



Processus stochastiques modélisation

Responsable UE : **Agnès Lagnoux** (lagnoux@univ-tlse2.fr)

Conception polycopié : **Claudie Hassenforder**

ISMAG

MASTER 2 - MI00451X

SOMMAIRE

INTRODUCTION	p01
Chapitre 1 : PROCESSUS DE MARKOV	
1.1 Généralités	p05
1.2 Chaînes de Markov à temps discret	p06
1.2.1 Matrice de transition et graphe d'une chaîne de Markov	p06
1.2.2 Exemples classiques de chaînes de Markov	p06
1.2.3 Classification des états	p07
1.2.4 Absorption par les classes récurrentes dans le cas fini	p10
1.2.5 Distribution stationnaire	p11
1.2.6 Comportement asymptotique	p15
1.3 Processus de Markov continus	p16
1.3.1 Régime transitoire	p16
1.3.2 Régime permanent	p18
Chapitre 2 : PROCESSUS DE POISSON	
2.1 Introduction	p19
2.2 Définitions et description du processus	p20
2.3 Caractérisation d'un processus par ses temps d'arrivée	p23
2.4 Propriétés supplémentaires	p24
2.4.1 Décomposition, superposition	p24
2.4.2 Processus de Poisson et loi binomiale	p25
2.4.3 Processus de Poisson et loi uniforme	p26
2.5 Processus de Poisson composés	p26
2.6 Processus non homogènes	p27
Chapitre 3 : PROCESSUS DE NAISSANCE ET DE MORT	
3.1 Etude générale	p29
3.1.1 Régime transitoire	p29
3.1.2 Régime permanent	p30
3.2 Étude de quelques cas particuliers	p30
3.2.1 Croissance pure, par immigration	p31
3.2.2 Croissance pure, par naissance	p31
3.2.3 Décroissance pure, par décès	p32
3.2.4 Processus de Yule-Ferry	p32
3.2.5 Modèle logistique, à taux non linéaires	p34
3.3 Problème de l'extinction de l'espèce	p35
Chapitre 4 : PROCESSUS DE RAMIFICATION	
4.1 Processus discret à 1 type	p40
4.2 Processus permanent à 1 type	p41
4.3 Processus discret à 2 types	p43
4.4 Processus permanent à 2 types	p44

Chapitre 5 : PREMIÈRES NOTIONS SUR LES FILES D'ATTENTE

5.1 Introduction	p47
5.2 La file simple	p47
5.2.1 Processus d'arrivée	p47
5.2.2 Temps de service	p48
5.2.3 Structure et discipline de la file	p48
5.2.4 Notation de Kendall	p49
5.2.5 Notion de classe de clients	p50
5.3 Les réseaux de files d'attente	p50
5.3.1 Les réseaux ouverts	p50
5.3.2 Les réseaux fermés	p51
5.3.3 Les réseaux multiclassés	p51
5.3.4 Les réseaux de files d'attente à capacité limitée	p52
5.3.5 Les réseaux de files d'attente ouverts à contrainte de population	p52
5.4 Quelques exemples de systèmes d'attente	p53
5.5 Paramètres de performances opérationnels	p53
5.5.1 Paramètres de performances en régime transitoire	p53
5.5.2 Paramètres de performances en régime stationnaire	p55
5.5.3 Stabilité	p56
5.5.4 Ergodicité	p56
5.5.5 La loi de Little	p58

Chapitre 6 : FILE D'ATTENTE UNIQUE

6.1 Files d'attente markoviennes	p60
6.1.1 Processus de naissance et de mort général	p60
3.1.2 La file $M/M/1$	p61
3.1.3 La file $M/M/1/K$	p63
3.1.4 La file $M/M/C$	p66
3.1.5 La file $M/M/\infty$	p68
3.2 Étude de la file $M/G/1$	p69
3.2.1 Introduction	p69
3.2.2 Analyse du régime permanent : méthode de la chaîne de Markov incluse	p70
3.2.3 Mise en oeuvre de l'analyse de la valeur moyenne	p72
3.3 La file $G/M/1$	p74
3.4 Extension à la file $G/G/1$	p76

Chapitre 7 : FIABILITÉ

7.1 Introduction	p79
7.1.1 Définitions	p79
7.1.2 Lois utilisées	p79
7.2 Systèmes non réparables	p80
7.2.1 Généralités	p80
7.2.2 Systèmes sans redondance	p80
7.2.3 Systèmes avec redondance	p80
7.3 Systèmes réparables	p81
7.3.1 Introduction	p81
7.3.2 Méthode des processus stochastiques	p81

Introduction

L'origine des études sur les phénomènes d'attente remonte aux années 1909-1920 avec les travaux de A.K. Erlang concernant le réseau téléphonique de Copenhague. La théorie mathématique s'est ensuite développée notamment grâce aux contributions de Palm, Kolmogorov, Khintchine, Pollaczek,... et fait actuellement toujours l'objet de nombreuses publications scientifiques. Cette théorie s'est ensuite étendue à de nombreux champs d'application comme la gestion de stocks, les télécommunications en général, la fiabilité de systèmes complexes,...

Les problèmes liés à l'attente dans un centre de service sont omniprésents dans notre société. Les exemples ne manquent pas :

- attente à un guichet (caisse dans un supermarché, administration),
- trafic urbain ou aérien,
- réseaux téléphoniques,
- circulation de pièces dans un atelier,
- programmes dans un système informatique,...

Il est devenu inconcevable de construire un système quelconque (que ce soit un système informatique, un réseau de communication, un système de production ou un système de la vie quotidienne) sans avoir auparavant fait d'analyse des performances. La pression des enjeux économiques est telle actuellement que l'on ne peut aboutir à un système sous-dimensionné et que l'on doit éviter au maximum le surdimensionnement. Construire un système adapté, respectant le plus possible les objectifs du cahier des charges est une démarche qui passe obligatoirement par une étape de modélisation et d'analyse des performances.

En plus des modélisations analytiques, les simulations sur calculateurs permettront des évaluations relativement précises, mais demandant parfois des temps de calcul qui peuvent être importants si l'on veut reproduire correctement les phénomènes aléatoires et avoir atteint un régime permanent.

Une condition nécessaire pour dimensionner un centre de service est qu'il soit capable d'absorber le débit moyen de clients prévu, condition très facile à vérifier par de simples calculs de débits moyens. Mais, même avec un système correctement dimensionné, le caractère aléatoire des arrivées et des temps de service rend les attentes impossibles à éviter complètement.

La théorie des processus aléatoires concerne l'étude mathématique de phénomènes physiques, biologiques ou économiques évoluant dans le temps, et dont l'évolution est de caractère aléatoire, c'est-à-dire non prévisible avec certitude.

Pour définir un processus aléatoire, il faut :

1- Un espace des temps T ($T \subset \mathbb{R}_+$)

Les deux espaces des temps les plus utilisés sont :

- $T = \mathbb{N}$: le processus est dit *discret* ; on regarde ce qu'il se passe à chaque unité de temps, ou bien on fait une suite d'opérations et on regarde ce qu'il se passe à chaque opération (ex : lancer d'une pièce).

- $T = \mathbb{R}_+$: le processus est dit *continu* : on garde les yeux fixés sur un système qui évolue dans le temps à partir d'un instant t_0 que l'on prend pour origine des temps ($t = 0$).

2- Un espace des états E

L'ensemble E peut être :

- discret : c'est-à-dire fini ou dénombrable. Il sera, dans ce cas, souvent pratique d'identifier E avec une partie de \mathbb{N} ou de \mathbb{Z} .
- non discret : par exemple $E = \mathbb{R}$ ou $E \subset \mathbb{R}^2$ (partie du plan) ou $E \subset \mathbb{R}^3$ (partie de l'espace)

3- Une famille de variables aléatoires $(X_t)_{t \in T}$.

Ces variables aléatoires sont toutes définies sur un même espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans l'espace des états E .

Ainsi, à chaque instant $t \in T$, on associe, non pas une valeur déterministe (comme dans le calcul d'une trajectoire mécanique) mais une valeur aléatoire décrite par une variable aléatoire X_t à valeurs dans E .

La variable aléatoire X_t peut représenter les résultats d'essais successifs comme par exemple, le jet d'une pièce à pile ou face, ou des observations successives sur une caractéristiques d'une population.

Le processus aléatoire est la famille de variables aléatoires $(X_t)_{t \in T}$.

Un processus aléatoire est une généralisation d'un vecteur aléatoire. Comme dans le cas du vecteur aléatoire, la connaissance de la loi de X_t pour tout $t \in T$ est loin de caractériser le processus.

En particulier, elle ne donne aucune information sur le passage de t à $t + \Delta t$ et donc, sur l'évolution du processus.

Ce qui jouera le plus grand rôle dans l'étude des processus aléatoires, ce sont les probabilités de transition.

Si B_1 et B_2 sont des parties de E , on note $P([X_{t+\Delta t} \in B_2] | [X_t \in B_1])$ la probabilité de transition de B_1 à B_2 entre t et $t + \Delta t$ (c'est-à-dire la probabilité d'être dans B_2 à l'instant $t + \Delta t$ sachant qu'on était dans B_1 à l'instant t).

Les éléments principaux qui différencient les processus aléatoires généraux sont l'espace des états E , l'espace des temps T et les relations de dépendance entre les X_t .

4- Quelques relations de dépendance.

- Processus de comptage :

Un processus de comptage $(N_t)_{t \in T}$, où $T \subset \mathbb{R}$, est un processus croissant (si $s \leq t$, alors

$N_s \leq N_t$), à valeurs dans $E = \mathbb{N}$.

• Processus à accroissements indépendants :

Un processus croissant $(X_t)_{t \geq 0}$ est dit à accroissements indépendants si pour tout $n \in \mathbb{N}^*$, et pour tous t_1, \dots, t_n tels que $t_1 < t_2 < \dots < t_n$, les accroissements $X_{t_1} - X_0, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont des variables aléatoires indépendantes.

• Processus homogène dans le temps :

Le processus $(X_t)_{t \geq 0}$ est dit homogène, si pour tout t et pour tout s , la loi de $X_{t+s} - X_s$ ne dépend pas de s .

• Processus de Markov :

Le processus $(X_t)_{t \geq 0}$ est dit de Markov, si pour tout $n \in \mathbb{N}$, pour tous t_1, \dots, t_n, t_{n+1} tels que $t_1 < t_2 < \dots < t_n < t_{n+1}$:

$$P([X_{t_{n+1}} = e_{n+1}] | [X_{t_1} = e_1] \cap \dots \cap [X_{t_n} = e_n]) = P([X_{t_{n+1}} = e_{n+1}] | [X_{t_n} = e_n])$$

c'est-à-dire que seul le passé le plus proche est pris en compte pour déterminer les probabilités d'occupation de chaque état.

Exemples de processus aléatoires

a) Signal télégraphique

Ce processus, utilisé en théorie de la communication, rend compte de l'état d'occupation d'une ligne.

A_k : instant du début de la $k^{\text{ième}}$ communication ;

D_k : instant de la fin de la $k^{\text{ième}}$ communication.

$$T = \mathbb{R}_+ ; E = \{-1, 1\}.$$

$X_t = 1$ si la ligne est libre, c'est-à-dire s'il existe k tel que $D_k \leq t < A_{k+1}$;

$X_t = -1$ si la ligne est occupée, c'est-à-dire s'il existe k tel que $A_k \leq t < D_k$.

On a là, un processus (dit à créneaux), markovien, à accroissements indépendants, homogène dans le temps.

Problèmes qui se posent :

- Trouver la loi de X_t ;
- Calculer $P([X_t = e_j] | [X_s = e_i])$ pour $s \leq t$ et pour $e_i, e_j \in E$;
- Existe-t-il une loi limite de la loi de X_t quand t tend vers l'infini ?

b) Processus de ramification

Ce processus est utilisé pour suivre l'évolution de certaines populations animales ou cellulaires, ou bien d'un nom chez les humains.

Chaque individu génère, indépendamment des autres, un nombre aléatoire de descendants.

X_n est le nombre d'individus total à la $n^{\text{ième}}$ génération.

$$E = \mathbb{N} ; T = \mathbb{N}.$$

Le processus est markovien.

Problèmes qui se posent :

- Déterminer la loi de X_n ;
- Trouver une relation entre X_n et X_{n+1} ;
- Existe-t-il une probabilité non nulle d'extinction de l'espèce ?

c) **Files d'attente**

Ces processus sont utilisés en recherche opérationnelle.

Des clients se présentent à des guichets à des temps aléatoires, pour y recevoir des services de durée aléatoire (ex : banque, magasin, péage d'autoroute...)

X_t est le nombre de clients en attente à l'instant t .

$$E = \mathbb{N} ; T = \mathbb{R}_+.$$

Problèmes qui se posent :

- Existence d'un régime stationnaire, nombre moyen de clients en attente, durée moyenne de l'attente d'un client ;
- Détermination du nombre minimal de guichets assurant un bon écoulement de la file.

d) **Mouvement Brownien**

Dans un gaz, chaque particule est soumise aux impacts incessants de ses voisines. Les collisions provoquent des déplacements.

X_t est la position d'une particule donnée à l'instant t .

$X_t - X_s$ est le déplacement pendant $[s, t[$: c'est la somme d'un grand nombre de petits déplacements indépendants et il sera légitime de supposer que sa loi est une loi Normale.

$$E \subset \mathbb{R} ; T = \mathbb{R}_+.$$

Le processus est à accroissements indépendants, homogène dans le temps.

Ce cours a pour objectif de donner quelques éléments de théorie sur les processus aléatoires généraux : processus de Markov, de Poisson, de naissance et de mort, puis sur les systèmes d'attentes : files uniques et réseaux, avec entre temps une application à la fiabilité.

Chapitre 1

Processus de Markov

1.1 Généralités

Le processus de Markov fournit un outil simple de modélisation d'une classe particulière de systèmes à espace d'états discret. L'analyse des processus de Markov est un préliminaire nécessaire à l'étude des systèmes de files d'attente.

Définition : Le processus $(X_t)_{t \geq 0}$ est dit de Markov, si

a) **axiome de Markov** : pour tous $t_1 < t_2 < \dots < t_n < t_{n+1}$, pour tous x_1, \dots, x_{n+1} :

$$P([X_{t_{n+1}} = x_{n+1}] | [X_{t_1} = x_1] \cap \dots \cap [X_{t_n} = x_n]) = P([X_{t_{n+1}} = x_{n+1}] | [X_{t_n} = x_n])$$

b) **axiome d'homogénéité** : pour tous s et t , pour tous $x, y \in E$, $P([X_{t+s} = y] | [X_s = x])$ ne dépend que de t (et non des instants s et $t + s$).

L'axiome de Markov traduit que la probabilité de n'importe quel comportement futur, le présent étant connu, n'est pas modifié par toute connaissance supplémentaire du passé.

Notations : On pose $p_{x,y}(t) = P([X_{t+s} = y] | [X_s = x]) = P([X_t = y] | [X_0 = x])$ et $P(t) = (p_{x,y}(t))_{x,y \in E}$. On pose aussi $\vec{\pi}(t) = P_{X_t}$ (vecteur ligne de composantes $\pi_x(t) = P([X_t = x])$).

Propriétés :

a) $P(t)$ est une matrice stochastique, *i.e.* $p_{x,y}(t) \geq 0$ et $\sum_y p_{x,y}(t) = 1$ pour tout x .

b) Pour tout s et pour tout t , $P(s+t) = P(s)P(t)$.

c) Pour tout s et pour tout t , $\vec{\pi}(s+t) = \vec{\pi}(s)P(t)$.

Preuve :

a) $p_{x,y}(t) \in [0, 1]$ car c'est une probabilité. De plus, la ligne x correspond à la loi $P_{X_0}^{[X_0=x]}$ et on a bien

$$\sum_y p_{x,y}(t) = \sum_y P^{[X_0=x]}([X_t = y]) = 1.$$

b) Pour calculer $p_{x,y}(t+s) = P^{[X_0=x]}([X_{t+s} = y])$, on va faire intervenir les différents états pouvant être occupés à l'instant t :

$$\begin{aligned} P^{[X_0=x]}([X_{t+s} = y]) &= \frac{P([X_0 = x] \cap [X_{t+s} = y])}{P([X_0 = x])} \\ &= \sum_{z \in E} \frac{P([X_0 = x] \cap [X_t = z] \cap [X_{t+s} = y])}{P([X_0 = x])} \\ &= \sum_{z \in E} \frac{P([X_{t+s} = y] | [X_0 = x] \cap [X_t = z]) P([X_0 = x] \cap [X_t = z])}{P([X_0 = x])} \\ &= \sum_{z \in E} P([X_{t+s} = y] | [X_t = z]) P([X_t = z] | [X_0 = x]) \end{aligned}$$

d'après l'axiome de Markov (X_0 n'apporte rien de plus que X_t pour déterminer la loi de X_{t+s}). Ainsi, on a

$$p_{x,y}(t+s) = \sum_{z \in E} p_{x,z}(t) p_{z,y}(s),$$

qui est le coefficient (x, y) de la matrice produit $P(t)P(s)$.

$$c) \pi_y(t+s) = P([X_{t+s} = y]) = \sum_{x \in E} P([X_{t+s} = y] | [X_t = x]) P([X_t = x]) \text{ c'est-à-dire}$$

$$\pi_y(t+s) = \sum_{x \in E} \pi_x(t) p_{x,y}(s).$$

□

1.2 Chaînes de Markov à temps discret

Pour cette classe particulière de processus, on ne s'intéresse à l'état du système qu'en des instants particuliers t_n de leur évolution. Cela peut se produire dans deux cas :

→ Soit on s'intéresse à l'état du système à intervalles de temps réguliers comme par exemple tous les jours ou toutes les heures. On a alors $t_n = n\tau$, où τ est l'unité de temps considérée. C'est le cas le plus simple à comprendre. L'état du système à l'étape n du processus est alors l'état du système au n -ième jour où à la n -ième heure.

→ Soit on s'intéresse à l'état du système juste après un événement : t_n est alors l'instant du n -ième événement. L'état du système à l'étape n du processus correspond alors à l'état juste après le n -ième changement d'état. Les instants effectifs de changements d'états ne sont alors plus équidistants.

Pour simplifier, on traitera ici le premier cas et on prendra $T = \mathbb{N}$.

1.2.1 Matrice de transition et graphe d'une chaîne de Markov

On notera P la matrice $P(1)$ définie précédemment (et $p_{x,y} = p_{x,y}(1)$). Comme, pour tous entiers m et n , on a $P(n+m) = P(n)P(m)$, on a en particulier, pour tout n , $P(n+1) = P(n)P(1) = P(n)P$ et donc, comme de plus, $P(0) = I = P^0$

$$\boxed{P(n) = P^n \text{ et } \vec{\pi}(n) = \vec{\pi}(0)P^n \text{ pour tout } n \in \mathbb{N}.}$$

Ainsi, la seule donnée de la matrice de transition P et de $\vec{\pi}(0)$ suffit à déterminer la loi de X_n pour tout $n \in \mathbb{N}$.

La matrice de transition P caractérise la chaîne et se prête bien aux calculs mais, pour mieux visualiser les transitions entre états, il est souvent utile de faire un graphe de la chaîne, équivalent à la donnée de P où :

- les états sont représentés par des points ;
- une probabilité de transition $p_{x,y} > 0$ est représentée par un arc orienté de x à y au dessus duquel est notée la valeur de $p_{x,y}$.

1.2.2 Exemples classiques de chaînes de Markov

a) Chemin aléatoire à une dimension

i) Un individu qui a bu fait, à chaque instant, de façon aléatoire, un pas en avant ou un pas en arrière. On a un graphe "en boudins" et une matrice de transition tridiagonale : $E = \mathbb{Z}$, $p_{i,i+1} = p_{i,i-1} = \frac{1}{2}$, $p_{i,j} = 0$ si $|i-j| \neq 1$.

ii) Ce même individu est maintenant dans un couloir fermé de chaque côté par une porte ($E = \{0, \dots, N\}$). Lorsqu'il heurte une porte, il est assomé : 0 et N sont des états absorbants (une fois qu'on y est, on y reste).

iii) Même cas de figure que ii) mais quand l'individu heurte une porte, celle-ci, au lieu de l'assomer, le renvoie un pas en arrière.

b) Chaîne à 2 états

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

On suppose que $\vec{\pi}(0) = (1, 0)$. Que vaut $\vec{\pi}(n)$?

On a $\vec{\pi}(n) = \vec{\pi}(0)P^n$ et pour calculer les puissances d'une matrice, on étudie ses éléments propres.

→ On sait que $\lambda_1 = 1$ est toujours valeur propre de P (associée au vecteur propre ${}^t(1, \dots, 1)$ car $\sum_{y \in E} p_{x,y} = 1$).

Pour trouver ici l'autre valeur propre, on utilise $\lambda_1 + \lambda_2 = \text{tr}P = 2 - \alpha - \beta$: $\lambda_2 = 1 - \alpha - \beta \neq 1$ si $P \neq I$: P est donc diagonalisable.

→ Recherche des vecteurs propres : pour $\lambda_2 = 1 - \alpha - \beta$, $(1 - \alpha)x_1 + \alpha x_2 = (1 - \alpha - \beta)x_1$ conduit à $\beta x_1 + \alpha x_2 = 0$.

→ On a donc, si $\Omega = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}$, alors $\Omega^{-1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix}$ et

$$\vec{\pi}(n) = \vec{\pi}(0)\Omega \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \Omega^{-1} = \left(\frac{\beta + \alpha(1 - \alpha - \beta)^n}{\alpha + \beta}, \frac{\alpha - \alpha(1 - \alpha - \beta)^n}{\alpha + \beta} \right).$$

1.2.3 Classification des états

Définition : Soit x et y deux états. On dit que x mène à y s'il existe $n \in \mathbb{N}$ tel que $p_{x,y}(n) > 0$. On dit que x et y communiquent si x mène à y et si y mène à x .

Remarque : Pour $n \geq 1$, $p_{x,y}(n) = \sum_{x_1, \dots, x_{n-1}} p_{x,x_1} p_{x_1,x_2} \cdots p_{x_{n-1},y}$ donc, si $p_{x,y}(n) > 0$, il existe x_1, \dots, x_{n-1} tels que $p_{x,x_1} > 0$, $p_{x_1,x_2} > 0, \dots, p_{x_{n-1},y} > 0$, c'est-à-dire que l'on peut trouver sur le graphe un "chemin" de x à y si $x \neq y$ (il peut même y en avoir plusieurs!). Attention toutefois, $p_{x,x}(0) = 1$, donc x mène toujours à x même s'il n'y a pas de chemin de x vers lui-même.

La relation de communication est une relation d'équivalence ; une chaîne à une seule classe est dite irréductible.

Définition : Un état x est dit non essentiel s'il existe un état y et $n \in \mathbb{N}$ tels que $p_{x,y}(n) > 0$ et $p_{y,x}(m) = 0$ pour tout $m \in \mathbb{N}$. Dans le cas contraire x est dit essentiel.

Les états d'une même classe de communication sont de même nature : on parlera de classe non essentielle (classe que l'on peut quitter) et classe essentielle (ou absorbante) (classe que l'on ne peut pas quitter).

Définition : La période $d(x)$ d'un état x est définie par :

$$d(x) = P.G.C.D.\{n \in \mathbb{N}^* ; p_{x,x}(n) > 0\}$$

(avec $d(x) = 0$ si pour tout $n \in \mathbb{N}^*$, $p_{x,x}(n) = 0$). Si $d(x) = 1$, x est dit apériodique.

Propriété : Tous les états d'une même classe de communication ont même période.

Preuve : Si x et x' communiquent, il existe k et l tels que $p_{x,x'}(k) > 0$ et $p_{x',x}(l) > 0$. Alors $p_{x,x}(k+l) > 0$.

Soit m tel que $p_{x',x'}(m) > 0$. Alors $p_{x,x}(k+m+l) > 0$. Donc $d(x)$ divise $k+l$ et $d(x)$ divise $k+m+l$; donc $d(x)$ divise m et ceci, pour tout m tel que $p_{x',x'}(m) > 0$, donc $d(x)$ divise $d(x')$. Comme x et x' jouent le même rôle, $d(x) = d(x')$.

□

Remarque : Une classe dont un élément admet une boucle, (c'est-à-dire $p_{x,x} > 0$), est obligatoirement apériodique (c'est-à-dire de période 1), mais ce n'est pas une condition nécessaire.

Exemple de chaîne périodique : marche aléatoire sur \mathbb{Z} .

Théorème : Si x et y sont dans une même classe de communication de période d , si $p_{x,y}(n) > 0$ et si $p_{x,y}(m) > 0$, alors d divise $m - n$.

Preuve : Comme y mène à x , il existe k tel que $p_{y,x}(k) > 0$. On a donc $p_{x,x}(m+k) > 0$ et $p_{x,x}(n+k) > 0$. Donc d divise $m+k$ et $n+k$, il divise donc la différence.

□

Ainsi, on peut partitionner les classes périodiques de la façon suivante :

Définition : Soit C une classe de communication de période d , et soit x_0 fixé dans C . On définit, pour $k \in \{0, 1, \dots, d-1\}$,

$$C'_k = \{y \in E ; \text{ si } p_{x_0,y}(n) > 0 \text{ alors } n \equiv k[d]\}.$$

Les C'_k sont appelées sous-classes cycliques de C .

Remarque : Les C'_k n'ont pas forcément le même nombre d'éléments, mais elles sont toutes non vides.

$$\text{On pose } T_x = \begin{cases} \min\{n \geq 1 ; X_n = x\} & \text{si il existe } m \geq 1 \text{ tel que } X_m = x \\ +\infty & \text{si } X_m \neq x \text{ pour tout } m \geq 1 \end{cases}.$$

Définition : Un état x est dit récurrent si $P^{[X_0=x]}([T_x < +\infty]) = 1$, c'est-à-dire, si partant de x , on repasse presque sûrement en x . Dans le cas contraire, x est dit transitoire.

On pose $f_{x,y}^{(n)} = P^{[X_0=x]}([T_y = n])$ pour $n \geq 1$ (et par convention $f_{x,y}^{(0)} = 0$) et

$$\mu_x = \mathbb{E}^{[X_0=x]}(T_x) = \sum_{n \geq 1} n f_{x,x}^{(n)}.$$

On a alors x récurrent si et seulement si $\sum_{n \geq 1} f_{x,x}^{(n)} = 1$, et μ_x représente le temps moyen de retour en x : il peut être infini, même si x est récurrent. On est donc conduit à une classification plus fine des états récurrents.

Définition : Un état récurrent est dit récurrent positif si $\mu_x < +\infty$. Dans le cas contraire, il est dit récurrent nul.

Propriété : Pour $n \geq 1$, on a $p_{x,y}(n) = \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k)$.

Preuve : Le processus passe de x à y en n étapes si et seulement s'il passe de x à y pour la première fois en k étapes ($0 \leq k \leq n$) et s'il passe ensuite de y à y en les $n-k$ étapes suivantes. Ces chemins, pour des k distincts, sont disjoints, et la probabilité d'un chemin pour k fixé est $f_{x,y}^{(k)} p_{y,y}(n-k)$.

Ce raisonnement intuitif peut être rendu rigoureux de la façon suivante.

Comme, pour $n \geq 1$,

$$[X_n = y] = \bigcup_{k=1}^{n-1} ([T_y = k] \cap [X_n = y]) \cup [T_y = n],$$

on en déduit, les $[T_y = k]$ étant disjoints,

$$\begin{aligned} P([X_n = y]/[X_0 = x]) &= \sum_{k=1}^{n-1} P([T_y = k] \cap [X_n = y]/[X_0 = x]) + f_{x,y}^{(n)} \\ &= \sum_{k=1}^{n-1} P([T_y = k]/[X_0 = x]) P([X_n = y]/[T_y = k] \cap [X_0 = x]) + f_{x,y}^{(n)} \end{aligned}$$

Or, pour $1 \leq k \leq n-1$, $[T_y = k] \cap [X_0 = x]$ est de la forme $A \cap [X_k = y]$ où A ne dépend que de X_0, \dots, X_{k-1} . Par conséquent,

$$P([X_n = y]/[T_y = k] \cap [X_0 = x]) = P([X_n = y]/A \cap [X_k = y]) = P([X_n = y]/[X_k = y]) = p_{y,y}(n-k).$$

Comme $P([X_n = y]/[X_0 = x]) = p_{x,y}(n)$ et $P([T_y = k]/[X_0 = x]) = f_{x,y}^{(k)}$, il en résulte

$$p_{x,y}(n) = \sum_{k=1}^{n-1} f_{x,y}^{(k)} p_{y,y}(n-k) + f_{x,y}^{(n)} = \sum_{k=1}^n f_{x,y}^{(k)} p_{y,y}(n-k) = \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k)$$

avec la convention $f_{x,y}^{(0)} = 0$.

□

Théorème (critère de récurrence) :

1) Un état y est récurrent si et seulement si $\sum_{n=0}^{+\infty} p_{y,y}(n) = +\infty$. On a alors $\sum_{n=0}^{+\infty} p_{x,y}(n) = +\infty$ pour tout x qui mène à y .

2) Un état y est transitoire si et seulement si $\sum_{n=0}^{+\infty} p_{y,y}(n) < +\infty$.

On a alors $\sum_{n=0}^{+\infty} p_{x,y}(n) < +\infty$ pour tout $x \in E$.

Preuve : On considère les séries entières $F_{x,y}(s) = \sum_{n=0}^{+\infty} s^n f_{x,y}^{(n)}$ et $P_{x,y}(s) = \sum_{n=0}^{+\infty} s^n p_{x,y}(n)$.

On établit que $P_{x,y}(s) = \delta_{x,y} + F_{x,y}(s)P_{y,y}(s)$.

En effet, $P_{x,y}(s) = \sum_{n=0}^{+\infty} s^n p_{x,y}(n) = p_{x,y}(0) + \sum_{n=1}^{+\infty} s^n \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k)$ avec $p_{x,y}(0) = \delta_{x,y}(0)$. De plus $f_{x,y}^{(0)} p_{y,y}(0) = 0$, donc on a bien

$$\sum_{n=1}^{+\infty} s^n \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k) = \sum_{n=0}^{+\infty} s^n \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k) = F_{x,y}(s)P_{y,y}(s)$$

d'après la définition du produit de deux séries.

En appliquant ceci à $x = y$, on obtient $P_{y,y}(s) = 1 + F_{y,y}(s)P_{y,y}(s)$, soit $P_{y,y}(s) = \frac{1}{1 - F_{y,y}(s)}$.

- Si y est récurrent, $f_{y,y} = 1$, soit, par le lemme d'Abel, $\lim_{s \rightarrow 1^-} F_{y,y}(s) = 1$ et donc $\lim_{s \rightarrow 1^-} P_{y,y}(s) = +\infty$.
- Si $x \neq y$ et si x mène à y , alors $F_{x,y}(1) \neq 0$, et, comme $P_{x,y}(s) = F_{x,y}(s)P_{y,y}(s)$, si $\lim_{s \rightarrow 1^-} P_{y,y}(s) = +\infty$, alors $\lim_{s \rightarrow 1^-} P_{x,y}(s) = +\infty$.
- Si y est transitoire, alors $\lim_{s \rightarrow 1^-} P_{y,y}(s) < +\infty$ et, comme on a $F_{x,y}(1) \leq 1$, on a $\lim_{s \rightarrow 1^-} P_{x,y}(s) < +\infty$, c'est-à-dire $\sum_{n=0}^{+\infty} p_{x,y}(n) < +\infty$.

□

Conséquence : Si y est transitoire, alors $\sum_{n=0}^{+\infty} p_{x,y}(n) < +\infty$ pour tout x , et, en particulier, $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$, mais la réciproque est fautive en général.

Propriété : Les états d'une même classe de communication sont, soit tous récurrents, soit tous transitoires.

Preuve : Si x et x' communiquent, il existe k et l tels que $p_{x,x'}(k) > 0$ et $p_{x',x}(l) > 0$. On a donc $p_{x,x}(k+n+l) \geq p_{x,x'}(k)p_{x',x'}(n)p_{x',x}(l)$ et donc $p_{x,x}(1) \geq p_{x,x'}(k)p_{x',x'}(1)p_{x',x}(l)$. D'où, si x' est récurrent, alors x l'est aussi, et, comme x et x' jouent le même rôle, si x est récurrent, alors x' l'est aussi.

□

Cas particulier des chaînes finies : Si E est fini, la classification se simplifie beaucoup.

Propriété : Si E est fini, essentiel équivaut à récurrent et il existe au moins un état récurrent.

