SMADJA David

Théorie ergodique : Premiers théorèmes et applications à l'étude des systèmes dynamiques

Mémoire d'Initiation à la Recherche

Sous la direction de Monsieur **Jacques Fejoz**

Cycle **P**luridisciplinaire d'**E**tudes **S**upérieures
Troisième année (L3)

A mon père,
qui a fait de moi celui que je suis aujourd'hui

Table des matières

Remerciements

Je mesure l'immense chance que j'ai eu d'avoir pu rédiger cette toute première ébauche de recherche sous la direction de mon professeur Jacques Fejoz. Merci infiniment à lui d'avoir pris sur son temps précieux pour m'avoir guidé lorsque j'en avais besoin, pour m'avoir conseillé lorsque le doute se faisait ressentir en période de concours. En outre, je ne peux passer sous silence comme son cours d'Intégration de Lebesgue au premier semestre m'a passionné, si bien que j'ai justement désiré d'en savoir plus avec ce présent document sur la théorie ergodique et ses applications à l'étude des systèmes dynamiques. Encore un grand merci à vous Mr Fejoz.

C'est l'occasion pour moi de remercier également mon professeur et responsable du CPES à Dauphine, Guillaume Vigéral, pour m'avoir permis d'intégrer cette année le CPES et ainsi rejoindre la grande famille PSL. Notamment, je voudrais également le remercier de m'avoir permis de suivre en candidat libre l'UE PSL de Mécanique Quantique au premier semestre qui m'a tout bonnement passionné.

Merci également à mon professeur et responsable de la Licence 3 Maths à Dauphine, Jimmy Lamboley, d'avoir accepté mon idée de faire passer des oraux blancs aux élèves désireux de passer les concours d'ingénieurs. Que ce projet puisse perdurer avec le temps et permette à davantage d'étudiants d'intégrer de grandes écoles, ce qui n'est qu'un gage de l'excellent niveau de Dauphine. Faire face à l'université et aux concours n'est pas chose facile et l'aide de nos professeurs fut véritablement précieuse.

Bien évidemment, aussi modeste qu'il est, ce mémoire n'aurait pas été possible sans tous ceux qui m'ont soutenu lors de cette dure année où au niveau soutenu des cours de L3 s'est rajouté le stress intense des concours. Mais le travail, la volonté et le soutien de mes proches ont fini par payer puisqu'au moment où j'achève ce mémoire, je sais depuis quelques jours que j'ai été admis à l'école Polytechnique. Un nouveau chapitre s'ouvre ainsi à moi, en espérant qu'il sera rempli d'heureuses rencontres et de belles découvertes mathématiques.

Merci tout d'abord à ma mère qui a ressenti mon stress plus que quiconque et m'a appuyé et encouragé comme personne. Cette école, c'est à elle que je la dédie ! Merci à ma sœur qui a toujours le mot pour me faire rire et me redonner le sourire, à mon frère qui sait si bien m'encourager et m'écouter des heures durant monologuer sur la force des probabilités sans m'interrompre par politesse. Vous trois êtes ma plus belle raison, ma plus belle motivation ; ceux que j'ai envie de rendre les plus fiers au monde.

Merci à tous mes amis dauphinois qui ont toujours cru en moi (parfois trop haha !). Merci à Leslie d'être toujours aussi attentive et bienveillante à mon égard. À Lucie pour m'avoir autant materné et encouragé. À Cyrine d'avoir passé autant d'heures au téléphone à me soutenir. À Eytan et Dorone, mes auteurs préférés, de savoir me redonner confiance en moi (en plus de m'avoir fait découvrir la trépidante vie barcelonaise). À Amina pour les milliards de textos, messages, snapshats farfelus qu'on s'est envoyé pour se donner courage et tous ceux qu'on continuera à s'envoyer à Palaisau. À Flavian pour m'avoir donné l'envie de porter l'uniforme depuis Hong-Kong. À Jean-Charles qui a cru en mes capacités de nageur émérite. À Emma pour toujours savoir mettre l'ambiance pendant les cours de recherche (et faire les meilleurs gâteaux au chocolat du monde). À Quentin et Nicolas, mes grands amis et super coachs pour m'avoir fait passé les meilleurs footings de ma vie où l'on pouvait discuter de tout et n'importe quoi. À Hugo pour tous nos oraux d'entraînement. Merci à Raphaël. À Ilan. À Gabrielle. À Kevin. À Nathan. À Clément. À Rayanne. À Julia. À Ruth. À Pablo. À Nelly. À tous ceux que j'ai pu oublier ; vous avez tous une grande place dans mon cœur.

Merci à toute ma famille qui ne comprend pas forcément pourquoi un tel engouement pour les équations et théorèmes. Merci à tous mes professeurs qui m'ont donné la passion de ce que je fais.

Merci à Julie qui est de loin la meilleure coach de l'Univers ! Si je suis arrivé jusque là, c'est grâce à elle.

Et enfin, merci à mon écureuil Haïku. Que les nuits à métaphysiquer sur Birkhoff jusqu'à 3 heures du matin auraient été moins belles sans sa présence !

Chapitre 1

Systèmes dynamiques mesurables et l'hypothèse ergodique

"Chaque partie réfléchit le Tout"
Leibniz

1.1 L'hypothèse ergodique de Boltzmann

Du grec $\varepsilon\rho\gamma o\zeta$, "énergie", et $o\delta o\zeta$, "chemin".
L'hypothèse ergodique est née avec la théorie cinétique des gaz et la physique statistique dans la seconde moitié du XIXe siècle. Elle fut formulée initialement par Ludwig Boltzmann en 1871.

A cette époque, les travaux de Boltzmann sur la cinétique des gaz le mènent à penser que les particules qui constituent un gaz peuvent être considérées comme des copies les unes des autres ayant toutes le même comportement aléatoire. La vitesse moyenne des particules se calcule en sommant les vitesses de toutes les particules à un instant donné mais sous l'hypothèse ergodique de Boltzmann, on pourrait aussi calculer cette moyenne en mesurant les vitesses à différents instants d'**une seule** particule.

L'hypothèse d'ergodicité revient à dire que les deux méthodes de calcul sont équivalentes. Intuitivement, on dit qu'un système est "ergodique" si on ne peut pas "casser" un espace X en au moins deux ensembles de mesure non nulle sur lesquelles les particules ont un comportement différent. Ainsi sous l'hypothèse d'ergodicité :

Les moyennes temporelles coïncident avec les moyennes spatiales

C'est ce qu'on s'efforcera de montrer dans ce papier à la lumière des théorèmes ergodiques de Neumann, Birkhoff et Kingman. Puis dans un deuxième temps, on s'intéressera aux applications de ces théorèmes en théorie des probabilités, en théorie des systèmes dynamiques et en théorie du chaos.

1.2 Transformations mesurables

Dans tout ce qui suit on se donne X un espace d'étude, $\mathcal{A}$ une tribu sur X et μ une probabilité sur $\mathcal{A}$. On dit que $(X, \mathcal{A}, \mu)$ est un espace de probabilité ou un espace probabilisé. On ne s'intéresse qu'à des parties mesurables de X c'est-à-dire appartenant à $\mathcal{A}$.

On considère par la suite une application mesurable $T : X \longrightarrow X$ qui préserve la mesure, c'est-à-dire que $\forall E,\ T\mu(E) = \mu(E)$ où $T\mu$ désigne la mesure image de μ par T définie par : $\forall E,\ T\mu(E) = \mu(T^{-1}(E))$.

On dit que T est une transformation de l'espace X et on dit que $(X, \mathcal{A}, \mu$,T) est un système dynamique mesuré. La transformation T associe à un élément $x \in X$ l'élément $T(x)$ après une transformation, l'élément $T(T(x)) = T \circ T(x) = T^2(x)$ après deux transformations et ainsi de suite. À chaque x, l'application T associe donc la suite, que l'on appelle "trajectoire" $(x, T(x), T^2(x)...)$.

Rien ne stipule à première vue que T soit inversible. Néanmoins l'expression $T^{-1}(E)$ a toujours un sens $\forall E \in \mathcal{A}$: $T^{-1}(E) = \{x \in X,\ T(x) \in E\}$. Et on définit $\forall k \geq 1, T^{-k} = T^{-1} \circ T^{-1} \circ ... \circ T^{-1}$ (composée k fois).

Enfin, une observable, $f : X \to \mathbb{R}$ est une application mesurable qui à un élément $x \in X$ plus ou moins transformé par T associe une certaine valeur d'observation, par exemple, la température ou l'énergie d'un ensemble d'éléments donnés comme les particules d'un gaz.

