Théorème ergodique de Birkhoff

Mon développement

Soit $(X, \mathcal{B}, \mu, T)$ un système dynamique probabilisé. Pour tous $f \in L^1(X, \mathcal{B}, \mu)$, $x \in X$, $N \in \mathbb{N}^*$, on appelle moyenne de Birkhoff de f en x la quantité :

$$S_N f(x) = \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n(x).$$

Lemme. Soit $f \in L^1(X, \mathcal{B}, \mu)$. On note $\overline{f}$ (resp. $\underline{f}$) la limite supérieure (resp. inférieure) de $S_N f$ quand $N \to +\infty$. Les fonctions $\overline{f}$ et $\underline{f}$ sont mesurables et T-invariantes.

$\overline{f}$ et $\underline{f}$ sont mesurables car les $S_N f$ le sont. De plus, pour tout $N \in \mathbb{N}^*$, on a

$$S_N f \circ T = \frac{N+1}{N} S_{N+1} f - \frac{1}{N} f \ .$$

Donc, en passant à la limite inférieure et supérieure, on obtient : $\overline{f} \circ T = \overline{f}$ et $\underline{f} \circ T = \underline{f}$. □

Théorème. Soit $(X, \mathcal{B}, \mu, T)$ un système dynamique probabilisé et soit $f \in L^1(X, \mathcal{B}, \mu)$.
- La suite $(S_N f)_{N \in \mathbb{N}^*}$ converge μ-presque partout de limite notée F,
- $F \in L^1(X, \mathcal{B}, \mu)$,
- F est T-invariante,
- $\int_X F \, d\mu = \int_X f \, d\mu$.

• Soit $B \in \mathcal{B}$. On note $f = \mathbf{1}_B$. On a clairement $0 \leq \underline{f} \leq \overline{f} \leq 1$. Soit $\varepsilon > 0$.
On note pour tout $x \in X$, $\tau(x) = \inf\{N \in \mathbb{N}^* \,|\, S_N f(x) \geq \overline{f}(x) - \varepsilon\}$.
On note pour tout $p \in \mathbb{N}^*$, $A_p = \{x \in X \,|\, \tau(x) \leq p\}$. La suite $(A_p^c)_{p \in \mathbb{N}^*}$ est décroissante et on a $\mu\left(\bigcap_{p \in \mathbb{N}^*} A_p^c\right) = 0$ car τ est fini presque partout. Il existe donc $p \in \mathbb{N}^*$ tel que $\mu(A_p^c) \leq \varepsilon$.
On note $B' = B \cup A_p^c$ et $g = \mathbf{1}_{B'}$.
Enfin, on note pour tout $x \in X$, $\overline{\tau}(x) = \tau(x) \, \mathbf{1}_{A_p}(x) + \mathbf{1}_{A_p^c}(x)$.

Pour presque tout $x \in X$, on a :

$$\sum_{n=0}^{\overline{\tau}(x)-1} g(T^n(x)) \geq \overline{\tau}(x)(\overline{f}(x) - \varepsilon) \ . \tag{1}$$

En effet, si $x \notin A_p$, alors $\overline{\tau}(x) = 1$ et $g(x) = 1 \geq \overline{f}(x) - \varepsilon$.
D'autre part, si $x \in A_p$, alors $\overline{\tau}(x) = \tau(x)$ donc

$$\begin{aligned} \sum_{n=0}^{\tau(x)-1} g(T^n(x)) &\geq \sum_{n=0}^{\tau(x)-1} f(T^n(x)) \quad (\text{car } f \leq g) \\ &\geq \tau(x)(\overline{f}(x) - \varepsilon) \quad (\text{par définition de } \tau(x)). \end{aligned}$$

Dans les deux cas, on a bien la relation (1).

