DEA Algorithmique X/ENS/ENST/Paris VI/Paris VII/Paris XI
Filière Modélisation des réseaux
Cours de modélisation Markovienne

Alain Jean-Marie

11 juillet 2001

1 Introduction

L'objet de ce cours est de passer en revue quelques modèles probabilistes utilisés dans le domaine de l'*évaluation quantitative de performance* pour l'analyse des systèmes à événements discrets (en abrégé: SED) en général, et celle des réseaux en particulier. D'autres applications sont possibles dans le domaine de l'informatique et l'algorithmique comme l'analyse d'algorithmes probabilistes, séquentiels ou distribués.

La modélisation probabiliste des SED est un domaine vaste qui englobe de nombreuses techniques mathématiques et de nombreux résultats. Compte tenu de la durée de ce cours, j'ai choisi de mettre en valeur certains des aspects les plus en rapport avec le sujet principal du DEA (les algorithmes), à savoir:

- les objets combinatoires rencontrés: graphes, chemins, mots, arbres.
- les méthodes analytiques similaires à celles mises en œuvre en analyse d'algorithmes: fonctions génératrices, analyse asymptotique.

Du coup, les méthodes et résultats plus probabilistes sont peu abordés. Je renvoie à la bibliographie et au support de cours de F. Baccelli pour cet aspect du problème.

Sources/Bibliographie Les principales sources de ce cours sont:

- Pour les fonctions génératrices en général: le cours de Ph. Flajolet et B. Salvy, les livres de Sedgewick et Flajolet [12] et Wilf [16]. Pour les résultats en théorie des fonctions complexes, les livres de Titchmarsh [15] et Rudin [11].
- Pour les résultats sur les puissances de matrices, et le théorème de Perron-Frobenius, les livres de Seneta [13] et Minc [7].
- Pour les fondements en probabilités, les livres de Feller, surtout le premier volume [3].
- Pour les Chaînes de Markov en temps et espace discrets, les livres de Kemeny et Snell [5] et Kleinrock [6]. Pour les Chaînes de Markov en temps continu, les livres de Feller [3], Ross [10], ou Kleinrock [6].
- Pour les modèles de files d'attente, les livres de Takacs [14] et Kleinrock [6]. Pour les applications aux réseaux, les livres de Pujolle et Fdida [9].
- Pour les processus de branchement, le livre de Harris [4].
- Pour les réseaux de Petri, l'article de synthèse de Murata [8], les livres de Baccelli *et al.* [2] (pour le point de vue algébrique) et de Ajmone Marsan *et al.* [1] (pour le point de vue Markovien).

2 Préliminaires

La liaison entre les techniques de l'analyse d'algorithme et celles de l'évaluation de performances (systèmes à événements discrets stochastiques) est due en grande part à l'utilisation de *fonctions (ou: séries) génératrices.*

Soit X une variable aléatoire discrète, à valeurs dans $\mathbb{N}$. On définit sa *fonction génératrice* par:

$$F_X(z) \ = \ \mathbb{E}(z^X) \ = \ \sum_{n=0}^{\infty} \mathbb{P}(X=n) z^n \ . \tag{1}$$

Une fonction génératrice de probabilités a les propriétés suivantes:

- $F_X(z)$ existe sur le disque unité fermé: $\{z : |z| \leq 1\}$. En d'autres termes, le rayon de convergence de la série est au moins égal à 1.
- $F_X(z)$ est analytique dans le disque unité ouvert et continue sur le cercle unité. En particulier:

$$\lim_{z\to 1} F_X(z) \ = \ 1.$$

- En tant que fonction réelle, $F_X(x)$ est C^∞ sur $]-1,1[$, croissante et convexe sur $[0,1]$.
- Les probabilités de X sont les coefficients du développement en série entière de F_X en $z=0$. En utilisant la notation $[z^n]f(z)$ pour «le coefficient de z^n dans le développement en série de $f(z)$», on a:

$$\mathbb{P}(X=n) \ = \ [z^n] \ F_X(z) \ .$$

  Pratiquement, on peut utiliser trois méthodes:

  - calculer le développement en série de F_X à partir de développements connus. Ainsi, s'il est su que $F_X(z) = 2/(2-z)(3-z)$, on a:

$$\mathbb{P}(X=n) \ = \ [z^n] \left(\frac{2}{2-z} - \frac{2}{3-z} \right) \ = \ [z^n] \frac{1}{1-z/2} - [z^n] \frac{2/3}{1-z/3} \ = \ 2^{-n} - \frac{2}{3}\, 3^{-n} \ .$$

  - calculer le développement en série de F_X par la *formule de Taylor*:

$$\mathbb{P}(X=n) \ = \ \frac{1}{n!} \ \frac{d^n}{dz^n} F_X(z) \bigg|_{z=0} \ .$$

  - utiliser la *formule de Cauchy*:

$$\mathbb{P}(X=n) \ = \ \frac{1}{2i\pi} \int_{\mathcal{C}} F_X(z) \ \frac{\mathrm{d}z}{z^{n+1}} \ ,$$

    où $\mathcal{C}$ est un contour simple quelconque entourant le point $z=0$ dans le sens trigonométrique, et inclus dans le domaine d'analyticité de F_X.

- Si le k-ième moment de X existe, on a:

$$\mathbb{E}(X(X-1)\dots(X-k)) \ = \ \frac{d^n}{dz^n} F_X(z) \bigg|_{z=1} \ .$$

  Ce nombre s'appelle le k-ième *moment factoriel* de X. Les vrais moments s'en déduisent par combinaisons linéaires (voir: nombres de Stirling de première espèce).
- Si deux variables aléatoires X et Y sont indépendantes, alors:

$$F_{X+Y}(z) \ = \ \mathbb{E}(z^{X+Y}) \ = \ \mathbb{E}(z^X z^Y) \ = \ \mathbb{E}(z^X)\ \mathbb{E}(z^Y) \ = \ F_X(z)\ F_Y(z) \ .$$

Assez souvent, on a à considérer des séries de nombres qui sont des probabilités mais pas les probabilités d'une même distribution. Dans ce cas, on aura usage des propriétés suivantes:

Lemme 1. *Soit $\{a_n, n \in \mathbb{N}\}$ une suite de nombres complexes bornés. Alors la fonction*

$$f(z) \ = \ \sum_{n=0}^{\infty} a_n z^n$$

est analytique dans le cercle unité: $\{z \in \mathbb{C} \mid |z| < 1\}$.

Lemme 2 (Théorème de Pringsheim). *Soit $\{a_n, n \in \mathbb{N}\}$ une suite de nombres positifs, et $f(z) = \sum_{n=0}^{\infty} a_n z^n$. Si le rayon de convergence de cette série est $R < \infty$, alors $z = R$ est une singularité (dite: singularité dominante).*

3 Puissances de Matrices

Dans cette partie, on étudiera la suite des puissances d'une matrice, du point de vue du calcul, et du point de vue asymptotique. On établira que le comportement asymptotique est directement lié au *rayon spectral* de la matrice. Dans un deuxième temps, on regardera de près le cas des matrices positives.

Spectre et rayon spectral. Soit $A \in \mathcal{M}_{N\times N}(\mathbb{C})$ une matrice carrée à coefficients complexes. On rappelle que λ est une valeur propre de A si il existe un vecteur $x \neq 0$ tel que $Ax = \lambda x$. Un tel vecteur est dit vecteur propre droit associé à la valeur propre λ. Pour chaque valeur propre, il y a également des vecteurs propres gauches. L'ensemble de toutes les valeurs propres est appelé le *spectre* de la matrice. On le notera $\mathrm{sp}(A)$. Le *rayon spectral* de A est défini par:

$$\rho(A) \;=\; \max\{|\lambda| \mid \lambda \in \mathrm{sp}(A)\ \}\ .$$

On notera dans cette section les *différentes* valeurs propres par: $\lambda_1, \ldots, \lambda_p$, avec multiplicités respectives $\alpha_1, \ldots, \alpha_p$. On a donc: $\sum \alpha_i = N$.

Les valeurs propres de A sont les racines de son polynôme caractéristique:

$$\chi_A(x) \;=\; \det(A - xI) \;=\; \prod_{i=1}^{p} (\lambda_i - x)^{\alpha_i}\ . \tag{2}$$

Si $\alpha_i = 1$, on dit que la veleur propre λ_i est *simple*.

3.1 Matrices et graphes

Un *graphe orienté valué* est un triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ tel que $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ et $\mathcal{W} : \mathcal{E} \to \mathbb{C} \setminus \{0\}$.

À tout graphe valué on associe une matrice carrée réelle $A[\mathcal{G}]$, en numérotant les sommets de $\mathcal{V}$: $\mathcal{V} = \{v(1), \ldots, v(N)\}$ et en assignant la valeur $A_{i,j} = \mathcal{W}(v(i), v(j))$ si $(i,j) \in \mathcal{E}$ et $A_{i,j} = 0$ sinon.

Réciproquement, à toute matrice A carrée $N \times N$, on associe naturellement un graphe valué $\mathcal{G}[A]$:

- Les sommets de $\mathcal{V}$ sont $\{1, 2, \ldots, N\}$,
- L'arc (i,j) est dans $\mathcal{E}$ si $A_{i,j} \neq 0$.
- Cet arc a alors la valeur (ou: poids) $\mathcal{W}(i,j) = A_{i,j}$.

Le fait d'associer un graphe à une matrice permet d'énoncer quelques concepts importants dans le langage de la théorie des graphes.

Irréducibilité On dira qu'une matrice est *irréductible* (parfois: indécomposable) si son graphe associé $\mathcal{G}$ est fortement connexe. Dans le cas contraire, la matrice est réductible ou décomposable.

Périodicité Pour un sommet i du graphe $\mathcal{G}[A]$, soit $p(i)$ le p.g.c.d de toutes les longueurs des cycles de ce graphe passant par i. Le sommet i est dit *périodique de période* $d(i)$.

Soit d le p.p.c.m. des $d(i)$. Si $d \geq 2$, on dira que la matrice A est *périodique de période* d. Si $d = 1$, on dira que A est *apériodique*.

3.2 Matrices et Chemins

Un chemin d'un graphe $\mathcal{G}$ est une suite de sommets de ce graphe $\gamma = (i_1, \ldots, i_n)$ telle que $(i_j, i_{j+1}) \in \mathcal{E}$ pour $j \in \{1, \ldots, n-1\}$. Notre convention sera que la longueur d'un tel chemin, notée $|\gamma|$ est le nombre d'arêtes qu'il contient, soit $n - 1$. Ainsi, un chemin réduit à un seul sommet est de longueur 0.

Quand le graphe $\mathcal{G}$ est valué, on associe à tout chemin γ son poids (multiplicatif) $w(\gamma)$ comme suit:

$$\gamma = (i_1, \ldots, i_n) \quad\Longrightarrow\quad w(\gamma) \;=\; \prod_{j=1}^{n-1} \mathcal{W}(i_j, i_{j+1})\ . \tag{3}$$

Par convention, $w(\gamma) = 1$ si $|\gamma| = 0$.

Soit alors $\Gamma^n(i,j)$ l'ensemble des chemins de longueur n allant de i à j. Si A est la matrice associée au graphe $\mathcal{G}$, on a l'identité:

$$(A^n)_{i,j} \;=\; \sum_{\gamma \in \Gamma^n(i,j)} w(\gamma) \;=\; \sum_{(i_1,\ldots,i_n)\in\Gamma^n(i,j)} \mathcal{W}(i_1,i_2)\mathcal{W}(i_2,i_3)\ldots\mathcal{W}(i_{n-1},i_n)\ . \tag{4}$$

Noter la cohérence avec les conventions ci-dessus dans le cas $n=0$: l'ensemble $\Gamma^0(i,j)$ des chemins sans arcs allant de i à j est: a/ vide si $i \neq j$, et la somme vide valant 0 par convention, $A^0_{i,j}=0$; b/ réduit à (i,i) si $i=j$ et dans ce cas la somme vaut 1. On trouve bien $A^0 = \mathbf{I}$.

Dans de nombreux cas, il est intéressant d'associer à chaque arc (i,j) de $\mathcal{E}$ on une mesure *additive* $\mathcal{D}(i,j)$, que je nommerai ici "distance" telle que pour le chemin γ de (3), sa distance $d(\gamma)$ vaut:

$$d(\gamma) \;=\; \sum_{j=1}^{n-1} \mathcal{D}(i_j,i_{j+1})\ .$$

Dans ce cas, si on définit:

$$\mathcal{W}(i,j) \;=\; z^{\mathcal{D}(i,j)}\ ,$$

alors $w(\gamma)=z^{d(\gamma)}$. Puisque la matrice $A[\mathcal{G}]$ dépend du paramètre z, notons-la $A(z)$. On a:

$$(A(z)^n)_{i,j} \;=\; \sum_{\gamma\in\Gamma^n(i,j)} z^{d(\gamma)} \;=\; \sum_d z^d \times \#\{\text{chemins de longueur } n \text{ et de distance } d \text{ de } i \text{ à } j\}\ . \tag{5}$$

On obtient donc les séries génératrices de la distance des chemins allant de i à j ayant la longueur n. Nous reviendrons sur une généralisation de cette construction dans le paragraphe 4.4.

3.3 Puissances de matrices: exemples

3.3.1 Graphes et nombre de chemins

Soit $\mathcal{G}$ un graphe. La *matrice d'adjacence* de $\mathcal{G}$ est la matrice $A[\mathcal{G}]$ correspondant à la valuation des arcs: $\mathcal{W}(i,j)=1\ \forall (i,j)\in\mathcal{E}$.

Dans ce cas, tout chemin a un poids 1. L'interprétation de (4) est alors:

$$(A^n)_{i,j} \;=\; \#\Gamma^n(i,j) \;=\; \text{nombre de chemins de longueur } n \text{ de } i \text{ à } j\ . \tag{6}$$

Si on choisit la valuation:

$$\mathcal{W}(i,j)=1,\quad (i,j)\in\mathcal{E} \text{ ou } i=j,\qquad \mathcal{W}(i,j)=0 \text{ sinon}\ ,$$

on obtient pour $A[\mathcal{G}]$ la matrice d'adjacence du graphe obtenu de $\mathcal{G}$ en rajoutant les arcs (i,i) si nécessaire, et alors

$$(A^n)_{i,j} \;=\; \sum_{k=0}^{n} \#\Gamma^k(i,j) \;=\; \text{nombre de chemins de longueur} \le n \text{ de } i \text{ à } j\ . \tag{7}$$

3.3.2 Graphes et longueur de chemins

Si on admet que les valuations $\mathcal{W}$ prennent leurs valeurs dans $\mathbb{R}_{\max}$ ou $\mathbb{R}_{\min}$ et que les sommes et produits se comprennent dans l'algèbre correspondante (voir le cours du DEA ALGO sur les dioïdes) alors (4) s'interprète comme (par exemple pour $\mathbb{R}_{\max}$):

$$\begin{aligned}(A^n)_{i,j} \;&=\; \bigoplus_{\gamma\in\Gamma^n(i,j)} \mathcal{W}(i_1,i_2) \;+\; \mathcal{W}(i_2,i_3) \;+\; \ldots \;+\; \mathcal{W}(i_{n-1},i_n) \\ &=\; \text{distance sur le plus long des chemins de } n \text{ arêtes } i \text{ à } j\ .\end{aligned}$$

Similairement, dans $\mathbb{R}_{\min}$, on obtient le plus court chemin de n arêtes.

3.3.3 Récurrences linéaires homogènes

Soient $a_1, \dots, a_p$ des nombres complexes. Soit $\{x_n\}_{n\in\mathbb{N}}$ la suite définie par:

$$\begin{array}{rcll} x_n & = & a_1\, x_{n-1} \;+\; a_2\, x_{n-2} \;+\; \dots \;+\; a_p\, x_{n-p} & n \geq p \\ & & x_0, x_1, \dots, x_{p-1} \text{ donnés.} & \end{array}$$

Si on définit

$$y_n \;=\; \begin{pmatrix} x_{n-1} \\ \vdots \\ x_{n-p} \end{pmatrix} \qquad A \;=\; \begin{pmatrix} a_1 & a_2 & \dots & & a_p \\ 1 & 0 & \dots & & \\ 0 & 1 & \ddots & & \\ & & \ddots & & \\ & & & 1 & 0 \end{pmatrix}$$

alors les vecteurs y_n se calculent par récurrence:

$$y_n \;=\; A\, y_{n-1} \;=\; \dots \;=\; A^{n-p} y_p \;.$$

À l'ordre des lignes et des colonnes près, la matrice A est la matrice compagnon du polynôme

$$X^p - a_1 X^{p-1} - \dots - a_{p-1} X - a_p \;.$$

3.4 Puissances de matrices: asymptotiques et calcul exact

On s'intéresse donc à A^n, la puissance n-ième de A. On montre d'abord que les coefficients de A^n croissent ou décroissent, en module, au plus exponentiellement en n. Soit

$$\mu(A) \;=\; \max\{|a_{ij}|,\; 1 \leq i, j \leq N\} \;.$$

Lemme 3. *Il existe une constante c telle que*

$$\mu(A^n) \;\leq\; c(N\mu(A))^n \;. \tag{8}$$

La preuve découle facilement du fait que $\mu(AB) \leq N\mu(A)\mu(B)$.

