

Cours de

# Processus aléatoires

## CHAPITRE VII

### Chaînes de Markov

---

La retombée la plus usuelle de la "Recherche Opérationnelle contre le hasard" est la théorie des files d'attente, qui a des applications pratiques tout à fait remarquables : elle permet, par exemple, d'évaluer le temps moyen que l'on devra passer à "faire la queue" dans un bureau de poste, avant d'être servi ... Nous verrons qu'il existe un cheminement logique incontournable entre le présent chapitre, traitant des chaînes de Markov, et les deux suivants, qui concernent les processus de Markov et les files d'attente.

Le modèle de la chaîne de Markov permet de traduire l'évolution temporelle de certains "**processus stochastiques**" (i.e. régis par le hasard), supposés **sans mémoire**, dans un sens défini plus loin.

On considère un "système", susceptible de se trouver dans différents "états" ; à un instant quelconque, le système est dans un état et un seul.

On suppose connue la liste exhaustive des  $n$  états possibles, soit  $E_i$ ,  $i = 1, 2, \dots, n$ .

L'état du système n'est observé qu'à des instants discrets équidistants :  $t = k \cdot \tau$ , où  $k$  est un nombre entier. La suite des états observés aux instants successifs  $0, \tau, 2\tau, \dots$  forme une chaîne.

Pour que cette chaîne soit une **chaîne de Markov**, il faut que les changements d'état respectent la propriété suivante, traduisant l'absence de mémoire :

Si le système est dans l'état  $E_i$  à l'instant  $k \cdot \tau$ , il possède la probabilité  $p_{ij}(k)$  d'être dans l'état  $E_j$  à l'instant d'observation suivant,  $(k + 1) \cdot \tau$ .

La probabilité  $p_{ij}(k)$ , qui est une **probabilité conditionnelle**, est indépendante du "parcours" du système avant l'instant  $k \cdot \tau$  : tout le passé du système se résume ainsi dans l'état présent.

On suppose que la fréquence  $\left(\frac{1}{\tau}\right)$  des observations est suffisante pour qu'aucun changement d'état (on dit aussi "transition") ne puisse être occulté ; autrement dit, si le système est dans l'état  $E_i$  à l'instant  $k \cdot \tau$ , il n'y a que deux possibilités pour l'instant  $(k + 1) \cdot \tau$  :

- le système est encore dans l'état  $E_i$  ;
- le système est passé dans l'un des autres états, mais il y a eu **un seul** changement d'état entre  $k \cdot \tau$  et  $(k + 1) \cdot \tau$ .

En outre, on suppose généralement que  $p_{ij}$  est constante, indépendante de  $k$ , c'est-à-dire du temps : la chaîne de Markov est dite alors **homogène**. Les probabilités de transition entre états sont consignées dans une matrice, soit  $M$ , dénommée **matrice de transition**. Par construction,  $M$  est une matrice carrée d'ordre  $n$  ; la somme des termes d'une ligne quelconque de  $M$  vaut 1.

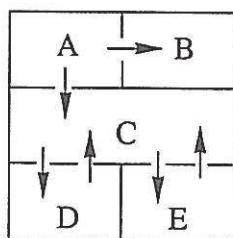
## Graph de transition

Les transitions entre états peuvent être représentées par un graphe, dénommé **graphe de transition**, obtenu comme suit :

- les sommets sont associés aux états ;
- les arcs sont associés aux transitions possibles entre deux instants successifs  $k \cdot \tau$  et  $(k + 1) \cdot \tau$  ;
- les valuations sont les probabilités de transition correspondantes.

### *Exemple :*

Un expérimentateur dispose d'une cage complexe (composée de cinq cases A, B, C, D et E) dans laquelle il enferme un animal afin d'étudier son comportement :



Il observe l'animal à des instants rapprochés ( $0, \tau, 2\tau, 3\tau, \dots$ ) entre lesquels le captif ne peut effectuer qu'une "transition" : aller vers une autre case ou décider de rester dans la case où il se trouve. Ainsi, entre deux instants voisins  $m\tau$  et  $(m+1)\tau$  ( $m \in \mathbb{N}$ ) : l'animal peut aller de A vers B avec la probabilité 0,5, de A vers C avec la probabilité 0,3, ou rester en A. De C, il peut aller en D avec la probabilité 0,4, ou vers E avec la probabilité 0,3, ou rester en C. De D, il peut aller vers C avec la probabilité 0,3, ou rester en D. Enfin de E, il peut aller vers C avec la probabilité 0,4, ou rester en E.

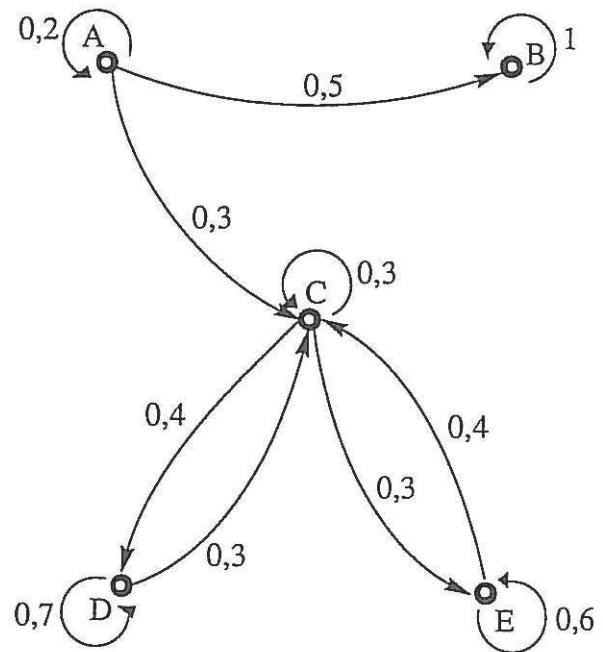
On s'intéresse à la position de l'animal à tout instant  $k\tau$  ( $k \in \mathbb{N}$ ). On peut associer à cette position aléatoire le modèle d'une chaîne de Markov.

