

# Etude de la vitesse de convergence des chaînes de Markov

Nicolas Boulanger et Vincent Vargas

Sujet proposé par Thierry BODINEAU

19 juin 2001

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Outils de base et théorème de Perron-Frobenius</b>	<b>2</b>
2.1	Théorème de Perron-Frobenius . . . . .	2
2.2	Un cadre naturel d'étude: les espaces de Hilbert . . . . .	4
2.3	Passage du discret au continu . . . . .	5
<b>3</b>	<b>Une méthode pour étudier la vitesse de convergence</b>	<b>7</b>
3.1	Définition du trou spectral . . . . .	7
3.2	Les théorèmes de convergence . . . . .	8
3.3	Une méthode d'approximation du trou spectral . . . . .	10
<b>4</b>	<b>Application d'une méthode graphique pour estimer le trou spectral</b>	<b>11</b>
4.1	Définitions de base . . . . .	11
4.2	Théorème . . . . .	11
4.3	Exemple . . . . .	12
<b>5</b>	<b>Conclusion</b>	<b>14</b>

# 1 Introduction

Le but de cet exposé est de présenter quelques méthodes pour déterminer les vitesses de convergence d'une chaîne de Markov apériodique sur un espace d'états finis vers sa probabilité stationnaire.

Pour cela, on utilise une idée classique en mathématiques à savoir transformer un problème à variable discrète en problème à variable continue où l'étude est facilitée par des outils conséquents tels que l'intégration.

## 2 Outils de base et théorème de Perron-Frobenius

Dans cette partie, on rappelle les définitions concernant les chaînes de Markov. Ensuite, on énonce le théorème de Perron-Frobenius qui affirme, sous certaines conditions, la convergence des chaînes de Markov vers la mesure stationnaire associée. On introduira ensuite un espace de Hilbert approprié à l'étude d'une chaîne de Markov. Enfin, on montrera comment on associe à une chaîne de Markov discrète une chaîne à temps continu qui a, sous certaines conditions, un comportement asymptotique analogue à celui de la chaîne discrète.

### 2.1 Théorème de Perron-Frobenius

On rappelle qu'une matrice stochastique est une matrice à coefficients positifs dont la somme des coefficients de chaque ligne vaut 1. On remarque qu'une matrice stochastique peut être assimilée à la matrice de transition d'une chaîne de Markov.

**Théorème 2.1 :**

*Soit  $\mathbf{M}$  une matrice stochastique de dimension  $n$  irréductible et apériodique. Alors il existe une unique probabilité stationnaire  $(m_j)_{1 \leq j \leq n}$ . De plus, on a*

$$\forall i, \forall j, \lim_{l \rightarrow \infty} M_{i,j}^l = m_j.$$

*Preuve :* On utilise deux lemmes intermédiaires :

**Lemme 2.1 :**

*Soit  $\mathbf{M}$  une matrice stochastique irréductible de dimension  $n$ . Alors  ${}^t\mathbf{M}$  admet 1 pour valeur propre et il existe un vecteur  $X$ , à coordonnées strictement positives tel que:  $\mathbf{Ker}({}^t\mathbf{M} - \mathbf{I}) = \mathbb{R}X$ .*

Nous admettons ce lemme qui se démontre par des techniques classiques d'algèbre linéaire.

**Lemme 2.2 (De Doeblin) :**

*Soit  $\mathbf{M}$  une matrice stochastique telle que il existe  $k \in \mathbb{N}^*$  et une mesure de probabilité  $(q_j)_{1 \leq j \leq n}$  sur  $\{1, \dots, n\}$  telle que la condition de Doeblin soit vérifiée, c'est à dire :*

$$\forall i, \forall j, M_{i,j}^k \geq cq_j \quad (\mathbf{D}).$$

Alors, en notant  $m = (m_j)_{1 \leq j \leq n}$  la probabilité stationnaire de  $\mathbf{M}$ , on a :

$$\forall i, \quad \sum_j |\mathbf{M}_{i,j} - m_j| \leq 2(1-c)^{[l/k]}, \quad (1)$$

où  $[ ]$  désigne la partie entière.

*Preuve :* Soit  $\mathbf{M}^\infty$  la matrice dont toutes les lignes sont égales à  $m$ , soit  $\mathbf{Q}$  la matrice dont toutes les rangées sont égales à  $(q_j)_{1 \leq j \leq n}$ . On pose :

$$\mathbf{N} = \frac{1}{1-c}(\mathbf{M}^k - c\mathbf{Q}), \quad \mathbf{N}^\infty = \frac{1}{1-c}(\mathbf{M}^\infty - c\mathbf{Q}).$$

Ces deux matrices sont stochastiques. Elles vérifient :

$$\mathbf{M}^k - \mathbf{M}^\infty = (1-c)(\mathbf{N} - \mathbf{N}^\infty)$$

et

$$\mathbf{M}^{kl} - \mathbf{M}^\infty = (\mathbf{M}^k - \mathbf{M}^\infty)^l = (1-c)^l(\mathbf{N} - \mathbf{N}^\infty)^l.$$