La borne (8) s'avère plutôt grossière en général! Le résultat suivant donne la valeur exacte de A^n en fonction des valeurs propres de A. Pour cette raison, on nomme ce type de résultat une *décomposition spectrale.*

Théorème 4 (Décomposition Spectrale). *i/ Soit A une matrice dont les valeurs propres sont $\lambda_1, \dots, \lambda_p$, avec multiplicités respectives $\alpha_1, \dots, \alpha_p$. Alors il existe p matrices $B_i(n, \lambda_i)$, $1 \leq i \leq p$, telles que chaque coefficient de B_i est un polynôme en λ_i de degré au plus n, dont les coefficients sont des polynômes en n de degré au plus $\alpha_i - 1$, et telles que:*

$$A^n \;=\; \sum_{i=1}^{p} B_i(n, \lambda_i) \;. \tag{9}$$

ii/ Dans le cas où λ_i est une valeur propre simple, alors $B_i(n, \lambda_i) = \lambda_i^n B_i$, où B_i est constante et de rang 1: il existe deux vecteurs propres gauches et droits pour la valeur propre λ_i, v_i et w_i, tels que $v_i.w_i = 1$ et $B_i = w_i.v_i$.

Preuve. Ce théorème peut se montrer de deux manières différentes dans la forme (mais pas tant que ça dans le fond). Il est intéressant de les comparer.

Méthode algébrique. Celle-ci se base sur la décomposition de Jordan: il existe une matrice inversible S telle que:

$$A \;=\; S \begin{pmatrix} J_{\nu_1}(\mu_1) & & & \\ & J_{\nu_2}(\mu_2) & & \\ & & \ddots & \\ & & & J_{\nu_k}(\mu_k) \end{pmatrix} S^{-1} \quad \text{et donc:} \quad A^n \;=\; S \begin{pmatrix} J_{\nu_1} & & & \\ & J_{\nu_2}^n & & \\ & & \ddots & \\ & & & J_{\nu_k}^n \end{pmatrix} S^{-1} \,. \tag{10}$$

Ici, les matrices $J_m(\mu)$, sont des blocs de Jordan de la forme:

$$J_m(\mu) \;=\; \begin{pmatrix} \mu & 1 & & \\ & \mu & \ddots & \\ & & \ddots & 1 \\ & & & \mu \end{pmatrix}$$

et les μ_m sont des valeurs propres parmi les λ_i. Attention: il peut y avoir plus de blocs de Jordan que de valeurs propres distinctes (*cf.* la matrice identité). Le résultat s'obtient en définissant

$$B_i(n,\lambda_i) \;=\; \sum_{\mu_m=\lambda_i} S \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & J_{\nu_m}(\mu_m)^n & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} S^{-1} \,.$$

Les matrices B_i ont les propriétés requises parce que S et S^{-1} sont constantes, et les puissances d'un bloc de Jordan ont la forme (en exercice):

$$((J_n(\lambda)^m))_{i,j} \;=\; \binom{m}{j-i} \lambda^{m-(j-i)} \; \mathbf{1}_{\{j\geq i\}} \,.$$

Cette matrice est bien un polynôme en λ de degré n, et un polynôme en n de degré $m-1$. Ceci prouve $i/$.

Dans le cas où λ_i est simple, le bloc de Jordan correspondant est de taille 1. On trouve alors que B_i est le produit de la colonne de S correspondant à la localisation de ce bloc, par la ligne de S^{-1} de même numéro. Or ces deux éléments sont précisément des vecteurs propres de A pour la valeur propre λ_1, et leur produit scalaire est 1 en vertu de l'identité $S^{-1}S = \mathbf{I}$. Donc $ii/$ est prouvé.

Méthode analytique. Définissons la série génératrice matricielle

$$A(z) \;=\; \sum_{n=0}^{\infty} A^n\, z^n \;=\; ((\sum_{n=0}^{\infty} A_{i,j} z^n))_{(i,j)} \,. \tag{11}$$

D'après le paragaphe 2, et grâce au Lemme 3, cette série est bien définie car elle converge au moins dans le cercle: $\mathcal{C}_0 = \{|z| < (N\mu(A))^{-1}\}$. On peut montrer que pour tout z dans le ce cercle, la matrice $\mathbf{I} - zA$ est inversible, et que:

$$A(z) \;=\; (\mathbf{I} - zA)^{-1} \;=\; \frac{1}{\det(\mathbf{I} - zA)}\, \mathrm{ajd}(\mathbf{I} - zA) \,. \tag{12}$$

On a noté $\mathrm{adj}(A)$ la matrice adjointe de A, telle que $A\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I$. On sait que les coefficients de cette matrice sont des mineurs d'ordre $n-1$ de A. Par conséquent, les coefficients de la matrice $\mathrm{ajd}(\mathbf{I} - zA)$ sont des polynômes de z de degré au plus $N-1$. D'autre part, d'après (2):

$$\det(\mathbf{I} - zA) \;=\; \det((-z)(A - \frac{1}{z}\mathbf{I})) \;=\; (-z)^N\, \chi_A(z) \;=\; \prod_{m=1}^{p} (1 - \lambda_m z)^{\alpha_m} \,.$$

C'est un polynôme de degré N si $\lambda = 0$ n'est pas valeur propre de A. Mais si A est singulière, alors le degré est réduit de l'ordre de la valeur propre 0.

Revenant maintenant à l'équation (12), on constate que les coefficients de la matrice $A(z)$ sont des fonctions rationnelles de z dont le dénominateur est $\det(\mathbf{I} - zA)$. On peut donc faire une décomposition en éléments simples de chacun de ces coefficients. La formule générale pour une telle décomposition est, $b(z)$ étant un polynôme de degré au plus $\leq N-1$:

$$\frac{b(z)}{\prod_{m=1}^{p}(1-\lambda_m z)^{\alpha_m}} \;=\; \nu(z) \;+\; \sum_{m=1,\lambda_m\neq 0}^{p}\sum_{q=1}^{\alpha_m}\frac{b_{m,q}}{(1-\lambda_m z)^q}\;,$$

avec $\nu(z)$ un polynôme de degré au plus $N-1$. Il ne reste qu'à extraire le coefficient de z^n dans le développement de Taylor de ce rapport:

$$[z^n]\,\frac{b(z)}{\prod_{m=1}^{p}(1-\lambda_m z)^{\alpha_m}} \;=\; [z^n]\,\nu(z) \;+\; \sum_{m=1}^{p}\sum_{q=1}^{\alpha_m}[z^n]\,\frac{b_{m,q}}{(1-\lambda_m z)^q} \;=\; \nu_n \;+\; \sum_{m=1}^{p}\sum_{q=1}^{\alpha_m} b_{m,q}\binom{n+q-1}{n}\lambda_m^n\;.$$

Le coefficient de z^n dans $\nu(z)$ est bien évidemment nul si n est assez grand. Le résultat $i/$s'ensuit.

Dans le cas où λ_i est simple, on a simplement dans le développement ci-dessus:

$$[z^n]\,\frac{b}{(1-\lambda_i z)} \;=\; b\lambda_i^n\;.$$

Ceci permet de conclure que la matrice B_i est une constante que multiplie λ_i^n, mais pour le reste de l'assertion $ii/$, il faut se donner plus de mal. □

En ce qui concerne le taux de croissance des coefficients de A^n, on déduit du Théorème 4:

Corollaire 5. *Soit A une matrice carrée. Pour tout $\varepsilon > 0$, on a:*

$$\mu(A^n) \;=\; \mathcal{O}((\rho(A)+\varepsilon)^n)\;. \tag{13}$$

Preuve. Étant donné que pour tout $p(n)$, polynôme de n on a:

$$\lim_{n\to\infty}\frac{p(n)\lambda^n}{(\lambda+\varepsilon)^n} \;=\; 0\;,$$

et que pour tout i, $|\lambda_i| \leq \rho(A)$, on voit facilement que le rapport $A_{i,j}/(\lambda+\varepsilon)^n$ est borné pour tout i,j. □

Autant le Corollaire 5 semble plus grossier que le Théorème 4, il est en pratique beaucoup plus facile à utiliser dans la mesure ou le spectre complet d'une matrice (nécessaire, entre autres, dans le théorème) est beaucoup plus difficile à trouver que son rayon spectral, pour lequel il existe des bornes. Par exemple, on a le lemme:

Lemme 6. *Soit A une matrice. Son rayon spectral $\rho(A)$ a la propriété:*

$$\min_i \sum_j |A_{i,j}| \;\leq\; \rho(A) \;\leq\; \max_i \sum_j |A_{i,j}|\;. \tag{14}$$

Preuve. Soit ρ une valeur propre et u un vecteur propre associé: $u \neq 0$ et $Au = \rho_u$. Soit i_0 tel que $|u_{i_0}| = \max_i |u_i|$. Alors pour toute ligne i,

$$\begin{aligned}\rho u_i \;&=\; \sum_j A_{i,j}u_j\\ |\rho|\;|u_i| \;&\leq\; \sum_j |A_{i,j}|\;|u_j| \;\leq\; |u_{i_0}|\;\sum_j |A_{i,j}| \;\leq\; |u_{i_0}|\;\max_i\sum_j |A_{i,j}|\;.\end{aligned}$$

En particulier pour $i = i_0$, on obtient l'inégalité de droite dans (14). Pour l'inégalité de gauche, on peut utiliser la fonction f_A de Collatz-Wielandt (voir l'appendice) et sa relation avec $\rho(A)$ (Équation (67)). On obtient:

$$\rho(A) \;=\; \max_{x\geq 0} f_A(x) \;\geq\; f_A(\mathbf{1}) \;=\; \min_i \sum_j |A_{i,j}| \;.$$

□

Parmi les applications du Corollaire 5, on trouve:

Corollaire 7. *Le rayon de convergence de la série* $A(z) = \sum_{n=0}^{\infty} z^n A^n$ *est* $[\rho(A)]^{-1}$*. Quand elle converge, cette série vaut* $A(z) = (\mathbf{I} - zA)^{-1}$.

3.5 Matrices positives

Le résultat fondamental est que le spectre d'une matrice positive irréductible ne peut pas être totalement arbitraire: son elément de module maximum est nécessairement réel. Ce résultat et d'autres qui lui sont reliés font l'objet du théorème de Perron-Frobenius.

Théorème 8 (Perron-Frobenius). *Soit A une matrice carrée positive, irréductible. Alors il existe une valeur propre de A, r qui a les propriétés suivantes:*

a/ $r > 0$*;*

b/ r est associée à un vecteur propre gauche g et un vecteur propre droit d strictement positifs*;*

c/ pour toute valeur propre λ de A, $|\lambda| \leq r$*;*

d/ L'espace propre associé à r est de dimension 1;

e/ Pour toute matrice B telle que $0 \leq B \leq A$, et toute valeur propre β de B, $|\beta| \leq r$, et l'égalité $|\beta| = r$ implique que $B = A$;

f/ r est une racine simple de χ_A, le polynôme caractéristique de A.

g/ Si A a la cyclicité d, alors les valeurs propres de A de module r sont exactement les ω^j, $0 \leq j < d$ avec $\omega = e^{2i\pi/d}$.

On qualifie r de *valeur propre de Perron-Frobenius* ou bien *valeur propre principale* ou bien *valeur propre dominante*.

La preuve (partielle) de ce grand théorème est proposée en appendice.

Puisque nous sommes intéressés par les puissances de A, nous déduisons immédiatement du Théorème de Perron-Frobenius que par exemple:

Théorème 9. *Soit A une matrice carrée réelle positive, irréductible. Alors*

$$\mu(A^n) \;=\; \mathcal{O}(\rho(A)^n) \;. \tag{15}$$

Mais on a en fait un résultat plus fin, que je formule en termes de graphes, et qu'on adaptera dans peine à des matrices positives.

Corollaire 10. *Soit un graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, fortement connexe, de matrice d'adjacence A. Soit $N_{i,j}$ le nombre de chemins de longueur n allant de i à j:*

i/ $N_{i,j} = \mathcal{O}(\rho(A)^n)$.

ii/ Si $\mathcal{G}$ est fortement connexe et apériodique, alors il existe un vecteur ligne v et un vecteur colonne w strictement positifs, tels que $v.w = 1$ et $N_{i,j} \sim w_i.v_j\rho(A)^n$.

4 Chaînes de Markov en temps discret

Les chaînes de Markov constituent l'outil principal pour la modélisation et la résolution de problèmes dynamiques stochastiques.

Dans cette section, je passe en revue quelques résultats sur les chaînes de Markov. Je suppose que le lecteur est modérément familier avec ce sujet.

4.1 Définitions

Soit $\mathcal{E}$ un ensemble au plus dénombrable. Un processus $\{X(n), n \in \mathbb{N}\}$ est une *chaîne de Markov en temps discret* (en abrégé: CMTD) si et seulement si:

i/ (propriété de Markov) pour tout $t \in \mathbb{N}$, et tout $t+2-$ uplet $(j_0, j_1, \ldots, j_t, j_{t+1})$ d'éléments de $\mathcal{E}$:

$$\mathbb{P}[X(t+1) = j_{t+1} | X(t) = j_t, \ldots, X(0) = j_0] = \mathbb{P}[X(t+1) = j_{t+1} | X(t) = j_t] \; ; \tag{16}$$

On dit que cette chaîne est *homogène* si de plus:

ii/ (homogénéité) pour tout $t \in \mathbb{N}$, et toute paire (i, j) de $\mathcal{E}$,

$$\mathbb{P}[X(t+1) = j | X(t) = i] \;=\; P_{i,j} \;, \tag{17}$$

indépendamment de t.

À partir de ce point et jusqu'au paragraphe 4.3, on supposera que l'espace d'états $\mathcal{E}$ est *fini*.

4.2 Chaînes de Markov à espace d'état fini

4.2.1 Chaînes de Markov et matrices

Les nombres $P_{i,j}, (i,j) \in \mathcal{E} \times \mathcal{E}$ sont les *probabilités de transition*, et la matrice $\mathbf{P}$ dont les composantes sont les $P_{i,j}$ est la *matrice de transition*. Cette matrice est stochastique, c'est à dire dont les coefficients sont positifs et telle que, $\mathbf{1}$ étant le vecteur (colonne) constitué de "1",

$$\mathbf{P}\;\mathbf{1} \;=\; \mathbf{1} \;. \tag{18}$$

Réciproquement, étant donnée une matrice stochastique $\mathbf{P}$, il est facile de définir un processus $\{X(n), n \in \mathbb{N}\}$ qui sera une chaîne de Markov homogène ayant $\mathbf{P}$ comme matrice de transition.

Une chaîne de Markov est donc caractérisée par sa matrice de transition.

4.2.2 Chaînes de Markov et graphes.

Comme à toute matrice carrée (voir paragraphe 3.1), on associe à une chaîne de Markov un graphe qui porte le nom de *diagramme de transition*. Ce graphe a bien entendu la propriété que la somme des poids des arcs sortant d'un sommet est 1:

$$\forall i, \qquad \sum_{j | (i,j) \in V} A_{i,j} \;=\; 1,$$

À noter qu'il est habituel d'étiqueter les sommets de ce graphe par les éléments de $\mathcal{E}$ plutôt que par $\{1, 2, \ldots\}$.

Les évolutions possibles de la chaîne de Markov se font le long des chemins du graphe $\mathcal{G}$.

On a coutume de donner les qualificatifs associés à une matrice dans le paragraphe 3.1 ces qualificatifs (irréductible, périodique, apériodique, etc.) à la chaîne elle-même.

Les composantes fortement connexes du graphe $\mathcal{G}$ sont appelées *classes de communication* de la chaîne.

Une classe de communication telle que aucun arc (*i.e.* aucune transition de la chaîne) n'en sort est dite *classe terminale*.

4.2.3 Probabilités transitoires

Le premier problème est de calculer la probabilité que la chaîne se trouve dans un état donné après n unités de temps. Ce sont les « probabilités transitoires ».

Soit donc

$$p(i, j; n) \;=\; \mathbb{P}(X(n) = j \mid X(0) = i) \;, \tag{19}$$

et soit $P(n)$ la matrice faite des $p(i, j; n), (i, j) \in \mathcal{E} \times \mathcal{E}$.

Les probabilités $p(i,j;n)$ sont "données" par le théorème:
Théorème 11. *Pour tout $n \in \mathbb{N}$:*

$$P(n) \;=\; \mathbf{P}^n \;.$$

Au vu de (4), on a:
Corollaire 12. *Pour tout $n \in \mathbb{N}$, $(i,j) \in \mathcal{E} \times \mathcal{E}$, on a:*

$$p(i,j;n) \;=\; \sum_{(i_1,\dots,i_{n-1}) \,\in\, \mathcal{E}^{n-1}} p_{i,i_1}\; p_{i_1,i_2}\; \cdots\; p_{i_{n-1},j} \;. \tag{20}$$

On peut interpréter ce résultat de la manière suivante: soit un chemin de $\mathcal{G}$ de longueur n: $w = i_0, i_1, \dots, i_n$. On associe a ce chemin le poids: $p_{i_0,i_1} \dots p_{i_{n-1},i_n}$. La probabilité $p(i,j;n)$ est alors la somme des poids de tous les chemins de longueur n allant de i à j.

