LAURENT GUILLOPÉ

Mesures invariantes sur des ensembles autosimilaires

Séminaire de Théorie spectrale et géométrie, tome 6 (1987-1988), p. 49-60
<http://www.numdam.org/item?id=TSG_1987-1988__6__49_0>

Séminaire de théorie spectrale et géométrie
CHAMBÉRY-GRENOBLE
1987–1988 (49–60)

MESURES INVARIANTES SUR DES ENSEMBLES AUTOSIMILAIRES

par *Laurent GUILLOPÉ*

Résumé : A toute famille S finie d'applications contractantes est associé un ensemble limite autosimilaire. Il existe une mesure invariante pour certains processus de Markov construits à partir de la famille S et des théorèmes récents d'approximation de cette mesure *à la Birkhoff* sont à la base de techniques de compression d'images.

Dans le développement actuel de l'infographie, le stockage des images est un problème aigu : par exemple, une seule image du satellite Spot correspond couramment à environ 30 MO de (= 31457280) caractères. Il y a ainsi une forte demande de méthodes efficaces de compression/décompression d'images : l'augmentation continuelle de la puissance des microprocesseurs permet d'envisager la mise en œuvre rapide de telles méthodes *a priori* complexes (au besoin avec l'aide de microprocesseurs spécialement conçus pour ces tâches).

Le cadre de l'image est un ensemble fini de pixels (*picture element*) à colorier *i.e.* (si on restreint aux images en noir et blanc) à noircir éventuellement. On le munit de la topologie discrète, de telle manière que la partie noire (ainsi que la partie blanche) apparaît comme un compact du cadre de l'image. Le problème de compression est donc le suivant :

Étant donné un espace compact E, coder (en termes économiques) un compact quelconque K de E.

Les techniques de compression de [3] (article qui fut à l'origine de cet exposé) sont basées sur l'idée suivante : muni de la distance de Hausdorff, l'espace $\mathcal{K}(E)$ des parties compactes de E est un espace compact. Une famille S d'applications contractantes induit sur $\mathcal{K}(E)$ une application contractante $\mathcal{S}(S)$; pourquoi ne pas voir tout compact comme point fixe $\mathcal{L}(S)$ de $\mathcal{S}(S)$, pour une famille S convenable? Etant donné un compact A, on peut construire une famille finie S_η telle que $\mathcal{S}(S_\eta)A$ soit η–proche de A, une variante du théorème du point fixe pour des applications contractantes dit alors que le point fixe $\mathcal{L}(S_\eta)$ est $\varepsilon(\eta)$–voisin de A (avec $\varepsilon(\eta) \to 0$ lorque $\eta \to 0$;

évidemment $\#S_\eta \to \infty$ aussi). Le schéma de compression est alors justifié simplement.

Dans les applications, E est une partie de l'espace euclidien (à deux dimensions) et S une famille finie de similitudes contractantes. On construit une bibliothèque $\mathcal{B}_\varepsilon$ de compacts (chacun codé par un nombre fini de similitudes) qui est ε–dense dans l'ensemble $\mathcal{K}(E)$. On est ramené ainsi, étant donné un compact K, à parcourir (la rapidité de la lecture de $\mathcal{B}_\varepsilon$ pose d'autres problèmes...) la bibliothèque $\mathcal{B}_\varepsilon$ pour découvrir un compact $\mathcal{L}(S)$, pour une certaine famille S, approchant K à ε–près.

Dans ce schéma, décomprimer une image, c'est produire le point fixe $\mathcal{L}(S)$ à partir de la famille S. En général $\mathcal{L}(S)$ a une structure très compliquée (cantorienne!) et décomprimer ne peut signifier que livrer une approximation de $\mathcal{L}(S)$. On montre facilement que $\mathcal{L}(S)$ est l'ensemble des points limites des trajectoires issues d'un point base x (quelconque) de E sous l'actions de S. Générer $\mathcal{L}(S)$ en tracant les trajectoires issues de x pose le problème du choix de la trajectoire : on quantifie ce choix en introduisant une chaîne de Markov associée à la famille S, déterminée par un vecteur de probabilité $\boldsymbol{p} = (p_s)_{s\in S}$ et indiquant la probabilité du choix de l'action de chaque élément de S à l'instant n ($n \in \mathbf{N}^*$). A cette chaîne markovienne, est associée une mesure stationnaire $\mu_{S,\boldsymbol{p}}$, dont l'unicité est garantie par les propriétés contractantes de la famille S : de plus, son support coïncide avec l'ensemble limite $\mathcal{L}(S)$.

On remplace ainsi le problème de l'approximation de l'ensemble $\mathcal{L}(S)$ par celui de la mesure $\mu_{S,\boldsymbol{p}}$ et il apparaît un théorème d'approximation *à la Birkhoff* : une orbite approche en moyenne (*i.e.* la moyenne des masses de Dirac le long des orbites à temps fini) la mesure $\mu_{S,\boldsymbol{p}}$ pour la topologie de la convergence vague et ce, pour presque toute trajectoire. Plus précisémment, on a le théorème :