On observe que :  $(\mathbf{N} - \mathbf{N}^\infty)^2 = (\mathbf{N} - \mathbf{N}^\infty)\mathbf{N}$  car  $\mathbf{N}^\infty$  a des colonnes constantes, donc  $\mathbf{P}\mathbf{N}^\infty = \mathbf{N}^\infty$  pour toute matrice stochastique  $\mathbf{P}$ . On en déduit que :  $(\mathbf{N} - \mathbf{N}^\infty)^l = (\mathbf{N} - \mathbf{N}^\infty)\mathbf{N}^{l-1}$ . Si on pose pour toute matrice  $\mathbf{A}$ ,  $\|\mathbf{A}\| = \max_i \sum_j |\mathbf{A}_{i,j}|$  (on a considéré une norme triple), alors on a :

$$\|\mathbf{M}^{kl} - \mathbf{M}^\infty\| \leq (1-c)^l \|\mathbf{N} - \mathbf{N}^\infty\| \|\mathbf{N}^{l-1}\|.$$

Comme  $\mathbf{N}$  est stochastique, on a  $\|\mathbf{N}\| = 1$  et  $\|\mathbf{N} - \mathbf{N}^\infty\| \leq 2$ . Donc,

$$\max_i \sum_j |\mathbf{M}_{i,j}^{kl} - \mathbf{M}^\infty| \leq 2(1-c)^l. \quad (1)$$

Le résultat découle alors du fait que  $l \rightarrow \|\mathbf{M}^l - \mathbf{M}^\infty\|$  est décroissante.

On retourne aux hypothèses du théorème de Perron-Frobenius : soit  $\mathbf{M}$  une matrice stochastique de dimension  $n$  irréductible et apériodique. Montrons qu'il existe  $k$  tel que  $\forall i, \forall j, \mathbf{M}_{i,j}^k > 0$ .

On note  $\forall i \in [1, n]$   $I(i) = \{n \in \mathbb{N}^*, \mathbf{M}^n(i, i) > 0\}$  ( $I(i)$  est clairement un semi-groupe). Montrons d'abord que  $\forall i \in [1, n]$  il existe un entier  $n_i$  tel que  $n_i$  et  $n_{i+1}$  appartiennent à  $I(i)$ . Soit  $\lambda_1, \dots, \lambda_k \in I(i)$  de pgcd 1; par le théorème de Bezout :

$$\exists u_1, \dots, u_k \in \mathbb{Z}, \quad u_1 \lambda_1 + \dots + u_k \lambda_k = 1$$

donc, en notant  $u_i = u_i^+ - u_i^-$ , on obtient :

$$u_1^+ \lambda_1 + \dots + u_k^+ \lambda_k = 1 + u_1^- \lambda_1 + \dots + u_k^- \lambda_k \in I(i).$$

On a alors :  $\forall n \geq n_i^2 - 1, \mathbf{M}^n(i, i) > 0$ .

En effet,  $\forall n \geq n_i^2 - 1$ , par division euclidienne,  $n$  s'écrit sous la forme  $n = kn_i + p$   $p \in \{1, \dots, n_i - 1\}$ . Nécessairement,  $p \geq n_x - 1$  donc  $n = p(n_i + 1) + (k - p)n_i \in I(i)$ , par la propriété de semi-groupe de  $I(i)$ .

Soit  $n_{i,j}$  un entier  $k$  quelconque tel que  $\mathbf{M}^k(i, j) > 0$  et  $M = \max_{i,j} \{n_i^2 - 1 + n_{i,j}\}$ .  $M$  vérifie la propriété voulue car on a :

$$\forall i, \forall j \in [1, n] \quad \mathbf{M}_{i,j}^M \geq \mathbf{M}_{i,j}^{M-n_{i,j}} \mathbf{M}_{i,j}^{n_{i,j}} > 0.$$

Nous sommes donc en mesure de démontrer le théorème de Perron Frobenius. Le lemme 1 nous assure l'existence d'une mesure stationnaire  $(m_j)_{1 \leq j \leq n}$ . Le théorème découle alors du lemme de Doeblin appliqué à  $c = \min_{i,j} \{\mathbf{M}_{i,j}^M / m_j\} \in ]0, 1]$  et  $(q_j)_{1 \leq j \leq n} = (m_j)_{1 \leq j \leq n}$ .

**Remarque 2.1 :** *Au cours de la démonstration, on a montré qu'il y avait convergence exponentielle des  $\mathbf{M}_{i,j}^l$  vers  $m_j$  (cf. l'égalité (1)).*

## 2.2 Un cadre naturel d'étude : les espaces de Hilbert

On peut modéliser une chaîne de Markov sur un ensemble fini  $\chi$ , en utilisant sa matrice de transition, comme une application linéaire sur les fonctions  $f$  définies sur  $\chi$  et à valeurs complexes. Le but de cette section est de munir l'espace des fonctions définies sur  $\chi$  d'un produit scalaire, afin de disposer des outils relatifs aux espaces hilbertiens pour étudier la chaîne de Markov vue comme une application linéaire.