Ici :

- le "système" est une cage à cinq compartiments, renfermant un animal ;
- les 5 "états" possibles, dénommés A, B, ..., E, correspondent à la présence de l'animal dans les compartiments de même nom.

La matrice et le graphe de transition sont les suivants :

	A	B	C	D	E
A	0,2	0,5	0,3	0	0
B	0	1	0	0	0
C	0	0	0,3	0,4	0,3
D	0	0	0,3	0,7	0
E	0	0	0,4	0	0,6



### Régime transitoire d'une chaîne de Markov

On suppose qu'initialement ( $t = 0$ ), le système est abandonné dans un état donné. Du fait de la nature stochastique des transitions, il n'est évidemment pas possible de prévoir l'état précis dans lequel se trouvera le système à  $t = k \cdot \tau$ .

Par contre, on peut calculer la probabilité de chacun des états à  $t = k \cdot \tau$ ; on désigne par  $\Pi_k$  un vecteur-ligne, de format  $(1, n)$ , regroupant ces probabilités.

La somme des  $n$  composantes du vecteur  $\Pi_k$  est, par définition, égale à 1.

Le calcul de  $\Pi_k$  peut être effectué en appliquant la relation de récurrence :

$$\boxed{\Pi_k = \Pi_{k-1} \cdot M} \quad (k \geq 1)$$

qui découle directement des définitions de  $\Pi_k$  et de  $M$ .

En multipliant membre à membre les relations de récurrence associées à des valeurs successives de  $k$ , on obtient la relation :

$$\boxed{\Pi_k = \Pi_0 \cdot M^k} \quad (k \geq 1)$$

qui permet un calcul direct de  $\Pi_k$ .

**Exemple :**

Reprenons l'exemple précédent, en supposant qu'à  $t = 0$ , l'animal est placé dans le compartiment A de la cage.

Les probabilités de trouver l'animal en A, B, C, D ou E à  $t = \tau$ , puis à  $t = 2\tau$ , sont données par :

$$\begin{cases} \Pi_1 = \Pi_0 \cdot M \\ \Pi_2 = \Pi_1 \cdot M \end{cases}$$

avec  $\Pi_0 = (1 \ 0 \ 0 \ 0 \ 0)$ .

On trouve : 
$$\begin{cases} \Pi_1 = (0,2 \ 0,5 \ 0,3 \ 0 \ 0) \\ \Pi_2 = (0,04 \ 0,6 \ 0,15 \ 0,12 \ 0,09) \end{cases}$$

(on vérifie que la somme des composantes de chaque vecteur est bien égale à 1).

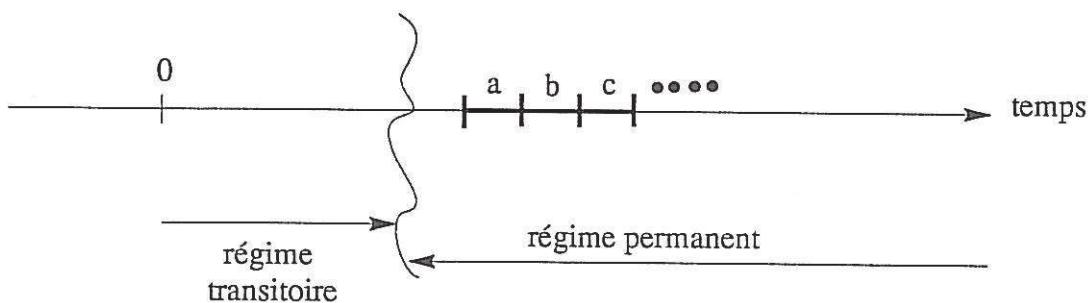
## Régime permanent d'une chaîne de Markov

On se pose ici la question suivante : au bout d'un "grand" nombre d'instants d'observation, existe-t-il une limite, soit  $\Pi_\infty$ , aux probabilités des différents états du système :

$$\Pi_\infty = \lim_{k \rightarrow +\infty} \Pi_k ?$$

Cette question est liée à l'**ergodicité** de la chaîne de Markov : on dit qu'**une chaîne de Markov est (fortement) ergodique si, lorsque k tend vers l'infini, le vecteur  $\Pi_k$  tend vers un vecteur constant,  $\Pi_\infty$ , indépendant de  $\Pi_0$ .**

Lorsque l'ergodicité a lieu, elle traduit une sorte de "**stabilisation macroscopique**" atteinte par la chaîne de Markov "en régime permanent". On peut, en effet, lui donner l'interprétation "statistique" suivante. Considérons, en régime permanent, des périodes d'observation successives du système, soit a, b, c ... :



on suppose que ces périodes a, b, c, ... sont de même durée, et que cette durée est "suffisante" vis-à-vis de  $\tau$  (par exemple  $100 \cdot \tau$ ).