THÉORÈME 0.1. — *Soit E compact, S une famille finie de contractions de E, $\boldsymbol{p} = (p_s)_{s\in S}$ un vecteur de probabilités et $\mu_{S,\boldsymbol{p}}$ l'unique mesure stationnaire sur E pour la probabilité de transition $t_{S,\boldsymbol{p}}(x, B) = \sum_{s\in S} p_s \chi_B(sx)$ de $x \in E$ au borélien B de E. Soit $\Omega = S^{\mathbf{N}^*}$, avec probabilité produit induite par le vecteur de probabilité $\boldsymbol{p}$.*

Il existe une partie $\tilde{\Omega}$ de Ω, de mesure pleine, telle que $\forall x \in E, \forall f \in \mathcal{C}_0(E), \forall \omega \in \tilde{\Omega}$,

$$\lim_{n\to\infty} \frac{f(\omega_1 x) + f(\omega_2\omega_1 x) + \ldots + f(\omega_n \ldots \omega_1 x)}{n} = \int_E f d\mu_{S,\boldsymbol{p}}. \tag{0.1}$$

Le théorème de Birkhoff donne la convergence (0.1) pour une trajectoire $(x, \omega_1 x, \omega_2\omega_1 x, \ldots)$ presque sûrement dans $E \times \Omega^{\mathbf{N}^*}$: l'hypothèse de contraction est utilisée de manière cruciale pour établir (0.1) pour *tout* point base et pour presque toute combinatoire $(\omega_1, \omega_2, \ldots)$ de la trajectoire.

Il nous a paru utile de rappeler quelques notions sur les processus et leur dynamique ainsi que sur les propriétés particulières des processus de Markov. Il est alors facile de démontrer le théorème 0.1, dont une généralisation à des familles contractantes en moyenne, avec une dynamique aléatoire (*i.e.* suivant un vecteur de probabilités) variant avec x, est démontrée dans l'article [4] de Elton. Nous traitons ici le cas des familles contractantes en moyenne avec p constant, renvoyant le lecteur à [4] pour la démonstration du résultat général : une fois établie l'unicité de la distribution initiale stationnaire, la démonstration utilise le théorème des martingales; on n'aura utilisé pour notre part que le théorème ergodique de Birkhoff (et la loi forte des grands nombres qu'on peut voir comme un de ses corollaires).

Merci à Jean Brossard pour sa lecture critique d'une version préliminaire de ce texte!

1. Point fixe dans $\mathcal{K}(E)$

Soit E un espace métrique complet et $\mathcal{K}(E)$ l'espace des parties compactes de E. On vérifie que l'application donnée par

$$d(A,B) = \sup_{a\in A, b\in B} (d(a,B), d(A,b))$$

définit une distance, dite distance de Haussdorff, sur $\mathcal{K}(E)$, telle que $(\mathcal{K}(E), d)$ soit complet et compact si E l'est.

On suppose désormais E compact et suffisamment régulier pour que les fonctions lipchitziennes soient denses dans l'espace $\mathcal{C}_0(E)$ des fonctions continues.

Soit $\mathcal{S}$ la transformation définie sur $\mathcal{K}(E)$ par $\mathcal{S}A = \cup_{s\in S} sA, A \in \mathcal{K}(E)$.

LEMME 1.1. — *Si les applications s sont lipchitziennes de rapport $|s|$, l'application $\mathcal{S}$ est lipchitzienne de rapport $|\mathcal{S}| = \sup_{s\in S} |s|$*

Preuve. — On a les inégalités

$$\begin{aligned} d(\cup_{s\in S} sA, \cup_{r\in S} rB) &= \sup_{r,s\in S, a\in A, b\in B} \Big(d(sa, \cup_{\tilde{s}\in S}\tilde{s}B), d(\cup_{\tilde{r}\in S}\tilde{r}A, rb)\Big) \\ &\leqslant \sup_{r,s\in S, a\in A, b\in B} (d(sa, sB), d(rb, rA)) \\ &\leqslant \sup_{s\in S, a\in A, b\in B} (|s|d(a,B), |s|d(A,b)) \leqslant |S|d(A,B). \qquad \square \end{aligned}$$

Exemples. — Soit s_h^a l'homothétie de centre a et de rapport h. Sur la droite, $\mathcal{L}(\{s_h^0, s_h^1\})$ est un ensemble de Cantor si $h < 1/2$ (avec pour $h = 1/3$ l'ensemble de Cantor exemplaire), alors que pour $h \geqslant 1/2$, on trouve toujours l'intervalle $[0,1]$. Un polytope K de sommets $\{P_j\}$ dans $\{\mathbf{R}^d\}$ est réalisé comme $\mathcal{L}(s_{h_j}^{F_j})$ en prenant des h_j inférieurs à 1, pas trop petits.

On note par $\mathcal{L}(S)$ le point fixe unique de $\mathcal{S}(S)$ dans $\mathcal{K}(E)$. Pour tout compact A, la suite $(\mathcal{S}^n A)$ converge vers $\mathcal{L}(S)$. En prenant $A = \{x_0\}$, on en déduit que l'ensemble des points d'adhérence des orbites de x_0 sous l'action de S est égal à $\mathcal{L}(S)$. Soit $x_\infty \in \mathcal{L}(S)$, point d'adhérence de la suite $(x_n = s_{i_{n,1}} \dots s_{i_{n,l(n)}} x_0)$. En usant du procédé diagonal, on peut en extraire une suite $(\tilde{x}_n)$ telle que $\tilde{i}_{n,k}$ soit constant $(= j_k)$ pour tout k. La suite $(s_{j_1} \dots s_{j_k} x_0)$ est de Cauchy :

$$d(s_{j_1} \dots s_{j_k} x_0, s_{j_1} \dots s_{j_{k+1}} x_0) \leqslant |S|^k d(x_0, s_{j_{k+1}} \dots s_{j_{k+1}} x_0)$$

et a même limite que la suite (x_n), limite qui ne dépend pas de l'origine x_0. On vient de démontrer la surjectivité de l'application Φ définie dans le lemme suivant (dont la démonstration est omise) :