Soit maintenant, pour $n \in \mathbb{N}$ et $j \in \mathcal{E}$,

$$\pi_n(j) \;=\; \mathbb{P}(X(n) = j).$$

Pour cette calculer cette probabilité, il est nécessaire de spécifier la *distribution initiale* du processus, c'est à dire:

$$\pi_0(j) \;=\; \mathbb{P}(X(0) = j), \qquad j \in \mathcal{E}.$$

On a alors:

$$\pi_n(j) \;=\; \sum_{i \in \mathcal{E}} \pi_0(i)\; p(i,j;n) \;.$$

Sous forme matricielle, et en notant $\boldsymbol{\pi}_n$ le vecteur (ligne) dont les coordonnées sont les $\pi_n(j)$, on a:

$$\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}_0\; \mathbf{P}^n \;. \tag{21}$$

4.2.4 Spectre des matrices stochastiques; Résolvante

Établissons d'abord les résultats spectraux propres aux matrices stochastiques:
Théorème 13. *Soit $\mathbf{P}$ une matrice stochastique.*

i/ Le rayon spectral de $\mathbf{P}$ est 1, et c'est une valeur propre.

ii/ Si $\mathbf{P}$ est irréductible, alors il existe un unique vecteur $\boldsymbol{\pi}$ de probabilités stationnaires *pour $\mathbf{P}$, c'est-à-dire tel que:*

$$\boldsymbol{\pi} \;=\; \boldsymbol{\pi}\; \mathbf{P} \;. \tag{22}$$

Ce vecteur est strictement positif.

iii/ Si $\mathbf{P}$ est irréductible et apériodique, alors toute valeur propre λ de $\mathbf{P}$ différente de 1 est telle que $|\lambda| < 1$.

Preuve. Si $\mathbf{P}$ est irréductible, le Théorème de Perron-Frobenius s'applique et il existe une valeur propre principale. D'après le Lemme 6, $\rho(\mathbf{P}) \leq 1$ puisque la somme des lignes de $\mathbf{P}$ vaut 1. D'autre part, 1 est bien valeur propre puisque $\mathbf{P1} = \mathbf{1}$. D'après P.-F. b/, il existe des vecteurs propres strictement positifs, et d'après d/, tous les vecteurs propres en sont multiples. Il existe donc une solution et une seule à (22) telle que $\boldsymbol{\pi}\mathbf{1} = 1$. Ceci prouve *ii/*.

Si $\mathbf{P}$ est réductible, alors $\mathbf{P}$ est équivalente à une matrice *blocs* (voir paragraphe 4.2.6):

$$\begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \\ \mathbf{0} & \mathbf{P}_{22} & \ddots & \\ & \mathbf{0} & \ddots & \\ & & \mathbf{0} & \mathbf{P}_{pp} \end{pmatrix} . \tag{23}$$

Le spectre de $\mathbf{P}$ est la réunion de ceux des matrices *irréductibles* de la diagonale: $\mathbf{P}_{11}, \ldots, \mathbf{P}_{pp}$. Ces matrices sont soit stochastiques ($\mathbf{P}_{pp}$ l'est obligatoirement), soit *sous-stochastiques*, c'est-à-dire telle que $\mathbf{P1} \leq \mathbf{1}$. Leur rayon spectral est alors soit 1 (en vertu de *ii/*), soit strictement inférieur à 1, en vertu de P.-F. e/. Ceci prouve *i/*.

L'assertion *iii/* découle de P.-F. g/.

□

D'après le Corollaire 7 et puisque $\rho(\mathbf{P}) = 1$, on sait que la série génératrice des puissances de $\mathbf{P}$:

$$P(z) \;=\; \sum_{n=0}^{\infty} \mathbf{P}^n \; z^n \tag{24}$$

converge dans le cercle unité et y vaut $(\mathbf{I} - z\mathbf{P})^{-1}$. La matrice $P(z)$ porte le nom de *résolvante*. D'après le Théorème de Pringsheim, on sait même que $z = 1$ est une singularité de la série.

La série de matrices $\mathbf{S} = P(1) = \sum_{n=0}^{\infty} \mathbf{P}^n$ converge dans $\mathbb{R} \cup \{+\infty\}$. Elle se nomme *série potentielle*. Puisque $P(z)$ a une singularité en $z = 1$, un élément au moins de cette matrice est infini.

Si $\mathbf{P}$ est ergodique (voir plus bas), on déduit du Théorème 15 que le résidu de $P(z)$ en $z = 1$ est la matrice $\mathbf{1}.\boldsymbol{\pi}$.

4.2.5 Comportement asymptotique

La question suivante est de savoir comment se comportent les nombres $p(i, j; n)$ et $\pi_n(j)$ quand n tend vers l'infini. Ou encore: quel est le comportement de $\mathbf{P}^n$ et $\boldsymbol{\pi}_0\mathbf{P}^n$ quand $n \to \infty$. On est donc en présence du problème étudié dans le Paragraphe 3.

Le premier résultat concerne la convergence des moyennes de Cesaro des puissances de $\mathbf{P}$ et des probabilités stationnaires.

Théorème 14. *Soit $\mathbf{P}$ une matrice stochastique irréductible et $\boldsymbol{\pi}$ son vecteur de probabilités stationnaires. Alors*

$$\sum_{m=0}^{n-1} \mathbf{P}^m \;=\; n\;\boldsymbol{\pi} + \mathcal{O}(1) \;,$$

et donc en particulier:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{P}^m \;=\; \mathbf{1}.\boldsymbol{\pi} \;. \tag{25}$$

Par conséquent, quel que soit le vecteur π_0, on a:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \boldsymbol{\pi}_m \;=\; \boldsymbol{\pi} \;. \tag{26}$$

Preuve. Considérons la série génératrice de sommes des puissances de $\mathbf{P}$:

$$\begin{aligned}
\sum_{n=0}^{\infty} z^n \sum_{m=0}^{n-1} \mathbf{P}^m \;&=\; \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} z^n \mathbf{P}^m \\
&=\; \frac{1}{1-z}\; P(z) \;,
\end{aligned}$$

où $P(z)$ est la résolvante, définie en (24). D'après la preuve "analytique" du Théorème 4 (de décomposition spectrale), et sachant que 1 est une valeur propre isolée de $\mathbf{P}$ (d'après le Théorème de Perron-Frobenius), on peut écrire:

$$P(z) \;=\; \frac{1}{1-z} \boldsymbol{\pi} \cdot \mathbf{1} \;+\; \sum_{\lambda_i \neq 1} \sum_{q=1}^{\alpha_i} \frac{B_{i,q}}{(1-\lambda_i z)^q} \;.$$

Par conséquent,

$$\frac{1}{1-z}\;P(z) \;=\; \frac{1}{(1-z)^2}\,\boldsymbol{\pi}\cdot\mathbf{1} \;+\; \sum_{\lambda_i\neq 1}\sum_{q=1}^{\alpha_i}\frac{1}{1-z}\;\frac{B_{i,q}}{(1-\lambda_i z)^q}\;.$$

On en déduit, par exemple en décomposant puis inversant les fractions rationnelles, que

$$\begin{aligned}\sum_{m=0}^{n}\mathbf{P}^m \;&=\; [z^n]\frac{1}{1-z} \;+\; \sum_{\lambda_i\neq 1}\sum_{q=1}^{\alpha_i}[z^n]\frac{1}{1-z}\;\frac{B_{i,q}}{(1-\lambda_i z)^q}\\ &=\; n+1 \;+\; \mathcal{O}(1)\;.\end{aligned}$$

Ceci prouve (25). Comme l'application $x \mapsto xA$ est continue en A, on en déduit que:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\boldsymbol{\pi}_m \;=\; \lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\boldsymbol{\pi}_0\mathbf{P}^m \;=\; \boldsymbol{\pi}_0\;\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{P}^m \;=\; \boldsymbol{\pi}_0.\mathbf{1}.\boldsymbol{\pi} \;=\; 1\times\boldsymbol{\pi}.$$

Ceci prouve (26). □

En supposant la matrice $\mathbf{P}$ *apériodique*, on obtient un résultat plus fort de convergence.

Théorème 15. *Soit $\mathbf{P}$ une matrice stochastique irréductible et apériodique. Soit $\boldsymbol{\pi}$ le vecteur propre gauche de $\mathbf{P}$ pour la valeur propre 1, tel que $\boldsymbol{\pi}.\mathbf{1} = 1$. Alors*

$$\lim_{n\to\infty}\mathbf{P}^n \;=\; \mathbf{1}.\boldsymbol{\pi}\;. \tag{27}$$

Par conséquent, quel que soit le vecteur $\boldsymbol{\pi}_0$, on a:

$$\lim_{n\to\infty}\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}\;. \tag{28}$$

La convergence est géométrique:

$$||\boldsymbol{\pi}_n - \boldsymbol{\pi}|| \;=\; \mathcal{O}(\rho^n) \tag{29}$$

pour tout ρ tel que $|\lambda_2| < \rho < 1$, où $|\lambda_2|$ est la valeur propre de $\mathbf{P}$ de plus grand module après 1.

Preuve. D'après le Théorème 13.*iii/*, les valeurs propres λ_m autres que 1 sont telles que $|\lambda_m| < 1$.

En utilisant maintenant le théorème de décomposition spectrale, on déduit de la partie *ii/* que pour $\lambda_1 = 1$, $B_1 = \mathbf{1}.\boldsymbol{\pi}$. Pour les autres valeurs propres, il est facile de voir que pour tout bloc de Jordan $J_i(\lambda)$ $i > 1$,

$$\lim_{n\to\infty}J_i(\lambda)^n \;=\; \mathbf{0}\;.$$

Ceci prouve (27). Comme précédemment, on en déduit par continuité que:

$$\lim_{n\to\infty}\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}_0\;\lim_{n\to\infty}\mathbf{P}^n \;=\; \boldsymbol{\pi}_0.\mathbf{1}.\boldsymbol{\pi} \;=\; 1\times\boldsymbol{\pi}.$$

Ceci prouve (28). Pour montrer (29), on applique l'argument de la preuve du Corollaire 5 au rapport J_i^n/ρ^n. □

Une chaîne de Markov dont la matrice satisfait aux conditions du Théorème 15 est dite *ergodique*.

On en conclut donc l'existence d'une distribution limite pour la chaîne de Markov (dans le cas ergodique). Cette distribution $\boldsymbol{\pi}$ est qualifiée de *stationnaire*. En effet, si la distribution *initiale* $\boldsymbol{\pi}_0$ vaut $\boldsymbol{\pi}$, alors pour tout n,

$$\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}_0\;\mathbf{P}^n \;=\; \boldsymbol{\pi}\;\mathbf{P}^n \;=\; \boldsymbol{\pi}\;.$$

4.2.6 Classification des états

Soit $\mathcal{G}$ le graphe de transition d'une matrice stochastique $\mathbf{P}$. On peut effectuer la décomposition en composantes fortement connexes de $\mathcal{G}$, et obtenir le graphes des classes de connexité, qui est un graphe orienté acyclique (DAG: directed acyclic graph).

Les sommets "finaux" de ce graphe, c'est-à-dire ceux qui n'ont aucun successeur, représentent des composantes fortement connexes de $\mathcal{G}$ que l'on nomme *classes terminales*. Les états de ces classes sont nommés *récurrents*, et les autres *transients*. Bien sûr, si $\mathbf{P}$ est irréductible, tous les états sont récurrents.

Cette dénomination des états est justifiée par le fait que la chaîne de Markov ne passe qu'un nombre fini de fois par un état transient, et par contre passe un nombre infini de fois par un état récurrent, si toutefois elle l'atteint. Pour se rendre compte de ceci, introduisons une variable aléatoire comptant le nombre de passages en i:

$$V_n^{(i)} \ = \ \sum_{m=0}^{n-1} \mathbf{1}_{\{X_m=i\}} \ .$$

En passant à l'espérance, on a clairement:

$$\mathbb{E}V_n^{(i)} \ = \ \sum_{m=0}^{n-1} \mathbb{P}(X_m = i) \ = \ \sum_{m=0}^{n-1} \boldsymbol{\pi}_0 \ \mathbf{P}^m \ \mathbf{e}_i \ ,$$

où $\mathbf{e}_i$ est le vecteur colonne $(0, \ldots, 1, \ldots, 0)'$, le "1" étant en ligne i.

La série de matrices $\mathbf{S} = \sum_{n=0}^{\infty} \mathbf{P}^n$ est la série potentielle déjà rencontrée au paragraphe 4.2.4. L'élément S_{ii} est donc le nombre moyen total de passages en i pour une chaîne partant de i.

Théorème 16. *Soit i un état de $\mathcal{E}$. Alors*

i/ si i est transient, alors $(\pi_n)_i = \mathcal{O}(\rho^n)$ pour un $\rho < 1$, et S_{ii} est fini;

ii/ si i est récurrent, alors S_{ii} est infini.

Preuve. Pour la partie *i/*, partitionons l'ensemble $\mathcal{E}$ en l'ensemble des états transients $\mathcal{T}$ et l'ensemble des états récurrents. En renumérotant les états si nécessaire, de telle sorte que les états de $\mathcal{T}$ précèdent ceux de $\mathcal{R}$, on peut écrire la matrice $\mathbf{P}$:

$$\mathbf{P} \ = \ \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{T} \end{pmatrix} \ .$$

La matrice $\mathbf{P}^n$ aura la même structure, et on voit facilement que les blocs diagonaux de cette matrice seront $\mathbf{Q}^n$ et $\mathbf{T}^n$. Or $\mathbf{Q}$ est sous-stochastique, et d'après P.-F. e/, $\rho(\mathbf{Q}) < 1$. Noter que ceci ne s'applique directement que si $\mathbf{Q}$ est irréductible. Mais si ce n'est pas le cas, il est possible de décomposer $\mathbf{Q}$ en blocs comme dans (23) et d'appliquer le principe sur les blocs diagonaux. Finalement, par le Corollaire 7, $\sum_n \mathbf{Q}^n$ converge.

Pour *ii/*, voir la preuve du Théorème 14. □

À noter le cas où un état i est tel que $P_{ii} = 1$. Un tel état est nommé *absorbant*. Il est nécessairement récurrent.

4.3 Chaînes de Markov à espace d'état dénombrable

Dans le cas où l'espace d'états est dénombrable, les définitions (16) et (17) ont toujours cours. Les notations matricielles sont également valides *bien que* le produit de matrices infinies ne soit pas défini en général. Ici, les produits de matrices et de vecteurs infinis sont bien définis car on ne manipule que des matrices à coefficients positifs et dont la somme sur les lignes est bornée (par 1).

La fonction génératrice $P(z)$ peut également se définir, mais on n'a plus pour elle la propriété (12). Par conséquent, le calcul des probabilités transitoires et l'existence des probabilités stationnaires se complique. L'étude de cette question sort du cadre de ce cours.

Le graphe de transition se définit comme auparavant, et on lui attribue les mêmes propriétés: irréducibilité, (a)périodicité.

On définit une classification des états, qui reprend celle établie pour les chaînes à espace d'état finie.

Un état de $\mathcal{E}$ est dit:

– transient si $S_{i,i} < \infty$,

– récurrent nul si $S_{i,i} = \infty$ et $(\mathbf{P}^n)_{i,i} \to 0$

– récurrent non nul (ou simplement: récurrent) si $S_{i,i} = \infty$ et $(\mathbf{P}^n)_{i,i} \not\to 0$.

On énonce le "théorème de solidarité" des états.

Théorème 17 (de solidarité). *Si $\mathbf{P}$ est irréductible, alors tous les états sont de même nature: (transients ou récurrents nuls ou récurrents positifs) et ont la même période.*

Donc si un état est récurrent non nul (respectivement, récurrent nul ou transient), tous les états le sont, et on donne le qualificatif à la chaîne de Markov, où à $\mathbf{P}$.

Enfin, on a un théorème de "classification" des comportements possibles.

Théorème 18. *Soit $\mathbf{P}$ une matrice stochastique infinie irréductible et apériodique. Deux possibilités seulement existent:*

- *soit le système $\mathbf{xP} = \mathbf{x}$ admet une solution $\mathbf{x}$ positive telle que $\sum_{i\in\mathcal{E}} x_i < \infty$, et donc un vecteur de probabilités stationnaires $\boldsymbol{\pi}$.*
  Alors $\boldsymbol{\pi}_n \to \boldsymbol{\pi}$ pour toute distribution initiale $\boldsymbol{\pi}_0$.
- *soit il n'existe aucune solution de $\mathbf{xP} = \mathbf{x}$ qui soit telle que $\sum_{i\in\mathcal{E}} |x_i| < \infty$.*
  Alors $\boldsymbol{\pi}_n \to \mathbf{0}$ pour toute distribution initiale $\boldsymbol{\pi}_0$.

La preuve classique de ces propriétés utilise des arguments probabilistes (par exemple, théorie du renouvellement; voir les références [3, 10]). Les arguments purement algébriques et analytiques utilisés dans le cas où $\mathcal{E}$ est fini ne sont plus appliquables en général.

4.4 Chaînes de Markov additives

4.4.1 Définitions

Supposons qu'a chaque transition de la chaîne de Markov, une "récompense" soit accumulée, en fonction de la transition effectuée. La récompense accumulée à la date n constitue un processus aléatoire qui a des propriétés intéressantes.

Plus formellement, soit $\{X(n), n \in \mathbb{N}\}$ une chaîne de Markov comme précedemment, et soit $\{Y(n), n \in \mathbb{N}\}$ le processus défini par

$$\begin{aligned} Y(0) &= 0 \\ Y(n+1) &= Y(n) + \xi_{X(n),X(n+1)}(n) \quad n \in \mathbb{N}, \end{aligned} \tag{30}$$

où pour tout $(i,j) \in V$, on a défini une suite $\{\xi_{i,j}(n), n \in \mathbb{N}\}$ de variables aléatoires *indépendantes* et identiquement distribuées. On supposera pour simplifier que les valeurs des v.a. $\xi_{i,j}$ sont entières, et donc celles de Y aussi.