Au cours du "régime permanent", le système continue à faire l'objet de mouvements incessants. Cependant, si l'on comptabilise, pour chacune des périodes a, b, c, ..., le temps cumulé passé par le système dans chacun des états, on constate que le temps cumulé passé dans un état donné est le même au cours des périodes successives. Autrement dit, le système passe désormais une proportion constante de son temps dans chacun des n états (les proportions étant données par les n composantes du vecteur  $\Pi_\infty$ ), bien que l'"**agitation microscopique**" se poursuive sans relâche à l'échelle des instants successifs  $k\tau$  d'observation.

Cette "**stabilisation macroscopique**" de la chaîne de Markov, lorsqu'elle a lieu, présente un grand intérêt, car elle confère au processus aléatoire une **prévisibilité** inespérée, qui peut être précieuse, en particulier en cas d'enjeu économique.

Il existe seulement des conditions suffisantes d'ergodicité, portant sur le graphe de transition :

La chaîne de Markov est ergodique si le graphe de transition possède les 3 propriétés suivantes :

- 1°) graphe fini (i.e. nombre fini de sommets)
- 2°) graphe **fortement connexe** (i.e. il existe au moins un chemin entre deux sommets quelconques)
- 3°) graphe comportant au moins une boucle (arc reliant un sommet à lui-même).

Lorsqu'il existe,  $\Pi_\infty$  se calcule par la relation :

$$\boxed{\Pi_\infty = \Pi_\infty \cdot M}$$

qui découle immédiatement de la relation de récurrence vue plus haut, par passage à la limite. En outre, la somme des composantes de  $\Pi_\infty$  est, bien entendu, égale à 1.

### **Exemples :**

- On reprend l'exemple précédent, en supposant cette fois que l'animal est placé dans la case C à  $t = 0$ . On constate alors que les cases A et B ne sont plus accessibles, si bien que le graphe de transition se simplifie sous la forme d'un graphe à 3 sommets, qui respecte les conditions suffisantes d'ergodicité.

Les composantes de  $\Pi_\infty$  sont tirées des relations :

$$\left\{ \begin{array}{l} (\Pi_\infty^C, \Pi_\infty^D, \Pi_\infty^E) = (\Pi_\infty^C, \Pi_\infty^D, \Pi_\infty^E) \cdot \begin{pmatrix} 0,3 & 0,4 & 0,3 \\ 0,3 & 0,7 & 0 \\ 0,4 & 0 & 0,6 \end{pmatrix} \\ \Pi_\infty^C + \Pi_\infty^D + \Pi_\infty^E = 1 \end{array} \right.$$

On trouve :  $\Pi_\infty^C = 12/37$ ,  $\Pi_\infty^D = 16/37$ , et  $\Pi_\infty^E = 9/37$ .

- Un autre exemple est proposé dans le fascicule "Exercices de Recherche Opérationnelle" (deuxième problème). Il permet d'illustrer :
  - l'intérêt économique du cas de la chaîne de Markov ergodique ;
  - deux phénomènes particuliers pouvant survenir en régime permanent : l'état permanent absorbant et le cycle périodique (cf question 4).

\* \* \* \* \*

# CHAPITRE VIII

## Processus de Markov

---

Ce terme est réservé à la généralisation du modèle markovien précédent au cas où le temps est continu, et non plus discrétisé en instants d'observation  $t = k \cdot \tau$ .

Pour que l'hypothèse markovienne de l'absence de mémoire soit respectée, il faut, dans le cas continu, supposer que les événements aléatoires, qui provoquent les changements d'"état" du "système", obéissent à une **loi de Poisson**.

Par exemple, lorsqu'on veut décrire l'arrivée des clients à un arrêt d'autobus, au guichet d'une banque, l'occurrence des pannes de composants électroniques, des appels à un central téléphonique, le passage des véhicules, en circulation fluide, en un point donné, sur une route, etc ..., on a fréquemment recours à une loi statistique de Poisson.

### Loi de Poisson

On suppose que les occurrences des événements - par exemple, les arrivées des clients - sont :

- indépendantes entre elles ;
- à flux stationnaire : ce qui se produit entre  $t$  et  $t + h$  ne dépend que de  $h$  ;
- non groupées : la probabilité de plus d'un événement entre  $t$  et  $t + h$  est très faible si  $h$  est petit : c'est un infiniment petit  $O(h)$ , c'est-à-dire  $O(h)/h$  tend vers 0, quand  $h$  tend vers 0 .

Les arrivées peuvent alors être modélisées par une loi de Poisson.

Par définition, les arrivées suivent une **loi de Poisson de taux  $\lambda$**  lorsque la probabilité d'observer **n arrivées pendant un temps t** est donnée par :

$$P_n(t) = \frac{(\lambda t)^n}{n!} \cdot e^{-\lambda t}$$

Parfois, l'occurrence d'un événement marque la fin d'une période : par exemple, la panne d'un composant électronique indique la fin de la vie de celui-ci ; ou le départ d'un client d'un guichet



coïncide avec l'achèvement d'un service. Dans ce cas, on s'intéresse souvent à une autre variable aléatoire : la "durée de vie" du composant, ou la "durée de service".

Lorsque l'occurrence d'évènement de fin est régi par une loi de Poisson de taux  $\lambda$ , la durée de vie associée suit une **loi exponentielle** de même taux. Les deux modes de raisonnement sont donc rigoureusement équivalents.

