

Master MPRI Paris 6
Fondements de la Modélisation des Réseaux (Module 2.17)
Cours de modélisation Markovienne

Alain Jean-Marie

Version 1.2 du 25 novembre 2010

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 de communication 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.

Ce texte met en valeur certains des aspects les plus en rapport avec :

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

Pour l'année 2010-2011, seule une partie de ce texte a été traitée dans le cours 2.17.

- la section 3 (chaînes de Markov en temps discret)
- la section 4 (chaînes de Markov en temps continu)
- la section 5 (réversibilité)
- la section 6 (files d'attente de base) pour la file $M/M/1$ et la file $M/M/1/K$,
- la section 7 (réseaux de files d'attente), *sans* les réseaux de Kelly.

Les autres parties sont laissées à titre informatif. Ces parties ne seront pas considérées pour l'évaluation.

Sources/Bibliographie Les principales sources de ce cours sont :

- Pour les fonctions génératrices en général : les livres de Sedgewick et Flajolet [10] et Wilf [14]. Pour les résultats en théorie des fonctions complexes, les livres de Titchmarsh [13] et Rudin [9].
- Pour les résultats sur les puissances de matrices, et le théorème de Perron-Frobenius, les livres de Seneta [11] et Minc [6].
- Pour les fondements en probabilités, les livres de Feller, surtout le premier volume [2].
- Pour les Chaînes de Markov en temps et espace discrets, les livres de Kemeny et Snell [4] et Kleinrock [5]. Pour les Chaînes de Markov en temps continu, les livres de Feller [2], Ross [8], ou Kleinrock [5]. Pour la réversibilité, le livre de Kelly [3].
- Pour les modèles de files d'attente, les livres de Takacs [12], Kleinrock [5] et Kelly [3]. Pour les réseaux de ressources, les supports de cours de François Baccelli et les articles de Bonald [1]. Pour les applications aux réseaux, les livres de Pujolle et Fdida [7].

2 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 $\text{sp}(A)$. Le *rayon spectral* de A est défini par :

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

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

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

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

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

2.1 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 $\mathcal{W} : \mathcal{E} \rightarrow \mathbb{C} \setminus \{0\}$.

À tout graphe valué on associe une matrice carrée réelle $A[\mathcal{G}]$, en numérotant les sommets de $\mathcal{V} : \mathcal{V} = \{v(1), \dots, v(N)\}$ 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 $\{1, 2, \dots, N\}$,
- 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.

Irréductibilité 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 $d(i)$ 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*.

2.2 Matrices et Chemins

Un chemin d'un graphe \mathcal{G} est une suite de sommets de ce graphe $\gamma = (i_1, \dots, i_n)$ telle que $(i_j, i_{j+1}) \in \mathcal{E}$ pour $j \in \{1, \dots, n-1\}$. 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 :

$$\gamma = (i_1, \dots, i_n) \implies w(\gamma) = \prod_{j=1}^{n-1} \mathcal{W}(i_j, i_{j+1}). \quad (2)$$

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 :

$$(A^n)_{i,j} = \sum_{\gamma \in \Gamma^n(i,j)} w(\gamma) = \sum_{(i_1, \dots, i_n) \in \Gamma^n(i,j)} \mathcal{W}(i_1, i_2) \mathcal{W}(i_2, i_3) \dots \mathcal{W}(i_{n-1}, i_n). \quad (3)$$

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, $A_{i,j}^0 = 0$; 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 (2), 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 :

$$(A(z)^n)_{i,j} = \sum_{\gamma \in \Gamma^n(i,j)} z^{d(\gamma)} = \sum_d z^d \times \#\{\text{chemins de longueur } n \text{ et de distance } d \text{ de } i \text{ à } j\}. \quad (4)$$

On obtient donc les séries génératrices de la distance des chemins allant de i à j ayant la longueur n .

2.3 Puissances de matrices : exemples

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 (3) est alors :

$$(A^n)_{i,j} = \#\Gamma^n(i, j) = \text{nombre de chemins de longueur } n \text{ de } i \text{ à } j. \quad (5)$$

Si on choisit la valuation :

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

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

$$(A^n)_{i,j} = \sum_{k=0}^n \#\Gamma^k(i, j) = \text{nombre de chemins de longueur } \leq n \text{ de } i \text{ à } j. \quad (6)$$

Réurrences linéaires homogènes

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

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \dots + a_p x_{n-p} \quad n \geq p \\ x_0, x_1, \dots, x_{p-1} \text{ donnés.}$$

Si on définit

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

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

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

2.4 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 1. *Il existe une constante c telle que, pour tout n ,*

$$\mu(A^n) \leq c(N\mu(A))^n . \quad (7)$$

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

La borne (7) 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 2 (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 :*

$$A^n = \sum_{i=1}^p B_i(n, \lambda_i) . \quad (8)$$

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 $v_i \cdot w_i = 1$ et $B_i = w_i \cdot v_i$.

Démonstration. 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 :

$$A = S \begin{pmatrix} J_{\nu_1}(\mu_1) & & & \\ & J_{\nu_2}(\mu_2) & & \\ & & \ddots & \\ & & & J_{\nu_k}(\mu_k) \end{pmatrix} S^{-1} \quad \text{et donc :} \quad A^n = S \begin{pmatrix} J_{\nu_1}^n & & & \\ & J_{\nu_2}^n & & \\ & & \ddots & \\ & & & J_{\nu_k}^n \end{pmatrix} S^{-1} . \quad (9)$$

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

$$A(z) = \sum_{n=0}^{\infty} A^n z^n = \left(\sum_{n=0}^{\infty} A_{i,j} z^n \right)_{(i,j)}. \quad (10)$$

D'après les résultats rappelés dans l'Appendice A, et grâce au Lemme 1, 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 :

$$A(z) = (\mathbf{I} - zA)^{-1} = \frac{1}{\det(\mathbf{I} - zA)} \text{adj}(\mathbf{I} - zA). \quad (11)$$

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

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

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 (11), 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. \square

En ce qui concerne le taux de croissance des coefficients de A^n , on déduit du Théorème 2 :

Corollaire 3. *Soit A une matrice carrée. Pour tout $\varepsilon > 0$, on a :*

$$\mu(A^n) = \mathcal{O}((\rho(A) + \varepsilon)^n). \quad (12)$$

Démonstration. Étant donné que pour tout $p(n)$, polynôme de n , tout $\lambda > 0$ et $\varepsilon > 0$, on a :

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

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

Autant le Corollaire 3 semble plus grossier que le Théorème 2, 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 4. *Soit A une matrice. Son rayon spectral $\rho(A)$ a la propriété :*

$$\min_i \sum_j |A_{i,j}| \leq \rho(A) \leq \max_i \sum_j |A_{i,j}|. \quad (13)$$

Démonstration. 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 (13). Pour l'inégalité de gauche, on peut utiliser la fonction f_A de Collatz-Wielandt (voir l'Appendice B) et sa relation avec $\rho(A)$ (Équation (77)). On obtient :

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

\square

Parmi les applications du Corollaire 3, on trouve :

Corollaire 5. *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 $A(z) = (\mathbf{I} - zA)^{-1}$.*

2.5 Matrices positives

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

Théorème 6 (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 période d , alors les valeurs propres de A de module r sont exactement les $r\omega^j$, $0 \leq j < d$ 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 B.

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

Théorème 7. *Soit A une matrice carrée réelle positive, irréductible. Alors*

$$\mu(A^n) = \mathcal{O}(\rho(A)^n) . \quad (14)$$

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 8. *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 aperiodique, alors il existe un vecteur ligne v et un vecteur colonne w strictement positifs, tels que $v.w = 1$ et $N_{i,j} \sim w_i.v_j \rho(A)^n$.

Considérons le nombre total de chemins partant du sommet i : $D_i = \sum_j N_{i,j}$. On a : $D_i \sim w_i(\sum_j v_j)\rho(A)^n$. En particulier, pour deux sommets $k \neq \ell$, on a $D_k/D_\ell = w_k/w_\ell$. Les nombres w_i s'interprètent donc comme des quantités relatives de chemins, partant des différents sommets i . Dans d'autres applications, on parlera de « poids relatif » du sommet i . Cette information est donc « contenue » dans le vecteur propre gauche de la matrice A , pour la valeur propre de Perron-Frobenius. De façon similaire, on interprète le vecteur propre droit v comme les quantités relatives de chemins arrivant aux différents sommets.

3 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.

3.1 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 $(j_0, j_1, \dots, j_t, j_{t+1})$ d'éléments de \mathcal{E} :

$$\mathbb{P}(X(t+1) = j_{t+1} | X(t) = j_t, \dots, X(0) = j_0) = \mathbb{P}(X(t+1) = j_{t+1} | X(t) = j_t) ; \quad (15)$$

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} ,

$$\mathbb{P}(X(t+1) = j | X(t) = i) = P_{i,j} , \quad (16)$$

indépendamment de t .

À partir de ce point et jusqu'au paragraphe 3.3, on supposera que l'espace d'états \mathcal{E} est *fini*.

3.2 Chaînes de Markov à espace d'état fini

3.2.1 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 »,

$$\mathbf{P} \mathbf{1} = \mathbf{1} . \quad (17)$$

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.

3.2.2 Chaînes de Markov et graphes.

Comme à toute matrice carrée (voir paragraphe 2.1), 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 :

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

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

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 2.1 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*.

3.2.3 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

$$p(i, j; n) = \mathbb{P}(X(n) = j \mid X(0) = i), \quad (18)$$

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 9. Pour tout $n \in \mathbb{N}$:

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

Au vu de (3), on a :

Corollaire 10. Pour tout $n \in \mathbb{N}$, $(i, j) \in \mathcal{E} \times \mathcal{E}$, on a :

$$p(i, j; n) = \sum_{(i_1, \dots, i_{n-1}) \in \mathcal{E}^{n-1}} p_{i, i_1} p_{i_1, i_2} \cdots p_{i_{n-1}, j} . \quad (19)$$

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 à ce chemin le poids : $p_{i_0, i_1} \cdots p_{i_{n-1}, i_n}$. 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 :

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

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 :

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \mathbf{P}^n . \quad (20)$$

3.2.4 Spectre des matrices stochastiques ; Résolvante

Établissons d'abord les résultats spectraux propres aux matrices stochastiques :

Théorème 11. 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 :

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P} . \quad (21)$$

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 $|\lambda| < 1$.

Démonstration. Si \mathbf{P} est irréductible, le Théorème de Perron-Frobenius s'applique et il existe une valeur propre principale. D'après le Lemme 4, $\rho(\mathbf{P}) \leq 1$ puisque la somme des lignes de \mathbf{P} vaut 1. D'autre part, 1 est bien valeur propre puisque $\mathbf{P}\mathbf{1} = \mathbf{1}$. 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 (21) telle que $\pi \mathbf{1} = 1$. Ceci prouve ii/.

Si \mathbf{P} est réductible, alors \mathbf{P} est équivalente à une matrice *blocs* (voir paragraphe 3.2.6) :

$$\begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & & \\ \mathbf{0} & \mathbf{P}_{22} & \ddots & & \\ & \mathbf{0} & \ddots & & \\ & & & \mathbf{0} & \mathbf{P}_{pp} \end{pmatrix}. \quad (22)$$

Le spectre de \mathbf{P} est la réunion de ceux des matrices *irréductibles* de la diagonale : $\mathbf{P}_{11}, \dots, \mathbf{P}_{pp}$. Ces matrices sont soit stochastiques (\mathbf{P}_{pp} l'est obligatoirement), soit *sous-stochastiques*, c'est-à-dire telle que $\mathbf{P}\mathbf{1} \leq \mathbf{1}$. 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 5 et puisque $\rho(\mathbf{P}) = 1$, on sait que la série génératrice des puissances de \mathbf{P} :

$$P(z) = \sum_{n=0}^{\infty} \mathbf{P}^n z^n \quad (23)$$

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 (Théorème 40), 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 13 que le résidu de $P(z)$ en $z = 1$ est la matrice $\mathbf{1}\pi$.

3.2.5 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 $\pi_0 \mathbf{P}^n$ quand $n \rightarrow \infty$. On est donc en présence du problème étudié dans le Paragraphe 2.

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

Théorème 12. *Soit \mathbf{P} une matrice stochastique irréductible et π son vecteur de probabilités stationnaires. Alors*

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

et donc en particulier :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{P}^m = \mathbf{1}\pi. \quad (24)$$

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} \pi_m = \pi. \quad (25)$$

Démonstration. 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 (23). D'après la preuve « analytique » du Théorème 2 (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 (24). Comme l'application $x \mapsto xA$ est continue en A , on en déduit que :

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

Ceci prouve (25). □

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

Théorème 13. *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 $\boldsymbol{\pi} \cdot \mathbf{1} = 1$. Alors*

$$\lim_{n \rightarrow \infty} \mathbf{P}^n = \mathbf{1} \cdot \boldsymbol{\pi}. \quad (26)$$

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

$$\lim_{n \rightarrow \infty} \boldsymbol{\pi}_n = \boldsymbol{\pi}. \quad (27)$$

La convergence est géométrique :

$$\|\boldsymbol{\pi}_n - \boldsymbol{\pi}\| = \mathcal{O}(\rho^n) \quad (28)$$

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.

Démonstration. D'après le Théorème 11.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 $\lambda_1 = 1$, $B_1 = \mathbf{1} \cdot \boldsymbol{\pi}$. Pour les autres valeurs propres, il est facile de voir que pour tout bloc de Jordan $J_i(\lambda)$ $i > 1$,

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

Ceci prouve (26). Comme précédemment, on en déduit par continuité que :

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

Ceci prouve (27). Pour montrer (28), on applique l'argument de la preuve du Corollaire 3 au rapport J_i^n / ρ^n . □

Une chaîne de Markov dont la matrice satisfait aux conditions du Théorème 13 est dite *primitive*. Dans ce cas, et pour un espace d'états fini, la chaîne de Markov satisfait la propriété stochastique d'*ergodicité*. Pour cette raison, et par abus de langage, on dit aussi que la matrice est . Cette dénomination est utilisée dans le reste de ce document, dans le cas fini.

On en conclut donc l'existence d'une distribution limite pour la chaîne de Markov (dans le cas ergodique). Cette distribution π est qualifiée de *stationnaire*. En effet, si la distribution *initiale* π_0 vaut π , alors pour tout n ,

$$\pi_n = \pi_0 \mathbf{P}^n = \pi \mathbf{P}^n = \pi .$$

3.2.6 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 :

$$\mathbb{E}V_n^{(i)} = \sum_{m=0}^{n-1} \mathbb{P}(X_m = i) = \sum_{m=0}^{n-1} \pi_0 \mathbf{P}^m \mathbf{e}_i ,$$

où \mathbf{e}_i est le vecteur colonne $(0, \dots, 1, \dots, 0)'$, 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 3.2.4. L'élément S_{ii} est donc le nombre moyen total de passages en i pour une chaîne partant de i .