Soit $x \in X$. On note $\nu_0 = 0$, et pour tout $i \in \mathbb{N}$, $\nu_{i+1} = \nu_i + \overline{\tau}(T^{\nu_i}(x))$.
Soit un entier $N \geq p$. Comme $(\nu_i)_{i\in\mathbb{N}}$ est strictement croissante et $\overline{\tau}$ est borné par p, il existe un unique $k \in \mathbb{N}$ tel que $\nu_k \leq N-1 < \nu_{k+1}$ et on a $\overline{\tau}(T^{\nu_k}(x)) \leq p$. Donc

$$\nu_k = \nu_{k+1} - \overline{\tau}(T^{\nu_k}(x)) \geq N - p \ . \tag{2}$$

On a donc pour presque tout $x \in X$:

$$\begin{aligned}
\sum_{n=0}^{N-1} g(T^n(x)) &= \sum_{i=0}^{k-1}\sum_{n=\nu_i}^{\nu_{i+1}-1} g(T^n(x)) + \sum_{n=\nu_k}^{N-1} g(T^n(x)) \\
&\geq \sum_{i=0}^{k-1}\sum_{n=0}^{\overline{\tau}(T^{\nu_i}(x))-1} g(T^n(T^{\nu_i}x)) + 0 \quad (\text{définition de } \nu_{i+1} \text{ et } g \geq 0) \\
&\geq \sum_{i=0}^{k-1} \overline{\tau}(T^{\nu_i}(x))(\overline{f}(T^{\nu_i}(x)) - \varepsilon) \quad (\text{d'après (1)}) \\
&= \sum_{i=0}^{k-1} (\nu_{i+1} - \nu_i)(\overline{f}(x) - \varepsilon) \quad (\text{définition de } \nu_{i+1} \text{ et } \overline{f}\circ T = \overline{f}) \\
&= \nu_k(\overline{f}(x) - \varepsilon) \\
&\geq (N-p)(\overline{f}(x) - \varepsilon) \quad (\text{d'après (2)}).
\end{aligned}$$

En divisant par N puis en intégrant sur X, on obtient, pour tout $N \geq p$:

$$\int_X S_N g \, d\mu \geq \frac{N-p}{N}\left(\int_X \overline{f}\, d\mu - \varepsilon\right) \ . \tag{3}$$

Or, la mesure μ est T-invariante, donc

$$\mu(B') = \int_X g \, d\mu = \int_X S_N g \, d\mu \ . \tag{4}$$

En utilisant (3) et (4) et en faisant tendre $N \to +\infty$, on obtient : $\mu(B') \geq \int_X \overline{f}\, d\mu - \varepsilon$. Or $\mu(B') \leq \mu(B) + \varepsilon$, d'où $\mu(B) \geq \int_X \overline{f}\, d\mu - 2\varepsilon$. Ceci étant vrai pour tout $\varepsilon > 0$, on obtient :

$$\mu(B) \geq \int_X \overline{f}\, d\mu \ .$$

On peut appliquer le même raisonnement à B^c. D'une part, $\mu(B^c) = 1 - \mu(B)$, et d'autre part, $\overline{\mathbf{1}_{B^c}} = \overline{1 - \mathbf{1}_B} = 1 - \underline{\mathbf{1}_B}$. Cela nous donne donc $1 - \mu(B) \geq \int_X 1 - \underline{f}\, d\mu$, soit :

$$\mu(B) \leq \int_X \underline{f}\, d\mu \ .$$

Finalement,

$$\mu(B) \leq \int_X \underline{f} d\mu \leq \int_X \overline{f} d\mu \leq \mu(B) \ .$$

Ces inégalités sont donc des égalités. Comme $\underline{f} \leq \overline{f}$, on en déduit que $\underline{f} = \overline{f}$ μ-presque partout, c'est-à-dire que $(S_N f)_{N\in\mathbb{N}^*}$ converge presque partout, de limite $F = \overline{f} = \underline{f}$. De plus, $0 \leq F \leq 1$ donc $F \in L^1(X, \mathcal{B}, \mu)$ et

$$\int_X F \, d\mu = \mu(B) = \int_X f \, d\mu \ .$$

• Soit f une fonction étagée. Il existe $r \in \mathbb{N}$, $a_1, \ldots, a_r \in \mathbb{C}$ et $A_1, \ldots, A_r \in \mathcal{B}$ disjoints deux à deux tels que $f = \sum_{i=1}^r a_i \mathbf{1}_{A_i}$.
D'après ce qui précède, pour tout $i \in \{1, \ldots, r\}$, $(S_N \mathbf{1}_{A_i})_{N\in\mathbb{N}^*}$ converge presque partout, de limite notée F_i, appartenant à $L^1(X, \mathcal{B}, \mu)$, telle que $\int_X F_i \, d\mu = \mu(A_i)$.
Donc la suite $(S_N f = \sum_{i=1}^r a_i S_N \mathbf{1}_{A_i})_{N\in\mathbb{N}^*}$ converge presque partout, de limite $F = \sum_{i=1}^r a_i F_i$ et