*Remarque :*

Pour le composant électronique, par exemple, la loi exponentielle, dite de "non-vieillissement", est une manifestation particulière de l'hypothèse markovienne de l'absence de mémoire : elle se traduit par un taux de mortalité constant du composant. Certains composants électroniques ont cette propriété remarquable, tout au long de leur vie utile ; par contre, ils peuvent connaître un taux important de mortalité infantile, et ce taux augmente également au bout d'un temps long ; d'où une "courbe en baignoire" du taux de mortalité en fonction du temps, classiquement rencontrée en fiabilité.

Dans la suite de ce cours, la loi de Poisson, ou la loi exponentielle, de taux  $\lambda$  se traduit de la manière suivante :

- la probabilité d'occurrence d'un évènement, pendant le temps ("petit")  $dt$ , est égale à  $\lambda \cdot dt$  ;
- la probabilité d'occurrence de  $n$  évènements, pendant le même temps  $dt$ , de la forme  $(\lambda \cdot dt)^n$ , est supposée négligeable devant la probabilité d'occurrence d'un seul évènement, quel que soit  $n > 1$ .

## Graphe de transition

L'occurrence d'un évènement s'accompagne d'un changement d'état du système. Compte tenu des considérations précédentes sur la loi de Poisson, supposée régir les évènements, on peut énoncer la propriété suivante, voisine de celle des chaînes de Markov :

Si le système est dans l'état  $E_i$  à l'instant  $t$ , il n'y a que deux possibilités pour l'instant  $t + dt$  :

- le système est encore dans l'état  $E_i$  ;
- le système est passé dans l'un des autres états, mais il y a eu **un seul** changement d'état entre  $t$  et  $t + dt$ .

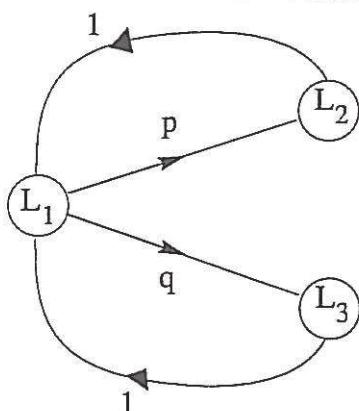
On peut alors représenter, dans le graphe de transition du processus de Markov, toutes les transitions possibles d'un état à un autre, entre  $t$  et  $t + dt$  (comme on le fait pour une chaîne de Markov, entre  $k\tau$  et  $(k+1)\cdot\tau$ ).

**Exemple :**

Dans un magasin de stockage, des composants sont stockés en trois lieux  $L_1$ ,  $L_2$  et  $L_3$ . Deux robots identiques permettent de prélever des composants en chacun de ces lieux, selon des programmes pré-mémorisés.

La durée aléatoire nécessaire pour un prélèvement en  $L_i$  ( $i = 1, 2, 3$ ) est régie par une loi exponentielle de taux  $\mu_i$ . Lorsque les deux robots se trouvent au même lieu, le dernier arrivé doit attendre la fin de l'opération du premier.

La circulation des robots s'effectue selon le schéma suivant :



Pour chaque robot, chaque programme commence par un prélèvement en  $L_1$ , puis le robot se dirige (en un temps négligeable par rapport à celui des prélèvements) vers  $L_2$  avec la probabilité  $p$ , ou vers  $L_3$  avec la probabilité  $q$  ( $p + q = 1$ ). Ensuite, il revient systématiquement en  $L_1$  (aussi en un temps négligeable), pour effectuer le programme suivant.

Chaque robot est d'utilisation permanente et comporte, à tout instant, plusieurs programmes de prélèvement en mémoire (ces programmes diffèrent entre eux par le choix de  $L_2$  ou  $L_3$ , et éventuellement par la nature du composant à prélever). Les pannes des robots, très rares en pratique, ne sont pas prises en compte ici.

Nous allons modéliser le système, constitué par les trois lieux et les deux robots, à l'aide d'un processus de Markov.

\*\*\*\*\*

On peut désigner les états par des triplets de la forme  $(k_1, k_2, k_3)$ , où  $k_i$  est le nombre de robots présents au lieu  $L_i$  ( $i = 1, 2, 3$ ).

D'évidence, on a :  $k_1 + k_2 + k_3 = 2$ . Il y a six états possibles : 200, 020, 002, 110, 101 et 011.

La probabilité pour qu'un robot se trouvant en  $L_1$ , à l'instant  $t$ , effectue un prélèvement, entre  $t$  et  $t + dt$ , vaut  $\mu_1 \cdot dt$ , par définition de la loi exponentielle de taux  $\mu_1$ . Ce robot a ainsi une probabilité  $p \cdot \mu_1 \cdot dt$  de se trouver en  $L_2$  à  $t + dt$ , et une probabilité  $q \cdot \mu_1 \cdot dt$  de se trouver en  $L_3$  à  $t + dt$ .

En conséquence, si l'on tient compte des deux robots, et si l'on suppose d'abord que le système se trouve dans l'état 200 à l'instant  $t$ , on voit qu'il n'y a que 3 possibilités, à l'instant  $t + dt$  :

- le système est passé dans l'état 110 (probabilité  $p \cdot \mu_1 \cdot dt$ ) ;
- le système est passé dans l'état 101 (probabilité  $q \cdot \mu_1 \cdot dt$ ) ;
- le système est encore dans l'état 200  
(probabilité  $1 - p\mu_1 dt - q\mu_1 dt = 1 - \mu_1 dt$ ).