On note  $\chi$  l'ensemble fini des états,  $\pi$  une probabilité sur  $\chi$  et  $l^2(\pi)$  l'espace de Hilbert des fonctions  $f : \chi \rightarrow \mathbb{C}$ ,  $l^2$ -intégrables, muni du produit scalaire :

$$\langle f, g \rangle = \sum_{x \in \chi} \pi(x) f(x) g(x).$$

Soit  $\mathbf{K}(x, y)$  la matrice dans la base canonique d'une application linéaire  $\mathbf{K} : l^2(\pi) \rightarrow l^2(\pi)$ . L'adjoint  $\mathbf{K}^*$  vérifie alors  $\mathbf{K}^*(x, y) = \pi(y) \overline{\mathbf{K}(y, x)} / \pi(x)$ . Donc  $\mathbf{K}$  est auto-adjoint si, et seulement si,  $\mathbf{K}(x, y) = \pi(y) \overline{\mathbf{K}(y, x)} / \pi(x)$ . On note aussi

$$\pi(f) = \sum_x f(x) \pi(x), \quad Var_\pi(f) = \sum_x |f(x) - \pi(f)|^2 \pi(x).$$

Voici le résultat essentiel de cette section :

**Lemme 2.3 (de diagonalisation) :**

*Soit  $\mathbf{K}$  une application linéaire auto-adjointe de  $l^2(\pi)$ . Alors  $\mathbf{K}$  est diagonalisable dans une base orthonormale de  $l^2(\pi)$  avec des valeurs propres réelles  $\beta_0 \geq \beta_1 \geq \dots \geq \beta_{n-1}$ .*

### 2.3 Passage du discret au continu

Dans ce chapitre et dans toute la suite, les matrices  $\mathbf{K}$  seront stochastiques, irréductibles et  $\pi$  désignera la mesure stationnaire de  $\mathbf{K}$ . Soit  $\mathbf{K}$  une telle matrice; on lui associe le semi-groupe à temps continu suivant :

$$\mathbf{H}_t = e^{-t(\mathbf{I}-\mathbf{K})} = e^{-t} \sum_0^{\infty} t^i \mathbf{K}^i / i!$$

On vérifie aisément qu'il s'agit d'un semi-groupe :

$$\begin{aligned} \mathbf{H}_{t+s} &= \mathbf{H}_t \mathbf{H}_s \\ \lim_{t \rightarrow 0} \mathbf{H}_t &= \mathbf{I} \end{aligned}$$

On note  $\mathbf{H}_t^x(y) = \mathbf{H}_t(x, y)$ . On observe aisément que  $\mathbf{H}_t^x$  est une probabilité. On introduit aussi les densités des mesures de probabilité  $\mathbf{H}_t^x$ ,  $\mathbf{K}_x^l$  par rapport à  $\pi$  :

$$k_x^l(y) = \mathbf{K}^l(x, y) / \pi(y)$$

et

$$h_t^x(y) = \mathbf{H}_t^x(y) / \pi(y)$$

On démontre à présent "l'équivalence" des comportements asymptotiques de  $\mathbf{K}^n$  et  $\mathbf{H}_t$  lorsque  $\mathbf{K}$  est réversible (c.à.d. auto-adjointe).

**Théorème 2.2** : Soit  $(\mathbf{K}, \pi)$  une matrice réversible et  $\beta_- = \max\{0, -1 + \lambda_{n-1}\}$ . Alors,

$$(1) \quad \|h_t^x - 1\|^2 \leq e^{-t} / \pi(x) + \|k_x^{[t/2]} - 1\|^2$$

$$(2) \quad \|k_x^N - 1\|^2 \leq \beta_-^{2m} (1 + \|h_l^x - 1\|^2) + \|h_N^x - 1\|^2 \quad \text{pour } N = m + l + 1$$

*Preuve :*

En considérant une base orthonormale  $(\psi_i)_{i \in [0, n-1]}$  de  $l^2(\pi)$  constituée de vecteurs propres de  $(I - K)$  associés aux valeurs propres  $(\lambda_i)_{i \in [0, n-1]}$ , vérifiant  $\psi_0 \equiv 1$ , on utilise le lemme suivant, qui est un résultat élémentaire d'algèbre euclidienne.

**Lemme 2.4** :

$$(1) \quad k^l(x, y) = \sum_{i=0}^{n-1} (1 - \lambda_i)^l \psi_i(x) \psi_i(y), \quad \|k_x^l - 1\|^2 = \sum_{i=1}^{n-1} (1 - \lambda_i)^{2l} |\psi_i(x)|^2$$

$$(2) \quad h_t(x, y) = \sum_{i=0}^{n-1} e^{-t\lambda_i} \psi_i(x) \psi_i(y), \quad \|h_t^x - 1\|^2 = \sum_{i=1}^{n-1} e^{-2t\lambda_i} |\psi_i(x)|^2$$

Ce lemme se démontre aisément en utilisant la décomposition dans la base orthonomale associée à  $\mathbf{I} - \mathbf{K}$ .

*Preuve du théorème :* Démontrons d'abord la première formule du théorème. On utilise les deux formules du lemme 2.4 en remarquant que

$$(1 - \lambda_i)^{2l} = e^{2l \ln(1 - \lambda_i)}$$

et que pour  $0 \leq x \leq 1/2$ , on a  $\ln(1 - x) \geq -2x$ .