Théorème 14. *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.

Démonstration. Pour la partie *i/*, partitionnons 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 (22) et d'appliquer le principe sur les blocs diagonaux. Finalement, par le Corollaire 5, $\sum_n \mathbf{Q}^n$ converge.

Pour *ii/*, voir la preuve du Théorème 12. □

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

3.3 Chaînes de Markov à espace d'état dénombrable

Dans le cas où l'espace d'états est dénombrable, les définitions (15) et (16) 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 (11). 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 : irréductibilité, (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 $(\mathbf{P}^n)_{i,i} \rightarrow 0$
- récurrent non nul (ou simplement : récurrent) si $S_{i,i} = \infty$ et $(\mathbf{P}^n)_{i,i} \not\rightarrow 0$.

On énonce le « théorème de solidarité » des états.

Théorème 15 (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 16. *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 $\boldsymbol{\pi}_n \rightarrow \boldsymbol{\pi}$ 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 $\boldsymbol{\pi}_n \rightarrow \mathbf{0}$ 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 [2, 8]). Les arguments purement algébriques et analytiques utilisés dans le cas où \mathcal{E} est fini ne sont plus applicables en général.

3.4 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 où 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 considérerons deux cas : le cas fini où il existe un maximum K , et le cas infini.

Dans le cas fini où 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 $\mathcal{E} = \{0, \dots, K\}$ ($K \leq +\infty$). Si K est fini, la matrice de transition est

$$\mathbf{P} = \begin{pmatrix} p & q & 0 & & \\ p & 0 & q & \ddots & \\ 0 & \ddots & \ddots & \ddots & \\ & & p & 0 & q \\ & & & p & q \end{pmatrix}. \quad (29)$$

Si K est infini, la matrice (29) est infinie avec la même structure et sans la dernière ligne.

Nous allons nous intéresser à diverses propriétés de ce processus.

3.4.1 Probabilités stationnaires, cas fini

La matrice \mathbf{P} est clairement irréductible et apériodique, donc ergodique. Le Théorème 13 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 :

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

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 :

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

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 dernière 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 :

$$\pi(z) = \frac{p\pi_0 - qz^{K+1}\pi_K}{p - qz}. \quad (30)$$

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 (30), le numérateur doit s'annuler. Ceci donne la condition :

$$p\pi_0 = q\left(\frac{p}{q}\right)^{K+1}\pi_K \iff \pi_K = \left(\frac{q}{p}\right)^K \pi_0.$$

On reporte dans (30) ce qui donne :

$$\pi(z) = \pi_0 \frac{p - p(qz/p)^{K+1}}{p - qz} = \pi_0 \frac{1 - (qz/p)^{K+1}}{1 - (qz/p)}. \quad (31)$$

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

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

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

$$\pi(z) = \frac{1 - q/p}{1 - (q/p)^{K+1}} \frac{1 - (qz/p)^{K+1}}{1 - (q/p)z} = \frac{1 - q/p}{1 - (q/p)^{K+1}} \sum_{k=0}^K \left(\frac{qz}{p}\right)^k. \quad (32)$$

On en déduit facilement les probabilités π_k :

$$\pi_k = [z^k]\pi(z) = \left(\frac{q}{p}\right)^k \frac{1 - q/p}{1 - (q/p)^{K+1}}. \quad (33)$$

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 (30). Dans le paragraphe suivant, on applique de nouveau cette technique dans des conditions moins triviales.

3.4.2 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 $n \geq 1$,

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

Le vecteur π_0 est supposé donné. On passe alors à la fonction génératrice (bidimensionnelle) :

$$\begin{aligned}
\mathcal{F}(x, z) &= \sum_{n=0}^{\infty} \sum_{k=0}^K z^n x^k \pi_{n,k} \\
&= \sum_{k=0}^K x^k \pi_{0,k} + \sum_{n=1}^{\infty} z^n \left(p\pi_{n-1,0} + p\pi_{n-1,1} + \sum_{k=1}^{K-1} x^k (q\pi_{n-1,k-1} + p\pi_{n-1,k+1}) \right. \\
&\quad \left. + x^K (q\pi_{n-1,K-1} + p\pi_{n-1,K}) \right) \\
&= \mathcal{F}(x, 0) + \sum_{n=0}^{\infty} z^{n+1} \left(p \sum_{k=0}^{K-1} x^k \pi_{n,k+1} + q \sum_{k=1}^K x^k \pi_{n,k-1} + p\pi_{n-1,0} + qx^K \pi_{n,K} \right) \\
&= \mathcal{F}(x, 0) + z \sum_{n=0}^{\infty} z^n \left(p \sum_{k=1}^K x^{k-1} \pi_{n,k} + q \sum_{k=0}^{K-1} x^{k+1} \pi_{n,k} + p\pi_{n-1,0} + qx^K \pi_{n,K} \right) \\
&= \mathcal{F}(x, 0) + \sum_{n=0}^{\infty} z^n \left(\frac{p}{x} \left(\sum_{k=0}^K x^k \pi_{n,k} - \pi_{n,0} \right) + qx \left(\sum_{k=0}^K x^k \pi_{n,k} - x^K \pi_{n,K} \right) \right. \\
&\quad \left. + p\pi_{n-1,0} + qx^K \pi_{n,K} \right) \\
&= \mathcal{F}(x, 0) + z \left(\frac{p}{x} + qx \right) \mathcal{F}(x, z) + z \left(1 - \frac{1}{x} \right) (p\mathcal{F}(0, z) - qx^{K+1} \mathcal{F}_K(z)) , \tag{34}
\end{aligned}$$

où on a noté :

$$\mathcal{F}_K(z) = \sum_{n=0}^{\infty} z^n \pi_{n,K} . \tag{35}$$

On déduit de (34) que la fonction \mathcal{F} vérifie l'équation fonctionnelle suivante :

$$(x - pz - qzx^2) \mathcal{F}(x, z) = x\mathcal{F}(x, 0) + z(x - 1)(p\mathcal{F}(0, z) - qx^{K+1} \mathcal{F}_K(z)) . \tag{36}$$

Dans cette équation, la fonction $\mathcal{F}(x, 0)$ est connue car donnée par π_0 , mais $\mathcal{F}(0, z)$ et $\mathcal{F}_K(z)$ sont des *fonctions inconnues*. Dans le membre droit, le polynôme :

$$K(x, z) = -qzx^2 + x - pz \tag{37}$$

joue un rôle déterminant. Ce polynôme est appelé 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'annule. Ceci donne une condition liant les fonctions inconnues $\mathcal{F}(0, z)$ et $\mathcal{F}_K(z)$. Avec assez de conditions, on va peut-être réussir à les déterminer.

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

$$x_1(z) = \frac{1}{2qz} \left(1 - \sqrt{1 - 4pqz^2} \right), \quad x_2(z) = \frac{1}{2qz} \left(1 + \sqrt{1 - 4pqz^2} \right), \tag{38}$$

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 (36), on obtient :

$$\begin{cases} 0 &= x_1(z)\mathcal{F}(x_1(z), 0) + z(x_1(z) - 1)(p\mathcal{F}(0, z) - qx_1(z)^{K+1} \mathcal{F}_K(z)) \\ 0 &= x_2(z)\mathcal{F}(x_2(z), 0) + z(x_2(z) - 1)(p\mathcal{F}(0, z) - qx_2(z)^{K+1} \mathcal{F}_K(z)) . \end{cases} \tag{39}$$

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 (36). Pour simplifier 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 (37)), on obtient :

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

3.4.3 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 16. Pour cela, on doit tenter de résoudre le système $\pi\mathbf{P} = \pi$. Celui-ci s'écrit :

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

Cette récurrence se résout de multiples façons, par exemple en utilisant la méthode analytique du paragraphe 3.4.1, et ses résultats (Équation (30)) en passant à la limite $K \rightarrow \infty$; ou encore en utilisant la construction de 2.3 (exercice instructif); ou simplement par récurrence en partant de la première équation.

Dans tous les cas, on obtient que *nécessairement*, la solution π 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 :

$$\pi_k = \left(\frac{q}{p} \right)^k \left(1 - \frac{q}{p} \right). \quad (40)$$

Il s'agit d'une *distribution géométrique*. D'après le Théorème 16, la distribution π_t converge vers elle pour tout π_0 . La chaîne est dans ce cas récurrente.

– $p \leq q$; alors il n'existe pas de solution sommable au système $\pi\mathbf{P} = \pi$. Donc π_t converge vers $\mathbf{0}$ pour tout π_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 3.4.2) révèle que si $p > q$, on a transience, alors que si $p = q$, on a récurrence nulle.

4 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.

4.1 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 :

$$F_X(x) = \mathbb{P}(X \leq x) = 1 - e^{-\lambda x} \quad (41)$$

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 $s, t > 0$,

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t) . \quad (42)$$

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

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(X \in [t, t + h[\mid X > t) = \lambda . \quad (43)$$

- 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 (au plus dénombrable) de variables.

Soit une suite aléatoire $T_0 \leq T_1 \leq \dots \leq T_n \leq T_{n+1} \leq \dots$. 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 $\{T_0, T_1 - T_0, \dots, T_{n+1} - T_n, \dots\}$ 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 $t_0 < t_1, < \dots < t_n < t_{n+1}$ et tout $n + 2$ -uplet $(j_0, j_1, \dots, j_n, j_{n+1})$ d'éléments de \mathcal{E} :

$$\mathbb{P}(X(t_{n+1}) = j_{n+1} \mid X(t_n) = j_n, \dots, X(t_0) = j_0) = \mathbb{P}(X(t_{n+1}) = j_{n+1} \mid X(t_n) = j_n) ; \quad (44)$$

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

$$\mathbb{P}(X(t+u) = j \mid X(s+u) = i) = \mathbb{P}(X(t) = j \mid X(s) = i) = P_{t-s}(i, j) , \quad (45)$$

indépendamment de t .

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

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

Notons que les chaînes de Markov sont la plupart du temps homogène, ce qui fait qu'on omet d'habitude de le préciser. Dans la suite de ce document, on ne rappellera le caractère homogène des chaînes de Markov que lorsque ce sera indispensable.

4.2 Équations de Chapman-Kolmogorov

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 :

$$P_{t+s}(i, j) = \sum_{k \in \mathcal{E}} P_t(i, k) P_s(k, j), \quad (46)$$

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-Kolmogorov.

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

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

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

$$\mathbf{Q} = \lim_{t \rightarrow 0} \frac{\mathbf{P}_t - \mathbf{I}}{t}. \quad (47)$$

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 : $\mathbf{Q}\mathbf{1} = \mathbf{0}$. 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 (46) se résout 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 $e^{t\mathbf{Q}}$.

4.3 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 17 (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 :*

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

La preuve de ce théorème sort du cadre de ce cours. C'est une conséquence des propriétés (42) et (43) 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 18 (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 :*

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

La preuve de ce théorème peut se déduire facilement du Théorème 17 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 .

4.4 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} :

$$\mathbf{P} = \mathbf{I} + \frac{1}{\nu} \mathbf{Q}. \tag{48}$$

On vérifie facilement que \mathbf{P} est une matrice positive telle que $\mathbf{P}\mathbf{1} = \mathbf{1}$, c'est-à-dire une matrice stochastique. C'est donc la matrice de transition d'une chaîne de Markov en temps discret, mettons $\{Y(n), n \in \mathbb{N}\}$. On vérifie également algébriquement que les distributions stationnaires des matrices \mathbf{P} et \mathbf{Q} sont les mêmes : puisque $\pi\mathbf{P} = \pi + \pi\mathbf{Q}/\nu$, la distribution π est stationnaire pour \mathbf{Q} si et seulement si elle l'est pour \mathbf{P} .

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

Théorème 19. *Soit un processus de Poisson de paramètre $\nu : \{T_n, n \in \mathbb{N}\}$. Alors le processus $\{X(t), t \in \mathbb{R}^+\}$ défini par :*

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

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 , $P_{ii} \neq 0$.

Démonstration. Calculons $P(t)_{i,j}$ pour le processus $X(t)$. Conditionnons sur le nombre d'arrivées du pro-

cessus de Poisson $\{T_n\}$ dans l'intervalle $[0, t[$. On obtient :

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

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 (48) 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 :

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

Le spectre de \mathbf{Q} est donc situé dans le demi-plan complexe $\{\Re(z) \leq 0\}$, 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 $\mathbf{Q}\mathbf{1} = \mathbf{0}$.

Enfin, du point de vue numérique, la preuve du Théorème 19 a fait apparaître la formule :

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

(valide pour tout $\nu \geq \max_i |Q(i, i)|$), qui est utile pour calculer numériquement les probabilités de transition de la chaîne de Markov.

4.5 Comportement asymptotique

Intéressons-nous aux probabilités stationnaires et à la convergence des probabilités π_t .

Dans le cas où \mathcal{E} est fini, on a un théorème similaire au Théorème 13.

Théorème 20. *Soit \mathbf{Q} un générateur infinitésimal irréductible. Soit π le vecteur propre gauche de \mathbf{Q} pour la valeur propre 0, tel que $\pi \cdot \mathbf{1} = 1$. Alors*

$$\lim_{t \rightarrow \infty} e^{\mathbf{Q}t} = \mathbf{1} \cdot \pi. \quad (49)$$

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

$$\lim_{t \rightarrow \infty} \pi_t = \pi. \quad (50)$$

La convergence est géométrique :

$$\|\pi_t - \pi\| = \mathcal{O}(e^{\phi t}) \quad (51)$$

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 16 :

Théorème 21. *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 π .

Alors $\pi_t \rightarrow \pi$ pour toute distribution initiale π_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 $\pi_t \rightarrow \mathbf{0}$ pour toute distribution initiale π_0 .

4.6 Chaîne incluse aux moment de saut

Soient $T_1, T_2, \dots, T_n, \dots$ les dates auxquelles se produisent une transition de la CMTD $\{X(t), t \geq 0\}$. Notons $X_n = \lim_{\varepsilon \rightarrow 0} X(T_n + \varepsilon)$ la valeur de l'état de la chaîne juste après le n -ième saut, c'est-à-dire entre le n -ième saut et le $n + 1$ -ième. Prenons aussi $X_0 = X(0)$.