Notre étude consiste à se demander s'il y a convergence (et si oui, **en quel sens !**) de la moyenne des observables. On notera donc $S_n(f)$, la fonction que l'on appelle aussi "moyenne de Birkhoff" telle que

$$S_n(f) = \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^i$$

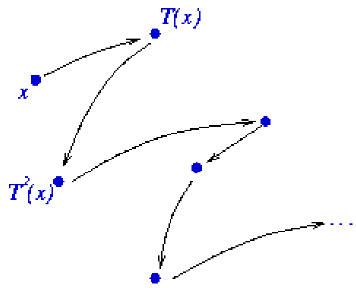

FIGURE 1.1 – Un point de l'espace et sa trajectoire

1.3 Propriétés ergodiques

Définition 1 (Ensemble T-*invariant*)**.** *Soit $T : X \longrightarrow X$ une transformation. Un ensemble E est dit T-invariant si $\mu(E \triangle T^{-1}(E)) = 0$ où $E \triangle T^{-1}(E)$ désigne la différence symétrique de E et $T^{-1}(E)$*

Alternativement, un ensemble E est dit T-invariant si et seulement si $\mathbb{1}_E = \mathbb{1}_{T^{-1}(E)}$ p.p

Proposition 1 (Tribu T-*invariante*)**.** *Les parties T-invariantes de X forment une tribu, notée $\mathcal{I} = \{E \subset X, \mu(E \triangle T^{-1}(E)) = 0\}$ et appelée la tribu invariante de T.*

Démonstration. Il suffit de vérifier les 3 conditions pour que $\mathcal{I}$ soit une tribu

. $T^{-1}(\varnothing) = \varnothing$. De ce fait, $\varnothing \triangle T^{-1}(\varnothing) = \varnothing \triangle \varnothing = \varnothing$ i.e. $\varnothing \in \mathcal{I}$ car $\mu(\varnothing) = 0$.

. Soit $E \in \mathcal{I}$, alors $E^c \triangle T^{-1}(E^c) = E^c \triangle (T^{-1}(E))^c = E \triangle (T^{-1}(E))$, i.e. $\mu(E^c \triangle T^{-1}(E^c)) = \mu(E \triangle T^{-1}(E)) = 0$. Donc $E^c \in \mathcal{I}$.

. Soit $(E_i)_{i \geq 1} \in \mathcal{I}\ \forall i \geq 1$, alors $\mu(\cup_{i\geq1} E_i \triangle T^{-1}(\cup_{i\geq1} E_i)) = \mu(\cup_{i\geq1}(E_i \triangle T^{-1}(E_i)) \leq \sum_{i\geq1} \mu((E_i \triangle T^{-1}(E_i)) = 0$ par σ-sous additivité. Donc $\cup_{i\geq1} E_i \in \mathcal{I}$.

Et donc $\mathcal{I}$ est une tribu. $\square$

On peut désormais introduire la notion de mesure ergodique :

Définition 2 (Mesure ergodique)**.** *On dit que le système mesuré $(X, \mathcal{A}, \mu, T)$ est ergodique si et seulement si les ensembles T-invariants sont de mesure nulle ou pleine.*

Remarque : Chez certains auteurs, il arrive qu'on puisse lire que *"T est ergodique"*, ou que *"μ est ergodique"* ou que *"T rend μ ergodique"*, ce qui ne pose aucun problème de compréhension apparent.

Proposition 2 (Équivalence de l'ergodicité, admis)**.** *Soit $(X, \mathcal{A}, \mu)$ un espace de probabilité et T qui préserve la mesure. Les propositions suivantes sont équivalentes :*

1) T rend μ ergodique

2) $\forall A \in \mathcal{A}$ tel que $\mu(A \triangle T^{-1}(A)) = 0$, $\mu(A) = 0$ ou $\mu(A) = 1$

3) $\forall A \in \mathcal{A}$ tel que $\mu(A) > 0$, $\mu(\cup_{i\geq0} T^{-n}(A)) = 1$

4) $\forall A \in \mathcal{A}, \forall B \in \mathcal{A}$ tel que $\mu(A) > 0$ et $\mu(B) > 0$, $\exists n > 0$ $\mu(T^{-n}(A \cap B)) > 0$

Proposition 3 (conséquences sur l'observable)**.** *Soit $(X, \mathcal{A}, \mu)$ un espace de probabilité et T qui préserve la mesure*
1) T rend μ ergodique
2) f est mesurable et pour presque tout $x \in X$, $f \circ T(x) = f(x) \Rightarrow f$ est constante presque partout

Démonstration. 1) $\Rightarrow$ 2) : Supposons T ergodique et f mesurable telle que $f \circ T = f$ presque partout.

Posons pour tout $k \in \mathbb{Z}$ et $n \in \mathbb{N}$, $X(k,n) := f^{-1}([k/2^n; (k+1)/2^n])$

Nous avons $T^{-1}(X(k,n)) \triangle X(k,n) \subset \{x \in X | f \circ T(x) \neq f(x)\}$ qui est par hypothèse négligeable donc de mesure nulle. Ainsi $\mu(T^{-1}(X(k,n)) \triangle X(k,n)) = 0$, et comme T rend μ ergodique on a $\mu(X(k,n)) = 0$ ou 1.

$\forall n$ fixé, $\cup_{k \in \mathbb{Z}} X(k,n) = X$ est une union disjointe, donc, par additivité de μ, $\exists! k_n$ tel que $\mu(X(k_n, n)) = 1$.

Soit $Y = \cap_{n \in \mathbb{N}} X(k_n, n)$, alors on a $\mu(X \setminus Y) = \mu(\cup_{n \in \mathbb{N}} X \setminus X(k_n, n)) \leq \sum_{n \in \mathbb{N}} \mu(X \setminus X(k_n, n)) = 0$.

Donc $\mu(Y) = 1$. De plus, f est constante sur Y. Donc f est constante presque partout.

2) $\Rightarrow$ 1) : Supposons que $E \in \mathcal{I}$, la tribu T-invariante, alors $\mathbb{1}_E$ est mesurable et $\mathbb{1}_E \circ T = \mathbb{1}_{T^{-1}(E)} = \mathbb{1}_E$ p.p.

Donc par hypothèse, $\mathbb{1}_E$ est constante presque partout.

Donc $\mathbb{1}_E = 0$ ou 1 p.p et $\mu(E) = \int_X \mathbb{1}_E \mathrm{d}\mu = 0$ ou 1, i.e. T rend μ ergodique. $\square$

Enfin, c'est surtout la propriété suivante qui nous intéressera car outil indispensable dans l'analyse d'une observable d'un système ergodique :

Proposition 4. *Soit $\mathcal{I}$ la tribu invariante de T*
f est $\mathcal{I}$-mesurable $\Leftrightarrow$ f est T-invariante : $f \circ T = f$ presque partout.

Démonstration. Supposons f T-invariante, alors $\forall E \in \mathcal{A}$, $(f \circ T)^{-1}(E) = T^{-1}(f^{-1}(E)) = f^{-1}(E)$

Et donc $\forall E \in \mathcal{A}$, $f^{-1}(E) \in \mathcal{I}$ i.e. f est $\mathcal{I}$-mesurable.

Réciproquement, montrons d'abord le résultat pour une fonction indicatrice. Soit donc $f = \mathbb{1}_C$ $\mathcal{I}$-mesurable avec $C \in \mathcal{A}$. Si elle est $\mathcal{I}$-mesurable, alors $\mathbb{1}_C = \mathbb{1}_{T^{-1}(C)}$ p.p. Or $\mathbb{1}_{T^{-1}(C)} = \mathbb{1}_C \circ T$, ce qui prouve le résultat lorsque f est une fonction indicatrice.

Le résultat s'en suit par linéarité pour toute fonction étagée positive de la forme $f = \sum_{i=0}^n a_i \mathbb{1}_{A_i}$.

L'existence d'une suite croissante de fonctions étagées positives pour toute fonction mesurable positive garantit le résultat pour toute fonction mesurable positive.

Enfin, en écrivant $f = f^+ - f^-$, où f^+ et f^- sont toutes deux des fonctions mesurables positives, on affirme le résultat pour toute fonction mesurable. $\square$

Chapitre 2

Le théorème ergodique de von Neumann

"para que se aprenda en el mundo que los que tienen bosque y agua (...) pueden ser sencillos y oscuros"
Testamento de Otoño, Pablo Neruda

2.1 Un Théorème de convergence en norme

Von Neumann, un des piliers de la théorie ergodique ainsi qu'un des pères fondateurs de la mécanique quantique, démontre son Théorème ergodique "en norme" en 1931. Bien que version plus faible que la version de Birkhoff, c'est un résultat digne d'intérêt sur la convergence L^2 de $S_n(f)$

Théorème 1 (Théorème ergodique de von Neumann)**.** *Soit* $(X, \mathcal{A}, \mu)$ *un espace mesuré,* $T : X \longrightarrow X$*, une transformation mesurable qui préserve* μ *et* $f \in \mathrm{L}^2(X)$*. Alors*

$$S_n(f) \xrightarrow[n\to\infty]{\mathrm{L}^2} Pf$$

où Pf *est le projecteur orthogonal de* f *sur le sous espace* $\mathrm{Inv} = \{f \in \mathrm{L}^2 \mid f \circ T = f\}$

On commence tout d'abord par montrer le lemme suivant :

Lemme 1 (Invariance par U et U^*)**.** *Soit* H *un espace de Hilbert,* $U : H \to H$*, une application linéaire de norme inférieure ou égale à 1, alors tout élément* g *de* H *est* U*-invariant si et seulement si il est* U^**-invariant, où* U^* *désigne l'adjoint de* U*, i.e. l'unique application* $U^* : H \to H$ *telle que* $\langle U^* f, g\rangle = \langle f, Ug\rangle \quad (\forall f, g \in H)$*.*

Démonstration. Supposons g U-invariant : $Ug = g$.
On écrit que :
$\|g - U^*g\|^2 = \|g\|^2 + \|U^*g\|^2 - 2\langle g, U^*g\rangle \leq 2\|g\|^2 - 2\langle g, U^*g\rangle = 2\langle g - Ug, g\rangle = 0.$
Ce qui montre bien que g est U^*-invariant. La réciproque s'obtient immédiatement en remplaçant U par U^* $\square$

On rappelle également la définition de convergences topologiques forte et faible.