$$\int_X F \, d\mu = \sum_{i=1}^r a_i \mu(A_i) = \int_X f \, d\mu \ .$$

• Soit $f \in L^1(X, \mathcal{B}, \mu)$ positive. Soit $\varepsilon > 0$. Il existe deux fonctions étagées positives f_1, f_2 telles que $0 \leq f_1 \leq f \leq f_2$ et $\int_X f_2 - f_1 \, d\mu \leq \varepsilon$. D'après le point précédent, on a d'une part : $\overline{f_1} = \underline{f_1}$ presque partout, $\int_X \underline{f_1} \, d\mu = \int_X f_1 \, d\mu$, et d'autre part : $\overline{f_2} = \underline{f_2}$ presque partout, $\int_X \overline{f_2} \, d\mu = \int_X f_2 \, d\mu$.
En utilisant le fait que $\underline{f_1} \leq \underline{f} \leq \overline{f} \leq \overline{f_2}$, on en déduit que

$$\int_X f_1 \, d\mu \leq \int_X \underline{f} \, d\mu \leq \int_X \overline{f} \, d\mu \leq \int_X f_2 \, d\mu$$

donc

$$0 \leq \int_X \overline{f} - \underline{f} \, d\mu \leq \int_X f_2 - f_1 \, d\mu \leq \varepsilon \ .$$

Ceci étant vrai pour tout $\varepsilon > 0$, on en déduit que $\int_X \overline{f} - \underline{f} d\mu = 0$ et que $\overline{f} = \underline{f}$ presque partout.
Donc $(S_N f)_{N\in\mathbb{N}^*}$ converge presque partout, de limite notée F, mesurable positive. De plus, avec les mêmes notations que ci-dessus, pour tout $\varepsilon > 0$, on a :

$$\int_X \underline{f} \, d\mu \leq \liminf_{N\to+\infty} \int_X S_N f \, d\mu = \int_X f \, d\mu \leq \int_X f_2 \, d\mu \leq \int_X f_1 \, d\mu + \varepsilon \leq \int_X \underline{f} \, d\mu + \varepsilon$$

en utilisant le lemme de Fatou et $\int_X S_N f \, d\mu = \int_X f \, d\mu$ pour tout $N \in \mathbb{N}^*$. Cela prouve que $F = \underline{f} \in L^1(X, \mathcal{B}, \mu)$ car $\int_X f \, d\mu < +\infty$, et que $\int_X F \, d\mu = \int_X f \, d\mu$. On a également $F \circ T = F$ d'après le lemme.

• Soit $f \in L^1(X, \mathcal{B}, \mu)$. On note f_+ (resp. f_-) sa partie positive (resp. négative). D'après le point précédent, les suites $(S_N f_+)_{N\in\mathbb{N}^*}$ et $(S_N f_-)_{N\in\mathbb{N}^*}$ convergent presque partout, de limites notées F_+ et F_-, appartenant à $L^1(X, \mathcal{B}, \mu)$, et on a $F_+ \circ T = F_+$, $F_- \circ T = F_-$, $\int_X F_+ d\mu = \int_X f_+ d\mu$ et $\int_X F_- d\mu = \int_X f_- d\mu$.
Donc $(S_N f = S_N f_+ - S_N f_-)_{N\in\mathbb{N}^*}$ converge presque partout, de limite $F_+ - F_-$, notée F. De plus, $F \in L^1(X, \mathcal{B}, \mu)$, $F \circ T = F$ et

$$\int_X F \, d\mu = \int_X F_+ \, d\mu - \int_X F_- \, d\mu = \int_X f_+ \, d\mu - \int_X f_- \, d\mu = \int_X f \, d\mu \ .$$

□

Corollaire. Avec les notations précédentes, si μ est T-ergodique, alors F est constante presque partout, égale à $\int_X f \, d\mu$.