La suite $\{X_n, n \in \mathbb{N}\}$ est une CMTD. Sa matrice de transition se calcule à partir du générateur \mathbf{Q} par (voir la Construction #1) :

$$\begin{aligned} p_{i,j} &= -\frac{q_{i,j}}{q_{i,i}}, & i \neq j \\ p_{i,i} &= 0. \end{aligned}$$

Cette chaîne de Markov en temps discret est irréductible si et seulement si la chaîne $\{X(t)\}$ l'est. Par contre, elle n'est pas nécessairement apériodique. Dans le cas où l'espace d'état est fini, le Théorème 11 garantit que \mathbf{P} admet une distribution stationnaire unique, que nous notons $\tilde{\pi}$. *Contrairement* au cas de l'uniformisation (paragraphe 4.4), il n'y a pas de raison pour que π et $\tilde{\pi}$ soient la même distribution. La relation entre les deux est donnée par :

$$\pi(i) = \frac{\tilde{\pi}(i)/q_{i,i}}{\sum_j \tilde{\pi}(j)/q_{j,j}}.$$

5 Distribution stationnaire d'une Chaîne de Markov et réversibilité

Dans cette partie, nous nous intéressons à la distribution stationnaire d'une chaîne de Markov et à son calcul sous forme explicite. Nous revenons d'abord sur les équations d'équilibre qui la caractérise, puis nous nous intéressons au cas des chaînes de Markov *réversibles*, pour lesquels le calcul est particulièrement simple. Nous concluons sur des propriétés liées à la réversibilité, qui sont également susceptibles d'accélérer le calcul de la distribution stationnaire.

Nous nous restreignons ici aux CMTD. Des équivalents existent pour les CMTD. Nous renvoyons au livre de Kelly [3] d'où ces résultats sont tirés.

5.1 Équations de balance

Considérons une CMTD irréductible sur un espace d'états dénombrable \mathcal{E} . D'après le Théorème 21, s'il existe une distribution stationnaire, elle est unique. Nous la noterons π par la suite.

Le vecteur π est solution du système d'équations linéaires :

$$\pi \mathbf{Q} = \mathbf{0}, \quad \pi \mathbf{1} = 1.$$

Si l'espace d'états \mathcal{E} est fini, c'est un problème d'algèbre linéaire classique. Il existe bien sûr des algorithmes pour le résoudre en machine, permettant ainsi une *solution numérique* du problème. Nous nous intéressons ici à une autre voie, la résolution *analytique*, avec comme objectif l'obtention de formules permettant de calculer π explicitement. À défaut, la compréhension de la structure des équations peut permettre de les résoudre plus rapidement qu'avec un algorithme de résolution standard. Cette compréhension est également indispensable dans le cas où l'espace d'états n'est pas fini.

L'équation matricielle $\pi \mathbf{Q} = \mathbf{0}$ est un système d'équations qui ont toutes la forme :

$$\sum_i \pi_i Q(i, j) = 0, \quad j \in \mathcal{E},$$

ou encore, en remplaçant $Q(j, j)$ par sa valeur,

$$\sum_{i \neq j} \pi_i Q(i, j) = \sum_{i \neq j} \pi_j Q(j, i). \quad (52)$$

Cette équation représente l'équilibre du « flux de probabilités » entrant dans l'état j (membre gauche) et sortant de l'état j (membre droit) dans l'état stationnaire du processus. On l'appelle *équation d'équilibre* (ou *équation de balance*, en reprenant la dénomination anglophone) pour l'état j . Nous l'appellerons « équation de balance locale » afin de la distinguer plus tard d'autres types d'équations. On peut démontrer que ce flux de probabilité est le nombre moyen de transitions entrant (ou sortant) de l'état j par unité de temps, en régime stationnaire.

Nous avons la propriété suivante.

Théorème 22. *Soit \mathcal{S} un sous-ensemble de \mathcal{E} , $\bar{\mathcal{S}}$ son complément, et π une distribution stationnaire d'un générateur infinitésimal \mathbf{Q} . Alors :*

$$\sum_{i \in \mathcal{S}} \sum_{j \in \bar{\mathcal{S}}} \pi_i Q(i, j) = \sum_{i \in \bar{\mathcal{S}}} \sum_{j \in \mathcal{S}} \pi_j Q(j, i). \quad (53)$$

Démonstration. En sommant l'Équation (52) pour j parcourant $\bar{\mathcal{S}}$, on obtient :

$$\sum_{j \in \bar{\mathcal{S}}} \sum_{i \in \mathcal{E}, i \neq j} \pi_i Q(i, j) = \sum_{j \in \bar{\mathcal{S}}} \sum_{i \in \mathcal{E}, i \neq j} \pi_j Q(j, i).$$

D'autre part, nous avons la relation

$$\sum_{j \in \bar{\mathcal{S}}} \sum_{i \in \bar{\mathcal{S}}, i \neq j} \pi_i Q(i, j) = \sum_{j \in \bar{\mathcal{S}}} \sum_{i \in \bar{\mathcal{S}}, i \neq j} \pi_j Q(j, i),$$

qui est évidente par Fubini. En faisant la différence de ces deux identités, on obtient l'Équation (53). \square

L'Équation (53) traduit l'équilibre des flux de probabilités entrant et sortant de l'ensemble \mathcal{S} . On appelle ce type d'équation une *équation de balance globale* associée à un ensemble d'états.

En utilisant astucieusement les « coupes » \mathcal{S} , il est parfois possible de simplifier énormément le calcul du vecteur π . La simplification maximale est obtenue quand le graphe de la chaîne de Markov a une structure d'arbre. Soit G le graphe non orienté qui supporte le générateur \mathbf{Q} , tel qu'une arête existe entre i et j si $Q(i, j) > 0$ ou $Q(j, i) > 0$. Notons que si \mathbf{Q} est irréductible, alors quand cette arête existe, $Q(i, j) > 0$ et $Q(j, i) > 0$.

Théorème 23. *Si le graphe G est un arbre, alors la distribution stationnaire π satisfait les équations (dites de balance détaillée) suivantes :*

$$\pi_i Q(i, j) = \pi_j Q(j, i) \quad \forall i, j. \quad (54)$$

Démonstration. Soit (i, j) une arête de G . Si on retire cette arête du graphe, cela crée deux composantes connexes. Il suffit de définir \mathcal{S} et $\bar{\mathcal{S}}$ comme ces deux composantes, puis d'appliquer le Théorème 22. \square

Nous verrons plus loin qu'il ne s'agit pas d'une condition nécessaire et suffisante : il existe des chaînes de Markov dont la distribution stationnaire satisfait les équations de balance détaillée, sans que le graphe de la chaîne soit un arbre.

Quand les équations de balance détaillée sont vraies, elles donnent la valeur explicite de la distribution stationnaire :

Théorème 24. *Supposons qu'un ensemble de nombres π_i , positifs et sommant à 1, satisfasse les équations de balance détaillée (54). Alors :*

i/ il s'agit de la distribution stationnaire.

ii/ Soit $i = i_0, i_1, \dots, i_m = j$ un chemin dans le graphe G . Alors :

$$\pi_j = \pi_i \prod_{\ell=0}^{m-1} \frac{Q(i_\ell, i_{\ell+1})}{Q(i_{\ell+1}, i_\ell)}. \quad (55)$$

En particulier, si ce chemin est un cycle ($i = j$), on doit avoir :

$$\prod_{\ell=0}^{m-1} Q(i_{\ell+1}, i_{\ell}) = \prod_{\ell=0}^{m-1} Q(i_{\ell}, i_{\ell+1}). \quad (56)$$

iii/ Réciproquement, si l'Équation (56) est satisfaite pour tout cycle, alors la distribution stationnaire de la chaîne satisfait les équations de balance détaillée.

L'identité (56), qui caractérise les chaînes de Markov qui ont la propriété de balance détaillée, est le *Critère de Kolmogorov*.

Démonstration. *i/.* Supposons qu'un ensemble de nombres π_i satisfaisant (54), existe. Alors en sommant ces équations pour i parcourant $\mathcal{E} \setminus \{j\}$, on obtient, pour chaque j fixé :

$$\sum_{i \neq j} \pi_i Q(i, j) = \sum_{i \neq j} \pi_j Q(j, i)$$

qui sont précisément les équations d'équilibre (52). Donc π est solution de $\pi \mathbf{Q} = \mathbf{0}$, et ses éléments somment à 1. C'est donc la distribution stationnaire.

ii/. L'identité s'obtient en multipliant les équations de balance détaillée

$$\pi_{i_{\ell}} Q(i_{\ell}, i_{\ell+1}) = \pi_{i_{\ell+1}} Q(i_{\ell+1}, i_{\ell})$$

pour ℓ allant de 0 à $m - 1$. En simplifiant les π_i présents dans les deux membres de l'égalité, il reste :

$$\pi_{i_0} \prod_{\ell=0}^{m-1} Q(i_{\ell}, i_{\ell+1}) = \pi_{i_m} \prod_{\ell=0}^{m-1} Q(i_{\ell+1}, i_{\ell}),$$

qui est équivalent à (55) si tous les $Q(i_{\ell+1}, i_{\ell})$ sont non nuls. Si $i_0 = i_m$, on trouve l'identité (56).

iii/. Supposons la propriété (56) vraie. Soit $i = i_0, i_1, \dots, i_m = j$ un chemin du graphe de la chaîne de Markov (au sens de diagramme de transition). On a, pour tout ℓ , $Q(i_{\ell+1}, i_{\ell}) > 0$ et donc $Q(i_{\ell}, i_{\ell+1}) > 0$. Définissons π_j par :

$$\pi_j = B \prod_{\ell=0}^{m-1} \frac{Q(i_{\ell}, i_{\ell+1})}{Q(i_{\ell+1}, i_{\ell})},$$

où B est une constante telle que la somme des π_j soit 1. La valeur de π_j ainsi définie ne dépend pas du chemin choisi, grâce à (56). En effet, soit $i = i'_0, i'_1, \dots, i'_q = j$ un second chemin. En écrivant le critère de Kolmogorov sur le cycle $i_0, i_1, \dots, i_m = i_q, i'_{q-1}, \dots, i'_1, i'_0 = i_0$, nous obtenons :

$$1 = \prod_{\ell=0}^{m-1} \frac{Q(i_{\ell}, i_{\ell+1})}{Q(i_{\ell+1}, i_{\ell})} \prod_{\ell=0}^{q-1} \frac{Q(i'_{\ell+1}, i'_{\ell})}{Q(i'_{\ell}, i'_{\ell+1})} = \frac{\pi_j^{(1)}}{B} \frac{B}{\pi_j^{(2)}}.$$

Les valeurs de π_j définies par les deux chemins sont donc les mêmes. D'autre part, cette famille de nombres vérifie les équations de balance détaillée. En effet, soit k un autre état, et $i = i'_0, i'_1, \dots, i'_q = k$ un chemin. Alors, selon le même principe :

$$\frac{\pi_j Q(j, k)}{\pi_k Q(k, j)} = \prod_{\ell=0}^{m-1} \frac{Q(i_{\ell}, i_{\ell+1})}{Q(i_{\ell+1}, i_{\ell})} \frac{Q(j, k)}{Q(k, j)} \prod_{\ell=0}^{q-1} \frac{Q(i'_{\ell+1}, i'_{\ell})}{Q(i'_{\ell}, i'_{\ell+1})}.$$

Ce produit vaut 1, en appliquant la propriété (56) au cycle $i = i_0, i_1, \dots, i_m = j, k = i'_q, i'_{q-1}, \dots, i'_0 = i$. \square

5.2 Réversibilité

Dans ce paragraphe, nous introduisons les notions de processus renversé et de réversibilité, et leur relation avec la balance détaillée.

Soit $\{X(t), t \in \mathbb{R}\}$ un processus stochastique. Le *processus renversé* est le processus $\{X(-t), t \in \mathbb{R}\}$. Un processus et son processus renversé ont les mêmes distributions stationnaires.

Théorème 25. *Supposons que le processus $\{X(t)\}$ soit une CMTC homogène de générateur \mathbf{Q} , et de distribution stationnaire π . Supposons que ce processus soit stationnaire. Alors le processus renversé est également une CMTC homogène et stationnaire, dont le générateur \mathbf{Q}' est donné par :*

$$Q'(i, j) = Q(j, i) \frac{\pi_j}{\pi_i}.$$

Démonstration. Soient i, j et k trois états et $t_1 < t_2 < t_3$ trois dates. Soit $X'(t)$ le processus renversé. Considérons la probabilité :

$$\begin{aligned} \mathbb{P}(X'(t_3) = j | X'(t_2) = i, X'(t_1) = k) &= \frac{\mathbb{P}(X'(t_3) = j, X'(t_2) = i, X'(t_1) = k)}{\mathbb{P}(X'(t_2) = i, X'(t_1) = k)} \\ &= \frac{\mathbb{P}(X(-t_3) = j, X(-t_2) = i, X(-t_1) = k)}{\mathbb{P}(X(-t_2) = i, X(-t_1) = k)} \\ &= \frac{\mathbb{P}(X(-t_1) = k | X(-t_3) = j, X(-t_2) = i) \mathbb{P}(X(-t_3) = j, X(-t_2) = i)}{\mathbb{P}(X(-t_1) = k | X(-t_2) = i) \mathbb{P}(X(-t_2) = i)} \\ &= \frac{\mathbb{P}(X(-t_1) = k | X(-t_2) = i) \mathbb{P}(X(-t_2) = i | X(-t_3) = j) \mathbb{P}(X(-t_3) = j)}{\mathbb{P}(X(-t_1) = k | X(-t_2) = i) \mathbb{P}(X(-t_2) = i)} \\ &= \mathbb{P}(X(-t_2) = i | X(-t_3) = j) \frac{\mathbb{P}(X(-t_3) = j)}{\mathbb{P}(X(-t_2) = i)} \\ &= \mathbb{P}(X(-t_2) = i | X(-t_3) = j) \frac{\pi_j}{\pi_i}. \end{aligned}$$

On a utilisé la propriété de Markov (ligne 3-ligne 4) puis la stationnarité pour le processus X (ligne 5-ligne 6). On conclut donc que la probabilité recherchée ne dépend pas de $X'(t_1)$, et donc le processus X' satisfait la propriété de Markov. Comme le processus X est supposé stationnaire, les probabilités de transition $\mathbb{P}(X'(t_3) = j | X'(t_2) = i)$ ne dépendent que de $t_3 - t_2$, c'est-à-dire que la chaîne de Markov est homogène. Enfin, en posant $t_3 = t + h$ et $t_2 = t$, on a :

$$\mathbb{P}(X'(t+h) = j | X'(t) = i) = \mathbb{P}(X(-t) = i | X'(-t-h) = j) \frac{\pi_j}{\pi_i}.$$

En laissant tendre h vers 0, on obtient, par définition du générateur (voir l'Équation (47)) :

$$Q'(i, j) = Q(j, i) \frac{\pi_j}{\pi_i},$$

ce qui est le résultat recherché. □

On notera, en relisant cette preuve, que si le processus X n'est pas stationnaire, alors le processus X' est bien Markovien, mais il n'est pas homogène : ses probabilités de transition dépendent du temps.