Définition 3 (convergences topologiques)**.** *Soit* H *un espace de Hilbert muni du produit scalaire* $\langle \cdot, \cdot\rangle$*. On peut munir* H *de deux topologies différentes*
· *la topologie forte : une suite* f_n *d'éléments de* H *converge fortement vers* $f \in H$ *si* $\|f_n - f\| \xrightarrow[n\to\infty]{} 0$
· *la topologie faible : une suite* f_n *d'éléments de* H *converge faiblement vers* $f \in H$ *si* $\langle f_n, g\rangle \xrightarrow[n\to\infty]{} \langle f, g\rangle \ \forall g \in H$*.*

Remarque : Soit $(X, \mathcal{A}, \mu)$, un espace mesuré. Prenons pour Hilbert H l'espace $\mathrm{L}^2(X)$. A partir d'une transformation $T : X \to X$ qui préserve la mesure μ, on définit U, une application linéaire $U : H \to H$ en posant $Uf = f \circ T$ de norme 1. En effet,

$$\|Uf\|^2 = \int (f \circ T).(f \circ T)\ \mathrm{d}\mu = \int f.f\ \mathrm{dT}\mu = \int f.f\ \mathrm{d}\mu = \|f\|^2 \quad \forall f \in H$$

On peut alors maintenant montrer le théorème ergodique en norme de von Neumann

Démonstration. On commence tout d'abord par décomposer f sous la forme $f = f^I + f^{I\perp}$ où $f^I \in \mathrm{Inv}$ et $f^{I\perp} \in \mathrm{Inv}^\perp$ et on rappelle que dans notre cas $U : \mathrm{L}^2 \to \mathrm{L}^2 : f \longmapsto f \circ T$.

Si f appartient à Inv, par récurrence immédiate $f \circ T^n = f \quad \forall n \geq 1$, et donc $S_n(f) = \frac{1}{n}\sum_{i=0}^{n-1} f = f$, i.e. $S_n(f)$ appartient à Inv.

Il suffit donc de montrer que $S_n(f)$ tend vers 0 pour $f \in \mathrm{Inv}^\perp$. Notons que

$$\|S_n(f)\|^2 = \langle f, \frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))\rangle$$

Il suffit donc de montrer que $\forall f \in \mathrm{Inv}^\perp$, $\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))$ converge faiblement vers 0.
$\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))$ appartient à $\mathrm{Inv}^\perp$ par stabilité linéaire du sous-espace.
On montre que $\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))$ tend faiblement vers 0 si on arrive à montrer qu'elle tend vers une application U-invariante (auquel cas, cette limite appartiendrait à la fois à $\mathrm{Inv}^\perp$ car le sous-espace est fermé et à Inv car U-invariante et à ce titre serait l'élément neutre de L^2 : la fonction nulle !).

On va plutôt montrer que les valeurs d'adhérences de $\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))$ sont U^*-invariantes, ce qui revient au même d'après le lemme 1. On écrit :

$$(I - U^*)\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f)) = \frac{1}{n}(I - U^*)\sum_{k=0}^{n-1} U^{*i} S_n(f) = \frac{1}{n}(I - U^{*n})(S_n(f))$$

Si bien qu'en choisissant une norme sous-multiplicative

$$\|(I - U^*)\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))\| \leq \frac{1}{n}\|I - U^{*n}\|.\|S_n(f)\| \leq \frac{2}{n}\|f\| \xrightarrow[n\to\infty]{} 0$$

Ce qui prouve que les valeurs d'adhérence sont bien U^*-invariantes et à ce titre appartiennent à Inv : $\frac{1}{n}\sum_{i=0}^{n-1} U^{*i}(S_n(f))$ tend donc faiblement vers 0, ce qui achève la preuve. $\square$

FIGURE 2.1 – John von Neumann

2.2 Implication probabiliste

Plaçons nous dans un contexte probabilisé : $(X, \mathcal{A}, \mu)$ est un espace de probabilité, $\mathcal{B}$ un ensemble d'éléments de $\mathcal{A}$, f mesurable est une variable aléatoire. Informellement, on appelle Espérance conditionnelle de f sachant $\mathcal{B}$ que l'on note $\mathbb{E}(f|\mathcal{B})$, la fonction mesurable qui se rapproche le plus de f aux vues des observations $\mathcal{B}$. L'espérance conditionnelle est définie pour toute variable aléatoire intégrable. Néanmoins, en se plaçant sur l'Hilbert L^2, on se ramène à une définition plus commode car l'espérance conditionnelle est alors la projection orthogonale sur le sous espace (qu'on admettra fermé) $\mathrm{L}^2(\mathcal{B}) = \mathrm{L}^2(\sigma(\mathcal{B}))$ pour le produit scalaire $\langle f, g\rangle = \int fg \,\mathrm{d}\mu$.

Le théorème ergodique de von Neumann devient alors :

Théorème 2 (Théorème ergodique en probabilité de von Neumann)**.** *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace de probabilité,* $T : \Omega \longrightarrow \Omega$*, une transformation mesurable qui préserve* $\mathbb{P}$ *et* $X \in \mathrm{L}^2(\Omega)$*. Alors*

$$\frac{1}{n}\sum_{i=0}^{n-1} X \circ T^i \xrightarrow[n\to\infty]{\mathrm{L}^2} \mathbb{E}(X|\mathcal{I})$$

où $\mathbb{E}(X|\mathcal{I})$ *est l'espérance conditionnelle de* X *sachant* $\mathcal{I}$*, la tribu invariante de* T*.*

Chapitre 3

Le théorème ergodique de Birkhoff, le théorème sous-additif de Kingman

"Every problem becomes very childish when once it is explained to you.
Here is an unexplained one."
The adventure of the Dancing Man, Arthur Conan Doyle

Dans ce chapitre on veut démontrer le célèbre théorème ergodique de Birkhoff, démontré en 1931. Version beaucoup plus forte que la version de von Neumann.

3.1 Le théorème de Birkhoff

Théorème 3 (Théorème de Birkhoff). *Soit f une fonction intégrable sur X, alors $\exists g \in \mathrm{L}^1(X)$ telle que*

$$S_n(f) \xrightarrow[n\to\infty]{p.p} g \quad \& \quad \int f = \int g$$

Ce théorème est lourd de conséquences. Il affirme que les sommes de Birkhoff convergent d'une manière plus puissante que ce que n'affirme von Neumann : $S_n(f)(x)$ converge presque partout vers une certaine valeur indépendante de x.

De ce théorème existent plusieurs preuves mais nous nous contenterons de n'en développer qu'une, toute récente. En réalité, nous démontrerons plutôt le théorème sous-additif de John Kingman, mathématicien anglais émérite toujours vivant, dont Birkhoff n'est qu'un corollaire.

3.2 Le théorème sous-additif de Kingman

Théorème 4 (Théorème sous-additif de Kingman). *Soient $f_n : X \longrightarrow \overline{\mathbb{R}}$, une suite de fonctions intégrables telle que f_1^+ est intégrable et vérifiant l'inéquation fonctionnelle*

$$f_{n+m} \leq f_m + f_n \circ T^m \quad \forall m, n \geq 1$$

Alors $\exists g : X \longrightarrow \overline{\mathbb{R}}$ telle que g^+ est intégrable et

$$\frac{f_n}{n} \xrightarrow[n\to\infty]{p.p} g \quad \& \quad \int g = \lim \frac{1}{n} \int f_n = \inf \frac{1}{n} \int f_n \in [-\infty, +\infty[$$

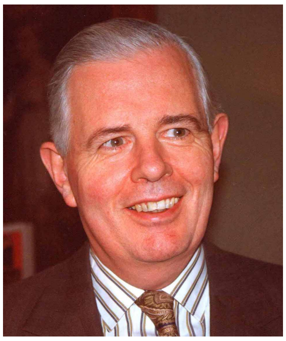

FIGURE 3.1 – John Kingman

$\widehat{f_n} = \sum_{k=0}^{n-1} f \circ T^i$ est clairement additive et à ce titre est un cas particulier de l'inéquation fonctionnelle :

$$\widehat{f_{n+m}} = \sum_{k=0}^{n+m-1} f \circ T^i = \sum_{k=0}^{m-1} f \circ T^i + \sum_{k=m}^{n+m-1} f \circ T^i = \sum_{k=0}^{m-1} f \circ T^i + \sum_{k=0}^{n-1} f \circ T^{m+i} = \sum_{k=0}^{m-1} f \circ T^i + (\sum_{k=0}^{n-1} f \circ T^i) \circ T^m = \widehat{f_m} + \widehat{f_n} \circ T^m$$

La preuve "rééditée" date de 2009 par Arthur Avila (médaille Fields 2014), remarquable de méthodisme et de simplicité

Démonstration.
Soit $f_n : X \to \mathbb{R}$ une suite de fonctions sous-additives telles que f_1^+ est dans L^1 (et donc de fait f_n^+).
$\int f_n$ est alors une suite de réels sous-additive et c'est un résultat classique que (voir preuve en annexe) :

$$\frac{1}{n} \int f_n \xrightarrow[n \to \infty]{} \mathrm{L} := \inf \frac{1}{n} \int f_n$$

On introduit $f^{\nearrow}, f^{\searrow} : X \to [-\infty, +\infty[$ mesurables définies par

$$f^{\nearrow} = \limsup_{n \to +\infty} \frac{f_n}{n} \qquad f^{\searrow} = \liminf_{n \to +\infty} \frac{f_n}{n}$$

Notre but est de montrer que

$$\int f^{\searrow} \geq L \geq \int f^{\nearrow}$$

Ce qui prouvera que $f^{\nearrow} = f^{\searrow}$ presque partout. On appellera alors g la limite commune de $f^{\nearrow}$ et $f^{\searrow}$.