Pour démontrer (2), on remarque que :

$$k^{2l+1}(x, x) = \sum_{i=0}^{n-1} (1 - \lambda_i)^{2l+1} |\psi_i(x)|^2$$

Cela entraîne que :

$$- \sum_{i:\lambda_i > 1} (1 - \lambda_i)^{2l+1} |\psi_i(x)|^2 \leq \sum_{i:\lambda_i < 1} (1 - \lambda_i)^{2l} |\psi_i(x)|^2.$$

D'où

$$\sum_{i:\lambda_i > 1} (1 - \lambda_i)^{2l+2} |\psi_i(x)|^2 \leq \sum_{i:\lambda_i < 1} (1 - \lambda_i)^{2l} |\psi_i(x)|^2.$$

Aussi  $\forall \lambda_i < 1$ , on a :

$$(1 - \lambda_i)^{2l} = e^{2l \log(1 - \lambda_i)} \leq e^{-2l\lambda_i}$$

D'où :

$$\sum_{i:\lambda_i < 1} (1 - \lambda_i)^{2l} |\psi_i(x)|^2 \leq \|h_l^x\|^2$$

et

$$\sum_{i \neq 0: \lambda_i < 1} (1 - \lambda_i)^{2l} |\psi_i(x)|^2 \leq \|h_l^x - 1\|^2.$$

On obtient alors, pour  $N = m + l + 1$  :

$$\begin{aligned} \|h_N^x - 1\|^2 &= \sum_i^{n-1} (1 - \lambda_i)^{2N} |\psi_i(x)|^2 \\ &= \sum_{i:\lambda_i > 1} (1 - \lambda_i)^{2N} |\psi_i(x)|^2 + \sum_{i \neq 0: \lambda_i < 1} (1 - \lambda_i)^{2N} |\psi_i(x)|^2 \\ &\leq \beta_-^{2m} \left( \sum_{i:\lambda_i > 1} (1 - \lambda_i)^{2l+2} |\psi_i(x)|^2 \right) + \sum_{i \neq 0: \lambda_i < 1} (1 - \lambda_i)^{2N} |\psi_i(x)|^2 \\ &\leq \beta_-^{2m} \|h_l^x\|^2 + \|h_N^x - 1\|^2 \\ &\leq \beta_-^{2m} (1 + \|h_l^x - 1\|^2) + \|h_N^x - 1\|^2. \end{aligned}$$

**Remarque 2.2 :** *Le comportement d'une chaîne discrète et de sa chaîne à temps continu associée peuvent être différents si la chaîne n'est pas réversible.*

### 3 Une méthode pour étudier la vitesse de convergence

Dans cette partie, on travaille sur des chaînes à temps continu. En premier lieu, on introduit la forme de Dirichlet associée à une chaîne et la notion fondamentale de trou spectral. Ensuite, on utilise des méthodes classiques d'analyse à variable réelle pour démontrer les théorèmes principaux de cette partie, qui donnent la convergence exponentielle vers la mesure stationnaire. Enfin, on donne une méthode d'approximation du trou spectral.

#### 3.1 Définition du trou spectral

**Définition 3.1 :**

Soit  $\mathbf{H}_t = e^{-t(\mathbf{I}-\mathbf{K})}$  la matrice de transition à temps continu associée à  $\mathbf{K}$ . On définit la forme de Dirichlet associée par :

$$\forall f \forall g \in l^2(\pi), \quad E(f, g) = \mathbf{Re}(\langle (\mathbf{I} - \mathbf{K})f, g \rangle)$$

**Lemme 3.1 :** L'opérateur de Dirichlet vérifie :

$$E(f, f) = \mathbf{Re}(\langle (\mathbf{I} - 1/2(\mathbf{K} + \mathbf{K}^*))f, f \rangle) = \frac{1}{2} \sum_{x,y} |f(x) - f(y)|^2 \mathbf{K}(x, y) \pi(x)$$

et

$$\frac{\partial}{\partial t} \|\mathbf{H}_t f\|_2^2 = -2E(\mathbf{H}_t f, \mathbf{H}_t f) \quad (2)$$

*Preuve :* La première égalité découle de  $\langle \mathbf{K}f, f \rangle = \langle f, \mathbf{K}^*f \rangle = \overline{\langle \mathbf{K}^*f, f \rangle}$ . Pour la seconde, on observe que :

$$\begin{aligned} E(f, f) &= \|f\|_2^2 - \mathbf{Re}(\langle \mathbf{K}f, f \rangle) \\ &= 1/2 \sum_{x,y} (|f(x)|^2 + |f(y)|^2 - 2\mathbf{Re}(\overline{f(x)}f(y))) \mathbf{K}(x, y) \pi(x) \\ &= 1/2 \sum_{x,y} |f(x) - f(y)|^2 \mathbf{K}(x, y) \pi(x) \end{aligned}$$

Enfin, pour la troisième :

$$\begin{aligned} \frac{\partial}{\partial t} \|\mathbf{H}_t f\|_2^2 &= \langle \frac{\partial}{\partial t} \mathbf{H}_t f, \mathbf{H}_t f \rangle + \langle \mathbf{H}_t f, \frac{\partial}{\partial t} \mathbf{H}_t f \rangle \\ &= \langle -(\mathbf{I} - \mathbf{K})\mathbf{H}_t f, \mathbf{H}_t f \rangle + \langle \mathbf{H}_t f, -(\mathbf{I} - \mathbf{K})\mathbf{H}_t f \rangle \\ &= -2E(\mathbf{H}_t f, \mathbf{H}_t f). \end{aligned}$$

**Remarque 3.1 :** La forme  $E$  associée à  $\mathbf{H}_t^*$  est identique à celle associée à  $\mathbf{H}_t$ .

Toutes les inégalités fondamentales de ce chapitre s'obtiennent à partir de (2) et de la résolution d'inéquations différentielles élémentaires.