Preuve :

- Si tous les états étaient transitoires, on aurait $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$ pour tout y et alors, comme la somme ne comporte qu'un nombre fini de termes, $\lim_{n \rightarrow +\infty} \sum_{y \in E} p_{x,y}(n) = \sum_{y \in E} \lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$, ce qui contredit $\sum_{y \in E} p_{x,y}(n) = 1$.

- Si x est non essentiel, la probabilité de ne pas revenir en x est non nulle, donc $f_{x,x} \neq 1$ et x est transitoire. Donc, récurrent implique toujours essentiel (même si E est infini).

- Si x est essentiel, la chaîne réduite à la classe de communication de x est une chaîne finie si E est fini, et admet donc au moins un état récurrent. Comme tous les états d'une même classe sont de même nature, x est bien récurrent.

□

1.2.4 Absorption par les classes récurrentes dans le cas fini

Si R_1, \dots, R_k sont les classes récurrentes de (X_n) et \mathcal{T} l'ensemble de ses états transitoires, on numérote les états de \mathcal{T} de 1 à t ; pour $j \in \mathcal{T}$, on note $a_j^{(n)}(R_i)$ la probabilité que, partant de l'état j , on arrive dans la classe R_i à la $n^{\text{ième}}$ étape et on pose $a_j(R_i) = \sum_{n=1}^{+\infty} a_j^{(n)}(R_i)$ (probabilité d'absorption de l'état transitoire j par la classe récurrente R_i). On pose $a^{(n)}(R_i) =$

$\begin{pmatrix} a_1^{(n)}(R_i) \\ \vdots \\ a_t^{(n)}(R_i) \end{pmatrix}$ et $a(R_i) = \sum_{n=1}^{+\infty} a^{(n)}(R_i)$. Enfin, on note $P_{\mathcal{T}}$ la matrice extraite de P en ne gardant que les lignes et les colonnes correspondant aux états transitoires.

Propriété : On a $(I - P_{\mathcal{T}})a(R_i) = a^{(1)}(R_i)$ avec $a_j^{(1)}(R_i) = \sum_{x \in R_i} p_{j,x}$.

Preuve : Pour $n \geq 2$, $a_j^{(n)}(R_i) = \sum_{x \notin R_i} p_{j,x} a_x^{(n-1)}(R_i)$.

Mais, si $x \in R_k$, avec $k \neq i$, alors $a_x^{(n-1)}(R_i) = 0$, car lorsqu'on est dans une classe récurrente, on y reste.

On a donc :

$$\begin{aligned} a_j(R_i) &= \sum_{n \geq 1} a_j^{(n)}(R_i) = a_j^{(1)}(R_i) + \sum_{n \geq 2} \sum_{x \in T} p_{j,x} a_x^{(n-1)}(R_i) \\ &= a_j^{(1)}(R_i) + \sum_{x \in T} p_{j,x} \sum_{n \geq 2} a_x^{(n-1)}(R_i) \\ &= a_j^{(1)}(R_i) + \sum_{x \in T} p_{j,x} a_x(R_i) \end{aligned}$$

ce qui donne la ligne j de la propriété. □

1.2.5 Distribution stationnaire

Définition : Une famille $\vec{\pi} = (\pi_x)_{x \in E}$ est dite distribution stationnaire d'une chaîne, de matrice de transition P , si c'est une probabilité qui vérifie $\vec{\pi} = \vec{\pi}P$.

Remarques :

- Si $\vec{\pi}(0) = \vec{\pi}$, alors, $\vec{\pi}(n) = \vec{\pi}$ pour tout n (preuve par récurrence sur n : étant donné que $\vec{\pi}(n+1) = \vec{\pi}(n)P$, si $\vec{\pi}(n) = \vec{\pi}$, alors $\vec{\pi}(n+1) = \vec{\pi}P = \vec{\pi}$.)

- Si (X_n) converge en loi, $\lim_{n \rightarrow +\infty} \vec{\pi}(n)$ est une probabilité notée $\vec{\pi}^{(\infty)}$ qui est distribution stationnaire de la chaîne. En effet, on a $\vec{\pi}(n+1) = \vec{\pi}(n)P$ pour tout $n \in \mathbb{N}$ et, en prenant la limite quand $n \rightarrow +\infty$, on obtient bien $\vec{\pi}^{(\infty)} = \vec{\pi}^{(\infty)}P$.

- $\vec{\pi}$ peut ne pas exister : si $E = \mathbb{N}$ et $p_{k,k+1} = 1$ pour tout $k \in \mathbb{N}$, $\vec{\pi} = \vec{\pi}P$ conduit à $\pi_0 = 0$ et $\pi_k = \pi_{k-1}$ pour tout $k \in \mathbb{N}^*$: on aurait alors $\pi_k = 0$ pour tout $k \in \mathbb{N}$ et $\vec{\pi}$ ne peut pas être une probabilité.

- Une chaîne peut admettre une infinité de distributions stationnaires.

Interprétations :

- $\pi_y = \sum_{x \in E} \pi_x p_{x,y}$ s'écrit aussi $\sum_{x \in E} \pi_y p_{y,x} = \sum_{x \in E} \pi_x p_{x,y}$ (car $\sum_{x \in E} p_{y,x} = 1$). On peut interpréter $\pi_y p_{y,x}$ comme le nombre moyen de transitions de y vers x par unité de temps et $\sum_{x \in E} \pi_y p_{y,x}$ s'interprète alors comme le flux moyen de sortie de l'état y . De même, $\sum_{x \in E} \pi_x p_{x,y}$ est le flux moyen d'entrée dans l'état y . Ainsi, en régime permanent :

pour tout état y , il y a égalité entre le flux sortant de y et le flux entrant dans y .

- π_y peut s'interpréter comme la proportion de temps passé dans l'état y .

Propriété : Si $\vec{\pi}$ est distribution stationnaire d'une chaîne irréductible, alors $\pi_x > 0$ pour tout $x \in E$.

Preuve : On a $\vec{\pi} = \vec{\pi}P^n$ pour tout $n \in \mathbb{N}$, donc $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$ pour tout état y .

Supposons qu'il existe y tel que $\pi_y = 0$. Alors, comme $\pi_x p_{x,y}(n) \geq 0$, on a, pour tout $x \in E$ et pour tout $n \in \mathbb{N}$, $\pi_x p_{x,y}(n) = 0$.

Or, la chaîne étant irréductible, il existe n tel que $\pi_x p_{x,y}(n) > 0$; d'où $\pi_x = 0$, et ceci, pour tout x . On a alors $\vec{\pi} = \vec{0}$, ce qui est impossible, puisque $\vec{\pi}$ est une probabilité. □

Si x est récurrent positif, on pose, pour tout $y \in E$, $\rho_x(y) = \mathbb{E}^{[X_0=x]} \left(\sum_{n \geq 0} \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right)$: c'est le nombre moyen de visites dans l'état y entre deux visites dans l'état x .

Propriété : Si x est un état récurrent positif, alors $\vec{\pi} = \left(\frac{\rho_x(y)}{\mu_x} \right)_{y \in E}$ est distribution stationnaire.

Preuve : On a

$$\begin{aligned} \rho_x(y) &= \mathbb{E}^{[X_0=x]} \left(\sum_{n \geq 0} \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right) = \sum_{n \geq 0} \mathbb{E}^{[X_0=x]} (\mathbb{I}_{[X_n=y] \cap [T_x > n]}) \\ &= \sum_{n \geq 0} P^{[X_0=x]} ([X_n = y] \cap [T_x > n]) \end{aligned}$$

Si $y \neq x$, on a $P^{[X_0=x]} ([X_0 = y] \cap [T_x > 0]) = 0$ et, pour $n \geq 1$, on a aussi $P^{[X_0=x]} ([X_n = y] \cap [T_x = n]) = 0$, de sorte que l'on a deux expressions pour $\rho_x(y)$:

$$\rho_x(y) = \sum_{n \geq 1} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) ; \quad (*)$$

$$\rho_x(y) = \sum_{n \geq 1} P^{[X_0=x]} ([X_n = y] \cap [T_x > n]).$$

D'autre part, $P^{[X_0=x]} ([X_0 = x] \cap [T_x > 0]) = 1$ et, pour $n \geq 1$, les probabilités $P^{[X_0=x]} ([X_n = x] \cap [T_x > n])$ sont nulles, de sorte que $\rho_x(x) = 1$. Comme x est récurrent, la relation $1 = f_{x,x} = \sum_{n \geq 1} P^{[X_0=x]} ([T_x = n])$ est vérifiée.

Or, elle peut encore s'écrire $1 = \rho_x(x) = \sum_{n \geq 1} P^{[X_0=x]} ([X_n = x] \cap [T_x > n - 1])$. Ainsi, (*) est vérifiée pour tout $y \in E$. On la réécrit :

$$\begin{aligned} \rho_x(y) &= P^{[X_0=x]} ([X_1 = y] \cap [T_x > 0]) + \sum_{n \geq 2} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{n \geq 2} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]). \end{aligned}$$

Maintenant, pour $n \geq 2$,

$$P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) = \sum_{z \neq x} P^{[X_0=x]} ([X_{n-1} = z] \cap [X_n = y] \cap [T_x > n - 1]) = K$$

avec $K = \sum_{z \neq x} P^{[X_0=x]} ([X_n = y]/[X_{n-1} = z] \cap [T_x > n - 1]) P^{[X_0=x]} ([X_{n-1} = z] \cap [T_x > n - 1])$.

Or,

$$[T_x > n - 1] \cap [X_{n-1} = z] = [X_1 \neq x] \cap \dots \cap [X_{n-1} \neq x] \cap [X_{n-1} = z] = \dots \cap [X_{n-2} \neq x] \cap [X_{n-1} = z],$$

et, par l'axiome de Markov

$$P^{[X_0=x]} ([X_n = y]/[X_{n-1} = z] \cap [T_x > n - 1]) = P^{[X_0=x]} ([X_n = y]/[X_{n-1} = z]).$$

Ainsi,

$$\begin{aligned} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) &= \sum_{z \neq x} P^{[X_0=x]} ([X_n = y]/[X_{n-1} = z]) P^{[X_0=x]} P([X_{n-1} = z] \cap [T_x > n - 1]) \\ &= \sum_{z \neq x} p_{z,y} P^{[X_0=x]} P([X_{n-1} = z] \cap [T_x > n - 1]). \end{aligned}$$

Par suite,

$$\begin{aligned} \rho_x(y) &= p_{x,y} + \sum_{n \geq 2} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{n \geq 2} \sum_{z \neq x} p_{z,y} P^{[X_0=x]} ([X_{n-1} = z] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{z \neq x} p_{z,y} \sum_{n \geq 2} P^{[X_0=x]} ([X_{n-1} = z] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{z \neq x} p_{z,y} \sum_{n \geq 1} P^{[X_0=x]} ([X_n = z] \cap [T_x > n]) \\ &= p_{x,y} + \sum_{z \neq x} p_{z,y} \rho_x(z) = \sum_z \rho_x(z) p_{z,y}. \end{aligned}$$

D'autre part, on a

$$\begin{aligned} \sum_y \rho_x(y) &= \sum_y \mathbb{E}^{[X_0=x]} \left(\sum_{n \geq 0} \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right) = \mathbb{E}^{[X_0=x]} \left(\sum_{n \geq 0} \sum_y \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right) \\ &= \mathbb{E}^{[X_0=x]} \left(\sum_{n \geq 0} \mathbb{I}_{[T_x > n]} \right) = \sum_{n \geq 0} P^{[X_0=x]} ([T_x > n]) = \mathbb{E}^{[X_0=x]} (T_x) = \mu_x. \end{aligned}$$

Si x est récurrent positif, alors $\mu_x < +\infty$ et on peut poser $\pi_y = \frac{\rho_x(y)}{\mu_x}$. Il en résulte bien que $\vec{\pi}$ est distribution stationnaire. □

Théorème :

- 1) Une chaîne irréductible admet une distribution stationnaire $\vec{\pi}$ si et seulement si tous ses états sont récurrents positifs. Dans ce cas, $\vec{\pi}$ est unique et $\pi_x = 1/\mu_x$ pour tout $x \in E$.
- 2) Une chaîne quelconque admet une distribution stationnaire si et seulement si elle possède une classe récurrente positive. Dans ce cas, si $\vec{\pi}$ est une distribution stationnaire et si $(R_i)_i$ sont les classes récurrentes positives, il existe des réels positifs λ_i de somme 1 tels que :

$$\pi_x = 0 \text{ si } x \text{ n'est pas récurrent positif et } \pi_x = \lambda_i/\mu_x \text{ si } x \in R_i.$$

Preuve :

1) • On montre facilement que l'existence d'une distribution stationnaire implique la récurrence de la chaîne. En effet, on a alors $\vec{\pi} = \vec{\pi} P^n$ pour tout $n \in \mathbb{N}$, d'où $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$ pour tout $y \in E$. Si tous les états étaient transitoires, on aurait, pour tout $x \in E$, $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$.

Comme $p_{x,y}(n) \leq 1$ et que $\sum_x \pi_x$ converge, on peut appliquer le théorème de convergence dominée de Lebesgue, c'est-à-dire intervertir limite et somme.

On a alors $\pi_y = \lim_{n \rightarrow +\infty} \sum_{x \in E} \pi_x p_{x,y}(n) = \sum_{x \in E} \lim_{n \rightarrow +\infty} p_{x,y}^{(n)} = 0$, ce qui est contradictoire.

• On montre ensuite que, si $\vec{\pi}$ est une distribution stationnaire, que l'on prend comme distribution initiale, alors $\pi_y \mu_y = 1$ pour tout état y .

En effet, $\mu_y = \mathbb{E}^{[X_0=y]}(T_y) = \sum_{n \geq 0} P^{[X_0=y]}([T_y > n]) = \sum_{n \geq 1} P^{[X_0=y]}([T_y \geq n])$; d'où

$$\pi_y \mu_y = \sum_{n \geq 1} P^{[X_0=y]}([T_y \geq n]) P([X_0 = y]) = \sum_{n \geq 1} P([X_0 = y] \cap [T_y \geq n]).$$

On pose $a_n = P([X_m \neq y \text{ pour } 0 \leq m \leq n])$. On a $P([X_0 = y] \cap [T_y \geq 1]) = P([X_0 = y]) = 1 - a_0$ et, pour $n \geq 2$,

$$\begin{aligned} P([X_0 = y] \cap [T_y > n - 1]) &= P([X_0 = y] \cap [X_m \neq y \text{ pour } 1 \leq m \leq n - 1]) \\ &= P([X_m \neq y \text{ pour } 1 \leq m \leq n - 1]) - P([X_m \neq y \text{ pour } 0 \leq m \leq n - 1]) \\ &= P([X_m \neq y \text{ pour } 0 \leq m \leq n - 2]) - P([X_m \neq y \text{ pour } 0 \leq m \leq n - 1]) \\ &= a_{n-2} - a_{n-1} \end{aligned}$$

d'après l'axiome d'homogénéité. Donc $\mu_y \pi_y = \sum_{n \geq 1} P([X_0 = y] \cap [T_y \geq n]) = 1 - \lim_{n \rightarrow +\infty} a_n$ et $\lim_{n \rightarrow +\infty} a_n = P([X_m \neq y \text{ pour tout } m \geq 0]) = 0$ car y est récurrent. Donc $\mu_y \pi_y = 1$ et, comme $\pi_y > 0$, on a $\mu_y < +\infty$ et $\pi_y = \frac{1}{\mu_y}$.

On a ainsi montré que tout état y est récurrent positif et que la distribution $\vec{\pi}$ est unique.

• Réciproquement, si les états sont récurrents positifs, on a une distribution stationnaire donnée par la propriété précédente, et, d'après ce que l'on vient de voir, cette distribution est donc unique et vérifie $\pi_y = \frac{1}{\mu_y}$ pour tout état y .

2) Si $\vec{\pi}$ est distribution stationnaire, alors $\vec{\pi} = \vec{\pi} P^n$ pour tout $n \in \mathbb{N}$. D'où $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$ pour tout $y \in E$.

• Si y est transitoire, on vérifie exactement comme dans 1) que $\pi_y = 0$.

• Si y est récurrent, alors $\pi_y = \sum_{x \in E} \pi_x p_{x,y} = \sum_{x \in \text{Cl}(y)} \pi_x p_{x,y}$ car si x est transitoire, $\pi_x = 0$ et si x est récurrent, dans une autre classe, $p_{x,y} = 0$.

En posant $C = \text{Cl}(y)$ et $\vec{\pi}(C) = \sum_{x \in \text{Cl}(y)} \pi_x$, on a :

→ soit $\vec{\pi}(C) = 0$ et alors $\pi_x = 0$ pour tout $x \in C$;

→ soit $\vec{\pi}(C) \neq 0$ et alors $\vec{\pi}_C = \left(\frac{\pi_x}{\pi(C)} \right)_{x \in C}$ est distribution stationnaire de la chaîne irréductible réduite

à C . Par conséquent, d'après l'unicité vue au 1), on a $\pi_x = \frac{\vec{\pi}(C)}{\mu_x}$ et x est récurrent positif. De plus, on a $\sum_C \vec{\pi}(C) = 1$.

• Réciproquement, $\sum_{x \in R_k} \frac{1}{\mu_x} = 1$ et $\sum_k \sum_{x \in R_k} \frac{\lambda_k}{\mu_x} = \sum_k \lambda_k = 1$.

De plus, $\pi_y = \sum_{x \in E} \pi_x p_{x,y} = \sum_{x \in \text{Cl}(y)} \pi_x p_{x,y}$.

□

Conséquences :

1) $\vec{\pi}$ n'est pas unique s'il existe plusieurs classes récurrentes positives.

2) Les états d'une même classe de communication sont, soit tous transitoires, soit tous récurrents positifs, soit tous récurrents nuls.

En effet, soit x récurrent positif. Alors x est essentiel et la chaîne réduite à la classe de communication de x est irréductible et admet une distribution stationnaire puisqu'elle possède un état récurrent positif. Donc, tous les états de la classe de communication de x sont récurrents positifs.

3) Si la chaîne est irréductible, le nombre moyen de visites à y entre deux passages par x vérifie $\rho_x(y) = \frac{\mu_x}{\mu_y}$.

En effet, $\vec{\pi}$ définie par $\pi_y = \frac{\rho_x(y)}{\mu_x}$ est distribution stationnaire et d'après l'unicité, $\pi_y = \frac{1}{\mu_y}$.

1.2.6 Comportement asymptotique

La loi conditionnelle de X_n sachant $[X_0 = x]$ est représentée par la famille $(p_{x,y}(n))_{y \in E}$. Pour avoir la convergence en loi, il est nécessaire que $\lim_{n \rightarrow +\infty} p_{x,y}(n)$ existe pour tout y mais, en général, ceci n'est pas suffisant car il n'est pas évident que la famille limite soit une probabilité. C'est pourquoi, on introduit la notion plus restrictive suivante :

Définition : On dit que la chaîne $(X_n)_{n \geq 0}$ de matrice de transition P admet une *distribution limite* si, pour tout $(x, y) \in E^2$, $p_{x,y}(n)$ admet une limite l_y indépendante de x , quand n tend vers $+\infty$, la famille $\vec{l} = (l_y)_{y \in E}$ étant non identiquement nulle. La famille $\vec{l} = (l_y)_{y \in E}$ est alors la distribution limite.

Remarque : Il n'est pas évident, *a priori*, que si \vec{l} existe, elle mérite le nom de distribution. Il est clair que $l_y \geq 0$, mais de $\sum_{y \in E} p_{x,y}(n) = 1$, on déduit *a priori* seulement que $\sum_{y \in E} l_y \leq 1$. Pour pouvoir aller plus loin, il va falloir admettre le théorème suivant, dit théorème ergodique :

Théorème ergodique : Si y est aperiodique ou non récurrent positif, alors $\lim_{n \rightarrow +\infty} p_{x,y}(n) = \frac{f_{x,y}}{\mu_y}$ pour tout $x \in E$.

Conséquences :

- En particulier, si y est non récurrent positif, alors $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$.
- Si $x = y$ aperiodique ou non récurrent positif, alors $\lim_{n \rightarrow +\infty} p_{y,y}(n) = \frac{1}{\mu_y}$. En effet, si y est récurrent, alors $f_{y,y} = 1$ et si y est non récurrent, alors $\mu_y = +\infty$ et comme $f_{y,y} \leq 1$, on a $\frac{f_{y,y}}{\mu_y} = \frac{1}{\mu_y} = 0$.

Remarque : Si y est récurrent positif de période $d \neq 1$, $\lim_{n \rightarrow +\infty} p_{x,y}(n)$ peut ne pas exister.

Propriété : Si \vec{l} est une distribution limite, alors \vec{l} est l'unique distribution stationnaire de la chaîne.

Preuve : Soit y tel que $\lim_{n \rightarrow +\infty} p_{x,y}(n) = l_y \neq 0$. Alors, nécessairement, y est récurrent positif, d'après le théorème ergodique. Il existe donc au moins une classe récurrente positive. Mais il ne peut pas en exister plusieurs car sinon, en prenant x dans une autre classe récurrente positive, on aurait $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$. Il existe donc une unique distribution stationnaire $\vec{\pi}$ avec $\pi_x = \frac{1}{\mu_x}$ pour tout état récurrent positif, et $\pi_x = 0$ sinon.

On a donc, pour tout état y , $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$ et, comme $p_{x,y}(n) \leq 1$ et que $\sum_{x \in E} \pi_x$ converge, on peut appliquer le théorème de convergence dominée de Lebesgue, c'est-à-dire intervertir limite et somme.

On a $\pi_y = \lim_{n \rightarrow +\infty} \sum_{x \in E} \pi_x p_{x,y}(n) = \sum_{x \in E} \pi_x \lim_{n \rightarrow +\infty} p_{x,y}(n) = l_y \sum_{x \in E} \pi_x = l_y$, donc $\vec{l} = \vec{\pi}$.

En particulier, on en déduit que \vec{l} est une loi de probabilité. □

Théorème : 1) La distribution limite existe si et seulement si la suite $(X_n)_{n \geq 0}$ converge en loi, la loi limite n'étant pas fonction de la loi initiale. La limite en loi est alors la distribution limite, qui est aussi l'unique distribution stationnaire de la chaîne.

2) La distribution limite existe si et seulement si il existe une seule classe récurrente apériodique C telle que, pour tout $x \in E$ et pour tout $y \in C$, la probabilité que, partant de x , la chaîne passe au moins une fois par y soit égale à 1.

Preuve : 1) résulte de la définition de la distribution limite et de la propriété précédente.

2) résulte essentiellement du théorème ergodique et du fait que, si y est récurrent périodique, $\lim_{n \rightarrow +\infty} p_{x,y}(n)$ n'existe pas. □

Remarque : Une chaîne peut converger en loi sans qu'il existe de distribution limite, mais, dans ce cas, la limite en loi dépend de la loi initiale : cela se produit s'il existe plusieurs classes récurrentes positives ou bien s'il existe une classe récurrente positive de période $d \neq 1$.

Pour terminer, on admettra le théorème suivant :

Définition : Une chaîne récurrente positive, de période d , admet une unique distribution stationnaire $(\pi_y)_{y \in E}$ telle que, si x et y sont dans la même sous-classe cyclique :

$$\lim_{n \rightarrow +\infty} p_{x,y}(nd) = d\pi_y.$$

1.3 Processus de Markov continus

1.3.1 Régime transitoire

Contrairement à ce qui se passe pour les chaînes de Markov à temps discret, on ne dispose pas ici d'un historique complet du processus : on observe celui-ci à certains instants dans le temps, choisis aussi nombreux que l'on veut et répartis comme on veut mais la notion d'"unité de temps" n'a plus de sens ici et la matrice $P = P(1)$ ne permet pas de déterminer $P(t)$ pour tout t .

L'idée est alors de considérer $P(h)$ lorsque $h \rightarrow 0$.

Grâce à $P(t+h) = P(t)P(h) = P(h)P(t)$ et à $P(0) = I$, on a :

$$\frac{P(t+h) - P(t)}{h} = P(t) \frac{P(h) - I}{h} = \frac{P(h) - I}{h} P(t).$$

Sous réserve d'existence des limites, si on pose $A = \lim_{h \rightarrow 0} \frac{P(h) - I}{h} = P'(0)$, on a alors :

$$P'(t) = P(t)A = AP(t) \text{ et } P'(0) = I.$$

Cette équation différentielle matricielle admet l'unique solution :

$$P(t) = e^{tA} = \sum_{n=0}^{+\infty} \frac{t^n}{n!} A^n.$$

Ainsi, on détermine A à partir de la famille $(P(t))_{t \geq 0}$, ($A = P'(0)$), mais réciproquement, la seule connaissance de A permet de retrouver tous les $P(t)$ ($P(t) = e^{At}$).

La matrice A est appelée le générateur infinitesimal du processus.

On pose $A = (a_{x,y})_{x,y \in E}$. Ainsi, $a_{x,y} = \lim_{h \rightarrow 0} \frac{p_{x,y}(h) - \delta_{x,y}}{h}$.

- Si $x \neq y$, $a_{x,y} \geq 0$ et $p_{x,y}(h) = a_{x,y}h + o(h)$
- Si $x = y$, $a_{x,x} \leq 0$ et $p_{x,x}(h) = 1 + a_{x,x}h + o(h)$.

On a ici, $\left[\text{pour tout } x \in E, \sum_{y \in E} a_{x,y} = 0 \right]$ (la somme sur chaque ligne de A vaut 0).

Propriété : Pour tout état $x \in E$, le temps passé dans x avant de le quitter suit la loi exponentielle $\mathcal{E}(\lambda_x)$ avec $\lambda_x = -a_{x,x} = \sum_{y \neq x} a_{x,y}$.

Preuve : Si T_x désigne ici le temps passé en x avant de le quitter, on a :

$$P([T_x > t+h]) = P([T_x > t+h]/[T_x > t])P([T_x > t])$$

Mais, d'après l'homogénéité dans le temps, on a $P([T_x > t+h]/[T_x > t]) = P([T_x > h])$. Ainsi, si on pose $G_x(t) = P([T_x > t])$, on obtient :

$$G_x(t+h) = G_x(t)G_x(h) \text{ avec } G_x(h) = p_{x,x}(h) + o(h) = 1 + a_{x,x}h + o(h),$$

donc $\frac{G_x(t+h) - G_x(t)}{h} = G_x(t) \left(a_{x,x} + \frac{o(h)}{h} \right)$ et, en faisant $h \rightarrow 0$, il vient :

$$G'_x(t) = a_{x,x}G_x(t) \text{ avec } G_x(0) = 1,$$

soit $G_x(t) = e^{a_{x,x}t}$ et $\left[P([T_x \leq t]) = 1 - e^{a_{x,x}t} \right]$.

□

L'évolution d'un processus de Markov à temps continu peut se voir comme une répétition de deux phases :

- on reste un certain temps (de loi exponentielle) dans un état ;
- lorsqu'on quitte cet état, on choisit l'état vers lequel on sort, cette destination ne dépendant ni du temps passé dans l'état, ni du chemin par lequel on est arrivé à l'état. On notera $p_{x,y}$ la probabilité de se rendre dans l'état y en quittant l'état x .

Propriété : Pour tout $(x, y) \in E^2$ tel que $x \neq y$, $p_{x,y} = \frac{a_{x,y}}{\sum_{z \neq x} a_{x,z}}$ (et $p_{x,x} = 0$).

Preuve : On a, si $\lambda_x = -a_{x,x} = \sum_{z \neq x} a_{x,z}$,

$$P([X_{t+h} \neq x]/[X_t = x]) = \lambda_x h + o(h)$$

et, pour $y \neq x$,

$$P([X_{t+h} = y]/[X_t = x]) = a_{x,y}h + o(h) = p_{x,y}\lambda_x h + o(h)$$

d'après la définition de $p_{x,y}$. C'est donc bien que $p_{x,y} = \frac{a_{x,y}}{\lambda_x} = \frac{a_{x,y}}{\sum_{z \neq x} a_{x,z}}$.

□.

Définition : La chaîne de Markov discrète de matrice de transition $P = (p_{x,y})_{x,y \in E}$ est appelée *chaîne de Markov induite* du processus.

1.3.2 Régime permanent

On rappelle que, si $\vec{\pi}(t)$ est la loi de X_t , i.e. $\vec{\pi}(t) = (\pi_x(t))_{x \in E}$ où $\pi_x(t) = P([X_t = x])$, on a :

$$\vec{\pi}(t) = \vec{\pi}(0)P(t).$$

Définition : On dit que le processus $(X_t)_{t \geq 0}$ est *stationnaire* si la loi de X_t est indépendante de t .

On a alors $\vec{\pi}(t) = \vec{\pi}(0)$ pour tout $t \geq 0$ et en dérivant l'équation précédente, on obtient :

$$0 = \vec{\pi}(0)P'(t) = \vec{\pi}(0)P(t)A = \vec{\pi}(t)A = \vec{\pi}(0)A.$$

Définition : On appelle *distribution stationnaire* toute probabilité $\vec{\pi}$ qui vérifie

$$\vec{\pi}A = \vec{0}.$$

Si (X_t) converge en loi et si $\vec{\pi}^{(\infty)} = \lim_{t \rightarrow +\infty} \vec{\pi}(t)$, alors $\vec{\pi}^{(\infty)}$ est distribution stationnaire du processus.

On a $(\vec{\pi}A)_y = \sum_{x \in E} \pi_x a_{x,y} = \sum_{x \neq y} \pi_x a_{x,y} + \pi_y a_{y,y}$ avec $a_{y,y} = -\sum_{x \neq y} a_{y,x}$ ainsi :

$$\vec{\pi}A = \vec{0} \text{ équivaut à } \sum_{x \neq y} \pi_x a_{x,y} = \sum_{x \neq y} \pi_y a_{y,x}$$

Cette relation traduit l'équilibre, en régime stationnaire du flux rentrant en y ($\sum_{x \neq y} \pi_x a_{x,y}$) et du flux sortant de y ($\sum_{x \neq y} \pi_y a_{y,x}$).

Chapitre 2

Processus de Poisson

2.1 Introduction

De nombreux phénomènes aléatoires se manifestent par des “arrivées” survenant une par une à des instants aléatoires successifs.

Exemples :

- arrivées d’appels à un central téléphonique ;
- impacts de micrométéorites sur un satellite ;
- passage de véhicules à un péage d’autoroute ;
- arrivées de clients à un guichet, occurrence d’accidents dans une ville, pannes de machines dans une usine...

De tels phénomènes peuvent se définir par la famille $(A_n)_{n \in \mathbb{N}^*}$ des temps d’arrivées qui sont des variables aléatoires. Mais on peut aussi le faire à partir du processus de comptage $(N_t)_{t \in \mathbb{R}_+}$, ou par la famille $(T_n)_{n \in \mathbb{N}^*}$ des intervalles de temps entre deux arrivées.

N_t est le nombre d’événements apparus jusqu’à l’instant t .

$N_{t+u} - N_u$ est le nombre d’événements apparus entre u et $u + t$.

L’espace des états du processus $(N_t)_{t \in \mathbb{R}_+}$ est $E = \mathbb{N}$ et l’espace des temps est $T = \mathbb{R}_+$. Le processus qui modélise convenablement les exemples cités est le processus de Poisson. On conviendra que $N_0 = 0$.

On note :

- A_n l’instant de réalisation du $n^{\text{ième}}$ événement ;
- T_n la durée séparant le $(n-1)^{\text{ième}}$ événement du $n^{\text{ième}}$ événement pour $n \geq 2$ et $T_1 = A_1$.

On a :

- $A_n = T_1 + T_2 + \dots + T_n$ pour tout $n \in \mathbb{N}^*$;
- $T_1 = A_1$ et $T_n = A_n - A_{n-1}$ pour tout $n \geq 2$.