Remarquons tout d'abord avant de commencer que $f^{\searrow}$ et $f^{\nearrow}$ sont T-invariantes. En effet par sous-additivité et passage à la limite

$$f^{\searrow} = \liminf_{n \to +\infty} \frac{f_n}{n} \leq \liminf_{n \to +\infty} \frac{f_1(x) + f_{n-1}(Tx)}{n} = f^{\searrow}(Tx)$$

De ce fait, $T^{-1}(\{f^{\searrow} \geq a\}) \subset \{f^{\searrow} \geq a\}$ $\forall a \in \overline{\mathbb{R}}$.

Or T préserve μ : de ce fait $\mu(T^{-1}(\{f^{\searrow} \geq a\})) = \mu(\{f^{\searrow} \geq a\})$ $\forall a \in \overline{\mathbb{R}}$.

L'inclusion et l'égalité des dimensions justifient donc $T^{-1}(\{f^{\searrow} \geq a\}) = \{f^{\searrow} \geq a\}$ $\forall a \in \overline{\mathbb{R}}$. Et ainsi, $f^{\searrow}$ est $\mathcal{I}$-mesurable. D'après la proposition 4, $f^{\searrow}$ est T-invariante i.e $f^{\searrow} \circ T = f^{\searrow}$ p.p. De même pour $f^{\nearrow}$ par un raisonnement analogue.

Commençons maintenant la preuve

Partie 1 : $\int f^{\searrow} \geq L$
Nous nous contenterons de montrer que $\int f^{\searrow} = L$

Commençons par simplifier l'hypothèse ("se donner ϵ-d'espace") en montrant le résultat dans le cas où

$$\exists C \in \mathbb{R} \text{ tel que } f_n \leq -Cn \quad \forall n$$

D'après le lemme de Fatou, $f^{\searrow}$ est intégrable avec $\int f^{\searrow} \leq L$.

Soit $\epsilon > 0$ fixé et considérons la suite croissante au sens de l'inclusion

$$E_k = \{x \in X,\ \exists j \in \{1, ..., k\} \text{ tel que } \frac{f_j(x)}{j} < f^{\searrow}(x) + \epsilon\}$$

Nous avons $\cup_{k\geq 1} E_k = X$. Soit la fonction intégrable

$$\psi_k = \begin{cases} f^{\searrow} + \epsilon & x \in\ E_k \\ f_1 & x \in\ E_k^c \end{cases}$$

Remarquez que $\psi_k \geq f^{\searrow} + \epsilon$

Nous allons montrer que pour presque tout x et $\forall n \geq k$

$$f_n(x) \leq \sum_{i=0}^{n-k-1} \psi_k(T^i x) + \sum_{i=n-k}^{n-1} (\psi_k \vee f_1)(T^i x)$$

Fixons $x \in X$. On introduit deux suites d'entiers $(m_j)_{j\in\mathbb{N}}$ et $(n_j)_{j\in\mathbb{N}^*}$ telles que

$$m_0 \leq n_1 < m_1 \leq n_2 < m_2 \leq ...$$

définies de la manière suivante :

Soit $m_0 = 0$. n_j est le plus petit des entiers supérieur ou égal à m_{j-1} tel que $T^{n_j} x \in E_k$ (un tel entier existe car $\cup_{k\geq 1} E_k = X$)

Par définition de E_k, on peut choisir m_j tel que $1 \leq m_j - n_j \leq k$ et

$$\frac{f_{m_j - n_j}(T^{n_j} x)}{m_j - n_j} < f^{\searrow}(x) + \epsilon$$

Soit $n \geq k$, on définit enfin ℓ, le plus grand entier tel que $m_\ell \leq n$. En utilisant la sous-additivité on obtient

$$f_n(x) \leq \sum_{i \in\ \cup_{j=0}^{\ell-1}[m_j, n_{j+1}[\cup[m_\ell, n[} f(T^i x) + \sum_{j=1}^{\ell} f_{m_j - n_j}(T^{n_j} x)$$

D'une part, $\forall i \in \cup_{j=0}^{\ell-1}[m_j, n_{j+1}[\cup[m_\ell, \inf(n, n_{\ell+1})[\ \ f(T^i x) = \psi_k(T^i x)$ (car $T^i x \in E_k^c$)

D'autre part, en utilisant la dernière inégalité, la T-invariance de $f^{\searrow}$ et le fait que $\psi_k \geq f^{\searrow} + \epsilon$, on a que

$$f_{m_j - n_j}(T^{n_j} x) \leq \sum_{i \in [n_j, m_j[} f^{\searrow}(T^i x) + \epsilon \leq \sum_{i \in [n_j, m_j[} \psi_k(T^i x)$$

On obtient alors le résultat en remarquant que $n_{\ell+1} > n - k$:

$$f_n(x) \leq \sum_{i=0}^{\inf(n_{\ell+1}, n-1)} \psi_k(T^i x) + \sum_{i=n_{\ell+1}}^{n-1} f_1(T^i x)$$

En intégrant le résultat, on obtient que $\int f_n \leq (n-k) \int \psi_k + k \int \psi_k \vee f_1$

En divisant par n et en passant à la limite $n \to \infty$ on obtient que $L \leq \int \psi_k$

Puis faisant tendre $k \to \infty$, on obtient que $L \leq \int f^{\searrow} + \epsilon$. Ceci $\forall \epsilon > 0$, ce qui conclut le résultat dans le cas où "$\exists C \in \mathbb{R}$ tel que $f_n \leq -Cn\ \forall n$"

Considérons maintenant le cas général. Pour $C \in \mathbb{R}$, définissons

$$f_n^{(C)} = f_n \vee (-Cn)$$

Alors la suite $f_n^{(C)}$ est sous-additive et

$$f^{\searrow(C)} := \liminf_{n\to+\infty} \frac{f_n^{(C)}}{n} = f^{\searrow} \vee (-C) \quad \& \quad f^{\nearrow(C)} := \limsup_{n\to+\infty} \frac{f_n^{(C)}}{n} = f^{\nearrow} \vee (-C)$$

De ce fait, en utilisant le théorème de convergence monotone et le fait que $\int f^{\searrow} = L$, on obtient

$$\int f^{\searrow} = \inf_C \int f^{\searrow(C)} = \inf_C \inf_n \frac{1}{n} \int f_n^{(C)} = \inf_n \inf_C \frac{1}{n} \int f_n^{(C)} = \inf_n \frac{1}{n} \int f_n = L$$

Partie 2 : $\int f^{\nearrow} \leq L$

On commence par démontrer deux courts lemmes

Lemme 2. *Soit $g : X \to \mathbb{R}$ une fonction intégrable. Alors* $\dfrac{g \circ T^n}{n} \to 0$ *p.p*

Démonstration. Il suffit de montrer que $\forall \epsilon > 0$, $\mu(\limsup A_n) = 0$ où $A_n = \{x | \; |g \circ T^n(x)| \geq \epsilon n\}$

Pour cela, il suffit de montrer que :$\sum_{n=1}^{\infty} \mu(\{|g \circ T^n| \geq \epsilon n\}) < \infty$ en vertu du lemme de Borel-Cantelli

$$\begin{aligned}\sum_{n=1}^{\infty} \mu(\{|g \circ T^n| \geq \epsilon n\}) &= \sum_{n=1}^{\infty} \mu(\{|g| \geq \epsilon n\}) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mu(\{k \leq \epsilon^{-1}|g| < k+1\}) \\ &= \sum_{k=1}^{\infty} k\mu(\{k \leq \epsilon^{-1}|g| < k+1\}) \leq \int_{\{|g| \geq \epsilon\}} \epsilon^{-1}|g| \, \mathrm{d}\mu < \infty\end{aligned}$$

□

Lemme 3.

$$\forall k \geq 1, \quad \limsup_{n\to\infty} \frac{f_{kn}}{n} = k \limsup_{n\to\infty} \frac{f_n}{n} \quad p.p$$

Démonstration. Le sens "$\leq$" est évident par sous-additivité.