Un processus stochastique $\{X(t), t \in \mathbb{R}\}$ est dit *réversible* si pour tout $\tau \in \mathbb{R}$, et pour toute suite t_1, \dots, t_k , le vecteur $(X(t_1), \dots, X(t_k))$ a la même distribution que $(X(\tau - t_1), \dots, X(\tau - t_k))$.

Un processus réversible est nécessairement stationnaire : en prenant $\tau = t_2 + t_1$, on déduit de la définition que $X(t_1)$ et $X(t_2)$ ont la même distribution quels que soient t_1 et t_2 . Le même raisonnement s'applique aux distributions jointes du processus.

Pour les chaînes de Markov, on a la propriété intéressante :

Théorème 26. *Une CMTC stationnaire est réversible si et seulement si sa distribution stationnaire et son générateur vérifient les équations de balance détaillée.*

Démonstration. Si la CMTC est réversible, son processus renversé a la même distribution qu'elle-même, donc le même générateur. D'après le Théorème 25, cela implique que, pour tout i, j :

$$Q(i, j) = Q(j, i) \frac{\pi_j}{\pi_i} .$$

Ceci est équivalent aux équations de balance détaillée (54). \square

On déduit par exemple de cette caractérisation et du Théorème 23 que les chaînes de Markov dont le graphe est en forme d'arbre sont réversibles. Parmi cette famille de processus, on trouve les *processus de naissance et de mort* (déjà évoqués en temps discret au Paragraphe 3.4), dont le graphe est en fait une chaîne.

Enfin, même si une CMTC n'est pas réversible, il peut être utile de considérer son processus renversé. Cela peut permettre de caractériser leur distribution stationnaire (qui est commune aux deux) plus facilement, comme énoncé dans cette propriété.

Théorème 27. *Soit $\{X(t)\}$ une CMTC de générateur \mathbf{Q} . S'il existe une famille de nombres π_i positifs sommant à 1, et une de nombres $Q'(i, j)$ tels que $\sum_{j \neq i} Q'(i, j) = \sum_{j \neq i} Q(i, j)$ pour tout i , qui satisfont :*

$$\pi_i Q(i, j) = \pi_j Q'(i, j) , \quad (57)$$

alors \mathbf{Q}' est le générateur du processus renversé de X , et π est leur distribution stationnaire commune.

Démonstration. En sommant l'équation (57) pour i parcourant l'espace d'états (sauf j), on obtient :

$$\sum_{i \neq j} \pi_i Q(i, j) = \sum_{i \neq j} \pi_j Q'(i, j) = \pi_j \sum_{i \neq j} Q'(i, j) = \pi_j \sum_{i \neq j} Q(i, j) .$$

Le vecteur π satisfait donc les équations d'équilibre, et c'est donc une distribution stationnaire. Par le Théorème 25, le générateur \mathbf{Q}' est donc celui du processus renversé. \square

5.3 Troncation de chaînes de Markov

Une propriété très utile des chaînes de Markov réversibles est d'être « résistantes » à la troncation. En effet, quand on restreint une chaîne de Markov à rester dans un sous-espace d'états en annulant les taux de transition qui en sortent, le nouveau processus est toujours une chaîne de Markov réversible, et sa distribution stationnaire se déduit de celle du processus complet, par renormalisation. Ceci est formalisé ci-dessous.

Soit $\mathcal{A} \subset \mathcal{E}$. Soit $\mathbf{Q}_{\mathcal{A}}$ la matrice obtenue de la façon suivante :

$$\begin{aligned} Q_{\mathcal{A}}(i, j) &= Q(i, j) , \quad \text{si } i \in \mathcal{A} \text{ et } j \in \mathcal{A}, i \neq j , \\ Q_{\mathcal{A}}(i, i) &= - \sum_{j \in \mathcal{A}} Q(i, j) , \quad \forall i \in \mathcal{A} \\ Q_{\mathcal{A}}(i, j) &= 0 \quad \text{sinon} . \end{aligned}$$

Cette matrice est le générateur infinitésimal d'une chaîne de Markov (notée $X_{\mathcal{A}}$ et qualifiée de « tronquée ») évoluant sur l'espace d'états \mathcal{A} . Ses propriétés sont résumées dans le théorème suivant.

Théorème 28. *Supposons que la chaîne $X_{\mathcal{A}}$ soit irréductible. Si une distribution stationnaire π de \mathbf{Q} existe, alors il en existe une aussi pour $\mathbf{Q}_{\mathcal{A}}$, $\pi_{\mathcal{A}}$. Les deux sont reliées par la relation : pour tout $i \in \mathcal{A}$,*

$$\pi_{\mathcal{A}}(i) = \frac{\pi(i)}{\sum_{j \in \mathcal{A}} \pi(j)} . \quad (58)$$

La chaîne $X_{\mathcal{A}}$ est réversible.

Démonstration. L'existence de $\pi_{\mathcal{A}}$ découle du Théorème 21. Comme la famille de probabilités $\pi_{\mathcal{A}}(i)$ définie par la formule (58) satisfait les équations de balance détaillée, cette famille est donc bien la distribution stationnaire de $X_{\mathcal{A}}$, qui est donc un processus réversible, par le Théorème 26 \square

5.4 Composition de chaînes de Markov

Dans certains cas, la modélisation d'un problème conduit à construire des chaînes de Markov « complexes » à partir de chaînes plus élémentaires. Nous étudions dans ce paragraphe une telle construction, qui a la propriété que la chaîne obtenue est réversible quand les chaînes de départ le sont.

Soient \mathcal{E}_1 et \mathcal{E}_2 deux espaces d'état, et soient $\mathbf{Q}^{(1)}$ et $\mathbf{Q}^{(2)}$ deux générateurs correspondant à des Chaînes de Markov sur ces deux espaces. Soient par ailleurs $V^{(1)}$ et $V^{(2)}$ deux matrices positives, dont les dimensions respectives sont celles de \mathcal{E}_1 et \mathcal{E}_2 . Considérons maintenant la chaîne de Markov sur l'espace d'état $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$, dont le générateur \mathbf{Q} est défini comme suit. Les éléments de \mathcal{E} sont des couples (i, j) où $i \in \mathcal{E}_1$ et $j \in \mathcal{E}_2$. Le taux de transition de (i, j) vers (p, q) est donc noté $Q_{(ij)(pq)}$. Supposons que sa valeur soit :

$$Q_{(ij)(pq)} = Q_{ip}^{(1)} V_{jq}^{(2)} + V_{ip}^{(1)} Q_{jq}^{(2)}. \quad (59)$$

On a alors la propriété suivante.

Théorème 29. *Supposons que $\mathbf{Q}^{(1)}$ et $\mathbf{Q}^{(2)}$ soient réversibles avec distributions stationnaires respectives $\pi^{(1)}$ et $\pi^{(2)}$. Supposons d'autre part que les matrices $V^{(1)}$ et $V^{(2)}$ vérifient, pour tout $i, j \in \mathcal{E}_1$ et $p, q \in \mathcal{E}_2$,*

$$\pi^{(1)}(i) V_{ip}^{(1)} = \pi^{(1)}(p) V_{pi}^{(1)} \quad \pi^{(2)}(j) V_{jq}^{(2)} = \pi^{(2)}(q) V_{qj}^{(2)}. \quad (60)$$

Alors le générateur \mathbf{Q} est réversible pour la distribution stationnaire π sur l'espace \mathcal{E} donnée par la formule :

$$\pi(i, j) = \pi^{(1)}(i) \pi^{(2)}(j). \quad (61)$$

Démonstration. En vertu du Théorème 26, il suffit de montrer que le générateur \mathbf{Q} et la distribution donnée en (61) vérifient les équations de balance détaillée. On a en effet a suite d'égalités :

$$\begin{aligned} \pi(i, j) Q_{(ij)(pq)} &= \pi^{(1)}(i) \pi^{(2)}(j) Q_{ip}^{(1)} V_{jq}^{(2)} + \pi^{(1)}(i) \pi^{(2)}(j) V_{ip}^{(1)} Q_{jq}^{(2)} \\ &= \pi^{(1)}(i) Q_{ip}^{(1)} \pi^{(2)}(j) V_{jq}^{(2)} + \pi^{(1)}(i) V_{ip}^{(1)} \pi^{(2)}(j) Q_{jq}^{(2)} \\ &= \pi^{(1)}(p) Q_{pi}^{(1)} \pi^{(2)}(q) V_{qj}^{(2)} + \pi^{(1)}(p) V_{pi}^{(1)} \pi^{(2)}(q) Q_{qj}^{(2)} \\ &= \pi(p, q) Q_{(pq)(ij)}. \end{aligned}$$

Dans le passage de la ligne 2 à la ligne 3, on a utilisé simultanément l'équation de balance détaillée pour les deux générateurs élémentaires, et la propriété (60). \square

On observera que si les matrices $V^{(1)}$ ou $V^{(2)}$ ne satisfont pas la condition (60), alors la chaîne de Markov de générateur \mathbf{Q} n'est plus nécessairement réversible, mais sa distribution stationnaire est toujours celle donnée par l'équation (61) (preuve en exercice).

Cette construction se généralise bien sûr à plus de deux dimensions.

Illustration. La condition algébrique imposée par (60) dans le Théorème 29 semble contraignante, mais elle est satisfaite dans un nombre de cas utiles en pratique.

Supposons par exemple que $V^{(1)}$ et $V^{(2)}$ soient *diagonales*. Dans ce cas, $V_{ip}^{(1)} = 0$ si $i \neq p$ et (60) se réduit au cas $i = p$, pour lequel cette condition est automatiquement vérifiée. De même pour la matrice $V^{(2)}$.

Dans ce cas, l'interprétation de la construction est que chaque chaîne élémentaire influence les taux de transition de l'autre par l'intermédiaire des matrices de *vitesse* V . Par exemple, si la chaîne numéro 1 est dans l'état i , le taux de transition de la chaîne 2 est multiplié par la valeur $v_i^{(1)} = V_{ii}^{(1)}$.

Dans le cas particulier où $V^{(1)} = \mathbf{I}$ (matrice identité), la chaîne 1 n'a pas d'influence sur l'évolution de la chaîne 2. Cette dernière évolue donc librement, et influence la vitesse de la chaîne 1. On parle parfois dans ce cas de « modulation » de la chaîne 1 par la chaîne 2.

Enfin, dans le cas où $V^{(1)} = V^{(2)} = \mathbf{I}$, les chaînes n'ont en fait aucune influence l'une sur l'autre et sont stochastiquement indépendantes.

5.5 Générateurs réversibles et diagonalisation

Nous concluons ce chapitre par l'énoncé d'une propriété algébrique intéressante des générateurs de chaînes réversibles.

Théorème 30. *Soit \mathbf{Q} le générateur infinitésimal d'une chaîne de Markov réversible sur un espace d'états \mathcal{E} fini. Alors \mathbf{Q} est diagonalisable et ses valeurs propres sont réelles et négatives.*

Démonstration. On a vu au paragraphe 4.4 que le spectre d'un générateur infinitésimal (réversible ou pas) était toujours contenu dans le demi-plan complexe $\{\Re(z) \leq 0\}$ et que 0 était toujours valeur propre. Il reste à voir que la matrice \mathbf{Q} est diagonalisable et ses valeurs propres réelles.

Sans perte de généralité, on peut se placer sur une composante fortement connexe du graphe de transition. Restreinte à cette composante, la matrice \mathbf{Q} est irréductible, et admet une unique distribution stationnaire π strictement positive. Définissons la matrice diagonale $\mathbf{W} = \text{diag}(\sqrt{\pi(i)^{-1}})_{i \in \mathcal{E}}$, puis la matrice :

$$\bar{\mathbf{Q}} = \mathbf{W}^{-1} \mathbf{Q} \mathbf{W},$$

dont les éléments valent :

$$\bar{Q}_{ij} = \frac{\sqrt{\pi(i)}}{\sqrt{\pi(j)}} Q_{ij}.$$

Or, l'équation de balance détaillée $\pi(i)Q_{ij} = \pi(j)Q_{ji}$ implique que :

$$\bar{Q}_{ij} = \frac{\pi(i)}{\sqrt{\pi(i)\pi(j)}} Q_{ij} = \frac{\pi(j)}{\sqrt{\pi(i)\pi(j)}} Q_{ji} = \frac{\sqrt{\pi(j)}}{\sqrt{\pi(i)}} Q_{ji} = \bar{Q}_{ji}.$$

La matrice $\bar{\mathbf{Q}}$ est donc symétrique, et il est bien connu qu'elle se diagonalise donc dans une base orthonormale. Il existe une matrice \mathbf{R} (qui peut être prise unitaire) telle que $\mathbf{R}^{-1} \bar{\mathbf{Q}} \mathbf{R} = \mathbf{D}$, avec \mathbf{D} diagonale réelle. Par conséquent, $(\mathbf{WR})^{-1} \mathbf{Q} \mathbf{WR} = \mathbf{D}$, ce qui prouve que \mathbf{Q} est diagonalisable dans \mathbb{R} . \square

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 « $M/M/1$ »

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 ; en particulier, la durée moyenne d'un service vaut $1/\mu$;
- il y a un serveur ;
- les clients sont servis un par un dans l'ordre de leur arrivée ;
- la capacité de la file d'attente est infinie : il y a toujours de la place pour les clients en attente.

Selon la *Notation de Kendall* (voir l'Appendice C), cette file est connue sous l'abréviation $M/M/1$.

Le processus $\{N(t), t \in \mathbb{R}^+\}$ est une chaîne de Markov en temps continu, et son espace d'état est \mathbb{N} . Ceci se déduit de la Construction #2, et de la *propriété sans mémoire* de la loi exponentielle. Celle-ci implique que, à tout instant, quelles que soient les dates d'arrivée du précédent client et celle du début du service en cours, la date d'arrivée du prochain client est après une durée distribuée exponentiellement de paramètre λ , et la prochaine date de fin de service est après une durée distribuée exponentiellement avec un paramètre μ .

Par conséquent, le générateur infinitésimal de la CMTC $\{N(t)\}$ s'écrit :

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

Ce processus est aussi *processus de naissance et de mort*, en temps continu. Le paramètre λ est le taux des naissances : le nombre moyen de naissances par unité de temps. De la même façon le paramètre μ est nommé taux de mort, ou plus couramment pour une file d'attente, le *taux de service*. Il s'agit en effet du nombre moyen de clients servis par seconde quand le serveur travaille en permanence. C'est une grandeur assimilable à une *vitesse* de service.

