\left[X_0\right].$$

C'est la loi forte des grands nombres, en remplaçant les hypothèses moyenne commune et indépendance par stationnarité. ♣

Remarque II.15. Comme dans l'introduction avec la loi des grands nombres, si on considère $f = \mathbf{1}_E$, avec $E \in \mathcal{A}$ alors

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{\{T^k \in E\}} \xrightarrow[n\to\infty]{\mu-\text{p.p}, L^1(\mu)} \mu(E).$$

Le temps de séjour empirique de T dans E converge vers la mesure de E. ♣

Reprenons nos exemples de la première partie pour voir ce que signifient les théorèmes ergodiques.

1. Le cas $X = {}^{\mathbb{Z}}/_{N\mathbb{Z}}$, muni de $\mu = \sharp$ et de $T(x) = x + 1$. Observons d'abord que T est bien ergodique puisque si $f \circ T = f$, cela signifie que que pour tout $x \in {}^{\mathbb{Z}}/_{N\mathbb{Z}}$, $f(x+1) = f(x)$, donc que $f(x) = f(0)$ est constante. Puisque $\sharp$ est finie, cela permet de conclure sur les théorèmes ergodiques : pour tout $f : {}^{\mathbb{Z}}/_{N\mathbb{Z}} \longrightarrow \mathbb{R}$

$$\frac{1}{n}\sum_{k=0}^{n-1} f(x+k) \xrightarrow[n\to\infty]{} \frac{1}{\sharp {}^{\mathbb{Z}}/_{N\mathbb{Z}}} \int_{{}^{\mathbb{Z}}/_{N\mathbb{Z}}} f(x)\ \mathrm{d}\sharp(x) = \frac{1}{N}\sum_{x\in {}^{\mathbb{Z}}/_{N\mathbb{Z}}} f(x).$$

On aurait pu le voir à la main, puisque pour n assez grand, la division euclidienne de $n-1$ par N donne

$$n - 1 = \left\lfloor \frac{n-1}{N} \right\rfloor N + r_n,$$

avec pour tout $n \geqslant 1$, $0 \leqslant r_n < N$, et on note $q_n = \left\lfloor \frac{n-1}{N} \right\rfloor$. Alors on a la partition

$$[\![0, n-1]\!] = \left(\bigsqcup_{k=0}^{q-1} [\![kN, kN + (N-1)]\!]\right) \sqcup [\![qN, qN + r]\!].$$

Ainsi, la somme de Birkhoff s'écrit aussi

$$\frac{1}{n}\sum_{k=0}^{n-1} f(x+k) = \frac{1}{n}\sum_{k=0}^{q_n}(f(kx) + f(kx+1) + \cdots + f(kx + (N-1)) + \frac{f(q_nN) + \cdots + f(q_nN + r_n)}{n}.$$

Puisqu'on est sur ${}^{\mathbb{Z}}/_{N\mathbb{Z}}$, cela se simplifie en

$$\frac{1}{n}\sum_{k=0}^{n-1} f(x+k) = \frac{q_n}{n}\sum_{x\in {}^{\mathbb{Z}}/_{N\mathbb{Z}}} f(x) + \frac{1}{n}\sum_{k=0}^{r_n} f(k).$$

Le premier terme tend vers $\frac{1}{N}\sum_{x\in {}^{\mathbb{Z}}/_{N\mathbb{Z}}} f(x)$ tandis que le deuxième tend vérifie

$$\left|\frac{1}{n}\sum_{k=0}^{r_n} f(k)\right| \leqslant \frac{1}{n}\sum_{x\in {}^{\mathbb{Z}}/_{N\mathbb{Z}}} |f(x)|$$

et tend vers 0 lorsque $[n \to \infty]$. On retrouve la convergence ponctuelle affirmée dans le théorème ergodique de Birkhoff. Notons que bien que l'exemple semble trivial, la convergence ne l'est pas (elle n'est pas stationnaire). On dispose même ici d'une vitesse de convergence en $\frac{1}{n}$. Nous verrons beaucoup plus tard qu'on n'en dispose pas en général.