Le processus $\{(X(n), Y(n)), n \in \mathbb{N}\}$ est une chaîne de Markov, qu'on nomme *chaîne de Markov additive.* Par contre, le processus $\{Y(n)\}$ n'est pas une chaîne de Markov en général.

4.4.2 Probabilités transitoires

On s'intéresse à la distribution de $Y(n)$ et à son comportement asymptotique. Pour cela, on passe aux fonctions génératrices et on introduit quelques notations matricielles. Soit pour commencer

$$G_{i,j}(z) = \mathbb{E}(z^{\xi_{i,j}}) .$$

Soient

$$\Psi_{i,j}(z) = \mathbb{E}(z^{\xi_{i,j}} \mathbf{1}_{\{X(n+1)=j\}} \mid X(n) = i) = p_{i,j}\, G_{i,j}(z) \qquad \text{et} \qquad \mathbf{\Psi}(z) = ((\Psi_{i,j}(z)))_{(i,j)\in\mathcal{E}\times\mathcal{E}}$$

$$\Psi^{(n)}_{i,j}(z) = \mathbb{E}(z^{Y(n)} \mathbf{1}_{\{X(n)=j\}} \mid X(0) = i), \qquad \text{et} \qquad \mathbf{\Psi}(z)^{(n)} = ((\Psi^{(n)}_{i,j}(z)))_{(i,j)\in\mathcal{E}\times\mathcal{E}} .$$

A noter que $\mathbf{\Psi}(1)$ est tout simplement $\mathbf{P}$, la matrice de transition de la chaîne $\{X(n)\}$. De même, $\mathbf{\Psi}^{(n)}(1)$ est la matrice $\mathbf{P}^n$. Soit enfin:

$$F_n(i;z) \;=\; \mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=i\}}) \quad \text{et} \quad \mathbf{F}_n(z) \;=\; (F_n(i;z))_{i\in\mathcal{E}} \;.$$

Le vecteur $\mathbf{F}_n(z)$ est s'exprime en fonction de $\mathbf{\Psi}$ et la distribution initiale de X: $\boldsymbol{\pi}_0$.

Lemme 19. *Pour tout $n \geq 0$, on a:*

$$\mathbf{\Psi}^{(n)}(z) \;=\; (\mathbf{\Psi}(z))^n$$

et par conséquent:

$$\mathbf{F}_n(z) \;=\; \boldsymbol{\pi}_0\; \mathbf{\Psi}(z)^n \;=\; \mathbf{F}_{n-1}\; \mathbf{\Psi}(z) \;. \tag{31}$$

Preuve. Pour i, j et n fixés, calculons donc la composante (i,j) de $\mathbf{\Psi}^{(n)}(z)$, soit: $\mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=j\}} \mid X(0)=i)$. Sachant que le processus $\{X(t)\}$ a suivi, entre les instants 0 et n, le chemin $\gamma = (i, i_1, i_2, \ldots, i_{n-1}, j)$, la valeur de $Y(n)$ est $\xi_{i,i_1}(0)+\xi_{i_1,i_2}(1)+\ldots+\xi_{i_{n-1},j}(n-1)$. Donc on a, en utilisant l'indépendance des variables aléatoires constituant cette somme:

$$\mathbb{E}(z^{Y(n)} \mid \gamma) \;=\; G_{i,i_1}(z) \times \ldots \times G_{i_{n-1},j}(z) \;.$$

Alors:

$$\begin{aligned}
\mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=j\}} \mid X(0)=i) \;&=\; \sum_{\gamma\in\Gamma^n(i,j)} \mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=j\}} \mid \gamma, X(0)=i)\mathbb{P}(\gamma \mid X(0)=i) \\
&=\; \sum_{\gamma\in\Gamma^n(i,j)} \mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=j\}} \mid \gamma)\; \mathbb{P}(\gamma) \\
&=\; \sum_{\gamma\in\Gamma^n(i,j)} \mathbb{E}(z^{Y(n)} \mid \gamma)\; p_{i,i_1}p_{i_1,i_2}\ldots p_{i_{n-1},j} \\
&=\; \sum_{\gamma\in\Gamma^n(i,j)} G_{i,i_1}(z)G_{i_1,i_2}\ldots G_{i_{n-1},j}(z)\; p_{i,i_1}p_{i_1,i_2}\ldots p_{i_{n-1},j} \\
&=\; \sum_{\gamma\in\Gamma^n(i,j)} \mathbf{\Psi}_{i,i_1}(z)\mathbf{\Psi}_{i_1,i_2}\ldots \mathbf{\Psi}_{i_{n-1},j}(z)
\end{aligned}$$

et le résultat s'ensuit en vertu de (4). Dans ce développement, on a restreint la somme à $\Gamma^n(i,j)$ car pour un autre chemin, soit sa probabilité est nulle sachant que $X(0)=i$, ou bien la fonction $\mathbf{1}_{\{X(n)=j\}}$ est nulle. On a également fait usage de la formule (20).

La preuve de (31) s'obtient en conditionnant sur l'état initial de la chaîne X:

$$\begin{aligned}
\mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=j\}}) \;&=\; \sum_{i\in\mathcal{E}} \boldsymbol{\pi}_0(i)\; \mathbb{E}(z^{Y(n)}\mathbf{1}_{\{X(n)=j\}}|X(0)=i) \\
&=\; (\boldsymbol{\pi}_0\; \mathbf{\Psi}(z)^n)_j \;.
\end{aligned}$$

□

On déduit de (31) que par exemple:

$$\begin{aligned}
\mathbb{E}(Y(n)) \;&=\; \frac{d}{dz}\mathbf{F}_n(1)\;\mathbf{1} \;=\; \boldsymbol{\pi}_0 \sum_{k=0}^{n-1} \mathbf{P}^k\mathbf{\Psi}'(1)\mathbf{P}^{n-1-k}\mathbf{1} \\
&=\; \boldsymbol{\pi}_0 \sum_{k=0}^{n-1} \mathbf{P}^k\mathbf{\Psi}'(1)\;\mathbf{1} \;.
\end{aligned}$$

En particulier, la moyenne de $Y(n)$, $\{X(n)\}$ étant *stationnaire* vaut:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\pi}}(Y(n)) &= \boldsymbol{\pi} \sum_{k=0}^{n-1} \mathbf{P}^k \boldsymbol{\Psi}'(1)\ \mathbf{1} \ = \ \sum_{k=0}^{n-1} \boldsymbol{\pi}\ \boldsymbol{\Psi}'(1)\ .\mathbf{1} \\ &= n\ \boldsymbol{\pi}\boldsymbol{\Psi}'(1)\mathbf{1}\ .\end{aligned}$$

La récompense moyenne accumulée à l'étape n croît (ou décroît) donc linéairement avec n.

4.4.3 Comportement asymptotique

D'après (32), $\mathbb{E}Y(n)$ croît donc linéairement, quand la chaîne X sous-jacente est dans son état stationnaire. Dans le cas où X n'est pas stationnaire, ceci reste vrai asymptotiquement:

Théorème 20. *Si la chaîne $\{X(n), n \in \mathbb{N}\}$ est irréductible, soit $\boldsymbol{\pi}$ son vecteur de probabilités stationnaires. Soit $\gamma = \boldsymbol{\pi}\boldsymbol{\Psi}'(1)\mathbf{1}$. Alors:*

$$\mathbb{E}Y(n) \ = \ n\gamma \ + \ \mathcal{O}(1)\ .$$

Preuve. On a de ce qui précède:

$$\mathbb{E}Y(n) \ = \ \boldsymbol{\pi}_0 \sum_{k=0}^{n-1} \mathbf{P}^k\ \boldsymbol{\Psi}'(1)\ \mathbf{1}.$$

La chaîne étant irréductible, le Théorème 14 s'applique et on obtient le résultat attendu. □

Le taux de croissance γ se calcule en fait de diverses façons:

Lemme 21. *Soit $\{X(n), Y(n)\}$ une chaîne de Markov additive telle que la chaîne $\{X(n), n \in \mathbb{N}\}$ est irréductible. Soit $\lambda_1(z)$ la valeur propre de plus grand module de $\boldsymbol{\Psi}(z)$ et soit $\chi_{\boldsymbol{\Psi}(z)}(x) = \det(\boldsymbol{\Psi}(z) - x\mathbf{I})$. Alors*

$$\begin{aligned}\gamma &= \sum_{i\in\mathcal{E}} \pi(i) \sum_{j\in\mathcal{E}} P_{i,j}\ \mathbb{E}\xi_{i,j} & (32)\\ &= \lambda'(1) & (33)\\ &= -\ \frac{\partial\chi_{\boldsymbol{\Psi}(z)}}{\partial z}(x)\ /\ \frac{\partial\chi_{\boldsymbol{\Psi}(z)}}{\partial x}(x)\bigg|_{x=z=1}\ . & (34)\end{aligned}$$

Preuve. La formule (32) s'obtient en développant le produit matriciel qui définit γ, sachant que:

$$\boldsymbol{\Psi}'(1)_{i,j} \ = \ \frac{d}{dz}\mathbb{E}(z^{\xi_{i,j}})\bigg|_{z=1} \ = \ \mathbb{E}\xi_{i,j}\ .$$

La formule (34) se déduit de (33): étant donné que

$$\chi_{M(z)}(x) \ = \ \prod_{i=1}^{p} (\lambda_i(z) - x)^{\alpha_i}\ ,$$

on trouve en formant les différentielles (logarithmiques) que:

$$\begin{aligned}\frac{\partial\chi_{M(z)}}{\partial z}(x) &= -\ \chi_{M(z)}(x)\ \sum_{i=0}^{p} \alpha_i\lambda_i'(z)\ \frac{1}{\lambda_i(z) - x}\\ \frac{\partial\chi_{M(z)}}{\partial x}(x) &= -\ \chi_{M(z)}(x)\ \sum_{i=0}^{p} \alpha_i\ \frac{1}{\lambda_i(z) - x}\ .\end{aligned}$$

En faisant $x = \lambda_1(z)$, puis $z = 1$, on obtient (34). □

4.4.4 Fonctions génératrices

Il est possible et parfois utile de former la série génératrice des puissances $\mathbf{\Psi}^n$. Soit donc:

$$R(z,t) \;=\; \sum_{n=0}^{\infty} t^n \;\mathbf{\Psi}(z)^n \;=\; (\mathbf{I} - t\mathbf{\Psi}(z))^{-1} \;.$$

Cette série converge quand $t < \rho(\mathbf{\Psi}(z))^{-1}$. Quand $z = 1$, on retrouve bien sûr la résolvante: $R(1,t) = P(t)$.

À l'aide de cette matrice, on peut écrire (au moins formellement):

$$\begin{aligned}
\sum_{n=0}^{\infty} t^n F_n(z) &= \boldsymbol{\pi}_0 \; R(z,t) \\
\sum_{n=0}^{\infty} t^n F_n'(z) &= t\boldsymbol{\pi}_0 \; R(z,t)\mathbf{\Psi}'(z)R(z,t) \\
\sum_{n=0}^{\infty} t^n F_n''(z) &= 2t^2\boldsymbol{\pi}_0 R(z,t)\mathbf{\Psi}'(z)R(z,t)\mathbf{\Psi}'(z)R(z,t) \;+\; t\boldsymbol{\pi}_0 R(z,t)\mathbf{\Psi}''(z)R(z,t) \;.
\end{aligned}$$

Comme exemple d'application, considérons la série génératrice des seconds moments factoriels de $Y(n)$:

$$\begin{aligned}
\sum_{n=0}^{\infty} t^n \mathbb{E}(Y(n)(Y(n)-1)) &= \sum_{n=0}^{\infty} t^n F_n''(1)\; \mathbf{1} \\
&= 2t^2\boldsymbol{\pi}_0 R(1,t)\mathbf{\Psi}'(1)R(1,t)\mathbf{\Psi}'(1)R(1,t)\mathbf{1} \;+\; t\boldsymbol{\pi}_0 R(1,t)\mathbf{\Psi}''(1)R(1,t)\mathbf{1} \;.
\end{aligned}$$

Or, $P(t)\mathbf{1} = \mathbf{1}/(1-t)$ et $\boldsymbol{\pi}P(t) = \boldsymbol{\pi}/(1-t)$. Par conséquent, dans le cas où $\{X(n)\}$ est stationnaire,

$$\sum_{n=0}^{\infty} t^n \mathbb{E}_{\boldsymbol{\pi}}(Y(n)(Y(n)-1)) \;=\; 2\frac{t^2}{(1-t)^2}\boldsymbol{\pi}\mathbf{\Psi}'(1)P(t)\mathbf{\Psi}'(1)\mathbf{1} \;+\; \frac{t}{(1-t)^2}\boldsymbol{\pi}\mathbf{\Psi}''(1)\mathbf{1} \;.$$

On en conclut que:

$$\mathbb{E}_{\boldsymbol{\pi}}(Y(n)(Y(n)-1)) \;=\; n\boldsymbol{\pi}\mathbf{\Psi}''(1)\mathbf{1} \;+\; 2\boldsymbol{\pi}\mathbf{\Psi}'(1)\sum_{m=0}^{n-1}\sum_{p=0}^{m}\mathbf{P}^p\mathbf{\Psi}'(1)\mathbf{1} \;.$$

4.5 La marche aléatoire en dimension 1

Dans ce paragraphe, on analyse un système dynamique aléatoire central: la marche aléatoire (ou "promenade de l'ivrogne") sur l'ensemble des entiers. La description du processus est la suivante: à chaque unité de temps:

– ou bien il diminue de 1 avec probabilité p,

– ou bien le processus augmente de 1, avec probabilité $q = 1 - p$.

Dans le cas ou le processus se trouve dans l'état 0, il ne peut plus diminuer, et il reste donc dans l'état 0 avec probabilité q. Nous considèrerons deux cas: le cas fini où il existe un maximum K, et le cas infini.

Dans le cas fini ou le processus a atteint son maximum K, il ne peut plus augmenter et reste donc dans l'état K avec probabilité q.

Ce processus appartient à la famille des *processus de naissance et de mort*, ainsi nommée parce que l'état ne change que par ajout de 1 (naissance) ou retrait de 1 (mort).

Clairement, le processus $\{X(t), t \in \mathbb{N}\}$ est une chaîne de Markov homogène à espace d'états $\mathcal{E} = \{0, \ldots, K\}$ ($K \leq +\infty$). Si K est fini, la matrice de transition est

$$\mathbf{P} \;=\; \begin{pmatrix} p & q & 0 & & \\ p & 0 & q & \ddots & \\ 0 & \ddots & \ddots & \ddots & \\ & & p & 0 & q \\ & & & p & q \end{pmatrix} \;. \qquad (35)$$

Si K est infini, la matrice (35) est infinie avec la même structure et sans la dernière ligne.

Nous allons nous intéresser à diverses propriétés de ce processus.

4.5.1 Probabilités stationnaires, cas fini

La matrice $\mathbf{P}$ est clairement irréductible et apériodique, donc ergodique. Le Théorème 15 s'applique donc, et il existe un unique vecteur de probabilités stationnaires donné par l'équation $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. Les équations de ce système linéaire se listent comme:

$$\begin{aligned}
\pi_0 &= p\pi_0 + p\pi_1, \\
\pi_k &= q\pi_{k-1} + p\pi_{k+1}, \qquad 1 \leq k \leq K-1, \\
\pi_K &= q\pi_{K-1} + q\pi_K.
\end{aligned}$$

Des nombreuses manières de résoudre ce système, je vais utiliser la moins directe, mais la plus générale. Ceci me permettra d'introduire une méthode de calcul omniprésente dans ce type de problème.

On introduit la fonction génératrice de la loi de probabilité:

$$\pi(z) = \sum_{n=0}^{K} z^k \pi_k .$$

Noter que cette fonction est un polynôme de la variable z. On sait également (*cf.* les préliminaires) que $\pi(1) = 1$.