Pour l'autre sens, soit $k \geq 1$ fixé, pour tout entier n on écrit $n = km_n + r_n$ avec $1 \leq r_n \leq k$. Par sous-additivité :

$$f_n \leq f_{km_n} + h \circ T^{km_n} \quad \text{avec} \quad h = f_1^+ \vee ... \vee f_k^+$$

On a $\frac{m_n}{n} \xrightarrow[n\to\infty]{} \frac{1}{k}$. En particulier $m_n \to \infty$ lorsque $n \to \infty$

Du fait que $h \in \mathrm{L}^1$, le lemme 2 donne que $h \circ T^{km_n}/n \to 0$ p.p. On écrit alors que :

$$\frac{f_{km_n}}{n} + \frac{h \circ T^{km_n}}{n} = \frac{f_{km_n}}{m_n}\frac{m_n}{n} + \frac{h \circ T^{km_n}}{n} \geq \frac{f_n}{n}$$

D'où le résultat en passant à la limite.

□

On peut donc maintenant prouver la deuxième partie du théorème en raisonnant comme précédemment. Supposons donc que $\exists C \in \mathbb{R}$ tel que $f_n \leq -Cn \;\; \forall n$

Fixons k et appelons F_n la n-ième somme de Birkhoff de $-f_k$ pour la transformation T, i.e $F_n = -\sum_{j=0}^{n-1} f_k \circ T^{jk}$. Alors la suite F_n est additive. De plus $F_1 = -f_k \leq Ck$ donc $F_1 \in \mathrm{L}^1$

On introduit $F^{\searrow} := \liminf_{n\to\infty} \frac{F_n}{n}$. La première partie de notre preuve donne que $\int F^{\searrow} \geq \lim \frac{1}{n} \int F_n$

Par T-invariance $\frac{1}{n} \int F_n = -\int f_k$

D'autre part, en utilisant le lemme 3

$$-F^{\searrow} = \limsup \frac{1}{n} \sum_{j=0}^{n-1} f_k \circ T^{jk} \geq \limsup \frac{f_{kn}}{n} = k \limsup \frac{f_n}{n} = kf^{\nearrow}$$

De ce fait, $\int f^{\nearrow} \leq \frac{-1}{k} \int F^{\searrow} \leq \frac{1}{k} \int f_k$

Ceci étant vrai $\forall k$, nous venons de prouver que $\int f^{\nearrow} \leq L$ sous l'hypothèse $\exists C \in \mathbb{R}^+$ tel que $f_n \leq -Cn \;\; \forall n$.

Traitons maintenant le cas général. Soit $f_n^{(C)}$ comme défini précédemment. Nous avons déjà montré que les fonctions $f^{\searrow(C)}$ et $f^{\nearrow(C)}$ comme définies précédemment ont la même intégrale. De ce fait $f^{\searrow(C)} = f^{\nearrow(C)}$ p.p

Du fait que $f^{\nearrow(C)} \to f^{\nearrow}$ et $f^{\searrow(C)} \to f^{\searrow}$ lorsque $C \to +\infty$, il vient que $f^{\searrow} = f^{\nearrow}$ p.p Ce qui conclut la preuve du théorème sous-additif de Kingman □

3.3 Avec l'hypothèse ergodique

Une fois les théorèmes de Neumann, Birkhoff et Kingman énoncés, on est en droit de se demander quelles conséquences en découleraient sous l'hypothèse d'ergodicité. En effet, depuis le début, nous utilisons simplement le fait que T préserve la mesure. Rappelons que le théorème de Neumann nous affirmait que $S_n(f)$ converge en norme L^2 vers $\mathbb{E}[f|\mathcal{I}]$, projection de f sur l'ensemble des fonctions T-invariantes. On va d'abord montrer que :

Proposition 5. *Sous l'hypothèse d'ergodicité, lorsque $\mu(X)$ est fini non nul*

$$\forall f \in \mathrm{L}^1, \qquad \mathbb{E}[f|\mathcal{I}] = \frac{1}{\mu(X)} \int_X f \ \mathrm{d}\mu \ \text{ p.p.}$$

Démonstration. C'est évident au vu de la Proposition 3.

Par définition, $\mathbb{E}[f|\mathcal{I}]$ est T-invariante. D'après la Proposition 3, elle est donc constante presque partout. Soit C cette constante. On peut alors écrire que $\mathbb{E}[f|\mathcal{I}] = C\mathbb{1}_E$ où $\mu(E) = \mu(X)$.

Ainsi, $\int_X \mathbb{E}[f|\mathcal{I}](x)\mathrm{d}\mu(x) = C\int_X \mathbb{1}_E(x)\mathrm{d}\mu(x) = C\mu(E) = C\mu(X) = \int_X f(x)\mathrm{d}\mu(x)$.
C'est ce que l'on voulait en rappelant que $\mathbb{E}[f|\mathcal{I}] = C$ p.p. □

Replaçons nous dans un espace de probabilité. Sous sa forme ergodique le théorème de Neumann prend alors une forme beaucoup plus intéressante car on voit alors bien apparaître la convergence L^2 de la moyenne temporelle vers la moyenne spatiale. Or une telle affirmation reste toujours vrai si l'on suppose f dans L^1, c'est-à-dire sous les hypothèses du théorème de Birkhoff. En effet, on va vu apparaître la convergence p.p. des moyennes temporelles vers une fonction g dont on a montré qu'elle est T-invariante. Sous l'hypothèse d'ergodicité elle est donc constante p.p. mais puisque son intégrale est égale à l'intégrale de f alors (raisonnement analogue de la preuve de la Proposition 5) g est égale p.p. à la moyenne spatiale de f. On obtient finalement le théorème ergodique de Birkhoff :

Théorème 5 (Théorème ergodique de Birkhoff)**.** *Soit f une fonction intégrable sur X avec T qui rende μ ergodique, alors*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^i = \int_X f \ \mathrm{d}\mu$$

On vient donc finalement de prouver notre affirmation initiale : à savoir que dans un système ergodique, "*Les moyennes temporelles coïncident avec les moyennes spatiales*".

Concluons ce chapitre avec une petite application de ce théorème : Supposons que le système probabilisé $(X, \mathcal{A}, \mu, \mathrm{T})$ soit ergodique et appliquons le théorème de Birkhoff à $f = \mathbb{1}_A$ où A est un ensemble mesurable de X, on obtient que

$$\text{p.p. } x \in X \quad \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A \circ T^i = \frac{1}{n}\sharp\{k \in \{0,1,...,n-1\} | T^k(x) \in A\} \xrightarrow[n\to\infty]{} \int_X \mathbb{1}_A \ \mathrm{d}\mu = \mu(A)$$

En d'autres termes, dans un système dynamique ergodique, le temps moyen passé par une trajectoire dans un ensemble A est donc le même pour toute trajectoire, il est proportionnel à la mesure de l'ensemble A.

Figure 3.2 – George David Birkhoff

Chapitre 4

Applications à la théorie des probabilités

"On entend par choses naturelles toutes celles qui mues continûment
par un principe qui leur est intime, arrivent à une certaine fin.
De chacun de ces principes, ne sort pas pour chaque espèce de chose un résultat identique,
de même qu'il n'en sort pas un résultat arbitraire ;
mais toujours le même principe tend au même résultat (...)
Quand c'est toujours ou du moins le plus ordinairement qu'une chose arrive,
ce n'est ni par accident, ni par hasard ;
or, dans la nature, les choses se produisent éternellement de la même façon"
La Physique, Aristote

On a déjà vu au chapitre 2 une première conséquence probabiliste de la théorie ergodique. En effet, on a vu que le théorème de convergence en norme de von Neumann impliquait qu'il y a convergence de $\frac{1}{n}\sum_{i=0}^{n-1} X \circ T^i$ vers l'espérance conditionnelle de X sachant $\mathcal{I}$, la tribu invariante de T.

On va voir dans ce chapitre qu'il existe une autre implication probabiliste de la théorie ergodique. Et pas des moindres. Nous allons ainsi montrer que le théorème ergodique de Birkhoff entraîne la célèbre "Loi forte des grands nombres" dûe en 1929 à Kolmogorov qu'on rappelle ici

Théorème 6 (Loi forte des grands nombres)**.** *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires iid (indépendantes et identiquement distribuées) et intégrables (i.e. $\mathbb{E}[|X_0|] < +\infty$) alors*

$$\frac{1}{n}\sum_{i=0}^{n-1} X_i \xrightarrow[n\to\infty]{p.p} \mathbb{E}[X_0]$$

4.1 Ergodicité du décalage

L'un des premiers réflexes lorsque on se donne un système dynamique mesuré est de vérifier si la transformation à laquelle on a affaire est ergodique. On va maintenant introduire une transformation d'un espace produit qu'on appellera "shift de Bernoulli" et montrer que celle-ci est ergodique.