LEMME 1.2. — *Soit $\Omega = S^{\mathbb{N}^*}$ avec la métrique d définie par $d(\omega, \tilde{\omega}) = \sum_{n=1}^{\infty} |S|^n \delta_{\omega_n, \tilde{\omega}_n}$ (où $\delta_{s,r}$ vaut 1 si $s = r$, 0 sinon). Pour ω dans Ω et x dans E, la suite $(\omega_1 \ldots \omega_n x)$ converge, avec une limite indépendante de x. L'application Φ de Ω dans $\mathcal{L}(S)$ définie par $\Phi(\omega) = \lim_{n\to\infty} \omega_1 \ldots \omega_n x$ est continue.*

Remarque. — Cette application n'est pas injective en général : pour $S = \{s^0_{1/2}, s^1_{1/2}\}$, on a $1/2 = \Phi(1,2,2,\ldots) = \Phi(2,1,1,\ldots)$; pour $S = \{s^0_{1/3}, s^1_{1/3}\}$, l'application Φ est injective.

Si les applications s ne sont contractantes qu'en moyenne, au sens où il existe un vecteur de probabilité p tel que $|S|_p = \prod_{s\in S} |s|^{p_s} < 1$, on ne peut affirmer l'existence d'un point fixe pour l'application $\mathcal{S}$. On peut néanmoins définir Φ presque partout sur Ω, comme limite des applications $Y_n^x(\omega) = \omega_1 \ldots \omega_n x$:

PROPOSITION 1.3. — *Soit $\Omega = S^{\mathbb{N}^*}$ muni de la mesure produit induite par le vecteur de probabilité $(p_s)_{s\in S}$ et $x \in E$. Alors, si (S, p) est contractant en moyenne, la suite $(Y_n^x(\omega))$ converge pour presque tout ω dans Ω (avec limite ne dépendant pas de x).*

Preuve. — Soit V_n la variable aléatoire sur Ω définie par $V_n(\omega) = \log |\omega_n|$. Les variables aléatoires V_n sont indépendantes, de même loi et d'espérance

$$\mathcal{E} = \sum_{s\in S} p_s \log |s| = \log |S|_p < 0.$$

D'après la loi forte des grands nombres, la suite $(V_1 + \ldots + V_n)/n$ converge presque sûrement vers $\mathcal{E}$: il existe une partie $\tilde{\Omega}$ de probabilité 1 telle que

$$\lim_{n\to\infty} \frac{V_1(\omega) + \ldots + V_n(\omega)}{n} = \mathcal{E}, \quad \omega \in \tilde{\Omega}.$$

Ainsi sur une partie $\tilde{\Omega}$ de mesure pleine, on a, si $t \in (|S|_p, 1)$, $\omega \in \tilde{\Omega}$, l'existence d'une constante $C(\omega)$ telle que

$$|\omega_1| \ldots |\omega_n| \leqslant C(\omega) t^n.$$

La suite $(Y_n^x(\omega))$ est donc de Cauchy, avec limite indépendante de x. □

2. Mesures invariantes relativement à une probabilité de transition

A la famille S et au vecteur de probabilités p, on associe la probabilité de transition $t_{S,p}(x, B) = \sum_{s\in S} p_s \chi_B(sx)$ de $x \in E$ au borélien B de E. On en déduit un opérateur $P_{S,p}$ sur l'espace des fonctions continues

$$P_{S,p} f(x) = \sum_{s\in S} p_s f(sx), \quad f \in \mathcal{C}_0(E),$$

et par dualité un opérateur, noté $t_{S,p*}$, sur l'espace des mesures

$$t_{S,p*}\mu = \sum_{s \in S} p_s s_* \mu, \quad \mu \in \mathcal{M}(E),$$

qui laisse invariant le convexe $\mathcal{M}_1(E)$ des mesures de probabilités sur E.

DÉFINITION 2.1. — *Tout point fixe de $t_{S,p*}$ dans $\mathcal{M}_1(E)$ est dit stationnaire relativement à la probabilité de transition $t_{S,p}$.*

PROPOSITION 2.2. — *Soit une famille S contractante en moyenne relativement au vecteur de probabilité p et μ_p la mesure produit $\Omega = S^{\mathbb{N}^*}$ associée au vecteur p. Alors $\Phi_* \mu_p$ est stationnaire relativement à la probabilité de transition $t_{S,p}$.*

Preuve. — L'application Φ entrelace le décalage à droite d_s défini par $d_s(\omega) = (s, \omega_1, \omega_2 \ldots)$ sur Ω et l'action de s sur E. D'autre part, la mesure μ_p est stationnaire par rapport à la probabilité de transition $t_{\{d_s\},p}$. Ainsi :

$$t_{S,p*}(\Phi_* \mu_p) = \sum_{s \in S} p_s s_* \Phi_* \mu_p = \sum_{s \in S} p_s \Phi_* d_{s*} \mu_p = \Phi_* \Big(\sum_{s \in S} p_s d_{s*} \mu_p \Big) = \Phi_* \mu_p. \quad \square$$

L'unicité est donnée, dans le cas d'une famille contractante, à l'aide du lemme suivant et de son corollaire (qui donne aussi l'existence de la mesure invariante comme point fixe de la transformation $t_{S,p*}$).