(un évènement au plus s'est produit pendant le temps  $dt$ ).

On a ainsi déterminé les arcs du graphe de transition issus du sommet 200.

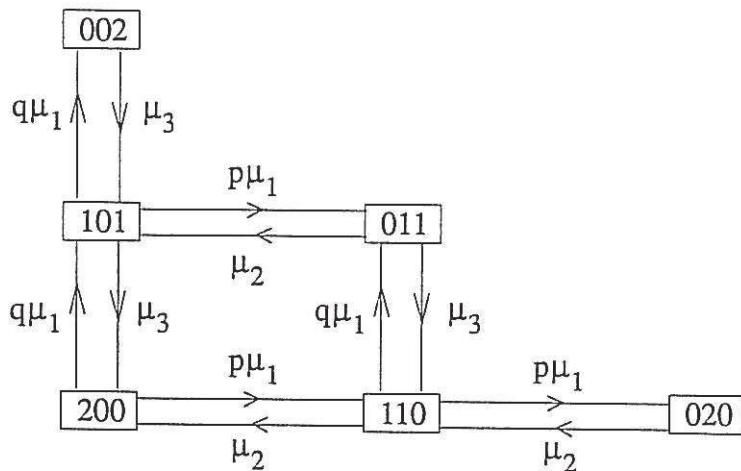
En raisonnant de même, successivement, à partir des 5 autres états, on obtient les autres arcs.

*Remarque :*

On constate que les valuations de tous les arcs reliant deux sommets distincts sont proportionnelles à  $dt$ . En outre, les valuations des boucles, reliant chaque sommet à lui-même, n'apportent pas d'information, puisque chaque boucle est simplement valuée par le complément à 1 de la somme des valuations des autres arcs issus du sommet en question. C'est pourquoi on préfère souvent représenter le "graphe de transition simplifié" du processus de Markov, obtenu :

- en omettant les boucles ;
- en valuant chaque arc par son "taux de probabilité", c'est-à-dire sa probabilité de transition divisée par  $dt$ .

Dans l'exemple considéré, le graphe de transition simplifié est le suivant :



On remarque que les arcs sont de 4 types :  $\uparrow$ ,  $\downarrow$ ,  $\rightarrow$  et  $\leftarrow$ , et que tous les arcs du même type sont valués par le même taux de probabilité.

### Régime permanent d'un processus de Markov

Un processus de Markov est **ergodique** si la probabilité d'observer le système dans chacun des états, à un instant  $t$  quelconque, tend vers une valeur donnée, indépendante de l'état initial, lorsque  $t$  tend vers l'infini.

Les **conditions suffisantes d'ergodicité** portant sur le graphe de transition sont les mêmes que pour une chaîne de Markov.

Lorsque l'ergodicité a lieu, le vecteur  $\pi_\infty$ , qui rassemble les probabilités de présence du système dans chacun des états en "régime permanent", peut être commodément calculé en exploitant le "**théorème des coupes**", qui s'énonce comme suit :

Etant donné un groupe quelconque de sommets du graphe de transition, isolé par une ligne fermée, dénommée "coupe", la probabilité en régime permanent de sortir, pendant  $dt$ , à l'extérieur de la coupe, est égale à la probabilité, pendant le même temps, d'y entrer.

L'image de la **sortie** de la coupe est explicite : elle correspond à l'occurrence d'une transition, depuis un état situé à l'intérieur de la coupe, vers un état situé à l'extérieur ; graphiquement, les entrées-sorties se font en empruntant les arcs du graphe de transition qui sont coupés par la coupe.

Chaque application du théorème des coupes procure une relation entre certaines composantes du vecteur  $\pi_\infty$ . Compte tenu de la relation de normalisation (somme des n composantes de  $\pi_\infty$  égale à 1), il faut appliquer  $(n - 1)$  fois le théorème, à  $(n - 1)$  coupes différentes, pour disposer d'un nombre d'équations suffisant pour l'extraction des n inconnues.

Le choix de ces  $(n - 1)$  coupes, a priori arbitraire, doit être guidé par la simplicité du système d'équations résultant.

**Exemple :**

Revenons au problème des robots dans le magasin de stockage.

On constate que le graphe de transition satisfait aux 3 conditions suffisantes d'ergodicité :

- il est fini (6 sommets) ;
- il est fortement connexe ;
- il comporte au moins une boucle (il a une boucle en chaque sommet).

Pour éviter d'avoir à résoudre un système de 6 équations à 6 inconnues, nous nous intéressons au cas particulier :  $p = q = \frac{1}{2}$  et  $\mu_1 = \mu_2 = \mu_3 = \mu$ . Dans ce cas, les lieux  $L_2$  et  $L_3$  sont

interchangeables, ce qui entraîne : 
$$\left| \begin{array}{l} \pi_\infty^{020} = \pi_\infty^{002} \\ \pi_\infty^{110} = \pi_\infty^{101} \end{array} \right.$$

Le problème ne comporte alors plus que 4 inconnues, notées comme suit :

$$\left| \begin{array}{l} a = \pi_\infty^{020} = \pi_\infty^{002} \\ b = \pi_\infty^{110} = \pi_\infty^{101} \\ c = \pi_\infty^{011} \\ d = \pi_\infty^{200} \end{array} \right.$$

Pour résoudre le problème, il faut appliquer 3 fois le théorème des coupes ; nous choisissons, par exemple, 3 coupes isolant chacune un seul sommet du graphe de transition, en l'occurrence 002, 011 et 200.