Définition 4 (Shift de Bernoulli)**.** *Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé. On définit le shift de Bernoulli sur l'espace produit $(X^{\mathbb{N}}, \mathcal{A}^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$ comme étant la transformation T telle que $T(\{x_i\}) = \{x_{i+1}\}$*

Proposition 6. *Le shift de Bernoulli est une transformation mesurable, qui préserve la mesure, et ergodique.*

Démonstration. Le caractère mesurable et préservant la mesure du shift sont évidents. Montrons donc que T est ergodique. Soit f une fonction intégrable T-invariante. On aura montrer que T est ergodique si f est constante presque partout (Proposition 3).

Pour tout $\epsilon > 0$, on peut trouver $g \in \mathrm{L}^1$ qui ne dépend que d'un nombre fini de coordonnées et telle que $\|g - f\|_1 < \epsilon/4$. (preuve en annexe)

g n'est pas invariante mais elle satisfait l'estimation :
$\|g - g \circ T^n\|_1 \leq \|g - f\|_1 + \|f - f \circ T^n\|_1 + \|f \circ T^n - g \circ T^n\|_1 \leq \epsilon/2$

$$\begin{aligned}
\text{Or } \ \|g - g\circ T^n\|_1 &= \int |g(x_0,...,x_{n-1}) - g(x_n,...,x_{2n-1})| \ \mathrm{d}\mu(x_0)...\mathrm{d}\mu(x_{2n-1}) \\
&= \int |g(x_0,...,x_{n-1}) - g(y_0,...,y_{n-1})| \ \mathrm{d}\mu(x_0)...\mathrm{d}\mu(x_{n-1})\mathrm{d}\mu(y_0)...\mathrm{d}\mu(y_{n-1}) \\
&= \int |g(x) - g(y)|\mathrm{d}\mu^{\otimes\mathbb{N}}(x)\mathrm{d}\mu^{\otimes\mathbb{N}}(y)
\end{aligned}$$

Et ainsi $\int |g(x) - g(y)|\mathrm{d}\mu^{\otimes\mathbb{N}}(x)\mathrm{d}\mu^{\otimes\mathbb{N}}(y) \leq \epsilon/2$.

$$\begin{aligned}
\text{On a alors } \ \int |f(x) - f(y)|\mathrm{d}\mu^{\otimes\mathbb{N}}(x)\mathrm{d}\mu^{\otimes\mathbb{N}}(y) &\leq \int |f(x) - g(x)|\mathrm{d}\mu^{\otimes\mathbb{N}}(x)\mathrm{d}\mu^{\otimes\mathbb{N}}(y) + \int |g(x) - g(y)|\mathrm{d}\mu^{\otimes\mathbb{N}}(x)\mathrm{d}\mu^{\otimes\mathbb{N}}(y) \\
&\quad + \int |g(y) - f(y)|\mathrm{d}\mu^{\otimes\mathbb{N}}(x)\mathrm{d}\mu^{\otimes\mathbb{N}}(y) \\
&\leq \epsilon/2 + 2\|f - g\|_1 \leq \epsilon
\end{aligned}$$

Ceci montre que $f(x) = f(y)$ pour μ-presque tout $(x, y) \in \Omega^{\mathbb{N}} \times \Omega^{\mathbb{N}}$. D'après le théorème de Fubini, on peut alors trouver $y_0 \in \Omega^{\mathbb{N}}$ tel que l'ensemble des $x \in \Omega^{\mathbb{N}}$ satisfaisant $f(x) = f(y_0)$ soit de mesure 1.

f est alors constante presque partout, ce qui achève la preuve avec la Proposition 3 : T est ergodique. $\square$

4.2 La loi forte des grands nombres par Birkhoff

On veut ici montrer comment la loi des grands nombres se déduit du théorème ergodique. Notons $\mathbb{R}^{\mathbb{N}}$, l'espace des suites réelles muni de la tribu des boréliens associée. Soit (X_i) une suite de variables aléatoires intégrables définies sur un espace probabilisé $(\Omega, \mathcal{B}, \mathbb{P})$.

Soit l'application $\psi : \Omega \to \mathbb{R}^{\mathbb{N}} : \omega \mapsto \{X_i(\omega)\}_{i\in\mathbb{N}}$. Et on pose $\mu = \psi\mathbb{P}$: mesure image de $\mathbb{P}$ par ψ.

Dire que les X_i sont des variables aléatoires identiquement distribuées revient à dire que μ coïncide avec $(\psi\mathbb{P}_{X_0})^{\otimes\mathbb{N}}$ où P_{X_0} est la loi de X_0

Soit l'application $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R} : \{x_i(\omega)\}_{i\in\mathbb{N}} \mapsto x_0$ et T le décalage de Bernoulli sur l'espace $\mathbb{R}^{\mathbb{N}}$.

Remarquons que $f \circ T^k \circ \psi = X_k$ et donc que les moyennes ergodiques de f et les moyennes des X_i sont reliés par

$$(\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k)\circ\psi = \frac{1}{n}\sum_{k=0}^{n-1} X_k$$

Appliquons le théorème ergodique de Birkhoff à la fonction f, au décalage T défini sur $\mathbb{R}^{\mathbb{N}}$ et à la probabilité μ. On a alors la convergence presque sûre des moyennes de f relativement à la mesure μ, ce qui équivaut à la convergence presque sûre des moyennes des X_i relativement à $\mathbb{P}$.

Enfin, en remarquant que $\mathbb{E}[X_0] = \int X_0 \mathrm{d}\mathbb{P} = \int f\, d\mu$, on a démontré ce que l'on voulait :

Corollaire 1 (Loi forte des grands nombres)**.** *Soient $(\Omega, \mathcal{B}, \mathbb{P})$ un espace probabilisé, $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires définies sur cet espace, intégrables, identiquement distribuées et indépendantes. Alors pour presque tout $\omega \in \Omega$:*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_i(\omega) \xrightarrow[n\to\infty]{} \mathbb{E}[X_0]$$

Chapitre 5

Introduction à l'étude d'un système dynamique : Etude des exposants de Lyapunov

"Le battement des ailes d'un papillon au Brésil peut-il provoquer un ouragan au Texas ?"
Titre d'une conférence de Lorenz en 1973

5.1 De la théorie ergodique à la théorie du chaos

Le but essentiel de l'étude d'un système dynamique est de réussir à prédire en fonction des caractéristiques du système (conditions initiales, transformation auquel il est soumis, etc...) l'évolution de celui-ci au cours du temps et plus spécifiquement à long terme. Dans un cadre simplifié à l'extrême, on peut ainsi voir apparaître des solutions régulières et des attracteurs : prenez par exemple le cas du système différentiel modélisant l'évolution d'une population soumise à l'équation logistique avec condition initiale $\begin{cases} s'(t) = s(t)(1-s(t)) - ps(t) \\ s(0) = s_0 \end{cases}$

Néanmoins, dans la pratique, l'étude de systèmes soumis à une dynamique est loin d'être évident. Cela étant surtout dû à deux causes : un trop grand nombre de données paramétrisant le système (pensez par exemple à la météorologie : pression, humidité, hydrométrie, force du vent, etc...) mais c'est surtout le fait que celles ci sont des paramètres de dimension infinie (ce sont toutes des fonctions de l'espace). Il n'est donc pas possible de modéliser un système en rendant exactement compte des phénomènes qui agissent sur lui. Et quand bien même on s'accordait le droit de simplifier le modèle en travaillant sur des données de dimension finie, la détermination explicite des solutions en fonction du temps est vouée à l'échec.

En 1962, Edward Lorenz, météorologue et mathématicien de formation, se permet de caricaturer l'équation de Navier-Stokes, de la simplifier à l'extrême, de faire comme si l'atmosphère ne dépendait que de trois paramètres, alors qu'il en faudrait une quantité colossale. Dans son atmosphère atrophiée réduite à ses trois coordonnées, Lorenz peut faire tourner son ordinateur et calculer les solutions numériques qui sont censées décrire le mouvement. C'est alors qu'il constate que la moindre modification dans son atmosphère simpliste (ajouter par exemple 0.0000001 à l'une des trois coordonnées initiales) entraîne dans le mouvement atmosphérique un changement considérable. C'est le phénomène de la "dépendance sensible aux conditions initiales", le paradigme de la théorie du chaos.

L'attracteur de Lorenz représente une trajectoire de l'équation de Lorenz simplifiée, dans l'espace tridimensionnel. Ces courbes tournent comme des folles, tantôt par la gauche, tantôt par la droite, et il semble impossible de prévoir si un tour à droite sera suivi par un autre tour à droite ou à gauche. Et pourtant, pour une condition initiale donnée, pour une atmosphère donnée, il y a un futur bien défini. En revanche deux points proches dans l'espace tridimensionnel, si proches qu'on ne les distingue peut-être pas sur la figure, définissent des trajectoires qui commenceront par être proches mais qui peuvent finir par se séparer de manière importante. Ainsi, si on ne connaît un point qu'avec une certaine incertitude, aussi petite soit-elle, la prédiction de l'avenir devient illusoire. Les mathématiciens de l'époque se sont alors penchés sur la question de savoir s'il était possible d'étudier la manière avec laquelle l'incertitude grandit, il s'avéra alors que la théorie ergodique était riche d'enseignement. Ainsi fut introduit l'indice de Lyapunov supérieur également appelé "taux de croissance exponentielle des normes".