**Définition 3.2 :**

Soit  $\mathbf{K}$  de forme de Dirichlet associée  $E$ . Le trou spectral  $\lambda = \lambda(\mathbf{K})$  est défini par :

$$\lambda = \min\left\{\frac{E(f, f)}{\text{Var}_\pi(f)}; \text{Var}_\pi(f) \neq 0\right\}$$

**Remarque 3.2 :** On montre aisément que  $\lambda = \min\{E(f, f); \|f\|_2 = 1, \pi(f) = 0\}$ . Donc, en diagonalisant  $\mathbf{I} - 1/2(\mathbf{K} + \mathbf{K}^*)$  dans une base orthonormale, on en déduit que  $\lambda$  est la plus petite valeur propre non nulle de  $\mathbf{I} - 1/2(\mathbf{K} + \mathbf{K}^*)$ .

**3.2 Les théorèmes de convergence****Théorème 3.1 :**

Soit  $\mathbf{K}$  un noyau markovien de trou spectral  $\lambda = \lambda(\mathbf{K})$ . On a alors :

$$\forall f \in l^2(\pi), \|\mathbf{H}_t f - \pi(f)\|_2^2 \leq e^{-2\lambda t} \text{Var}_\pi(f)$$

*Preuve :* Soit  $u(t) = \text{Var}_\pi(\mathbf{H}_t(f))$ . Par (2), on obtient  $u'(t) \leq -2\lambda u(t)$  d'où  $u(t) \leq \exp^{-2\lambda t} u(0)$ . Le résultat découle alors du fait que  $u(0) = \text{Var}_\pi(f)$ .

**Théorème 3.2 :**

Soit  $\mathbf{K}$  un noyau markovien de trou spectral  $\lambda$ . On a alors, en utilisant les notations du premier chapitre :

$$\|h_t^x - 1\|_2 \leq \sqrt{1/\pi(x)} e^{-\lambda t}$$

et

$$|\mathbf{H}_t(x, y) - \pi(y)| \leq \sqrt{\pi(y)/\pi(x)} e^{-\lambda t} \quad (**)$$

*Preuve :* Soit  $\mathbf{H}_t^*$  l'adjoint de  $\mathbf{H}_t$ . C'est un semi-groupe markovien de trou spectral  $\lambda(\mathbf{K}^*) = \lambda(\mathbf{K})$ . Si l'on pose  $\delta_x(y) = 1/\pi(x)$  si  $y = x$  et  $\delta_x(y) = 0$  sinon. Alors :  $h_t^x(y) = \frac{\mathbf{H}_t^*(y)}{\pi(y)} = \mathbf{H}_t^* \delta_x(y)$  donc, par application du lemme précédent à  $\mathbf{K}^*$ , on obtient :

$$\|\mathbf{H}_t^* \delta_x - 1\|_2^2 \leq e^{-2\lambda t} \text{Var}_\pi(\delta_x).$$

On a aussi :  $\|h_t^x - 1\|_2 \leq \sqrt{\frac{1-\pi(x)}{\pi(x)}} e^{-\lambda t} \leq \frac{1}{\sqrt{\pi(x)}} e^{-\lambda t}$ , d'où la première partie de l'énoncé. Enfin, on a par la propriété de semi-groupe :

$$\begin{aligned} |h_t(x, y) - 1| &= \left| \sum_z (h_{t/2}(x, z) - 1)(h_{t/2}(z, y) - 1)\pi(z) \right| \\ &\leq \|h_{t/2}^x - 1\|_2 \|h_{t/2}^{*y} - 1\|_2 \\ &\leq \frac{1}{\sqrt{\pi(x)\pi(y)}} \exp^{-\lambda t}. \end{aligned}$$

D'où, en multipliant par  $\pi(y)$ , on obtient l'inégalité désirée.

**Remarque 3.3 :** On a aucune estimation sur  $\sqrt{\pi(y)/\pi(x)}$ . On fait donc parfois appel à d'autres inégalités pour avoir de meilleures estimations pour des  $t$  raisonnables.)

**Remarque 3.4 :** *Le trou spectral est rarement calculable explicitement.*

**Exemple 3.1 :** *Marche aléatoire sur un segment :*

Soit  $\chi = \{0, \dots, n\}$ . On considère la matrice de transition:  $\mathbf{K}(x, y) = 1/2$  si  $y = x \pm 1$ ,  $(x, y) = (0, 0)$  ou  $(n, n)$  et  $\mathbf{K}(x, y) = 0$  sinon.

On cherche à résoudre l'équation aux valeurs propres :  $(\mathbf{I} - \mathbf{K})X = \lambda X$ . En notant  $X = (x_j)_{0 \leq j \leq n}$ , l'équation devient :

$$\begin{cases} \frac{1}{2}x_0 - \frac{1}{2}x_1 = \lambda x_0 \\ \forall n \geq 1, -\frac{1}{2}x_{n-1} + x_n - \frac{1}{2}x_{n+1} = \lambda x_n \\ -\frac{1}{2}x_{n-1} + \frac{1}{2}x_n = \lambda x_n \end{cases}$$

soit

$$\begin{cases} x_1 = (1 - 2\lambda)x_0 \\ \forall n \geq 1, x_{n+1} + 2(\lambda - 1)x_n + x_{n-1} = 0 \\ x_{n-1} = (1 - 2\lambda)x_n \end{cases}$$

On introduit le polynôme caractéristique  $P$  de la relation de récurrence  $x_{n+1} + 2(\lambda - 1)x_n + x_{n-1} = 0$  :  $P(x) = x^2 + 2(\lambda - 1)x + 1$ .