LEMME 2.3. — *Soit $\mathrm{Lip}^1(E)$ l'ensemble des applications lipchitziennes sur E de rapport de Lipchitz au plus égal à 1. En posant*

$$d_L(\mu, \nu) = \sup_{f \in \mathrm{Lip}^1(E)} |\mu(f) - \nu(f)|, \quad \mu, \nu \in \mathcal{M}_1(E),$$

on définit une distance sur $\mathcal{M}_1(E)$ induisant la topologie de la convergence vague sur $\mathcal{M}_1(E)$.

COROLLAIRE 2.4. — *L'application $t_{S,p*}$ sur $(\mathcal{M}_1(E), d_L)$ est contractante de rapport $|S|$.*

Preuve du corollaire. — On a

$$d_L(t_{S,p*}\mu, t_{S,p} * \mu) = \sup_{f \in \mathrm{Lip}^1(E)} \Big(| \sum_{s \in S} p_s(\mu(fos) - \nu(fos)) \Big)$$

Pour f dans $\mathrm{Lip}^1(E)$, fos est lipschitzienne de rapport inférieur à $|s|$, donc $|\mu(fos) - \nu(fos)| \leqslant |s| d_L(\mu, \nu)$, et le corollaire en résulte. $\square$

Preuve du lemme. — Soit D le diamètre de E et $\mathcal{E}_M = \mathrm{Lip}^1(E) \cap \{\|\varphi\|_\infty \leqslant M\}$. On a $d_L(\mu, \nu) = \sup_{\varphi \in \mathcal{E}_\infty} |\mu(\varphi) - \nu(\varphi)| = \sup_{\varphi \in \mathcal{E}_D} |\mu(\varphi) - \nu(\varphi)|$. En effet, l'oscillation de $\varphi \in \mathrm{Lip}^1$ est au plus D; ainsi $\varphi - \varphi(x_0)$ est dans $\mathcal{E}_D$ et on a

$(\mu - \nu)(\varphi) = (\mu - \nu)(\varphi - \varphi(x_0))$. On vérifie alors que d_L est finie et définit bien une distance sur $\mathcal{M}_1(E)$.

Du fait de la densité des fonctions lipchitziennes dans $\mathcal{C}_0(E)$, il est clair que la topologie de la convergence vague est plus fine que celle déterminée par la distande d_L.

Inversement, soit $\varepsilon > 0$, E_ε (V_ε *resp.*) une ε–approximation finie de E (de $[-D, D]$ *resp.*) et φ_ε la restriction de φ à E_ε. On construit alors une famille finie $\mathcal{E}_D^\varepsilon$ ($\subset \mathcal{E}_D$) qui approche tout élément de $\mathcal{E}_D$ à 6ε près; si ψ est un élément de $\mathcal{E}_D^\varepsilon$ à distance au plus 6ε de φ, on a alors :

$$|(\mu - \nu)(\varphi)| \leqslant |(\mu - \nu)(\varphi - \psi)| + |(\mu - \nu)(\psi)| \leqslant 2.6\varepsilon + |(\mu - \nu)(\psi)|.$$

On en déduit que le voisinage ouvert $U_\varepsilon(\mu) = \{\nu, |(\mu - \nu)(\varphi)| < \varepsilon, \varphi \in \mathcal{E}_D^\varepsilon\}$ est inclus dans la boule centrée en μ de rayon 13ε pour la distance d_L, ce qui permet de conclure. □

3. Dynamique des processus (à temps discret)

Par dynamique sur un espace **E**, on entend la donnée d'une transformation T de **E** (on suppose ici le temps discret; dans le cas d'un temps continu, la donnée consisterait en la donnée d'un semi–flot de **E**). On étudie alors l'action de T vis–à–vis des structures de **E** : topologique, mesurable, mesurée.... L'objet de ce paragraphe est de montrer comment la donnée d'un processus sur Ω à valeurs dans E est équivalente à la donnée d'une dynamique sur une espace **E** et d'expliciter quelques notions dynamiques (telles celles d'événement invariant ou d'ergodicité) en termes de processus.

A. Définitions et propriétés générales.

Un processus X (à temps discret, sous–entendu désormais) sur l'espace mesurable Ω (muni, en général, d'une mesure de probabilité, notée ν dans la suite) et à valeurs dans E est une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}^*}$ sur Ω à valeurs dans E. De la mesure ν sur Ω, sont dérivées les mesures $\nu_n = (X_n)_*\nu$, distributions de X au temps n sur E : en fait, l'étude du processus X passe plus par celle des distributions ν_n que par celle de la mesure ν (*cf.* ci–dessous les quelques résultat sur les processus de Markov)!