On va à présent substituer les équations d'équilibre dans la définition de $\pi(z)$ et réarranger les termes pour faire apparaître une équation fonctionnelle impliquant $\pi(z)$. On obtient:

$$\begin{aligned}
\pi(z) &= \pi_0 + \sum_{k=1}^{K-1} z^k \pi_k + z^K \pi_K \\
&= p\pi_0 + p\pi_1 + \sum_{k=1}^{K-1} z^k (q\pi_{k-1} + p\pi_{k+1}) + z^K (q\pi_{K-1} + q\pi_K) \\
&= p\pi_0 + p\sum_{k=0}^{K-1} z^k \pi_{k+1} + q\sum_{k=1}^{K} z^k \pi_{k-1} + qz^K \pi_K \\
&= p\pi_0 + p\sum_{k=1}^{K} z^{k-1} \pi_k + q\sum_{k=0}^{K-1} z^{k+1} \pi_k + qz^K \pi_K \\
&= p\pi_0 + \frac{p}{z}(\pi(z) - \pi_0) + qz(\pi(z) - z^K \pi_K) + qz^K \pi_K
\end{aligned}$$

Par conséquent, $\pi(z)$ vérifie (on multiplie par z cette derniere relation):

$$\pi(z)(-qz^2 + z - p) = p\pi_0(z-1) - q\pi_K z^{K+1}(z-1) .$$

Le terme $(z-1)$ se factorise du membre gauche et on arrive à:

$$\pi(z) = \frac{p\pi_0 - qz^{K+1}\pi_K}{p - qz} . \tag{36}$$

Le problème serait résolu si on connaissait les nombres π_0 et π_K. On va les déterminer grâce aux informations connues sur $\pi(z)$. Puisque la fonction $\pi(z)$ est un polynôme, elle n'a pas de singularité dans le plan complexe. Donc si on fait $z = p/q$ dans le membre droit de (36), le numérateur doit s'annuller. Ceci donne la condition:

$$p\pi_0 = q(\frac{p}{q})^{K+1}\pi_K \quad \Longleftrightarrow \quad \pi_K = (\frac{q}{p})^K \pi_0 .$$

On reporte dans (36) ce qui donne:

$$\pi(z) \;=\; \pi_0\; \frac{p - p(qz/p)^{K+1}}{p - qz} \;=\; \pi_0\; \frac{1-(qz/p)^{K+1}}{1-(qz/p)}\ . \tag{37}$$

Enfin, pour déterminer la valeur de π_0, on utilise la condition dite de normalisation: $\pi(1) = 1$. On a alors:

$$1 \;=\; \pi_0\; \frac{1-(q/p)^{K+1}}{1-q/p} \qquad \Longleftrightarrow \qquad \begin{cases} \pi_0 \;=\; \dfrac{1-q/p}{1-(q/p)^{K+1}} & \text{si } p \neq q \\ \pi_0 \;=\; \dfrac{1}{K+1} & \text{si } p = q = 1/2\ . \end{cases}$$

Finalement, la fonction génératrice que nous cherchions est, dans le cas $p \neq q$,

$$\pi(z) \;=\; \frac{1-q/p}{1-(q/p)^{K+1}}\; \frac{1-(qz/p)^{K+1}}{1-(q/p)z} \;=\; \frac{1-q/p}{1-(q/p)^{K+1}} \sum_{k=0}^{K} \left(\frac{qz}{p}\right)^k\ . \tag{38}$$

On en déduit facilement les probabilités π_k:

$$\pi_k \;=\; [z^k]\pi(z) \;=\; \left(\frac{q}{p}\right)^k\; \frac{1-q/p}{1-(q/p)^{K+1}}\ . \tag{39}$$

À nouveau, il n'était pas strictement nécessaire de faire tout ce calcul pour arriver à cette conclusion (surtout si on connaît déjà le résultat!) mais cette méthode a l'avantage d'être générale et d'aboutir à la solution dans des cas plus compliqués.

Ce type de raisonnement porte le nom de *méthode du noyau*, que nous utilisons encore ci-dessous sous une forme moins triviale. Le noyau est ici le polynôme $p - qz$ qui apparaît dans le membre droit de l'équation fonctionnelle, et au dénominateur de (36). Dans le paragraphe suivant, on applique de nouveau cette technique dans des conditions moins triviales.

4.5.2 Probabilités transitoires, cas fini

Nous allons à présent nous attaquer au calcul des probabilités transitoires de ce processus, c'est à dire des:

$$\pi_{n,k} \;=\; \mathbb{P}(X(n) = k)\ , \quad n \in \mathbb{N},\ k \in \mathcal{E}.$$

Comme pour le calcul des probabilités stationnaires, le point de départ est l'écriture de la récurrence $\boldsymbol{\pi}_n \;=\; \boldsymbol{\pi}_{n-1}\mathbf{P}$ sous la forme: pour tout $n \geq 1$,

$$\begin{aligned} \pi_{n,0} &= p\pi_{n-1,0} + p\pi_{n-1,1}, \\ \pi_{n,k} &= q\pi_{n-1,k-1} + p\pi_{n-1,k+1}, \qquad 1 \leq k \leq K-1, \\ \pi_{n,K} &= q\pi_{n-1,K-1} + q\pi_{n-1,K}. \end{aligned}$$

Le vecteur $\boldsymbol{\pi}_0$ est supposé donné. On passe alors a la fonction génératrice (bidimensionnelle):

$$
\begin{aligned}
\mathcal{F}(x,z) &= \sum_{n=0}^{\infty}\sum_{k=0}^{K} z^n x^k \pi_{n,k} \\
&= \sum_{k=0}^{K} x^k \pi_{0,k} \;+\; \sum_{n=1}^{\infty} z^n \Big(p\pi_{n-1,0} + p\pi_{n-1,1} + \sum_{k=1}^{K-1} x^k (q\pi_{n-1,k-1} + p\pi_{n-1,k+1}) \\
&\qquad\qquad + x^K (q\pi_{n-1,K-1} + q\pi_{n-1,K}) \Big) \\
&= \mathcal{F}(x,0) \;+\; \sum_{n=0}^{\infty} z^{n+1} \left(p \sum_{k=0}^{K-1} x^k \pi_{n,k+1} + q \sum_{k=1}^{K} x^k \pi_{n,k-1} + p\pi_{n-1,0} + qx^K \pi_{n,K} \right) \\
&= \mathcal{F}(x,0) \;+\; z \sum_{n=0}^{\infty} z^n \left(p \sum_{k=1}^{K} x^{k-1} \pi_{n,k} + q \sum_{k=0}^{K-1} x^{k+1} \pi_{n,k} + p\pi_{n-1,0} + qx^K \pi_{n,K} \right) \\
&= \mathcal{F}(x,0) \;+\; \sum_{n=0}^{\infty} z^n \left(\frac{p}{x} \left(\sum_{k=0}^{K} x^k \pi_{n,k} - \pi_{n,0} \right) + qx \left(\sum_{k=0}^{K} x^k \pi_{n,k} - x^K \pi_{n,K} \right) + p\pi_{n-1,0} + qx^K \pi_{n,K} \right) \\
&= \mathcal{F}(x,0) \;+\; z\left(\frac{p}{x} + qx\right) \mathcal{F}(x,z) \;+\; z(1-\frac{1}{x})(p\mathcal{F}(0,z) \;-\; qx^{K+1}\mathcal{F}_K(z)) \;,
\end{aligned}
\tag{40}
$$

où on a noté:

$$
\mathcal{F}_K(z) \;=\; \sum_{n=0}^{\infty} z^n \pi_{n,K} \;. \tag{41}
$$

On déduit de (40) que la fonction $\mathcal{F}$ vérifie l'équation fonctionnelle suivante:

$$
(x - pz - qzx^2)\; \mathcal{F}(x,z) \;=\; x\mathcal{F}(x,0) + z(x-1)\big(p\mathcal{F}(0,z) \;-\; qx^{K+1}\mathcal{F}_K(z)\big) \;. \tag{42}
$$

Dans cette équation, la fonction $\mathcal{F}(x,0)$ est connue car donnée par $\boldsymbol{\pi}_0$, mais $\mathcal{F}(0,z)$ et $\mathcal{F}_K(z)$ sont des *fonctions inconnues*. Dans le membre droit, le polynôme:

$$
K(x,z) \;=\; qzx^2 + x - pz \tag{43}
$$

joue un rôle déterminant. Ce polynôme est appellé le *noyau* de l'équation. L'argument de la méthode du noyau est le suivant: pour tout point (x_0, z_0) tel que $K(x_0, z_0) = 0$ *et* $\mathcal{F}$ est analytique (ne tend pas vers l'infini) en ce point, alors le membre droit s'annulle. Ceci donne une condition liant les fonctions inconnues $\mathcal{F}(0,z)$ et $\mathcal{F}_K(z)$. Avec assez de conditions, on va peut-etre réussir à les déterminer.

On commence donc par résoudre $K(x,z) = 0$ en x. On obtient deux fonctions

$$
x_1(z) \;=\; \frac{1}{2qz}\left(-1 + \sqrt{1 - 4pqz^2}\right), \qquad x_2(z) \;=\; \frac{1}{2qz}\left(-1 - \sqrt{1 - 4pqz^2}\right), \tag{44}
$$

qui sont telles que $K(x_1(z), z) = K(x_2(z), z) = 0$ pour tout complexe $z \neq 0$.

La fonction $\mathcal{F}$, elle, est définie pour tout $x \in \mathbb{C}$ (c'est un polynôme en x) et $|z| < 1$. Donc si on prend z dans le cercle unité que qu'on fait $x = x_1(z)$ ou bien $x = x_2(z)$ dans (42), on obtient:

$$
\left\{ \begin{array}{lcl} 0 &=& x_1(z)\mathcal{F}(x_1(z),0) + z(x_1(z)-1)\big(p\mathcal{F}(0,z) \;-\; qx_1(z)^{K+1}\mathcal{F}_K(z)\big) \\ 0 &=& x_2(z)\mathcal{F}(x_2(z),0) + z(x_2(z)-1)\big(p\mathcal{F}(0,z) \;-\; qx_2(z)^{K+1}\mathcal{F}_K(z)\big) \;. \end{array} \right. \tag{45}
$$

Nous sommes en présence d'un système de deux équations à deux inconnues: les nombres $\mathcal{F}(0,z)$ et $\mathcal{F}_K(z)$. On en déduit les valeurs de ces fonctions, puis celle de $\mathcal{F}(x,z)$ en reportant dans (42). Pour simplier les expressions, je suppose que $X(0) = k$ avec probabilité 1, et donc que $\mathcal{F}(x,0) = x^k$. En utilisant d'autre part que $x_1(z)x_2(z) = p/q$ (d'après (43)), on obtient:

$$
\begin{aligned}
pz\mathcal{F}(0,z) &= \frac{1}{x_1^{K+1} - x_2^{K+1}} \left(\frac{x_1^{k+1}x_2^{K+1}}{1-x_1} - \frac{x_2^{k+1}x_1^{K+1}}{1-x_2} \right) \\
&= \frac{1}{x_1^{K+1} - x_2^{K+1}} \left(\frac{p}{q}\right)^{k+1} \left(\frac{x_2^{K-k}}{1-x_1} - \frac{x_1^{K-k}}{1-x_2} \right) \\
qz\mathcal{F}_K(z) &= \frac{1}{x_1^{K+1} - x_2^{K+1}} \left(\frac{x_1^{k+1}}{1-x_1} - \frac{x_2^{k+1}}{1-x_2} \right) .
\end{aligned}
$$

4.5.3 Probabilités stationnaires, cas infini

Dans le cas où $K = \infty$, le processus $\{X(t), t \in \mathbb{N}\}$ est toujours irréductible et apériodique. On peut donc appliquer le Théorème 18. Pour cela, on doit tenter de résoudre le système $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. Celui-ci s'écrit:

$$
\begin{aligned}
\pi_0 &= p\pi_0 + p\pi_1 , \\
\pi_k &= q\pi_{k-1} + p\pi_{k+1} , \qquad 1 \le k
\end{aligned}
$$

Cette récurrence se résoud de multiples façons, par exemple en utilisant la méthode analytique du paragraphe 4.5.1, et ses résultats (Équation (36)) en passant à la limite $K \to \infty$; ou encore en utilisant la construction de 3.3.3 (exercice instructif); ou simplement par récurrence en partant de la première équation.

Dans tous les cas, on obtient que *nécessairement*, la solution $\boldsymbol{\pi}$ est de la forme:

$$\pi_k = \left(\frac{q}{p}\right)^k \pi_0 .$$

Deux cas apparaissent donc:

- $p > q$; alors la série $\sum \pi_k$ converge et il existe donc une probabilité stationnaire donnée par:

$$\pi_k = \left(\frac{q}{p}\right)^k (1 - \frac{q}{p}) . \tag{46}$$

  Il s'agit d'une *distribution géométrique*. D'après le Théorème 18, la distribution $\boldsymbol{\pi}_t$ converge vers elle pour tout $\boldsymbol{\pi}_0$. La chaîne est dans ce cas récurrente.
- $p \le q$; alors il n'existe pas de solution sommable au système $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. Donc $\boldsymbol{\pi}_t$ converge vers $\mathbf{0}$ pour tout $\boldsymbol{\pi}_0$. La chaîne est dans ce cas soit transiente, soit récurrente nulle. L'analyse des probabilités transitoires (en exercice, en reprenant les calculs du paragraphe 4.5.2) révèle que si $p > q$, on a transience, alors que si $p = q$, on a récurrence nulle.

5 Chaînes de Markov en temps continu

Les concepts introduits dans le cas discret ont tous une analogie dans le cas continu. Ce paragraphe propose un survol des résultats de base dans ce domaine, plus à titre de référence que d'étude dans la mesure où le sujet ne sera qu'effleuré en cours. Les processus de Markov en temps continu sont étroitement liés à une distribution particulière: la distribution exponentielle. Ses propriétés sont donc brièvement listées.

5.1 Définitions

Une variable aléatoire X est dite avoir une distribution *exponentielle* de paramètre $\lambda > 0$ (ou bien être exponentiellement distribuée de paramètre λ), si sa fonction de répartition est:

$$F_X(x) = \mathbb{P}(X \le x) = 1 - e^{-\lambda x} \tag{47}$$

pour $x \geq 0$ et $F_X(x) = 0$ pour $x \leq 0$. On notera pour raccourcir que $X \sim \text{Exp}(\lambda)$.

Parmi les nombreuses propriétés merveilleuses de cette distribution, notons:

- si $X \sim \text{Exp}(\lambda)$, alors $\mathbb{E}X = \lambda^{-1}$, $\mathbb{E}X^k = k!\lambda^{-k}$ et $\text{var}(X) = \lambda^{-2}$.
- la distribution exponentielle est *sans mémoire*: pour tous $s, t > 0$,

$$\mathbb{P}(X > s + t \mid X > s) \;=\; \mathbb{P}(X > t) \; . \tag{48}$$

- la distribution exponentielle a un taux de défaillance constant: pour tout $t \geq 0$,

$$\lim_{h \to 0} \frac{1}{h} \, \mathbb{P}(X \in [t, t+h[\mid X > t) \;=\; \lambda \; . \tag{49}$$

- la famille des distributions exponentielles est stable pour la minimisation: si $X_1 \sim \text{Exp}(\lambda_1)$, $X_2 \sim \text{Exp}(\lambda_2)$ et X_1 et X_2 sont indépendantes alors $\min\{X_1, X_2\} \sim \text{Exp}(\lambda_1 + \lambda_2)$. De plus, la probabilité que ce minimum soit réalisé par X_i est $\lambda_i/(\lambda_1 + \lambda_2)$. Ces propriétés se généralisent à un nombre quelconque de variables.

Soit une suite aléatoire $T_0 \leq T_1 \leq \ldots \leq T_n \leq T_{n+1} \leq \ldots$. Le processus stochastique $\{X(t), t \in \mathbb{R}^+\}$ défini par

$$X(t) \;=\; \#\{n \mid T_n \leq t\} \;=\; \sum_{n=0}^{\infty} \mathbf{1}_{\{T_n \leq t\}}$$

(c'est un *processus de comptage*) est un *Processus de Poisson* de paramètre λ si la suite $\{T_0, T_1 - T_0, \ldots, T_{n+1} - T_n, \ldots\}$ est une suite de variables aléatoires indépendantes et de distribution exponentielle de paramètre λ. Parmi les nombreuses propriétés merveilleuses de ce processus, notons:

- le nombre de points dans tout intervalle de longueur u a une distribution de Poisson de paramètre λu: $\mathbb{P}(X(t+u) - X(t) = k) \;=\; (\lambda u)^k e^{-\lambda u}/k!$. En particulier, sa moyenne est λu. Pour cette raison, λ est aussi nommé le *taux d'arrivée* du processus.
- les nombres de points du processus dans deux intervalles disjoints sont indépendants. En particulier, le processus de Poisson est sans mémoire: pour $t_1 > t_2 > t_3 > t_4$, k et j:

$$\mathbb{P}(X(t_1) - X(t_2) = k \mid X(t_3) - X(t_4) = j) \;=\; \mathbb{P}(X(t_1) - X(t_2) = k) \; .$$

Soit $\mathcal{E}$ un ensemble au plus dénombrable. Un processus $\{X(t), t \in \mathbb{R}^+\}$ à valeurs dans $\mathcal{E}$ est une *chaîne de Markov en temps continu* (en abrégé: CMTC) ou un *processus de Markov* homogène si et seulement si:

i/ (propriété de Markov) pour tout $n \in \mathbb{N}$, tout $n+2$– uplet de réels $t_0 < t_1, < \ldots < t_n < t_{n+1}$ et tout $n+2$– uplet $(j_0, j_1, \ldots, j_n, j_{n+1})$ d'éléments de $\mathcal{E}$:

$$\mathbb{P}[X(t_{n+1}) = j_{n+1} | X(t_n) = j_n, \ldots, X(t_0) = j_0] = \mathbb{P}[X(t_{n+1}) = j_{n+1} | X(t_n) = j_n] \; ; \tag{50}$$

ii/ (homogénéité) pour tout réels s, t et u, et toute paire (i, j) de $\mathcal{E}$,

$$\mathbb{P}[X(t+u) = j | X(s+u) = i] \;=\; \mathbb{P}[X(t) = j | X(s) = i] \;=\; P_{t-s}(i, j) \; , \tag{51}$$

indépendamment de t.