On a  $\Delta' = \lambda(\lambda - 2)$ . On cherche donc, pour commencer, les valeurs propres  $\lambda \in [0, 2]$  ( $\Delta' \leq 0$ ). On peut donc écrire  $\lambda$  sous la forme  $\lambda = 1 - \cos(\theta)$ .

Alors,  $P(x) = x^2 - 2\cos(\theta)x + 1 = (x - e^{i\theta})(x - e^{-i\theta})$ . On a donc  $\forall p \in [0, n]$ ,  $x_p = a_0 e^{ip\theta} + e^{-ip\theta}$ . Les conditions :

$$\begin{cases} x_1 = (1 - 2\lambda)x_0 \\ x_{n-1} = (1 - 2\lambda)x_n \end{cases}$$

donnent

$$\begin{cases} a_0(1 - e^{-i\theta}) + b_0(1 - e^{i\theta}) = 0 \\ a_0 e^{in\theta}(1 - e^{i\theta}) + b_0 e^{-in\theta}(1 - e^{-i\theta}) = 0 \end{cases}$$

L'équation aux valeurs propres admet une solution non nulle si et seulement si le système ci-dessus admet une solution  $(a_0, b_0) \neq (0, 0)$ , c'est à dire si et seulement si le déterminant du système est nul. Après calcul, on obtient :

$$e^{2in\theta} = 1$$

Donc  $\exists j \in [0, n]$ ,  $\theta = \frac{j\pi}{n+1}$ . Finalement, on obtient  $n + 1$  valeurs propres distinctes  $(1 - \cos(\frac{j\pi}{n+1}))_{0 \leq j \leq n}$ , donc le trou spectral  $\lambda$  vérifie  $\lambda = 1 - \cos(\frac{\pi}{n+1})$ .

**Exemple 3.2 :** On prend maintenant  $\chi = \{0, 1\}^n$  et  $\mathbf{K}(x, y) = 0$  sauf si  $\sum_i |x_i - y_i| = 1$  (dans ce cas,  $\mathbf{K}(x, y) = \frac{1}{n}$ ). On vérifie aisément que  $(\chi_y : x \rightarrow (-1)^{y \cdot x})_{y \in \{0, 1\}^n}$  forme une base orthonormale de vecteurs propres associés aux valeurs propres  $(2|y|/n)_{y \in \{0, 1\}^n}$  donc  $\lambda = 2/n$ .

### 3.3 Une méthode d'approximation du trou spectral

Il existe de nombreuses méthodes d'approximation; dans cette partie, nous en présentons une particulièrement simple.

**Théorème 3.3 :**

Soit  $(\mathbf{K}, \pi), (\mathbf{K}', \pi')$  deux chaînes de Markov définies respectivement sur les ensembles finis  $\chi$  et  $\chi'$ . On suppose qu'il existe une application linéaire

$$\begin{cases} l^2(\chi, \pi) \rightarrow l^2(\chi', \pi') \\ f \rightarrow \tilde{f} \end{cases}$$

et qu'il existe des constantes  $A, B, a > 0$  telles que :

$$\forall f \in l^2(\chi, \pi), E'(\tilde{f}, \tilde{f}) \leq AE(f, f), aVar_{\pi}(f) \leq Var_{\pi'}(\tilde{f}) + BE(f, f).$$

$$\text{On a alors : } \frac{a\lambda'}{A+B\lambda} \leq \lambda$$

*Preuve :* Le théorème découle immédiatement de la série d'inégalités :

$$\begin{aligned} aVar_{\pi}(f) &\leq Var_{\pi'}(\tilde{f}) + BE(f, f) \\ &\leq \frac{1}{\lambda'} E'(\tilde{f}, \tilde{f}) + BE(f, f) \\ &\leq \left(\frac{A}{\lambda'} + B\right) E(f, f) \end{aligned}$$

**Exemple 3.3 :** Soit  $\chi = \{0, 1\}^n$ . On pose  $|x - y| = \sum_i |x_i - y_i|$ . Soit  $\tau : \chi \rightarrow \chi$  l'application définie par  $\tau(x) = y$  avec :

$$y_i = x_{i-1}, 1 < i \leq n, y_1 = x_n.$$

On considère la chaîne :

$$\mathbf{K}(x, y) = \begin{cases} 1/(n+1) & \text{si } |x - y| = 1 \\ 1/(n+1) & \text{si } y = \tau(x) \\ 0 & \text{sinon} \end{cases}$$

On vérifie aisément que la distribution uniforme  $\pi \equiv 2^{-n}$  est la mesure stationnaire de  $\mathbf{K}$ . Nous allons étudier cette chaîne, par comparaison avec celle de l'exemple 3.2. Manifestement, avec les notations précédentes, on a :

$$E' \leq \frac{1}{n+1} E(f, f)$$

Par application du théorème, comme  $\lambda' = 2/n$  (calculé dans l'exemple 3.2), on obtient :

$$\lambda \geq \frac{2}{n+1}.$$

## 4 Application d'une méthode graphique pour estimer le trou spectral

Le but de ce chapitre est d'introduire un outil géométrique permettant d'estimer l'écart spectral.