Le processus $\{N(t)\}$ ressemble beaucoup au processus étudié au Paragraphe 3.4.3. Effectivement, 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 (29) dans sa forme infinie. On peut bien sûr se servir de cette remarque pour calculer la distribution stationnaire de la chaîne, si on a commencé par étudier la marche aléatoire en dimension 1.

Mais disons qu'on ne dispose pas de ces résultats. Il est possible d'appliquer les résultats de la Section 5. Le graphe de la chaîne est en effet un arbre, et le Théorème 23 s'applique. Les équations de balance détaillées sont donc satisfaites, ainsi que la réversibilité (Théorème 26).

En particulier, nous avons (Théorème 24) une expression de chaque probabilité en fonction de π_0 :

$$\pi_n = \pi_0 \prod_{\ell=0}^{n-1} \frac{\lambda}{\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^n.$$

Seule la valeur de π_0 est inconnue. Pour la déterminer, on utilise l'équation de *normalisation* :

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n.$$

La discussion de la convergence de cette série conduit, grâce au Théorème 21, à la conclusion suivante :

Théorème 31. *Dans une file d'attente M/M/1 avec débit d'arrivée λ et durée moyenne de service $1/\mu$:*

– si $\lambda < \mu$, il existe une unique distribution stationnaire π pour le processus $\{N(t)\}$, qui est géométrique de loi :

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

– Si $\lambda \geq \mu$, il n'existe pas de distribution de probabilité stationnaire pour le processus $\{N(t)\}$.

La condition $\lambda < \mu$ est dite *condition de stabilité* de la file M/M/1. Elle s'interprète par la nécessité 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.

Par application du Théorème 26, on conclut que quand il est stationnaire, le processus $\{N(t), t \in \mathbb{R}\}$ est réversible. Comme dans le processus renversé, ce qui était les départs de clients devient les arrivées de clients, on obtient le résultat suivant (Théorème de Burke).

Théorème 32. *Soit une file d'attente M/M/1 stationnaire. Alors le processus de départ des clients est un processus de Poisson de paramètre λ .*

des C clients en service aura lieu après une durée $\sim \text{Exp}(\mu)$. Les différentes dates de fin de service sont indépendantes. La première date de fin de service est donc obtenue comme le minimum de C variables aléatoires indépendantes $\text{Exp}(\mu)$, et c'est donc une v.a. $\sim \text{Exp}(C\mu)$. Si il y a $n < C$ clients présents, seuls n serveurs sont utilisés, et le raisonnement précédent dit que la fin de service aura lieu après une durée $\sim \text{Exp}(n\mu)$.

En résumé, le taux de transition $Q(n, n-1)$ vaut $\min(n, C)\mu$. Le taux de transition $Q(n, n+1)$ vaut λ comme précédemment.

On est donc toujours en présence d'un processus de naissance et de mort, et en appliquant le Théorème 24, *ii/*, on trouve pour la distribution stationnaire :

$$\pi_n = \pi_0 \prod_{\ell=0}^{n-1} \frac{\lambda}{(\ell+1)\mu} = \pi_0 \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n, \quad \text{si } n \leq C,$$

et

$$\pi_n = \pi_C \prod_{\ell=C}^{n-1} \frac{\lambda}{C\mu} = \pi_C \left(\frac{\lambda}{C\mu}\right)^{n-C} = \pi_0 \frac{C^C}{C!} \left(\frac{\lambda}{C\mu}\right)^n \quad \text{si } n \geq C.$$

Pour déterminer π_0 , on utilise l'équation de normalisation, et on doit avoir :

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \left[\sum_{n=0}^{C-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{C^C}{C!} \sum_{n=C}^{\infty} \left(\frac{\lambda}{C\mu}\right)^n \right].$$

Cette série converge si et seulement si $\lambda < C\mu$, ce qui est la condition de stabilité de cette file d'attente. Sous cette condition, le calcul donne pour π_0 :

$$\pi_0 = \left[\sum_{n=0}^{C-1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{C^C}{C!} \left(\frac{\lambda}{C\mu}\right)^C \frac{1}{1 - (\lambda/C\mu)} \right]^{-1}.$$

6.4 La file « $M/M/\infty$ »

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 :

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

De nouveau, cette chaîne est réversible, et la résolution des équations de balance détaillée conduit à conclure que :

$$\pi_n = \pi_0 \prod_{\ell=0}^{n-1} \frac{\lambda}{(\ell+1)\mu} = \pi_0 \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n.$$

En introduisant la notation $\rho = \lambda/\mu$, l'équation de normalisation donne :

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \sum_{n=0}^{\infty} \frac{1}{n!} \rho^n = \pi_0 e^\rho.$$

On en déduit que $\pi_0 = e^{-\rho}$. Autrement dit, la distribution stationnaire de la file $M/M/\infty$ 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 n'y a donc pas de condition de stabilité pour la file $M/M/\infty$.

Il se trouve que cette distribution de Poisson de paramètre ρ est la distribution stationnaire de la file $M/GI/\infty$. 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*.

6.5 Calcul de probabilités stationnaires par récurrences

Dans ce paragraphe, nous illustrons à travers un exemple une technique qui permet de calculer la distribution stationnaire de certaines chaînes de Markov qui ne sont pas réversibles. Quand les transitions ont une structure régulière, la solution peut parfois être calculée par récurrences, et celles-ci résolues de façon explicite.

L'exemple résolu ici est représenté dans le Figure 1.

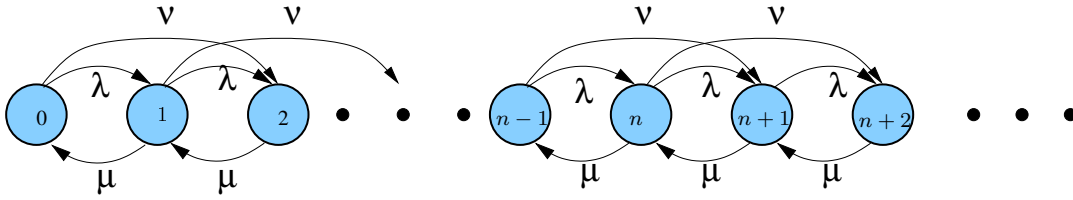


FIG. 1 – Une CMTC avec une structure régulière

Les équations de balance locales sont :

$$(\lambda + \nu)\pi(0) = \mu\pi(1) \quad (62)$$

$$(\lambda + \nu + \mu)\pi(1) = \mu\pi(2) \quad (63)$$

$$(\lambda + \nu + \mu)\pi(n) = \mu\pi(n+1) + \nu\pi(n-2), \quad n \geq 2. \quad (64)$$

Cette dernière équation peut être vue comme une récurrence pour calculer $\pi(n+1)$ en fonction de $\pi(n)$ et $\pi(n-2)$. C'est une récurrence linéaire à coefficient constants, d'ordre 3.

Si on applique les équations de balance globale à l'ensemble $\mathcal{S} = \{0, \dots, n\}$, on obtient à la place : l'équation (62) et

$$(\lambda + \nu)\pi(n) + \nu\pi(n-1) = \mu\pi(n+1), \quad n \geq 1. \quad (65)$$

C'est toujours une récurrence linéaire à coefficients constants, mais elle est d'ordre 2.

Il est bien connu que la solution générale de cette récurrence (65) est de la forme :

$$\pi(n) = A\alpha_1^n + B\alpha_2^n,$$

où α_1 et α_2 sont les racines de l'équation caractéristique

$$\mu\alpha^2 - (\lambda + \nu)\alpha - \nu = 0.$$

La solution du problème du calcul de π est complète si nous pouvons déterminer les constantes A et B . Pour cela, on utilise les deux équations qui ne l'ont pas encore été : l'équation de balance (62) et l'équation de normalisation $\pi \mathbf{1} = 1$. Étant donnée la forme générale pour $\pi(n)$ et si $|\alpha_1| < 1$ et $|\alpha_2| < 1$, on doit avoir :

$$1 = \frac{A}{1 - \alpha_1} + \frac{B}{1 - \alpha_2}.$$

L'équation (62) se réécrit comme :

$$(\lambda + \nu)(A + B) = \mu(A\alpha_1 + B\alpha_2).$$

L'analyse du polynôme caractéristique donne les résultats suivants : une des racines α_1 est toujours dans l'intervalle $] -1, 0[$ si $\nu > 0$, est vaut 0 si $\nu = 0$. L'autre racine α_2 est positive, et $\alpha_2 < 1$ si et seulement si $\mu > \lambda + 2\nu$. Si cette *condition de stabilité* n'est pas satisfaisante, il n'y a pas de distribution stationnaire.

Supposons donc que $\mu > \lambda + 2\nu$. L'expression finale complète pour la solution peut être exprimée comme :

$$\begin{aligned}\pi(n) &= A\alpha_1^n + B\alpha_2^n \\ A &= \frac{\mu - 2\nu - \lambda}{2\mu} \left(1 - \frac{\lambda + \nu}{\Delta}\right) \\ B &= -A \frac{\lambda + \nu - \mu\alpha_1}{\lambda + \nu - \mu\alpha_2} \\ \Delta &= \sqrt{(\lambda + \nu)^2 + 4\mu\nu} \\ \alpha_1 &= \frac{\lambda + \nu - \Delta}{2\mu} \\ \alpha_2 &= \frac{\lambda + \nu + \Delta}{2\mu}.\end{aligned}$$

Cette solution se réduit à celle de la file $M/M/1$ si $\nu = 0$.

6.6 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 du 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 :

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

dans ce système. Alors :

$$\bar{N} = \lambda \bar{T}. \tag{66}$$

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 :

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

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

$$\bar{T} = \frac{1}{\mu},$$

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, on donne 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 temps de réponse d'un paquet dans l'Internet, ou celui du passage aux guichets d'une poste, ne se résolvent 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 Section 6 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 $\vec{\lambda}^0 = (\lambda_1^0, \dots, \lambda_N^0)$ des taux d'arrivées extérieures dans chaque file,
- le vecteur (μ_1, \dots, μ_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 $1, 2, \dots, N$ sont tous *transients*. Ceci revient à dire que tout client a toujours un ou plusieurs chemins qui mènent à l'extérieur, et finit donc fatalement par en prendre un (voir Théorème 14.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 §3.2.5).

Des règles d'évolution énoncées ci-dessus, et en utilisant la Construction #2, on obtient facilement que le processus $N(t) = (N_1(t), \dots, N_N(t))$ 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 : $\mathbf{n} = (n_1, \dots, n_N)$ et $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ pour le i -ième vecteur unitaire, on décrit le générateur \mathbf{Q} par le tableau de transitions :

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

On a alors :

Théorème 34 (de Jackson pour réseaux ouverts). *Soit le vecteur $\vec{\lambda} = (\lambda_1, \dots, \lambda_N)$ solution du système linéaire :*

$$\vec{\lambda} = \vec{\lambda}^0 + \vec{\lambda} \mathbf{R} . \tag{67}$$

Si la condition (de stabilité) :

$$\forall 1 \leq i \leq N, \quad \lambda_i < \mu_i$$

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

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

Démonstration. La preuve du Théorème de Jackson et des théorèmes similaires peut se faire par vérification directe : il suffit juste de vérifier que la distribution proposée satisfait aux équations d'équilibre. Dans le cas

présent, on doit alors avoir pour tout vecteur \mathbf{n} :

$$\begin{aligned} \boldsymbol{\pi}(\mathbf{n}) \left(\sum_{i=0}^N \mu_i \mathbf{1}_{\{n_i > 0\}} + \lambda_i^0 \right) &= \sum_{j=0}^N \sum_{i=0}^N \boldsymbol{\pi}(\mathbf{n} + \mathbf{e}_j - \mathbf{e}_i) \mu_j r_{j,i} \mathbf{1}_{\{n_i > 0\}} \\ &+ \sum_{j=1}^N \boldsymbol{\pi}(\mathbf{n} + \mathbf{e}_j) \mu_j r_{j,0} + \sum_{j=0}^N \boldsymbol{\pi}(\mathbf{n} - \mathbf{e}_j) \lambda_j^0 \mathbf{1}_{\{n_j > 0\}} . \end{aligned} \quad (68)$$

Ce qu'on peut vérifier avec un peu de patience en faisant usage de (67). L'unicité découle du fait que la chaîne est irréductible (justifier!) et du Théorème 21. Une autre méthode de preuve sera utilisée dans le paragraphe suivant pour les réseaux de Kelly. \square

Exercice : justifier le système (67) qui définit $\bar{\lambda}$.

Ces équations portent le nom d'*équations de trafic*. Elles admettent une solution unique sous la condition faite sur \mathbf{R} . Le fait que cette solution soit positive peut être déduit du Corollaire 5. Le vecteur λ ainsi défini est le vecteur des *débîts* des clients entrant dans chacune des files d'attente.

D'après le Théorème 34, la distribution stationnaire de la file i est donnée par :

$$\mathbb{P}(n_i) = \left(1 - \frac{\lambda_i}{\mu_i} \right) \left(\frac{\lambda_i}{\mu_i} \right)^{n_i} .$$

On reconnaît la distribution stationnaire de la file $M/M/1$ avec pour taux d'arrivée λ_i et temps de service moyen μ_i . La probabilité d'un état du réseau s'obtient comme produit des probabilités marginales : on dit que cette distribution a une « forme produit ». Ainsi, dans l'état stationnaire, l'état de chaque file d'attente est celui d'une file $M/M/1$, indépendante du reste du réseau.

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 (μ_1, \dots, μ_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 précédemment, $\{N(t), t \in \mathbb{R}\}$ 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 3.2.2).

Théorème 35 (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 :*

$$\boldsymbol{\theta} = \boldsymbol{\theta} \mathbf{R} . \quad (69)$$

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 :

$$G = \left[\sum_{\sum n_i = K} p(n_1, \dots, n_N) \right]^{-1}. \quad (70)$$

Plusieurs remarques sur ces résultats.

- L'équation (69) a une infinité de solutions. Cependant, comme l'ensemble des solutions forme une droite vectorielle, on peut facilement vérifier que la distribution obtenue ne dépend pas du choix de la solution (non nulle) de cette équation.
- Pour le même raison, et contrairement au cas ouvert, l'équation de trafic (69) 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 (69).

- 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 (70). 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 Basket, Chandy, Muntz et Palacios.

Il n'est pas question ici d'entrer dans les détails de ces réseaux. Le lecteur intéressé se reportera à la bibliographie (par exemple [7]). Dans le paragraphe suivant, nous verrons comment modéliser certaines de ces variantes.

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 de production. Pour ce type de systèmes, on dispose des *Réseaux de Kelly*.