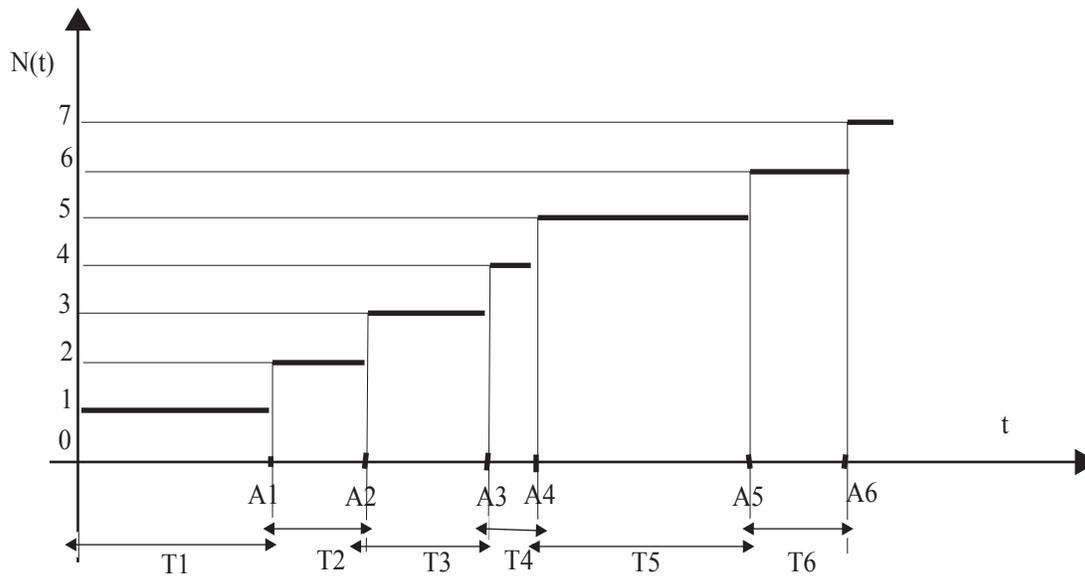
Ainsi, la connaissance de la famille $(A_n)_{n \in \mathbb{N}^*}$ équivaut à celle de la famille $(T_n)_{n \in \mathbb{N}^*}$.

D’autre part, $A_n \leq t$ signifie que le $n^{\text{ième}}$ événement a eu lieu à l’instant t ou avant, c’est-à-dire qu’à l’instant t , au moins n événements ont eu lieu, c’est-à-dire que $N_t \geq n$. Ainsi

$$F_{A_n}(t) = P([A_n \leq t]) = P([N_t \geq n]) = 1 - \sum_{k=0}^{n-1} P([N_t = k])$$

et

$$P([N_t = n]) = P([N_t \geq n]) - P([N_t \geq n+1]) = P([A_n \leq t]) - P([A_{n+1} \leq t]).$$



Par conséquent, la connaissance de $(N_t)_{t \in \mathbb{R}_+}$ équivaut à celle de la famille $(A_n)_{n \in \mathbb{N}^*}$.

2.2 Définition et description du processus

Soit $(N_t)_{t \in \mathbb{R}_+}$ un processus aléatoire à valeurs dans \mathbb{N} tel que $N_0 = 0$.

Définition : Le processus $(N_t)_{t \in \mathbb{R}_+}$ est appelé processus de comptage si c'est un processus croissant, c'est-à-dire si pour tout $s \leq t$, $N_s \leq N_t$. La variable aléatoire $N_t - N_s$ est alors appelée accroissement du processus sur $]s, t]$.

Définition : Un processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ est appelé processus à accroissements indépendants si pour tout $n \in \mathbb{N}^*$ et pour tous t_1, \dots, t_n tels que $t_1 < t_2 < \dots < t_n$, les accroissements $N_{t_1} - N_0, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ sont des variables aléatoires indépendantes.

Les processus vérifiant cette hypothèse sont assez nombreux : il semble en effet assez naturel que les arrivées se produisant dans des intervalles disjoints ne soient pas liées entre elles.

Définition : Le processus est dit stationnaire (ou homogène dans le temps), si pour tout s et pour tout t , l'accroissement $N_{t+s} - N_s$ a même loi que N_t .

Remarque : Cette propriété est semblable à l'axiome d'homogénéité pour les chaînes de Markov : seule la durée écoulée entre deux instants (et non pas les deux instants eux-mêmes) compte pour déterminer la loi de l'accroissement.

Si cette propriété semble relativement naturelle pour certains types d'arrivées, comme par exemple pour les pannes survenant dans un parc de machines qui fonctionnent en continu et toujours de la même façon, elle ne sera certainement pas réalisée pour des processus représentant des arrivées à "pics saisonniers" (par exemple le nombre de passages de véhicules au péage autoroutier de Bandol du 14 juillet au 15 août ne suit pas la même loi que celui des véhicules du

Définition : Un processus à accroissements indépendants stationnaire $(N_t)_{t \in \mathbb{R}_+}$ est dit à événements rares si $\lim_{h \rightarrow 0_+} P([N_h > 0]) = 0$ et si $\lim_{h \rightarrow 0_+} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$.

Remarque : La deuxième propriété traduit l'improbabilité d'arrivées simultanées. Si on pose $f(t) = P([N_t = 0])$ (et donc $f(0) = 1$), la première hypothèse traduit la continuité de f en 0, car $P([N_h > 0]) = 1 - f(h)$ et donc $\lim_{h \rightarrow 0_+} f(h) = f(0)$.

Définition : Un processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ tel que $N_0 = 0$ est un processus de Poisson si
 C1 : $(N_t)_{t \in \mathbb{R}_+}$ est stationnaire
 C2 : $(N_t)_{t \in \mathbb{R}_+}$ est un processus à accroissements indépendants ;
 C3 : $(N_t)_{t \in \mathbb{R}_+}$ est un processus à événements rares .

Le nom donné au processus de Poisson s'explique par ce qui suit :

Propriété : Un processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ tel que $N_0 = 0$ est un processus de Poisson si et seulement si :
 C1 : $(N_t)_{t \in \mathbb{R}_+}$ est stationnaire ;
 C2 : $(N_t)_{t \in \mathbb{R}_+}$ est un processus à accroissements indépendants ;
 C3' : il existe $\lambda > 0$ tel que, pour tout $t \geq 0$, la variable aléatoire N_t suit la loi de Poisson de paramètre λt .

Preuve : Un processus qui vérifie C3' vérifie également C3.

En effet, si on a C3', alors

$$P([N_h = 0]) = e^{-\lambda h} = 1 - \lambda h + o(h)$$

d'où $P([N_h > 0]) = \lambda h + o(h)$ et $\lim_{h \rightarrow 0} P([N_h > 0]) = 0$. De plus, $P([N_h = 1]) = e^{-\lambda h} \lambda h = \lambda h + o(h) \sim \lambda h$ et

$$P([N_h > 1]) = 1 - P([N_h = 0]) - P([N_h = 1]) = o(h)$$

et donc $\lim_{h \rightarrow 0_+} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$.

On va maintenant montrer que, pour un processus de Poisson, l'axiome C3' est vérifié.

$$P([N_{t+s} = k]) = \sum_{i=0}^k P([N_{t+s} = k] \cap [N_t = i]) \tag{1}$$

$$= \sum_{i=0}^k P([N_t = i] \cap [N_{t+s} - N_t = k - i]) \tag{2}$$

$$= \sum_{i=0}^k P([N_t = i]) P([N_{t+s} - N_t = k - i]) \tag{3}$$

$$= \sum_{i=0}^k P([N_t = i]) P([N_s = k - i]) \tag{4}$$

(On passe de (2) à (3) en utilisant C2 et de (3) à (4) en utilisant C1.)

On pose alors $P([N_t = k]) = \pi_k(t)$ et on va établir que $\pi_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ par récurrence sur k .

Étude pour $k = 0$: $\pi_0(t) = f(t)$ avec f qui vérifie

$$f(t + s) = f(t)f(s) \tag{*}$$

f est continue sur \mathbb{R}_+ car $\lim_{h \rightarrow 0} (f(t+h) - f(t)) = \lim_{h \rightarrow 0} f(t)(f(h) - 1) = 0$.

Or, les seules solutions continues de (*), à valeurs dans \mathbb{R}_+ , sont les fonctions de la forme $f(t) = e^{at}$.

(En effet, $\psi = \ln f$ vérifie $\psi(s+t) = \psi(t) + \psi(s)$. En particulier, $\psi(nt) = n\psi(t)$, puis, avec $t = \frac{1}{n}$, $\psi\left(\frac{1}{n}\right) = \frac{1}{n}\psi(1)$, puis $\psi\left(\frac{m}{n}\right) = \frac{m}{n}\psi(1)$. On a alors $\psi(r) = r\psi(1)$ pour tout $r \in \mathbb{Q}$ et on conclut que $\psi(x) = x\psi(1)$ par densité de \mathbb{Q} dans \mathbb{R} ($x = \lim r_n$) et grâce à la continuité de f ($f(x) = \lim f(r_n)$). Ainsi $f(x) = e^{\psi(x)} = e^{x\psi(1)}$.

Ici f est à valeurs dans $[0, 1]$, non identiquement nulle, donc $a \leq 0$. Le cas $a = 0$ n'a pas d'intérêt car alors on aurait $P([N_t = 0]) = 1$ pour tout t . Donc, $a < 0$ et il existe $\lambda = -a$ tel que

$$f(t) = \pi_0(t) = P([N_t = 0]) = e^{-\lambda t}.$$

Récurrence sur k : On suppose maintenant que $\pi_i(t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!}$ pour tout $t \geq 0$ et pour tout $i \leq k$ et on va montrer que $\pi_{k+1}(t) = e^{-\lambda t} \frac{(\lambda t)^{k+1}}{(k+1)!}$.

Pour cela on va établir une équation différentielle du premier ordre vérifiée par $\pi_{k+1}(t)$: on regarde combien d'arrivées se sont produites jusqu'à l'instant $t+h$ en faisant intervenir le nombre d'arrivées qui se sont produites jusqu'à l'instant t .

$$[N_{t+h} = k+1] = \bigcup_{i=0}^{k+1} [N_t = i] \cap [N_{t+h} - N_t = k+1-i]$$

et donc, toujours en utilisant C2 puis C1, il vient

$$P([N_{t+h} = k+1]) = \sum_{i=0}^{k+1} P([N_t = i])P([N_{t+h} - N_t = k+1-i]) = \sum_{i=0}^{k+1} P([N_t = i])P([N_h = k+1-i])$$

soit, avec la notation introduite ici

$$\pi_{k+1}(t+h) = \sum_{i=0}^{k+1} \pi_i(t)\pi_{k+1-i}(h)$$

Or, pour $i \leq k-1$, on a $k+1-i \geq 2$; d'où $P([N_h = k+1-i]) = \pi_{k+1-i}(h) = o(\pi_1(h))$ car $\lim_{h \rightarrow 0} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$ et

$$\pi_{k+1}(t+h) = \sum_{i=0}^{k-1} \pi_i(t)\pi_{k+1-i}(h) + \pi_1(h)\pi_k(t) + \pi_0(t)\pi_{k+1}(h) = \pi_1(h)\pi_k(t) + \pi_0(h)\pi_{k+1}(t) + o(\pi_1(h)).$$

Avec $\pi_0(h) = 1 - \pi_1(h) + o(\pi_1(h))$, on obtient alors :

$$\pi_{k+1}(t+h) - \pi_{k+1}(t) = \pi_1(h) [\pi_k(t) - \pi_{k+1}(t)] + o(h).$$

Donc $\frac{\pi_{k+1}(t+h) - \pi_{k+1}(t)}{h} = \frac{\pi_1(h)}{h} [\pi_k(t) - \pi_{k+1}(t)] + \frac{o(h)}{h}$ et

$$\pi'_{k+1}(t) = \lim_{h \rightarrow 0} \frac{\pi_1(h)}{h} [\pi_k(t) - \pi_{k+1}(t)].$$

Or $\pi_1(h) = 1 - \pi_0(h) + o(\pi_1(h))$ donc $\lim_{h \rightarrow 0} \frac{1 - \pi_0(h)}{\pi_1(h)} = 1 = \lim_{h \rightarrow 0} \frac{1 - \pi_0(h)}{h} \times \frac{h}{\pi_1(h)}$ et comme

$$\lim_{h \rightarrow 0} \frac{1 - \pi_0(h)}{h} = \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h}}{h} = \lambda,$$

c'est donc que $\lim_{h \rightarrow 0} \frac{\pi_1(h)}{h} = \lambda$, d'où $\pi'_{k+1}(t) = \lambda [\pi_k(t) - \pi_{k+1}(t)]$, soit :

$$\boxed{\pi'_{k+1}(t) + \lambda \pi_{k+1}(t) = \lambda \pi_k(t).}$$

Pour résoudre cette équation différentielle linéaire du premier ordre, on applique la méthode dite "de variation de la constante".

On a donc $\pi_{k+1}(t) = C(t)e^{-\lambda t}$ avec

$$C'(t)e^{-\lambda t} = \lambda\pi_k(t) = e^{-\lambda t} \frac{\lambda^{k+1}t^k}{k!}.$$

D'où $C'(t) = \frac{\lambda^{k+1}t^k}{k!}$ et

$$C(t) - C(0) = \int_0^t C'(u)du = \frac{\lambda^{k+1}}{k!} \int_0^t u^k du = \frac{\lambda^{k+1}}{k!} \left[\frac{u^{k+1}}{k+1} \right]_0^t = \frac{(\lambda t)^{k+1}}{(k+1)!}$$

avec $\pi_{k+1}(0) = C(0) = P([N_0 = k+1]) = 0$ car $N_0 = 0$. On a donc bien $\pi_{k+1}(t) = e^{-\lambda t} \frac{(\lambda t)^{k+1}}{(k+1)!}$. \square

Remarque :

Pour tous s et t et pour tous $i \leq k$:

$$\begin{aligned} P([N_{t+s} = k] \cap [N_s = i]) &= P([N_s = i] \cap [N_{t+s} - N_s = k - i]) \\ &= P([N_s = i])P([N_{t+s} - N_s = k - i]) \\ &= P([N_s = i])P([N_t = k - i]). \end{aligned}$$

Ainsi,

$$\begin{aligned} P([N_{t+s} = k]/[N_s = i]) &= \frac{P([N_s = i] \cap [N_{t+s} = k])}{P([N_s = i])} \\ &= P([N_t = k - i]) = e^{-\lambda t} \frac{(\lambda t)^{k-i}}{(k-i)!}. \end{aligned}$$

Ceci détermine les probabilités de transition de i à k entre les instants s et $t + s$.

2.3 Caractérisation d'un processus de Poisson par ses temps d'arrivée

Soit A_n l'instant de la $n^{\text{ième}}$ arrivée : $A_n = \inf\{t \geq 0 ; N_t = n\}$ et T_n le $n^{\text{ième}}$ temps d'attente pour $n \in \mathbb{N}^*$: $T_n = A_n - A_{n-1}$ (en convenant $A_0 = 0$).

On a $A_n = \sum_{i=1}^n T_i$ et $N_t = \max\{n \geq 0 ; A_n \leq t\}$.

Théorème : $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson de paramètre λ si et seulement si les variables aléatoires T_n sont indépendantes de même loi exponentielle $\mathcal{E}(\lambda)$ (de densité

$$f_{T_n}(t) = \lambda e^{-\lambda t} \mathbb{1}_{]0, +\infty[}(t)).$$

Preuve :

$$P([T_1 > t]) = P([N_t = 0]) = e^{-\lambda t} = 1 - F_1(t)$$

où F_1 est la fonction de répartition de T_1 . On a donc bien T_1 qui suit la loi exponentielle $\mathcal{E}(\lambda)$.

$$\begin{aligned} P^{[T_1=t_1]}([T_2 > t]) &= P([N_{t_1+t} = 1]/[N_{t_1} = 1] \cap [N_s = 0 \text{ pour tout } s < t_1]) \\ &= P([N_{t_1+t} - N_{t_1} = 0]/[N_{t_1} = 1] \cap [N_s = 0 \text{ pour tout } s < t_1]) \\ &= P([N_{t_1+t} - N_{t_1} = 0]) \end{aligned}$$

d'après l'indépendance des accroissements.

Or $P([N_{t_1+t} - N_{t_1} = 0]) = P([N_{t_1+t-t_1} = 0]) = P([N_t = 0])$ d'après la stationnarité ; et c'est aussi $e^{-\lambda t}$ car N_t suit la loi de Poisson $\mathcal{P}(\lambda t)$.

Donc T_2 est bien indépendante de T_1 et de même loi exponentielle $\mathcal{E}(\lambda)$.

De façon plus générale,

$$\begin{aligned} P([T_k > t] / [T_1 = t_1] \cap \dots \cap [T_{k-1} = t_{k-1}]) &= P([N_{t_{k-1}+t} - N_{t_{k-1}} = 0]) \\ &= P([N_t = 0]) = e^{-\lambda t}. \end{aligned}$$

Donc T_k est indépendante de T_1, \dots, T_{k-1} et de même loi exponentielle $\mathcal{E}(\lambda)$.

La réciproque sera admise. □

Conséquence : Les variables aléatoires A_n suivent la loi Gamma $\gamma(\lambda, n)$ (ou loi d'Erlang), de densité définie par

$$f_{A_n}(t) = \frac{\lambda^n}{(n-1)!} e^{-\lambda t} t^{n-1} \mathbb{I}_{]0, +\infty[}(t).$$

Propriété : Si $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson de paramètre λ , le temps aléatoire U qui sépare un instant θ du prochain événement et le temps aléatoire V qui sépare θ du dernier événement suivent la loi exponentielle $\mathcal{E}(\lambda)$.

Preuve :

$$P([U > x]) = P([N_{\theta+x} - N_\theta = 0]) = P([N_x = 0]) = e^{-\lambda x}$$

car $[U > x]$ signifie que pendant la durée x qui suit θ , il n'y a aucune arrivée. De même,

$$P([V > x]) = P([N_\theta - N_{\theta-x} = 0]) = P([N_x = 0]) = e^{-\lambda x}$$

car $[V > x]$ signifie que pendant la durée x qui précède θ , il n'y a eu aucune arrivée. □

Remarque : On a alors $\mathbb{E}(U + V) = \mathbb{E}(U) + \mathbb{E}(V) = \frac{2}{\lambda}$ alors que $\mathbb{E}(T_n) = \frac{1}{\lambda}$ pour tout $n \in \mathbb{N}^*$. C'est donc que $U + V = T_{N_\theta}$ n'a pas même loi que les T_n alors que sur $[N_\theta = n]$, on a $T_{N_\theta} = T_n$.

On peut terminer ce paragraphe en remarquant que $\mathbb{E}(N_1) = \lambda$ et $\mathbb{E}(T_n) = \frac{1}{\lambda}$. Ainsi, plus λ est grand, plus le nombre moyen d'arrivées par unité de temps est important, et plus l'intervalle entre 2 arrivées est court, ce qui semblait *a priori* évident. Pour cette raison, on appelle également le paramètre λ l'*intensité* du processus.

2.4 Propriétés supplémentaires

2.4.1 Décomposition, superposition

On peut décomposer en deux un processus de Poisson suivant un certain critère.

Exemples :

Compteur à fonctionnement aléatoire : Des particules arrivant vers un compteur forment un processus de Poisson de paramètre $\lambda = 6$ part/mn. Chaque particule a la probabilité $2/3$ d'être enregistrée par le compteur. Les particules enregistrées forment un processus de Poisson de paramètre $\lambda = 6 \times \frac{2}{3} = 4$ part/mn.

De même,

les voitures qui passent devant une station service forment un processus de Poisson de paramètre λ . Chaque voiture a la probabilité p de s'arrêter à la station. Alors, les voitures qui s'arrêtent à la station forment un processus de Poisson de paramètre λp et les voitures qui passent devant la station sans s'arrêter forment un processus de Poisson de paramètre $\lambda(1-p)$. De plus, ces deux processus sont indépendants.

Preuve : Ici, on va simplement montrer la propriété C3' pour le processus $(N_t^{(1)})_{t \geq 0}$ des voitures qui s'arrêtent à la station.

$$\begin{aligned}
 P([N_t^{(1)} = k]) &= \sum_{n=k}^{+\infty} P([N_t^{(1)} = k] / [N_t = n]) P([N_t = n]) \\
 &= \sum_{n=k}^{+\infty} C_n^k p^k (1-p)^{n-k} e^{-\lambda t} \frac{(\lambda t)^n}{n!} = e^{-\lambda t} \sum_{n=k}^{+\infty} p^k (1-p)^{n-k} \frac{(\lambda t)^n}{k!(n-k)!} \\
 &= e^{-\lambda t} \frac{p^k}{k!} \sum_{m=0}^{+\infty} \frac{(\lambda t)^{m+k} (1-p)^m}{m!} \quad (\text{en posant } m = n - k); \\
 &= e^{-\lambda t} \frac{(\lambda p t)^k}{k!} e^{\lambda(1-p)t} = e^{-\lambda p t} \frac{(\lambda p t)^k}{k!}.
 \end{aligned}$$

□

Inversement, on pourrait montrer que

la somme de deux processus de Poisson indépendants de paramètres respectifs λ_1 et λ_2 est un processus de Poisson de paramètre $\lambda_1 + \lambda_2$.

On a alors :

Propriété : Si $(N_t^{(1)})_{t \geq 0}$ et $(N_t^{(2)})_{t \geq 0}$ sont deux processus de Poisson indépendants de paramètres respectifs λ_1 et λ_2 , alors la loi conditionnelle de $N_t^{(1)}$ sachant $[N_t^{(1)} + N_t^{(2)} = n]$ est la loi binomiale $\mathcal{B}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$.

Preuve :

$$\begin{aligned}
 P([N_t^{(1)} = k] / [N_t^{(1)} + N_t^{(2)} = n]) &= \frac{P([N_t^{(1)} = k] \cap [N_t^{(2)} = n - k])}{P([N_t^{(1)} + N_t^{(2)} = n])} = \frac{P([N_t^{(1)} = k])P([N_t^{(2)} = n - k])}{P([N_t^{(1)} + N_t^{(2)} = n])} \\
 &= \frac{e^{-\lambda_1 t} \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_2 t} \frac{(\lambda_2 t)^{n-k}}{(n-k)!}}{e^{-(\lambda_1 + \lambda_2)t} \frac{((\lambda_1 + \lambda_2)t)^n}{n!}} = C_n^k \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n}.
 \end{aligned}$$

□

2.4.2 Processus de Poisson et loi binomiale.

Propriété : Pour $s \leq t$, la loi conditionnelle de N_s sachant $[N_t = n]$ est la loi binomiale $\mathcal{B}\left(n, \frac{s}{t}\right)$.

Preuve :

$$\begin{aligned}
P^{[N_t=n]}([N_s=k]) &= \frac{P([N_s=k] \cap [N_t=n])}{P([N_t=n])} = \frac{P([N_s=k] \cap [N_t - N_s = n - k])}{P([N_t=n])} \\
&= \frac{P([N_s=k])P([N_t - N_s = n - k])}{P([N_t=n])} = \frac{P([N_s=k])P([N_{t-s} = n - k])}{P([N_t=n])} \\
&= \frac{e^{-\lambda s} \frac{(\lambda s)^k}{k!} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{n-k}}{(n-k)!}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} = C_n^k \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}.
\end{aligned}$$

□

2.4.3 Processus de Poisson et loi uniforme

La loi de (A_1, A_2, \dots, A_n) sachant $[N_t = n]$ est celle de $(U_1^*, U_2^*, \dots, U_n^*)$; U_i^* étant la $i^{\text{ième}}$ plus petite valeur parmi les U_1, U_2, \dots, U_n , variables aléatoires indépendantes de même loi uniforme sur $[0, t]$. (En particulier $U_1^* = \min(U_1, \dots, U_n)$ et $U_n^* = \max(U_1, \dots, U_n)$). Cette loi a pour densité f^* définie par :

$$f^*(s_1, s_2, \dots, s_n) = \begin{cases} \frac{n!}{t^n} & \text{si } 0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq t; \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, pour simuler les n premiers temps d'arrivée d'un processus de Poisson sachant que $[N_t = n]$, il suffit de tirer n nombres au hasard entre 0 et 1 (touche "random" de la calculatrice), de les ranger par ordre croissant, puis de les multiplier par t en faisant éventuellement les conversions en format horaire adéquat (heures, minutes, seconde).

2.5 Processus de Poisson composés

Pour un processus de Poisson, la condition C3 entraîne l'impossibilité d'une réalisation simultanée de deux événements ou plus. Si on supprime cette condition, les événements peuvent se produire par "grappes", l'effectif de la grappe étant lui-même aléatoire.

Exemples :

→ Arrivées d'avions dans un aéroport : chaque avion transporte un certain nombre de passagers ;

→ Trafic routier : chaque accident engendre un certain nombre de blessés...

Les instants d'occurrence des grappes d'événements sont les A_n d'un processus de Poisson de paramètre λ .

À chaque instant A_n est associée une variable aléatoire Y_n désignant le nombre d'événements se produisant à l'instant A_n .

On suppose les Y_n indépendantes et de même loi.

Z_t est le nombre d'événements apparus jusqu'à l'instant t .

On a la relation fondamentale suivante :

$$Z_t = Y_1 + Y_2 + \dots + Y_{N_t}.$$

Cette relation permet, en particulier, de déterminer la fonction génératrice de Z_t et donc son espérance et sa variance.

Propriété : $G_{Z_t}(s) = \exp(\lambda t(G_Y(s) - 1))$; $\mathbb{E}(Z_t) = \lambda t \mathbb{E}(Y)$ et $\text{var}(Z_t) = \lambda t \mathbb{E}(Y^2)$.

Preuve : $G_{Z_t}(s) = \mathbb{E}\left(\mathbb{E}^{N_t}\left(s^{Z_t}\right)\right)$ avec

$$\mathbb{E}^{[N_t=n]}\left(s^{Z_t}\right) = \mathbb{E}^{[N_t=n]}\left(s^{Y_1+\dots+Y_n}\right) = (G_Y(s))^n$$

car les Y_i sont indépendantes entre elles et indépendantes de N_t .

Donc $\mathbb{E}^{N_t}\left(s^{Z_t}\right) = (G_Y(s))^{N_t}$ et

$$\mathbb{E}\left(s^{Z_t}\right) = \mathbb{E}\left(\mathbb{E}^{N_t}\left(s^{Z_t}\right)\right) = \mathbb{E}\left(G_Y(s)^{N_t}\right) = G_{N_t}(G_Y(s)) = \exp(\lambda t(G_Y(s) - 1))$$

car N_t suit la loi de Poisson de paramètre λt . On a alors

$$G'_{Z_t}(s) = \lambda t G'_Y(s) G_{Z_t}(s) \text{ et } G''_{Z_t}(s) = \lambda t G_{Z_t}(s) (G''_Y(s) + \lambda t (G'_Y(s))^2).$$

En prenant la limite quand s tend vers 1, on obtient $\mathbb{E}(Z_t) = \lambda t \mathbb{E}(Y)$, et

$$\lim_{s \rightarrow 1^-} G''_{Z_t}(s) = \lambda t ((\text{var}(Y) - \mathbb{E}(Y) + \mathbb{E}(Y)^2) + \lambda t (\mathbb{E}(Y))^2) = \text{var}(Z_t) - \mathbb{E}(Z_t) + (\mathbb{E}(Z_t))^2$$

d'où $\text{var}(Z_t) = \lambda t (\text{var}(Y) + \mathbb{E}(Y)^2) = \lambda t \mathbb{E}(Y^2)$. □

Exemple :

Le nombre d'accidents par jour dans une ville suit un processus de Poisson de paramètre 2 et le nombre de personnes impliquées suit la loi géométrique de paramètre $\frac{1}{2}$. Quelle est la moyenne et la variance du nombre de personnes accidentées durant une semaine ?

$\mathbb{E}(Y) = 2$ et $\text{var}(Y) = 2 = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$ d'où $\mathbb{E}(Y^2) = 6$ et, pour $t = 7$, $\mathbb{E}(Z_t) = 2 \times 7 \times 2 = 28$ et $\text{var}(Z_t) = 2 \times 7 \times 6 = 84$. □

Remarque :

On a $\mathbb{E}(N_t) = \lambda t$, donc, par exemple, dans le cas de l'avion, le nombre moyen de passagers arrivés jusqu'à l'instant t est égal au nombre moyen $\mathbb{E}(N_t)$ d'avions arrivés jusqu'à t multiplié par le nombre moyen de passagers par avion $\mathbb{E}(Y)$, ce qui paraît évident.

2.6 Processus non homogènes

Soit $(X_t)_{t \geq 0}$ un processus de comptage à accroissements indépendants vérifiant $X_0 = 0$,

$$P([X_{t+h} = i + 1] / [X_t = i]) = \lambda(t)h + o(h) \text{ et } P([X_{t+h} = i] / [X_t = i]) = 1 - \lambda(t)h + o(h).$$

On note A_n le $n^{\text{ième}}$ temps d'arrivée du processus.

Propriété : X_t suit la loi de Poisson de paramètre $\Lambda(t) = \int_0^t \lambda(u) du$ et $T_1 = A_1$ suit la loi de densité $t \mapsto \lambda(t) \exp\left[-\int_0^t \lambda(u) du\right] \mathbb{1}_{]0, +\infty[}(t)$.

Remarques :

[1] Pour un processus de Poisson (X_t) d'intensité λ , on a

$$P([X_{t+h} = i + 1] / [X_t = i]) = \lambda h + o(h) \text{ et } P([X_{t+h} = i] / [X_t = i]) = 1 - \lambda h + o(h).$$

c'est-à-dire que λ ne dépend pas de t .

[2] Le processus étudié ici est une généralisation du processus de Poisson. Pour $\lambda(t) = \lambda$ constant, on retrouve que X_t suit la loi de Poisson de paramètre $\Lambda(t) = \lambda \times t$ et que T_1 suit la loi exponentielle de paramètre λ .

Preuve : On reprend la méthode utilisée pour montrer que, si (X_t) est un processus à accroissements indépendants, stationnaire et à événements rares, alors X_t suit la loi de Poisson $\mathcal{P}(\lambda t)$: on pose $\pi_n(t) = P([X_t = n])$ puis, en considérant $\pi_n(t+h)$ et en passant à la limite quand $h \rightarrow 0$, on va établir une équation différentielle. $P([X_{t+h} = 0]) = P([X_t = 0] \cap [X_{t+h} - X_t = 0]) = P([X_t = 0])(1 - \lambda(t)h + o(h))$, soit

$$\pi_0(t+h) - \pi_0(t) = -\lambda(t)h\pi_0(t) + o(h).$$

D'autre part, pour $n \geq 1$,

$$\begin{aligned} P([X_{t+h} = n]) &= \sum_k P([X_t = k] \cap [X_{t+h} - X_t = n - k]) \\ &= P([X_t = n])(1 - \lambda(t)h) + P([X_t = n - 1])\lambda(t)h + o(h) \end{aligned}$$

soit $\pi_n(t+h) - \pi_n(t) = (-\lambda(t)\pi_n(t) + \lambda(t)\pi_{n-1}(t))h + o(h)$. On a donc, en divisant par h et en faisant $h \rightarrow 0$:

$$\boxed{\begin{cases} \pi'_0(t) + \lambda(t)\pi_0(t) = 0 \\ \pi'_n(t) + \lambda(t)\pi_n(t) = \lambda(t)\pi_{n-1}(t) \end{cases} \text{ avec } \pi_0(0) = 1 \text{ et } \pi_n(0) = 0 \text{ pour } n \geq 1.}$$

On résout alors, soit en utilisant les équations différentielles du premier ordre (et la méthode de variation de la constante), soit à l'aide de la fonction génératrice de X_t :

$$G(z, t) = \sum_{n=0}^{+\infty} z^n \pi_n(t).$$

Première méthode :

$$\frac{\pi'_0(t)}{\pi_0(t)} = [\ln(\pi_0(t))]' = -\lambda(t) \text{ et } \ln(\pi_0(t)) - \ln(\pi_0(0)) = -\int_0^t \lambda(u)du.$$

Ainsi, $\pi_0(t) = P([X_t = 0]) = \exp(-\Lambda(t))$. D'autre part, par la méthode de variation de la constante,

$$\pi_n(t) = C(t) \exp(-\Lambda(t)) \text{ avec } C'(t) \exp(-\Lambda(t)) = \lambda(t)\pi_{n-1}(t)$$

et par récurrence, si on suppose que $\pi_{n-1}(t) = \exp(-\Lambda(t)) \frac{(\Lambda(t))^{n-1}}{(n-1)!}$, on a alors $C'(t) = \lambda(t) \frac{(\Lambda(t))^{n-1}}{(n-1)!}$ et, comme $C(0) = 0$ ($\pi_n(0) = 0$ pour $n \geq 1$), on a alors $C(t) = \frac{(\Lambda(t))^n}{n!}$ et on a bien :

$$\boxed{P([X_t = n]) = \pi_n(t) = \exp(-\Lambda(t)) \frac{(\Lambda(t))^n}{n!}.$$

Deuxième méthode : On a $\frac{\partial G}{\partial t}(z, t) + \lambda(t)G(z, t) = \lambda(t) \sum_{n=1}^{+\infty} z^n \pi_{n-1}(t) = \lambda(t)zG(z, t)$, soit $\frac{G'}{G} = (z-1)\lambda$ et,

comme $G(0) = \pi_0(0) = 1$, $G(z, t) = \exp \left[(z-1) \int_0^t \lambda(u)du \right]$. On reconnaît bien là la fonction génératrice de la

loi de Poisson de paramètre $\Lambda(t) = \int_0^t \lambda(u)du$.