On note par $(\mathbf{E}, \nu_X)$ l'espace $E^{\mathbb{N}^*}$ muni de la mesure $\Psi_*\nu$, où Ψ est l'application de Ω dans **E** induite par le processus et définie par $\Psi(\omega) = (X_n(\omega))_{n\geqslant 1}$. Sur **E**, on définit l'action de décalage (à gauche) $S((x_n)_{n\geqslant 1}) = (x_{n+1})_{n\geqslant 1}$. Inversement, si T est une transformation opérant sur l'espace de probabilité $(\mathbf{E}, \mu)$, on lui associe le processus $X = (X_n = T^n)$ à valeurs dans **E** et où l'espace du hasard est **E** lui même.

L'étude du processus X est en fait l'étude des événements déterminés par les variables X_n *i.e.* de la tribu engendrée par les X_n, E étant supposé muni de la tribu borélienne standard : les événements de la forme $\{(X_1, X_2, \ldots) \in G\}$, pour un borélien G de **E**, jouent ainsi un rôle particulier (et notamment lorsque G est cylindrique).

Si T, opérant sur **E**, laisse μ invariante ($T_*\mu = \mu$), alors pour $A_1, \ldots, A_n$ parties de $\mathbf{E}$, on a :

$$\begin{aligned}\mu(X_1 \in A_1, \dots, X_n \in A_n) &= \mu(T^{-1}A_1 \cap \dots \cap T^{-n}A_n)\\ &= \mu(T^{-k}(T^{-1}A_1 \cap \dots \cap T^{-n}A_n))\\ &= \mu(T^{-(k+1)}A_1 \cap \dots \cap T^{-(k+n)}A_n)\\ &= \mu(X_{k+1} \in A_1, \dots, X_{k+n} \in A_n).\end{aligned}$$

ce qui amène la définition :

DÉFINITION 3.1. — *Soit X un processus sur Ω à valeurs dans $\mathbf{E}$ et ν une mesure sur Ω. Le processus X est stationnaire pour la mesure ν si la loi de $(X_1, X_2, \dots)$ est la même que celle de $(X_k, X_{k+1}, \dots)$ pour tout k.*

Si le processus X est stationnaire, vérifions que la mesure ν_X sur $\mathbf{E}$ est invariante sous l'action du décalage S :

$$\begin{aligned}S_*\nu_X(A_1 \times A_2 \times \dots \times A_n \times \Omega \times \dots) &= \nu_X(\Omega \times A_1 \times A_2 \times \dots \times A_n \times \Omega \times \dots)\\ &= \nu(X_2 \in A_1, \dots, X_{n+1} \in A_n)\\ &= \nu(X_1 \in A_1, \dots, X_n \in A_n) \qquad \text{par stationnarité}\\ &= \nu_X(A_1 \times A_2 \times \dots \times A_n \times \Omega \times \dots).\end{aligned}$$

Deux événements A et B sont dits presque–égaux si leur différence symétrique est de mesure nulle (et on notera $A \simeq B$), l'événement A est dit T–presque invariant si $A \simeq T^{-1}A$.

Si G ($\subset \mathbf{E}$) est presque invariant sous l'action du décalage S, alors

$$\Gamma = \{(X_1, X_2, \dots) \in S^{-1}G\} \simeq \{(X_2, X_3, \dots) \in G\},$$

ce qui motive la définition suivante :

DÉFINITION 3.2. — *Un événement Γ est presque invariant pour le processus X s'il existe G dans $\mathbf{E}$ tel que $\Gamma \simeq \{(X_k, X_{k+1}, \dots) \in G\}$ pour tout entier k.*

On a alors la définition classique de processus ergodique relativement à une mesure ν :

DÉFINITION 3.3. — *Un processus ν–stationnaire X est ergodique si tout événement presque invariant sous l'action de ce processus est de mesure nulle ou pleine.*

et la non moins classique proposition :

PROPOSITION 3.4. — *Soit X un processus sur E et une mesure ν stationnaire vis–à–vis de X, extrémale dans le convexe $\mathcal{M}_X(\Omega)$ des mesures stationnaires pour le processus X. Alors ν est ergodique sous l'action du processus X.*

Preuve. — *Ab absurdo*, soit Δ un événement invariant non trivial et soit ν_Δ la mesure définie sur Ω par $\nu_\Delta(\Gamma) = \nu(\Gamma \cap \Delta)/\nu(\Delta)$. L'événement Δ est de la forme

$\Delta \simeq \{(X_{k+1}, X_{k+2}, \dots) \in D\}$ avec D invariant sous l'action de S. Soit G un événement de $\mathbf{E}$.

$$\begin{aligned}
\nu(\Delta)\nu_\Delta((X_1, X_2, \dots) \in G) &= \nu\big(((X_1, X_2, \dots) \in G) \cap \Delta\big) \\
&= \nu\big((X_1, X_2, \dots) \in G \cap D\big) \\
&= \nu\big((X_2, X_3, \dots) \in G \cap D\big) \quad \text{par stationnarité} \\
&= \nu\big(((X_2, X_3, \dots) \in G) \cap ((X_2, X_3, \dots) \in D)\big) \\
&= \nu\big(((X_2, X_3, \dots) \in G) \cap \Delta\big) \\
&= \nu(\Delta)\nu_\Delta\big((X_2, X_3, \dots) \in G\big).
\end{aligned}$$

ainsi ν_Δ est stationnaire, de même que $\nu_{^c\Delta}$: ν, barycentre de ν_Δ et $\nu_{^c\Delta}$, n'est pas extrémal, ce qui contredit l'hypothèse. □

LEMME 3.5. — *Soit* (X, Ω) *un processus à valeurs dans* E, $\phi : E \to F$ *mesurable.*

(i) Si (X, Ω, ν) *est stationnaire,* $(\phi(X), \Omega, \nu)$ *l'est aussi.*

(ii) Si (X, Ω, ν) *est ergodique,* $(\phi(X), \Omega, \nu)$ *l'est aussi.*

Preuve. — (i) On vérifie simplement la propriété de stationnarité pour des événements dans la tribu engendrée par les $\phi(X_n)$, tribu incluse dans celle engendrée par les X_n.