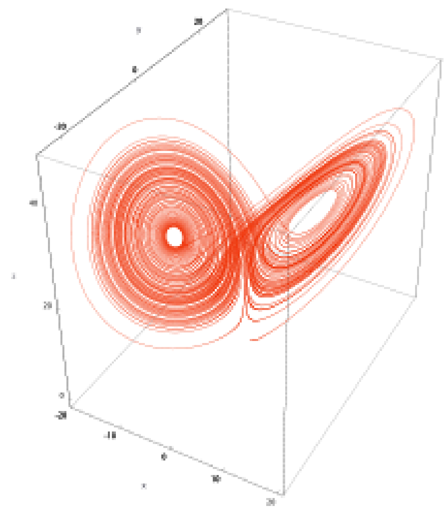

FIGURE 5.1 – L'attracteur de Lorenz

5.2 L'exposant de Lyapunov supérieur

Supposons que l'on se donne un point p de l'espace X, dont le mouvement g est connu. Supposons maintenant qu'on fasse subir à ce point une déviation ξ au temps $t = 0$, la formule de Taylor au premier ordre dit que l'on peut raisonnablement affirmer que l'on trouvera le point p au temps $t = 1$ dans un voisinage de $g(p)$ d'écartement $\mathrm{d}g(p).\xi$ (i.e. dans la boule centrée en $g(p)$ et de rayon $\mathrm{d}g(p).\xi$). A l'instant $t = 2$, on trouvera alors p dans un voisinage de $g^2(p)$ d'écartement $\mathrm{d}g(g(p)).\mathrm{d}g(p).\xi$ et ainsi de suite. Le but est alors d'essayer de prédire l'écart avec lequel p s'écarte de sa trajectoire initiale au cours du temps.

Formalisons un peu cela : Soit $(X, \mathcal{A}, \mu)$ un espace probabilisé et $T : X \to X$ une transformation ergodique. Dans le cas de notre exemple, T est entièrement déterminé par g et sa dérivée.

Soit $A : X \to \mathrm{SL}(2, \mathbb{R})$ (où $\mathrm{SL}(2, \mathbb{R})$ désigne l'ensemble des matrices de $M_2(\mathbb{R})$ de déterminant égal à 1 muni de la norme subordonnée à la norme euclidienne de $\mathbb{R}^2$ i.e. $\|A\| = \sup_{v\neq 0} \frac{\|Av\|}{\|v\|}$), une fonction mesurable satisfaisant la condition d'intégrabilité :

$$\int \log \|A\| \mathrm{d}\mu < \infty$$

De plus, on pose :

$$N(A) = \log(\frac{\|A\| + \|A\|^{-1}}{2}) \quad \text{pour} A \in \mathrm{SL}(2, \mathbb{R})$$

Et on rappelle la rotation d'angle θ :

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{pour } \theta \in \mathbb{R}$$

On définit pour $x \in X$ et $n \in \mathbb{N}$,

$$A^n(x) = A(T^{n-1}x)...A(x)$$

La fonction A est appelée un "cocycle linéaire". Dans ces conditions, il existe un nombre $\lambda^+(A) \geq 0$, appelé "l'exposant de Lyapunov supérieur" défini par

$$\lambda^+(A) = \lim_{n\to+\infty} \frac{1}{n} \log \|A^n(x)\| \quad \text{pour } \mu\text{−presque tout } x \in X$$

C'est justement le théorème sous-additif de Kingman qui nous assure l'existence de l'exposant de Lyapunov supérieur. En effet, la norme étant sous-multiplicative, la suite $u_n(x) = \log \|A^n(x)\|$ vérifie l'inéquation fonctionnelle pour tout x :

$$u_{n+m}(x) = \log \|A^{n+m}(x)\| = \log \| \prod_{i=0}^{n+m-1} A(T^i x)\| = \log \| \prod_{i=0}^{m-1} A(T^i x) \prod_{i=m}^{n+m-1} A(T^i x)\| \leq \log \| \prod_{i=0}^{m-1} A(T^i x)\|\| \prod_{i=m}^{n+m-1} A(T^i x)\|$$

$$= \log \| \prod_{i=0}^{m-1} A(T^i x)\| + \log \| \prod_{i=m}^{n+m-1} A(T^i x)\| = \log \| \prod_{i=0}^{m-1} A(T^i x)\| + \log \| \prod_{i=0}^{n-1} A(T^i \circ T^m x)\| = u_n(x) + u_m \circ T^m(x)$$

Et ainsi, sous l'hypothèse que $u_1(x) = \log \|A(x)\|$ soit intégrable alors la suite u_n/n converge presque partout. On aimerait néanmoins en savoir un peu plus sur $\lambda^+(A)$. Notamment, depuis qu'il a été introduit, les mathématiciens tentent tant bien que mal de savoir si $\lambda^+(A) > 0$. Or une telle affirmation n'a jamais été prouvée et pourtant elle prouverait bien le caractère exponentiel de la divergence de vecteur, but majeur de la théorie du chaos.

Plutôt que de travailler sur la définition même de $\lambda^+(A)$, Avilla et Bochi ont été désireux de faire une sorte de moyenne de celui-ci. Ainsi, pour $\theta \in \mathbb{R}$, on définit un cocycle AR_θ par $(AR_\theta)(x) = A(x)R_\theta$. Clairement, $\theta \to \lambda^+(AR_\theta)$ est une fonction mesurable et alors

Théorème 7 (Moyenne de Lyapunov). *Si T, μ et A sont définis comme précédemment alors*

$$\frac{1}{2\pi}\int_0^{2\pi} \lambda^+(AR_\theta)\mathrm{d}\theta = \int_X \log\left(\frac{\|A(x)\| + \|A(x)\|^{-1}}{2}\right)\mathrm{d}\mu(x)$$

Pour démontrer ce théorème, on devra admettre le suivant :

Théorème 8. *Soient $A_1, ..., A_n \in \mathrm{SL}(2, \mathbb{R})$. Alors*

$$\frac{1}{2\pi}\int_0^{2\pi} N(A_n R_\theta ... A_1 R_\theta)\ \mathrm{d}\theta = \sum_{j=1}^{n} N(A_j)$$

Démonstration. Remarquez tout d'abord que $\log \|A\| - \log 2 < N(A) \leq \log \|A\|$.

En effet $\|A\| > 0$ car $A \in \mathrm{SL}(2, \mathbb{R})$ (et donc $A \neq 0$), ceci prouve le premier sens de l'inégalité. Pour le second sens il faut remarquer que, puisque $A \in \mathrm{SL}(2, \mathbb{R})$ alors $\rho(A) \geq 1$ où $\rho(A)$ désigne le rayon spectral de A, (i.e. le maximum des valeurs propres de A). Or, il est évident que $\rho(A) \leq \|A\|$. Ainsi, $\|A\| \geq 1$ pour $A \in \mathrm{SL}(2, \mathbb{R})$. Or, on vérifie aisément que $x + x^{-1} \leq 2x$ pour $x \geq 1$ d'où l'autre sens de l'inégalité.

Utilisant le fait que $N(A) \leq \log \|A\| < \log 2 + N(A)$, il vient par le Théorème 8 que

$$\sum_{j=0}^{n-1} N(A(T^j(x))) \leq \frac{1}{2\pi}\int_0^{2\pi} \log \|(AR_\theta)^n(x)\|\mathrm{d}\theta \leq \log 2 + \sum_{j=0}^{n-1} N(A(T^j(x)))$$

En divisant par n, le théorème de Birkhoff donne que

$$\lim_{n\to\infty} \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{n}\log \|(AR_\theta)^n(x)\|\mathrm{d}\theta = \int_X N(A(x))\mathrm{d}\mu(x) \quad \text{pour } \mu\text{−presque tout } x \in X$$

Pour conclure la preuve, il ne reste plus qu'à vérifier que le théorème de convergence dominée s'applique. On a

$$\frac{1}{n}\log \|(AR_\theta)^n(x)\| \leq \frac{1}{n}\sum_{j=0}^{n-1} \log \|AR_\theta(T^j x)\| = f_n(x)$$

$(f_n)_{n\geq 1}$ est une suite de moyennes de Birkhoff de la fonction $\log \|A\| \in \mathrm{L}^1$. En particulier, $f_n(x)$ est bornée pour presque tout x et le théorème de convergence dominée s'applique donc. Ce qui conclut la preuve. ☐

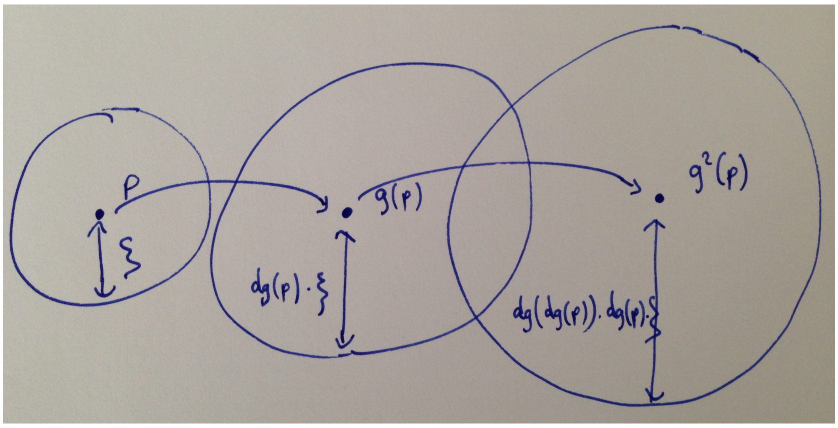

FIGURE 5.2 – Le phénomène de divergence exponentielle de vecteur

Conclusion

Dans ce papier, on s'est intéressé à définir les éléments essentiels de la théorie ergodique pour un système dynamique donné. Puis, on s'est intéressé à trois théorèmes ergodiques en particulier : ceux de Neumann, Birkhoff et Kingman, le but premier étant de statuer sur la convergence des moyennes temporelles $S_n(f)$. On s'est ainsi rendu compte grâce au théorème de Kingman que celles-ci convergent presque sûrement vers $\mathbb{E}[f|\mathcal{I}]$. Dans le cas où le système est ergodique, on a donc vu apparaître la convergence des moyennes temporelles vers la moyenne spatiale de l'observable f en question.