Pour la loi de T_1 , $P([T_1 > t]) = P([X_t = 0]) = \exp(-\Lambda(t)) = 1 - F_{T_1}(t)$ pour $t > 0$. On a donc bien

$$\boxed{f_{T_1}(t) = \lambda(t) \exp(-\Lambda(t)) \text{ pour } t > 0.}$$

Chapitre 3

Processus de naissance et de mort

3.1 Généralités

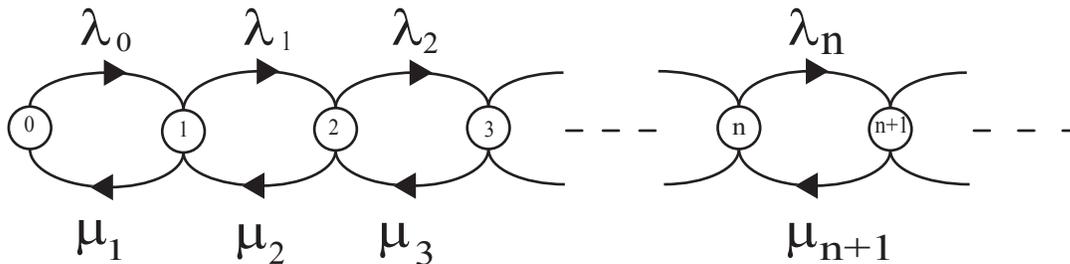
Utilisés plus particulièrement en biologie, démographie, physique, sociologie, pour rendre compte de l'évolution de la taille d'une population, les processus de naissance et de mort sont des processus de Markov continus ($T = \mathbb{R}_+$), à valeurs dans $E = \mathbb{N}$ tels que les seules transitions non négligeables possibles à partir de k soient vers $k + 1$ ou vers $k - 1$. Le générateur infinitésimal du processus est donc une matrice dite "tridiagonale" $A = (a_{i,j})_{i,j \in \mathbb{N}}$ vérifiant $a_{i,j} = 0$ si $|i - j| \geq 2$.

On posera $a_{k,k+1} = \lambda_k$ et $a_{k,k-1} = \mu_k$ pour $k \geq 1$ (et $a_{0,1} = \lambda_0$) : λ_k représente le taux de naissance à partir de l'état k et μ_k le taux de mort à partir de l'état k .

Les files d'attente de type Markovien (M/M) sont des cas particuliers très importants de processus de naissance et de mort. Leur étude complète sera effectuée dans le chapitre 6.

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & (0) \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ & & \mu_3 & \ddots & \ddots \\ (0) & & \ddots & \ddots & \ddots \end{pmatrix}$$

de graphe des taux :



Le graphe des taux de transition est constitué de "boudins", les arcs vers la droite représentant les taux de naissance, et ceux vers la gauche les taux de mort. Ces processus sont l'analogie des "chemins aléatoires" dans le cas où l'échelle des temps est continue.

3.1.1 Régime transitoire

On rappelle que si $P(t) = (p_{i,j}(t))$, où $p_{i,j}(t) = P([X_{u+t} = j] / [X_u = i])$, alors

$$P(t) = e^{At} = \sum_{k=0}^{+\infty} \frac{t^k}{k!} A^k.$$

La loi de X_t est alors donnée, en théorie, par $\vec{\pi}(t) = \vec{\pi}(0)P(t)$. (On notera ici, par commodité d'écriture, $\vec{\pi}(t) = (\pi_n(t))_{n \in \mathbb{N}}$, avec $\pi_n(t) = P([X_t = n])$).

Malheureusement, le calcul de e^{At} s'avère bien souvent très compliqué (puissances de matrice, puis série à sommer!). On préférera, en général, garder l'expression différentielle, même si celle-ci est souvent tout aussi difficile à résoudre.

Équations de Kolmogorov : On a $\vec{\pi}'(t) = \vec{\pi}(0)P'(t)$ qui redonne, en dérivant,

$$\vec{\pi}'(t) = \vec{\pi}(0)P'(t) = \vec{\pi}(0)P(t)A = \vec{\pi}(t)A,$$

et donc $\pi_j'(t) = \sum_i \pi_i(t)a_{i,j}$, qui donne ici le système suivant, dit d'*équations de Kolmogorov* :

$$\boxed{\begin{cases} \pi_0'(t) = -\lambda_0\pi_0(t) + \mu_1\pi_1(t) \\ \pi_n'(t) = \lambda_{n-1}\pi_{n-1}(t) - (\lambda_n + \mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t) \text{ pour } n \geq 1 \end{cases}}$$

3.1.2 Régime permanent

Lorsque, quand $t \rightarrow +\infty$, les limites $\pi_n = \lim_{t \rightarrow +\infty} \pi_n(t)$ existent et sont indépendantes de l'état initial du processus, on a alors $\vec{\pi}' = \vec{0}$ et $\vec{\pi}A = \vec{0}$ i.e. $\vec{\pi}$ est distribution stationnaire du processus si le régime permanent existe. Ceci se traduit par les équations dites "de balance"

$$\boxed{\begin{cases} 0 = -\lambda_0\pi_0 + \mu_1\pi_1 \\ 0 = \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} \text{ pour } n \geq 1 \end{cases}}$$

que l'on retrouve en écrivant en chaque état l'égalité du flux entrant et du flux sortant :

- état 0 : $\mu_1\pi_1 = \lambda_0\pi_0$;
- état 1 : $\mu_2\pi_2 + \lambda_0\pi_0 = \mu_1\pi_1 + \lambda_1\pi_1$;
- \vdots
- état n : $\mu_{n+1}\pi_{n+1} + \lambda_{n-1}\pi_{n-1} = \mu_n\pi_n + \lambda_n\pi_n$;
- \vdots

auxquelles il faut ajouter l'équation $\sum_{n=0}^{+\infty} \pi_n = 1$ pour que $\vec{\pi}$ définisse bien une probabilité.

Ces équations se simplifient successivement pour donner finalement des égalités "boudins par boudins" :

$$\begin{cases} \mu_1\pi_1 = \lambda_0\pi_0 \\ \mu_2\pi_2 = \lambda_1\pi_1 \\ \vdots \\ \mu_n\pi_n = \lambda_{n-1}\pi_{n-1} \\ \vdots \end{cases}.$$

On en déduit alors $\pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \pi_0$ pour $n \geq 1$, avec $\pi_0 \left(1 + \sum_{n=1}^{+\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \right) = 1$.

3.2 Étude de quelques cas particuliers

On étudie ici les cas particuliers des populations où les taux de naissance et de mort sont linéaires :

$$\lambda_n = n\lambda + \alpha$$

→ α représente le taux d'immigration : arrivées venant de l'extérieur. Il est supposé constant (et donc indépendant du nombre d'individus déjà présents).

→ λ représente le taux de naissance : chaque individu est susceptible de donner naissance à un nouvel individu avec le taux λ et s'il y a déjà n individus, la probabilité qu'il y ait une naissance sur l'intervalle $[t, t + h[$ est alors $n\lambda h + o(h)$.

$$\mu_n = n\mu + \beta \text{ si } n \geq 1$$

→ β représente le taux d'émigration : départs vers l'extérieur. Il est supposé constant (sauf s'il n'y a personne, auquel cas il est nul).

→ μ représente le taux de mort : chaque individu est susceptible de mourir avec le taux μ et s'il y a déjà n individus, la probabilité qu'il y ait une mort sur l'intervalle $[t, t + h[$ est alors $n\mu h + o(h)$.

3.2.1 Croissance pure, par immigration

C'est le cas $\lambda_n = \alpha$ et $\mu_n = 0$ pour tout $n \in \mathbb{N}$.

$$\begin{cases} \pi'_0(t) = -\alpha\pi_0(t) \\ \pi'_n(t) = \alpha\pi_{n-1}(t) - \alpha\pi_n(t) \text{ pour } n \geq 1 \end{cases} .$$

On retrouve les équations vérifiées par le processus de Poisson (N_t) de paramètre α . En

particulier X_t suit la loi de Poisson $\mathcal{P}(\alpha t) : \pi_n(t) = e^{-\alpha t} \frac{(\alpha t)^n}{n!}$.

3.2.2 Croissance pure, par naissance

C'est le cas $\lambda_n = n\lambda$ et $\mu_n = 0$ pour tout $n \in \mathbb{N}$.

Ceci n'a de sens que si $X_0 \geq 1$. On supposera que $X_0 = 1$.

$$\begin{cases} \pi'_0(t) = 0 \\ \pi'_n(t) = (n-1)\lambda\pi_{n-1}(t) - n\lambda\pi_n(t) \text{ pour } n \geq 1 \end{cases} .$$

Propriétés : X_t suit la loi géométrique $\mathcal{G}(e^{-\lambda t}) : \pi_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$.

Preuve : On peut trouver $\pi_n(t)$ par récurrence ascendante sur n (matrice A triangulaire supérieure !)

→ $\pi'_0(t) = 0$ donc $\pi_0(t) = \pi_0(0) = 0$.

→ $\pi'_1(t) = -\lambda\pi_1(t)$ donc $\pi_1(t) = Ce^{-\lambda t}$ avec $C = \pi_1(0) = 1$.

→ Supposons que $\pi_{n-1}(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-2}$ pour $n \geq 2$. On a

$$\pi'_n(t) + n\lambda\pi_n(t) = (n-1)\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-2}.$$

C'est une équation différentielle linéaire du premier ordre avec second membre donc, pour la résoudre, on applique la méthode de la variation de la constante.

$\pi_n(t) = C'(t)e^{-n\lambda t}$ avec $C'(t)e^{-n\lambda t} = (n-1)\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-2}$, soit

$$\begin{aligned} C'(t) &= (n-1)\lambda e^{(n-1)\lambda t}(1 - e^{-\lambda t})^{n-2} \\ &= (n-1)\lambda e^{\lambda t}(e^{\lambda t} - 1)^{n-2} = \frac{d}{dt} \left((e^{\lambda t} - 1)^{n-1} \right) \end{aligned}$$

donc $C(t) - C(0) = (e^{\lambda t} - 1)^{n-1}$ et $C(0) = p_n(0) = 0$ pour $n \geq 2$. Finalement,

$$\pi_n(t) = (e^{\lambda t} - 1)^{n-1} e^{-n\lambda t} = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1}.$$

X_t suit donc la loi géométrique de paramètre $e^{-\lambda t}$ et, en particulier, $\mathbb{E}(X_t) = e^{\lambda t}$

3.2.3 Décroissance pure, par décès

C'est le cas $\lambda_n = 0$ et $\mu_n = n\mu$ pour tout $n \in \mathbb{N}$.

On suppose $X_0 = N \geq 1$. Ce modèle décrit l'évolution d'un système composé de N dispositifs indépendants non réparables, dont les durées de fonctionnement obéissent à une même loi exponentielle de paramètre μ (durée de vie moyenne $1/\mu$).

Dans ce cas, il est immédiat que $\pi_n(t) = C_N^n e^{-n\mu t} (1 - e^{-\mu t})^{N-n}$ pour $n \in \{0, 1, \dots, N\}$.

En effet, la probabilité qu'un dispositif de durée de vie T fonctionne encore à l'instant t est $P([T > t]) = e^{-\mu t}$ (donc la probabilité qu'il ne fonctionne plus est $(1 - e^{-\mu t})$). Comme $\pi_n(t)$ est la probabilité qu'il y ait exactement n dispositifs qui fonctionnent à l'instant t et qu'il y ait exactement C_N^n façons de choisir les n qui fonctionnent, on a bien le résultat.

On peut le retrouver avec les équations de Kolmogorov :

$$\begin{cases} \pi'_n(t) = -n\mu\pi_n(t) + (n+1)\mu\pi_{n+1}(t) \text{ pour } n \in \{0, 1, \dots, N-1\} \\ \pi'_N(t) = -N\mu\pi_N(t) \end{cases}.$$

Propriétés : X_t suit la loi binomiale $\mathcal{B}(N, e^{-\mu t})$: $\pi_n(t) = C_N^n e^{-n\mu t} (1 - e^{-\mu t})^{N-n}$.

Preuve : On peut trouver $\pi_n(t)$ par récurrence descendante sur n (matrice A triangulaire inférieure !)

→ $\pi'_N(t) + N\mu\pi_N(t) = 0$ donc $\pi_N(t) = C e^{-N\mu t}$ avec $C = \pi_N(0) = 1$.

→ Supposons maintenant que $\pi_{n+1}(t) = C_N^{n+1} (e^{-\mu t})^{n+1} (1 - e^{-\mu t})^{N-n-1}$ pour $n \in \{1, \dots, N-1\}$. On a

$$\pi'_n(t) + n\mu\pi_n(t) = (n+1)\mu C_N^{n+1} (e^{-\mu t})^{n+1} (1 - e^{-\mu t})^{N-n-1}.$$

On applique la méthode de la variation de la constante : $\pi_n(t) = C(t)e^{-n\mu t}$ avec

$$\begin{aligned} C'(t) &= e^{n\mu t} \times (n+1)\mu \frac{N!}{(n+1)!(N-n-1)!} e^{-(n+1)\mu t} (1 - e^{-\mu t})^{N-n-1} \\ &= (N-n)\mu C_N^n e^{-\mu t} (1 - e^{-\mu t})^{N-n-1} = \frac{d}{dt} \left(C_N^n (1 - e^{-\mu t})^{N-n} \right) \end{aligned}$$

donc $C(t) - C(0) = C_N^n (1 - e^{-\mu t})^{N-n}$ avec $C(0) = \pi_n(0) = 0$ pour $n < N$ et

$$\pi_n(t) = C_N^n e^{-n\mu t} (1 - e^{-\mu t})^{N-n}.$$

X_t suit donc la loi binomiale $\mathcal{B}(N, e^{-\mu t})$ et, en particulier, $\mathbb{E}(X_t) = N e^{-\mu t}$.

3.2.4 Processus de Yule-Ferry

C'est le cas $\lambda_n = n\lambda$ et $\mu_n = n\mu$ pour tout $n \in \mathbb{N}$.

$$\begin{cases} \pi'_0(t) = \mu_1\pi_1(t) \\ \pi'_n(t) = (n-1)\lambda\pi_{n-1}(t) - n(\lambda + \mu)\pi_n(t) + (n+1)\mu\pi_{n+1}(t) \text{ pour } n \geq 1 \end{cases}.$$

Dans ce cas, la matrice A n'est plus triangulaire et les $\pi_n(t)$ sont plus difficiles à déterminer. On peut toutefois déterminer $\mathbb{E}(X_t)$ dans un cadre un peu plus général (autorisant l'immigration mais pas l'émigration :

Propriétés : Si $\lambda_n = n\lambda + \alpha$ et $\mu_n = n\mu$ et si $X_0 = N$, alors

$$M(t) = \mathbb{E}(X_t) = \begin{cases} \alpha t + N & \text{si } \lambda = \mu \\ N e^{(\lambda-\mu)t} + \frac{\alpha}{\lambda-\mu} (e^{(\lambda-\mu)t} - 1) & \text{si } \lambda \neq \mu \end{cases}$$

Preuve : On écrit les équations de Kolmogorov :

$$\begin{cases} \pi'_0(t) = -\alpha\pi_0(t) + \mu\pi_1(t) \\ \pi'_n(t) = ((n-1)\lambda + \alpha)\pi_{n-1}(t) - (n(\lambda + \mu) + \alpha)\pi_n(t) + (n+1)\mu\pi_{n+1}(t) \text{ pour } n \geq 1 \end{cases}$$

On pose $G(s, t) = \mathbb{E}(s^{X_t}) = \sum_{n=0}^{+\infty} s^n \pi_n(t)$. On a alors $\partial_2 G(s, t) = \sum_{n=0}^{+\infty} s^n \pi'_n(t)$ qui s'écrit ici :

$$\begin{aligned} \partial_2 G(s, t) &= -\alpha\pi_0(t) + \mu\pi_1(t) + \sum_{n=1}^{+\infty} s^n [(n-1)\lambda + \alpha]\pi_{n-1}(t) - (n(\lambda + \mu) + \alpha)\pi_n(t) + (n+1)\mu\pi_{n+1}(t) \\ &= -\alpha\pi_0(t) + \mu\pi_1(t) + \sum_{n=0}^{+\infty} s^{n+1} (n\lambda + \alpha)\pi_n(t) - \sum_{n=1}^{+\infty} (n(\lambda + \mu) + \alpha)s^n \pi_n(t) + \sum_{n=2}^{+\infty} n\mu s^{n-1} \pi_n(t) \end{aligned}$$

avec $\sum_{n=0}^{+\infty} s^{n+1} (n\lambda + \alpha)\pi_n(t) = s^2 \lambda \sum_{n=1}^{+\infty} n s^{n-1} \pi_n(t) + \alpha s G(s, t) = s^2 \lambda \partial_1 G(s, t) + \alpha s G(s, t)$,

$$\sum_{n=1}^{+\infty} (n(\lambda + \mu) + \alpha)s^n \pi_n(t) = s(\lambda + \mu)\partial_1 G(s, t) + \alpha G(s, t) - \alpha\pi_0(t),$$

$$\sum_{n=2}^{+\infty} n\mu s^{n-1} \pi_n(t) = \mu \partial_1 G(s, t) - \mu_1 \pi_1(t)$$

d'où finalement $\partial_2 G(s, t) = (s^2 \lambda - s(\lambda + \mu) + \mu)\partial_1 G(s, t) + \alpha(s-1)G(s, t)$.

Cette équation se résout grâce à la théorie des équations aux dérivées partielles mais l'étude de ce type de problèmes n'est pas à l'ordre du jour et on va ici se contenter d'en déduire la taille moyenne de la population à l'instant t : $M(t) = \mathbb{E}(X_t)$.

On remarque que $M(t) = \sum_{n=1}^{+\infty} n\pi_n(t) = \partial_1 G(1_-, t)$ et on va trouver une équation différentielle linéaire du premier ordre dont M est solution. Pour cela, on dérive l'équation précédente par rapport à s et on prend la limite quand $s \rightarrow 1_-$:

$$\partial_1 \partial_2 G(s, t) = (2s\lambda - (\lambda + \mu))\partial_1 G(s, t) + (s^2 \lambda - s(\lambda + \mu) + \mu)\partial_1^2 G(s, t) + \alpha G(s, t) + \alpha(s-1)\partial_1 G(s, t)$$

Or $\partial_1 G(1_-, t) = M(t)$, $\partial_1 \partial_2 G(1_-, t) = \partial_2 \partial_1 G(1_-, t) = M'(t)$ et $G(1_-, t) = 1$ donnent :

$$M'(t) = (\lambda - \mu)M(t) + \alpha$$

- Si $\lambda = \mu$, $M'(t) = \alpha$ et $M(t) = \alpha t + N$ car $M(0) = N$.
- Si $\lambda \neq \mu$, l'équation sans second membre a pour solution $Ce^{(\lambda-\mu)t}$ et une solution particulière de l'équation complète est $-\frac{\alpha}{\lambda-\mu}$ donc $M(t) = Ce^{(\lambda-\mu)t} - \frac{\alpha}{\lambda-\mu}$ avec $M(0) = N = C - \frac{\alpha}{\lambda-\mu}$ et finalement

$$M(t) = Ne^{(\lambda-\mu)t} + \frac{\alpha}{\lambda-\mu} (e^{(\lambda-\mu)t} - 1)$$

3.2.5 Cas particuliers à taux non linéaires

[1] Un modèle logistique : On considère une population dont la taille X_t est comprise entre 2 entiers N_1 et N_2 et dont les taux de naissance et de mort par individu sont

$$\lambda = \alpha(N_2 - X_t) \quad \text{et} \quad \mu = \beta(X_t - N_1)$$

(plus il y a de monde, plus le taux de naissance est faible et le taux de mort fort ; moins il y a de monde, plus le taux de naissance est fort et le taux de mort faible : les individus ont besoin de place pour se développer !).

On suppose que les membres de cette population agissent indépendamment les uns des autres. Les taux de naissance et de mort résultant pour la population sont alors

$$\lambda_n = n\alpha(N_2 - n) \quad \text{et} \quad \mu_n = n\beta(n - N_1).$$

Propriétés : Un tel processus admet une distribution stationnaire définie par :

$$p_{N_1+k} = \frac{C}{N_1+k} C_{N_2-N_1}^k \left(\frac{\alpha}{\beta}\right)^k \quad \text{pour } k \in \{0, \dots, N_2 - N_1\}$$

$$\text{avec } C = \left[\sum_{j=0}^{N_2-N_1} \frac{1}{N_1+j} C_{N_2-N_1}^j \left(\frac{\alpha}{\beta}\right)^j \right]^{-1}.$$

Preuve : On pose $\pi_k = p_{N_1+k}$ on écrit les équations de "balance" :

$$\lambda_{m-1}\pi_{m-1} = \mu_m\pi_m \quad \text{pour } m \in \{N_1+1, \dots, N_2\}.$$

Avec $m = N_1 + k$, on a $\lambda_m = \alpha(N_1 + k)(N_2 - N_1 - k)$ et $\mu_m = \beta k(N_1 + k)$. Ainsi, on a :

$$\begin{aligned} \pi_1 &= \frac{\alpha}{\beta} \frac{N_1(N_2 - N_1)}{1(N_1 + 1)} \pi_0 \\ \pi_2 &= \left(\frac{\alpha}{\beta}\right)^2 \frac{N_1(N_1 + 1)(N_2 - N_1)(N_2 - N_1 - 1)}{1.2.(N_1 + 1)(N_1 + 2)} \pi_0 = \left(\frac{\alpha}{\beta}\right)^2 \frac{N_1}{N_1 + 2} C_{N_2-N_1}^2 \pi_0 \\ &\vdots \\ \pi_k &= \left(\frac{\alpha}{\beta}\right)^k \frac{N_1(N_1 + 1) \cdots (N_1 + k - 1)(N_2 - N_1) \cdots (N_2 - N_1 - k + 1)}{k!(N_1 + 1) \cdots (N_1 + k)} \pi_0 = \left(\frac{\alpha}{\beta}\right)^k \frac{N_1}{N_1 + k} C_{N_2-N_1}^k \pi_0 \end{aligned}$$

puis on détermine π_0 en utilisant $\pi_0 + \pi_1 + \dots + \pi_{N_2-N_1} = 1$.

[2] Un modèle génétique : On considère ici une population composée de N individus qui ont, soit un gène type a ou un gène type A . L'état du processus X_t représente le nombre d'individus a à l'instant t . On suppose que la probabilité que l'état change pendant l'intervalle de temps $[t, t + h[$ est $\lambda h + o(h)$ indépendante des valeurs de X_t et que la probabilité que 2 changements ou plus se produisent pendant un tel intervalle est $o(h)$.

Les changements dans la structure de la population s'effectuent de la façon suivante : un individu sera remplacé par un autre choisi au hasard dans la population ; c'est-à-dire que, si $X_t = n$, on choisira pour être remplacé un type a avec la probabilité n/N et un type A avec la probabilité $1 - n/N$. On parlera de cette étape comme de la mort. Ensuite, la naissance prend place suivant la règle ci dessous :

un autre choix est fait au hasard dans la population pour déterminer le type du nouvel individu remplaçant de celui qui est mort. Le modèle suppose que le type du nouvel individu

soit altéré au cours de la naissance. De façon plus précise, soit γ_1 la probabilité qu'un type a donne par mutation un type A et γ_2 la probabilité qu'un type A donne par mutation un type a . La probabilité que le nouvel individu ajouté à la population soit de type a est

$$\frac{n}{N}(1 - \gamma_1) + \left(1 - \frac{n}{N}\right) \gamma_2$$

(on choisit un type a et il reste a ou bien on choisit un type A qui change).

Le taux de naissance correspondant à une augmentation d'une unité des individus de type a , on doit avoir disparition d'un type A et apparition d'un type a soit

$$\lambda_n = \lambda \left(1 - \frac{n}{N}\right) \left[\frac{n}{N}(1 - \gamma_1) + \left(1 - \frac{n}{N}\right) \gamma_2\right].$$

De même, la probabilité que le nouvel individu ajouté à la population soit de type A est

$$\frac{n}{N} \gamma_1 + \left(1 - \frac{n}{N}\right) (1 - \gamma_2)$$

(on choisit un type a et il change ou bien on choisit un type A qui reste A).

Le taux de mort correspondant à une diminution d'une unité des individus de type a , on doit avoir disparition d'un type a et apparition d'un type A soit

$$\mu_n = \lambda \frac{n}{N} \left[\frac{n}{N} \gamma_1 + \left(1 - \frac{n}{N}\right) (1 - \gamma_2)\right].$$

3.3 Problème de l'extinction de l'espèce

Lorsque l'on étudie une population telle que $\lambda_0 = 0$ (quand il n'y a plus personne, c'est définitif !), on peut chercher la probabilité qu'elle a de disparaître un jour, et, si elle doit disparaître, le temps qu'elle va mettre pour disparaître. On est donc conduit à considérer les probabilités :

→ a_k : probabilité d'extinction de l'espèce si la taille initiale est k ;
ainsi que les variables aléatoires :

→ T_k : temps écoulé jusqu'à l'extinction de l'espèce si la taille initiale est k .

On est dans le cas, où, si $X_t = 0$, alors $X_\tau = 0$ pour $\tau \geq t$ et

$$P([T_k \leq t]) = P^{[X_0=k]}([X_t = 0]) = p_{k,0}(t).$$

La fonction de répartition de T_k est donc $F_{T_k} : t \mapsto P_{k,0}(t) = \pi_0(t)$, où π_0 est déterminé par les équations de Kolmogorov, avec la condition initiale $\pi_k(0) = 1$.

De plus, $a_k = P([T_k < +\infty]) = \lim_{t \rightarrow +\infty} P^{[X_0=k]}([X_t = 0])$.

En pratique, la détermination du régime transitoire est bien souvent compliquée : ainsi, on se contentera parfois, à défaut d'avoir la loi de T_k , de déterminer $\tau_k = \mathbb{E}(T_k)$: temps moyen jusqu'à l'extinction, lorsque la taille initiale est k .

Pour la détermination de a_k , on considèrera la chaîne de Markov induite telle qu'on l'a définie dans le chapitre sur les processus de Markov. On a ici $p_{n,n+1} = \frac{\lambda_n}{\lambda_n + \mu_n} = p_n$ et $p_{n,n-1} = \frac{\mu_n}{\lambda_n + \mu_n} = q_n$ et 0 est une classe absorbante, alors que \mathbb{N}^* constitue une classe transitoire.

On a alors les résultats suivants :

Théorème : On pose $d_j = \frac{q_1 \cdots q_j}{p_1 \cdots p_j} = \frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j}$ et $\rho_j = \frac{\lambda_1 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j} = \frac{1}{d_j \lambda_j}$.

- La probabilité d'extinction à partir de la taille initiale k est :

$$a_k = \begin{cases} \frac{\sum_{j=k}^{+\infty} d_j}{1 + \sum_{j=1}^{+\infty} d_j} & \text{si } \sum_{j=1}^{+\infty} d_j \text{ converge} \\ 1 & \text{sinon} \end{cases}$$
- Le temps moyen jusqu'à l'extinction est :

$$\tau_k = \begin{cases} \sum_{j=1}^{+\infty} \rho_j + \sum_{i=1}^{k-1} d_i \sum_{j=i+1}^{+\infty} \rho_j & \text{si } \sum_{j=1}^{+\infty} \rho_j \text{ converge} \\ +\infty & \text{sinon} \end{cases}$$

Preuve : La probabilité d'extinction, c'est-à-dire d'absorption par l'état 0, est déterminée en détail dans l'exercice qui suit :

Exercice : Un joueur décide de faire des paris jusqu'à qu'il soit ruiné ou bien qu'il ait atteint N euro. S'il dispose de j euro quand il effectue un pari, alors, à ce pari, il gagne 1 euro avec la probabilité p_j ou bien il perd 1 euro avec la probabilité $q_j = 1 - p_j$. On note a_k la probabilité de finir ruiné avec une fortune initiale de k euro et $a_k^{(n)}$ la probabilité de finir ruiné en n paris.

- Modéliser le problème par une chaîne de Markov et faire le graphe correspondant.
- Déterminer a_N et a_0 .
- Montrer que :

- $a_1^{(1)} = q_1$ et $a_1^{(n)} = p_1 a_2^{(n-1)}$ si $n \geq 2$;
- pour $j \geq 2$, $a_j^{(1)} = 0$ et $a_j^{(n)} = p_j a_{j+1}^{(n-1)} + q_j a_{j-1}^{(n-1)}$ si $n \geq 2$.

- En déduire que :

- $a_1 = q_1 + p_1 a_2$ et $a_j = p_j a_{j+1} + q_j a_{j-1}$ si $j \geq 2$, puis
- $p_j (a_j - a_{j+1}) = q_j (a_{j-1} - a_j)$ pour $1 \leq j \leq N - 1$;
- $a_j - a_{j+1} = d_j (a_0 - a_1)$ pour $1 \leq j \leq N - 1$, si on pose $d_j = \frac{q_1 \cdots q_j}{p_1 \cdots p_j}$.

- Vérifier que $a_k = \sum_{j=k}^{N-1} (a_j - a_{j+1})$ pour $1 \leq k \leq N - 1$ et que $1 = \sum_{j=0}^{N-1} (a_j - a_{j+1})$, et en déduire que

$$a_k = \frac{\sum_{j=k}^{N-1} d_j}{1 + \sum_{j=1}^{N-1} d_j}.$$

- En déduire, en faisant $N \rightarrow +\infty$, la probabilité de ruine du joueur.

a) Le processus est bien invariant dans le temps et la connaissance du processus à un instant suffit pour déterminer la probabilité d'occupation d'un état à l'instant suivant. On a bien une chaîne de Markov, de matrice de transition $P = (p_{i,j})$ où $p_{i,i+1} = p_i$, $p_{i,i-1} = q_i$ et $p_{i,j} = 0$ si $|i-j| \neq 1$.

b) $a_N = 0$ et $a_0 = 1$.

c) • $a_1^{(1)} = q_1$ (passage de 1 à 0 en 1 étape) et $a_1^{(n)} = p_1 a_2^{(n-1)}$ si $n \geq 2$ car alors, on est obligé de commencer par un passage de 1 à 2 pour ne pas être absorbé directement ; il faudra ensuite $n-1$ étapes pour passer de 2 à 0.

• pour $j \geq 2$, $a_j^{(1)} = 0$ car on ne peut pas passer directement de j à 0.

Pour déterminer $a_j^{(n)}$, on décompose en 2 cas suivant le premier déplacement : soit on commence par passer de j à $j+1$ (avec la probabilité p_j), soit on commence par passer de j à $j-1$ (avec la probabilité q_j) ; il restera alors $n-1$ déplacements à effectuer pour se rendre en 0 et $a_j^{(n)} = p_j a_{j+1}^{(n-1)} + q_j a_{j-1}^{(n-1)}$ si $n \geq 2$.

d) • Il faut maintenant utiliser $a_j = \sum_{n=1}^{+\infty} a_j^{(n)}$:

$$a_1 = a_1^{(1)} + \sum_{n \geq 2} a_1^{(n)} = q_1 + p_1 \sum_{n \geq 2} a_2^{(n-1)}, \text{ soit } a_1 = q_1 + p_1 a_2.$$

$$\text{De même, si } j \geq 2, a_j = a_j^{(1)} + \sum_{n \geq 2} a_j^{(n)} = p_j \sum_{n \geq 2} a_{j+1}^{(n-1)} + q_j \sum_{n \geq 2} a_{j-1}^{(n-1)}, \text{ d'où } a_j = p_j a_{j+1} + q_j a_{j-1} \text{ si } j \geq 2.$$

• On a alors, comme $a_j = (p_j + q_j)a_j$, $p_j(a_j - a_{j+1}) = q_j(a_{j-1} - a_j)$ pour $1 \leq j \leq N-1$ (car $a_j = p_j a_{j+1} + q_j a_{j-1}$ est encore vrai pour $j=1$ puisque $a_0 = 1$).