(ii) Soit Γ invariant pour le processus $\phi(X)$. Il existe G borélien de $\mathbf{F}$ tel que $\Gamma \simeq \{(\phi(X_1), \dots \phi(X_k), \dots) \in G\} \simeq \{(X_1, \dots, X_k \dots) \in \Phi^{-1}(G)\}$ où Φ est l'application produit induite par ϕ entre $\mathbf{E}$ et $\mathbf{F}$. Ainsi Γ, invariant pour le processus X, est de mesure nulle ou pleine. □

B. Processus de Markov.

Rappelons la définition des probabilités conditionnelles ([6]) :

DÉFINITION 3.6. — *Soit* (Ω, ν) *un espace mesuré. La probabilité conditionnelle de* $\{Y \in A\}$ *relativement aux variables aléatoires* $Y_1, \dots, Y_{k-1}$ *est toute fonction* $g_A(\omega)$ *mesurable relativement à la tribu (complète)* $\mathcal{F}(Y_1, \dots, Y_{k-1})$ *engendrée par les* $Y_1, \dots, Y_{k-1}$ *telle que*

$$\nu(Y \in A, \omega \in \Lambda) = \int_\Lambda g_A(\omega) d\nu(\omega)\ , \ \forall \Lambda \in \mathcal{F}(Y_1, \dots, Y_{k-1}),$$

et qu'on notera $m(Y \in A | Y_1, \dots, Y_{k-1})$.

On a alors la définition d'un processus de Markov (à temps discret) :

DÉFINITION 3.7. — *Un processus* X *sur l'espace mesuré* (Ω, ν) *est dit de Markov si* $m(X_n \in A | X_1, \dots, X_{n-1}) = m(X_n \in A | X_{n-1})$ *à tout instant* n.

Une variable aléatoire $\mathcal{F}(X_{n-1})$–mesurable est égale presque partout à une variable du type $g(X_{n-1})$ où g est une fonction définie sur E. On désignera par $m(.|X_{n-1}=x)$ une telle fonction pour la variable aléatoire $m(.|X_{n-1})$. On n'envisagera dans la suite que des processus de Markov dont les probabilités de transition $m(X_n \in A|X_{n-1}=x)$ sont indépendantes du temps : $m(X_n \in A|X_{n-1}=x) = t(x,A)$.

A tout processus de Markov X à valeurs dans E et de distribution initiale $\nu_1 = X_{1*}\nu$ est associé l'espace de représentation $\mathbf{E}$ avec mesure $\nu_{\mathbf{E}}$ définie sur les événements cylindriques par

$$\nu_{\mathbf{E}}(A_1 \times \ldots \times A_n) = \int_{A_1}\left[\int_{A_2}\cdots\left[\int_{A_n} t(x_{n-1},dx_n)\right]\ldots t(x_1,dx_2)\right]d\nu_1(x_1)$$

et portant le processus $\tilde{X}$ donné par les applications coordonnées. Les propriétés markoviennes de X et $\tilde{X}$ étant les mêmes, on peut supposer dans la suite X défini sur son espace de représentation, la mesure $\nu_{\mathbf{E}}$ étant déterminée par la donnée de la distribution initiale ν_1 et la probabilité de transition t.

A la probabilité de transition t est associé l'opérateur t_* sur $\mathcal{M}_1(E)$ défini par $t_*\mu(B) = \int_E t(x,B)d\mu(x)$. Si T est une tranformation de E et t_* la probabilité de transition donnée par $t(x,B) = \chi_B(Tx)$, t_* est l'opération habituelle d'image directe T_*, duale de la composition par T sur $\mathcal{C}_0(E)$ si T est continue (*cf.* partie 2).

LEMME 3.8. — *Le processus de Markov (X,t,ν_1) est stationnaire si et seulement si ν_1 est stationnaire pour la probabilité de transition t.*

Preuve. — Si X_n est stationnaire, les X_i ont même loi. Or on a $\nu_2 = t_*\nu_1$ par hypothèse markovienne, aussi ν_1 est stationnaire sous l'action de t_*.