De même que dans le cas du temps discret, on s'intéresse aux probabilités d'état:

$$p_t(n) \;=\; \mathbb{P}[X(t) = n], \qquad t \in \mathbb{R}^+, \; n \in \mathcal{E} \; .$$

5.2 Equations de Chapman-Kolmogoroff

De la propriété de Markov, et de l'homogénéité, on déduit que pour toute paire de réels s et t, et toute paire d'états (i, j), on a:

$$P_{t+s}(i,j) \;=\; \sum_{k\in\mathcal{E}} P_t(i,k)\; P_s(k,j)\;, \tag{52}$$

ou encore, sous forme matricielle,

$$\mathbf{P}_{t+s} \;=\; \mathbf{P}_t\;\mathbf{P}_s\;,$$

ou $\mathbf{P}_u$ est la matrice dont les composantes sont les $P_u(i,j)$, $(i,j) \in \mathcal{E}^2$. Ces équations portent le nom de Chapman-Kolmogoroff.

Noter au passage que le calcul sur les matrices infinies reste valide au même titre que dans le cas du temps discret.

Si le processus $\{X(t)\}$ est suffisamment "régulier", alors

$$\lim_{t\to 0} \mathbf{P}_t \;=\; \mathbf{I}$$

et qu'il existe une matrice $\mathbf{Q}$ donnée par:

$$\mathbf{Q} \;=\; \lim_{t\to 0} \frac{\mathbf{P}_t - I}{t}\;.$$

Cette matrice est appelée le *générateur infinitésimal* du processus. Elle a la propriété que pour tout $i \in \mathcal{E}$: $\sum_{j\in\mathcal{E}} Q(i,j) = Q(i,i) + \sum_{i\neq j} Q(i,j) = 0$. Sous forme matricielle, cela revient à dire: $\mathbf{Q1} = \mathbf{0}$. La construction pratique et l'interprétation d'un générateur infinitésimal sont décrits au paragraphe suivant.

Dans le cas où $\mathcal{E}$ est fini (ou «formellement» sinon), l'équation (52) se résoud en:

$$\mathbf{P}_t \;=\; e^{t\mathbf{Q}} \;=\; \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{Q}^n,$$

et les probabilités d'état sont alors données par:

$$\mathbf{p}_t \;=\; \mathbf{p}_0\mathbf{P}_t \;=\; \mathbf{p}_0 e^{t\mathbf{Q}}\;.$$

Donc, de même que dans le cas discret, il est *a priori* possible de calculer les valeurs de ces probabilités pour tout t, au prix du calcul de $e^{t\mathbf{Q}}$.

5.3 Construction du générateur

À la différence de la matrice de transition des chaînes de Markov en temps discret, le générateur infinitésimal n'est pas directement donné dans la définition du processus. Or, comme c'est l'objet central de l'analyse du comportement du processus, il importe de savoir le calculer. En pratique, le générateur se déduit d'une description de l'évolution du processus.

Théorème 22 (Construction #1). *Soit $\{X(t), t \geq 0\}$ un processus tel que: une fois entré dans l'état i, il y reste une durée aléatoire de distribution exponentielle de paramètre τ_i puis il saute vers l'état $j \neq i$ avec la probabilité $p_{i,j}$ ($\sum_j p_{i,j} = 1$). Alors le processus $\{X(t), t \geq 0\}$ est une chaîne de Markov en temps continu, dont le générateur infinitésimal $\mathbf{Q}$ est donné par:*

$$Q(i,i) \;=\; -\tau_i \qquad Q(i,j) \;=\; \tau_i\; p_{i,j}\;, i \neq j\;.$$

La preuve de ce théorème sort du cadre de ce cours. C'est une conséquence des propriétés (48) et (49) de la distribution exponentielle.

Cette construction permet de trouver le générateur d'un processus de Markov dont on décrirait le fonctionnement ainsi, mais donne surtout une façon de générer des trajectoires d'un tel processus.

Dans la pratique, pour prouver qu'un processus est une chaîne de Markov en temps continu et trouver son générateur, on se sert plus souvent de la construction suivante:

Théorème 23 (Construction # 2). *Soit $\{X(t), t \geq 0\}$ un processus tel que: quand il se trouve dans l'état i, il peut en sortir pour aller dans l'état $j \neq i$ après un temps $Y_{i,j}$ exponentiellement distribué de paramètre $\nu_{i,j}$. On suppose que $\sum_j \nu_{i,j} < \infty$. Le processus reste effectivement jusqu'à la date $\min_j\{Y_{i,j}\}$ puis saute vers l'état k qui réalise le minimum. Alors le processus $\{X(t), t \geq 0\}$ est une chaîne de Markov en temps continu, dont le générateur infinitésimal $\mathbf{Q}$ est donné par:*

$$Q(i,i) \;=\; -\sum_j \nu_{i,j} \qquad Q(i,j) \;=\; \nu_{i,j} \;, i \neq j \;.$$

La preuve de ce théorème peut se déduire facilement du Théorème 22 et de la propriété du minimum de variables exponentiellement distribuées indépendantes.

En vertu de cette construction, on peut interpréter chaque $Q(i,j)$ comme le paramètre d'une variable aléatoire exponentielle. Pour cela, on dit que $Q(i,j)$ est le *taux de transition* de l'état i vers l'état j.

5.4 Uniformisation

Le lien entre les chaînes de Markov en temps continu et en temps discret se fait par le processus d'*uniformisation*. Uniformiser une chaîne de Markov revient à réaliser un couplage entre cette chaîne et une certaine chaîne de Markov en temps discret par l'intermédiaire d'un processus de Poisson.

Considérons une chaîne de Markov en temps continu de générateur $\mathbf{Q}$, et soit un nombre ν tel que

$$\nu \geq \max_i\{|Q(i,i)|\} \;.$$

Soit ensuite la matrice $\mathbf{P}$:

$$\mathbf{P} \;=\; \mathbf{I} \;+\; \frac{1}{\nu}\,\mathbf{Q} \;. \tag{53}$$

On vérifie facilement que $\mathbf{P}$ est une matrice positive telle que $\mathbf{P1} = \mathbf{1}$, c'est à dire une matrice stochastique. C'est donc la matrice de transition d'une chaîne de Markov en temps discret, mettons $\{Y(n), n \in \mathbb{N}\}$.

Du point de vue stochastique, le lien entre les deux processus est le suivant:

Théorème 24. *Soit un processus de Poisson de paramètre ν: $\{T_n, n \in \mathbb{N}\}$. Alors le processus $\{X(t), t \in \mathbb{R}^+\}$ défini par:*

$$X(t) \;=\; Y(n) \qquad T_n \leq t < T_{n+1}$$

est une Chaîne de Markov en temps continu dont le générateur est $\mathbf{Q}$.

En d'autres termes: le processus $\{X(t)\}$ peut être vu comme changeant d'état aux instants $T(n)$ d'un processus de Poisson d'intensité constante ν, la matrice de transition étant $\mathbf{P}$. Il peut se faire que le processus ne change en fait pas d'état, dans le cas où pour un état i, $P_{ii} \neq 0$.

Preuve. Calculons $P(t)_{i,j}$ pour le processus $X(t)$. Conditionnons sur le nombre d'arrivées du processus de Poisson $\{T_n\}$ dans l'intervalle $[0,t[$. On obtient:

$$\begin{aligned}
\mathbb{P}(X(t) = j | X(0) = i) \;&=\; \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; \mathbb{P}(X(t) = j \mid X(0) = i, k \text{ sauts de } Y \text{ dans } [0,t[) \\
&=\; \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; \mathbb{P}(Y(k) = j \mid X(0) = i) \\
&=\; \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\; e^{-\nu t}\; (\mathbf{P})^k_{i,j} \;.
\end{aligned}$$

Donc on a:

$$P(t) \;=\; \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!}\, e^{-\nu t}\, \mathbf{P}^k \;=\; e^{\nu \mathbf{P}}\, e^{-\nu \mathbf{I}} \;=\; e^{\nu(\mathbf{P}-\mathbf{I})} \;.$$

On a donc bien $P(t) = e^{\mathbf{Q}t}$. □

Du point de vue algébrique, si on prend dans (53) un nombre $\nu > \max\{|Q(i,i)|\}$, alors la matrice $\mathbf{P}$ est irréductible si $\mathbf{Q}$ l'est, et $\mathbf{P}$ est apériodique. D'autre part, le spectre de $\mathbf{P}$ et celui de $\mathbf{Q}$ sont reliés:

$$\lambda \in \mathrm{sp}(\mathbf{Q}) \iff 1 + \frac{\lambda}{\nu} \in \mathrm{sp}(\mathbf{P}) \;.$$

Le spectre de $\mathbf{Q}$ est donc situé dans le demi-plan complexe $\{\Re(z) \le 0\}$, dans un cercle de rayon ν centré au point $-\nu$. En particulier, 0 est toujours une valeur propre de $\mathbf{Q}$. Nous le savions déjà puisque $\mathbf{Q1} = \mathbf{0}$.

5.5 Comportement asymptotique

Intéressons-nous aux probabilités stationnaires et à la convergence des probabilités $\boldsymbol{\pi}_t$.

Dans le cas où $\mathcal{E}$ est fini, on a un théorème similaire au Théorème 15.

Théorème 25. *Soit $\mathbf{Q}$ un générateur infinitésimal irréductible. Soit $\boldsymbol{\pi}$ le vecteur propre gauche de $\mathbf{Q}$ pour la valeur propre 0, tel que $\boldsymbol{\pi}.\mathbf{1} = 1$. Alors*

$$\lim_{t\to\infty} e^{\mathbf{Q}t} \;=\; \mathbf{1}.\boldsymbol{\pi} \;. \tag{54}$$

Par conséquent, quel que soit le vecteur $\boldsymbol{\pi}_0$, on a:

$$\lim_{t\to\infty} \boldsymbol{\pi}_t \;=\; \boldsymbol{\pi} \;. \tag{55}$$

La convergence est géométrique:

$$||\boldsymbol{\pi}_t - \boldsymbol{\pi}|| \;=\; \mathcal{O}(e^{\phi t}) \tag{56}$$

pour tout ϕ tel que $\Re(\lambda_2) < \phi < 0$, où λ_2 est la valeur propre de $\mathbf{Q}$ de plus grande partie réelle après 0.

Dans le cas où $\mathcal{E}$ est infini, on a un résultat similaire au Théorème 18:

Théorème 26. *Soit $\mathbf{Q}$ un générateur infinitésimal infini et irréductible. Deux possibilités seulement existent:*

- *soit le système $\mathbf{xQ} = \mathbf{0}$ admet une solution $\mathbf{x}$ positive telle que $\sum_{i\in\mathcal{E}} x_i < \infty$, et donc un vecteur de probabilités stationnaires $\boldsymbol{\pi}$.*
  Alors $\boldsymbol{\pi}_t \to \boldsymbol{\pi}$ pour toute distribution initiale $\boldsymbol{\pi}_0$.
- *soit il n'existe aucune solution de $\mathbf{xQ} = \mathbf{0}$ qui soit telle que $\sum_{i\in\mathcal{E}} |x_i| < \infty$.*
  Alors $\boldsymbol{\pi}_t \to \mathbf{0}$ pour toute distribution initiale $\boldsymbol{\pi}_0$.

6 Éléments de théorie des files d'attente markoviennes

On va faire une première application de la théorie des chaînes de Markov à l'étude de quelques problèmes de la théorie des files d'attente.

6.1 La file "$M/M/1/K$"

Soit une file d'attente ayant le comportement suivant:

- Les clients arrivent selon un processus de Poisson de taux λ;
- chaque client reçoit une quantité de service dont la distribution est exponentielle de paramètre μ;
- les clients sont servis un par un dans l'ordre de leur arrivée,

– il y a au plus K clients en attente; si un nouveau client se présente alors que la file d'attente est pleine, il est rejeté.

Soit $N(t)$ le nombre de clients dans la file à l'instant t. Alors en vertu de la construction # 2 (Théorème 23), le générateur infinitésimal de ce processus est:

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & & \\ \mu & -\lambda-\mu & \lambda & \ddots & \\ 0 & \ddots & \ddots & \ddots & \\ & & \mu & -\lambda-\mu & \lambda \\ & & & \mu & -\mu \end{pmatrix} .$$

Si on pose $\nu = \lambda + \mu$ et qu'on uniformise $\mathbf{Q}$ selon le procédé du paragraphe 5.4, alors en notant:

$$p = \frac{\mu}{\lambda+\mu} , \qquad q = \frac{\lambda}{\lambda+\mu} ,$$

on trouve que la chaîne a pour matrice de transition la matrice $\mathbf{P}$ de (35).

On déduit donc des théorèmes généraux et des calculs du paragraphe 4.5 la propriété suivante de la file d'attente M/M/1/K.

Théorème 27. *Le processus $N(t)$ admet une distribution stationnaire $\boldsymbol{\pi}$ donnée par: si $\lambda \neq \mu$,*

$$\pi_k = \left(\frac{\lambda}{\mu}\right)^k \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}} ,$$

et si $\lambda = \mu$:

$$\pi_k = \frac{1}{K+1} ,$$

pour $k = 0, 1, \dots, K$.

6.2 La file "$M/M/1$"

La file d'attente $M/M/1$ a la même description que la file $M/M/1/K$ sauf que la restriction sur le nombre de clients dans la file d'attente est abandonnée. On dit que c'est une file à capacité illimitée.

Le processus $\{N(t), t \in \mathbb{R}^+\}$ est toujours une Chaîne de Markov en temps continu, mais son espace d'état est maintenant $\mathbb{N}$. Son générateur infinitésimal s'écrit:

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & & \\ \mu & -\lambda-\mu & \lambda & \ddots & \\ 0 & \mu & -\lambda-\mu & \lambda & \ddots \\ & \ddots & \ddots & \ddots & \end{pmatrix} .$$

Ce processus est aussi *processus de naissance et de mort*, en temps continu. Il ressemble beaucoup au processus étudié au paragraphe 4.5.3. En fait, comme pour la file $M/M/1/K$ vue précédemment, l'uniformisation de ce processus avec la valeur $\nu = \lambda + \mu$ et l'utilisation des variables $p = \mu/(\lambda+\mu)$ et $q = 1 - p$, donne la matrice de (35) dans sa forme infinie. On peut bien sûr se servir de cette remarque pour calculer la distribution stationnaire de la chaîne.

Mais disons qu'on ne dispose pas de ces résultats. Le calcul direct de la probabilité stationnaire s'appuie ici sur le Théorème 26. Une distribution stationnaire pour le processus $\{N(t)\}$ doit vérifier: $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$. Ce système linéaire n'est autre que:

$$\begin{cases} \lambda\pi_0 = \mu\pi_1 & \\ (\lambda+\mu)\pi_k = \lambda\pi_{k-1} + \mu\pi_{k+1} & k \geq 1 . \end{cases}$$

Comme au paragraphe 4.5.1, on peut résoudre cette récurrence par la méthode des séries génératrices (bien que d'autres méthodes, comme "intuition+récurrence" soient plus faciles dans ce cas).

On introduit donc la fonction génératrice de la loi de probabilité:

$$\pi(z) \;=\; \sum_{n=0}^{\infty} z^k \pi_k \;.$$

Noter que cette fonction n'est plus un polynôme de la variable z. Par contre, on sait (Lemme 1) que c'est une fonction analytique au moins dans le disque unité, et que $\pi(1) = 1$. Après substitution des équations d'équilibre dans la définition de $\pi(z)$ et simplification, on aboutit à:

$$\pi(z) \;=\; \frac{\mu\pi_0}{\mu - \lambda z} \;, \tag{57}$$

ce qui n'est guère surprenant dans la mesure où c'est ce qu'on obtiendrait en laissant $K \to \infty$ dans (36) pour un z de module strictement inférieur à 1.

Ici, seul π_0 est inconnu et la condition $\pi(1) = 1$ donne directement $\pi_0 \;=\; 1 - \lambda/\mu$. On conclut donc quant a la distribution stationnaire $\boldsymbol{\pi}$:

- Si $\lambda < \mu$, il existe une unique solution $\boldsymbol{\pi}$ de $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$:

$$\pi_k \;=\; \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^k \;, k \geq 0 \;.$$

- Si $\lambda \geq \mu$, il n'existe aucune fonction analytique dans le disque unité, non identiquement nulle et qui soit solution de (57). Par conséquent, il n'existe pas de distribution de probabilité $\boldsymbol{\pi}$ solution de $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$.

La condition $\lambda < \mu$ est dite *condition de stabilité* de la file $M/M/1$. Elle s'interprète par la necessité qu'en moyenne, la quantité de clients qui arrive au système dans une unité de temps doit être inférieure à la quantité moyenne de clients que le serveur peut servir dans ce même temps.