2. L'application doublante est ergodique. Pour cela, on se donne $f \in L^2(0,1)$ telle que $f \circ T = f$, et on développe f en série de Fourier. La condition $f \circ T = f$ donne alors pour presque tout $0 \leqslant x < \frac{1}{2}$:

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\mathrm{i}\pi nx} = \sum_{n\in\mathbb{Z}} a_n e^{2\mathrm{i}\pi n\cdot(2x)} = \sum_{n\in 2\mathbb{Z}} a_{\frac{n}{2}} e^{2\mathrm{i}\pi n\cdot x} = f \circ T(x).$$

Par unicité du développement en série de FOURIER, les a_{2n+1} sont nuls et $a_{2n} = a_n = 0$, lorsqu'on itère les divisions par 2, pour $n \neq 0$. Il suit que f est constante égale à a_0, donc l'application doublante T est ergodique. Notons au passage qu'on peut aussi interpréter cette application comme un décalage. En effet, si on développe x en base 2, $x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{2^n}$ alors l'application

$$\epsilon : \left(\begin{array}{ccc} [0,1) & \longrightarrow & \{0,1\}^{\mathbb{N}^*} \\ x & \longmapsto & (\varepsilon_n(x))_{n\geqslant 1} \end{array} \right)$$

est une bijection. De plus, si $\sigma : (u_n)_n \longmapsto (u_{n+1})_n$ est l'opérateur de décalage, on a en fait $\epsilon \circ T = \sigma \circ \epsilon$. On dit que T est σ sont isomorphes (au sens de la théorie ergodique).

Ainsi, les théories ergodiques nous indiquent que les sommes de BIRKHOFF associées convergent vers $\int_0^1 f \, dx$. Le prouver à la main est une autre affaire...

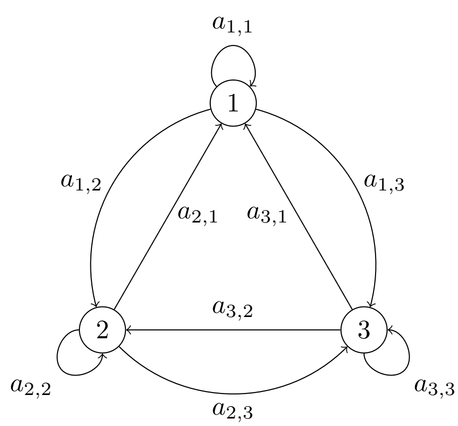
3. La rotation R_θ est ergodique si et seulement si θ est irrationnel.

4. Le décalage de BERNOUILLI est ergodique : si on se donne B cylindrique telle que $\{\sigma \in B\} = B$, $B = \bigcap_{j=1}^d \{\pi_{n_j} = x_j\}$ alors pour $M > \max\{n_1, \dots, n_d\}$, on a $\{n_1, \cdots, n_d\} \cap \{n_1 + M, \cdots, n_d + M\} = \varnothing$, on a

$$\mu(B \cap \{\sigma^M \in B\}) = \mu(B)\mu(\sigma^M \in B) = \mu(B)^2,$$

soit encore $\mu(B) = \mu(B)^2$, donc $\mu(B) \in \{0,1\}$, d'où l'ergodicité.

5. Le décalage de MARKOV est ergodique si et seulement si A est *irréductible* : cela signifie que pour tout $x, y \in \mathfrak{A}$, il existe $k \geqslant 1$ telle que $[A^k]_{x,y} > 0$. Pour le comprendre, rappelons qu'une chaîne de MARKOV vue par exemple sur $\mathfrak{A} = \{1,2,3\}$ peut se résumer à la représentation suivante :



Chaque $a_{x,y}$ représente la probabilité qu'on passe de l'état x à y en une seule étape. $[A^k]_{x,y}$ représente quand à lui la probabilité de passer de x à y en k étapes. Supposer l'existence d'un tel k signifie qu'il existe une probabilité non nulle qu'on aboutisse de l'état x à l'état y après un certain nombre d'étapes. Si un tel k n'existe pas, le graphe n'est plus connexe, et on observe deux dynamiques dans deux graphes différents. On perd alors, par définition, l'ergodicité.