Réciproquement,

$$\begin{aligned}
&\nu(X_k \in A_1,\ldots,X_{k+n} \in A_n)\\
&\quad = \int_X\int_X\cdots\int_{A_1}\cdots\int_{A_n} t(x_{n+k-1},dx_{n+k})\ldots t(x_1,dx_2)\ldots d\nu_1(x_1)\\
&\quad = \int_{A_1}\cdots\int_{A_n} t(x_{n+k-1},dx_{n+k})\ldots (t_*)^k d\nu_1(x_k)\\
&\quad = \int_{A_1}\cdots\int_{A_n} t(x_{n+k-1},dx_{n+k})\ldots d\nu_1(x_k)\\
&\quad = \nu(X_1 \in A_1,\ldots,X_n \in A_n)
\end{aligned}$$

Le processus X est donc stationnaire □

PROPOSITION 3.9. — *Soit (X,ν_1,t) un processus de Markov. Si ν_1 est extrémale dans le convexe des mesures t_*–stationnaires, le processus de Markov (X,ν_1,t) est ergodique.*

Preuve. — Toute mesure stationnaire pour le processus de Markov X avec probabilité de transition t est déterminée de manière unique par la distribution initiale

ν_1 et la probabilité de transition t. Comme pour la proposition 3.4, on raisonne par l'absurde, supposant l'existence d'un événement presque invariant Δ de mesure non triviale. Nous admettrons la proposition (non triviale) :

PROPOSITION 3.10 ([6, p 197]). — *Soit Δ un événement presque invariant du processus de Markov X. Alors, il existe un borélien D de E tel que $\Delta = \{X_i \in D\}, \forall i \geqslant 1$.*

On vérifie alors que ν_Δ définie par $\nu_\Delta(A) = \nu(A \cap \Delta)/\nu(\Delta)$ est markovienne pour le processus X avec probabilité de transition t :

$$\begin{aligned}
\nu_\Delta((X_{n+1}, \dots, X_1) \in A_{n+1} \times \dots \times A_1)&\\
&= \nu\big(((X_{n+1}, \dots, X_1) \in A_{n+1} \times \dots \times A_1) \cap \Delta\big)/\nu(\Delta)\\
&= \nu(X_{n+1} \in A_{n+1} \cap D, \dots, X_1 \in A_1 \cap D)/\nu(\Delta)\\
&= \int_{\{X_n \in A_n \cap D, \dots, X_1 \in A_1 \cap D\}} t(X_n, A_{n+1}) d\nu / \nu(\Delta)\\
&= \int_{\{X_n \in A_n, \dots, X_1 \in A_1\}} t(X_n, A_{n+1}) d\nu_\Delta
\end{aligned}$$

de distribution initiale ν_D avec $\nu_D(A) = \nu(A \cap D)/\nu(D)$ (vu $\nu_1(D) = \nu(\Delta)$). On en déduit que ν_D est t_*–invariante et que ν_1 n'est pas extrémale dans le convexe des mesures t_*–invariantes. □

C. Théorème de Birkhoff.

Contentons-nous de citer le théorème de Birkhoff suivant ([6]) :

THÉORÈME 3.11. — *Soit (E, μ) un espace mesuré, T une transformation de E laissant μ invariante et f dans $L^1(E, \mu)$. Alors, $(f \circ T + \dots + f \circ T^n)/n$ converge presque partout sur E. Si en outre T est ergodique, la limite est constante, égale à l'intégrale $\int_E f d\mu$.*

On a un résultat analogue pour les processus stationnaires :

COROLLAIRE 3.12. — *Soit X un processus stationnaire sur l'espace mesuré (Ω, ν) à valeurs dans E. Soit f une fonction intégrable sur E relativement à la distribution $\nu_1 = X_{1*}\nu$. Alors $(f \circ X_1 + \dots + f \circ X_n)/n$ converge presque partout sur Ω. Si en outre, le processus est ergodique, la limite est constante égale à $\int_E f d\nu_1$).*

Remarque. — La loi forte des grands nombres peut s'obtenir à partir du théorème précédent en remarquant que le décalage sur **E** est ergodique (d'après la loi dite 0–1).

4. Le théorème principal

Soient $E, S\ldots$ les objets introduits dans la partie 1. Nous devons commencer par démontrer le résultat d'unicité suivant :

PROPOSITION 4.1. — *Supposons la famille S contractante en moyenne relativement au vecteur de probabilité p. La mesure $\mu_{S,p}$ est l'unique point fixe de $t_{S,p*}$ dans $\mathcal{M}_1(E)$.*

Preuve. — Si chaque application de S est contractante, on a vu l'unicité de $\mu_{S,p}$ dans la partie 2.

Dans le cas de contraction en moyenne, on va montrer que $\mu_{S,p}$ est un attracteur vague pour l'action de $t_{S,p}$ sur $\mathcal{M}_1(E)$, l'unicité du point fixe de $t_{S,p}$ en résultera alors.

La mesure $(t_{S,p})^n_*\mu$ ($\mu \in \mathcal{M}_1(E)$) est la distribution au temps n du processus de Markov sur E de distribution initiale μ et probabilité de transition $t_{S,p}$. Soit Z^x le processus de Markov de distribution initiale $\delta(x)$ (masse de Dirac en x) : $Z^x_n(\omega) = \omega_n \ldots \omega_1 x$ si on prend $(\Omega = S^{\mathbb{N}^*}, \mu_p$ comme espace de représentation. Dans la partie 1, on a introduit la variable Y^x_n définie par $Y^x_n(\omega) = \omega_1 \ldots \omega_n x$. Les variables Y^x_n et Z^x_n ont même loi, ainsi pour f continue sur E,

$$< (t_{S,p})^n_*\delta(x), f >= E[f(Z^x_n)] = E[f(Y^x_n)]$$

or, d'après le définition de Φ (*cf.* lemme 1.2 et proposition 1.3), on a

$$\lim_{n\to\infty} E[f(Y^x_n)] = E[f(\Phi)] = \int_E f d\mu_{S,p},$$

d'où la convergence vague de $(t_{S,p*})^n\delta(x)$ vers $\mu_{S,p}$.