6.3 La file "M/M/∞"

Dans cette file d'attente, il y a une infinité de serveurs. Le processus $\{N(t), t \in \mathbb{R}^+\}$ est toujours une Chaîne de Markov en temps continu. Son générateur infinitésimal s'écrit:

$$\mathbf{Q} \;=\; \begin{pmatrix} -\lambda & \lambda & 0 & & \\ \mu & -\lambda-\mu & \lambda & \ddots & \\ 0 & 2\mu & -\lambda-2\mu & \lambda & \ddots \\ & \ddots & \ddots & \ddots & \end{pmatrix} .$$

Toujours en admettant que la théorie établie s'applique dans le cas présent (on a: $\sup_i\{|Q_{ii}|\} = \infty$), une distribution stationnaire pour le processus $\{N(t)\}$ doit vérifier:

$$\left\{ \begin{array}{rcll} \lambda\pi_0 & = & \mu\pi_1 & \\ (\lambda + k\mu)\pi_k & = & \lambda\pi_{k-1} + (k+1)\mu\pi_{k+1} & \quad k \geq 1 \;. \end{array} \right.$$

Le calcul de la fonction génératrice $\pi(z) = \sum_{n=0}^{\infty} z^k \pi_k$ nous mène ici à l'équation différentielle (linéaire et homogène):

$$\lambda\pi(z) + \mu z \pi'(z) \;=\; \lambda z \pi(z) + \mu\pi'(z)$$

soit donc:

$$\pi'(z) \;=\; \frac{\lambda}{\mu}\,\pi(z) \;.$$

La solution de cette équation est bien sûr, en rajoutant la condition $\pi(1) = 1$ et en notant $\rho = \lambda/\mu$:

$$\pi(z) \;=\; e^{\rho(z-1)} \;.$$

Autrement dit, la distribution stationnaire de la file M/M/∞ est:

$$\pi(n) \;=\; \frac{\rho^n}{n!}\, e^{-\rho}, \quad n \geq 0 \;,$$

c'est-à-dire une distribution de Poisson de paramètre ρ. Cette distribution existe toujours, pourvu que $\mu > 0$.

Il se trouve que cette distribution de Poisson de paramètre ρ est la distribution stationnaire de la file M/GI/∞. Le résultat ne dépend donc pas de la distribution du temps de service, mais uniquement de sa moyenne. Ce phénomène (rare) porte le nom d'*insensibilité*.

6.4 Les temps de réponse et la formule de Little

Jusqu'ici, le calcul s'est concentré sur le nombre de clients dans le système, sa distribution et sa moyenne. D'autres mesures de performances sont intéressantes, et en particulier le *temps de réponse*, défini comme le temps qui s'écoule entre l'entrée su client dans le système et sa sortie définitive.

Si on ne s'intéresse qu'à la moyenne de cette quantité, on peut utiliser la formule de Little.

Soit un système de files d'attente *stable*. Soit:

- $\overline{N}$ le nombre moyen de clients,
- $\overline{T}$ le temps moyen de réponse,
- λ le débit entrant (et sortant) des clients,

dans ce système. Alors:

$$\overline{N} \;=\; \lambda\, \overline{T} \;. \tag{58}$$

Bien sûr, ce résultat ne s'applique que des conditions précises, mais ceci sort du cadre de ce cours. En pratique, du moment que le système est "stable", on peut appliquer le résultat.

Si on revient sur les files d'attente précédemment analysées, on trouve que pour la $M/M/1$, on a:

$$\overline{T} \;=\; \frac{1}{\mu - \lambda} \;.$$

Pour la $M/M/\infty$, on a:

$$\overline{T} \;=\; \frac{1}{\mu} \;,$$

ce qui est naturel puisque dans ce cas, temps de réponse et temps de service coïncident.

7 Les Réseaux de files d'attente

Dans ce paragraphe, je donne sans détails ni rigeur mathématique, les principaux résultats qui sont connus sur les *réseaux* de files d'attente.

Pour attractifs qu'ils soient, les modèles à file d'attente unique ne permettent pas de modéliser un système un peu complexe. Par exemple, le problème du passage aux guichets d'une poste ne se résoud pas avec une seule file, car les serveurs sont indépendants, en général non interchangeables, et chacun a sa propre file d'attente. L'idée est donc de représenter chaque attente par une file simple, et de connecter ces files entre elles dans un réseau.

Les questions que l'on se pose sont toujours les mêmes: étant donnés les paramètres du modèle, quelle est la distribution du nombre de clients dans un site donné (ou à défaut sa moyenne), que valent les temps d'attente, quelles sont les conditions de stabilité du système, etc...

Le problème est évidemment plus compliqué que dans le cas d'une file simple. On ne peut pas espérer obtenir de résultats sans faire d'hypothèses restrictives (mais le moins possible) sur les lois d'arrivée, de service, le type des files simples qui constituent le réseau, le *routage* (la façon dont les clients se déplacent dans le réseau) ou la topologie de ce réseau. Dans certains cas que nous allons voir, les résultats obtenus dans la première section peuvent être généralisés.

7.1 Les réseaux de Jackson

Les réseaux de Jackson ont pour particularité la spécification du routage des clients de manière *aléatoire*. Les paramètres d'un tel réseau sont:

- le nombre N de stations,
- le vecteur $\vec{\lambda}^0 = (\lambda_1^0, \ldots, \lambda_N^0)$ des taux d'arrivées extérieures dans chaque file,
- le vecteur $(\mu_1, \ldots, \mu_N)$ des taux de service,
- la matrice carrée $N \times N$ de routage interne R, dont chaque composante $r_{i,j}$ est la probabilité qu'un client sortant de i aille vers j. La probabilité qu'il sorte du système est alors $r_{i,0} = 1 - \sum_j r_{i,j}$.

La matrice $\mathbf{R}$ est une matrice sous-stochastique. On peut la "compléter" en une matrice stochastique en rajoutant un état "0" absorbant, et les probabilités de routage $r_{i,0}$ définies ci-dessus. On fait l'hypothèse que dans cette matrice, les états $1, 2, \ldots, N$ sont tous *transients*. Ceci revient à dire que tout client a toujours un chemin qui mène à l'extérieur, et finit donc fatalement par le prendre (voir Théorème 16.*i/*).

Donc par exemple, si $\mathbf{R}$ est irréductible, on demande qu'il existe au moins un $r_{i,0} > 0$. Ou bien, si $\mathbf{R}$ se décompose, que tout bloc diagonal ait cette propriété (voir §4.2.5).

Des règles d'évolution énoncées ci-dessus, et en utilisant la Construction #2, on obtient facilement que le processus $N(t) = (N_1(t), \ldots, N_N(t))$ qui décrit le nombre de clients dans chaque file d'attente du réseau au temps t, est une Chaîne de Markov en temps continu sur $\mathbb{N}^N$. En utilisant une notation vectorielle: $\mathbf{n} = (n_1, \ldots, n_N)$ et $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ pour le i-ième vecteur unitaire, on décrit le générateur $\mathbf{Q}$ par le tableau de transitions:

$$\begin{array}{lcll} \mathbf{n} & \to & \mathbf{n} + \mathbf{e}_i & \text{avec taux } \lambda_i^0 \\ \mathbf{n} & \to & \mathbf{n} - \mathbf{e}_i & \text{avec taux } \mu_i r_{i0} \quad n_i > 0 \\ \mathbf{n} & \to & \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j & \text{avec taux } \mu_i r_{ij} \quad n_i > 0 \ . \end{array}$$

On a alors:

Théorème 28 (de Jackson pour réseaux ouverts). *Soit le vecteur $\vec{\lambda} = (\lambda_1, \ldots, \lambda_N)$ solution du système linéaire:*

$$\vec{\lambda} \;=\; \vec{\lambda}^0 \;+\; \vec{\lambda}\,\mathbf{R}\ . \tag{59}$$

Si la condition (de stabilité):

$$\forall 1 \leq i \leq N, \qquad \lambda_i \;<\; \mu_i.$$

est satisfaite, alors le processus $\{N(t), t \in \mathbb{R}\}$ admet une distribution stationnaire unique donnée par:

$$p(n_1, \ldots, n_N) \;=\; \prod_{i=1}^{N} \left(1 - \frac{\lambda_i}{\mu_i}\right) \left(\frac{\lambda_i}{\mu_i}\right)^{n_i} \ .$$

Preuve. La preuve du Théorème de Jackson et des théorèmes similaires se fait en général en vérifiant simplement que la distribution proposée satisfait aux équations d'équilibre. On doit alors avoir pour tout vecteur $\mathbf{n}$:

$$\begin{aligned} \boldsymbol{\pi}(\mathbf{n}) \left(\sum_{i=0}^{N} \mu_i \mathbf{1}_{\{n_i>0\}} + \lambda_i^0 \right) \;&=\; \sum_{j=0}^{N}\sum_{i=0}^{N} \boldsymbol{\pi}(\mathbf{n} + \mathbf{e}_j - \mathbf{e}_i)\mu_i r_{j,i} \mathbf{1}_{\{n_i>0\}} \\ &\quad + \sum_{j=1}^{N} \boldsymbol{\pi}(\mathbf{n} + \mathbf{e}_j)\mu_j r_{j,0} \;+\; \sum_{j=0}^{N} \boldsymbol{\pi}(\mathbf{n} - \mathbf{e}_j)\lambda_j^0 \mathbf{1}_{\{n_j>0\}}\ . \end{aligned} \tag{60}$$

Ce qu'on peut vérifier avec un peu de patience en faisant usage de (59). L'unicité découle du fait que la chaîne est irréductible (justifier!) et du Théorème 26. □

Exercice: justifier le système (59) qui définit $\vec{\lambda}$.

Ces équations portent le nom d'*équations de trafic*. Elles admettent une solution unique sous la condition faite sur $\mathbf{R}$. Le vecteur λ ainsi défini est le vecteur des *débits* des clients entrant dans chacune des files d'attente.

Réseaux fermés. Il existe aussi des systèmes dans lesquels le nombre de clients est fixé. Un exemple classique est le problème machines/réparateurs, où les machines d'un site (en nombre fini, hélas) sont susceptibles de tomber en panne et sont réparées par des personnes compétentes (en nombre fini aussi); les réparateurs sont alors représentés par des serveurs et les machines par des clients, qui alternativement passent par une période de bon fonctionnement, puis vont se faire réparer par un réparateur libre.

Les paramètres d'un tel réseau sont:
- le nombre N de stations,
- le vecteur $(\mu_1, \dots, \mu_N)$ des taux de service,
- la matrice carrée $N \times N$ de routage interne R, stochastique,
- le nombre K de clients présents.

Comme aucun client n'entre ni ne sort, le nombre de clients total dans le réseau est conservé. Comme précedemment, $\{N(t), t \in \mathbb{R}\}$ est une CMTC, mais son espace d'états est (au plus): $\mathcal{E} = \{\mathbf{n} \in \mathbb{N}^N \mid n_1 + \dots + n_N = K\}$.

On suppose que $\mathbf{R}$ est irréductible, de telle sorte que chaque client visite toutes les files d'attente du réseau. Si $\mathbf{R}$ est réductible, alors on analyse séparément les classes terminales (voir 4.2.2).

Théorème 29 (Jackson pour les réseaux fermés). *Soit un réseau de Jackson fermé, contenant K clients. Soit $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ une solution positive du système linéaire:*

$$\boldsymbol{\theta} \;=\; \boldsymbol{\theta}\, \mathbf{R} \,. \tag{61}$$

Le processus $\{N(t), t \in \mathbb{R}\}$ admet une distribution de probabilité stationnaire donnée par:

$$p(n_1, \dots, n_N) \;=\; G \prod_{i=1}^{N} \left(\frac{\theta_i}{\mu_i}\right)^{n_i} \mathbf{1}_{\{n_1+\dots+n_N=K\}} \;,$$

où:

$$G \;=\; \left[\sum_{\sum n_i = K} p(n_1, \dots, n_N)\right]^{-1} \,. \tag{62}$$

Plusieures remarques sur ces résultats.

– L'équation (61) a une infinité de solutions. Cependant, comme l'ensemble des solutions forme une droite vectorielle, on peut facilement vérifier que le résultat ne dépend pas du choix de la solution.

– Pour le même raison, et contrairement au cas ouvert, l'équation de trafic (61) ne donne pas la valeur exacte du débit entrant dans chaque file d'attente. Pour obtenir cette valeur, on utilise l'équation:

$$\lambda_i \;=\; \mu_i \; \mathbb{P}\{n_i > 0\} \,.$$

Il suffit d'en calculer un pour obtenir le vecteur entier si on utilise (61).

– Le calcul des probabilités et des autres mesures de performances (nombre moyen de clients, débits) passe par le calcul de la constante de normalisation G par (62). Il s'agit d'une somme avec un nombre de termes $\binom{N+K-1}{N}$, ce qui peut faire beaucoup. Il existe toutefois des algorithmes récursifs rapides pour ce calcul.

Il est possible de généraliser ces résultats en affaiblissant les hypothèses sur les disciplines et les lois de service. Ainsi, le résultat principal, à savoir l'existence d'une *forme produit*, reste valable quand les stations sont gouvernées par certaines valeurs plus générales de la discipline de service et de la loi de service. Il est également permis d'attribuer des classes aux clients, et des lois de service dépendant de la classe. Ces réseaux généralisant ceux de Jackson sont dits “BCMP” d'après les initiales de leurs découvreurs.

Il n'est pas question ici d'entrer dans les détails de ces réseaux. Le lecteur intéressé se reportera à la bibliographie (par exemple [9]).

7.2 Les réseaux de Kelly. Les modèles à commutation de paquets.

Les réseaux de Jackson et BCMP ont comme caractéristique principale un routage aléatoire. Un client suit dans le réseau un chemin qui n'est pas déterministe.

Mais il existe des systèmes où, au contraire, les clients doivent suivre une route bien définie dans le réseau. Par exemple, les paquets de voix/données dans un réseau de communication numérique, ou bien les pièces à usiner dans un atelier flexible.

Pour ce type de systèmes, on dispose des *Réseaux de Kelly.* Dans ce modèle, on suppose que les clients appartiennent à des classes, et qu'à chaque classe k est affecté une route dans le réseau $r_k = (r_k^1, \ldots, r_k^{n_k})$, qui est la suite des serveurs que le client doit visiter avant de sortir. Eventuellement, cette liste peut contenir plusieurs fois la même station. Comme pour les réseaux de Jackson, les clients arrivent suivant des processus de Poisson, et les serveurs délivrent des temps de service exponentiels.

Considérons un réseau de N files. Soit, comme plus haut, μ_i le taux de service de la file i. Soit K le nombre de classes, et λ_k le taux d'arrivées de clients de la classe k.

Connaissant les taux d'arrivées des clients de chaque classe, et tous les trajets correspondants, on peut calculer le flux $\hat{\lambda}_{ik}$ de clients de classe k entrant dans la file d'attente i:

$$\hat{\lambda}_{ik} \;=\; \lambda_k \;\times\; (\text{nombre de } i \text{ dans } r_k),$$

puis le flux total, toutes classes confondues:

$$\hat{\lambda}_i \;=\; \sum_k \hat{\lambda}_{ik} \;.$$

L'espace d'état du réseau peut être représenté par l'ensemble $\mathcal{M}_{NK}$ des matrices $N \times K$ à coefficients entiers: le coefficient m_{ik} représente alors le nombre de clients de classe k dans la file d'attente i.

On a alors un théorème donnant la probabilité d'être dans un état donné:

Théorème 30 (de Kelly). *Soit M une matrice de $\mathcal{M}_{NK}$. Si la condition de stabilité*

$$\hat{\lambda}_i \;<\; \mu_i \qquad \forall i$$

est satisfaite, alors la probabilité stationnaire que le réseau se trouve dans l'état M existe et vaut:

$$\mathbb{P}(M) \;=\; \prod_{i=1}^{N} \left(1 - \frac{\hat{\lambda}_i}{\mu_i}\right) \left(\sum_{k=1}^{K} m_{ik}\right)! \prod_{k=1}^{K} \frac{1}{m_{ik}!} \left(\frac{\hat{\lambda}_{ik}}{\mu_i}\right)^{m_{ik}} .$$

Quelques remarques sur cette formule:

- C'est toujours une forme produit, et chaque terme est la probabilité d'état de la file i en isolation.
- En fait, s'il n'y a qu'une classe, c'est la même formule que dans le Théorème de Jackson. Le mécanisme de routage (et la façon de calculer les débits) n'est pourtant pas le même.
- Le processus $\{M(t), t \in \mathbb{R}\}$ n'est pas, en général, une chaîne de Markov. En effet, étant donné le routage des clients selon leur route et même l'étape de leur route, il est nécessaire de savoir quel client est en service pour prévoir l'évolution du système. Par extension, il faut en fait l'ordre des clients dans la file d'attente.

Exercice combinatoire: la quantité $(\sum_1^K m_{ik})!/\prod_1^K m_{ik}!$ représente un certain nombre de configurations: lequel?

A partir de cette formule, on peut calculer toutes les statistiques que l'on veut, comme par exemple le nombre moyen de clients dans la file i:

$$\overline{N}_i \;=\; \frac{\hat{\lambda}_i}{\mu_i - \hat{\lambda}_i}$$

et de là, à l'aide de la formule de Little (58), le temps moyen passé par un client de classe k dans le réseau:

$$\overline{T}_k \;=\; \frac{1}{\lambda_k} \sum_{j=1}^{n_k} \frac{\hat{\lambda}_{r_k^j,k}}{\mu_{r_k^j} - \hat{\lambda}_{r_k^j}} \;.$$

Exercice calculatoire: quel est le nombre moyen de clients de classe k dans la file i?

Comme pour les réseaux de Jackson, il existe une version du théorème de Kelly pour les réseaux fermés, où les clients de classe k suivent le chemin r_k de manière cyclique. Les résultats sont similaires à ceux des réseaux de Jackson, avec l'apparition de la constante de normalisation G. Nous n'entrerons pas dans les détails.

8 Processus de Branchement

Les processus de branchement sont une classe particulière de chaînes de Markov. Ils sont utiles dans la modélisation de procesus naturels qui peuvent être vus comme des arbres aléatoires. Ces arbres modélisent également le déroulement de certains algorithmes ou de certains programmes parallèles.