7.2.1 Modélisation

Dans ce modèle, on suppose que les clients appartiennent à des « classes », et qu'à chaque classe k est affectée une route dans le réseau $r_k = (r_k^1, \dots, r_k^{n_k})$, qui est la suite des serveurs que le client doit visiter avant de sortir. Éventuellement, cette liste peut contenir plusieurs fois la même station. Les clients de la classe k arrivent suivant des processus de Poisson, et rejoignent la première file d'attente de leur route. Soit K le nombre de classes, et λ_k le taux d'arrivées de clients de la classe k .

Comme dans les réseaux de Jackson, les serveurs des files d'attente délivrent des temps de service exponentiellement distribués, qu'on supposera tous de paramètre 1, et les services sont supposés être indépendants d'une file à l'autre. Par contre, Kelly introduit d'autres possibilités dans le modèle de file d'attente :

- le *taux de service* de la file i quand il y a n clients présents est $\mu_i(n) > 0$,
- la proportion de ce taux de service consacré au client en position ℓ dans la file, sachant que n clients sont présents, est $\gamma_i(\ell, n)$,
- quand un client rejoint la file i , il est placé en position ℓ avec probabilité $\delta_i(\ell, n)$.

Donc certains paramètres sont autorisés à dépendre du nombre (total) de clients dans la file.

La combinaison de la notion de taux (ou vitesse) de service et de proportion de service ne modifie pas le caractère Markovien du processus. En effet, si un client de la file i est en position ℓ , le temps de service R qui lui reste à faire est distribué exponentiellement de paramètre 1 (toujours en vertu de la propriété sans mémoire de la loi exponentielle). C'est-à-dire que $\mathbb{P}(R \leq x) = 1 - \exp(-x)$. Mais la vitesse du serveur qui lui est dédiée est $\mu_i(n_i)\gamma_i(\ell, n_i)$. Par conséquent, la durée D qui devrait s'écouler jusqu'à la fin de service vaut $D = R/\mu_i(n_i)\gamma_i(\ell, n_i)$. La distribution de cette durée se calcule comme :

$$\mathbb{P}(D \leq x) = \mathbb{P}(R \leq \mu_i(n_i)\gamma_i(\ell, n_i)x) = 1 - \exp(-\mu_i(n_i)\gamma_i(\ell, n_i)x) .$$

Elle est donc bien exponentielle, avec comme paramètre : $\mu_i(n_i)\gamma_i(\ell, n_i)$.

Cette modélisation inclut bien la file d'attente à un serveur, FIFO et taux de service constant : il faut choisir

$$\mu_i(n) \equiv \mu_i , \quad \gamma_i(1, n) = 1 , \gamma_i(\ell, n) = 0 \text{ si } \ell > 1, \quad \delta_i(n, n) = 1 .$$

Mais elle permet aussi d'inclure les serveurs multiples (comme la file $M/M/K$, y compris le nombre infini de serveurs comme dans la file $M/M/\infty$).

Il faut prendre :

$$\mu_i(n) = \min(n, K) \mu_i , \quad \gamma_i(\ell, n) = \begin{cases} 1/n & \text{si } n \leq K, \\ 1/K & \text{si } n > K \text{ et } \ell \leq K, \\ 0 & \text{si } n > K \text{ et } \ell > K, \end{cases} \quad \delta_i(n, n) = 1 .$$

On a également la discipline Processor Sharing :

$$\mu_i(n) \equiv \mu_i , \quad \gamma_i(\ell, n) = 1/n , \text{ pour tout } 1 \leq \ell \leq n, \quad \delta_i(n, n) = 1 .$$

7.2.2 Analyse

Nous passons maintenant à l'étude du processus des clients dans les différentes files d'attente du réseau.

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 :

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

Comme pour les équations de trafic (67), ces équations permettent de déterminer le débit global d'arrivée des clients dans chaque file.

Considérons un réseau de N files. L'état de la file i peut être représenté par la liste des description des clients qui sont dans cette file. La description d'un client contient le numéro de sa route, et le numéro de son étape courante dans cette route. On a donc, comme état du *contenu* d'une file d'attente, un vecteur de paires :

$$\mathbf{c} = ((r(1), s(1)), \dots, (r(n_i), s(n_i))) ,$$

où $r(\ell)$ est la route (ou la classe) du client en position ℓ , et $s(\ell)$ le numéro de l'étape de cette route dans laquelle se trouve le client (cette information n'est utile que si la route passe plusieurs fois par la même file). Si nous notons \mathbf{c}_i l'état de la file i , l'état global du système sera $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_N)$.

Considérons maintenant les transitions possibles dans ce système. Pour simplifier la notation, nous noterons $T_{i,\ell,j,m}\mathbf{C}$ l'état obtenu à partir de l'état \mathbf{C} , suite à la transition du client en position ℓ dans la file i vers la position m de la file j . Nous noterons par le numéro « 0 » l'extérieur du système. Ainsi, $T_{i,\ell,0,-}$ désigne le départ définitif du client en position ℓ dans la file i . Pour les nouvelles arrivées, il est nécessaire de spécifier

la classe du client. On utilisera donc la notation $T_{0,-;j,m}^{(k)}$ pour exprimer que le client qui arrive dans la file j est de classe k .

Tout d'abord, il doit être clair que les transitions d'état dans ce processus ont lieu après des durées qui sont distribuées exponentiellement, et indépendantes du passé. Ceci se déduit du fait que les processus d'arrivée de l'extérieur sont des processus de Poisson, et que les durées de service sont distribuées exponentiellement. La propriété d'être sans mémoire de la distribution exponentielle fait le reste. L'état que nous avons choisi permet de déterminer la valeur du taux de transition, comme nous le verrons un peu plus bas.

Théorème 36. *Soit*

$$b_i = \left(\sum_{n=0}^{\infty} \frac{(\hat{\lambda}_i)^n}{\prod_{p=1}^n \mu_i(p)} \right)^{-1} .$$

Si la condition de stabilité

$$b_i > 0, \quad \forall i \tag{71}$$

est satisfaite, alors la probabilité stationnaire que le réseau se trouve dans l'état $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_N)$ existe et vaut :

$$\mathbb{P}(\mathbf{C}) = \prod_{i=1}^N \mathbb{P}(\mathbf{c}_i) ,$$

et les probabilités marginales des états des files individuelles $\mathbf{c}_i = ((r(1), s(1)), \dots, (r(n_i), s(n_i)))$ sont données par

$$\mathbb{P}(\mathbf{c}_i) = b_i \prod_{p=1}^{n_i} \frac{\lambda_{r(p)}}{\mu_i(p)} . \tag{72}$$

Pour prouver ce théorème, plutôt que de vérifier directement que la distribution proposée est solution des équations d'équilibre (comme au Paragraphe 7.1), on va se servir du Théorème 27. Pour cela, il faut exhiber une probabilité stationnaire candidate (elle est proposée dans l'énoncé du théorème), et des taux de transition candidats pour le processus renversé. Dans le processus renversé, les clients devraient arriver sur la route k avec un taux λ_k , mais commencer par la dernière file de la route, et suivre les étapes à l'envers. Dans chaque file i , le taux de service est $\mu_i(n_i)$ et les clients qui partent le font avec probabilité $\delta_i(n_i, \ell)$ de la position ℓ . Cela est équivalent à supposer que le serveur consacre une proportion $\delta_i(n_i, \ell)$ de sa capacité à la position ℓ de la file d'attente. Quand un client arrive dans la file i , il le fait en position ℓ avec probabilité $\gamma_i(n_i, \ell)$. En conclusion, le processus renversé candidat serait similaire au processus direct, mais avec les routes retournées et les fonctions δ_i et γ_i permutées.

L'étape suivante est d'identifier les taux de transition d'état à état. Nous allons pour cela considérer chaque transition possible $T_{i,\ell;j,m}$. Notons $Q(\mathbf{C}; i, \ell, j, m)$ le taux de transition de l'état \mathbf{C} vers l'état $T_{i,\ell;j,m}\mathbf{C}$ suite à un transfert du client en position ℓ de la file i vers la position m de la file j (i et j peuvent valoir 0, comme convenu plus haut). De la même façon, on notera $Q'(\dots)$ les taux de transition pour le (candidat) processus renversé. Dernière convention pour la notation : $\mathbf{C}.c_i$, $\mathbf{C}.n_i$, $\mathbf{C}.c_i.r(m)$ et des formes similaires seront utilisées pour dénoter le contenu de la file i , le nombre de clients dans cette file, le numéro de la route du client en position m dans la file i , etc., relativement à l'état \mathbf{C} .

On trouve : pour le processus direct (taux de transition $Q(\cdot)$),

État d'origine	Transition	Taux Q
\mathbf{C}	$T_{0,-;j,m}^{(k)}$	$\lambda_k \delta_j(m, \mathbf{C}.n_j + 1)$
\mathbf{C}	$T_{i,\ell;j,m}$	$\mu_i(\mathbf{C}.n_i) \gamma_i(\ell, \mathbf{C}.n_i) \delta_j(m, \mathbf{C}.n_j + 1)$
\mathbf{C}	$T_{i,\ell;0,-}$	$\mu_i(\mathbf{C}.n_i) \gamma_i(\ell, \mathbf{C}.n_i)$

Bien entendu, ces taux de transitions ne sont valables que si la transition est cohérente avec les règles du routage. Pour $T_{i,\ell;j,m}$, il faut que l'étape suivante du client situé en position ℓ dans la file i soit la file j . Pour $T_{i,\ell;0,-}$, il faut que ce client soit dans la dernière étape de sa route. Dans les autres cas, le taux de transition est nul. Pour le processus renversé (taux de transition $Q'(\cdot)$), on trouve :

État d'origine	Transition	Taux Q'
$T_{0,-;j,m}^{(k)} \mathbf{C}$	$T_{j,m;0,-}$	$\mu_j(\mathbf{C}.n_j + 1)\delta_j(m, \mathbf{C}.n_j + 1)$
$T_{i,\ell;j,m} \mathbf{C}$	$T_{j,m;i,\ell}$	$\mu_j(\mathbf{C}.n_j + 1)\delta_j(m, \mathbf{C}.n_j + 1)\gamma_i(\ell, \mathbf{C}.n_i)$
$T_{i,\ell;0,-} \mathbf{C}$	$T_{0,-;i,\ell}^{(k)}$	$\lambda_k \gamma_i(\ell, \mathbf{C}.n_i)$

Considérons une transition interne du réseau de files d'attente : $i \neq 0$ et $j \neq 0$. On va d'abord prouver que si $\mathbf{D} = T_{i,\ell;j,m} \mathbf{C}$, alors $\pi(\mathbf{C})Q(\mathbf{C}; i, \ell, j, m) = \pi(\mathbf{D})Q'(\mathbf{D}; j, m, i, \ell)$. En effet, on a la suite d'équivalences :

$$\begin{aligned}
\pi(\mathbf{C})Q(\mathbf{C}; i, \ell, j, m) &= \pi(\mathbf{D})Q'(\mathbf{D}; j, m, i, \ell) \\
\pi(\mathbf{C}.c_i)\pi(\mathbf{C}.c_j)Q(\mathbf{C}; i, \ell, j, m) &= \pi(\mathbf{D}.c_i)\pi(\mathbf{D}.c_j)Q'(\mathbf{D}; j, m, i, \ell) \\
b_i \prod_{p=1}^{\mathbf{C}.n_i} \frac{\lambda_{\mathbf{C}.c_i.r(p)}}{\mu_i(p)} b_j \prod_{p=1}^{\mathbf{C}.n_j} \frac{\lambda_{\mathbf{C}.c_j.r(p)}}{\mu_j(p)} Q(\mathbf{C}; i, \ell, j, m) &= b_i \prod_{p=1}^{\mathbf{D}.n_i} \frac{\lambda_{\mathbf{D}.c_i.r(p)}}{\mu_i(p)} b_j \prod_{p=1}^{\mathbf{D}.n_j} \frac{\lambda_{\mathbf{D}.c_j.r(p)}}{\mu_j(p)} Q'(\mathbf{D}; j, m, i, \ell) \\
\frac{\lambda_{\mathbf{C}.c_i.r(\ell)}}{\mu_i(\mathbf{C}.n_i)} Q(\mathbf{C}; i, \ell, j, m) &= \frac{\lambda_{\mathbf{D}.c_j.r(m)}}{\mu_j(\mathbf{D}.n_j)} Q'(\mathbf{D}; j, m, i, \ell) .
\end{aligned}$$

Puisque la transition est interne, $\mathbf{D}.n_j = \mathbf{C}.n_j + 1$. D'autre part, puisqu'il s'agit du même client, il s'agit de la même route et $\mathbf{C}.c_i.r(\ell) = \mathbf{D}.c_j.r(m)$. En remplaçant les valeurs des taux de transition, prises dans les tableaux ci-dessus, on a :

$$\begin{aligned}
\frac{1}{\mu_i(\mathbf{C}.n_i)} \mu_i(\mathbf{C}.n_i) \gamma_i(\ell, \mathbf{C}.n_i) \delta_j(m, \mathbf{C}.n_j + 1) &= \frac{1}{\mu_j(\mathbf{C}.n_i + 1)} \mu_j(\mathbf{C}.n_j + 1) \delta_j(m, \mathbf{C}.n_j + 1) \gamma_i(\ell, \mathbf{C}.n_i) \\
\gamma_i(\ell, \mathbf{C}.n_i) \delta_j(m, \mathbf{C}.n_j + 1) &= \gamma_i(\ell, \mathbf{C}.n_i) \delta_j(m, \mathbf{C}.n_j + 1) .
\end{aligned}$$

L'identité initiale est donc vérifiée. Soient maintenant \mathbf{C} et \mathbf{D} deux états. Il est possible que plusieurs transitions mènent du premier état au second. En sommant, on a :

$$\begin{aligned}
\pi(\mathbf{C})Q(\mathbf{C}, \mathbf{D}) &= \sum_{i,\ell,j,m | T_{i,\ell;j,m} \mathbf{C} = \mathbf{D}} \pi(\mathbf{C})Q(\mathbf{C}; i, \ell, j, m) = \sum_{i,\ell,j,m | T_{i,\ell;j,m} \mathbf{C} = \mathbf{D}} \pi(\mathbf{D})Q'(\mathbf{D}; j, m, i, \ell) \\
&= \pi(\mathbf{D})Q'(\mathbf{D}, \mathbf{C}) .
\end{aligned}$$

L'équation de « fausse » balance détaillée (57) est donc établie dans ce cas. Le raisonnement est le même s'il s'agit d'un couple d'états obtenu grâce à une transition de ou vers l'extérieur du réseau.