Pour μ quelconque dans $\mathcal{M}_1(E)$, considérant le processus de Markov Z^μ de distribution initiale μ, on a

$$\begin{aligned} < (t_{S,p})^n_*\mu, f > &= E[f(Z^\mu_n)] = \int_E m[f(Z^\mu_n)|Z_1 = x]d\mu(x) \\ &= \int_E E[f(Z^x_n)]d\mu(x) \to \int_E (\int_E f d\mu_{S,p})d\mu(x) = \int_E f d\mu_{S,p}, \end{aligned}$$

d'où le résultat. □

THÉORÈME 4.2. — *Soit E compact, S une famille d'applications lipchitziennes de E, famille finie et contractante relativement au vecteur de probabilité $p = (p_s)_{s\in S}$ et $\mu_{S,p}$ l'unique mesure stationnaire sur E pour la probabilité de transition $t_{S,p}(x, B) = \sum_{s\in S} p_s \chi_B(sx)$. Soit $\Omega = S^{\mathbb{N}^*}$, avec probabilité produit μ_p induite par le vecteur de probabilité p. Il existe une partie $\tilde{\Omega}$ de mesure pleine telle que $\forall x \in E, \forall f \in C^0(E), \forall\omega \in \tilde{\Omega}$,*

$$\lim_{n\to\infty} \frac{f(\omega_1 x) + f(\omega_2\omega_1 x) + \ldots + f(\omega_n \ldots \omega_1 x)}{n} = \int_E f d\mu_{S,p}.$$

Preuve. — En utilisant la séparabilité de l'espace des fonctions continues sue E, il suffit de montrer le théorème pour une f continue donnée quelconque.

Le processus de Markov Z sur E de distribution initiale $\mu_{S,p}$ et de probabilité de transition $t_{S,p}$ est ergodique, vu l'unicité de $\mu_{S,p}$; il en est de même pour le processus $f(Z)$ d'après le lemme 3.5. On peut prendre comme espace de représentation $\tilde{\mathbf{E}} = E \times \Omega$ avec comme mesure ν_Z la mesure produit $\mu_{S,p} \otimes \mu_p$ et applications (Z_n) definies par, si $\tilde{x} = (x, \omega) \in \tilde{\mathbf{E}}$, $Z_0(\tilde{x}) = x, Z_n(\tilde{x}) = Z_n^x(\omega), n > 0$.

Ainsi, d'après le théorème de Birkhoff, il existe une partie C de $\tilde{\mathbf{E}}$, de mesure pleine telle que, si $\tilde{x} = (x, \omega) \in C$:

$$\frac{\sum_{k=1}^{n} f(Z_k \tilde{x})}{n} = \frac{f(\omega_1 x) \ldots + f(\omega_n \ldots \omega_1 x)}{n} \to \int_E f d\mu_{S,p}.$$

On a

$$\begin{aligned} 1 = \nu_Z(\tilde{x} \in C) &= \int_E m[\tilde{x} \in C | Z_0(\tilde{x}) = x] d\mu_{S,p}(x) \\ &= \int_E \mu_p[(\omega_1, \omega_2 \ldots) | (x, \omega_1 x, \omega_2, \omega_1 x, \ldots) \in C] d\mu_{S,p}(x) \end{aligned}$$

Ainsi, pour un x_0, on a $\mu_p(B_{x_0}) = 1$, avec $B_{x_0} = \{\omega \in \Omega | (x_0, \omega_1 x_0, \omega_2 \omega_1 x_0, \ldots) \in C\}$. Soit B_1 une partie de mesure pleine de Ω (dont l'existence a été établie lors de la preuve de la proposition 1.3) telle que

$$|\omega_1| \ldots |\omega_n| \leqslant C t^n, (t < 1)$$

Alors, pour $x \in E$ et $\omega \in B_1$, vu la continuité uniforme de f sur E, la moyenne $S^n(x) = (f(\omega_1 x) + \ldots + f(\omega_n \ldots \omega_1 x))/n$ converge, avec même limite, en même temps que la moyenne $S^n(x_0)$.

Ainsi sur la partie $B_{x_0} \cap B_1$ ($\subset \Omega$) de mesure 1, pour tout x dans E,

$$\frac{f(\omega_1 x) \ldots + f(\omega_n \ldots \omega_1 x)}{n} \to \int_E f d\mu_{S,p}. \qquad \square$$

Références

[1] BARNSLEY M.F, DEMKO S. — *Iterated functions systems and the global construction of fractals*,Proc. R. Soc. London. A, **399** (1985), 243–275.

[2] BARNSLEY M.F., ELTON J. — A new class of Markov processes for image encoding, prépublication, 1987.

[3] BARNSLEY M.F., SLOAN A.D. — *A better way to compress images*,Byte Magazine, **13.1** (1987), 215–223.

[4] ELTON J. — *An ergodic theorem for iterated maps*,Erg. Th. & Dynam. Sys., **7** (1987), 481–488.

[5] HUTCHINSON J. — *Fractals and self-similarity*,Indiana U. J. of Math., **30** (1981), 713–747.

[6] STOUT W. — *Almost sure convergence*, Academic Press, New York, 1974.

Laurent GUILLOPÉ
INSTITUT FOURIER
Laboratoire de Mathématiques
BP 74
38402 St MARTIN D'HÈRES Cedex (France)