### 4.1 Définitions de base

Soit  $\mathbf{K}$  une chaîne de Markov irréductible sur un ensemble fini  $\chi$ .

On dit qu'un sous-ensemble ( $\mathbb{A} \subset \chi \times \chi$ ) est adapté à  $\mathbf{K}$  si  $\mathbb{A}$  est symétrique (c'est à dire  $(x, y) \in \mathbb{A} \implies (y, x) \in \mathbb{A}$ ),  $(\chi, \mathbb{A})$  est connexe et  $(x, y) \in \mathbb{A} \implies \pi(x)\mathbf{K}(x, y) + \pi(y)\mathbf{K}(y, x) > 0$ .

Soit  $\mathbf{K}$  une chaîne de Markov irréductible sur  $\chi$  avec pour mesure stationnaire  $\pi$ . Pour  $e = (x, y) \in \chi \times \chi$ , on pose:  $\nabla f(e) = f(y) - f(x)$  et

$$Q(e) = \left(\frac{1}{2}(\mathbf{K}(x, y)\pi(x) + \mathbf{K}(y, x)\pi(y))\right).$$

On remarque que par définition,  $E(f, f) = \frac{1}{2} \sum_{e \in \chi \times \chi} |\nabla f(e)|^2 Q(e)$ .

Un chemin  $\gamma$  dans  $(\chi, \mathbb{A})$  est une suite finie  $(x_0, \dots, x_k)$  qui vérifie  $(x_{i-1}, x_i) \in \mathbb{A}$ , pour  $i = 1, \dots, k$  la longueur du chemin est  $|\gamma| = k$ . Soit  $\Gamma$  l'ensemble de tous les chemins  $\gamma$  de  $(\chi, \mathbb{A})$  sans couple répété. Pour toute paire  $(x, y) \in \chi \times \chi$ , on pose  $\Gamma(x, y) = \{\gamma = (x_0, \dots, x_k) \in \Gamma : x = x_0, x = x_k\}$ .

On appelle inégalité de Poincaré une inégalité du type  $\forall f, \text{Var}_\pi(f) \leq C\varepsilon(f, f)$ . Donc, d'après la définition du trou spectral, c'est équivalent à  $\lambda \leq \frac{1}{C}$ . On utilise donc ces inégalités pour minorer  $\lambda$ .

#### Définition 4.1 :

Une fonction-poids  $\omega$  est une fonction positive

$$\omega : \mathbb{A} \longrightarrow (0, \infty).$$

La  $\omega$  – longueur d'un chemin  $\gamma$  dans  $\Gamma$  est

$$|\gamma|_\omega = \sum_{e \in \gamma} \frac{1}{\omega(e)}.$$

Pour  $\omega \equiv 1$ , la longueur d'un chemin coïncide avec la longueur usuelle.

### 4.2 Théorème

#### Théorème 4.1 :

Soit  $\mathbf{K}$  une chaîne irréductible avec pour mesure stationnaire  $\pi$  sur un ensemble fini  $\chi$ . Soit  $\mathbb{A}$  un sous-ensemble adapté de couples et  $\omega$  une fonction-poids. Pour tout  $(x, y) \in \chi \times \chi$ , on choisit un chemin  $\gamma(x, y) \in \Gamma(x, y)$ . Alors

$\lambda \geq \frac{1}{A(\omega)}$ , où

$$A(\omega) = \max_{e \in \mathbb{A}} \left\{ \frac{\omega(e)}{Q(e)} \sum_{x, y, \gamma(x, y) \ni e} |\gamma(x, y)|_{\omega} \pi(x) \pi(y) \right\}.$$

*Preuve :* On introduit la fonction  $f$  telle que  $\forall (x, y) \in \chi \times \chi$ , on a :

$$f(y) - f(x) = \sum_{e \in \gamma(x, y)} \nabla f(e),$$

et en utilisant Cauchy-Schwarz, on obtient :

$$\begin{aligned} |f(x) - f(y)|^2 &\leq \sum_{e \in \gamma(x, y)} (\omega(e))^{-1} \sum_{e \in \gamma(x, y)} \omega(e) |\nabla f(e)|^2 \\ &\leq |\gamma(x, y)|_{\omega} \sum_{e \in \gamma(x, y)} \omega(e) |\nabla f(e)|^2. \end{aligned}$$

En multipliant par  $\frac{1}{2} \pi(x) \pi(y)$ , et, en sommant sur tous les  $x$  et  $y$ , on obtient :

$$\frac{1}{2} \sum_{x, y} |f(x) - f(y)|^2 \pi(x) \pi(y) \leq \frac{1}{2} \sum_{x, y} |\gamma(x, y)|_{\omega} \sum_{e \in \gamma(x, y)} \omega(e) |\nabla f(e)|^2 \pi(x) \pi(y)$$