• $a_j - a_{j+1} = \frac{q_j}{p_j}(a_{j-1} - a_j)$ donne $a_1 - a_2 = \frac{q_1}{p_1}(a_1 - a_0)$, puis $a_2 - a_3 = \frac{q_2}{p_2} \frac{q_1}{p_1}(a_1 - a_0)$ et de proche en proche, $a_j - a_{j+1} = d_j(a_0 - a_1)$ pour $1 \leq j \leq N-1$, si on pose $d_j = \frac{q_1 \cdots q_j}{p_1 \cdots p_j}$.

e) $\sum_{j=k}^{N-1} (a_j - a_{j+1})$ est une "somme télescopique" : tout se simplifie sauf le premier et le dernier terme qui vaut $a_N = 0$ et donc $\sum_{j=k}^{N-1} (a_j - a_{j+1}) = a_k$ pour $1 \leq k \leq N-1$.

Le principe reste le même si l'on part de $j=0$. On a alors $\sum_{j=0}^{N-1} (a_j - a_{j+1}) = a_0 - a_N = 1$.

Ainsi $a_k = \frac{\sum_{j=k}^{N-1} (a_j - a_{j+1})}{\sum_{j=0}^{N-1} (a_j - a_{j+1})}$ et, comme $(a_j - a_{j+1}) = d_j(a_0 - a_1)$ pour $j \geq 1$, on a bien, en simplifiant

numérateur et dénominateur par $(a_0 - a_1)$, on a bien $a_k = \frac{\sum_{j=k}^{N-1} d_j}{1 + \sum_{j=1}^{N-1} d_j}$.

f) Si le joueur n'arrête que lorsqu'il est ruiné, cela revient à considérer que $N \rightarrow +\infty$ et on a alors

$$a_k = \frac{\sum_{j=k}^{+\infty} d_j}{1 + \sum_{j=1}^{+\infty} d_j} \text{ si } \sum_{j=1}^{+\infty} d_j \text{ converge.}$$

Pour le temps moyen d'absorption, on procède de même. On rappelle que le temps moyen de séjour dans un état j est $\frac{1}{\lambda_j + \mu_j}$, car la distribution de ce temps suit la loi exponentielle $\mathcal{E}(\lambda_j + \mu_j)$.

En considérant les états $j + 1$ et $j - 1$ qui peuvent résulter de la première transition, on a alors :

$$\tau_j = \frac{1}{\lambda_j + \mu_j} + \frac{\lambda_j}{\lambda_j + \mu_j} \tau_{j+1} + \frac{\mu_j}{\lambda_j + \mu_j} \tau_{j-1} \text{ pour } j \geq 1 \quad (3.1)$$

où par convention $\tau_0 = 0$. En posant $z_j = \tau_j - \tau_{j+1}$ et en réordonnant (3.1) on obtient :

$$z_j = \frac{1}{\lambda_j} + \frac{\mu_j}{\lambda_j} z_{j-1} \text{ pour } j \geq 1. \quad (3.2)$$

En itérant cette relation, il vient

$$z_k = \frac{1}{\lambda_k} + \frac{\mu_k}{\lambda_k} \frac{1}{\lambda_{k-1}} + \frac{\mu_k \mu_{k-1}}{\lambda_k \lambda_{k-1}} z_{k-2} \dots$$

... et on termine en suivant exactement la même démarche que dans l'exercice précédent.

Chapitre 4

Processus de ramification

Utilisé plus particulièrement en démographie, en génétique ou en physique nucléaire, c'est le modèle mathématique représentant le développement dans le temps d'une population.

Chaque individu génère, indépendamment des autres, un nombre aléatoire de descendants.

X_t est le nombre d'individus total à l'instant t dans le cas d'un processus continu (espace des temps $T = \mathbb{R}_+$) ; dans le cas discret ($T = \mathbb{N}$), X_n est le nombre total d'individus à la $n^{\text{ième}}$ génération.

Les hypothèses :

- que les individus agissent indépendamment les uns des autres ;
 - que la durée de vie d'individus de même type est la même ou suit la même loi ;
 - que la loi de reproduction est indépendante de l'époque et du nombre d'individus existants ;
- entraînent que le processus $(X_t)_{t \geq 0}$ (ou $(X_n)_{n \in \mathbb{N}}$) est markovien.

Problèmes qui se posent :

- Déterminer la loi de X_t ;
- Connaître la taille moyenne de la population à l'instant t (et éventuellement, la dispersion autour de cette valeur moyenne) (moins précis, mais plus parlant que la loi)
- Existe-t-il une probabilité non nulle d'extinction de l'espèce ?

Pour déterminer la loi de X_t , on exprimera en général sa fonction génératrice, qui la caractérise, et qui est définie par :

$$G_{X_t}(s) = \sum_{j=0}^{+\infty} s^j P([X_t = j]) = \mathbb{E}(s^{X_t}).$$

On notera $G(s, t) = \mathbb{E}^{[X_0=1]}(s^{X_t}) = \sum_{j=0}^{+\infty} s^j p_{1,j}(t)$.

Rappel sur les fonctions génératrices :

Si X est une variable aléatoire à valeurs dans \mathbb{N} , on pose :

$$G_X(s) = \sum_{k=0}^{+\infty} s^k P([X = k]).$$

G_X est une série entière de rayon ≥ 1 car $G_X(1) = \sum_{k=0}^{+\infty} P([X = k]) = 1$.

La fonction génératrice caractérise la loi. Lorsqu'elle a été déterminée, il est alors très facile de déterminer l'espérance et la variance de X . En effet,

$$\mathbb{E}(X) = \lim_{s \rightarrow 1^-} G'_X(s) = G'_X(1_-) \text{ et } \text{var}(X) = G''_X(1_-) + G'_X(1_-) - (G'_X(1_-))^2.$$

De plus, si X et X' sont deux variables aléatoires indépendantes à valeurs dans \mathbb{N} , on a

$$G_{X+X'}(s) = G_X(s) \times G_{X'}(s).$$

Le processus de ramification $(X_t)_{t \geq 0}$ (discret ou continu) étant un processus de Markov, on a la relation fondamentale suivante :

Théorème : $G(s, t + \tau) = G(G(s, \tau), t).$

Preuve : $G(s, t + \tau) = \sum_{j=0}^{+\infty} s^j p_{1,j}(t + \tau).$

Mais, comme $(X_t)_{t \geq 0}$ est Markovien, on a $P(t + \tau) = P(t)P(\tau)$ et

$$\sum_{j=0}^{+\infty} s^j p_{1,j}(t + \tau) = \sum_{j=0}^{+\infty} s^j \sum_{k=0}^{+\infty} p_{1,k}(t) p_{k,j}(\tau) = \sum_{k=0}^{+\infty} p_{1,k}(t) \sum_{j=0}^{+\infty} s^j p_{k,j}(\tau) = \sum_{k=0}^{+\infty} p_{1,k}(t) \mathbb{E}^{[X_0=k]}(s^{X_\tau})$$

Mais, si $X_0 = k$, $X_\tau = X_{\tau,1} + \dots + X_{\tau,k}$ où $X_{\tau,i}$ représente le nombre de descendants au temps τ du i -ième individu initial. Les $X_{\tau,i}$ étant k variables indépendantes de même loi, on a alors $\mathbb{E}^{[X_0=k]}(s^{X_\tau}) = G(s, \tau)^k$, puis

$$G(s, t + \tau) = \sum_{k=0}^{+\infty} G(s, \tau)^k p_{1,k}(t) = G(G(s, \tau), t).$$

□

4.1 Processus discret à 1 type

Exemples :

a) Survivance des noms de famille

On suppose que seuls les hommes conservent et transmettent leur nom.

b) Multiplicateur à électrons

On dresse sur la trajectoire d'un électron une série de plaques : lorsque l'électron rencontre une plaque, il engendre un nombre aléatoire de nouveaux électrons.

On note X_n le nombre d'individus à la n -ième génération.

Si Y est le nombre de descendants directs d'un individu, on note G la fonction génératrice de Y et on définit G_n par récurrence en posant

$$G_1 = G \text{ et pour tout } n \geq 1, G_{n+1} = G_n \circ G = G \circ G_n.$$

On pose également

$$m = \mathbb{E}(Y) \text{ et } \sigma^2 = \text{var}(Y).$$

Théorème : $\mathbb{E}^{[X_r=k]}(s^{X_{r+n}}) = (G_n(s))^k.$

Preuve : Grâce à l'homogénéité dans le temps, on a $\mathbb{E}^{[X_r=k]}(s^{X_{r+n}}) = \mathbb{E}^{[X_0=k]}(s^{X_n}) = G(s, n)^k.$

Mais $G(s, 1) = G(s) = G_1(s)$ et, si $G(s, n) = G_n(s)$, d'après la relation fondamentale,

$$G(s, n + 1) = G(G(s, n), 1) = G(G_n(s)) = G_{n+1}(s) ;$$

c'est donc bien que $G(s, n) = G_n(s)$ pour tout $n \geq 1$.

□

Conséquences :

$$[1] \mathbb{E}^{[X_r=k]}(X_{n+r}) = km^n.$$

$$[2] \text{ Si } X_0 = 1, \text{ alors } \mathbb{E}(X_n) = m^n \text{ et } \text{var}(X_n) = \begin{cases} \sigma^2 m^{n-1} \frac{m^n - 1}{m - 1} & \text{si } m \neq 1 \\ n\sigma^2 & \text{si } m = 1 \end{cases}.$$

(Ceci se démontre par récurrence, en dérivant une, puis deux fois $G_{n+1} = G_n \circ G$ comme une fonction composée et en prenant la valeur en $s = 1$.)

Interprétation de [2] :

- si $m < 1$, $\mathbb{E}(X_n) \rightarrow 0$, $\text{var}(X_n) \rightarrow 0$: moyenne de plus en plus significative ;
- si $m > 1$, $\mathbb{E}(X_n) \rightarrow +\infty$, $\text{var}(X_n) \rightarrow +\infty$: moyenne peu significative ;
- si $m = 1$, $\mathbb{E}(X_n) = 1$, $\text{var}(X_n) \rightarrow +\infty$: beaucoup de familles disparaissent, d'autres prolifèrent énormément.

Probabilité d'extinction :

On désire déterminer la probabilité que la population finisse par disparaître. Pour cela, il faut supposer que $P([Y = 0]) \in]0, 1[$, car, si $P([Y = 0]) = 1$, la population est sûre de disparaître et si $P([Y = 0]) = 0$, il n'y aura jamais extinction.

Théorème : Soit $\pi_0 = \lim_{n \rightarrow +\infty} P([X_n = 0])$. Alors $\pi_0 = 1$ si et seulement si $m \leq 1$ (c'est-à-dire $G'_Y(1) \leq 1$) et, si $m > 1$, π_0 est la plus petite solution positive de $G(s) = s$.

Preuve : $\pi(n)_0 = P([X_n = 0]) = G_n(0)$ si $X_0 = 1$.

Si $X_n = 0$, alors $X_{n+1} = 0$, donc $G_n(0) \leq G_{n+1}(0)$ et $(\pi(n)_0)_n$ est croissante et majorée par 1, donc converge vers une limite π_0 .

D'autre part, $\pi(n+1)_0 = G_{n+1}(0) = G(G_n(0)) = G(\pi(n)_0)$ et, comme G est continue, $\pi_0 = G(\pi_0)$ par passage à la limite.

Si s_0 est le plus petit point fixe de G positif, on a $0 \leq s_0$ donc, par croissance de G , $G(0) \leq G(s_0) = s_0$, et par applications successives de G , $G_n(0) \leq s_0$ et, par passage à la limite, $\pi_0 \leq s_0$, donc $\pi_0 = s_0$ car π_0 est un point fixe positif de G .

□

4.2 Processus permanent à 1 type

Le processus de ramification traité précédemment est limité en ce sens que les instants de génération sont fixes. Bien que certains phénomènes, en particulier des essais expérimentaux, la généalogie ou la génétique, s'adaptent à cette situation, la plupart des processus reproductifs naturels se produisent de façon continue dans le temps. Il est donc intéressant d'établir une version à temps continu des processus de ramification.

Soit X_t le nombre d'individus à l'instant t : $(X_t)_{t \geq 0}$ est un processus de Markov continu.

On pose $p_{kj}(t) = P^{[X_u=k]}([X_{u+t} = j])$. On a alors

$$p_{11}(h) = 1 + a_{11}h + o(h) \text{ et } p_{1j}(h) = a_{1j}h + o(h) \text{ si } j \neq 1,$$

avec $a_{11} \leq 0$, $a_{1j} \geq 0$ pour $j \neq 1$ et $\sum_j a_{1j} = 0$. On posera, pour simplifier, $a_{1j} = a_j$.

On a $P^{[X_t=1]}([X_{t+h} \neq 1]) = \left(\sum_{k \neq 1} a_k \right) h + o(h)$ et ainsi, un individu “vit” un temps exponentiel $\mathcal{E}(\sum_{k \neq 1} a_k)$ avant de se transformer (on dit qu’il se transforme au temps t , si à cet instant, on obtient 0 ou bien un nombre ≥ 2 d’individus). On peut ainsi considérer qu’un individu produit, à la fin de sa vie, un nombre D de descendants, avec D à valeurs dans $\mathbb{N} \setminus \{1\}$ ($D = 1$ n’est pas pris en compte car seules les variations de la taille de la population importent). La loi de D est donnée par

$$P([D = j]) = \frac{a_j}{\sum_{k \neq 1} a_k} \text{ pour } j \in \mathbb{N} \setminus \{1\}.$$

On a alors, si $j \neq 1$, $P^{[X_t=n]}([X_{t+h} = n - 1 + j]) = na_jh + o(h)$ car le changement ne peut venir que de l’un des n individus et celui-ci est alors remplacé par j nouveaux individus.

Remarque : Si $a_0 = \mu$, $a_1 = -\lambda - \mu$, $a_2 = \lambda$ et $a_k = 0$ pour $k \geq 3$, on retrouve un processus de naissance et de mort à croissance et décroissance linéaires.

On pose $u(s) = \sum_{j=0}^{+\infty} a_j s^j$: u est parfois appelée fonction génératrice instantannée. Alors :

Théorème : On a les équations aux dérivées partielles suivantes :

$$(4.1) \begin{cases} \partial_2 G(s, \tau) = u(G(s, \tau)) \\ G(s, 0) = s \end{cases} \quad \text{et} \quad (4.2) \begin{cases} \partial_2 G(s, t) = u(s) \partial_1(G(s, t)) \\ G(s, 0) = s \end{cases} .$$

Preuve : On part de la relation fondamentale $G(s, t + \tau) = G(G(s, \tau), t)$ et du développement limité, pour une courte période h :

$$G(s, h) = \sum_{j=0}^{+\infty} s^j p_{1j}(h) = s + u(s)h + o(h).$$

• En prenant $t = h$, on a alors :
 $G(s, \tau + h) = G(G(s, \tau), h) = G(s, \tau) + u(G(s, \tau))h + o(h)$, puis :

$$\partial_2 G(s, \tau) = u(G(s, \tau)).$$

• En prenant $\tau = h$, on a également :
 $G(s, t + h) = G(s + u(s)h + o(h), t) = G(s, t) + u(s)h \partial_1 G(s, t) + o(h)$, puis :

$$\partial_2 G(s, t) = u(s) \partial_1 G(s, t).$$

□

On en déduit en particulier :

Théorème : Le nombre moyen d'individus à l'instant t en partant de 1 est :

$$\mathbb{E}^{[X_0=1]}(X_t) = e^{u'(1)t}$$

Preuve : En dérivant la première équation de (4.2) par rapport à s , on obtient :

$$\partial_1 \partial_2 G(s, t) = u(s) \partial_1^2 G(s, t) + u'(s) \partial_1 G(s, t)$$

puis, avec $s = 1$, et en utilisant $u(1) = 0$,

$$\partial_1 \partial_2 G(s, t) = \partial_2 \partial_1 G(1, t) = u'(1) \partial_1 G(1, t).$$

Or on a $m(t) = \partial_1 G(1, t)$, donc $m'(t) = \partial_2 \partial_1 G(1, t) = u'(1)m(t)$, puis $m(t) = Ce^{u'(1)t}$ et, comme $m(0) = 1$, on obtient $C = 1$ d'où le résultat. □

Interprétation :

- si $u'(1) < 0$, alors $m_t \rightarrow 0$ (c'est l'analogie de $m < 0$) ;
- si $u'(1) = 0$, alors $m_t = 1$ pour tout t ;
- si $u'(1) > 0$, alors $m_t \rightarrow +\infty$.

Probabilité d'extinction :

Théorème : Soit $\pi_0 = \lim_{t \rightarrow +\infty} P([X_t = 0])$. Alors $\pi_0 = 1$ si et seulement si $u'(1) \leq 0$; π_0 est la plus petite solution positive de $u(s) = 0$.

Preuve : $\pi_0(t) = P([X_t = 0]) = G(0, t)$ si $X_0 = 1$.

Si $X_t = 0$, alors pour $\tau \geq t$, $X_\tau = 0$ et $\pi_0(t) \leq \pi_0(\tau)$ donc $t \mapsto \pi_0(t)$ est une fonction croissante et majorée par 1 ; elle admet une limite π_0 en $+\infty$.

$$\pi_0 = \lim_{t \rightarrow +\infty} G(0, t) = \lim_{t \rightarrow +\infty} G(0, t + \tau) = \lim_{t \rightarrow +\infty} G(G(0, t), \tau) = G(\pi_0, \tau)$$

par continuité de G . Ainsi, $G(\pi_0, \tau)$ est indépendant de τ et $\partial_2 G(\pi_0, \tau) = 0$.

Par la relation (4.2), on a alors $u(\pi_0) \partial_1 G(\pi_0, \tau) = 0$ pour tout τ . En particulier, pour $\tau = 0$, $G(s, 0) = \mathbb{E}^{[X_0=1]}(s^{X_0}) = s$, donc $\partial_1 G(s, 0) = 1$ et $u(\pi_0) = 0$.

Si s_0 est la plus petite solution de $u(s) = 0$ dans $[0, 1]$, alors $G(s_0, \tau)$ est indépendant de τ , car $\partial_2 G(s_0, \tau) = 0$. Donc $G(s_0, \tau) = G(s_0, 0) = s_0$.

De plus, $s \mapsto G(s, t) = \sum_j s^j p_{1,j}(t)$ est croissante, donc $G(0, \tau) \leq G(s_0, \tau) = s_0$ pour tout τ , donc, par passage à la limite quand $\tau \rightarrow +\infty$, $\pi_0 \leq s_0$, donc $\pi_0 = s_0$ car $\pi_0 \in [0, 1]$ et $u(\pi_0) = 0$. □

4.3 Processus discret à 2 types

Dans une population, on considère une caractéristique permettant de classer les individus en 2 types (par exemple, les gauchers et les droitiers chez les humains). Chaque individu de type (1) peut donner naissance à des individus de type (1) et de type (2) et chaque individu de type (2) peut donner naissance à des individus de type (1) et de type (2).

À chaque instant n , on s'intéresse, non seulement à la taille de la population, mais aussi au nombre d'individus de type (1) et de type (2).

On suppose qu'un individu de type (1) génère, à la fin d'une période donnée, un nombre aléatoire $Y^{(1)}$ de descendants de type (1), un nombre aléatoire $Y^{(2)}$ de descendants de type (2), et qu'un individu de type (2) génère, à la fin d'une même période, un nombre aléatoire $Z^{(1)}$ de descendants de type (1), un nombre aléatoire $Z^{(2)}$ de descendants de type (2). On suppose également, comme pour les processus à 1 type :

- que les individus agissent indépendamment les uns des autres ;
- que la durée de vie de chaque individu est la même.

On note $X_n = (X_n^{(1)}, X_n^{(2)})$ le nombre d'individus à la n -ième génération.

Si $X_n = (k_1, k_2)$, alors $X_{n+1}^{(i)} = \left(Y_1^{(i)} + \dots + Y_{k_1}^{(i)} \right) + \left(Z_1^{(i)} + \dots + Z_{k_2}^{(i)} \right)$ pour $i \in \{1, 2\}$. On a alors

$$p_{(k_1, k_2)(j_1, j_2)}(n) = P^{[X_n = (k_1, k_2)]}([X_{n+1} = (j_1, j_2)]).$$

On pose alors $e_1 = (1, 0)$, $e_2 = (0, 1)$ et pour $i \in \{1, 2\}$, $p_{(j_1, j_2)}^{(i)} = p_{e_i, (j_1, j_2)}(1)$ (probabilité qu'un individu de type (i) ait j_1 rejetons de type (1) et j_2 rejetons de type (2)).

On définit également :

$$G^{(i)}(s_1, s_2) = \mathbb{E}^{[X_0 = e_i]} \left(s_1^{X_1^{(1)}} s_2^{X_1^{(2)}} \right) = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} p_{(j_1, j_2)}^{(i)} \text{ et } G_n^{(i)}(s_1, s_2) = \mathbb{E}^{[X_0 = e_i]} \left(s_1^{X_n^{(1)}} s_2^{X_n^{(2)}} \right).$$

Théorème :

$$\mathbb{E}^{[X_0 = (k_1, k_2)]} \left(s_1^{X_n^{(1)}} s_2^{X_n^{(2)}} \right) = \left(G_n^{(1)}(s_1, s_2) \right)^{k_1} \left(G_n^{(2)}(s_1, s_2) \right)^{k_2} = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} p_{(k_1, k_2)(j_1, j_2)}(n).$$

La fonction génératrice détermine entièrement la loi et permet, en particulier, d'en déduire la taille moyenne de la population à la n -ième génération.

Pour cela, on pose $M = (m_{ij})$, où m_{ij} désigne le nombre moyen de rejetons de type (j) d'un individu de type (i) , c'est-à-dire :

$$m_{ij} = \mathbb{E}^{[X_0 = e_i]}(X_1^{(j)}) = \partial_j G^{(i)}(1, 1).$$

On a alors :

Théorème : $\mathbb{E}^{[X_r = (k_1, k_2)]}(X_{n+r}^{(1)}, X_{n+r}^{(2)}) = (k_1, k_2) M^n$.

Probabilité d'extinction :

Théorème : Soit $\pi_0 = (\pi_0^{(1)}, \pi_0^{(2)})$ où $\pi_0^{(i)} = \lim_{n \rightarrow +\infty} P^{[X_0 = e_i]}([X_n = (0, 0)])$ et soit ρ la valeur propre de plus grande valeur absolue de $M = (m_{ij})$. Alors $\pi_0 = (1, 1)$ si et seulement si $\rho \leq 1$ et si $\rho > 1$, π_0 est la plus petite solution positive de $\begin{cases} s_1 = G^{(1)}(s_1, s_2) \\ s_2 = G^{(2)}(s_1, s_2) \end{cases}$.

4.4 Processus permanent à 2 types

Dans cette partie, on généralise à la fois le passage du discret au continu, et le passage de 1 type à 2 types. Les notations, bien que lourdes, demeurent cohérentes.

Ainsi, on notera $X_t^{(i)}$ le nombre d'individus de type (i) à l'instant t , $e_1 = (1, 0)$, $e_2 = (0, 1)$. On note également :

$$p_{(k_1, k_2)(j_1, j_2)}(t) = P^{[X_u = (k_1, k_2)]}([X_{u+t} = (j_1, j_2)]) \text{ et } p_{e_i, (j_1, j_2)}(t) = p_{(j_1, j_2)}^{(i)}(t).$$

On suppose que $p_{(j_1, j_2)}^{(i)}(h) = \delta_{e_i, (j_1, j_2)} + a_{(j_1, j_2)}^{(i)}h + o(h)$ avec $a_{e_i}^{(i)} \leq 0$, $a_{(j_1, j_2)}^{(e_i)} \geq 0$ sinon, et $\sum_{j_1, j_2} a_{(j_1, j_2)}^{(e_i)} = 0$, c'est-à-dire que chaque individu de type (i) donne naissance, pendant $[t, t + h[$ à j_1 rejetons de types (1) et j_2 rejetons de type (2) avec la probabilité

$$p_{(j_1, j_2)}^{(i)}(h) = \delta_{e_i, (j_1, j_2)} + a_{(j_1, j_2)}^{(i)}h + o(h).$$

On pose encore $u^{(i)}(s_1, s_2) = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} a_{(j_1, j_2)}^{(i)}$ les fonctions génératrices instantannées, et $G^{(i)}(s_1, s_2, t) = \mathbb{E}^{[X_0=e_i]} \left(s_1^{X_t^{(1)}} s_2^{X_t^{(2)}} \right) = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} p_{(j_1, j_2)}^{(i)}(t)$. On a alors :

$$\mathbb{E}^{[X_0=(k_1, k_2)]} \left(s_1^{X_t^{(1)}} s_2^{X_t^{(2)}} \right) = \left(G^{(1)}(s_1, s_2, t) \right)^{k_1} \left(G^{(2)}(s_1, s_2, t) \right)^{k_2} = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} p_{(k_1, k_2)(j_1, j_2)}(t).$$

On a ici encore, du fait que (X_t) est un processus de Markov, la relation fondamentale :

Théorème : $G^{(i)}(s_1, s_2, t + \tau) = G^{(i)} \left(G^{(1)}(s_1, s_2, \tau), G^{(2)}(s_1, s_2, \tau), t \right)$

dont on peut déduire, exactement comme dans le cas à 1 type, les équations aux dérivées partielles suivantes :

$$(1) \begin{cases} \partial_3 G^{(i)}(s_1, s_2, \tau) = u^{(i)} \left(G^{(1)}(s_1, s_2, \tau), G^{(2)}(s_1, s_2, \tau) \right) \\ G^{(i)}(s_1, s_2, 0) = s_i \end{cases} .$$

$$(2) \begin{cases} \partial_3 G^{(i)}(s_1, s_2, t) = u^{(1)}(s_1, s_2) \partial_1 G^{(i)}(s_1, s_2, t) + u^{(2)}(s_1, s_2) \partial_2 G^{(i)}(s_1, s_2, t) \\ G^{(i)}(s_1, s_2, 0) = s_i \end{cases} ;$$

Exemples

a) Processus de ramification avec immigration

On considère ici que $P^{[X_t=0]}([X_{t+h} = j]) = \delta_{0,j} + b_j h + o(h)$ avec $b_0 \leq 0$ et les autres $b_j \geq 0$. Plus précisément, on note :

- b_j le taux d'immigration de j individus ;
- a_j le taux de naissance de j individus.

Pour résoudre ce genre de problème, on considère 2 types d'individus : les réels et les fictifs, et on s'intéressera évidemment à la taille de la population réelle. (Les fictifs constituent la population qui nous envoie les immigrants. On peut considérer qu'elle est constituée d'un seul individu, pouvant produire un nombre quelconque de rejetons). On pose alors :

$$a_{j_1, j_2}^{(1)} = \begin{cases} 0 & \text{si } j_2 \neq 0 \\ a_{j_1} & \text{si } j_2 = 0 \end{cases} \quad \text{et} \quad a_{j_1, j_2}^{(2)} = \begin{cases} 0 & \text{si } j_2 \neq 1 \\ b_{j_1} & \text{si } j_2 = 1 \end{cases} .$$

On en déduit, si $u(s) = \sum_j a_j s^j$ et $v(s) = \sum_j b_j s^j$:

$$u^{(1)}(s_1, s_2) = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} a_{j_1, j_2}^{(1)} = \sum_{j_1} s_1^{j_1} a_{j_1} = u(s_1) ;$$

$$u^{(2)}(s_1, s_2) = \sum_{j_1, j_2} s_1^{j_1} s_2^{j_2} a_{j_1, j_2}^{(2)} = \sum_{j_1} s_1^{j_1} s_2 b_{j_1} = s_2 v(s_1).$$

b) Remplacement d'une particule de durée de vie de loi gamma $\gamma(\lambda, 2)$ par 2 particules

Cet exemple permet de résoudre un problème particulier de processus permanents pour lesquels la durée de vie de chaque individu n'est pas de loi exponentielle (processus qui n'est plus markovien). Pour se ramener à ce que l'on connaît, on introduit 2 phases indépendantes et successives dans la vie de chaque particule, chaque phase suivant la loi exponentielle $\mathcal{E}(\lambda)$.

$X_t^{(i)}$ désigne ici le nombre de particules dans la phase i à l'instant t .

On s'intéresse évidemment à $X_t^{(1)} + X_t^{(2)}$.

Pour simplifier, on prend $\lambda = 1$. On a alors

$$a_{j_1, j_2}^{(1)} = \begin{cases} -1 & \text{si } (j_1, j_2) = (1, 0) \\ 1 & \text{si } (j_1, j_2) = (0, 1) \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad a_{j_1, j_2}^{(2)} = \begin{cases} -1 & \text{si } (j_1, j_2) = (0, 1) \\ 1 & \text{si } (j_1, j_2) = (2, 0) \\ 0 & \text{sinon} \end{cases} .$$

On en déduit $\boxed{u^{(1)}(s_1, s_2) = s_2 - s_1}$ et $\boxed{u^{(2)}(s_1, s_2) = s_1^2 - s_2}$.

Chapitre 5

Premières notions sur les files d'attente

5.1 Introduction

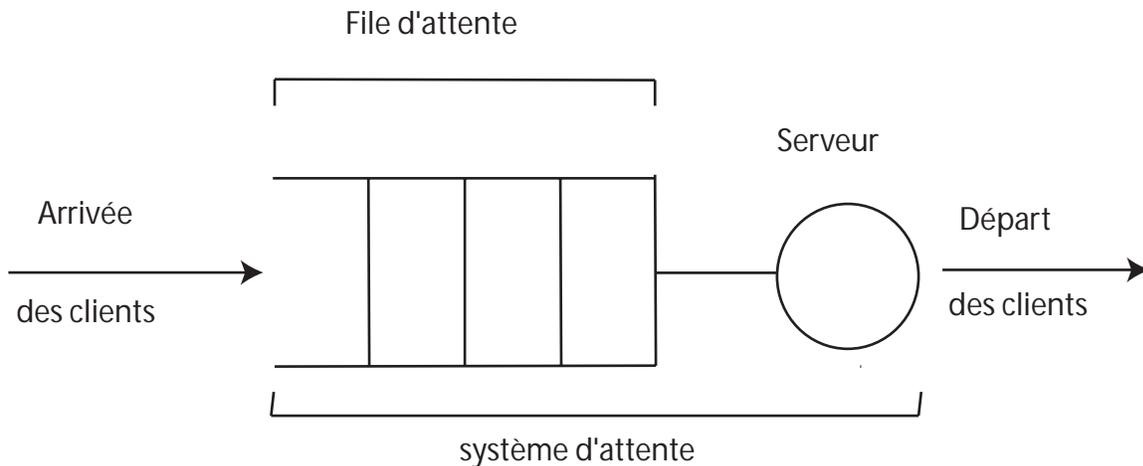
Les files d'attente peuvent être considérées comme un phénomène caractéristique de la vie contemporaine. On les rencontre dans les domaines d'activité les plus divers (guichet de poste, trafic routier, central téléphonique, atelier de réparation,...). L'étude mathématique des phénomènes d'attente constitue un champ d'application important des processus stochastiques.

On parle de phénomène d'attente chaque fois que certaines unités appelées "clients" se présentent d'une manière aléatoire à des "stations" afin de recevoir un service dont la durée est généralement aléatoire.

L'objectif de ce cours est d'étudier la structure et de calculer des valeurs caractéristiques permettant de décrire les performances de tels systèmes.

5.2. La file simple

Une file simple (ou station) est un système constitué d'un ou plusieurs serveurs et d'un espace d'attente. Les clients arrivent de l'extérieur, patientent éventuellement dans la file d'attente, reçoivent un service, puis quittent la station. Afin de spécifier complètement une file simple, on doit caractériser le processus d'arrivée des clients, le temps de service ainsi que la structure et la discipline de service de la file d'attente.



5.2.1 Processus d'arrivée

L'arrivée des clients à la station sera décrite à l'aide d'un processus stochastique de comptage $(N_t)_{t \geq 0}$. [On notera indifféremment N_t ou bien $N(t)$].

Si A_n désigne la variable aléatoire mesurant l'instant d'arrivée du $n^{\text{ième}}$ client dans le système, on aura ainsi : $A_0 = 0$ (par convention) et $A_n = \inf\{t ; N_t = n\}$.

Si T_n désigne la variable aléatoire mesurant le temps séparant l'arrivée du $(n - 1)$ ^{ième} client et du n ^{ième} client, on a alors : $T_n = A_n - A_{n-1}$.

Définition : Un processus de comptage $(N_t)_{t \geq 0}$ est un *processus de renouvellement* si et seulement si les variables aléatoires (T_n) sont des variables indépendantes et identiquement distribuées. La loi décrivant le temps d'interarrivée suffit alors à caractériser le processus de renouvellement.

La plupart du temps, l'arrivée des clients à une file simple est supposée décrite par un processus de renouvellement. Le processus d'arrivée le plus simple et le plus couramment employé est le processus de Poisson. C'est un processus de renouvellement qui est tel que les interarrivées sont distribuées selon une loi exponentielle.