On obtient :

- théorème appliqué à la coupe 002 :  $a\mu_3 dt = bq\mu_1 dt$ , soit :  $a\mu_3 = bq\mu_1$
- théorème appliqué à la coupe 011 :  $c\mu_2 + c\mu_3 = bp\mu_1 + bq\mu_1$
- théorème appliqué à la coupe 200 :  $dq\mu_1 + dp\mu_1 = b\mu_3 + b\mu_2$
- relation de normalisation :  $2a + 2b + c + d = 1$

En tenant compte des hypothèses  $p = q = \frac{1}{2}$  et  $\mu_1 = \mu_2 = \mu_3 = \mu$ , le système d'équations

s'écrit :

$$\begin{cases} a = \frac{b}{2} \\ 2c = b \\ d = 2b \\ 2a + 2b + c + d = 1 \end{cases}$$

d'où l'on tire :

$$\pi_{\infty}^{020} = \pi_{\infty}^{002} = a = \frac{1}{11}$$

$$\pi_{\infty}^{110} = \pi_{\infty}^{101} = b = \frac{2}{11}$$

$$\pi_{\infty}^{011} = c = \frac{1}{11}$$

$$\pi_{\infty}^{200} = d = \frac{4}{11}$$

Comme pour les chaînes de Markov, l'ergodicité des processus de Markov leur confère une sorte de prévisibilité, qui présente un grand intérêt. Dans l'exemple considéré, on peut, par exemple, en tirer l'évaluation du nombre moyen  $n_i$  de robots se trouvant en  $L_i$ , en régime permanent :

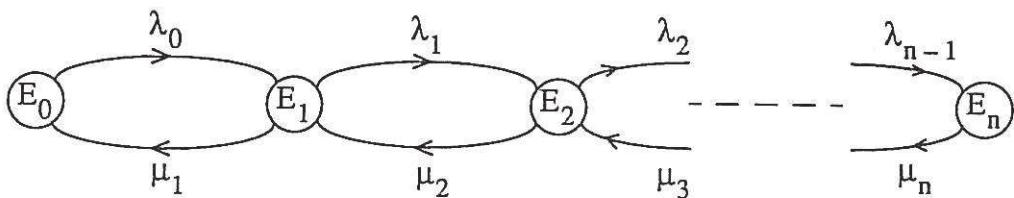
$$\begin{cases} n_1 = 2 \cdot \pi_{\infty}^{200} + 1 \cdot \pi_{\infty}^{110} + 1 \cdot \pi_{\infty}^{101} = \frac{12}{11} \\ n_2 = 1 \cdot \pi_{\infty}^{011} + 1 \cdot \pi_{\infty}^{110} + 2 \cdot \pi_{\infty}^{020} = \frac{5}{11} \\ n_3 = 1 \cdot \pi_{\infty}^{011} + 1 \cdot \pi_{\infty}^{101} + 2 \cdot \pi_{\infty}^{002} = \frac{5}{11} \end{cases}$$

On vérifie que l'on a :  $n_1 + n_2 + n_3 = 2$ , qui traduit la présence de deux robots dans le magasin.

## Processus de Markov particulier : le processus de naissance et de mort

Lorsque le système considéré est un ensemble de personnes situées en un lieu (par exemple toutes celles qui se trouvent à l'intérieur d'un bureau de poste), dont les arrivées et les départs sont régis par des lois de Poisson, les différents états possibles du système, notés  $E_k$ , sont définis par la présence de  $k$  personnes dans le lieu, et la dynamique du système peut être modélisée par un processus de Markov particulier, le "**processus de naissance et de mort**". En effet, le graphe de transition simplifié présente, dans ce cas, une caractéristique remarquable : chaque sommet  $E_k$ ,  $k > 1$ , n'est relié qu'à deux sommets,  $E_{k-1}$  et  $E_{k+1}$ , si bien que les sommets peuvent être disposés en ligne droite ; le premier sommet,  $E_0$ , et le dernier,  $E_n$  (qui n'existe que si le système est "**fermé**" : le lieu a une capacité d'accueil maximale fixée, soit  $n$  personnes), n'ont évidemment qu'un seul voisin.

Tous les processus de ce type ont donc un graphe de transition simplifié (définition donnée plus haut) de la même forme suivante :



(cas fermé :  $n + 1$  sommets)

La transition  $E_k \rightarrow E_{k+1}$ , valuée par le taux  $\lambda_k$ , est dénommée "naissance", et la transition  $E_k \rightarrow E_{k-1}$ , valuée par le taux  $\mu_k$ , est dénommée "mort".

On voit que les processus de naissance et de mort ne diffèrent les uns des autres que par le nombre de "maillons" du graphe de transition, et les valeurs numériques des valuations des arcs. Avec les notations ci-dessus, on peut donc établir, une fois pour toutes, des résultats généraux. La portée de ces résultats est accrue par le fait qu'un processus de naissance et de mort induit, généralement, une ou plusieurs "files d'attente" (dont l'étude fait l'objet du chapitre suivant).