Lemme II.16 : Ergodicité et récurrence

Soient $(X, \mathcal{A}, \mu)$ un espace mesuré, et $T : X \longrightarrow X$ une application μ-invariante ergodique. Alors pour tout $A \in \mathcal{A}$ tel que $\mu(A) > 0$,

$$\mu\left(X \ \Delta \ \bigcup_{n\in\mathbb{N}} \{T^n \in A\}\right) = 0.$$

Autrement dit, $X = \bigcup_{n\in\mathbb{N}} \{T^n \in A\}$ modulo les négligeables pour μ.

Démonstration du lemme : Notons $\tilde{X} = \bigcup_{n\in\mathbb{N}} \{T^n \in A\}$. Alors

$$\{T \in \tilde{X}\} = \bigcup_{n\in\mathbb{N}} \{T^{n+1} \in A\} \subset \tilde{X}.$$

Or, puisque T est μ-invariante, $\mu(\tilde{X}) = \mu(T \in \tilde{X})$. Ainsi,

$$\mu\left(\tilde{X} \ \Delta \ \{T \in \tilde{X}\}\right) = 0.$$

Puisque T est ergodique, on en déduit que $\mu(\tilde{X}) = 0$ ou que $\mu(X \backslash \tilde{X}) = 0$. Puisqu'on a supposé que $\mu(A) > 0$, on a bien $\mu(X \backslash \tilde{X}) = 0$, donc $X = \tilde{X}$ modulo les négligeables pour μ. □

Théorème II.17 : Lemme de Kac

Soit $(X, \mathcal{A}, \mu)$ un espace mesurable fini. Soit $T : X \longrightarrow X$ une application μ-invariante ergodique. Alors, si $n_A \overset{\text{def.}}{=} \inf\{n \geqslant 1, T^n \in A\}$, on a

$$\frac{1}{\mu(A)} \int_A n_A(x) \, \mathrm{d}\mu(x) = \frac{\mu(X)}{\mu(A)}.$$

Autrement dit, le temps moyen de retour vers A est donné par l'inverse de la moyenne temporelle de A.

Démonstration : Comme pour le théorème ergodique maximal, proposons plusieurs démonstrations de ce résultat.

[Démo 1 : *via* les grattes-ciels de Kakutani] On prouve le théorème uniquement dans le cas où T est bijective. Dans ce cas, T^{-1} est aussi ergodique, et par lemme précédent, on a alors, modulo les négligeables

$$X = \bigcup_{j=0}^{+\infty} T^j(A).$$

Or, en utilisant le fait que $\{n_A = \infty\}$ soit négligeable, on peut conclure (avec un jeu par des divisions euclidiennes) qu'on a en fait, modulo les négligeables

$$X \overset{(\mho)}{=} \bigsqcup_{n=1}^{+\infty} \bigsqcup_{k=0}^{n-1} T^k(\{n_A = n\} \cap A).$$

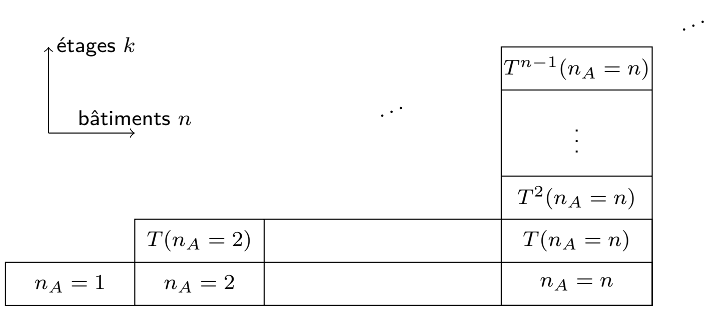
Il suffit de prendre la mesure de ceci pour conclure :

$$\mu(X) = \sum_{n=1}^{+\infty} \sum_{k=0}^{n-1} \mu((\{n_A = n\} \cap A)).$$

Puisque T est invariante, T^{-1} aussi, et

$$\mu(X) = \sum_{n=1}^{+\infty} n\mu((\{n_A = n\} \cap A)) = \int_A n_A \, \mathrm{d}\mu.$$

On peut représenter l'égalité $(\mho)$ *via* ce qu'on appelle les *gratte-ciels de Kakutani* : presque tout $x \in X$ vérifie $n_A(x) < \infty$. On place alors à l'étage 0 tous les x de A selon la valeur de $n_A(x)$. On a créé le rez-de-chaussée de tous les gratte-ciels. Ensuite, pour $n \geqslant 1$, les x du bâtiment n prennent l'ascenseur pour arriver à l'étage $n-1$: on fait agir T à chaque étage. A l'étage $n-1$, faire agir T revient à ce que x prenne un téléporteur vers un rez-de-chaussée d'un autre bâtiment.