On note λ le taux des arrivées : $1/\lambda$ est l'intervalle moyen entre deux arrivées consécutives.

5.2.2 Temps de service

Considérons tout d'abord une file à serveur unique.

On note D_n la variable aléatoire mesurant l'instant de départ du n ^{ième} client du système et Y_n la variable aléatoire mesurant le temps de service du n ^{ième} client (temps séparant le début et la fin du service).

Un instant de départ correspond toujours à une fin de service, mais ne correspond pas forcément à un début de service. Il se peut en effet qu'un client qui quitte la station laisse celle-ci vide. Le serveur est alors inoccupé jusqu'à l'arrivée du prochain client.

On considèrera uniquement des stations dont les temps de service consécutifs sont décrits par des variables Y_n indépendantes et identiquement distribuées (i.i.d.).

On note μ le taux de service : $1/\mu$ est la durée moyenne de service.

La distribution du temps de service la plus simple à étudier est la distribution exponentielle. Cependant, la propriété "sans mémoire" de la loi exponentielle fait que celle-ci n'est généralement pas très réaliste pour modéliser les phénomènes réels. On est donc souvent obligé de recourir à d'autres distributions de service.

5.2.3 Structure et discipline de la file

Nombre de serveurs

Une station peut disposer de plusieurs serveurs en parallèle. Soit C le nombre de serveurs. Dès qu'un client arrive à la station, soit il y a un serveur de libre et le client entre instantanément en service, soit tous les serveurs sont occupés et le client se place dans la file en attente de libération d'un des serveurs. La plupart du temps, les serveurs sont supposés identiques (ils possèdent donc la même distribution) et indépendants les uns des autres.

Une station particulière est la station IS (infinite servers) dans laquelle le nombre de serveurs est infini. Cette station ne comporte donc pas de file d'attente. Dès qu'un client s'y présente, il trouve en effet instantanément un serveur disponible et entre donc directement en service. Elle permet de représenter des systèmes pour lesquels le nombre de serveurs est toujours supérieur

au nombre de clients qui peuvent s'y trouver.

Capacité de la file

La capacité de la file à accueillir des clients en attente de service peut être finie ou infinie. Soit K la capacité de la file (incluant le ou les clients en service). Une file à capacité illimitée vérifie $K = +\infty$. Lorsque la capacité de la file est limitée et qu'un client arrive alors que cette dernière est pleine, le client est perdu.

Discipline de service

La discipline de service détermine l'ordre dans lequel les clients sont rangés dans la file et y sont retirés pour recevoir un service. Les disciplines les plus courantes sont :

- FIFO (first in, first out) ou FCFS (first come first served) ou PAPS (premier arrivé, premier servi) : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs. Dans la première, le premier client arrivé sera le premier à quitter la file alors que dans la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui. En français, le terme PAPS comporte une ambiguïté, puisqu'il ne peut différencier une file "premier arrivé, premier servi" d'une file "premier arrivé, premier sorti".

- LIFO (last in, first out) ou LCFS (last come, first served) ou DAPS (dernier arrivé, premier servi). Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). À nouveau, les disciplines LIFO et LCFS ne sont équivalentes que pour une file monoserveur.

- RANDOM (aléatoire) Le prochain client qui sera servi est choisi aléatoirement dans la file d'attente :

- Round-Robin (cyclique). Tous les clients de la file d'attente entrent en service à tour de rôle, effectuant un quantum Q de leur temps de service et sont replacés dans la file, jusqu'à ce que leur service soit totalement accompli. Cette discipline de service a été introduite afin de modéliser des systèmes informatiques ;

- PS (Processor Sharing). C'est le cas limite de la distribution Round-Robin lorsque le quantum de temps Q tend vers 0. Tous les clients sont servis en même temps, mais avec une vitesse inversement proportionnelle au nombre de clients simultanément présents. Si le taux du serveur est égal à μ et qu'à un instant donné il y a n clients à la station, tous les clients sont donc servis simultanément avec un taux $\frac{\mu}{n}$ (Attention, dire que les n clients sont servis simultanément ne signifie absolument pas qu'ils seront libérés simultanément)

5.2.4 Notations de Kendall

La notation de Kendall normalise la description d'une file simple :

$$T/Y/C/K/m/Z$$

avec

T : distribution d'interarrivée

Y : distribution de service

C : nombre de serveurs

K : capacité de la file
 m : population des usagers
 Z : discipline de service.

Lorsque les trois derniers éléments de la notation de Kendall ne sont pas précisés, il est sous-entendu que $K = +\infty$, $m = +\infty$, et $Z = \text{FIFO}$. Le paramètre m précise le nombre maximum d'usagers susceptibles d'arriver dans la file, cette dernière étant "plongée" dans un monde fermé contenant m clients.

5.2.5 Notion de classes de clients

Une file d'attente peut être parcourue par différentes classes de clients. Ces différentes classes se distingueront par :

- des processus d'arrivée différents ;
- des temps de service différents ;
- un ordonnancement dans la file d'attente fonction de leur classe.

Pour définir une file multiclassées, il faut définir pour chaque classe de clients le processus d'arrivée et la distribution du temps de service associés. Il faut également préciser comment les clients des différentes classes s'ordonnent dans la file d'attente. Cela permet d'introduire de nouvelles disciplines de service dites à *priorité* : PR (préemption) et HOL (hold on line). Par exemple, dans une file parcourue par deux classes de clients, les clients de la classe 1 seront alors toujours "devant" les clients de la classe 2. Pour caractériser une discipline de service prioritaire, il faut de plus préciser si le service est préemptible : lorsqu'un client de classe 2 est en service et qu'un client de classe 1 arrive à la file, est-ce que ce dernier doit attendre que le client de classe 2 ait terminé son service (sans préemption : HOL) ou est-ce que le client de la classe 2 est instantanément remis dans la file (avec préemption : PR) afin de laisser le serveur disponible pour le client de classe 1 ? Dans ce dernier cas, il faut encore distinguer deux sous-cas. Est-ce que le client de classe 2 relégué mémorise le temps de service accompli ou pas ?

5.3. Les réseaux de files d'attente

Un réseau de files d'attente est un ensemble de files simples (stations) interconnectées. Soit M le nombre de stations du réseau.

5.3.1 Les réseaux ouverts

Dans un réseau de files d'attente ouvert, les clients arrivent de l'extérieur, circulent dans le réseau à travers les différentes stations, puis quittent le réseau. Le nombre de clients pouvant se trouver à un instant donné dans un réseau ouvert n'est donc pas limité. Afin de spécifier complètement un réseau ouvert, il faut bien sûr caractériser chaque station, mais également le processus d'arrivée des clients et le routage (cheminement) des clients dans le réseau.

Processus d'arrivée

Le processus d'arrivée des clients dans le réseau sera décrit, comme pour une file simple, à l'aide d'un processus de renouvellement (et sera donc caractérisé par la distribution du temps d'interarrivée). Si l'arrivée des clients suit un processus de Poisson, les interarrivées sont exponentielles et sont caractérisées par un unique paramètre : le taux d'arrivée λ . Dans le cas d'un

processus d'arrivée non poissonien, ce paramètre reste intéressant, puisqu'il indique le nombre moyen de clients qui arrivent dans le système par unité de temps, mais devient insuffisant pour caractériser parfaitement l'arrivée des clients.

Il faut en plus préciser, lorsqu'un client arrive dans le réseau, à quelle file il se rend. On caractérisera la plupart du temps le routage d'entrée de façon probabiliste : soit p_{0i} la probabilité pour qu'un client qui arrive se rende à la station i (les probabilités p_{0i} sont bien sûr telles que $\sum_{i=1}^M p_{0i} = 1$). Si les arrivées dans le système sont poissonniennes de taux λ , on montre que le processus d'arrivée des clients (venant uniquement de l'extérieur) à la station i est poissonien de taux $p_{0i}\lambda$.

Routage des clients

Lorsqu'un client termine son service à une station, il faut préciser où ce client va se rendre : soit à une autre station, soit à l'extérieur (le client quitte alors le réseau). À nouveau, le routage des clients est très souvent caractérisé de façon probabiliste : soit p_{ij} la probabilité pour qu'un client qui quitte la station i se rende à la station j et soit p_{i0} la probabilité pour qu'un client qui quitte la station i quitte le système. Les p_{ij} sont tels que $\sum_{j=0}^M p_{ij} = 1$.

Il existe cependant d'autres types de routages :

- le routage vers la file la plus courte (routage dynamique) : un client quittant une station choisira, parmi toutes les destinations possibles, la station qui comporte le moins de clients ;
- le routage cyclique (routage déterministe) : les clients quittant une station choisiront à tour de rôle chacune des stations parmi toutes les destinations possibles.

5.3.2 Les réseaux fermés

Dans un réseau de files d'attente fermé, les clients sont en nombre constant. Soit N le nombre total de clients du système. Il n'y a donc pas d'arrivée ni de départ de clients. La spécification d'un réseau fermé se réduit donc à celle des différentes stations et à celle du routage des clients.

Par un mécanisme de routage probabiliste, on définit p_{ij} la probabilité qu'un client qui quitte la station i se rende à la station j . Les p_{ij} sont tels que $\sum_{j=1}^M p_{ij} = 1$.

5.3.3 Les réseaux multiclassés

Comme pour les files simples, les réseaux de files d'attente peuvent être parcourus par différentes classes de clients. Soit R le nombre de classes de clients. Ces différentes classes se distingueront par :

- des processus d'arrivée différents (si le réseau est ouvert)
- des comportements différents à chaque station (service et discipline de service)
- des routages différents dans le réseau.

On est alors amené à caractériser pour chaque classe r :

- pour un réseau ouvert, le processus d'arrivée (pour un processus d'arrivée poissonien, il suffit alors de donner le taux d'arrivée λ_r des clients de classe r) ;
- pour un réseau fermé, le nombre total N_r de clients de classe r ;

- le routage de clients. Si on se limite aux routages probabilistes, on définit p_{rij} la probabilité pour qu'un client de classe r qui quitte la station i se rende à la station j . (Si i ou j est égal à 0, cela fait la différence à l' "extérieur" d'un réseau ouvert.)

La notion de réseaux multiclassés nous permet d'introduire la notion de réseau mixte qui est un réseau ouvert vis à vis de certaines classes et fermé vis à vis des autres classes.

On peut également autoriser certains clients à changer de classe lors de leur cheminement dans le réseau. On définit alors $p_{ri,sj}$ la probabilité pour qu'un client de classe r qui quitte la station i se rende à la station j et se transforme en un client de classe s .

5.3.4 Les réseaux de files d'attente à capacité limitée

Les différentes stations du réseau peuvent avoir des capacités limitées. Lorsqu'une file est pleine, plus aucun client ne peut y entrer. Cela introduit des blocages dans les autres stations amonts et éventuellement des pertes de clients à l'entrée du système (si celui-ci est ouvert).

On distingue principalement deux types de blocage : le blocage avant service et le blocage après service.

Dans un blocage avant service (ou blocage de type "réseau de communication"), un client voulant commencer son service à une station donnée doit tout d'abord s'assurer qu'il y a une place de libre dans la station de destination. Si c'est le cas, son service commence. Dans le cas contraire, le serveur de la station est bloqué et le client doit attendre la libération d'une place en aval avant de commencer son service.

Dans un mécanisme de blocage après service (ou blocage de type "système de production"), un client commence sans attendre son service dès l'instant où le serveur est disponible. Ce n'est qu'à la fin de son service qu'un blocage peut survenir. Si la station de destination est pleine, le client reste au niveau du serveur qui se trouve alors bloqué, jusqu'à ce qu'une place se libère en aval.

5.3.5 Les réseaux de files d'attente ouverts à contrainte de population

Certains réseaux de files d'attente, bien qu'étant des modèles ouverts, peuvent être soumis à une limite supérieure sur le nombre total de clients pouvant s'y trouver simultanément. Cette "contrainte de population" implique que le réseau n'est ni réellement un modèle ouvert, puisque le nombre de clients qui peuvent s'y trouver est limité, ni réellement un réseau fermé, puisque le nombre total de clients dans le système n'est pas constant. On parlera de "modèle ouvert à contrainte de population". Lorsqu'un client arrive dans le réseau alors que celui-ci est plein (la contrainte de population est atteinte), deux cas peuvent être envisagés. Soit le client est "rejeté", ce qui rejoint le modèle de la section précédente, soit le client est "mémorisé" et se place en attente dans une file externe (généralement FIFO). Par la suite, on ne s'intéressera qu'au cas où le client est mémorisé.

Un système ouvert à contrainte de population est souvent modélisé à l'aide d'un formalisme de type "sémaphore". Une file de "jetons" contenant initialement N jetons est alors associée à la file externe des clients. Lorsqu'un client arrive alors qu'il reste un jeton de libre, il prend le jeton et entre instantanément dans le système. Il conserve alors le jeton pendant tout son séjour dans le système et le libère dès qu'il quitte le système. Le jeton revient alors instantanément dans la file des jetons et devient à nouveau disponible pour un autre client. Lorsqu'un client arrive alors qu'il n'y a aucun jeton de libre, il se place dans la file externe (des clients) en attente de libération d'un jeton. Le nombre initial N de jetons impose donc une limite supérieure sur le nombre total de clients pouvant se trouver simultanément dans le système.

5.4. Quelques exemples de systèmes d'attente

- T/Y/C : 1 file et C stations offrant le même service ; le client qui attend va à la première qui se libère.

Exemple : toilettes d'un lieu public, guichets d'une poste...

- T/Y/∞ : infinité de serveurs ; pas d'attente.

Exemple : lignes téléphoniques.

- 2 stations différentes, l'une plus rapide que l'autre.

Si les 2 stations sont libres, le client va dans l'une avec la probabilité p et dans l'autre avec la probabilité $1 - p$.

→ Si $p = \frac{1}{2}$, clients de passage : ils ne voient pas la différence.

→ Si $p = 1$, pour tous les clients, la rapidité prime.

→ On a en fait $p \in [0, 1]$: la rapidité peut ne pas être le seul critère (qualité du service, jolie ou gentille caissière...)

- Systèmes à perte T/Y/C/C pas d'attente

Exemple : Central téléphonique où les appels non traités sont rejetés.

- Systèmes T/Y/C/K où $K > C$: longueur de la file limitée, dépendant de la capacité.

Exemple : Station service, salle d'attente...

- Systèmes fermés : nombre constant de clients.

Exemple : m machines en marche ou en panne, C ouvriers pour les réparer, n en panne.

→ $m > C$: cas le plus fréquent ;

→ $m < C$: peu intéressant car il y aurait des ouvriers jamais occupés ;

→ $m = C$: un ouvrier par machine.

- Réseaux : plusieurs stations offrent des services différents, chacune ayant sa file.

→ réseau ouvert : on peut venir de l'extérieur ou repartir à l'extérieur ; il se peut que chaque station offre le même service mais chacune à sa file.

Exemple : Remontées mécaniques au ski, caisses de grandes surfaces.

→ réseau fermé : pas d'échange avec l'extérieur.

Exemple : Palettes dans un atelier.

5.5 Paramètres de performances opérationnels

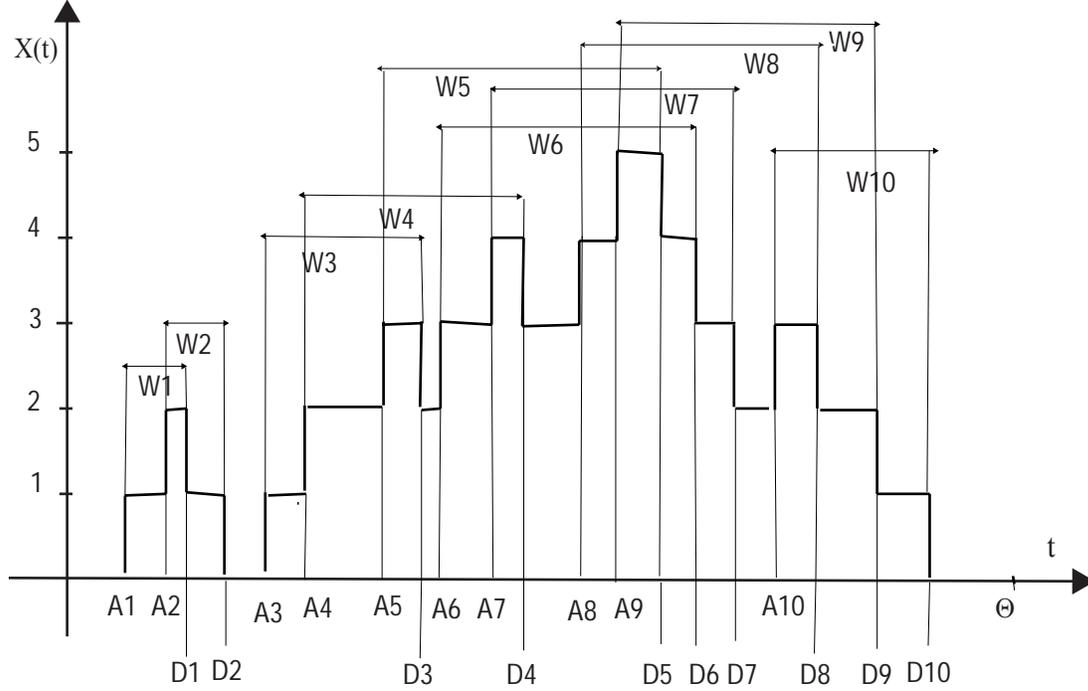
5.5.1 Paramètres de performances en régime transitoire

Considérons le comportement du système sur une période de temps donnée, par exemple entre $t = 0$ et $t = \Theta$. Soit X_t le nombre total de clients dans le système à l'instant t . S'intéresser au comportement du système sur l'intervalle de temps $[0, \Theta]$ revient à considérer le "régime transitoire" du système.

Définissons les "paramètres opérationnels" suivants :

W_k : temps de séjour du $k^{\text{ième}}$ client dans le système : $W_k = D_k - A_k$.

Θ : temps total de l'observation.



$T(n, \Theta)$: temps total pendant lequel le système contient n clients ; on a bien sûr :

$$\sum_{n \geq 0} T(n, \Theta) = \Theta.$$

$P(n, \Theta) = \frac{T(n, \Theta)}{\Theta}$: proportion de temps pendant laquelle le système contient n clients.

$\alpha(\Theta)$: nombre de clients arrivant dans le système pendant la période $[0, \Theta]$.

$\delta(\Theta)$: nombre de clients quittant le système pendant la période $[0, \Theta]$.

À partir de ces quantités, on définit les paramètres de performances en régime transitoire suivants :

Débit moyen d'entrée : Le débit moyen d'entrée est le nombre moyen de clients arrivés dans le système par unité de temps. Sur la période d'observation $[0, \Theta]$, c'est donc :

$$d_e(\Theta) = \frac{\alpha(\Theta)}{\Theta} \quad (5.1)$$

Débit moyen de sortie : Le débit moyen de sortie est le nombre moyen de clients ayant quitté le système par unité de temps. Sur la période d'observation $[0, \Theta]$, c'est donc :

$$d_s(\Theta) = \frac{\delta(\Theta)}{\Theta} \quad (5.2)$$

Nombre moyen de clients : Le nombre moyen de clients présents dans le système est la moyenne temporelle de X_t (ou $X(t)$) sur la période d'observation $[0, \Theta]$, c'est donc l'aire sous la courbe de X_t :

$$L(\Theta) = \frac{1}{\Theta} \sum_{n=0}^{+\infty} nT(n, \Theta) = \sum_{n=0}^{+\infty} nP(n, \Theta) \quad (5.3)$$

Temps moyen de séjour : Le temps moyen de séjour d'un client dans le système est, par définition, la moyenne arithmétique des temps de séjour des clients arrivés dans le système pendant l'intervalle de temps $[0, \Theta]$:

$$W(\Theta) = \frac{1}{\alpha(\Theta)} \sum_{k=1}^{\alpha(\Theta)} W_k \quad (5.4)$$

Taux d'utilisateur U : Pour une file simple comportant un unique serveur, en plus des paramètres précédents, débit moyen d'entrée, débit moyen de sortie, nombre moyen de clients, temps moyen de séjour, on peut définir un paramètre de performances supplémentaire : le taux d'utilisation du serveur. Il est défini comme la proportion de temps pendant laquelle le serveur est occupé sur l'intervalle de temps $[0, \Theta]$:

$$U(\Theta) = \sum_{n=1}^{+\infty} P(n, \Theta) = 1 - P(0, \Theta) \quad (5.5)$$

Cas des réseaux de files d'attente

Pour un réseau de files d'attente, on peut considérer les paramètres de performances du réseau tout entier : débit moyen d'entrée dans le réseau, temps moyen de séjour dans le réseau. Notons que ces paramètres n'ont d'intérêt que pour un réseau ouvert (ou mixte). En effet, si le réseau est fermé, le débit moyen d'entrée et le débit moyen de sortie sont nuls, le nombre de clients est en permanence égal à la population du réseau (N) et le temps moyen de séjour est égal à la durée d'observation Θ (infinie en régime permanent). Pour un réseau multiclassés ouvert (ou mixte), on pourra s'intéresser aux paramètres de performances, par classe ou toutes classes confondues.

On peut également considérer les paramètres de performances pour chacune des stations du réseau : débit moyen d'entrée dans la station i (d_{ei}), débit moyen de sortie de la station i (d_{si}), nombre moyen de clients dans la station i (L_i), temps moyen de séjour dans la station i (W_i). À nouveau, pour un réseau multiclassés, on pourra s'intéresser aux paramètres de performances de chaque classe ou toutes classes confondues.

5.5.2 Paramètres de performances en régime permanent

Toutes les quantités précédentes définissent les performances du système en régime transitoire (au bout d'un temps Θ fini). En régime permanent, on s'intéressera à l'existence et aux valeurs (éventuelles) des limites lorsque Θ tend vers l'infini de tous ces paramètres :

$$d_e = \lim_{\Theta \rightarrow +\infty} d_e(\Theta) ; d_s = \lim_{\Theta \rightarrow +\infty} d_s(\Theta)$$

$$L = \lim_{\Theta \rightarrow +\infty} L(\Theta) ; W = \lim_{\Theta \rightarrow +\infty} W(\Theta)$$

$$U = \lim_{\Theta \rightarrow +\infty} U(\Theta)$$

5.5.3 Stabilité

Définition : Un système est *stable* si et seulement si le débit moyen asymptotique de sortie des clients du système est égal au débit moyen d'entrée des clients dans le système :

$$\lim_{\Theta \rightarrow +\infty} d_s(\Theta) = \lim_{\Theta \rightarrow +\infty} d_e(\Theta) = d$$

D'après les relations précédentes, cela implique que le nombre total de clients arrivés dans le système pendant l'intervalle $[0, \Theta]$, $\alpha(\Theta)$ ne soit pas croître plus rapidement que le nombre total de clients ayant quitté le système $\delta(\Theta)$, lorsque Θ tend vers l'infini :

$$\lim_{\Theta \rightarrow +\infty} \frac{\delta(\Theta)}{\alpha(\Theta)} = 1.$$

Exemple : Afin de bien comprendre la notion de stabilité d'une file simple, nous allons considérer l'exemple d'une file $D/D/1$ (les interarrivées et les services sont déterministes, ce qui signifie que les clients arrivent exactement toutes les $t_a = \frac{1}{\lambda}$ unités de temps et restent tous en service pendant un temps $t_s = \frac{1}{\mu}$).

Considérons tout d'abord le cas où $t_a > t_s$ ($\lambda < \mu$), par exemple : $t_a = 10$ s, soit $\lambda = 0,1$ clients/seconde, $t_s = 5$ s, soit $\mu = 0,2$ clients/seconde.

Sur la période d'observation $[0, 100[$, le débit moyen de sortie $d_s = \frac{\delta(\Theta)}{\Theta} = \frac{10}{100} = 0,1$ est bien égal au débit moyen d'entrée $d_e = \lambda$. La file est stable : lorsque t tend vers l'infini, le nombre de clients dans la file X_t reste fini (et dans ce cas égal en alternance à 0 ou à 1).

Considérons maintenant le cas où $t_a < t_s$ ($\lambda > \mu$), par exemple : $t_a = 10$ s, soit $\lambda = 0,1$ clients/seconde, $t_s = 20$ s, soit $\mu = 0,05$ clients/seconde.

Sur la période d'observation $[0, 100[$, le débit moyen de sortie $d_s = \frac{\delta(\Theta)}{\Theta} = \frac{5}{100} = 0,05$ est différent du débit moyen d'entrée $d_e = \lambda$. La file est instable : lorsque t tend vers l'infini, le nombre de clients dans la file, X_t tend vers l'infini.

Le problème est exactement le même si, au lieu de considérer des temps déterministes, on considère des temps aléatoires (et donc si l'on considère une file $G/G/1$, pour laquelle la loi d'interarrivée et la loi de service ont des distributions générales quelconques). Les clients s'accumulent dans la file et le serveur n'arrive jamais à "rattraper" son retard. Ce phénomène se produit lorsque $t_a < t_s$ (donc lorsque $\lambda > \mu$) et est connu sous le nom d'instabilité.

5.5.4 Ergodicité

L'ergodicité est une notion très importante dans le domaine des processus stochastiques.

D'un côté, l'analyse opérationnelle s'intéresse à une évolution particulière d'un système entre deux instants $t = 0$ et $t = \Theta$. Nous avons vu que faire tendre Θ vers l'infini et considérer les limites de tous les paramètres de performances opérationnels, revient à s'intéresser au régime permanent du système. En fait, cela revient à s'intéresser au régime permanent d'une évolution particulière du système. Il est alors possible d'étudier différentes évolutions du système. La question que l'on se pose immédiatement est bien sûr : toutes ses réalisations ont-elles le même

comportement asymptotique ? En d'autres termes, tous les paramètres de performances considérés ont-ils la même limite quelle que soit l'évolution du système considérée ?

D'un autre côté, l'analyse stochastique va associer au système des variables aléatoires et des processus stochastiques :

A_k : variable aléatoire mesurant l'instant d'arrivée du $k^{\text{ième}}$ client dans le système ;

D_k : variable aléatoire mesurant l'instant de départ du $k^{\text{ième}}$ client du système ;

W_k : variable aléatoire mesurant le temps de séjour du $k^{\text{ième}}$ client dans le système :

$$W_k = D_k - A_k ;$$

(α_t) : processus mesurant le nombre de clients arrivés dans le système à l'instant t .

(δ_t) : processus mesurant le nombre de clients ayant quitté le système à l'instant t .

(X_t) : processus stochastique mesurant le nombre de clients dans le système à l'instant t :

$$X_t = \alpha_t - \delta_t.$$

$\pi_n(t)$: probabilité pour que le système contienne n clients à l'instant t : $\pi_n(t) = P([X_t = n])$.

On peut alors, comme cela a été fait dans le cadre de l'analyse opérationnelle, calculer tous les paramètres de performance stochastiques, en régime transitoire et en régime permanent. Le nombre moyen de clients présents dans le système à l'instant t se calcule, par exemple, de la façon suivante :

$$L(t) = \sum_{n=0}^{+\infty} n\pi_n(t).$$

La question que l'on se pose est alors : comment se relient les paramètres de performances stochastiques aux paramètres de performances opérationnels ? Par exemple, si l'on s'intéresse à caractériser combien de temps un individu passe, en moyenne dans sa vie, à dormir, on s'intéresse bien à la proportion de temps pendant laquelle il dort et non à la probabilité qu'à un instant quelconque, il soit en train de dormir ! Par ailleurs, les techniques de simulation étudient une réalisation particulière du processus et fournissent donc des paramètres de performances opérationnels.

La notion d'ergodicité nous permet de définir une classe de systèmes pour laquelle toutes les réalisations particulières de l'évolution du système sont asymptotiquement et statistiquement identiques, c'est-à-dire :

Définition : Un système est ergodique si et seulement si, quelle que soit la réalisation particulière étudiée du processus stochastique :

$$\lim_{\Theta \rightarrow +\infty} \sum_{n=0}^{+\infty} n^k P(n, \Theta) = \lim_{t \rightarrow +\infty} \sum_{n=0}^{+\infty} n^k \pi_n(t) \text{ pour tout } k = 1, 2, \dots$$

Cela implique que tous les paramètres de performances opérationnels en régime permanent (mesurés ou calculés par simulation à partir de n'importe quelle réalisation particulière du processus) sont égaux aux paramètres de performances stochastiques en régime permanent (obtenus à partir d'une étude analytique de performances). Même si cela n'est pas parfaitement rigoureux, on peut donc considérer cette propriété comme définition équivalente à l'ergodicité.

En choisissant comme paramètres de performances les proportions de temps passé par le système dans l'état n et les probabilités associées pour que le système contienne n clients, on obtient que si le système est ergodique :

$$\lim_{\Theta \rightarrow +\infty} P(n, \Theta) = \lim_{t \rightarrow +\infty} \pi_n(t)$$

Dans un système ergodique, on pourra donc confondre en régime permanent, proportions de temps passé dans un état et probabilités d'être dans cet état. On parlera indifféremment de "probabilités en régime permanent", de "probabilités à l'équilibre" ou de "probabilités stationnaires". Tous les paramètres de performances fournis par une analyse stochastique de notre système seront égaux aux paramètres de performances opérationnels, c'est-à-dire à ceux que l'on pourrait "observer" sur n'importe quelle réalisation particulière (suffisamment longue) de l'évolution de notre système. On parlera, de la même façon, des "performances stationnaires" du système.

Notons toutefois qu'il existe des systèmes non ergodiques. Par exemple, une chaîne de Markov non irréductible constitue un système non ergodique (certaines réalisations conduisent dans une sous-chaîne absorbante, d'autres dans une autre). Une chaîne périodique est un autre exemple d'un système non ergodique. Les probabilités stationnaires n'existent pas mais on est cependant capable de déterminer les proportions de temps passé dans chaque état de la chaîne. Notons finalement qu'un système instable n'est également pas un système ergodique puisque la limite lorsque Θ tend vers l'infini du nombre moyen de clients $L(\Theta)$ n'existe pas (est infinie).

5.5.5 La loi de Little

La loi de Little est une relation très générale qui s'applique à une grande classe de systèmes. Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'interarrivées, temps de service,...) n'est nécessaire. La seule condition d'application de la loi de Little est que le système soit stable. Le débit du système est alors indifféremment, soit le débit d'entrée, soit le débit de sortie : $d_s = d_e = d$. La loi de Little s'exprime telle que dans la propriété suivante :

Théorème : (Formule de Little) Le nombre moyen de clients, le temps moyen passé dans le système et le débit moyen d'un système stable en régime permanent se relient de la façon suivante :

$$L = W \times d$$

Preuve :

Considérons, dans un premier temps, une durée d'observation Θ qui est telle que le système est vide au début et à la fin de l'observation. Dans ces conditions, le nombre de clients qui ont quitté le système pendant $[0, \Theta]$ est égal au nombre de clients qui y sont arrivés : $\delta(\Theta) = \alpha(\Theta)$. Les quantités $L(\Theta)$ et $W(\Theta)$ peuvent alors être décrites de la façon suivante :

$$L(\Theta) = \sum_{n=0}^{+\infty} nP(n, \Theta) = \frac{1}{\Theta} \sum_{n=0}^{+\infty} nT(n, \Theta)$$

$$W(\Theta) = \frac{1}{\delta(\Theta)} \sum_{k=1}^{\delta(\Theta)} W_k$$

La formule de Little repose sur le résultat suivant :

$$\sum_{n=0}^{+\infty} nT(n, \Theta) = \sum_{k=1}^{\delta(\Theta)} W_k.$$

On peut démontrer formellement que les deux sommations intervenant dans cette égalité représentent deux façons de calculer l'aire sous la courbe de X_t .

Comme le débit $d(\Theta)$ du système est le rapport du nombre de clients sortis $\delta(\Theta)$ sur le temps moyen d'observations Θ , on déduit immédiatement des trois relations donnant $L(\Theta)$, $W(\Theta)$ et $d(\Theta)$ que : $L(\Theta) = W(\Theta) \times d(\Theta)$. On peut finalement s'affranchir de l'hypothèse que le système est vide au début et à la fin de l'observation, en faisant tendre Θ vers l'infini et en se rappelant que si le système est stable, $\lim_{\Theta \rightarrow +\infty} \frac{\delta(\Theta)}{\alpha(\Theta)} = 1$.

□

La loi de Little a une très grande importance dans l'analyse des systèmes de files d'attente. Elle permet de déduire l'une des trois quantités (L, W, d) en fonction de la connaissance des deux autres. Elle s'applique sous des formes très diverses. Considérons ici le cas d'une file simple comportant un unique serveur et montrons que la loi de Little peut s'appliquer de différentes façons.