Le principal résultat concerne l'ergodicité. Les conditions suffisantes classiques sont ici, visiblement, satisfaites. En notant  $\pi_{\infty}^k$  la composante du vecteur  $\pi_{\infty}$  relative à l'état  $E_k$ , le calcul aboutit au résultat suivant (admis) :

$$\boxed{\pi_{\infty}^k = \pi_{\infty}^0 \cdot \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{j=1}^k \mu_j}}, \quad k = 1, \dots, n$$

Il en résulte que :

$$\pi_{\infty}^k = \pi_{\infty}^{k-1} \cdot \frac{\lambda_{k-1}}{\mu_k}, \quad k = 1, \dots, n$$

Ces relations permettent d'exprimer  $\pi_{\infty}^k$ ,  $k = 1, \dots, n$ , en fonction de  $\pi_{\infty}^0$ .

En écrivant ensuite :  $\sum_{k=0}^n \pi_{\infty}^k = 1$ , on extrait la valeur numérique de  $\pi_{\infty}^0$ , d'où celles des autres composantes de  $\pi_{\infty}$ .

*Exemple :*

On se reportera à l'exercice figurant dans le fascicule "Exercices de Recherche Opérationnelle" : étude du service comptable d'une entreprise.

L'exercice montre comment le calcul du "régime permanent" permet, en particulier, d'évaluer précisément le temps moyen perdu en attente par les comptables souhaitant accéder à un terminal informatique. En exploitant le modèle de processus de naissance et de mort proposé, le chef d'entreprise pourra ainsi déterminer, chiffres en mains, le nombre de terminaux approprié.

\* \* \* \* \*

# CHAPITRE IX

## Files d'attente

---

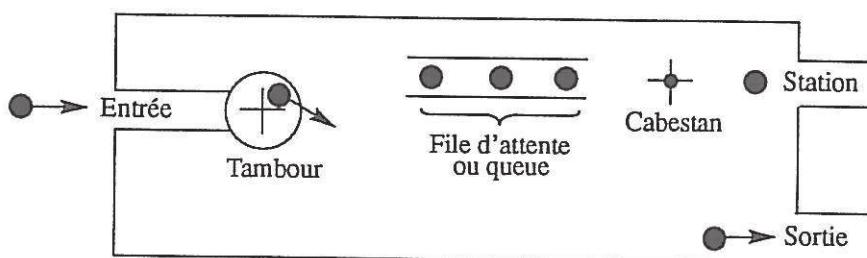
### Généralités sur les phénomènes d'attente

Les phénomènes d'attente sont hélas d'observation courante dans la vie quotidienne. Quand nous nous rendons à la poste, à la gare, à la banque, bien souvent nous devons "faire la queue" pour obtenir des timbres, un billet, de l'argent. On a l'habitude d'appeler **clients** les individus qui constituent la file d'attente et **station** le guichet où un **serveur** leur procure un service déterminé.

La file d'attente ne se manifeste pas toujours d'une manière physique : par exemple, si, dans un atelier, des machines tombées en panne "attendent" la visite et les soins du mécanicien-réparateur, elles ne se constituent pas en file. De même, les personnes qui sollicitent une communication par l'intermédiaire d'un standard téléphonique. Il est néanmoins facile, à chaque fois, de déterminer quels sont les clients, la ou les station (s), le ou les serveur (s).

Remarquons qu'il s'agit de phénomènes stochastiques. En effet, les clients arrivent en général au hasard ; d'autre part, la durée d'occupation de la station par chaque client n'est en général pas constante. Il suffit d'ailleurs que l'un de ces deux caractéristiques se présente comme un phénomène aléatoire pour qu'il s'agisse déjà d'un problème stochastique.

Considérons le cas très simple (figure ci-dessous) où il n'existe qu'une station, donc qu'une file d'attente (en l'absence de priorités). Chaque client pénètre dans l'enceinte où fonctionne la station en passant par un tambour : il se dirige aussitôt vers la file d'attente et y prend la dernière place : dès qu'un client a reçu le service qu'il attendait de la station, il s'en va ; aussitôt, le premier de ceux qui attendaient dans la file fait fonctionner le cabestan qui se trouve devant lui et entre en contact avec le serveur.



Cette matérialisation du passage des clients d'un état dans un autre est simplement utile pour préciser le vocabulaire employé dans la théorie des phénomènes d'attente.

On a, en effet, coutume d'appeler **file d'attente** l'ensemble des clients qui attendent d'être servis, à l'exclusion de celui qui est en train de se faire servir. On nomme **système d'attente** l'ensemble des clients qui font la queue, y compris celui qui se fait servir. Le **phénomène d'attente** s'étend à tous les clients possibles (dans le cas de systèmes bouclés, où les mêmes clients reviennent plus tard à l'entrée - par exemple les machines qui tombent en panne dans un atelier -, le nombre des clients est, en général, fini). Ces appellations se généralisent et prennent surtout leur intérêt dans les situations où coexistent plusieurs stations et plusieurs files d'attente.

Nous distinguerons donc deux types principaux de systèmes d'attente : le **système ouvert**, dans lequel le nombre de clients potentiels est très élevé (cas des guichets publics, des grands magasins, etc ...), au point qu'on peut considérer qu'il est illimité ; le **système d'attente fermé**, dans lequel le nombre de clients est limité (cas d'un atelier dans lequel existe un service de réparation de machines) ; dans les deux cas il peut y avoir une ou plusieurs files d'attente.

Dans la fourniture de services (aéroports, gares, ...), deux catégories de problèmes se posent :

- **les problèmes de conception** : dimensionnement de locaux, nombre de guichets, capacité de réseaux, ...
- **les problèmes d'exploitation** : élaboration de règles de fonctionnement, utilisation judicieuse des personnels, ...