[Démo 2] L'idée de cette preuve provient de [Kre85]. D'après le théorème de récurrence de Poincaré, $\mu(\{n_A = +\infty\}) = 0$, si bien qu'on a la décomposition $X = \bigcup_{n=1}^{\infty} \{n_A = n\}$ modulo les négligeables pour μ. Ainsi,

$$\int_A n_A \, \mathrm{d}\mu = \sum_{n=1}^{+\infty} \int_{A \cap \{n_A = n\}} n_A \, \mathrm{d}\mu = \sum_{n=1}^{+\infty} n\mu(\{n_A = n\} \cap A).$$

Ainsi, on doit étudier $\{n_A = n\} \cap A$ pour essayer de comprendre comment se comporte sa mesure.

- Pour cela, écrivons d'abord que par définition

$$\{n_A = n\} = \bigcap_{k=1}^{n-1} \{T^k \notin A\} \cap \{T^n \in A\}.$$

Puis, par propriété des images réciproques

$$\{n_A = n\} = \left\{ T \in A^c \cap \left(\bigcap_{k=1}^{n-2} \{T^k \notin A\} \cap \{T^{n-1} \in A\} \right) \right\}.$$

Autrement dit, en termes de n_A, on a la relation de récurrence

$$\{n_A = n\} = \{T \in A^c \cap \{n_A = n-1\}\}.$$

Puisque T est μ-invariante, on en déduit que

$$\mu(n_A = n) = \mu(\{n_A = n-1\} \cap A^c).$$

Ou encore que

$$\mu(\{n_A = n\} \cap A) = \mu(\{n_A = n-1\} \cap A^c) - \mu(\{n_A = n\} \cap A^c).$$

Ainsi, on a alors pour tout $N \geqslant 1$:

$$\sum_{p=1}^{N} \mu(\{n_A = p+n\} \cap A) = \mu(\{n_A = n\} \cap A^c) - \mu(\{n_A = n+N\} \cap A^c).$$

Or, $\mu(n_A = n+N)$ tend vers 0 lorsque N tend vers $+\infty$, puisqu'il s'agit d'un terme général d'une série convergente. Ainsi, on a l'égalité pour tout $n \geqslant 1$:

$$\sum_{p=1}^{+\infty} \mu(\{n_A = p+n\} \cap A) = \mu(\{n_A = n\} \cap A^c).$$

Pour $n = 0$, cette égalité se réduit à

$$\sum_{p=1}^{+\infty} \mu(\{n_A = p\} \cap A) = \mu(A).$$

- Revenons alors sur le calcul de $\int_A n_A \, \mathrm{d}\mu$. L'astuce ici réside dans l'interprétation du facteur n qui apparaît devant μ pour faire apparaître ce qu'on vient de déterminer précédemment.

$$\int_A n_A \, \mathrm{d}\mu = \sum_{n=1}^{+\infty} \sum_{p=1}^{n} \mu(\{n_A = n\} \cap A).$$

Et on intervertit les sommes.

$$\int_A n_A \, \mathrm{d}\mu = \sum_{p=1}^{+\infty} \sum_{n=p}^{+\infty} \mu(\{n_A = n\} \cap A).$$

D'après le point précédent,

$$\int_A n_A \, \mathrm{d}\mu = \mu(A) + \sum_{p=1}^{+\infty} \mu(\{n_A = p\} \cap A^c).$$

On interprète le tout en terme d'image réciproque de T.

$$\int_A n_A \, \mathrm{d}\mu = \mu(A) + \sum_{p=1}^{+\infty} \mu\left(\bigcap_{k=0}^{p-1} \{T^k \notin A\} \cap \{T^p \in A\} \right).$$

On peut regrouper en une seule somme.