On a vu que la loi de Little nous dit qu'il existe une relation entre le nombre moyen de clients dans la file (en attente ou en service) et le temps moyen total de séjour d'un client dans la file (temps d'attente + temps de service) :

$$L = W \times d \quad (5.6)$$

La loi de Little peut aussi s'appliquer en considérant uniquement l'attente dans la queue (sans le service). Elle permet alors de relier le nombre moyen de clients en attente L_q , au temps moyen d'attente d'un client avant service W_q , par la relation :

$$L_q = W_q \times d \quad (5.7)$$

Enfin, on peut appliquer la loi de Little en ne considérant que le serveur. Dans ce cas, elle relie le nombre moyen de clients en service L_S , au temps moyen de séjour d'un client dans le serveur qui n'est rien d'autre que le temps moyen de service $\frac{1}{\mu}$, par la relation

$$L_S = \frac{1}{\mu} \times d$$

Comme il n'y a jamais plus d'un client en service, L_S s'exprime simplement :

$$L_S(\Theta) = 0P(0, \Theta) + 1[P(1, \Theta) + P(2, \Theta) + P(3, \Theta) + \dots] = 1 - P(0, \Theta).$$

L_S n'est donc rien d'autre que le taux d'utilisation du serveur, U , défini comme étant la probabilité que le serveur soit occupé (ou la proportion de temps pendant laquelle le serveur est occupé). Ainsi, dans ce cas, la loi de Little s'écrit :

$$U = \frac{1}{\mu} \times d \quad (5.8)$$

On a obtenu trois relations en appliquant la loi de Little successivement au système entier, à la file d'attente seule et, enfin, au serveur seul. Ces trois relations ne sont bien sûr pas indépendantes. On peut en effet déduire l'une d'entre elles à partir des deux autres en remarquant que :

$$W = W_q + \frac{1}{\mu} \text{ et } L = L_q + L_S = L_q + U \quad (5.9)$$

Chapitre 6

File d'attente unique

L'étude mathématique d'un système d'attente se fait le plus souvent par l'introduction d'un processus stochastique approprié.

→ En premier lieu, on s'intéresse au nombre X_t de clients se trouvant dans le système à l'instant t . En fonction des quantités qui définissent la structure du système, on cherche à calculer :

- les probabilités d'état $\pi_n(t) = P([X_t = n])$ qui définissent le *régime transitoire* du processus $(X_t)_{t \geq 0}$;
- le *régime stationnaire* du processus, défini par

$$\pi_n = \lim_{t \rightarrow +\infty} \pi_n(t) = \lim_{t \rightarrow +\infty} P([X_t = n]).$$

→ À partir de la distribution stationnaire du processus $(X_t)_{t \geq 0}$, on pourra obtenir d'autres caractéristiques d'exploitation du système telles que :

- le nombre moyen L de clients dans le système ;
- le nombre moyen L_q de clients dans la file d'attente ;
- la durée d'attente moyenne W_q d'un client ;
- la durée de séjour moyenne W dans le système (attente + service) ;
- le taux d'occupation des postes de service ;
- le pourcentage de clients n'ayant pu être servis ;
- la durée d'une période d'activité, c'est-à-dire de l'intervalle de temps pendant lequel il y a toujours au moins un client dans le système...

Il faut toutefois constater que le calcul explicite du régime transitoire s'avère pénible, voire impossible, pour la plupart des modèles considérés. Mis à part certains modèles particulièrement faciles à traiter, nous nous contenterons donc par la suite de déterminer le régime stationnaire d'un phénomène d'attente.

Les résultats analytiques portant essentiellement sur des valeurs moyennes, en particulier, on n'a pas accès à leurs valeurs minimum et maximum qui peuvent être importantes par exemple dans l'optique du dimensionnement de zones de stockage. Dans ce cas, la connaissance des probabilités d'état π_n [notées aussi $\pi(n)$ dorénavant], peut être très utile.

6.1 Files d'attente markoviennes

Les files d'attentes markoviennes sont celles pour lesquelles les interarrivées et les durées de service sont exponentielles. Leur notation de Kendall sera de la forme $M/M/\dots$ (M comme markovien...)

6.1.1 Processus de naissance et de mort général

L'étude de ce processus a été faite au chapitre précédent. Rappelons qu'il est caractérisé par un nombre n d'entités qui évolue de la façon suivante :

- les arrivées et les départs d'entités obéissent à des lois exponentielles de taux respectifs $\lambda(n)$ et $\mu(n)$

- la probabilité pour que deux événements se produisent dans un intervalle de temps dt est négligeable (hypothèse de régularité : deux événements ne peuvent pas se produire en même temps).

Il y a transition vers un état voisin, soit par l'arrivée d'un client (naissance), soit par le départ d'un client (mort).

Si $\pi_n(t)$ est la probabilité pour qu'il y ait n clients dans le système à l'instant t , l'équation de Kolmogorov s'écrit, pour $n > 0$:

$$\pi_n(t + dt) = (1 - (\lambda_n + \mu_n) dt)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t) dt + \lambda_{n-1}\pi_{n-1}(t) dt + o(dt)$$

c'est-à-dire, en faisant tendre dt vers 0, pour $n > 0$:

$$\frac{d}{dt}\pi_n(t) = -(\lambda_n + \mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t) + \lambda_{n-1}\pi_{n-1}(t).$$

De la même façon, on obtient pour $n = 0$:

$$\frac{d}{dt}\pi_0(t) = -\lambda_0\pi_0(t) + \mu_1\pi_1(t).$$

Il est pratiquement impossible de calculer l'expression générale de $\pi_n(t)$, si ce n'est par simulation à partir des valeurs $\pi_n(0)$ qui caractérisent l'état initial du système.

Toutefois, si l'on suppose qu'un régime permanent parvient à s'établir, les probabilités deviennent indépendantes du temps et on a alors, comme on l'a vu au chapitre 3 :

$$\pi_n = \pi_0 \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}.$$

avec

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{+\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}}.$$

Le processus de Poisson est un cas particulier du processus de naissance et de mort pour lequel $\mu_n = 0$ et $\lambda_n = Cte = \lambda$ mais, dans ce cas, il n'y a pas de régime stationnaire.

Les équations différentielles s'écrivent alors :

$$\frac{d}{dt}\pi_0(t) = -\lambda\pi_0(t)$$

d'où $\pi_0(t) = e^{-\lambda t}$.

$$\frac{d}{dt}\pi_n(t) = -\lambda(\pi_n(t) - \pi_{n-1}(t))$$

dont la solution est $\pi_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$ (voir chapitre 2).

6.1.2 La file $M/M/1$

Cette file est caractérisée par une arrivée poissonnienne de taux λ et une durée de service exponentielle de taux μ .

On pose $\rho = \lambda/\mu$.

La file peut être considérée comme un processus de naissance et de mort, pour lequel :

$$\lambda_n = \lambda$$

$$\mu_n = \begin{cases} \mu & \text{si } n \neq 0 \\ 0 & \text{si } n = 0 \end{cases}$$

La probabilité d'état (en tenant compte que $\rho < 1$ pour qu'il y ait un régime permanent) est donnée par :

$$\begin{cases} \pi_n = \pi_0 \rho^n \\ \pi_0 = \frac{1}{\sum_{n=0}^{+\infty} \rho^n} = 1 - \rho \end{cases}$$

donc

$$\pi_n = (1 - \rho) \rho^n.$$

Tous les paramètres de performances sont calculés dans le cas où la file est stable ($\lambda < \mu$, c'est-à-dire $\rho < 1$) et pour le régime stationnaire de la file.

Débit d

Ici $d = \lambda$ car $\lambda_n = \lambda$ pour tout $n \geq 0$. Une autre façon de voir les choses est de remarquer que le service s'effectue avec un taux μ dans chaque état où le système contient au moins un client :

$$d = \text{Proba}([\text{file non vide}])\mu = \sum_{n=1}^{\infty} \pi_n \mu = [1 - \pi_0]\mu = \rho\mu = \lambda$$

On retrouve bien que si la file est stable, le débit moyen de sortie est égal au débit moyen d'entrée.

Taux d'utilisation du serveur U

Par définition, le taux d'utilisation est la probabilité pour que le serveur de la file soit occupé

$$U = \sum_{n=1}^{+\infty} \pi_n = 1 - \pi_0 = \rho = \frac{\lambda}{\mu}.$$

Nombre moyen de clients L

Le nombre moyen de clients se calcule à partir des probabilités stationnaires de la façon suivante :

$$\begin{aligned} L &= \sum_{n=1}^{+\infty} n\pi_n = \sum_{n=1}^{+\infty} n(1 - \rho)\rho^n = \rho(1 - \rho) \sum_{n=0}^{+\infty} (n+1)\rho^n = \rho(1 - \rho)(1 + 2\rho + 3\rho^2 + \dots) \\ &= \rho(1 - \rho) \frac{d}{d\rho} (\rho + \rho^2 + \rho^3 + \dots) = \rho(1 - \rho) \frac{d}{d\rho} \left(\frac{1}{1 - \rho} - 1 \right) \end{aligned}$$

soit

$$L = \frac{\rho}{1 - \rho}.$$

Temps moyen de séjour W

Ce paramètre est obtenu en utilisant la loi de Little :

$$W = \frac{L}{d} = \frac{1}{\mu(1-\rho)}$$

qui peut se décomposer en :

$$W = \frac{1}{\mu} + \frac{\rho}{\mu(1-\rho)}.$$

On en déduit le *temps moyen passé dans la file d'attente* W_q :

$$W_q = \frac{\rho}{\mu(1-\rho)}.$$

Nombre moyen de clients dans la file d'attente L_q :

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho}.$$

Preuve des formules de Little dans le cas $M/M/1$:

On cherche le temps moyen d'attente $W_q = \mathbb{E}(T_q)$ d'un individu : celui-ci est fonction du nombre de clients déjà présents lorsqu'il arrive.

Soit E_n l'événement "il y a n clients dans le système lorsque l'individu arrive". On a alors

$$W_q = \mathbb{E}(T_q) = \sum_{n=0}^{+\infty} \mathbb{E}(T_q/E_n)P(E_n)$$

avec $\mathbb{E}(T_q/E_n) = \frac{n}{\mu}$ (absence de mémoire de la loi exponentielle) et $P(E_n) = \pi_n = (1-\rho)\rho^n$ où $\rho = \frac{\lambda}{\mu}$.

On a alors $W_q = \sum_{n=1}^{+\infty} \frac{n}{\mu}(1-\rho)\rho^n = \frac{L}{\mu}$ et $W = W_q + \frac{1}{\mu} = \frac{L+1}{\mu}$.

Or $L_q = \sum_{n=1}^{+\infty} (n-1)\pi_n = \sum_{n=1}^{+\infty} n\pi_n - \sum_{n=1}^{+\infty} \pi_n = L - (1-\pi_0) = L - \rho$.

Comme $L = \sum_{n=1}^{+\infty} n\pi_n = \frac{\rho}{1-\rho}$, $L+1 = \frac{1}{1-\rho} = \mu W$, donc $L = \rho(L+1) = \rho\mu W$ et on a bien $L = \lambda W$.

De même, $L_q = L - \frac{\lambda}{\mu} = \lambda \left(W - \frac{1}{\mu} \right) = \lambda W_q$.

□

6.1.3 La file $M/M/1/K$

On considère un système à serveur simple identique à la file $M/M/1$ excepté que la capacité de la file d'attente est finie. On a donc toujours les hypothèses suivantes : le processus d'arrivée des clients dans la file est un processus de Poisson de taux λ et le temps de service d'un client est une variable aléatoire exponentielle de taux μ . Soit K la capacité de la file d'attente : c'est le nombre maximal de clients qui peuvent être présents dans le système, soit en attente, soit en service. Quand un client arrive alors qu'il y a déjà K clients présents dans le système, il est perdu. Ce système est connu sous le nom de file $M/M/1/K$. L'espace d'états E est maintenant fini : $E = \{0, 1, 2, \dots, K\}$. La capacité de la file étant limitée, même si les clients arrivent en moyenne beaucoup plus vite que ce que le serveur de la file est capable de traiter, dès que celle-ci

est pleine, les clients qui se présentent sont rejetés. Le nombre de clients dans la file ne peut donc jamais “partir” à l’infini. De plus, dès qu’un client est autorisé à entrer, il sortira un jour et son temps de séjour dans la file est fini, puisqu’il correspond au temps de service de tous les clients devant lui et que ce nombre est limité par K . Sur un temps très long, le débit de sortie sera donc bien égal au débit d’entrée, ce qui correspond bien à la stabilité inconditionnelle du système.

Le processus de naissance et de mort modélisant ce type de file d’attente est alors défini de la façon suivante :

$$\lambda_n = \begin{cases} \lambda & \text{si } n < K \\ 0 & \text{si } n = K \end{cases}$$

$$\mu_n = \begin{cases} \mu & \text{si } n \neq 0 \\ 0 & \text{si } n = 0 \end{cases}$$

L’intégration de l’équation récurrente permettant de calculer π_n se fait alors comme suit :

$$\pi_n = \pi_0 \rho^n \text{ pour } n \leq K$$

$$\pi_n = 0 \text{ pour } n > K$$

$$\pi_0 = \frac{1}{\sum_{n=0}^K \rho^n} = \frac{1 - \rho}{1 - \rho^{K+1}} \text{ si } \lambda \neq \mu \text{ (et } \frac{1}{K+1} \text{ si } \lambda = \mu).$$

Débit d : Le débit du système peut être calculé de deux manières équivalentes : soit en mesurant le taux de départ des clients en sortie du serveur, d_s , soit en mesurant le taux d’arrivée effectif des clients acceptés dans le système d_e . On s’attend bien sûr à obtenir l’égalité de ces deux débits.

Le débit en sortie du serveur est égal à μ dès l’instant où la file n’est pas vide :

$$d_s = \text{Proba}([\text{ file non vide }])\mu = \sum_{n=1}^K \pi_n \mu = [1 - \pi_0]\mu = \frac{\rho - \rho^{K+1}}{1 - \rho^{K+1}} \mu$$

Le débit effectif d’entrée dans la file est égal à λ dès l’instant qu’un client arrive lorsque la file n’est pas pleine :

$$d_e = \text{Proba}([\text{ file non pleine aux instants d’arrivée }])\lambda = \sum_{n=0}^{K-1} \pi_n \lambda = [1 - \pi_K]\lambda = \frac{1 - \rho^K}{1 - \rho^{K+1}} \lambda.$$

Puisque $\rho = \frac{\lambda}{\mu}$, on a bien $d_e = d_s = d$, où d est le débit moyen de la file (d’entrée ou de sortie) :

$$d = \frac{1 - \rho^K}{1 - \rho^{K+1}} \lambda$$

Notons que, lorsque K tend vers l’infini, on retrouve bien les résultats de la $M/M/1$, c’est-à-dire $d = \lambda$, à condition que $\rho < 1$, ce qui correspond à la condition de stabilité de la $M/M/1$.

Taux d’utilisation du serveur $U(K)$

$$U(K) = \sum_{n=1}^K \pi_n = 1 - \pi_0 = \frac{\rho - \rho^{K+1}}{1 - \rho^{K+1}} = \rho \frac{1 - \rho^K}{1 - \rho^{K+1}}$$

Ainsi, dans le cas d'une file à capacité limitée, le taux d'utilisation n'est plus égal à ρ . En effet, le taux d'utilisation est toujours égal au rapport du débit moyen d'entrée sur le taux moyen de service (loi de Little) : $U = \frac{d}{\mu}$. Mais ici d n'est plus égal à λ .

Remarquons que, lorsque $K \rightarrow +\infty$, $U(K)$ tend vers ρ si $\rho < 1$ et vers 1 si $\rho > 1$.

Nombre moyen de clients L

$$\begin{aligned} L &= \sum_{n=0}^K n\pi_n = \frac{1-\rho}{1-\rho^{K+1}} \sum_{n=0}^K n\rho^n = \frac{\rho(1-\rho)}{1-\rho^{K+1}} \sum_{n=1}^K n\rho^{n-1} \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \frac{d}{d\rho} \left(\frac{1-\rho^{K+1}}{1-\rho} - 1 \right) = \frac{\rho(1-\rho)}{1-\rho^{K+1}} \frac{1-(K+1)\rho^K + K\rho^{K+1}}{(1-\rho)^2} \\ &= \frac{\rho}{1-\rho} \frac{1-(K+1)\rho^K + K\rho^{K+1}}{1-\rho^{K+1}} \end{aligned}$$

À nouveau, lorsque K tend vers l'infini et $\rho < 1$, on retrouve les résultats de la $M/M/1$:

$$L = \frac{\rho}{1-\rho}$$

Temps moyen de séjour W

On considère ici le temps moyen de séjour d'un client effectivement admis dans la file d'attente. Cette quantité peut être obtenue par application de la loi de Little :

$$W = \frac{L}{d}$$

Cas particulier où $\lambda = \mu$: $\pi_n = \pi_0 = \frac{1}{K+1}$ pour $n \leq K$.

$$U(K) = 1 - \pi_0 = \frac{K}{K+1} \text{ et } d = \frac{K\lambda}{K+1}.$$

$$\text{Nombre moyen de clients dans la file d'attente : } L = \sum_{n=0}^K \frac{n}{K+1} = \frac{K}{2}.$$

$$\text{Temps moyen dans le système d'attente } W = \frac{L}{d} = \frac{K+1}{2\lambda}.$$

Preuve des formules de Little dans le cas $M/M/1/K$:

On a ici $\mathbb{E}(T_q) = \sum_{n=1}^{K-1} \frac{n}{\mu} \pi_n^* = \frac{1}{\mu} \left(\frac{L - K\pi_K}{1 - \pi_K} \right)$, avec $\pi_n^* = \frac{\pi_n}{1 - \pi_K}$ (client rejeté s'il y en a déjà K).

$$\begin{aligned} L - K\pi_K &= \rho \frac{1 - (K+1)\rho^K + K\rho^{K+1} - K\rho^{K-1}(1-\rho)^2}{(1-\rho)(1-\rho^{K+1})} \\ &= \rho \frac{1 - (K+1)\rho^K + K\rho^{K+1} - K\rho^{K-1}(1-2\rho+\rho^2)}{(1-\rho)(1-\rho^{K+1})} = \rho \frac{1 - K\rho^{K-1} + (K-1)\rho^K}{(1-\rho)(1-\rho^{K+1})} \\ 1 - \pi_K &= 1 - \frac{\rho^K(1-\rho)}{1-\rho^{K+1}} = \frac{1-\rho^{K+1}-\rho^K+\rho^{K+1}}{1-\rho^{K+1}} = \frac{1-\rho^K}{1-\rho^{K+1}} \end{aligned}$$

$$\begin{aligned}
W &= W_q + \frac{1}{\mu} = \frac{1}{\mu} \left[\frac{\rho(1 - K\rho^{K-1} + (K-1)\rho^K)}{(1-\rho)(1-\rho^K)} + \frac{1 - \rho - \rho^K + \rho^{K+1}}{(1-\rho)(1-\rho^K)} \right] \\
&= \frac{1}{\mu} \left[\frac{1 - (K+1)\rho^K + K\rho^{K+1}}{(1-\rho)(1-\rho^K)} \right] = \frac{L(1 - \rho^{K+1})}{\rho\mu(1 - \rho^K)} = \frac{L}{\lambda(1 - \pi_K)}
\end{aligned}$$

Or $d = \lambda \sum_{n=0}^{K-1} \pi_n = \lambda(1 - \pi_K)$ et on a bien $L = d \times W$.

De même, $L - L_q = \sum_{n=1}^K n\pi_n - \sum_{n=1}^K (n-1)\pi_n = \sum_{n=1}^K \pi_n = 1 - \pi_0$ donc, comme $\bar{\lambda} = \bar{\mu} = \mu(1 - \pi_0) = d$,

$$L_q = L - (1 - \pi_0) = d \times W - \frac{d}{\mu} = d \left(W - \frac{1}{\mu} \right) = d \times W_q.$$

□

6.1.4 La file $M/M/C$

On considère un système identique à la file $M/M/1$ excepté qu'il comporte C serveurs identiques et indépendants les uns des autres. On conserve les hypothèses : processus d'arrivée des clients poissonien de taux λ et temps de service exponentiel de taux μ (pour chacun des serveurs). Ce système est connu sous le nom de file $M/M/C$. L'espace d'états E est, comme pour la $M/M/1$ infini : $E = \{0, 1, 2, \dots\}$. On a un processus de naissance et de mort de taux :

$$\begin{aligned}
\lambda_n &= \lambda \\
\mu_n &= \begin{cases} 0 & \text{si } n = 0 \\ n\mu & \text{si } 0 < n < C \\ C\mu & \text{si } n \geq C \end{cases}
\end{aligned}$$

En effet, lorsque le processus est dans un état $n < C$, tous les clients sont en service et sont donc susceptibles de quitter la file. Pour passer de n clients à $n - 1$ clients en un temps dt , il faut qu'un des n clients termine son service et que les autres ne terminent pas le leur, ceci pouvant se produire pour le premier, le deuxième, ..., ou le n -ième client. Pour être précis, il faut également rajouter qu'aucun client n'arrive pendant ce temps dt . La propriété caractéristique de la loi exponentielle nous dit que la probabilité pour qu'un client termine son service en un temps dt est $\mu dt + o(dt)$, la probabilité pour qu'un client ne termine pas son service est donc $1 - \mu dt + o(dt)$ et la probabilité pour qu'aucun client n'arrive est $1 - \lambda dt + o(dt)$. La probabilité recherchée se calcule donc de la façon suivante :

$$p_{n,n-1}(dt) = \left(\sum_{j=1}^n (\mu dt + o(dt))(1 - \mu dt + o(dt))^{n-1} \right) (1 - \lambda dt + o(dt))$$

Un développement limité au premier ordre nous donne immédiatement que

$$p_{n,n-1}(dt) = n\mu dt + o(dt)$$

Le taux de transition de l'état n vers l'état $n - 1$ est donc égal à $n\mu$. De la même façon, lorsque $n \geq C$, seuls C clients sont en service et sont donc susceptibles de quitter la file, donc de faire passer le processus de l'état n à l'état $n - 1$. Le taux de transition correspondant est donc égal à $C\mu$. Dans tous les cas, une transition d'un état n vers un état $n + 1$ correspond à une arrivée

de client, soit en un temps dt , à une probabilité $\lambda dt + o(dt)$. Le taux de transition est donc égal à λ .

La condition de stabilité est ici $\lambda < C\mu$ et exprime le fait que le nombre moyen de clients qui arrivent à la file par unité de temps doit être inférieur au nombre moyen de clients que les serveurs de la file sont capables de traiter par unité de temps.

On peut calculer π_n comme suit :

$$\begin{cases} \pi_{n-1}\lambda = \pi_n n\mu & \text{pour } n = 1, \dots, C-1 \\ \pi_{n-1}\lambda = \pi_n C\mu & \text{pour } n = C, C+1, \dots \end{cases}$$

soit $\begin{cases} \pi_n = \frac{\rho}{n} \pi_{n-1} & \text{pour } n = 1, \dots, C-1 \\ \pi_n = \frac{\rho}{C} \pi_{n-1} & \text{pour } n = C, C+1, \dots \end{cases}$ où $\rho = \frac{\lambda}{\mu}$.

On peut alors exprimer toutes les probabilités en fonction de π_0 :

$$\begin{cases} \pi_n = \frac{\rho^n}{n!} \pi_0 & \text{pour } n = 1, \dots, C-1 \\ \pi_n = \frac{\rho^n}{C! C^{n-C}} \pi_0 & \text{pour } n = C, C+1, \dots \end{cases}$$

La condition de normalisation nous permet de calculer la probabilité π_0 , à condition bien sûr que cette série converge. On peut aisément vérifier que la condition de convergence de cette série est identique à la condition de stabilité de la file, soit $\lambda < C\mu$.

$$\pi_0 = \frac{1}{\sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C}{(C-1)!(C-\rho)}}$$

Lorsque $C = 1$, on retrouve bien les résultats de la file $M/M/1$:

$$\pi_n = (1 - \rho)\rho^n$$

Tous les paramètres de performances peuvent se calculer dans le cas où la file est stable ($\lambda < C\mu$) donc $\rho < C$).

Débit d

Le service s'effectue avec un taux $n\mu$ dans chaque état où le système contient moins de C clients et avec un taux $C\mu$ dans chaque état où le système contient plus de C clients :

$$d = \sum_{n=1}^{C-1} \pi_n n\mu + \sum_{n=C}^{+\infty} \pi_n C\mu$$

En remplaçant les expressions obtenues pour les probabilités π_n et π_0 , on retrouve bien que la file est stable, le débit moyen de sortie est égal au débit moyen d'entrée :

$$d = \lambda.$$

Pour la file $M/M/C$, il est plus simple (au niveau des calculs mis en jeu) de calculer d'abord le temps moyen de séjour et d'en déduire le nombre moyen de clients.

Temps moyen de séjour W

Le temps moyen de séjour d'un client se décompose en un temps moyen dans la file d'attente, plus un temps moyen de service. Il suffit alors d'appliquer la loi de Little à la seule file :

$$W = W_q + S = \frac{L_q}{d} + \frac{1}{\mu} = \frac{L_q}{\lambda} + \frac{1}{\mu}.$$

Il reste alors à calculer le nombre moyen de clients en attente dans la file, L_q :

$$\begin{aligned} L_q &= \sum_{n=C}^{+\infty} (n-C)\pi_n = \sum_{n=C}^{+\infty} (n-C) \frac{\rho^n}{C!C^{n-C}} \pi_0 = \frac{\rho^{C+1}}{C!C} \sum_{n=C}^{+\infty} (n-C) \left(\frac{\rho}{C}\right)^{n-C-1} \pi_0 \\ &= \frac{\rho^{C+1}}{C!C} \frac{1}{\left(1 - \frac{\rho}{C}\right)^2} \pi_0 = \frac{\rho^{C+1}}{(C-1)!(C-\rho)^2} \pi_0 \end{aligned}$$

On en déduit l'expression du temps moyen de séjour

$$W = \frac{\rho^C}{\mu(C-1)!(C-\rho)^2} \pi_0 + \frac{1}{\mu}$$

Nombre moyen de clients L

Le nombre moyen de clients s'obtient alors par application de la loi de Little à l'ensemble de la file :

$$L = W \times d = W \times \lambda = \frac{\rho^{C+1}}{(C-1)!(C-\rho)^2} \pi_0 + \rho$$

6.1.5 La file $M/M/\infty$

On considère un système composé d'un nombre illimité de serveurs identiques et indépendants les uns des autres. Dès qu'un client arrive, il rentre donc instantanément en service. Dans cette file particulière, il n'y a donc pas d'attente. On suppose toujours que le processus d'arrivée des clients est poissonien de taux λ et que les temps de service sont exponentiels de taux μ (pour tous les serveurs). Ce système est connu sous le nom de file $M/M/\infty$.

Comme cela a été fait pour la file $M/M/C$, on peut facilement démontrer que le taux de transition d'un état n quelconque vers l'état $n-1$ est égal à $n\mu$ et correspond au taux de sortie d'un des n clients en service. De même, le taux de transition d'un état n vers l'état $n+1$ est égal à λ et correspond au taux d'arrivée d'un client.

De façon intuitive, la capacité de traitement de la file est infinie puisque tout nouveau client se présentant à l'entrée de la file est instantanément traité. La condition de stabilité exprimant que "le nombre moyen de client arrivant à la file par unité de temps doit être inférieure à la capacité de traitement de la file" est donc toujours satisfaite.

Soit π_n la probabilité stationnaire d'être dans l'état n . Les équations d'équilibre nous donnent

$$\pi_{n-1}\lambda = \pi_n n\mu \text{ pour } n = 1, 2, \dots$$

soit $\pi_n = \frac{\rho}{n} \pi_{n-1}$ pour $n = 1, 2, \dots$, où $\rho = \frac{\lambda}{\mu}$.

On peut alors exprimer toutes les probabilités en fonction de π_0 :

$$\pi_n = \frac{\rho^n}{n!} \pi_0 \text{ pour } n = 1, 2, \dots$$

La condition de normalisation nous donne alors immédiatement π_0 :

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

Notons que la série $\sum_{n=0}^{+\infty} \frac{\rho^n}{n!}$ converge pour toutes valeurs de ρ (donc de λ et de μ), ce qui est cohérent avec la stabilité inconditionnelle de la file. On obtient finalement :

$$\pi_n = \frac{\rho^n}{n!} e^{-\rho} \text{ pour } n = 1, 2, \dots$$

Débit d

Le service s'effectue avec un taux $n\mu$ dans chaque état où le système contient n clients :

$$d = \sum_{n=1}^{+\infty} \pi_n n \mu = e^{-\rho} \sum_{n=1}^{+\infty} \frac{\rho^n}{(n-1)!} \mu = e^{-\rho} \rho e^{\rho} \mu = \rho \mu = \lambda.$$

On retrouve la stabilité inconditionnelle de la file.

Nombre moyen de clients L

$$L = \sum_{n=1}^{+\infty} n \pi_n = e^{-\rho} \sum_{n=1}^{+\infty} \frac{\rho^n}{(n-1)!} = e^{-\rho} \rho e^{\rho} = \rho$$

Temps moyen de séjour W

Intuitivement, le temps moyen passé dans le système est réduit au temps moyen de service, soit $\frac{1}{\mu}$. On peut redémontrer ce résultat en utilisant la loi de Little :

$$W = \frac{L}{d} = \frac{\rho}{\lambda} = \frac{1}{\mu}$$

6.2 Étude de la file $M/G/1$

6.2.1 Introduction

On revient à un système formé d'une file FIFO à capacité illimitée et d'un seul serveur. Le processus d'arrivée des clients dans la file est toujours supposé poissonien de taux λ mais, maintenant, le temps de service Y d'un client est distribué selon une loi qui n'est plus supposée exponentielle. Ce système est connu sous le nom de file $M/G/1$. En fait, on suppose implicitement que les services successifs sont indépendants les uns des autres et distribués selon la même loi (donc que les variables aléatoires mesurant le temps de service des différents clients sont i.i.d.). Il faudrait alors parler d'une file $M/GI/1$, "GI" faisant référence à des lois générales et indépendantes les unes des autres.

Une file simple comportant un unique serveur est stable si le nombre moyen de clients qui arrivent à la file par unité de temps est inférieur au nombre moyen de clients que le serveur de

la station est capable de traiter, soit $\lambda < \mu$ où $\mu = \frac{1}{m}$ est le taux moyen de service de la file (avec $m = \mathbb{E}(Y)$).

6.2.2 Analyse du régime permanent : Méthode de la chaîne de Markov incluse.

Pour étudier simplement ce système, le service n'étant plus exponentiel, il ne suffit donc plus de savoir qu'un client est en service, pour prédire quand ce service va se terminer. Il faut en plus savoir depuis combien de temps le service a commencé.

Par exemple, dans le cas d'un service déterministe de durée 10 secondes, si le service a commencé depuis 2 secondes, il se terminera exactement dans 8 secondes alors que, s'il a commencé depuis 9 secondes, il se terminera au bout d'une seconde.

Cette information est indispensable pour prédire l'évolution future de l'état du serveur, donc du système.

L'idée est de ne s'intéresser qu'à des instants particuliers de l'évolution du système $t_k = D_{k+}$, $k = 1, 2, \dots$, correspondant aux instants de fin de service (instants "juste après" le départ d'un client).

Considérons le processus $(N_k)_{k \geq 1} = (X_{t_k})_{k \geq 1}$ où X_{t_k} est le nombre de clients juste après le départ du $k^{\text{ième}}$ client.

Le processus $(N_k)_{k \geq 1}$ est une chaîne de Markov ergodique pouvant être facilement étudiée. En particulier, on pourra obtenir les probabilités $\pi_n(k) = P([N_k = n])$ pour que le départ du $k^{\text{ième}}$ client laisse derrière lui n clients, ainsi que $\pi_n = \lim_{k \rightarrow +\infty} \pi_n(k)$.

On va maintenant déterminer la matrice de transition $P = (q_{i,j})$ de la chaîne de Markov $(N_k)_{k \geq 1}$ dite *chaîne de Markov incluse* du processus $(X_t)_{t \geq 0}$.