Le membre de gauche est égal à  $Var_{\pi}(f)$ . Le membre de droite vaut :

$$\frac{1}{2} \sum_{e \in \mathbb{A}} \left\{ \frac{\omega(e)}{Q(e)} \sum_{x, y, \gamma(x, y) \ni e} |\gamma(x, y)|_{\omega} \pi(x) \pi(y) \right\} |\nabla f(e)|^2 Q(e)$$

d'où :

$$\frac{Var_{\pi}(f)}{E(f, f)} \leq A(\omega).$$

### 4.3 Exemple

Nous considérons une marche aléatoire sur deux grilles carrées de longueur  $n$  connectées en un de leurs coins. Ainsi  $\chi$  est la réunion de  $\{0, \dots, n\}^2$  et  $\{-n, \dots, 0\}^2$ . On considère la matrice de transition markovienne :

$$K(x, y) = \begin{cases} 0 & \text{si } |x - y| > 1 \\ \frac{1}{4} & \text{si } |x - y| = 1 \\ 0 & \text{si } x = y \text{ à l'intérieur,} \\ \frac{1}{4} & \text{si } x = y \text{ sur un bord mais pas sur un coin,} \\ \frac{1}{2} & \text{si } x = y \text{ sur un coin,} \end{cases}$$

C'est tout simplement une marche aléatoire sur  $\xi$ . Il s'agit de minorer l'écart spectral pour avoir une idée de la vitesse de convergence de cette chaîne de Markov. La probabilité stationnaire est la probabilité uniforme  $\pi \equiv \frac{1}{2(n+1)^2-1}$ , et  $\frac{1}{Q(e)} = 4(2(n+1)^2 - 1)$  si  $e \in \mathbb{A}$ . ( $\mathbb{A} = \{(x, y), |x - y| = 1\}$ )

Nous allons maintenant étudier une certaine catégorie de chemins. Pour des raisons pratiques, pour l'instant, nous considérons des points à coordonnées positives. La structure de graphe sur  $\chi$  induit une distance  $d(x, y)$ . Nous considérons aussi la distance euclidienne  $|x - y|$ .

Nous définissons maintenant les chemins géodésiques  $\gamma(x, 0)$  de  $x$  à 0 tels que  $\forall z \in \gamma(x, 0)$ , la distance euclidienne entre  $z$  et le segment  $[0, x]$  est au plus  $\frac{1}{\sqrt{2}}$ . Soit  $e = (u, v) \in \mathbb{A}$  avec  $d(0, v) = i, d(0, u) = i + 1$ .

On peut montrer, par un dénombrement basé sur de la géométrie élémentaire, l'inégalité:

$$\#\{x : \gamma(x, 0) \ni e\} \leq \frac{8(n+1)^2}{i+1}$$

À partir de là, on définit le chemin  $\gamma(x, y)$  entre  $x$  et  $y$  en mettant bout à bout  $\gamma(x, 0)$  et  $\gamma(y, 0)$ . On considère la fonction de poids  $w$ , définie par  $w(e) = i + 1$  si  $e$  est à la distance  $i$  de 0. On a alors:

$$|\gamma(x, y)|_\omega \leq 2 \sum_0^{2n-1} \frac{1}{i+1} \leq 2 \ln(2n+1)$$

et

$$\begin{aligned} \#\{(x, y) : \gamma(x, y) \ni e\} &\leq (2(n+1)^2 - 1) \#\{z : \gamma(z, 0) \ni e\} \\ &\leq 4(n+1)^2(2(n+1)^2 - 1)/(i+1) \end{aligned}$$

Le théorème précédent permet alors de conclure:

$$\begin{aligned} A &\leq \frac{4 \max_{x,y} |\gamma(x, y)|_\omega}{(2(n+1)^2 - 1)} \max_e \{w(e) \#\{(x, y) : \gamma(x, y) \ni e\}\} \\ &\leq 32(n+1)^2 \ln(2n+1) \end{aligned}$$

ce qui implique

$$\lambda \geq (32(n+1)^2 \ln(2n+1))^{-1}$$

Vérifions enfin que notre minoration a le bon ordre de grandeur: on utilise la fonction test  $f$  définie par  $f(x) = \text{sgn}(x) \ln(1 + d(0, x))$  où  $\text{sgn}(x)$  vaut 1, 0, -1 selon le signe de la somme des coordonnées de  $x$ . On a  $\pi(f) = 0$  et

$$\|f\|_2^2 \geq \frac{n(n+1)}{2(n+1)^2 - 1} (\ln(n+1))^2$$

, et

$$\begin{aligned} E(f, f) &\leq \frac{1}{2(2(n+1)^2 - 1)} \sum_0^{2n-1} ((i+1) \wedge (2n-i+1)) (\ln(i+2) - \ln(i+1))^2 \\ &\leq \frac{1}{2(2(n+1)^2 - 1)} \sum_0^{n-1} \frac{1}{i+1} \\ &\leq \frac{\ln(n+1)}{2(n+1)^2 - 1} \end{aligned}$$

On a donc  $\lambda \leq \frac{1}{n(n+1)\ln(n+1)}$ , ce qui montre que l'inégalité précédente a le bon ordre de grandeur.

## 5 Conclusion

Nous avons donc associé à une chaîne de Markov  $\mathbf{K}$  irréductible une chaîne à temps continu, à savoir  $\mathbf{H}_t = \exp^{-t(\mathbf{I}-\mathbf{K})}$ , pour lequel nous avons démontré l'équivalence des comportements asymptotiques dans le cas réversible. Grâce à la notion de trou spectral, nous avons montré la convergence exponentielle des coefficients de  $\mathbf{H}_t$  vers la mesure stationnaire.

Enfin, dans la dernière partie, nous avons présenté une méthode géométrique pour minorer le trou spectral, rarement calculable explicitement.

### Bibliographie:

-Saloff-Coste L. (1996) *Lectures on probability Theory and Statistics.* , Bernard P. Lectures on finite Markov chains , 301-407.