Dans les deux cas, les simulations reposant sur la théorie des files d'attente permettent de reproduire la réalité, et de prendre des décisions.

### **Loi des arrivées. Loi des services.**

Revenons au phénomène d'attente le plus simple où il n'existe qu'une queue et qu'une station.

Le cas intéressant est celui où deux phénomènes aléatoires sont conjugués :

- 1°) les clients arrivent au hasard ;
- 2°) le temps passé à la station par chaque client est aléatoire.

Par exemple, les automobilistes arrivent au hasard à un parc (cela veut dire : l'heure d'arrivée de chacun est aléatoire) ; ils y stationnent un temps variable qui ne peut être prévu pour chacun d'eux (c'est encore une variable aléatoire).

Peut-on caractériser statistiquement la manière dont les clients arrivent au parc et la durée de leur stationnement ?

L'expérience montre que, dans beaucoup de phénomènes d'attente, les lois des arrivées et des services sont, respectivement, **poissonniennes** et **exponentielles**. Ce ne sont évidemment pas les seules formes que peuvent affecter ces lois, mais ce sont les plus fréquentes et aussi les plus simples à employer pour obtenir un exposé facile des principes et de la théorie des phénomènes d'attente.

### Notation de Kendall d'une file d'attente

On note A/B/S (N) (discipline), une file d'attente où A est le code de la loi des arrivées, B, celui de la loi des services ; S est le nombre de guichets (tous identiques), fonctionnant en parallèle ; N désigne le nombre maximal de clients admis dans le système (il est infini si cette indication est omise) ; puis on précise la discipline du service, sauf si elle est : "Premier arrivé, Premier servi" (en anglais FIFO : "first in first out"), auquel cas on l'omet.

Les principaux codes employés pour A et B sont :

- M pour des lois markoviennes, c'est-à-dire une loi de Poisson pour des arrivées et une loi exponentielle pour des services ;
- D (pour déterministe), pour des arrivées régulièrement espacées d'un temps  $1/\lambda$  ou pour des services de durée constante  $1/\mu$  ;
- G pour une loi générale (quelconque, incluant donc les précédentes).

Dans la suite du chapitre, nous traiterons seulement le modèle M/M/S, qui correspond à des arrivées selon une loi de Poisson, des durées de service de loi exponentielle, et l'existence de S serveurs. Le système est **ouvert** dans le cas strict M/M/S et **fermé** dans le cas M/M/S/N. Ce modèle joue un rôle prédominant, du fait de sa simplicité et de l'étendue de son champ d'application.

### File d'attente M/M/S

L'étude de cette file d'attente se ramène à celle du **processus de naissance et de mort** associé.

#### *Exemple :*

Un exemple est proposé dans le fascicule "Exercices de Recherche Opérationnelle" : étude d'un centre d'information comportant trois bureaux. La file d'attente analysée est une file M/M/3/5.

En appliquant les résultats du chapitre précédent, on peut calculer des grandeurs caractéristiques de la file d'attente, notamment - dans le cas ergodique - le **nombre moyen  $\bar{n}$  de clients en attente** et le **temps moyen  $\bar{t}_f$  d'attente d'un client** dans la file (il existe un abaque donnant  $\bar{t}_f$  en fonction du nombre  $S$  de guichets, du taux  $\lambda$  des arrivées des clients et du taux  $\mu$  de service de chaque guichet).

### *Cas particulier $S = 1$*

Les résultats sont plus simples, puisque le processus de naissance et de mort associé n'a alors qu'un taux de naissance ( $\lambda_i = \lambda, \forall i$ ) et un taux de mort ( $\mu_j = \mu, \forall j$ ).

Il convient de distinguer deux cas :

- *cas fermé : file M/M/1/N*

On retrouve le résultat du chapitre précédent ( $n = N$ ). L'ergodicité a lieu **sans condition**, et l'on a :

$$\left\{ \begin{array}{l} \pi_\infty^k = \pi_\infty^0 \cdot \left(\frac{\lambda}{\mu}\right)^k, \quad k = 1, \dots, N \\ \pi_\infty^0 = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{N+1}} \end{array} \right.$$

- *cas ouvert : file M/M/1 ( $\infty$ )*

L'ergodicité exige une condition : **le taux  $\lambda$  d'arrivée des clients doit être strictement inférieur au taux  $\mu$  de service**, pour éviter l'engorgement (qui se traduirait par une file d'attente de longueur infinie).

Dans le cas  $\boxed{\frac{\lambda}{\mu} < 1}$ , on obtient :

$$\left\{ \begin{array}{l} \pi_\infty^k = \pi_\infty^0 \cdot \left(\frac{\lambda}{\mu}\right)^k, \quad k = 1, 2, \dots \\ \pi_\infty^0 = 1 - \frac{\lambda}{\mu} \end{array} \right.$$

## IX (5)

Pour  $\bar{n}$  et  $\bar{t}_f$  - qui n'ont de sens que dans le cas  $\boxed{\frac{\lambda}{\mu} < 1}$  - on trouve :

$$\left\{ \begin{array}{l} \bar{n} = \frac{\lambda}{\mu - \lambda} \\ \bar{t}_f = \frac{\lambda}{\mu(\mu - \lambda)} \end{array} \right.$$

\* \* \* \* \*