On note α_c la probabilité que c clients arrivent pendant un temps de service. On a

$$\alpha_c = \int_0^{+\infty} P([c \text{ arrivées pendant } t] / Y = t) f_Y(t) dt = \int_0^{+\infty} \frac{(\lambda t)^c}{c!} e^{-\lambda t} f_Y(t) dt$$

En effet, les arrivées sont poissonniennes de taux λ . Si le temps de service était de durée constante t , la probabilité pour que c clients arrivent pendant un temps de service t serait égale à $\frac{(\lambda t)^c}{c!} e^{-\lambda t}$, par définition du processus de Poisson. Comme le temps de service est distribué selon une loi générale de densité f_Y , la durée de temps de service est égale à t à dt près avec une probabilité égale à $f_Y(t) dt$.

Notons alors $q_{i,j}$ la probabilité de transition de l'état i vers l'état j . On a :

$$\begin{cases} q_{0,j} = \alpha_j & \text{si } j \geq 0 \\ q_{i,j} = \alpha_{j-i+1} & \text{si } 1 \leq i \leq j+1 \\ q_{i,j} = 0 & \text{sinon.} \end{cases}$$

En effet,

→ si $N_k = 0$, $[N_{k+1} = j]$ correspond à l'arrivée de j clients pendant le service du $(k+1)^{\text{ième}}$ client ;

→ si $N_k = i$, on a $[N_{k+1} = j]$ si $N_{k+1} - N_k = j - i$, ce qui correspond à l'arrivée de $j - i + 1$ clients, car il ne faut pas oublier que le $(k+1)^{\text{ième}}$ client vient de partir.

La matrice de la chaîne de Markov incluse est donc :

$$P = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ 0 & \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ 0 & 0 & \alpha_0 & \alpha_1 & \alpha_2 & \cdots \\ 0 & 0 & 0 & \alpha_0 & \alpha_1 & \cdots \\ 0 & 0 & 0 & 0 & \alpha_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

On pourrait montrer que cette chaîne de Markov est irréductible, apériodique et récurrente positive si elle est stable (c'est-à-dire si $\lambda < \mu$). On peut alors affirmer que le vecteur $\vec{\pi}$ des probabilités stationnaires existe et est solution du système $\vec{\pi} = \vec{\pi}P$ (avec bien sûr $\sum_{n=0}^{+\infty} \pi_n = 1$).

La chaîne de Markov (N_k) est irréductible et admet donc une unique distribution stationnaire, qui est aussi distribution limite de X_t . Le théorème suivant permet, entre autre, de déterminer L :

Théorème : Si $\Pi(z) = \sum_{n=0}^{+\infty} \pi_n z^n$, si $A(z) = \sum_{n=0}^{+\infty} \alpha_n z^n$ et si $\rho = \frac{\lambda}{\mu}$, alors :

$$\Pi(z) = \frac{(1-\rho)A(z)(z-1)}{z-A(z)} \quad \text{et} \quad L = \rho + \frac{\rho^2 + \lambda^2 \text{var}(Y)}{2(1-\rho)}.$$

Preuve : La distribution stationnaire $\vec{\pi}$ doit vérifier $\vec{\pi} = \vec{\pi}P$ soit $\pi_j = \sum_{i=0}^{+\infty} \pi_i q_{i,j}$, ce qui s'écrit également

$$\pi_j = \alpha_j \pi_0 + \sum_{i=1}^{j+1} \alpha_{j-i+1} \pi_i = \alpha_j \pi_0 + \sum_{i=0}^{j+1} \alpha_{j-i+1} \pi_i - \alpha_{j+1} \pi_0.$$

Si l'on multiplie cette équation par z^j et si l'on somme sur j , on a :

$$\Pi(z) = \pi_0 A(z) + \frac{1}{z} \sum_{j=0}^{+\infty} a_{j+1} z^{j+1} - \frac{\pi_0}{z} (A(z) - \alpha_0),$$

où $a_j = \sum_{i=0}^j \alpha_{j-i} \pi_i$ et donc $\frac{1}{z} \sum_{j=0}^{+\infty} a_{j+1} z^{j+1} = \frac{1}{z} (A(z)\Pi(z) - \alpha_0 \pi_0)$. Ainsi,

$$\Pi(z) = \frac{\pi_0 A(z)(z-1)}{z-A(z)}.$$

On fait alors un développement limité de $A(z)$ à l'ordre 2 en 1 :

$$A(1+h) = A(1) + A'(1)h + \frac{1}{2}A''(1)h^2 + o(h^2)$$

puis

$$\begin{aligned} \Pi(1+h) &= \pi_0 h \frac{1 + A'(1)h + \frac{1}{2}A''(1)h^2 + o(h^2)}{1+h-1-A'(1)h-\frac{1}{2}A''(1)h^2+o(h^2)} = \pi_0 \frac{1 + A'(1)h + \frac{1}{2}A''(1)h^2 + o(h^2)}{1 - A'(1)h - \frac{1}{2}A''(1)h^2 + o(h^2)} \\ &= \frac{\pi_0}{1-A'(1)} [1 + A'(1)h + o(h)] \left[1 - \frac{A''(1)}{2(1-A'(1))}h + o(h) \right]^{-1} \\ &= \frac{\pi_0}{1-A'(1)} \left[1 + \left(A'(1) + \frac{A''(1)}{2(1-A'(1))} \right) h + o(h) \right]. \end{aligned}$$

On en déduit, comme $\Pi(1+h) = 1 + h\Pi'(1) + o(h) = 1 + Lh + o(h)$, que $\pi_0 = 1 - A'(1)$ et que $L = A'(1) + \frac{A''(1)}{2(1-A'(1))}$. Or

$$A(z) = \sum_{j=0}^{+\infty} \alpha_j z^j = \sum_{j=0}^{+\infty} z^j \int_0^{+\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} f_Y(t) dt = \int_0^{+\infty} e^{-\lambda t} e^{\lambda z t} f_Y(t) dt$$

donc $A'(1) = \lambda \int_0^{+\infty} t f_Y(t) dt = \lambda \mathbb{E}(Y) = \frac{\lambda}{\mu} = \rho$ et

$$A''(1) = \lambda^2 \int_0^{+\infty} t^2 f_Y(t) dt = \lambda^2 \mathbb{E}(Y^2) = \lambda^2 (\text{var}(Y) + \mathbb{E}(Y)^2) = \lambda^2 \text{var}(Y) + \rho^2,$$

ce qui donne bien le résultat. □

Détermination des autres paramètres de performances

Débit : $d = \lambda$

Taux d'utilisation du serveur U : $U = \sum_{n=1}^{+\infty} \pi_n = 1 - \pi_0 = \rho = \frac{\lambda}{\mu}$

Temps moyen de séjour W : d'après la formule de Little

$$W = \frac{1}{\mu} + \lambda \frac{\mathbb{E}(Y)^2 + \text{var}(Y)}{2(1 - \rho)}$$

Remarque : Tous les paramètres de performances moyens d , U , L , W ne dépendent donc que des deux premiers moments de la loi de service ainsi bien sûr que du taux d'arrivée λ , bien que les deux premiers moments d'une loi ne suffisent pas à la caractériser, (des variables aléatoires différentes pouvant, en effet avoir leurs deux premiers moments identiques).

6.2.3 Mise en oeuvre de l'analyse de la valeur moyenne

On calcule directement les performances moyennes de la file. Si L_q est le nombre moyen de clients dans la file d'attente seule :

$$L_q = \sum_{n=0}^{+\infty} n \text{Proba}([n \text{ clients dans la file}]) = \sum_{n=0}^{+\infty} n \pi_{n+1} = \sum_{n=1}^{+\infty} (n-1) \pi_n$$

Lorsqu'un client arrive dans le système :

- soit il trouve 0 client et son temps d'attente dans la file est nul ;
- soit il trouve n clients et son temps d'attente dans la file est : le temps moyen résiduel du client en service t_r + le temps de service des $n-1$ clients en attente devant lui dans la file.

Le temps résiduel du client en service est défini comme le temps qu'il reste au serveur, au moment où un client arrive dans la file, pour terminer son service. On en déduit l'expression du temps moyen d'attente dans la file :

$$W_q = \pi_0 \times 0 + \sum_{n=1}^{+\infty} \pi_n [t_r + (n-1)m]$$

En combinant les deux relations précédentes, et en appliquant la formule de Little à la file, on obtient :

$$W_q = (1 - \pi_0)t_r + m \sum_{n=1}^{+\infty} (n-1)\pi_n = (1 - \pi_0)t_r + mL_q = \rho t_r + m\lambda W_q = \rho t_r + \rho W_q$$

soit $W_q = \frac{\rho}{1 - \rho} t_r$.

Il ne reste plus qu'à calculer le temps moyen résiduel de service.

Propriété : Dans une file $M/G/1$, le temps moyen résiduel de service t_r "vu" par l'arrivée d'un client dans la file est

$$t_r = \frac{m}{2} \left(1 + \frac{\text{var}(Y)}{m^2} \right)$$

Preuve : Le processus d'arrivée des clients dans la file est Poissonien.

Vis-à-vis du client en service, un client arrive donc dans la file à un instant quelconque dans le temps et a plus de chance de "tomber" sur un service long que sur un service court. Plus précisément, la probabilité de tomber sur un service de longueur t (à dt près) est proportionnelle à t , ainsi bien sûr qu'à la fréquence avec laquelle un service de longueur t (à dt près) a lieu. On note Z la variable aléatoire mesurant la durée d'un service vue par l'arrivée d'un client (qui donc est différente de la variable aléatoire Y mesurant la durée d'un service). La probabilité $f_Z(t) dt$ s'obtient donc de la façon suivante :

$$f_Z(t) dt = K t f_Y(t) dt$$

où K est la constante de proportionnalité.

Afin d'obtenir la valeur de K , il suffit d'intégrer les deux côtés de l'équation :

$$1 = \int_0^{+\infty} f_Z(t) dt = K \int_0^{+\infty} t f_Y(t) dt = K m$$

où m est le temps moyen de service. On en déduit donc que $K = \frac{1}{m}$.

Dès l'instant qu'un client arrive et tombe sur un service de longueur t , on peut supposer qu'il a autant de chance de tomber au début, au milieu ou à la fin du service. En d'autres termes, la variable aléatoire mesurant le temps résiduel de service, conditionnée par le fait que le service est de longueur t (à dt près), est uniforme sur l'intervalle $[0, t]$. Le temps résiduel de service, toujours conditionné par le fait que le service est de longueur t est donc égal à $\frac{t}{2}$.

On en déduit le temps moyen résiduel de service (inconditionnel) :

$$t_r = \int_0^{+\infty} \frac{t}{2} f_Z(t) dt = \frac{1}{2m} \int_0^{+\infty} t^2 f_Y(t) dt = \frac{m_2}{2m}$$

où $m_2 = \mathbb{E}(Y^2) = m^2 + \text{var}(Y)$ est le moment d'ordre 2 de la loi de service. D'où le résultat attendu :

$$t_r = \frac{m}{2} \left(1 + \frac{\text{var}(Y)}{m^2} \right).$$

□

Remarque : Pour une file $M/M/1$, $\text{var}(Y) = \frac{1}{\mu^2} = m^2$ donc $t_r = m$, ce qui est logique, vu le caractère sans mémoire de la loi exponentielle. Pour une file $M/D/1$, $\text{var}(Y) = 0$ et donc $t_r = \frac{m}{2}$, ce qui est tout à fait intuitif, car si le service est toujours de durée égale à m , un client arrivant à un instant quelconque devra attendre, en moyenne, un temps $\frac{m}{2}$ avant que le service se termine.

6.3 La file $G/M/1$

On considère toujours un système formé d'une file FIFO à capacité illimitée et d'un seul serveur. Le processus d'arrivée des clients dans la file n'est, cette fois-ci, plus supposé poissonien. En revanche, le service d'un client est supposé exponentiel de taux μ . Ce système est connu sous le nom de file $G/M/1$. Comme pour la file $M/G/1$, il faudrait en fait parler d'une file $GI/M/1$ car on suppose implicitement que le processus d'arrivée des clients est un processus de renouvellement, donc que les interarrivées successives sont indépendantes les unes des autres et distribuées selon une même loi. Soit T la variable aléatoire mesurant le temps séparant l'arrivée de deux clients consécutifs. On notera f_T la densité de la loi de T , A^* sa transformée de Laplace et $m_k = \mathbb{E}(T^k)$ pour tout $k \geq 1$, avec $m_1 = m$.

On rappelle qu'une file simple est stable si le nombre moyen de clients qui arrivent à la file par unité de temps est inférieur au nombre moyen de clients que le serveur de la station est capable de traiter, soit $\lambda < \mu$ où $\lambda = \frac{1}{m}$ est le taux moyen d'arrivée des clients dans la file.

L'idée est ici de ne s'intéresser qu'aux instants d'arrivée des clients. Considérons le processus $(N_k) = (X_{A_k-})_{k \geq 1}$: nombre de clients "juste avant" l'arrivée du $k^{\text{ième}}$ client. Le processus $(N_k)_{k \geq 1}$ est une chaîne de Markov discrète pouvant être facilement étudiée. En particulier, on pourra obtenir les probabilités $\pi_n(k) = P([N_k = n])$ pour que le $k^{\text{ième}}$ client trouve en arrivant n clients dans la file. Si la chaîne est ergodique, on pourra également calculer les probabilités stationnaires aux instants d'arrivée : $\pi_n = \lim_{k \rightarrow +\infty} \pi_n(k)$.

On note β_c la probabilité pour que c clients terminent leur service pendant un temps d'arrivée. On a alors :

$$\beta_c = \int_0^{+\infty} \frac{(\mu t)^c}{c!} e^{-\mu t} f_T(t) dt$$

En effet, les services consécutifs étant des variables aléatoires exponentielles de taux μ et indépendantes les unes des autres, tant qu'il reste des clients dans la file, le processus de sortie des clients est donc un processus de Poisson de taux μ . Si le temps d'interarrivée était de durée constante t , la probabilité pour que c clients quittent la file pendant un temps t serait donc égale à $\frac{(\mu t)^c}{c!} e^{-\mu t}$. Comme le temps d'interarrivée est distribué selon une loi générale de densité de probabilité f_T , la durée séparant l'arrivée consécutive de deux clients est égale à t à dt près avec une probabilité $f_T(t) dt$.

La matrice de la chaîne de Markov incluse est alors :

$$P = \begin{pmatrix} \beta_1 & \beta_0 & 0 & 0 & 0 & \cdots \\ \beta_2 & \beta_1 & \beta_0 & 0 & 0 & \cdots \\ \beta_3 & \beta_2 & \beta_1 & \beta_0 & 0 & \cdots \\ \beta_4 & \beta_3 & \beta_2 & \beta_1 & \beta_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

En effet, si $N_k = i$, le $k^{\text{ième}}$ client a trouvé en arrivant i clients dans le système. Le client suivant trouvera alors $i+1$ clients si pendant l'interarrivée, aucun client n'a terminé son service, i clients dans la file si un client a terminé son service et, de façon générale $i+1-c$ clients dans la file si pendant le temps séparant l'arrivée du $k^{\text{ième}}$ client de celle du $(k+1)^{\text{ième}}$ client, c clients ont terminé leur service et quitté la file. Ainsi, si $q_{i,j}$ est la probabilité de transition de l'état i vers l'état j , on a $q_{i,j} = 0$ si $j > i+1$ et

$$q_{i,j} = \beta_{i+1-j} \text{ pour tout } j \leq i + 1$$

On pourrait alors montrer que le vecteur $\vec{\pi}$ des probabilités stationnaires existe et est solution du système $\vec{\pi} = \vec{\pi}P$. Ce système s'écrit :

$$\pi_k = \sum_{i=k-1}^{+\infty} \pi_i \beta_{i-k+1} \text{ pour } k = 0, 1, \dots$$

On a alors le résultat suivant :

Propriété : Les probabilités stationnaires π_n possèdent une distribution géométrique

$$\pi_n = (1 - \sigma)\sigma^n$$

où σ est l'unique solution strictement comprise entre 0 et 1 de l'équation

$$\sigma = A^*(\mu - \mu\sigma).$$

Preuve : On cherche une solution du système de la forme $\pi_n = K\sigma^n$ où K est une constante :

$$K\sigma^k = \sum_{i=k-1}^{+\infty} K\sigma^i \beta_{i-k+1}$$

En divisant les deux côtés de l'égalité par $K\sigma^{k-1}$, on obtient :

$$\sigma = \sum_{i=k-1}^{+\infty} \sigma^{i-k+1} \beta_{i-k+1} = \sum_{j=0}^{+\infty} \sigma^j \beta_j$$

En posant $T(z) = \sum_{k=0}^{+\infty} \beta_k z^k$ la transformée en z des probabilités β_k , on a immédiatement que σ est solution de l'équation $\sigma = T(\sigma)$. On admettra que cette équation possède une unique solution telle que $0 < \sigma < 1$.

Par ailleurs, un calcul similaire à celui effectué dans la preuve de la propriété 3 nous permet de relier $T(z)$ à la transformée de Laplace de la loi d'interarrivée, de la façon suivante : $T(z) = A^*(\mu - \mu z)$. σ est donc l'unique solution de l'équation $\sigma = A^*(\mu - \mu\sigma)$.

Il reste alors à déterminer l'expression de la constante K . La condition de normalisation des probabilités π_n nous donne immédiatement $K = 1 - \sigma$.

□

Pour calculer les probabilités stationnaires de la file $G/M/1$, il suffit donc de calculer la transformée de Laplace de la fonction densité de probabilité f_T de la loi générale d'interarrivée et de résoudre l'équation $\sigma = A^*(\mu - \mu\sigma)$. Notons que $\sigma = 1$ est toujours solution de cette équation (car la normalisation de f_T implique que $A^*(0) = 1$). Cela permet dans tous les cas de diminuer de un le degré de l'équation engendrée.

Nous allons appliquer cette relation à l'exemple particulier le plus simple d'une file $G/M/1$: celui d'une file $M/M/1$. Le temps d'interarrivée est distribué selon une loi exponentielle de taux λ . La transformée de Laplace de la distribution d'interarrivée est donc $A^*(s) = \frac{\lambda}{\lambda + s}$. On résout alors le système $\sigma = A^*(\mu - \mu\sigma)$:

$$\sigma = \frac{\lambda}{\lambda + \mu - \mu\sigma} \text{ soit } \mu\sigma^2 - (\lambda + \mu)\sigma + \lambda = 0$$

dont la seule racine strictement comprise entre 0 et 1 est $\sigma = \rho = \frac{\lambda}{\mu}$. On retrouve alors le résultat de la file $M/M/1$: $\pi_n = (1 - \rho)\rho^n$.

Les paramètres de performances sont toujours calculés lorsque la file est stable ($\lambda < \mu$) et pour le régime stationnaire.

Débit $d = \lambda$

Taux d'utilisation du serveur : $U = \mathbb{E}(Y) \times d = \frac{\lambda}{\mu}$ (loi de Little appliquée au serveur)

Temps moyen de séjour : $W = \frac{1}{\mu} + \frac{\sigma}{\mu(1 - \sigma)}$

Preuve directe : On calcule le temps moyen d'attente dans la file W_q .

Lorsqu'un client arrive, il trouve, avec une probabilité π_n , n clients dans le système, donc $n - 1$ clients en attente et un client en service. Il devra donc attendre que le client en service termine son service, puis que les $n - 1$ clients devant lui dans la file effectuent un service complet. La propriété sans mémoire de la loi exponentielle nous dit que le temps moyen résiduel de service est égal au temps moyen de service, soit $\frac{1}{\mu}$. W_q s'exprime donc :

$$W_q = \sum_{n=0}^{+\infty} n \frac{1}{\mu} \pi_n = \frac{1 - \sigma}{\mu} \sum_{n=0}^{+\infty} n \sigma^n = \frac{\sigma}{\mu(1 - \sigma)}.$$

Le temps moyen de séjour W est finalement égal à $W = W_q + \mathbb{E}(Y) = \frac{\sigma}{\mu(1 - \sigma)} + \frac{1}{\mu}$.

Nombre moyen de clients : $L = W \times d = \rho + \frac{\sigma\rho}{1 - \sigma}$ (par la formule de Little)

6.4 Extensions à la file $G/G/1$

Nous indiquons ici, sous forme d'exercice, une méthode pour traiter le cas général, méthode qui fournit également le régime transitoire. Malheureusement, les intégrales rencontrées sont la plupart du temps impossibles à calculer : il ne reste plus alors qu'à utiliser cette méthode avec l'aide d'un ordinateur...

Cet exercice, traite en particulier complètement le cas $M/M/1$, et même le cas d'une impatience "*a posteriori*" des clients (où les clients se sont donnés une limite τ_0 de temps dans le système à ne pas dépasser).

Pour résoudre cet exercice, il est nécessaire d'avoir quelques connaissances élémentaires sur la transformée de Laplace, dont nous rappelons d'abord les principales propriétés.

Rappel sur la transformée de Laplace :

Pour $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, et pour $p > 0$, on pose :

$$L(f)(p) = \bar{f}(p) = \int_0^{+\infty} e^{-px} f(x) dx.$$

Propriétés :

1) Si $f : x \mapsto 1$, $L(f)(p) = \frac{1}{p}$.

2) Si $f : x \mapsto e^{-\lambda x} g(x)$, $L(f)(p) = L(g)(p + \lambda)$; en particulier, avec $g = 1$, $L(f)(p) = \frac{1}{p + \lambda}$.

3) $L(f')(p) = -f(0) + pL(f)(p)$.

4) $L(f * g) = L(f)L(g)$, si $f * g(x) = \int_0^x f(u)g(x - u)du$.

Notations :

Pour le $k^{\text{ième}}$ client qui se présente dans le système, on note

- A_k l'instant de son arrivée ;
- D_k l'instant de son départ ;
- $\tau_k = A_k - A_{k-1}$ l'intervalle de temps entre la $(k - 1)^{\text{ième}}$ arrivée et la $k^{\text{ième}}$;
- W_{qk} le temps d'attente dans la file, éventuellement nul, du $k^{\text{ième}}$ client ;
- Y_k le temps de service du $k^{\text{ième}}$ client ;
- W_k le temps passé dans le système par le $k^{\text{ième}}$ client.

On suppose que la loi de τ_k est indépendante de k , de densité a , et que la loi de Y_k est indépendante de k , de densité v , de fonction de répartition H . On note F_k la fonction de répartition de W_{qk} et G_k la fonction de répartition de W_k .

1) Exprimer W_k en fonction de W_{qk} et de Y_k . En déduire que, pour tout $x > 0$,

$$G_k(x) = \int_0^x F_k(x - t)v(t)dt.$$

2) Exprimer W_{qk} en fonction de W_{k-1} et de τ_k . En déduire que, pour tout $x > 0$,

$$F_k(x) = \int_0^{+\infty} G_{k-1}(x + t)a(t)dt.$$

3) Déterminer $F_1(x)$ et montrer que les résultats précédents permettraient de déterminer $F_k(x)$ et $G_k(x)$ pour tout k . Écrire les équations vérifiées par F et G en régime permanent.

4) On se propose de résoudre ces équations dans le cas $M/M/1$.

i) Montrer que $F(x) = e^{\lambda x} F(0) - \lambda e^{\lambda x} \int_0^x e^{-\lambda t} G(t)dt$ pour tout $x > 0$, puis que

$$\bar{F}(p) = \frac{F(0) - \lambda \bar{G}(p)}{p - \lambda} \text{ et que } \bar{G}(p) = \frac{\mu \bar{F}(p)}{\mu + p}.$$

ii) En déduire que $\bar{F}(p) = \frac{p + \mu}{p(p + \mu - \lambda)} F(0)$, puis que $F(x) = \left(\mu - \lambda e^{-(\mu - \lambda)x} \right) \frac{F(0)}{\mu - \lambda}$. Montrer enfin que $F(x) = 1 - \frac{\lambda}{\mu} e^{-(\mu - \lambda)x}$ et que G est la fonction de répartition d'une loi exponentielle

de paramètre $\mu - \lambda$.

5) Dans le cas avec impatience “a posteriori”, on pose $\tilde{W}_k = W_{qk} + \tilde{Y}_k = \min(W_k, \tau_0)$. Exprimer \tilde{Y}_k en fonction de Y_k , de τ_0 et de W_{qk} . En déduire qu’en régime permanent :

$$G(x) = \begin{cases} \int_0^x H(x-t)dF(t) & \text{si } x \leq \tau_0 \\ 1 & \text{si } x > \tau_0 \end{cases}$$

et en déduire la probabilité qu’une personne quitte le système avant la fin de son service.

Chapitre 7

Fiabilité

7.1 Introduction

La théorie de la fiabilité a pour objectif d'étudier l'aptitude de dispositifs techniques (machines, équipements,...), à accomplir une fonction requise, dans des conditions données, durant un temps donné. Actuellement, c'est une discipline à part entière. Prévoir la fiabilité d'un système est essentielle pour des problèmes de sécurité (systèmes de freinage, systèmes nucléaires, systèmes informatiques...). La quasi-impossibilité de réparer certains matériels (satellites), les problèmes économiques (coûts des défaillances, gestion du personnel de maintenance, maintenance des stocks des pièces de rechange...) rendent nécessaire la connaissance de la fiabilité des systèmes utilisés.

Les défaillances se produisant généralement de façon aléatoire, il est logique de faire appel au calcul des probabilités pour étudier des problèmes de fiabilité. Ainsi, nous définissons la fiabilité d'un dispositif comme étant sa probabilité de fonctionner correctement pendant une durée donnée, ou, ce qui revient au même, la probabilité qu'aucune défaillance ne se produise pendant cette durée.

7.1.1 Définitions

La variable aléatoire T , généralement absolument continue, représente la durée de vie d'un dispositif (ou la durée de bon fonctionnement jusqu'à sa première panne).

- Fiabilité à l'instant t : $R(t) = P([T > t])$.
- Fonction de répartition de T : $F(t) = P([T \leq t])$.
- Densité de défaillance : f , densité de la loi de T .
- Taux de défaillance instantané : $\lambda(t) = \lim_{h \rightarrow 0} \frac{P([T > t] \cap [t < T < t + h])}{h}$.

Propriétés : L'une de ces données suffit à caractériser la loi de T .

Preuve : On a $R(t) = 1 - F(t)$, $F(t) = \int_0^t f(u)du$, $R(t) = \int_t^{+\infty} f(u)du$ et $f(t) = F'(t) = -R'(t)$.

De plus, $\lambda(t) = \frac{F'(t)}{1 - F(t)} = -\frac{R'(t)}{R(t)}$ et $R(t) = \exp\left[-\int_0^t \lambda(u) du\right]$.

□

7.1.2 Loix utilisées

- Loi exponentielle $\mathcal{E}(\lambda)$: $f(t) = \lambda e^{-\lambda t} \mathbb{I}_{]0, +\infty[}(t)$, $\tau = \mathbb{E}(T) = \frac{1}{\lambda}$, $\lambda(t) = \lambda$ (taux de défaillance constant, caractérise les dispositifs sans usure).
- Loi de Weibull : $f(t) = \lambda \beta (t - t_0)^{\beta-1} e^{-\lambda(t-t_0)^\beta}$, $\mathbb{E}(T) = \Gamma(1 + \beta^{-1})(\lambda(t - t_0))^{-\beta^{-1}}$.
- Loi gamma $\gamma(\lambda, a)$: $f(t) = \frac{\lambda^a}{\Gamma(a)} t^{a-1} e^{-\lambda t} \mathbb{I}_{]0, +\infty[}(t)$, $\mathbb{E}(T) = \frac{a}{\lambda}$.

- Loi normale $\mathcal{N}(m, \sigma^2)$: $f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}$, $\mathbb{E}(T) = m$.
- Loi log-normale : $f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - m)^2}{2\sigma^2}} \mathbb{I}_{]0, +\infty[}(t)$, $\mathbb{E}(T) = e^{m + \frac{\sigma^2}{2}}$.

7.2 Systèmes non réparables

7.2.1 Généralités

On appelle système tout assemblage de composants, dont on suppose en général (mais pas toujours !) que les pannes se produisent indépendamment les unes des autres.

Un système non réparable est un système pour lequel aucune réparation de composants défectueux n'est envisageable.

À l'exception des systèmes en série, les systèmes ont généralement des structures redondantes : un ou plusieurs composants peuvent tomber en panne sans que le système ne cesse de fonctionner. En renforçant la redondance du système, on augmente sa fiabilité. Il existe 2 types de redondance :

- redondance active (réserve chaude) : tous les composants fonctionnent en même temps.
- redondance passive (réserve froide) : il existe des composants en attente, qui ne peuvent pas tomber en panne tant qu'ils ne sont pas mis en marche.

7.2.2 Systèmes sans redondance

Un système de n éléments en série ne fonctionne que si les n éléments fonctionnent.

$$T = \min(T_1, \dots, T_n) \quad \text{et} \quad R(t) = \prod_{i=1}^n R_i(t).$$

Sa fiabilité est plus faible que celle du composant le moins fiable.

7.2.3 Systèmes avec redondance

- Système en parallèle : fonctionne si au moins l'un de ses n éléments fonctionne.

$$T = \max(T_1, \dots, T_n) \quad \text{et} \quad F(t) = \prod_{i=1}^n F_i(t).$$

Sa fiabilité est supérieure à celle du composant le plus fiable.

- Système k -de- n : fonctionne si au moins k de ses n éléments fonctionnent.
- Système mixte : association de systèmes en série et en parallèle, ou d'autres systèmes à réserve chaude.
- Système à redondance passive (système à commutation) : un seul élément est en service à la fois. Lorsqu'il tombe en panne, il est immédiatement remplacé.

$$T = T_1 + \dots + T_n.$$

7.3 Systèmes réparables

7.3.1 Introduction

Dans le but d'augmenter la fiabilité d'un système à redondance, on peut envisager de réparer les composants qui tombent en panne. On admettra généralement qu'un composant réparé se comporte comme un composant neuf. Pour résoudre ce type de problème, il est nécessaire de connaître, non seulement la loi de la durée de bon fonctionnement, mais en plus, celle de la durée de réparation.

Lorsque le nombre de réparateurs s est inférieur au nombre de composants du dispositif, une file d'attente de composants en panne peut se former. Il s'agit d'un système fermé d'attente (nombre de clients limité au n composants du dispositif), dont le taux d'entrée varie en fonction de l'état dans lequel le système se trouve.

Pour un système réparable, il est nécessaire d'introduire une autre notion probabiliste : celle de disponibilité : la disponibilité à l'instant t est la probabilité $D(t)$ que le dispositif fonctionne à l'instant t . La différence avec la fiabilité est que, pour la disponibilité, il a pu être en panne avant l'instant t , mais à l'instant t , il fonctionne.

Remarque : $D(t) \geq R(t)$ et, dans le cas d'un système non réparable, $D(t) = R(t)$.

7.3.2 Méthode des processus stochastiques

Il est nécessaire de définir avec précision tous les états possibles dans lesquels le système peut se trouver. On étudie alors le processus $(X_t)_{t \geq 0}$, où X_t est l'état dans lequel le processus se trouve à l'instant t . Un certain nombre de ces états (sous-ensemble E_F de E) correspondent au fonctionnement du dispositif, les autres (sous-ensemble E_D de E) correspondent au cas où le dispositif est défectueux (ne peut pas fonctionner).

On a alors $D(t) = \sum_{k \in E_F} \pi_k(t)$.

Pour calculer la fiabilité, on considère le système non réparable correspondant en rendant les états de E_D absorbants (taux de sortie de ces états rendus nuls).

La résolution de tels problèmes s'effectue comme dans les chapitres précédents. Ainsi :

→ on commence par déterminer le graphe des taux de transition ;

→ on écrit les équations de Kolmogorov : en chaque état k , la variation de flux est égale à la différence (flux entrant - flux sortant), c'est-à-dire :

$$\pi'_k(t) = \sum_{i \neq k} a_{i,k} \pi_i(t) - \sum_{j \neq k} a_{k,j} \pi_k(t)$$

($a_{i,j}$ représentant le taux de transition de i vers j) ;

→ Le système différentiel obtenu est généralement difficile à résoudre. Pourtant, même si, pour la disponibilité, on peut parfois se contenter du régime stationnaire, il est le plus souvent

nécessaire de connaître la fiabilité ou la disponibilité à un instant t donné. C'est pourquoi, on cherche généralement à résoudre le système à l'aide des transformées de Laplace.

En effet, étant donné que $L(\pi'_k)(p) = pL(\pi_k)(p) - \pi_k(0)$, le système différentiel, se transforme en système linéaire, beaucoup plus simple à résoudre. On obtient les $L(\pi_k)(p)$ sous forme de fractions rationnelles de p , que l'on décompose en éléments simples, afin de "remonter" à $\pi_k(t)$, grâce à $\frac{1}{p - \alpha} = L(t \mapsto e^{\alpha t})(p)$ notamment.

La "technique" évoquée ici est développée dans les exercices correspondant à ce chapitre.