Dans un deuxième temps, on a vu en quoi la théorie ergodique pouvait être riche de conséquences en théorie des probabilités. Celle-ci a également beaucoup d'applications dans l'étude des systèmes dynamiques qui n'ont pas pu être traités, faute de temps (notamment les applications mélangeantes, les conjugaisons, etc...). On a cependant brièvement expliqué en quoi la théorie ergodique permettait de faire le lien entre l'étude des systèmes dynamiques et la théorie du chaos, discipline toute récente, dont Poincaré est un des piliers, destinée à étudier le comportement chaotique de certains systèmes particuliers. On a ainsi mis en évidence que les théorèmes de Birkhoff et Kingman assuraient l'existence d'exposants de Lyapunov sur des ensembles de mesure pleine et pointé du doigt les autres résultats pas encore découverts à ce jour qui pourront faire avancer la discipline.

En effet, et c'est sur cela que l'on concluera, la théorie ergodique est une belle théorie mais surtout une théorie riche et jeune dont beaucoup de résultats restent encore à percer. Le lecteur averti pourra s'il le désire consulter la longue liste d'"Open problems in Dynamical systems and Ergodic theory" à l'adresse https ://www.math.iupui.edu/ mmisiure/open/. En espérant que cela puisse vous inspirer autant que j'ai pris plaisir à rédiger ce premier document sur cette belle théorie ergodique.

Annexes

A - Sur les suites sous-additives

Résultat : Soit $(u_n)_{n\geq 0}$ une suite réelle vérifiant $u_{n+m} \leq u_m + u_n$ $\forall (m,n) \in \mathbb{N}\times\mathbb{N}$ alors $(\dfrac{u_n}{n})_{n\geq 1}$ converge vers $\ell := \inf_{n\geq 1} \dfrac{u_n}{n} \in \mathbb{R}\cup\{-\infty\}$

Preuve : De l'hypothèse, on déduit immédiatement que, pour tout entier n, $u_{2n} \leq 2u_n$, puis par une récurrence simple, que $u_{mn} \leq mu_n$ pour m et n dans $\mathbb{N}^*$.

Commençons par le cas où ℓ est réel. Pour tout $\epsilon > 0$, on peut trouver m tel que $\ell \leq \frac{u_m}{m} \leq \ell + \epsilon$. Si n est un entier quelconque, on en fait la division euclidienne par m et on obtient $n = qm + r$ avec $0 \leq r < m$. On a alors $u_n \leq u_{mq} + u_r \leq qu_m + u_r$ et donc, en divisant par n,

$$\ell \leq \frac{u_n}{n} \leq \frac{qu_m}{n} + \frac{u_r}{n} \leq \frac{(\ell+\epsilon)qm}{n} + \frac{M}{n} \leq (\ell+\epsilon)(1-\frac{r}{n}) + \frac{M}{n}$$

où $M = \max_{0\leq k<m} u_k$. Puisque $0 \leq r < m$, $\frac{r}{n}$ tend vers 0 lorsque n tend vers $+\infty$ et $(\ell+\epsilon)(1-\frac{r}{n}) + \frac{M}{n}$ tend vers $\ell + \epsilon$. On a donc pour n assez grand,

$$\ell \leq \frac{u_n}{n} \leq +2\epsilon$$

ce qui prouve que $\lim_{n\to+\infty} \frac{u_n}{n} = \ell$. Le cas où $\ell = -\infty$ se traite de même. Pour tout $A < 0$, il existe un m tel que $\frac{u_m}{m} \leq A$. On obtient comme précédemment, pour n assez grand :

$$\frac{u_n}{n} \leq A(1-\frac{n}{r}) + \frac{M}{n} \leq \frac{A}{2}$$

B - Sur l'approximation d'une fonction L^1

Résultat : Toute fonction intégrable peut être approchée de manière explicite par des fonctions qui ne dépendent que d'un nombre fini de coordonnées. Énoncé plus précisément : soient $(X, \mathcal{A}, \mu)$ un espace mesuré et $(\mathcal{B}_i)_{i\geq 1}$ une suite de σ-algèbres incluses dans $\mathcal{A}$, croissante pour l'inclusion : $\mathcal{B}_i \subset \mathcal{B}_j$ si $i < j$. Notons $\mathcal{B}$ la σ-algèbre engendrée par les $\mathcal{B}_i$. Alors pour tout $f \in \mathrm{L}^2$:

$$\mathbb{E}[f|\mathcal{B}_n] \xrightarrow[n\to\infty]{\mathrm{L}^2} \mathbb{E}[f|\mathcal{B}]$$

Preuve : Afin d'alléger les notations, on suppose que $\mathcal{B} = \mathcal{A}$, le cas général se déduisant en remplaçant f par $\mathbb{E}[f|\mathcal{B}]$. L'espérance conditionnelle de f relativement à la tribu $\mathcal{B}_n$ est la projection orthogonale sur $\mathrm{L}^2(X, \mathcal{B}_n)$. Notons π_∞ la limite de ces projections. Il faut montrer que cette limite est égale à l'identité. Comme les fonctions étagées sont denses dans les fonctions L^2, il suffit de vérifier que la famille d'ensembles suivants coïncide avec $\mathcal{B}$: $\mathcal{B}' = \{B \in \mathcal{B} \mid \pi_\infty \mathbb{1}_B = \mathbb{1}_B\}$

Cette classe contient l'union des $\mathcal{B}_i$, qui est une algèbre de parties de X. Vérifions que $\mathcal{B}'$ est une classe monotone. Elle est bien invariante par passage au complémentaire, montrons qu'elle est invariante par union dénombrable croissante. Soit B_n une suite croissante d'éléments de $\mathcal{B}'$; la suite $\mathbb{1}_{B_n}$ converge vers $\mathbb{1}_{\cup B_n}$ en norme L^2 par le théorème de convergence dominée. On en déduit : $\pi_\infty(\mathbb{1}_{\cup B_n}) = \lim \pi_\infty(\mathbb{1}_{B_n}) = \lim \mathbb{1}_{B_n} = \mathbb{1}_{\cup B_n}$. D'après le lemme de la classe monotone $\mathcal{B}'$ contient la σ-algèbre engendrée par les $\mathcal{B}_n$. Elle coïncide avec $\mathcal{B}$ et π_∞ est égale à l'identité.

Pour illustrer ce résultat, considérons un espace probabilisé $(X, \mathcal{A}, \mu)$ et plaçons-nous sur l'espace produit $(X^{\mathbb{N}}, \mathcal{A}^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$. Soit $p_n : X^{\mathbb{N}} \to X^n$ la projection sur les n premières coordonnées et posons $\mathcal{B}_n = p_n^{-1}(\mathcal{A}^{\otimes\mathbb{N}})$. Les σ-algèbres $\mathcal{B}_n$ engendrent $\mathcal{A}^{\otimes\mathbb{N}}$ par définition du produit tensoriel. Nous avons donc pour tout $f \in \mathrm{L}^2(X^{\mathbb{N}})$:

$$\mathbb{E}[f|\mathcal{B}_n] \xrightarrow[n\to\infty]{\mathrm{L}^2} f$$

La fonction f est ainsi approchée par des fonctions qui ne dépendent que d'un nombre fini de coordonnées. Ce résultat est encore vrai avec une fonction intégrable, car l'espace L^2 est dense dans L^1.

Bibliographie

◊ Yves Coudène. Théorie ergodique et systèmes dynamiques. CNRS éditions.

◊ A. Avilla and J.Bochi. On the subadditive ergodic theorem. Unpublished. 2009

◊ A.Avilla and J.Bochi. A formula with some applications to the theory of Lyapunov exponents. 2002

◊ Antoine Brunel, "Ergodique Théorie", Encyclopædia Universalis.

◊ Emmanuel Rousseaux Voisin Nathalie. Théorie ergodique et applications. 2007.