Il y a de nombreuses variantes. Dans sa version la plus simple (processus de Galton-Watson), le processus se décrit de la manière suivante.

Soit X une distribution discrète à valeurs dans $\mathbb{N}$, de moyenne m, de variance σ^2 et de fonction génératrice ϕ. On considère que dans une population, un individu peut donner naissance à un nombre aléatoire de descendants directs, distribués selon la distribution de X. Les descendances (directes) d'individus distincts sont stochastiquement indépendantes.

8.1 Distribution du nombre de descendants par génération

On suppose qu'il y a un individu unique au départ, et on dénombre ses descendants par générations.

Soit $Z(n)$ le nombre de descendants à la génération n. On voit facilement que le processus $\{Z(n), n \in \mathbb{N}\}$ est un chaîne de Markov à valeurs dans $\mathbb{N}$. Les fonctions génératrices des $Z(n)$ vérifient:

$$\begin{aligned}
F_{n+1}(z) \ := \ \mathbb{E}(z^{Z(n+1)}) \ &= \ \sum_{k=0}^{\infty} \mathbb{P}(Z(n)=k)\ \mathbb{E}(z^{Z(n+1)} | Z(n)=k) \\
&= \ \sum_{k=0}^{\infty} \mathbb{P}(Z(n)=k)\ \mathbb{E}(z^{\sum_{j=1}^{k} X_j}) \\
&= \ \sum_{k=0}^{\infty} \mathbb{P}(Z(n)=k)\ \phi(z)^k \\
&= \ F_n(\phi(z)) \ .
\end{aligned}$$

Par conséquent, par récurrence:

$$F_n(z) \ = \ \underbrace{\phi \circ \phi \circ \ldots \circ \phi}_{n \text{ fois}}(z) \ = \ \phi^{(n)}(z) \ , \quad \text{et} \quad F_0(z) \ = \ 1. \tag{63}$$

En général, on ne sait pas résoudre les équations fonctionnelles récurrentes obtenues. On peut cependant en extraire des informations sur les moments (ci-après) ou de nature asymptotique (§8.4).

En dérivant (63) par rapport à z, on déduit une récurrence qui conduit aux premiers moments:

$$\mathbb{E}(Z(n)) \ = \ m^n, \qquad \mathrm{Var}(Z(n)) \ = \ \sigma^2 \ \frac{m^n(m^n-1)}{m(m-1)} \ . \tag{64}$$

Si $m=1$, il faut lire: $\mathrm{Var}(Z(n)) \ = \ n\sigma^2$.

8.2 Distribution du nombre total de descendants

Soit $T(n) = \sum_{m=0}^{n} Z(m)$ le nombre total de descendants dans l'arbre jusqu'à la génération n. On s'intéresse à la distribution limite de la variable aléatoire $T(n)$ quand $n \to \infty$. Cette limite existe car la suite

$T(n)$ est stochastiquement croissante. Mais elle peut être “impropre”, c'est-à-dire donner une probabilité non nulle à l'infini.

En utilisant (64), il est clair que:

$$\mathbb{E}T(n) \;=\; \frac{m^{n+1}-1}{m-1} \qquad (n+1 \text{ si } m=1)\;.$$

Par conséquent, si n croît, alors:

- si $m \geq 1$, le nombre moyen de descendants tend vers l'infini, ce qui suggère que l'arbre devient infini avec une probabilité non nulle, ou bien qu'il reste fini mais avec une taille d'espérance infinie;
- si $m < 1$, le nombre moyen de descendants converge vers $1/(1-m)$, ce qui implique que l'arbre se termine avec probabilité 1.

Soit T la variable aléatoire qui est la taille de l'arbre quand n tend vers l'infini. Comme le processus de génération est markovien, chaque descendant de la première génération donne naissance à un arbre de même distribution que l'arbre complet. Les sous-arbres engendrés sont indépendants. On a donc, en distribution:

$$T \;=_d\; 1 \;+\; \sum_{i=1}^{X} T_i\;,$$

où les T_i sont des v.a. indépendantes de même distribution que T. Par conséquent, en passant aux fonctions génératrices:

$$T^*(z) \;:=\; \mathbb{E}(z^T) \;=\; z\;\phi(T^*(z))\;. \tag{65}$$

8.3 Exemples

Bien que l'équation (63) soit difficile à résoudre en général, on peut parfois calculer exactement la distribution de $Z(n)$.

Le cas géométrique. Si X a une distribution géométrique: $\mathbb{P}(X=k)=\rho^k(1-\rho)$, de moyenne $m = \rho/(1-\rho)$. Comme:

$$\phi(z) \;=\; \frac{1-\rho}{1-\rho z}\;,$$

on prouve par récurrence que

$$F_n(z) \;=\; (1-\rho)\;\frac{(1-\rho)^{n-1}+a_n(1-z)}{(1-\rho)^n+b_n(1-z)}\;.$$

En exercice, trouver les suites a_n et b_n, puis vérifier que $\mathbb{E}Z(n)=m^n$.

D'après (65), la fonction génératrice du nombre total de descendants est solution de:

$$T^*(z) \;=\; z\;\frac{1-\rho}{1-\rho T^*(z)}\;.$$

Si $m \leq 1$ (c'est-à-dire $\rho \leq 1/2$), il existe une unique solution de cette équation qui soit une fonction analytique dans le disque unité:

$$T^*(z) \;=\; \frac{1-\sqrt{1-4\rho(1-\rho)z}}{2\rho}\;.$$

En exercice: inverser cette fonction génératrice pour trouver $\mathbb{P}(T=k)$.

Le cas des chaînes. Si par exemple X vaut 1 avec proba m et 0 avec proba $1-m$, ce qui correspond à $\phi(z) = 1 - m + mz$, alors:

$$\begin{aligned} F_2(z) &= (1 - m + m(1 - m + mz)) = 1 - m^2 + m^2 z \\ F_3(z) &= (1 - m + m(1 - m^2 + m^2 z)) = 1 - m^3 + m^3 z \\ F_n(z) &= 1 - m^n + m^n z, \quad n \geq 0. \end{aligned}$$

Pour la taille totale de la chaîne, on trouve que:

$$T^*(z) = \frac{1-m}{1-mz} .$$

Dans ce cas, on retrouve que l'"arbre" est une chaîne de longueur géométriquement distribuée.

Arbres binaires. Si enfin X vaut 2 ou 0 avec probas $m/2$ et $1 - m/2$, on a:

$$F_n(z) = 1 - \frac{m}{2} + \frac{m}{2} F_{n-1}(z)^2 .$$

Quand $m < 1$, la solution de (65) s'écrit ici:

$$\begin{aligned} T^*(z) &= \frac{1 - \sqrt{1 - m(2-m)z^2}}{mz} \\ &= \sum_{k=0}^{\infty} \frac{1}{k+1} \binom{2k}{k} \left(\frac{m}{2}\right)^k \left(1 - \frac{m}{2}\right)^{k+1} z^{2k+1} . \end{aligned}$$

Cette formule était à prévoir car il est clair que la probabilité que l'arbre ait k nœuds internes, et donc $k+1$ feuilles (soit en tout $2k+1$ nœuds) est la même quelle que soit sa forme, et vaut:

$$\mathbb{P}(k \text{ nœuds avec deux descendants, } k+1 \text{ sans descendants}) = \left(\frac{m}{2}\right)^k \left(1 - \frac{m}{2}\right)^{k+1} .$$

Or il est bien connu qu'il y a

$$\frac{1}{1+k} \binom{2k}{k}$$

tels arbres (nombre de Catalan).

Observer que quand $m = 1$, la distribution de T est propre et sa fonction génératrice est:

$$T^*(z) = \frac{1 - \sqrt{1 - z^2}}{z} .$$

Cette variable aléatoire a une espérance infinie.

8.4 Propriétés asymptotiques

La première question que l'on se pose sur un arbre aléatoire est celle de sa finitude.

La *probabilité d'extinction* de l'arbre, π_e est la probabilité qu'il existe un n tel que $Z(n) = 0$. Clairement, si $Z(n) = 0$ alors $Z(m) = 0$ pour tout $m \geq n$.

L'événement $\{\exists n | Z(n) = 0\}$ est égal à $\cup_n \{Z(n) = 0\}$. Comme les événements $\{Z(n) = 0\}$ sont inclus les uns dans les autres, on a:

$$\pi_e = \lim_{n \to \infty} \mathbb{P}(Z(n) = 0) .$$

La probabilité $\mathbb{P}(Z(n) = 0)$ vaut, d'après (63),

$$\pi_e = \lim_{n \to \infty} \phi^{(n)}(0) . \tag{66}$$

On a alors le résultat:

Théorème 31. *Si $m \leq 1$, alors $\pi_e = 1$ et le processus de branchement s'éteint presque sûrement. Si $m > 1$, alors la probabilité d'extinction est la plus petite solution x de l'équation*

$$\phi(x) \; = \; x \; .$$

Preuve. La fonction ϕ est continue, croissante et convexe sur l'intervalle $[0, 1]$, et $\phi(1) = 1$. Deux cas se présentent. Soit $\phi(0) = 0$. Dans ce cas, chaque individu a toujours au moins un descendant et l'arbre ne peut pas s'éteindre. Dans ce cas, on a forcément $m \geq 1$. Dans l'autre cas, $\phi(0) > 0$. Puisque ϕ est croissante, on a $\phi^{(2)}(0) > \phi(0)$ et la suite $\phi^{(n)}(0)$ est donc croissante et bornée par 1. Elle converge donc vers le plus petit point fixe de ϕ qui soit > 0. Si $m = \phi'(0) < 1$, alors $\phi(x) < x$ au voisinage de $x = 1$, donc il existe un point fixe de ϕ dans $]0, 1[$. Par contre, si $m \geq 1$, comme ϕ est convexe, le plus petit point fixe de ϕ est 1. □

Exemple: dans le cas où X vaut 2 ou 0 avec probas $m/2$ et $1-m/2$, on trouve que $\pi_e = 1-m/2+(m/2)\pi_e^2$, d'où:

$$\pi_e \; = \; \frac{2-m}{m} \; ,$$

dans le cas $m \geq 1$.

Références

[1] M. Ajmone Marsan, G. Balbo, and G. Conte. *Performance models of multiprocessor systems.* MIT Press, 1986.

[2] F. Baccelli, G. Cohen, G.J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity.* Wiley, 1992.

[3] W. Feller. *An introduction to probability theory and its applications*, volume I. J. Wiley & Sons, New York, troisième edition, 1968.

[4] T.E. Harris. *The theory of branching processes.* Dover, 1989.

[5] J.G. Kemeny and Snell J.L. *Finite Markov Chains*, volume 76. Springer-Verlag, 1976.

[6] L. Kleinrock. *Queueing Systems*, volume I, Theory. J. Wiley & Sons, New York, 1975.

[7] H. Minc. *Nonnegative Matrices.* J. Wiley & Sons, 1988.

[8] T. Murata. Petri nets: properties, analysis and applications. *Proceedings of the IEEE*, 77:541–580, 1989.

[9] G. Pujolle and S. Fdida. *Modèles de Systèmes et de Réseaux – Tome I: Performance.* Eyrolles, 1989.

[10] S. Ross. *Stochastic Processes.* Wiley & Sons, 1996.

[11] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1974.

[12] R. Sedgewick and Ph. Flajolet. *Analysis of Algorithms.* Addison-Wesley, 1996.

[13] E. Seneta. *Non-negative Matrices and Markov chains.* Springer-Verlag, second edition, 1981.

[14] L. Takács. *An introduction to the theory of queues.* Oxford University Press, New York, 1962.

[15] E.C. Titchmarsh. *The Theory of Functions.* Oxford University Press, second edition, 1939.

[16] H.S. Wilf. *Generatingfunctionology.* Academic Press, second edition, 1994.

A Preuve du théorème de Perron-Frobenius

Dans tout ce qui suit, la matrice A est irréductible.

Une remarque est particulièrement utile: la matrice $(I+A)^{n-1}$ est strictement positive. En effet, le graphe $\mathcal{G}[I+A]$ est le graphe $\mathcal{G}[A]$ auquel on a rajouté des arcs «en boucle» $i \to i$. Si A est irréductible, il existe un chemin de i à j dans $\mathcal{G}[A]$. Ce chemin a évidemment au plus $n-1$ arcs. En utilisant autant de ces boucles que nécessaire, il est possible de prolonger ce chemin jusqu'à la longueur $n-1$ dans $\mathcal{G}[I+A]$.

Preuve de a/. On introduit la fonction de Collatz-Wielandt [7]:

$$f_A(x) \;=\; \min_{i \;|\; x_i\neq 0} \frac{(Ax)_i}{x_i}\;,$$

définie sur $(\mathbb{R}^+)^n \setminus \mathbf{0}$. On a tout d'abord:

$$f_A(x) \;=\; \max\{\rho \in \mathbb{R} \mid Ax - \rho x \geq 0\}\;. \tag{67}$$

En effet, si ρ est tel que $Ax \geq \rho x$ alors pour tout i, tel que $x_i \neq 0$ $(Ax)_i/x_i \geq \rho$ (car alors $x_i > 0$) donc $f_A(x) \geq \rho$. Ceci est vrai pour tout ρ tel que $Ax - \rho x \geq 0$ donc $f_A(x) \geq \max\{\rho \in \mathbb{R} | Ax - \rho x \geq 0\}$. Réciproquement, si $r = f_A(x)$, alors $(Ax)_i \geq rx_i$ pour tout i, y compris ceux tels que $x_i = 0$. Donc $f_A(x) \in \{\rho \in \mathbb{R} | Ax - \rho x \geq 0\}$, et par conséquent, $f_A(x) \leq \max\{\rho \in \mathbb{R} | Ax - \rho x \geq 0\}$.

La fonction f_A est *homogène*, c'est à dire que $f_A(\lambda x) = f_A(x)$ pour tout $\lambda > 0$. D'autre part, f_A atteint un maximum sur le simplexe $S_n = \{x \in (\mathbb{R}^+)^n \mid \sum_i x_1 = 1\}$. En effet, sur l'*intérieur* de S_n, f_A est le minimum de fonctions rationnelles donc continues, donc f_A est continue. Ce n'est plus vrai à la frontière quand $x_i \to 0$. Par contre, pour tout vecteur $x \in S_n$, le vecteur $y = (I + A)^{n-1}x$ a les propriétés suivantes: a/ $y > 0$, puisque $(I + A)^{n-1} > 0$ et x a au moins une composante strictement positive et b/ $f_A(y) \geq f_A(x)$: en effet, d'après (67), pour tout x, $Ax \geq f_A(x)x$. En multipliant par $(I + A)^{n-1}$, on a, puisque A et $I + A$ commutent:

$$(I + A)^{n-1}Ax \geq f_A(x)(I + A)^{n-1}x \quad\Longrightarrow\quad Ay \geq f_A(x)\; y \quad\Longrightarrow\quad f_A(y) \geq f_A(x),$$

toujours d'après (67). Soit donc G_n l'image de S_n par $(I + A)^{n-1}$. L'ensemble G_n est un compact contenu dans $]0, +\infty[^n$ et donc f_A y est continue et atteint son maximum, disons en y^0. Soit x^0 la projection de y^0 sur S_n: $x^0 = y^0/\sum y_i^0$. Alors $f_A(x^0) = f_A(y^0)$. Soit $x \in S_n$. Alors $f_A(x) \leq f_A(y) \leq f_A(y^0) \leq f_A(x^0)$. Donc le maximum de f_A sur S_n est atteint en x^0.

Montrons à présent que $r = \max_{x\in S_n} f_A(x)$ est une valeur propre de A. Soit x^0 tel que $r = f_A(x^0)$. Alors $y = Ax^0 - rx^0 \geq 0$. Supposons qu'il n'y ait pas égalité: il existe un i tel que $y_i > 0$. Alors puisque $(I + A)^{n-1}$ est strictement positive, $(I + A)^{n-1}y > 0$. Or, $(I + A)^{n-1}(Ax^0 - rx^0) = A((I + A)^{n-1}x^0) - r((I + A)^{n-1}x^0)$. Donc en vertu de (67) il existe un vecteur $y^0 = (I + A)^{n-1}x^0$ et un $\varepsilon > 0$ tels que $f_A(y^0) = r + \varepsilon$ ce qui est une contradiction. Donc finalement, r est une valeur propre, et x_0 est un vecteur propre associé. On a $r > 0$ car $r \geq f_A(\mathbf{1}) > 0$.

Preuve de b/. Soit x le vecteur propre trouvé en $a/$, tel que $Ax = rx$. Alors $(I + A)^{n-1}x = (1 + r)^{n-1}x$. Mais $(I + A)^{n-1}x > 0$ donc $x > 0$.

Preuve de d/. Soit y tel que $Ay = ry$. Alors comme A est positive, $r|y| = |Ay| \leq A|y|$. Donc d'après (67), on a en fait $r|y| = A|y|$. Donc (voir la preuve du b/): $|y| > 0$. A présent, soient u et v deux vecteurs propres pour la valeur propre r qui sont indépendants. Toute combinaison linéaire $\alpha u + \beta v$ est également vecteur propre. Mais on vient de prouver qu'un vecteur propre pour r non nul n'a pas de coordonnée nulle. Or on peut toujours trouver α et β pour annuler une composante de $\alpha u + \beta v$. Contradiction.

Table des matières

Index