$$\int_A n_A \, \mathrm{d}\mu = \sum_{p=0}^{+\infty} \mu\left(\bigcap_{k=0}^{p-1} \{T^k \notin A\} \cap \{T^p \in A\} \right).$$

Enfin, comme on l'a remarqué en début de démonstration, on a

$$\bigsqcup_{p=0}^{\infty}\left(\bigcap_{k=0}^{p-1}\{T^k \notin A\}\right) \cap \{T^p \in A\} = \bigcup_{n\in\mathbb{N}}\{T^n \in A\}.$$

Par le lemme précédent, on conclut que

$$\int_A n_A \, \mathrm{d}\mu = \mu\left(\bigcup_{n\in\mathbb{N}}\{T^n \in A\}\right) = \mu(X).$$

[Démo 3 : *via* le théorème ergodique] On utilise le fait suivant : si $A \in \mathcal{A}$ vérifie $\mu(A) > 0$ et si on considère $T_A : x \in A \longmapsto T^{n_A(x)}(x)$ l'application induite alors sur $(A, \mathcal{A}_{|_A}, \mu_{|_A})$, T_A est une application ergodique. D'après le théorème ergodique de Birkhoff, on a alors

$$\frac{1}{n}\sum_{k=0}^{n-1} n_A \circ T_A^k \xrightarrow[n\to\infty]{\mu-\text{p.p.}} \frac{\mu(X)}{\mu(A)}\int_A n_A \, \mathrm{d}\mu.$$

Si $N_n(x) \overset{\text{def.}}{=} \sum_{k=0}^{n-1} n_A \circ T_A^k(x)$, pour $x \in A$, alors on a pour presque tout $x \in A$

$$T_A^n(x) = T^{N_n(x)}(x).$$

Ainsi, pour tout $n \geqslant 1$, et presque tout $x \in A$

$$n = \sharp\{1 \leqslant k \leqslant N_n(x), T^k(x) \in A\}.$$

Or, par le théorème ergodique pour T, pour presque tout $x \in A$:

$$\frac{1}{N_n(x)}\sum_{k=1}^{N_n(x)} \mathbf{1}_{\{T^k(x)\in A\}} \xrightarrow[n\to\infty]{} \mu(A).$$

On a alors prouvé que d'une part $\frac{N_n}{n}$ converge μ-presque partout sur A vers $\frac{\mu(X)}{\mu(A)}\int_A n_A \, \mathrm{d}\mu$, et d'autre part que presque partout sur A, $\frac{n}{N_n}$ converge vers $\mu(A)$. Par unicité, on conclut sur l'égalité presque partout. □

III Absence de vitesse de convergence dans le théorème ergodique

IV Ergodicité du Mouvement Brownien

Références

[Der54] C. Derman. « Ergodic Property of the Brownian Motion Process ». In : *Proceedings of the National Academy of Sciences of the United States of America Vol. 40, No. 12 (Dec. 15, 1954), pp. 1155-1158 (4 pages)* (1954).

[EW11] M. Einsiedler et T. Ward. *Ergodic Theory with a view towards Number Theory*. Springer, 2011.

[JR79] A. del Junco et J. Rosenblatt. « Counterexamples in Ergodic Theory and Number Theory ». In : *Math. Ann. 245, 185–197* (1979).

[Kle20] A. Klenke. *Probability Theory : a comprehensive course*. Springer, 2020.

[KR53] G. Kallianpur et H. Robbins. « Ergodic Property of the Brownian Motion Process ». In : *Proceedings of the National Academy of Sciences of the United States of America Vol. 39, No. 6 (Jun. 15, 1953), pp. 525-533 (9 pages)* (1953).

[Kre85] U. Krengel. *Ergodic Theorems*. de Gruyter, 1985.

[Neu32] J. von Neumann. « Proof of the quasi-ergodicity hypothesis ». In : *Proceedings of the National Academy of Sciences, Vol. 18 No. 1 pp. 70-82* (1932).

[Pet83] K. Petersen. *Ergodic Theory*. Cambridge studies in advanced mathematics 2, 1983.

[Wal82] P. Walters. *An Introduction to Ergodic Theory*. Springer, 1982.