Il ne reste qu'à montrer que $\sum_{\mathbf{D} \neq \mathbf{C}} Q(\mathbf{C}, \mathbf{D}) = \sum_{\mathbf{D} \neq \mathbf{C}} Q'(\mathbf{C}, \mathbf{D})$. Par inspection des tableaux de transition, il est clair que le taux global des transitions hors de l'état \mathbf{C} vaut :

$$-Q(\mathbf{C}, \mathbf{C}) = \sum_{\mathbf{D} \neq \mathbf{C}} Q(\mathbf{C}, \mathbf{D}) = \sum_{k=1}^K \lambda_k + \sum_{i=1}^N \mu_i(\mathbf{C}.n_i) = -Q'(\mathbf{C}, \mathbf{C}) .$$

Le Théorème 27 s'applique donc, et la distribution proposée est bien la distribution stationnaire du processus. Ceci conclut la preuve du Théorème 36.

État agrégé. Le Théorème de Kelly permet d'obtenir la distribution détaillée de l'état du réseau, mais cette forme n'est pas la plus pratique pour les applications. En examinant la probabilité donnée par l'Équation (72), on s'aperçoit qu'elle ne dépend pas de l'*ordre* des clients dans la file d'attente i , mais seulement du *nombre* de clients de chaque route. Ceci donne l'idée de représenter l'état du réseau par l'ensemble des nombres m_{ik} , $1 \leq i \leq N$, $1 \leq k \leq K$, représentant le nombre de clients de classe k présents dans la file i . Un état global est alors une matrice $N \times K$ à valeurs entières. Soit $\mathcal{M}_{N,K}$ l'ensemble de ces matrices.

La probabilité d'être dans un état donné est décrite dans le théorème suivant.

Corollaire 37. Soit \mathbf{M} une matrice de \mathcal{M}_{NK} . Si la condition de stabilité (71) est satisfaite, alors la probabilité stationnaire que le réseau se trouve dans l'état \mathbf{M} existe et vaut :

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

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.
- S'il n'y a qu'une classe et que les files ont un serveur simple ($\mu_i(p) = \mu_i, \forall p$), 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 $\{\mathbf{M}(t), t \in \mathbb{R}\}$ 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 et calculer les taux de transition. Or l'état $\mathbf{M}(t)$ ne contient pas cette information.

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

7.2.3 Applications

A partir du Théorème 37, on peut calculer toutes les statistiques dont on a besoin, comme par exemple le nombre moyen de clients dans la file i , pour la classe k ,

$$\bar{N}_{ik} = \frac{\hat{\lambda}_{ik}}{\mu_i - \hat{\lambda}_i},$$

ou globalement :

$$\bar{N}_i = \frac{\hat{\lambda}_i}{\mu_i - \hat{\lambda}_i}.$$

De là, à l'aide de la formule de Little (66), le temps moyen passé par un client de classe k dans le réseau :

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

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.

7.3 Réseaux de Ressources

Nous présentons maintenant une autre classe de modèles Markoviens dont la distribution stationnaire peut être calculée facilement.

Par rapport aux réseaux de Jackson et Kelly, leur caractéristique est que les « clients » de ce système occupent plusieurs ressources simultanément, mais qu'il n'y a pas d'attente. Pour cette raison nous les appelons ici « réseaux de ressources ».

Dans les applications à l'Ingénierie des Réseaux, ces modèles servent pour étudier les réseaux dont les communications sont « orientées connexion » (avec réservation de ressources : canaux de communication, espace de stockage dans les routeurs etc.) ou bien pour modéliser les flots de données dans leur ensemble comme les « flux TCP ».

7.3.1 Modélisation

Soit un ensemble de R ressources de différents types numérotés $1, 2, \dots, r, \dots, R$. On considère des clients de différentes classes numérotées $1, 2, \dots, c, \dots, C$. Un client de classe c demande une quantité A_{cr} de ressources de type r , où $A_{cr} \in \mathbb{N}$.

On notera $\mathbf{x} = (x_1, \dots, x_C)$ le vecteur du nombre de clients de chaque classe présents dans le système. Ce nombre évolue en fonction d'arrivées et de départs, comme dans une file d'attente classique. Les règles sont :

- Les clients de classe c arrivent selon un Processus de Poisson dont le taux dépend du nombre de clients. Soit $\lambda_c(\mathbf{x})$ ce taux. Cela signifie que, conditionnellement au fait que l'état courant du système est \mathbf{x} , la distribution de la date de la prochaine arrivée est exponentielle de paramètre $\lambda_c(\mathbf{x})$.
- Le taux de service des clients de classe c dépend également du nombre de clients. Soit $\mu_c(\mathbf{x})$ ce taux. Cela signifie que, conditionnellement au fait que l'état courant du système est \mathbf{x} , la distribution de la date du prochain départ est exponentielle de paramètre $\mu_c(\mathbf{x})$.

Sous ces hypothèses, le processus $\{\mathbf{x}(t)\}$ est une CMTC sur $\mathcal{E} = \mathbb{N}^C$ avec le tableau de transitions :

$$\begin{aligned} \mathbf{x} &\rightarrow \mathbf{x} + \mathbf{e}_c && \text{avec taux } \lambda_c(\mathbf{x}) \\ \mathbf{x} &\rightarrow \mathbf{x} - \mathbf{e}_c && \text{avec taux } \mu_c(\mathbf{x}) \quad \text{si } n_c > 0 \end{aligned}$$

Cette CMTC est irréductible sur \mathbb{N}^C si $\lambda_c(\mathbf{x}) > 0$ et $\mu_c(\mathbf{x}) > 0$ pour tout \mathbf{x} mais ce n'est pas forcément le cas dans les applications. On considèrera donc que \mathcal{E} se réduit à la composante connexe de l'état $\mathbf{0} \in \mathbb{N}^C$.

Sous certaines hypothèses sur les taux d'arrivée et de service, la distribution stationnaire de cette chaîne se calcule simplement. Les réseaux couverts par le théorème sont nommés « réseaux de Whittle ».

Théorème 38 (Whittle). *Supposons qu'il existe deux fonctions $\Lambda, \Phi : \mathcal{E} \rightarrow]0, +\infty[$, telles que pour tout $\mathbf{x} \in \mathcal{E}$ et tout c ,*

$$\Lambda(\mathbf{x})\lambda_c(\mathbf{x}) = \Lambda(\mathbf{x} + \mathbf{e}_c) \tag{73}$$

$$\Phi(\mathbf{x}) = \Phi(\mathbf{x} + \mathbf{e}_c)\mu_c(\mathbf{x} + \mathbf{e}_c), \tag{74}$$

et telles que :

$$\sum_{\mathbf{x} \in \mathcal{E}} \Lambda(\mathbf{x})\Phi(\mathbf{x}) < +\infty.$$

Alors le processus $\{\mathbf{x}(t)\}$ admet comme distribution stationnaire

$$\pi(\mathbf{x}) = \frac{\Lambda(\mathbf{x})\Phi(\mathbf{x})}{\sum_{\mathbf{y} \in \mathcal{E}} \Lambda(\mathbf{y})\Phi(\mathbf{y})}. \tag{75}$$

Démonstration. Nous allons montrer que sous les hypothèses du théorème, le processus $\{\mathbf{x}(t)\}$ vérifie les équations de balance détaillée pour la distribution proposée.

En effet : soient deux états reliés par une transition de taux non nul. Ces deux états s'écrivent nécessairement \mathbf{x} et $\mathbf{x} + \mathbf{e}_c$ pour une certaine classe c . On a alors les équivalences en utilisant (73)–(75) :

$$\begin{aligned} \pi(\mathbf{x})Q_{\mathbf{x}, \mathbf{x} + \mathbf{e}_c} &= \pi(\mathbf{x} + \mathbf{e}_c)Q_{\mathbf{x} + \mathbf{e}_c, \mathbf{x}} \\ \Lambda(\mathbf{x})\Phi(\mathbf{x})\lambda_c(\mathbf{x}) &= \Lambda(\mathbf{x} + \mathbf{e}_c)\Phi(\mathbf{x} + \mathbf{e}_c)\mu_c(\mathbf{x} + \mathbf{e}_c) \\ \Lambda(\mathbf{x} + \mathbf{e}_c)\Phi(\mathbf{x}) &= \Lambda(\mathbf{x} + \mathbf{e}_c)\Phi(\mathbf{x}) \end{aligned}$$

et cette dernière est une identité. □

Exemples. Les exemples suivants appliquent le Théorème 38 dans des cas où l'interaction entre les classes est de plus en plus forte.

Classes autonomes. Considérons d'abord le cas où les arrivées et départs dans la classe c ne dépendent que de la population dans cette classe :

$$\lambda_c(\mathbf{x}) = \tau_c(x_c), \quad \mu_c(\mathbf{x}) = \mu_c(x_c).$$

Dans ce cas, les différents processus $\{x_c(t)\}$ sont indépendants les uns des autres, chacun étant en fait un processus de naissance et de mort. Nous sommes une des situations étudiées au paragraphe 5.4. Le Théorème 38 s'applique avec les fonctions :

$$\Lambda(\mathbf{x}) = \prod_{c=1}^C \prod_{i=0}^{x_c-1} \tau_c(i) \quad \Phi(\mathbf{x}) = \prod_{c=1}^C \prod_{i=1}^{x_c} (\mu_c(i))^{-1},$$

où un produit sur un ensemble vide vaut 1 selon la convention habituelle. Avec cette définition, on retrouve bien que :

$$\Lambda(\mathbf{x})\Phi(\mathbf{x}) = \prod_{c=1}^C \prod_{i=0}^{x_c-1} \frac{\tau_c(i)}{\mu_c(i+1)}.$$

Sous la *condition de stabilité* que, pour toute classe c , $G_c := \sum_n \prod_{i=0}^{n-1} \tau_c(i)/\mu_c(i+1) < +\infty$, on voit ensuite que

$$\sum_{\mathbf{x} \in \mathcal{E}} \Lambda(\mathbf{x})\Phi(\mathbf{x}) = \prod_{c=1}^C G_c$$

et on retrouve finalement que la distribution stationnaire est le produit des distributions marginales :

$$\pi(\mathbf{x}) = \prod_{c=1}^C \frac{1}{G_c} \prod_{i=0}^{x_c-1} \frac{\tau_c(i)}{\mu_c(i+1)}.$$

Classes autonomes avec contraintes de ressources. Supposons maintenant que le nombre de ressources dans chaque type soit limité : notons L_r ce nombre pour la ressource r . Supposons d'autre part que pour le processus décrit précédemment, les clients de la classe c ne peuvent être acceptés que s'il y a assez de ressources disponibles dans chaque type de ressource. Dans le cas contraire, le client est rejeté.¹ La condition d'acceptation pour ce client, dans l'état courant \mathbf{x} , est :

$$\forall r, \quad A_{cr} \leq L_r - \sum_{c'} x_{c'} A_{c'r},$$

que nous noterons sous forme vectorielle : $(\mathbf{x} + \mathbf{e}_c)\mathbf{A} \leq \mathbf{L}$. Les taux de transitions pour ce nouveau processus sont inchangés pour les départs, mais maintenant il faut prendre :

$$\lambda_c(\mathbf{x}) = \tau_c(x_c) \mathbf{1}_{\{(\mathbf{x} + \mathbf{e}_c)\mathbf{A} \leq \mathbf{L}\}}.$$

Le Théorème 38 s'applique avec la nouvelle fonction Λ :

$$\Lambda(\mathbf{x}) = \prod_{c=1}^C \prod_{i=0}^{x_c-1} \tau_c(i) \mathbf{1}_{\{\mathbf{x}\mathbf{A} \leq \mathbf{L}\}}.$$

L'espace d'états (nous avons convenu que c'était la composante connexe du vecteur $\mathbf{0}$) est maintenant un sous ensemble de $\{\mathbf{x} \in \mathbb{N}^C \mid \mathbf{x}\mathbf{A} \leq \mathbf{L}\}$. Selon les valeurs de la matrice \mathbf{A} , tous les états de cet ensemble ne sont pas forcément atteignables depuis $\mathbf{0}$. La *constante de normalisation* $\sum_{\mathbf{y}} \Lambda(\mathbf{y})\Phi(\mathbf{y})$ n'est en général pas évidente à calculer, mais il existe des algorithmes qui peuvent le faire sans énumérer l'espace d'états.

¹Jusqu'ici nous n'avions pas utilisé cette notion de ressources qui justifie le nom des réseaux de ce paragraphe.

Cette situation correspond à la *truncation* du processus décrit plus haut. Nous savions grâce au Théorème 28 que tronquer un processus réversible préserve la réversibilité. Le résultat du Théorème 38 n'est donc pas une nouveauté dans ce cas non plus.

Choix d'une ressource au hasard. Supposons maintenant qu'il n'y a qu'un type de ressource ($R = 1$) en quantité L . Les clients de classe c arrivent selon un processus de Poisson de taux constant τ_c et demandent un nombre $A_{c1} = c$ d'unités de cette ressource. Ils choisissent les c unités au hasard, uniformément dans l'ensemble des L ressources, et sont acceptés si et seulement si toutes ces ressources sont disponibles. Une fois acceptés, les clients de classe c sont traités par un serveur unique et le temps de service est $\text{Exp}(\mu_c)$.

Le taux d'arrivée effectif des clients de classe c dans l'état \mathbf{x} , compte-tenu de la probabilité qu'un client réussisse à trouver c ressources libres, est donné par :

$$\lambda_c(\mathbf{x}) = \tau_c \frac{\binom{L-\mathbf{x}\cdot\mathbf{A}}{c}}{\binom{L}{c}} .$$

On vérifie directement que le Théorème 38 peut être appliqué avec les fonctions :

$$\Lambda(\mathbf{x}) = \prod_{c=1}^C \left(\tau_c \frac{(L-c)!}{L!} \right)^{x_c} \frac{1}{(L-\mathbf{x}\cdot\mathbf{A})!} \quad \Phi(\mathbf{x}) = \prod_{c=1}^C \mu_c^{-x_c} .$$

Une façon plus constructive de vérifier que le processus est bien réversible est d'appliquer le critère de Kolmogorov. Sur une maille carrée du graphe de transitions, on doit vérifier l'égalité des produits sur les cycles : $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{e}_i \rightarrow \mathbf{x} + \mathbf{e}_i + \mathbf{e}_j \rightarrow \mathbf{x} + \mathbf{e}_j \rightarrow \mathbf{x}$ et $\mathbf{x} \rightarrow \mathbf{x} + \mathbf{e}_j \rightarrow \mathbf{x} + \mathbf{e}_i + \mathbf{e}_j \rightarrow \mathbf{x} + \mathbf{e}_i \rightarrow \mathbf{x}$. On devrait donc avoir :

$$\tau_i \frac{\binom{L-\mathbf{x}\cdot\mathbf{A}}{i}}{\binom{L}{i}} \tau_j \frac{\binom{L-\mathbf{x}\cdot\mathbf{A}-i}{j}}{\binom{L}{j}} \mu_i \mu_j = \tau_j \frac{\binom{L-\mathbf{x}\cdot\mathbf{A}}{j}}{\binom{L}{j}} \tau_i \frac{\binom{L-\mathbf{x}\cdot\mathbf{A}-j}{i}}{\binom{L}{i}} \mu_j \mu_i ,$$

soit donc, avec $\ell = \mathbf{x}\cdot\mathbf{A}$:

$$\binom{L-\ell}{i} \binom{L-\ell-i}{j} = \binom{L-\ell}{j} \binom{L-\ell-j}{i} ,$$

et on vérifie facilement cette identité combinatoire.