F est T-invariante et μ est T-ergodique. D'après la caractérisation des mesures ergodiques, F est constante presque partout. Comme μ est une mesure de probabilité, on a $F = \int_X F\,d\mu$ p.p. Donc $F = \int_X f\,d\mu$ p.p. d'après le théorème ergodique de Birkhoff. □

Références

J'ai utilisé [CLFM97, p. 179]. On peut aussi consulter [PY98, pp. 102, 111], [Wal00], [Dud02, p. 268], [Shi84, p. 381].

Leçons correspondantes

J'utilise ce développement pour les leçons 223, 226, 235. On peut également l'utiliser pour les leçons 234, 238, 241, 250.

Remarques

- Attention, les notations ne sont pas les mêmes que celles de [CLFM97].
- Dans ce développement, on a identifié un élément de L^1 avec un de ses représentants comme on le fait habituellement.
- Pour ne pas dépasser le temps imparti, il ne faut pas détailler le passage des fonctions indicatrices aux fonctions étagées, ni celui des fonctions positives aux fonctions intégrables, ceux-ci étant de simples conséquences de la linéarité.
- Il ne faut pas hésiter à être pédagogue, car vu de loin la preuve peut paraître obscure et lourde en notations, et on a parfois du mal à voir ce qui se passe.
- Ce développement nécessite d'être à l'aise avec les bases de la théorie des systèmes dynamiques, car il utilise beaucoup de propriétés et de techniques classiques. Par ailleurs, il ne faut pas oublier de mettre tout ce qu'il faut dans le plan de la leçon, puisqu'on est totalement hors du programme de l'agrégation. En particulier, il faut donner les définitions de système dynamique probabilisé, fonction T-invariante, mesure T-ergodique et donner les caractérisations des mesures invariantes et ergodiques.
- Il ne faut pas non plus oublier de donner des applications à ce théorème, pour justifier tout le mal qu'on se donne. Si on a investi dans ce domaine, on peut parler de fractions continues, exemple non trivial mais intéressant de système dynamique. Le problème est que tout cela prend de la place dans un plan limité à trois pages.
- Des exemples plus ou moins simples de systèmes dynamiques sont traités dans [FGN07] par exemple, et peuvent trouver leur place dans les leçons sur les suites sans nécessairement évoquer la notion de système dynamique.
- En utilisant le décalage de Bernoulli, ce théorème permet de retrouver la loi des grands nombres. On peut donc voir ce théorème comme une généralisation de la loi forte des grands nombres, et donc le présenter dans la leçon 250.

Questions possibles

1. Démontrer la caractérisation des mesures invariantes, utilisée à plusieurs reprises dans le développement : μ est T-invariante si et seulement si $\forall n \in \mathbb{N}$, $\forall f \in L^1(X, \mathcal{B}, \mu)$, on a : $\int_X f \circ T^n\,d\mu = \int_X f\,d\mu$.
2. Donner une idée de la démonstration de la caractérisation des mesures ergodiques, utilisée dans le corollaire : μ est T-ergodique si et seulement si pour tout $f \in L^1(X, \mathcal{B}, \mu)$ telle que $f \circ T = f$ μ-p.p., f est constante μ-presque partout.

3. Donner une application du théorème ergodique de Birkhoff.
4. Déterminer une condition nécessaire et suffisante pour que la mesure de Lebesgue soit ergodique pour la rotation d'angle α sur le cercle $\mathbb{S}^1$. Déterminer, à l'aide du théorème, la fréquence d'apparition du 1 dans la suite des premiers chiffres des puissances de 2.

Références

[CLFM97] Antoine Chambert-Loir, Stéfane Fermigier et Vincent Maillot : *Exercices de mathématiques pour l'agrégation : Analyse 1.* Masson, 1997.

[Dud02] Richard M. Dudley : *Real analysis and probability.* Cambridge University Press, 2002.

[FGN07] Serge Francinou, Hervé Gianella et Serge Nicolas : *Exercices de mathématiques, Oraux X-ENS, Analyse 1.* Cassini, 2007.

[PY98] Mark Pollicott et Michiko Yuri : *Dynamical systems and ergodic theory.* Cambridge University Press, 1998.

[Shi84] Albert Shiryayev : *Probability.* Springer, 1984.

[Wal00] Peter Walters : *An introduction to ergodic theory.* Springer, 2000.