Bibliographie

- [1] T. Bonald. Insensitive queueing models for communication networks. In *Proc. ValueTools'09*, Pisa, Italia, 2009.
- [2] W. Feller. *An introduction to probability theory and its applications*, volume I. J. Wiley & Sons, New York, troisième édition, 1968.
- [3] F. Kelly. *Reversibility and Stochastic Networks*. J. Wiley & Sons, Chichester, 1979. Disponible sur le site web de l'auteur.
- [4] J.G. Kemeny and Snell J.L. *Finite Markov Chains*, volume 76. Springer-Verlag, 1976.
- [5] L. Kleinrock. *Queueing Systems*, volume I, Theory. J. Wiley & Sons, New York, 1975.
- [6] H. Minc. *Nonnegative Matrices*. J. Wiley & Sons, 1988.
- [7] G. Pujolle and S. Fdida. *Modèles de Systèmes et de Réseaux – Tome I : Performance*. Eyrolles, 1989.
- [8] S. Ross. *Stochastic Processes*. Wiley & Sons, 1996.
- [9] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1974.
- [10] R. Sedgewick and Ph. Flajolet. *Analysis of Algorithms*. Addison-Wesley, 1996. Voir les sites web des auteurs.
- [11] E. Seneta. *Non-negative Matrices and Markov chains*. Springer-Verlag, second edition, 1981.
- [12] L. Takács. *An introduction to the theory of queues*. Oxford University Press, New York, 1962.
- [13] E.C. Titchmarsh. *The Theory of Functions*. Oxford University Press, second edition, 1939.
- [14] H.S. Wilf. *Generatingfunctionology*. Academic Press, second edition, 1994. Disponible sur le site web de l'auteur.

A Fonctions génératrices en calcul des probabilités

Les méthodes de l'évaluation de performances (systèmes à événements discrets stochastiques) font souvent appel aux *fonctions (ou : séries) génératrices*. Soit X une variable aléatoire discrète, à valeurs dans \mathbb{N} . On définit sa *fonction génératrice de probabilités* par :

$$F_X(z) = \mathbb{E}(z^X) = \sum_{n=0}^{\infty} \mathbb{P}(X = n)z^n . \quad (76)$$

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 :

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

- 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* :

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

- utiliser la *formule de Cauchy* :

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

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 :

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

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 39. 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 40 (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).

B 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» $i \rightarrow i$. 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 [6] :

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

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

$$f_A(x) = \max\{\rho \in \mathbb{R} \mid Ax - \rho x \geq 0\}. \quad (77)$$

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 $f_A(x) \geq \max\{\rho \in \mathbb{R} \mid Ax - \rho x \geq 0\}$. Réciproquement, si $r = f_A(x)$, alors $(Ax)_i \geq r x_i$ pour tout i , y compris ceux tels que $x_i = 0$. Donc $f_A(x) \in \{\rho \in \mathbb{R} \mid Ax - \rho x \geq 0\}$, et par conséquent, $f_A(x) \leq \max\{\rho \in \mathbb{R} \mid Ax - \rho x \geq 0\}$.

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 $S_n = \{x \in (\mathbb{R}^+)^n \mid \sum_i x_i = 1\}$. 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 $x_i \rightarrow 0$. 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 (77), pour tout x , $Ax \geq f_A(x)x$. En multipliant par $(I + A)^{n-1}$, on a, puisque A et $I + A$ commutent :

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

toujours d'après (77). 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 (77) 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 (77), 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.

C La notation de Kendall

Cet appendice est consacré à la description rapide du modèle de files d'attente couramment utilisé.

Les files d'attente ont différents constituants, et leurs propriétés quantitatives dépendent en général de certaines hypothèses de nature probabiliste. Pour pouvoir s'y retrouver, il est utile de pouvoir résumer rapidement les caractéristiques principales de la file. On utilise pour cela la **notation de Kendall**. C'est un symbole de la forme

$$A/ S/ P/ K/ D$$

où :

- A désigne la **loi des inter-arrivées**. Le processus d'arrivée des clients étant représenté par une suite de dates $a_1 \leq a_2 \leq \dots \leq a_n \leq a_{n+1} \leq \dots$, l'inter-arrivée $\tau_n = a_{n+1} - a_n$ est la durée qui sépare deux arrivées consécutives. On suppose que c'est une variable aléatoire, avec une certaine loi. Selon la nature de cette loi, A peut prendre la valeur (entre autres) :
 - M pour la distribution exponentielle (M comme Markov, ou « memoryless », voir au paragraphe 4.1) ; cette loi de probabilité est telle que :

$$\mathbb{P}(\tau_n \leq x) = 1 - e^{-\lambda x} ,$$

où $\lambda > 0$ est un paramètre. Les variables aléatoires τ_n sont de plus supposées indépendantes. Dans ce cas, le processus d'arrivée est un appelé un *processus de Poisson* de paramètre λ . On calcule que son débit est λ .

- D pour la distribution déterministe ; dans ce cas, le processus d'arrivée est périodique.
 - G (ou GI) pour une distribution générale (Générale Indépendante : les variables aléatoires τ_n sont supposées indépendantes).
- Cette liste n'est pas exhaustive.
- S désigne la **loi de service**. Selon la nature de cette loi, S peut prendre la valeur
 - M pour la distribution exponentielle. Dans ce cas, on a :

$$\mathbb{P}(\sigma_n \leq x) = 1 - e^{-\mu x} ,$$

pour un paramètre $\mu > 0$, et la moyenne du temps de service est donnée par $\bar{\sigma} = 1/\mu$.

- D pour la distribution déterministe.
- G (ou GI) pour une distribution générale.
- P désigne le **nombre de serveurs**. P peut prendre n'importe quelle valeur entière, y compris l'infini. Par défaut, c'est 1.
- K désigne la **capacité du système**, c'est-à-dire le nombre maximum de clients qui peuvent être présents simultanément, en comptant le ou les clients en service. K peut prendre n'importe quelle valeur supérieure à P , y compris l'infini. Par défaut, c'est l'infini.

Note : il arrive que K désigne en fait la capacité de la salle d'attente, hors serveur. Notre convention sera de compter les serveurs.

– D désigne la **discipline de service**, ou encore : **politique de service**, qui détermine comment sont servis les clients.

D peut prendre les valeurs suivantes.

- FIFO (First-In-First-Out) soit, en français, PAPS (Premier Arrivé Premier Servi) : service dans l'ordre défini par les arrivées. On rencontre également l'acronyme FCFS (First-Come-First-Served).
- LIFO (Last-In-First-Out) soit, en français, DAPS (Dernier Arrivé Premier Servi) : comportement de pile.
- PS (Processor Sharing) : le serveur est partagé entre tous les clients présents. Si les clients ont des classes, plusieurs variantes existent : le processeur peut être partagé de façon égalitaire entre les classes (puis FIFO dans chaque classe, par exemple), ou de façon pondérée, avec certaines classes favorisées vis à vis d'autres.
- Priorités : il y a plusieurs classes de clients, chaque classe avec un niveau de priorité. C'est le client qui a la priorité la plus haute qui est servi en premier. S'il y en a plusieurs, c'est le premier arrivé (discipline FIFO par classe).

La discipline FIFO est le défaut.

Les disciplines LIFO et Priorité ont des variantes préemptives et non-préemptives, selon que le client en service est interrompu si arrive un client plus prioritaire, ou non. Dans le cas préemptif, on distingue encore le cas « continuation » (en anglais : « resume ») où un client interrompu reprend son service là où il en était avant l'interruption, du cas « recommencement » (en anglais : « restart ») où il reprend son service à zéro. Dans ce dernier cas, le travail effectué au moment de l'interruption est perdu.

Les symboles K et D admettent des valeurs par défaut. Ainsi,

$M/M/1/K$ est équivalent à $M/M/1/K/FIFO$

et

$M/M/1$ est l'abréviation de $M/M/1/\infty/FIFO$.

Table des matières

1	Introduction	1
2	Puissances de Matrices	2
2.1	Matrices et graphes	2
2.2	Matrices et Chemins	2
2.3	Puissances de matrices : exemples	3
2.4	Puissances de matrices : asymptotiques et calcul exact	4
2.5	Matrices positives	7
3	Chaînes de Markov en temps discret	8
3.1	Définitions	8
3.2	Chaînes de Markov à espace d'état fini	8
3.2.1	Chaînes de Markov et matrices	8
3.2.2	Chaînes de Markov et graphes.	8
3.2.3	Probabilités transitoires	9
3.2.4	Spectre des matrices stochastiques ; Résolvante	9
3.2.5	Comportement asymptotique	10
3.2.6	Classification des états	12
3.3	Chaînes de Markov à espace d'état dénombrable	13
3.4	La marche aléatoire en dimension 1	13
3.4.1	Probabilités stationnaires, cas fini	14
3.4.2	Probabilités transitoires, cas fini	15
3.4.3	Probabilités stationnaires, cas infini	17
4	Chaînes de Markov en temps continu	17
4.1	Définitions	17
4.2	Équations de Chapman-Kolmogorov	19
4.3	Construction du générateur	19
4.4	Uniformisation	20
4.5	Comportement asymptotique	21
4.6	Chaîne incluse aux moment de saut	22
5	Distribution stationnaire d'une Chaîne de Markov et réversibilité	22
5.1	Équations de balance	22
5.2	Réversibilité	25
5.3	Troncation de chaînes de Markov	26
5.4	Composition de chaînes de Markov	27
5.5	Générateurs réversibles et diagonalisation	28
6	Éléments de théorie des files d'attente markoviennes	28
6.1	La file « $M/M/1$ »	28
6.2	La file « $M/M/1/K$ »	30
6.3	La file « $M/M/C$ »	30
6.4	La file « $M/M/\infty$ »	31
6.5	Calcul de probabilités stationnaires par récurrences	32
6.6	Les temps de réponse et la formule de Little	33

7	Les Réseaux de files d'attente	33
7.1	Les réseaux de Jackson	34
7.2	Les réseaux de Kelly. Les modèles à commutation de paquets.	36
7.2.1	Modélisation	36
7.2.2	Analyse	37
7.2.3	Applications	40
7.3	Réseaux de Ressources	40
7.3.1	Modélisation	41
A	Fonctions génératrices en calcul des probabilités	45
B	Preuve du théorème de Perron-Frobenius	46
C	La notation de Kendall	47

Index

- absorbant
 - état, 12
- chaînes de Markov
 - troncation, 43
- chaînes de Markov
 - composition, 27
 - constructions, 19
 - en forme d'arbre, 23
 - en temps continu, 17
 - définition, 18
 - en temps discret, 8
 - définition, 8
 - incluses, 22
 - réversibles, 26
 - troncation, 26
 - uniformisées, 20
- classe de clients, 36
- classe terminale, 8
- Collatz-Wielandt, 6, 46
- condition de stabilité, 29, 31, 32, 34, 38, 42
- critère de Kolmogorov, 24, 43
- débit, 35–37
- décomposition spectrale, 4
- diagramme de transition, 8
- distribution
 - à forme produit, 35, 42
 - de Poisson, 18, 32
 - déterministe, 47
 - exponentielle, 17, 37
 - géométrique, 17, 29
 - initiale, 9
 - stationnaire, 12
- équation
 - caractéristique d'une récurrence, 32
 - de normalisation, 15, 29, 31, 32, 42
- équations
 - d'équilibre, 23
 - de balance détaillée, 23, 26, 29, 41
 - de balance globale, 23
 - de balance locale, 23
 - de Chapman-Kolmogorov, 19
 - de trafic, 35–37
- exponentielle
 - distribution, 17
 - variable aléatoire, 47
- file d'attente
 - $M/M/1$, 28, 35
 - $M/M/1/K$, 30
 - $M/M/C$, 30, 37
 - $M/M/\infty$, 31, 37
- fonction
 - de Collatz-Wielandt, 46
 - génératrice de probabilités, 45
- forme produit, 35, 40
- formule de Little, 33, 40
- générateur infinitésimal, 19
 - construction, 19
 - du processus renversé, 25
- géométrique
 - convergence, 11, 21
 - distribution, 17
- insensibilité, 32
- irréductibilité, 2
- Jordan, décomposition de, 4
- Kendall, notation de, 28, 30, 47
- Kolmogorov
 - critère de, 24, 43
 - équations de Chapman-Kolmogorov, 19
- Little, formule de, 33, 40
- marche aléatoire, 13
- matrice
 - apériodique, 2
 - compagnon, 4
 - d'adjacence, 3
 - de transition, 8
 - de routage, 34
 - ergodique, 12
 - irréductible, 2
 - primitive, 12
 - périodique, 2
 - rayon spectral, 2
 - sous-stochastique, 10, 34
 - stochastique, 8
- méthode du noyau, 15, 16
- moment factoriel, 45
- Perron-Frobenius
 - pour matrices stochastiques, 9
 - théorème de, 7
- Poisson

- distribution, 18, 32
- processus, 18, 47
- probabilités
 - de transition, 8
 - stationnaires, 9
 - transitoires, 9
- processor sharing, 37, 48
- processus
 - d'arrivée, 47
 - de comptage, 18
 - de naissance et de mort, 13, 26, 29, 42
 - de Poisson, 18, 47
 - renversé, 25, 29
 - réversible, 25
- propriété
 - sans mémoire, 18, 28, 30, 37, 38
- périodicité, 2
- récurrent, état, 12, 13
- récurrent nul, état, 13
- réseau
 - BCMP, 36
 - de Jackson fermé, 35
 - de Jackson ouvert, 34
 - de Kelly, 36
 - de ressources, 40
- résolvante, 10
- réversibilité, 25
- route, 36
- série
 - génératrice, 45
 - potentielle, 10
- taux
 - d'arrivée, 18
 - de croissance, 6
 - de défaillance, 18
 - de service, 29, 36
 - de transition, 20
- théorème
 - de Burke, 29
 - de décomposition spectrale, 4
 - de Jackson, 34, 35
 - de Kelly, 38, 39
 - de Perron-Frobenius, 7
 - de Pringsheim, 46
 - de solidarité, 13
 - de Whittle, 41
- transient, état, 12, 13
- uniformisation, 20, 29
- vitesse de service, 29, 37